INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

SURVEILLANCE VIBRATIONS SHOCKS NOISE

Vibes in action

8-10 JULY 2019

INSA LYON

Proceedings

BOOK OF PROCEEDINGS

SURVISHNO 2019

A joint organisation of the conferences Surveillance, Vibration Shock and Noise, and Experimental Vibration Analysis

General Chair : J. Antoni, D. Rémond

Table of contents

Monday

Angular approaches

Cyclo-nonstationary analysis for bearing fault identification based on instantaneous angular speed estimation

Edgar F. Sierra-Alonso[1,2] Jerome Antoni[1] German Castellanos-Dominguez[2]
[1]Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon,
LVA EA677, F-69621, Villeurbanne, France
edgar-felipe.sierra-alonso@insa-lyon.fr
[2]Signal Processing and Recognition Group, Universidad Nacional de Colombia,
Km 9 way to magdalena, Manizales Caldas, Colombia

Abstract

Rolling Element Bearings (REB) are present in most of the rotating machines being the ones in charge of supporting the charge of the shaft, being the cosntant charge of the shaft a reason to make the REB prone to fail. Failures under real conditions such as variable speed/load are a subject of interest in the state of the art in digital signal processing of vibration signals, lastly, defined as a cyclo-nonstationary process due to the intrinsic cyclic behaviour of the REB and the nonstatinarity introduced by the variations of the Instantaneous Angular Speed (IAS). The most direct approach to deal with a REB failure under time-varying IAS, is to compensate the IAS transforming the vibration signal to the angular domain, then highlight the cyclo-stationary part of the signal in the angular domain. To obtain the IAS, the direct approach is to measure the IAS via an encoder to obtain the so-called tachometer signal, to place an encoder usually requires a modification of the machine. In cases where it is not possible, the IAS could be extracted directly from the vibration signal. To extract the IAS from a vibration signal is a hard task due to to the low Signal to Noise Ratio (SNR); consequently, a short-time approach methodology robust to noise for IAS estiamtion termed, Short Time Non-Linear Least Squares (STNLS) estimation is proposed. However, even with the IAS to identify the failure requires an additional step that is to highlight the impulsive behaviour, several techniques in the literature makes use directly or indirectly of the Spectral Kurtosis (SK) to highlight impulsive behaviour. The SK is designed to work under small variations of the IAS, even when the variation on the IAS could be compensated through the transformation to the angular domain; there are components angle-variant like the transfer function, that could mask the impulsive components. Thus, a short-angle method based on the SK named Short Time/Angle Spectral Kurtosis (STSK) is introduced. The STSK method is compared with the traditional approach outperforming in both numerical an a challenging case study of an aircraft engine. Similarly, the STNLS is tested on a numerical database for robustness to noise and in the real signal showing a Mean Square Error bellow 5×10^{-3}.

1 Introduction

The Instantaneous Angular Speed (IAS) alongside the acceleration data is one of the most important parameters to measure in condition monitoring. In our case study, to identify a rolling element bearing (REB) failure under variable speed is mandatory to deal with highly non-stationary conditions, mainly caused by the variation of the speed, that modifies the underlying cyclo-stationary process. Due to for REB failures the failure frequency is a function of the angular speed, it is impossible to identify a failure without this parameter [1]; consequently, the IAS is one of the most important parameters to measure. However, the speed measurement is sensitive to disturbances like loss of samples or artifacts[2]. Alternatively, the IAS can be extracted directly from the vibration signal, where this is the most challenging situation. Due to the multi-component nature of the signal, where different families of harmonics may coexist, alongside with the interaction between the orders and the structural resonances of the machine, and the low Signal to Noise Ratio (SNR) where noise comprises any component in the signal which is not of direct interest for the analysis [3]. The importance of the IAS measurement/extraction is of the continuous interest in the literature for REB fault detection under variable IAS; the subject was recently addressed in [4] where is concluded that the uniform angular re-sampling using

the IAS profile is the most common pre-process to identify A REB failure under variable speed. As it is expected to deal with low SNRs, the proposed approach must be robust. For such a reason, it is proposed the use of a Short-Time Non-Linear Least Squares (STNLS) method to estimate the IAS, assuming that a signal is stationary in a short-time segment.

A mandatory pre-process when dealing with vibration signals under variable speed is to transform the signal into the angular domain. The signal is transform throughout the uniform angular re-sampling a technique that transform a signal vibration signal in order to have the same amount of samples per rotation [5]. Nevertheless, as stated in [3] a vibration signal is affected by noise, where the noise is understood as any component in the signal which is not of direct interest for the analysis. Such noise is not entirely removed by any method even by the Spectral Kurtosis (SK). The SK is a low computational and effective tool to highlight impulsive behaviour in a seemly constant speed scenario, like in the case of a cyclo-stationary process [6]. Besides the SK, the traditional approach is to apply the envelope analysis to an angular re-sampled vibration signal [1]; however, the traditional approach is effective only in the case of small fluctuations[7]. Therefore, there is a need to develop more robust, automatic algorithms, suitable for different operating conditions, which leads to the relaxation of the constraint of the constant speed [8]. Recently, [9] proposed an extension of the cyclic spectral correlation for a time-varying speed scenario. However, it is assumed that time-dependent components are independent of the operating speed, which may be acceptable for modest speed variations; thus, its compensation constitutes an emerging field of investigation.

For such a reason, to highlight a REB failure under variable speed under highly non-stationary conditions. It is proposed in the present work a parametric methodology named STNLS for IAS estimation short-time based, and after the angular re-sampling by means of the IAS. It is proposed a REB highlight method that makes use of a Short-Time/Angle frequency 2D filter based on the Spectral Kurtosis (STSK), given that a signal it is expected to be stationary regardless the domain (time or angle) if a window small enough is considered. The robustness of the STNLS method is tested in a simulated signal contaminated with different levels of two different types of noise (pink and white). Finally, the STNLS and STSK are successfully applied in a case study of an aircraft engine publicly available in[4].

2 Theoretical background

This section intrudes the theory about the IAS estimation and the REB failure detection. First, a multi-harmonic model is introduced, and grounded in that model it is proposed the STNLS IAS estimation procedure. Likewise, a model for a REB failure as a superposition of impulses under constant IAS is introduced to numerically prove the proposed STSK filter to highlight a cyclo-stationary process. To do so, the influence of the time-varying IAS on the signal with respect to the traditional model, can be seen as a change of variable, and its compensation is the traditional Computed Order Tracking (COT), with that model in mind it is introduced a Short-Time/Angle Spectral Kurtosis (STSK) filter.

2.1 IAS estimation

The IAS of a given shaft is defined as the fundamental frequency of a multi-harmonic sum measured in a time interval $\mathbf{T} = [t_1, t_2] \subset \mathbb{R}^+$ with a duration $T = t_2 - t_1$. Let us define $\{f_{i,k}(t)\}_{i,k=1}^{I,K}$ such that $f_{i,k}(t) \in \mathbb{R}^+$ is the set of instantaneous frequencies of interest[1], where $\{i,k\}$ denotes the i-th fundamental frequency of the k-th reference shaft. In general a vibration signal $x(t) \in \mathbb{C}$ for the IAS extraction task is modelled as:

$$x(t) = \sum_{i,k=1}^{I,K} a_{i,k}(t) e^{j\phi_{i,k}(t)} + \eta(t) \tag{1}$$

where $a_{i,k}(t) \in \mathbb{C}$ is the time-varying amplitude for the i-th harmonic and the k-th harmonic family, the Instantaneous Angular Displacement[2] $\phi_{i,k}(t) \in \mathbb{R}^+$ is defined as:

[1] the Instantaneous Angular Speed is an Instantaneous Frequency $f_{i,k}(t)$ for which the index $i = 1$, and it is in rad.

[2] the angular displacement is related to the physical phenomenon of interest but in Signal Processing in general it is the phase of the signal.

$$\phi_{i,k}(t) = \int_{t\in\mathbf{T}} f_{i,k}(t)dt \tag{2}$$

$$f_{i,k}(t) = if_{1,k}(t) \tag{3}$$

and $\eta(t) \sim \mathcal{N}(\mu, \sigma^2)$ is stationary Additive White Gaussian Noise (AWGN) with $\mu = 0$. The reader should notice that in practice the noise $\eta(t)$ comprises any component that is not of interest, but as the AWGN is the worst and most common noise that can be found on a signal, the noise in the Eq. (1) is modelled as such. Consequently, the stochastic signal $x(t)$ has two parts: the sum that is the deterministic part, and the stochastic part modelled as AWGN. Adding as a restriction that the signal has a dominant multi-harmonic family, the vibration signal is modelled as a signal from a machine with only one shaft, consequently, the Eq. (1) is rewritten as:

$$\hat{x}(t) = \sum_{i=1}^{I} a_i(t)e^{j\phi_i(t)} + \eta(t) \tag{4}$$

where $\phi_i(t) = i\phi_{1,1}(t)$ and $a_i(t) = a_{i,1}(t)$. Note that in general $\hat{x}(t) \neq x(t)$, using $\hat{x}(t)$ it can be estimated $\hat{f}_1(t)$ with a Non-linear Least Squares estimation procedure. As in theory $x(t)$ is a *quasi-stationary signal* of real domain and complex range[3], it is assumed that in a short-time window the stochastic signal is stationary, for such a reason a Short-Time Non-Linear Least Squares (STNLS) estimation procedure is studied in the present work.

2.1.1 Short-time Non-Linear Least Squares IAS estimation

As the signal $x(t)$ is said to be quasi-stationary on a short-time interval. Thus, the proposed STNLS procedure is the traditional Non-Linear Least Squares estimation, but, applied to all the obtained short-time segments through a sliding window [4]. As $x(t)$ is quasi-stationary if the interval in which $t \in [t_c - t_T/2, t_c + t_T/2] \subset \mathbb{R}^+$ is small enough, for convenience let us define a short-time signal in discrete time notation. A signal in discrete time is defined as $x[n] = x(n\Delta t)$ for a given sampling time $\Delta t \in \mathbb{R}^+$ and a time index $n \in \mathbb{N}$. Therefore, the n-th short-time segment is $\mathbf{x}[n] = [x[n], x[n+1], \cdots, x[n+L-1]]^T$, where L is the length of the segment, i.e, $\mathbf{x}[n] \in \mathbb{R}^{L\times 1}$. Please note that in this case a rectangular window function is used as recommended in [10]. Eq. (4) can be rewritten using a matrix notation, as follows:

$$\mathbf{x}[n] = \mathbf{Z}(\phi_1[n])\mathbf{a}[n] + \eta[n] \tag{5}$$

$$\mathbf{Z}[n] = [\mathbf{e}^{j\phi_1[n]}, \mathbf{e}^{j\phi_1[n+1]}, ..., \mathbf{e}^{j\phi_1[n+L-1]}]^T \tag{6}$$

$$\mathbf{e}^{j\phi_1[n]} = [e^{j1\phi_1[n]}, e^{j2\phi_1[n]}, ..., e^{jI\phi_1[n]}] \tag{7}$$

$$\mathbf{a}[n] = [a_1, a_2, \cdots, a_I]^T \tag{8}$$

$$\mathbf{a}[n] = [A_1 e^{j\psi_1}, A_2 e^{j\psi_2}, \cdots, A_I e^{j\psi_I}]^T \tag{9}$$

where $\mathbf{Z}(\phi_1[n]) \in \mathbb{C}^{L\times I}$ or shorted $\mathbf{Z}[n]$ is a Vandermonde matrix that has the non-linear exponential complex base, the amplitudes are $\mathbf{a}[n] \in \mathbb{C}^{I\times 1}$, being $\{A_i\}_{i=1}^I$ the set of real value magnitudes, $\{\psi_i\}_{i=1}^I$ the set of real value initial phases, and $\eta[n] \in \mathbb{C}^{L\times 1}$ is stationary AWGN. The STNLS method alongside the model described on Eqs. (5) to (9) are used to estimate the IAS, please note that as stated in [11], under the assumption of AWGN, the NLS method is equivalent to the maximum likelihood method. Consequently, the NLS is a maximum likelihood estimator for the considered model for the vibration signal. The estimation of the IAS is as follows:

$$\hat{\phi}_1[n] = \underset{\{\phi_{1k}[n]\}_{k=1}^K}{\arg\min} \|\mathbf{x}[n] - \mathbf{Z}(\phi_{1k}[n])\mathbf{a}[n]\|^2 \tag{10}$$

[3] the signal is complex in range for convenience to make use of the analytic signal in practice

[4] the simplest window function will be used on this work, i.e, the rectangular function $\Pi\left(\frac{t-t_c}{t_T}\right)$, a function centred at t_c with duration t_T, and a height of 1.

Eq. (10) has as argument $\phi_{1k}[n]$, where k here is the index of the set for the optimization problem not the k-th harmonic family. As baseline it will be assumed the IAS in a short-time segment to be constant, i.e. $\phi_1[n] = \beta_2[n]$, as in [11]. But this estimation has a main limitation, the compromise between the length of the segment L, the sampling period Δt, and the IAS bounded interval $f_1(t) \in [f_1, f_2]$. Attempting to relax the aforementioned constraints and to deal with a low SNR, it is proposed a linear model for $\phi_1[n] = \sum_{n \in N} f_1[n]\Delta t$, given the segment $\mathbf{x}[n]$:

$$\phi_1[n] = \beta_1 n^2/2 + \beta_2 n \tag{11}$$
$$= \beta_1[n] + \beta_2[n] \tag{12}$$

where β_1 is the quadratic term and β_2 an initial phase for the given segment, as a consequence the Eq. (10) is rewritten as:

$$\hat{\phi}_1[n] = \underset{\{\beta_{1,k_1}[n], \beta_{2,k_2}[n]\}_{k_1,k_2=1}^{K_1,K_2}}{\arg\min} \|\mathbf{x}[n] - \mathbf{Z}(\beta_1[n], \beta_2[n])\mathbf{a}[n]\|^2 \tag{13}$$

In order to simplify the notation the dependence on the IAS and the n-th time segment are removed from the matrix $\mathbf{Z}$, for the two cases two matrices are introduced, for the constant case $\mathbf{Z}_1[n] = \mathbf{Z}(0, \beta_2[n])$, and the linear approximation case $\mathbf{Z}_2[n] = \mathbf{Z}(\beta_1[n], \beta_2[n])$. To solve the optimization problem in the Eqs. (10) and (13) it is necessary to remove the dependence of the amplitudes $\mathbf{a}[n]$, to do so, a projection is done through its least squares estimation:

$$\hat{\mathbf{a}}_i[n] = (\mathbf{Z}_i[n]^H \mathbf{Z}_i[n])^{-1} \mathbf{Z}_i[n]^H \mathbf{x}[n] | i = 1, 2 \tag{14}$$

after removing the amplitude dependence algorithmic strategies are introduced to solve the optimization problem in Eqs. (10) and (13), it has to be noted that as it is a Non-Linear problem the immediate solution it is an exhaustive search building the set of possible solutions $\{\beta_{1,k_1}[n], \beta_{2,k_2}[n]\}_{k_1,k_2=1}^{K_1,K_2}$ and to choose the $\arg\min$. The two models for $\phi_1[n]$ and some further constrains to have faster convergence will be examined in the Section 3.

2.2 REB highlight: a short-time/angle SK filtering approach

Before introducing the proposed filtering scheme, let us model a REB fault. A REB Fault at constant IAS can be modelled as a superposition of periodic impulses, where T is the period of the fault frequency, i.e., the average time between impacts, the model is proposed in [12] as follows:

$$x(t) = \sum_{i=1}^{I} A_i s(t - iT - \tau_i) + \eta(t) \tag{15}$$

where $\{A_i\}_{i=1}^{I}$ is the set of real value amplitudes, $\{\eta(t)\}_{t \in \mathbb{R}}$ is AWGN as in the previous model (see Eq. (1)), the difference lies in the function $s(t)$ that models an impact and $\{\tau_i\}_{i=1}^{I}$ is a random process to introduce small variations in the period T, in the case of *time-varying IAS* the period is no longer constant is time-varying consequently, a time-dependent function $\bar{T}(\phi_1(t)) = \phi_1(t) - iT - \tau_i$ should be introduced to explain the influence of the IAS in the REB failure vibration signal. Consequently, the variation in the IAS can be seen as a Parametrization of $x(t)$. Consequently, a REB time-varying failure signal can be written as:

$$y(t) = \sum_{i=1}^{I} A_i s[\bar{T}(\phi_1(t))] + \eta(t) \tag{16}$$

where the function $\bar{T}(\phi_1(t))$ controls the time-varying fault period making the signal cyclo-nonstationary. In contrast, a typical cyclo-stationary signal is the model in Eq. (15). Therefore, it is necessary to transform the signal $y(t)$ to be as close as possible to $x(t)$, i.e, a vibration signal at constant speed. In practice the operation to arrive from $y(t)$ to $x(t)$ is performed through uniform angular re-sampling named also computed order tracking (COT) a technique described in [5]. Traditionally, in the literature of COT it is assumed to have $y(t)$, a signal

with a time-varying IAS, the signal in angular domain is $y(\Theta)$ which is similar to $x(t)$, i.e, a version of $y(t)$ where the IAS is constant, more precisely is equal to 1. Due to a vibration signal with an IAS of 1Hz it is equivalent to have the vibration signal in rotations, i.e, in angular domain.

An important identity of the angular domain is that in the case of constant speed it is trivial to see that if the speed is the unity, i.e, $\phi_1(t) = t$ then $y(t) = x(t)$, so considering any constant speed scenario in practice the angular re-sampling is not necessary, even more, the influence of the speed is nothing more than a scale in the time and/or frequency axis, considering a simple example be $y(\gamma)$ a signal in angular domain for $\phi_1(t) = ct$ the angular domain variable is $\gamma = ct$, the operation is inversely proportional (the orders $\Gamma = \frac{1}{c}f$) in frequency domain. After the COT the time axis is termed angular domain, and its frequency domain order domain.

In general the model is only valid after filtering the impulsive behaviour of the signal due to a REB failure under variable IAS is present the vibration signal in practice is a superposition of the two used models Eqs. (4) and (16), for such a reason if a REB failure scenario is considered is mandatory to consider a filtering step to highlight the failure. The most common filter is based on the SK, consequently, a brief description of the SK will be given in the next subsection.

2.2.1 Short Time/Angle SK

To deal with a signal with a time-varying IAS the angular domain is a mandatory step, consequently, the most general approach is to first transform the signal to the angular domain, for such a reason, let us assume that the signal $x[n]$ is in the angular domain regardless if the speed is constant or time-varying. Let us consider a L length segment $\mathbf{x}[n]$, being this the n-th segments of a signal $x[n]$ with more than L point, this notation of an n-th segment is mandatory to make the estimation of the SK. Usually the consideration of segments is given per se, due to it is necessary to compute the estimation of the spectral order monments for the SK, but in the present work the distinction is made due to it is assumed that the angle-varying structural response will produce an Angle-varying SK. Consequently, the SK will be computed per n-th segment, consequently an estimator based on the Short Time Fourier Transform (STFT) of the n-th segment for the SK (a detailed definition can be found on [6]), given a window function $w(n)$, and a time lag m is introduced. Let us define the STFT of $x[n]$ as:

$$\mathrm{STFT}_x[m, f_n] := \sum_{n=0}^{L} x[n]w[n-m]e^{-j2\pi f_n} \tag{17}$$

Based on this definition, the SK can be estimated through the second and fourth order empirical spectral moment of $\mathrm{STFT}_x[m, f_n]$ defined as:

$$\hat{S}_{2x}[f_n] := \left\langle |\mathrm{STFT}_x[m, f_n]|^2 \right\rangle_m \tag{18}$$

$$\hat{S}_{4x}[f_n] := \left\langle |\mathrm{STFT}_x[m, f_n]|^4 \right\rangle_m \tag{19}$$

being $\langle\cdot\rangle_m$ the average operator through the index m. The SK is finally defined utilizing its second and fourth order empirical moment as:

$$\mathrm{SK}[f_n] = \frac{\hat{S}_{4x}[f_n]}{\hat{S}^2_{2x}[f_n]} - 2 \tag{20}$$

Please note that $\mathrm{SK}[f_n]$ is the SK for the n-th segment of length L, for such a reason let us note the SK of the n-th segment as $\mathrm{SK}_n[f_m]$. As the model in Eq. (16) is a superposition of impulses it is necessary to use second-order moments to isolate the fault frequency, but those second-order moments are conditioned to a constant speed and stationary vibration signals, for such a reason to work with the signal in the angular domain is mandatory. The most common descriptor when dealing with REB failures is the spectral correlation, a bi-spectrum which average in frequency is equal to the envelope spectrum, a brief description is made in the following subsection for a more detailed explanation view [12].

2.2.2 Envelope analysis & spectral correlation

The relationship between the traditional envelope spectrum and the spectral correlation can be summarized as the equivalence between the spectrum of cyclic frequencies $M_x(\alpha)$, i.e., the spectrum obtained integrating the spectral correlation function $S_z(\alpha, f)$ through the frequency axis f, and the envelope spectrum

$\mathrm{F}_{t\to f}\{\mathbb{E}\{|z(t)|^2\}\}$, where $z(t)$ is the analytic signal[5] of $x(t)$ being $|z(t)|^2$ the squared envelope of $x(t)$. As the spectral correlation is the double Fourier transform of the auto-correlation function, let us define the auto-correlation function for a zero mean signal $z(t)$ as:

$$R_z(t,\tau) := \mathbb{E}\{z(t+\tau/2)z^*(t-\tau/2)\} \tag{21}$$

then the spectral correlation is defined as:

$$S_z(\alpha,f) = \underset{\tau\to f}{\mathrm{F}}\left\{\underset{t\to\alpha}{\mathrm{F}}\{R_z(t,\tau)\}\right\} \tag{22}$$

The procedure for the proof is straightforward, the main result used is that the signal is dominated by its stochastic part then $R_z(t,0) = \mathbb{E}\{|z(t)|^2\}$. With the definition given in Eq. (22) and its integral across the frequency axis, the relationship between the envelope spectrum and the spectral correlation function is:

$$\begin{aligned} M_z(\alpha) &= \int_{\mathbb{R}} S_z(\alpha,f)df && (23)\\ &= \underset{t\to\alpha}{\mathrm{F}}\{R_z(t,0)\} && (24)\\ &= \underset{t\to\alpha}{\mathrm{F}}\{\mathbb{E}\{|z(t)|^2\}\} && (25)\end{aligned}$$

the previous Eq. (24) is obtained making use of the Fubini's theorem and the definition of spectral correlation (view Eq. (22)). The Eq. (25) comprises the desired relationship, thus, the integral of the spectral correlation through the frequency axis is equal to the traditional envelope spectrum. Nevertheless, the envelope spectrum can be hard to read due to the noise that masks the impulsive behaviour, such behaviour could be highlighted making use of the SK. Yet the SK is designed for stationary processes and a high time-varying IAS will produce a non-stationary signal even in angular domain, consequently, a short-time/angle filtered envelope spectrum estimation procedure is proposed in the next subsection.

2.2.3 Short Time/Angle Filtered Envelope Spectrum: welch's based estimation

Before introducing the estimation of the squared envelope spectrum using the proposed Short Time/Angle SK (STSK), it is mandatory to highlight the impulsive behaviour to be as close as possible to the model given by Eq. (15), so, for a n-th segment of length $L \in \mathbb{N}$ of a real signal in angle domain noted as $\mathbf{x}[n] = [x[n], x[n+1], \cdots, x[n+L-1]]^T$, let us consider a filtered version using the SK as:

$$\mathbf{x}_{\mathrm{SK}}[n] = \underset{f_m\to n}{\mathrm{DFT}^{-1}}\{\mathbf{X}_n[f_m]\mathrm{SK}_n[f_m]\} \tag{26}$$

where $\mathbf{X}_n[f_m] := \mathrm{DFT}_{n\to f_m}\{\mathbf{x}[n]\}$ being $\mathrm{DFT}_{n\to f_m}\{\cdot\}$ the Discrete Fourier transform of a given signal, and $\mathrm{DFT}^{-1}{}_{f_m\to n}\{\cdot\}$ the inverse Fourier transform, it should be pointed out that the Eq. (26) in practice is a filter that makes use of the convolution theorem due to a convolution in time is a product in frequency. The result is a filtered n-th segment where the impulsive behaviour is highlighted, afterwards the envelope spectrum is estimated grounded on the Welch's method of periodograms as follows:

$$\hat{S}_{|\mathbf{z}_{\mathrm{SK}}|^2}[f_m] = \frac{1}{NU}\sum_{n=1}^{N}\left|\underset{n\to f_m}{\mathrm{DFT}}\{|\mathbf{z}_{\mathrm{SK}}[n]|^2\}\right|^2 \tag{27}$$

where $\mathbf{z}_{\mathrm{SK}}[n]$ is the analytical signal, a signal whose modulus is approximately the envelope[6] of $\mathbf{x}[n]$ for a detailed explanation on how good the approximation is given in [13], and $U = \sum_{n=1}^{L}|w[n]|^2$ is the energy of the window. The result is a short-angle filtered approximation of the envelope spectrum noted $\hat{S}_{|\mathbf{z}_{\mathrm{SK}}|^2}[f_m]$. Which it is expected to produce an improvement of the SNR, given that the proposal has an optimal angle-variant SK filter (angle-frequency 2D), using the Welch's method of periodograms in short segments that makes the computation of the filtered envelope spectrum faster with respect to the traditional approaches in the state of the art.

[5] an analytic signal is a signal without negative frequencies where $z(t) := x(t) + j\mathscr{H}\{x(t)\}$ being $\mathscr{H}\{\cdot\}$ the Hilbert transform.

[6] the instantaneous envelope and phase are well defined only for mono-component signal, then is advisable to pre-filter $z_k(t)$ in order to have a mono-component signal.

3 Experimental set-up

Two experiments are proposed, one in a numerical dataset and a case study of an aircraft engine, for both the numerical and the case study, the IAS is estimated employing the proposed STNLS and a fault in REB is highlighted by the 2D short-angle filter STSK. The robustness of the IAS STNLS estimator will be studied under the influence of two types of Noise, Pink and Additive White Gaussian Noise (AWGN). The robustness test is made only in a numerical case, due to, only in a numerical signal, the SNR can be controlled. For the proposed STSK filter, the REB highlighting is compared against two methodologies, a naive approach and the traditional state of the art approach. Regarding the case study, the IAS estimation is compared against the reference signal measured from the tachometer signal, and for the failure in a REB, a visual comparison is made as well as for the numerical case.

3.1 Numerical test: IAS estimation & REB highlight

Usually the models in the state of the art for time-varying IAS does not take into account the effect of the transfer function, consequently, to have a more realistic and challenging signal it is introduced the transfer function $h[n]$ in the model presented in the Eq. (4) as follows:

$$\hat{x}[n] = h[n] \circledast \sum_{i=1}^{I} a_i[n] e^{j\phi_i[n]} + \eta[n] \tag{28}$$

where $\circledast(\cdot,\cdot)$ represents the convolution operator between two functions, an example of the numerical signal with the influence of the transfer function as stated in the Eq. (28) is shown in the Fig. 1, to be even more challenging the signal presents a random discontinuity in all the orders comprised by a piecewise formulation, i.e., the amplitudes $a_i[n] = 0$ for all $i \in I$ and for all n in an interval $[n_1, n_2] \subset [0, L]$, where $n_1 \approx 2.5$fs and $n_2 = n_1 + 0.1$fs. In this case the interval is of 0.1sec, such discontinuity is as misleading for any IAS tracking algorithm as an impulse, like the one produced by a REB failure.

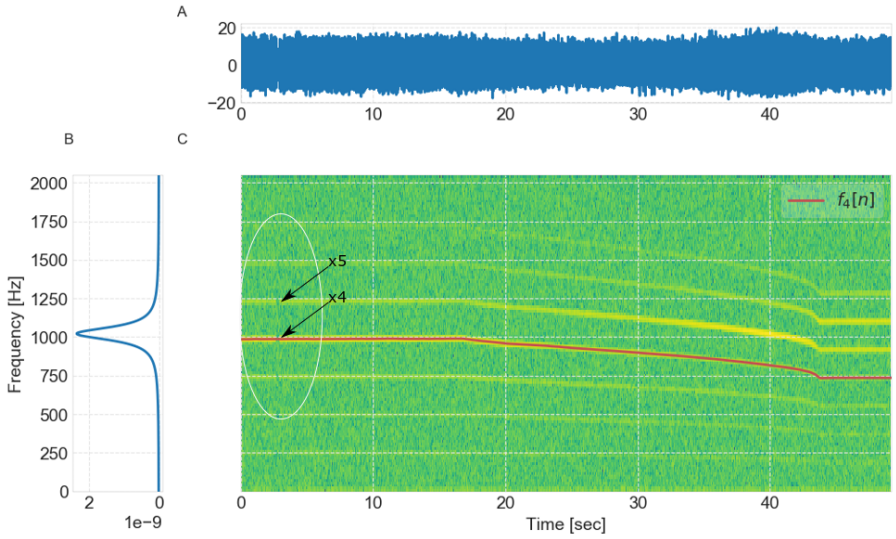


Figure 1: Numerical signal with AWGN of 3dB, a) in time, b) transfer function $H(f)$, c) spectrogram showing a random discontinuity of 0.1 sec reflected on all orders.

The considered sampling frequency fs is of 2^{12} samples per second, and the signal is contaminated with AWGN to accomplish a SNR of 3dB, also in order to be as realistic as possible the IAS $\phi_r[n]$ is from the real case study described in Section 3.2, only that the time indexes are compressed by a factor of 4, i.e., the used IAS $\phi_1[n] = \phi_r[4n]$. The compression is done for computational reasons.

3.1.1 IAS estimation: robustness test

In order to prove the robustness of the proposed IAS estimation methodology the numerical signal in the Eq. (28) will be contaminated with different levels of AWGN and Pink noise[7] with a SNR in the interval $[-10,10]$dB, and as baseline it is considered the IAS estimated with a null quadratic term, i.e., $\beta_1[n]=0$, as measure of estimation error it is used the Mean-Squared-Error (MSE), which is defined as:

$$\text{MSE}\{f_1[n],\hat{f}_1[n]\}=\frac{1}{N}\sum_{n=1}^{N}|f_1[n]-\hat{f}_1[n]|^2 \tag{29}$$

where $\hat{f}_1[n]\in\mathbb{R}^N$ is the estimated IAS, and $f_1[n]$ the real IAS profile for the entire signal of length N, being being $N>L$, and L has as length the amount of samples equivalent to 0.2sec. Recalling the minimisation problem in the Eq. (13) a minimisation algorithm should be used but as there is no an explicit Jacobian, the Nelder-Mead algorithm is selected. The parameters of the algorithm are: an initialization value ($\{\beta_{1,nd}[0],\beta_{2,nd}[0]\}$), the optimization function that will produce an scalar value using the logarithmic version of the Eq. (13) for a faster convergence. The amount of orders I, conditioned by the sampling frequency (all the possible orders are considered), and a tolerance for the convergence of 10^{-6}. The initialisation is made through the minimum of the cost function using a grid for the first segment $n=0$; nevertheless, the problem is non-convex as shown in the Fig. 2, for such a reason the initialisation is of extreme importance due to the functional to optimize could fall into a suboptimal solution.

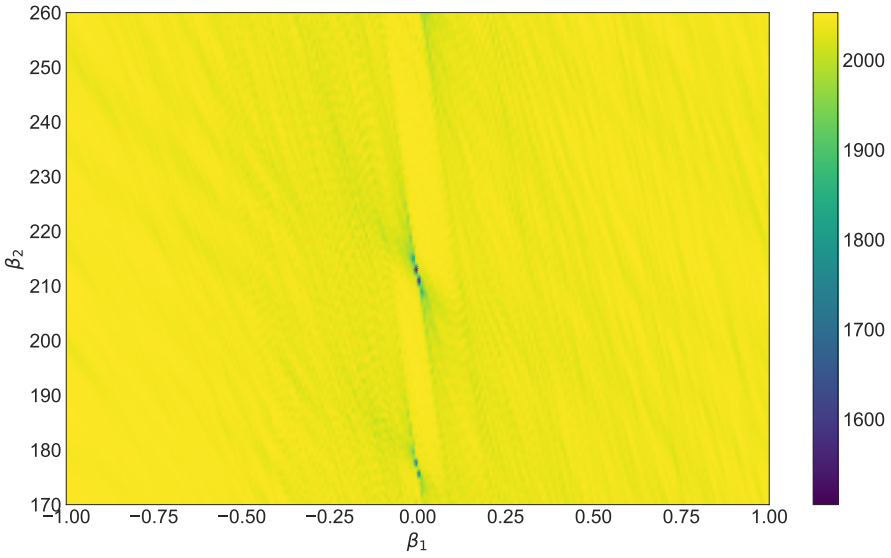


Figure 2: Cost function in the Eq. (13) for a segment of the the exemplary signal in Fig. 1 at 35sec.

As the optimisation is made per segment, it is assumed that the IAS does not variates significantly from the n-th segment with respect to the $n+1$-th segment, so after the initialization using the cost function, the parameters are initialized as $\beta_{i,nd}[n+1]=\hat{\beta}_i[n]$ for $i=1,2$ for $n>0$.

The resulting estimated parameters $\hat{\beta}_1[n]$, and $\hat{\beta}_2[n]$ for the piecewise estimation (for each n-th segment) for the robustness test ($[-10,10]$dB of SNR) are shown in the Figs. 3 and 4 for the AWGN and Pink Noise case respectively, alongside with the reconstruction of the IAS $\hat{f}_1[n]$ by means of the Eqs. (11) and (12) for the entire signal. The reconstruction is made taking into account that it is used an overlap of half of the points $L/2$. It should be noticed that when the quadratic term $\beta_1[n]$ is null the linear term $\beta_2[n]$ is in fact the IAS, being the IAS the derivative of the phase and the derivative of $\beta_2 n=\beta_2$ per each n-the segment. where the n in $\beta_2 n$ could take values from $[n+1,n+L-1]$. As can be seen in Figs. 3 and 4 all the estimations are superposed for different levels of noise where each colour represents a SNR level.

In the Fig. 3 is shown that the AWGN is the worst kind of noise to be expected, due to its presence in the whole spectrogram, meanwhile the estimation in the Fig. 4 contaminated with Pink Noise it is almost perfect due to the noise is decreasing with respect to the frequency, making the harmonic model extremely robust

[7] where the Pink Noise is defined in terms of its Power Spectral Density (PSD) as $\eta_{\text{PN}}[f_m]\approx 1/f_m$.

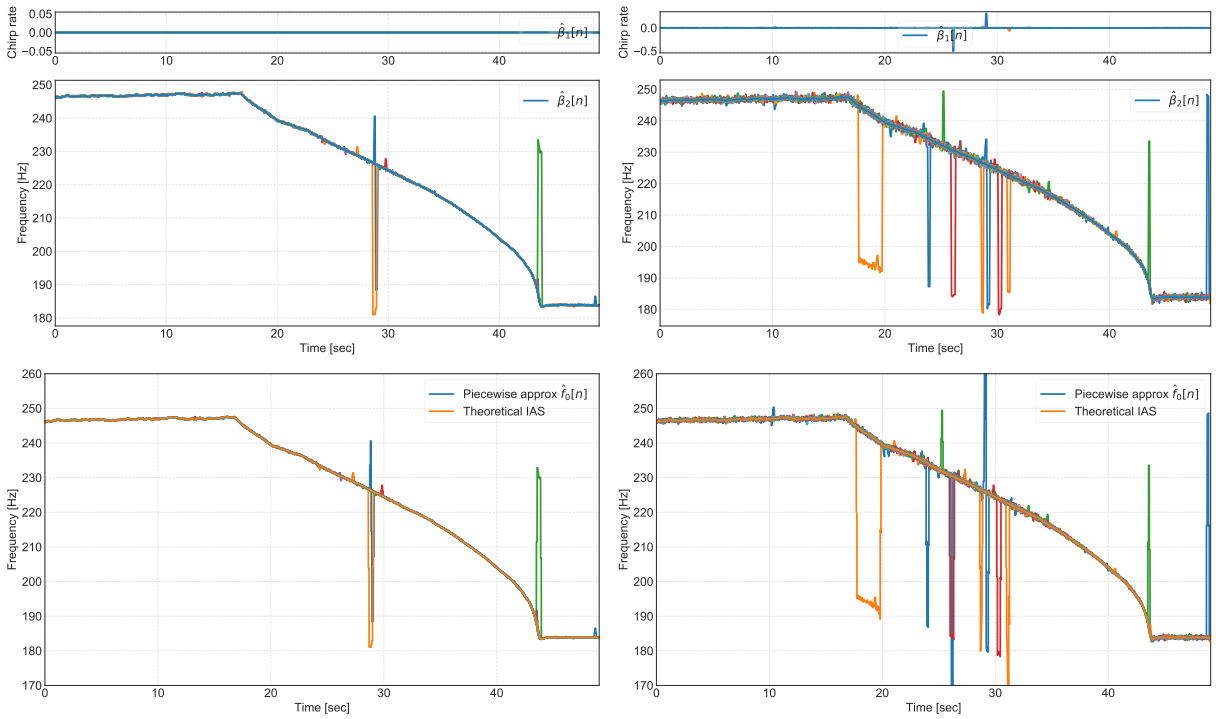


Figure 3: AWGN test, first-column:$\beta_1[n]$, $\beta_2[n]$, and IAS estimated no linear approximation, second-column: $\beta_1[n]$, $\beta_2[n]$ and IAS estimation after linear approximation, for all the considered signal polluted with noise in to achieve a for SNR in $[-10, 10]$dB.

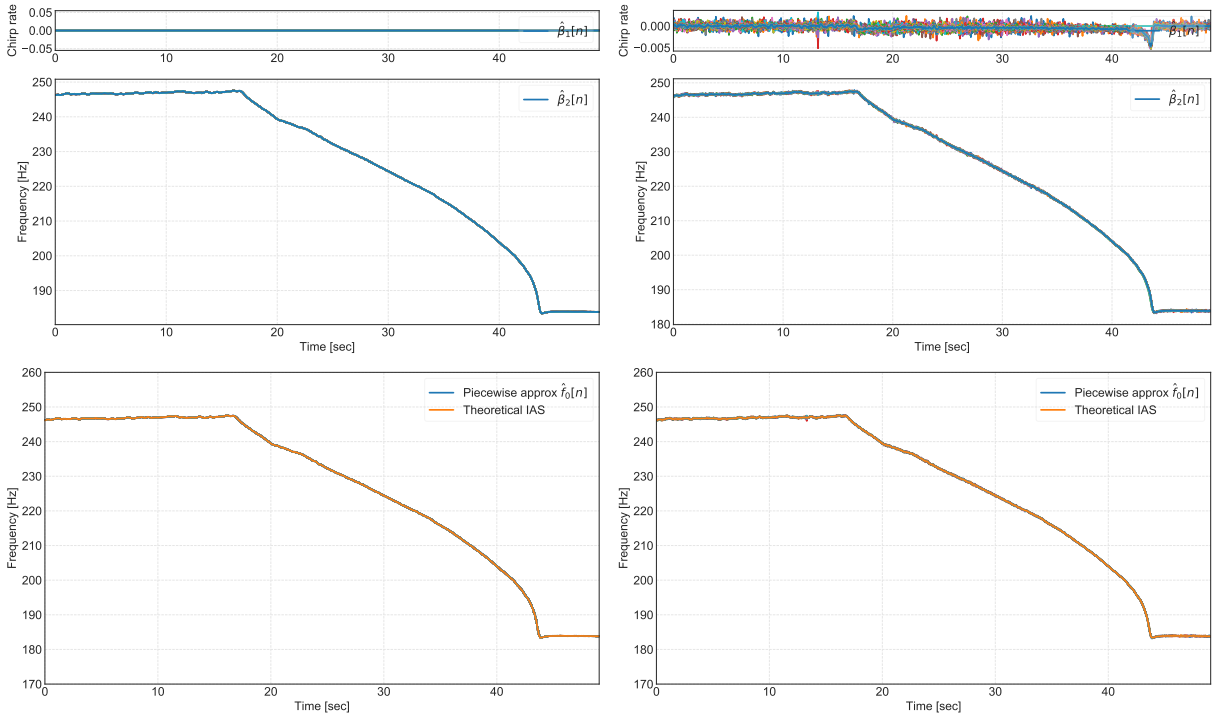


Figure 4: Pink noise test, first-column:$\beta_1[n]$, $\beta_2[n]$, and IAS estimated no linear approximation, second-column: $\beta_1[n]$, $\beta_2[n]$ and IAS estimation after linear approximation, for all the considered signal polluted with noise in to achieve a for SNR in $[-10, 10]$dB. Figure to compare with the Fig. 3

to Pink Noise. The Fig. 3 shows that for the AWGN case are two critical segments the one that comprises the segment with the discontinuity of 0.1sec around 30sec and the segment where the speed has the highest variation at 45sec.

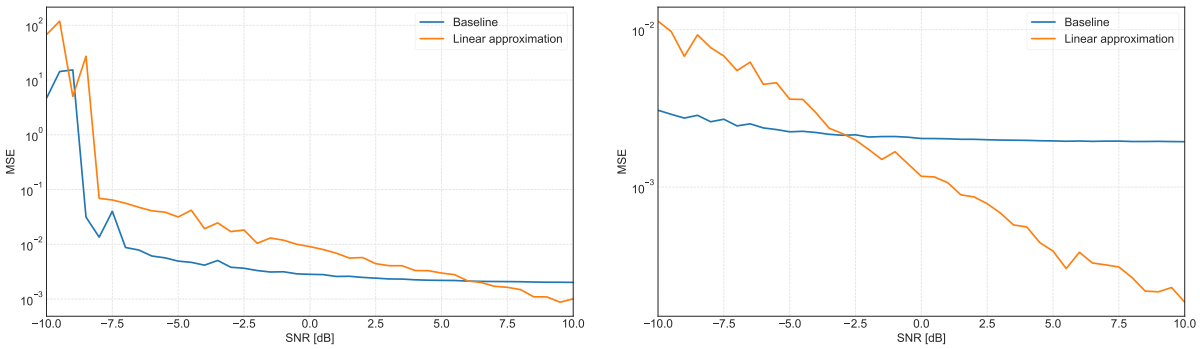


Figure 5: SNR vs MSE tests for an SNR between $[-10, 10]$dB, in the left the AWN scenario, and in the right the Pink-noise scenario, for the baseline and linear IF reconstruction of the IAS $\hat{f}_1[n]$ using Eqs. (11) and (12)

The Fig. 5 shows the MSE for the baseline, and the linear approximation for the two types of considered noise, as expected a decreasing behaviour of the MSE by respect the SNR is shown, for both the baseline and the linear approximation. Nevertheless, the linear approximation of the IAS with low SNR, i.e, less than 6dB for AWGN and -3dB for Pink Noise, is worst with respect to the baseline (quadratic term null). Due to the introduction of the quadratic term (the slope of the IAS) amplifies the errors at low SNR, in contrast with a good SNR improves the estimation. It should be pointed out that there are misleading errors (peaks in Figs. 3 and 4) those errors could be erased with an additional filter, like a median filter, yet it is not part of the scope of the work and taking into account the challenging theoretical signal, the results are encouraging, experimentally proving the robustness of the parametric methods.

3.1.2 REB failure detection

As it is not defined an standard scalar bearing fault indicator, the fault detection is usually made through visual examination for such a reason it will not be considered an SNR test in this subsection. The numerical signal in the Fig. 6 is modelled using the Eq. (16). The signal simulates a Ball Pass Frequency Inner-race (BPFI) fault with a fault frequency (with period T) of 5.875Hz, with a linear IAS profile creasing from 5Hz to 60Hz and contaminated with AWGN to achieve a SNR of 3dB.

The signal in time, the IAS profile, and its spectrum is shown in Fig. 6, as it is expected from the visual examination of the spectrum of the considered bearing fault, a second-order descriptor like the envelope spectrum should be used. Furthermore, as the speed variates the angular domain transform is a mandatory step. The resulting filtered envelope spectrum through the proposed STSK filtering approach are shown in Fig. 7, where the only parameters for the STSK are the size of the windows: for the computation of the standard SK is of 4096 samples and a larger window for the Short-angle SK of 65536 samples given a sampling frequency of 512 samples per revolution.

The Fig. 7 shows a comparison between two baseline methodologies and the STSK proposed approach, initially it is assumed a naive approach that does not take into account that the transfer function $h(t)$ (LTI) becomes angle-variant (Non-LTI), so the vibration signal $x[n]$ is re-sampled to the angular domain and after the state of the art SK filtering and posterior envelope spectrum estimation. With this naive approach the first fault harmonic x1 is almost destroyed. The second approach is to isolate with the SK the impulsive band, i.e, the frequency band with highest SK, assuming that the signal filtered by the SK is close to the model in the Eq. (16). This approach gives the expected decreasing spectrum, but surprisingly the highest energy is located in the second harmonic of the fault frequency x2. Finally, the proposed STSK approach is summarized in the Eq. (27) which is short-time/angle based, as it is short-angle based the transfer function will remain LTI, and the failure is better highlighted due to the SK filter is optimal segment based, i.e, each segment will follow the model Eq. (16), as result the first harmonic of the failure x1 is dominant and there is a decreasing distribution of the energy for the rest of the harmonics, making the failure highly diagnosable visually examining the filter

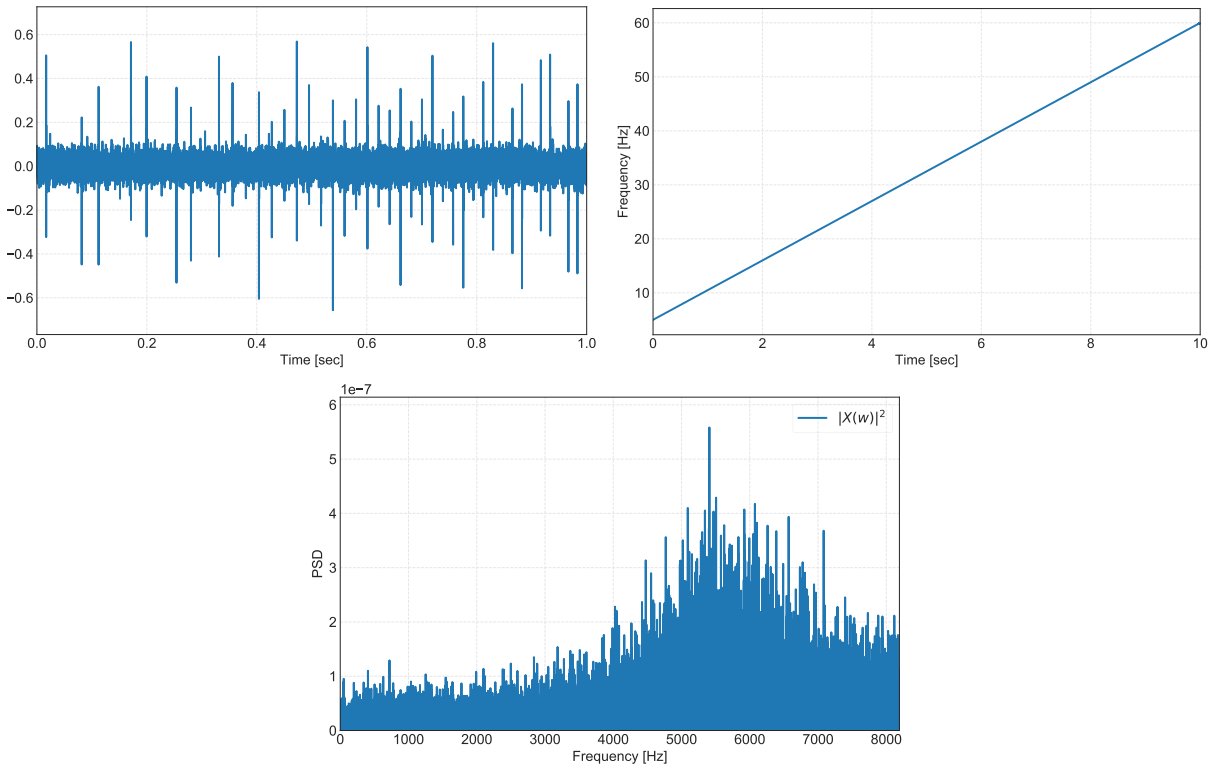


Figure 6: Top-row: left) numerical signal in time with a BPFI at 5.875Hz, right) IAS profile, bottom-row: spectrum of the simulated signal.

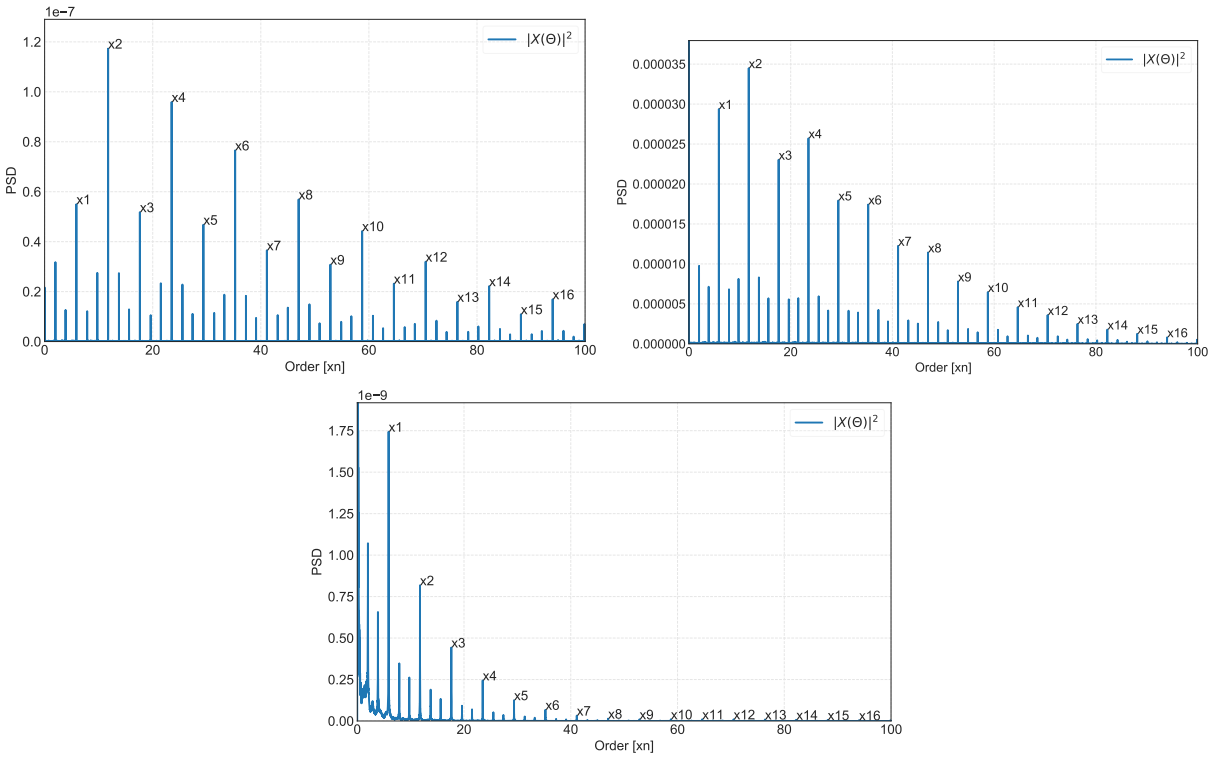


Figure 7: Top-row: baseline methodologies, left) angular domain transform, SK filtering, then envelope spectrum, right) SK filtering, angular domain transform, then envelope spectrum, bottom-row: proposal) angular domain transform, then, envelope spectrum convolved with a time-frequency dependent SK based filter.

envelope spectrum.

3.2 Case study: an aircraft engine

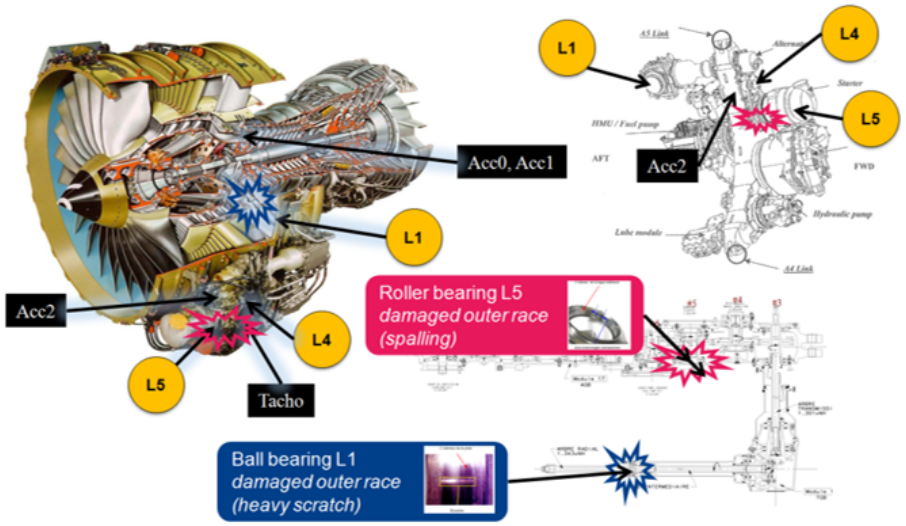


Figure 8: General overview of the engine and the accessory gearbox. Shafts (L1-L5) are identified by labels in amber colour (image taken from [4]).

The data was acquired during a ground test on a civil aircraft engine. The Fig. 8 gives a general overview of the engine with the damaged bearings and the sensors locations. The engine has two main shafts and an accessory gearbox with pieces of equipment such as pumps, filters, alternators, and starter. The accessory gearbox is linked to the high-pressure shaft HP by a radial drive shaft and a horizontal drive shaft. The records have a sampling frequency of 50kHz, it should be noted that the fault it is only in the record ACC2, and it is taken close to L5, for such a reason it is expected to view mainly the fault in outer race of that line, also, as there are several lines, L5 is chosen as reference for the fault frequencies, the table with the fault frequencies referenced to L5 are shown in Table 1. For a more detailed explanation (kinematics) of the dataset and the recorded signals please refer to [4].

	L1(L5)	L4(L5)	L5(L5)
Speed	1,34	0,984	1
Cage	0,55	0,40	0,43
Rolling element	3,46	2,44	3,56
Inner race	7,95	5,87	10,24
Outer race	5,45	3,97	7,76

Table 1: Table of fault frequencies of bearings for the supporting shafts L1, L4, L5, taking as reference the speed of L5.

3.2.1 IAS estimation

The short-time estimation of the IAS has only as critical parameter the window size L and the overlap is $L/2$, the window size is two times the period of the minimum expected frequency of 170Hz, there are have two records to study ACC1 normal condition and ACC2 with a fault. The spectrograms of the records are shown in Fig. 9, in the left column it is shown the signals with the entire frequency spectrum from 0 to 25kHz, and in the right the signals with a down-sampling of 16 times. For the record ACC2, it is not expected to extract accurately the IAS due to, the information from the down-sampled record is almost destroyed by the transfer function of the structure, i.e., a redistribution of the energy to a high-frequency band. A typical phenomenon when dealing

with bearing failures. For such a reason it makes more sense to extract the highest energy harmonic, that is the one related with the amount of teeth for gear located at L5, i.e., the harmonic 62 of the IAS.

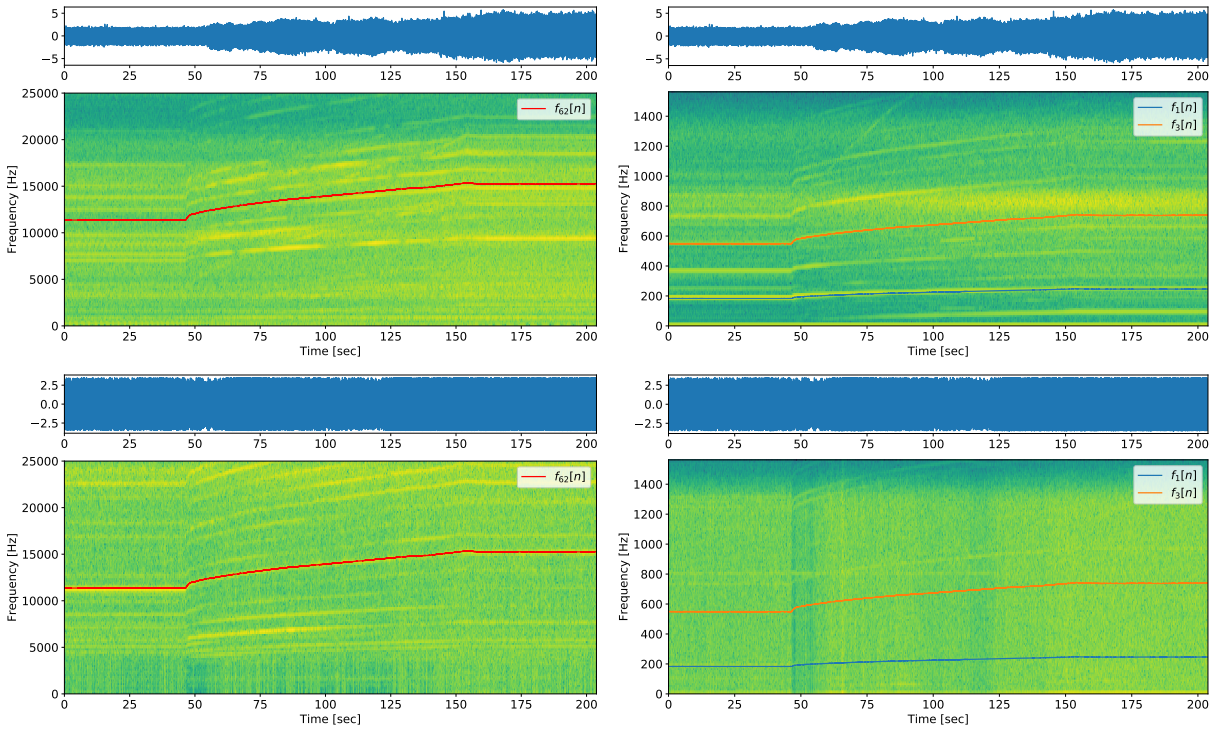


Figure 9: Top-row: spectrograms ACC1, left, full frequency range $[0,25]$kHz, right, down-sampled to $[0,1.56]$kHz, bottom-row: spectrograms for ACC2 same frequency ranges than in top figures.

As it is expected in the Fig. 10 is shown that the STNLS method is not able to extract the IAS from the ACC2, in contrast, the IAS extracted from the ACC1 is a perfect match visually comparing the IAS obtained from the tachometer signal, and the one obtained employing the proposed methodology. In terms of MSE the IAS obtained through the proposed piece-wise linear approximation is of 4.93×10^{-3} and 14.56×10^{-3} for the proposed methodology with and without quadratic term β_1 respectively. Similarly, without down-sampling the corresponding MSEs are: 2.15×10^{-3} and 2.94×10^{-3}, the improvement in the MSE it is expected, due to there are more points to compute the cost function. The results are consistent concerning the numerical test where it is verified that the influence of the quadratic term could introduce errors when dealing with low SNR.

As in the previous experiment Fig. 10 the IAS was not successfully extracted for the ACC2 record, consequently, the most natural approach is to extract the highest energy harmonic $\times 62$ after the visual examination of the Fig. 9, as a result it is shown in the Fig. 11 that the proposed methodology could extract the IAS even from a challenging record. Yet the results are far from the ones obtained for the record ACC1. Being the MSE for the record ACC2 of 3.21 and 3.19 for the estimation without and with the inclusion of the quadratic term β_1 respectively.

The critical point is after 50sec where the estimation partially lost the optimal point, and followed a sub-optimal as previously analysed in the numerical cost function in Fig. 2 the optimization problem is not convex and prone to fall into a sub-optimal. That point is critical disregard the inclusion of the quadratic term, for such a reason the IAS extraction on a record under a REB failure at high variable speed such as ACC2 continues to be a subject of research.

3.2.2 REB failure detection

In this subsection is done the angular re-sampling with the signal from the tachometer to not to bias the experiment, the signal was recorded located on HP tacho (L4), and the ACC2 record the only record with the bearing failure. In order to filter the vibration signal to obtain a signal as close as possible to the model Eq. (16), the experiment is the same as the previously done in the Section 3.1.2. As a result there are three scenarios to highlight the REB failure: first the naive approach, then, angular re-sampling, then SK filtering,

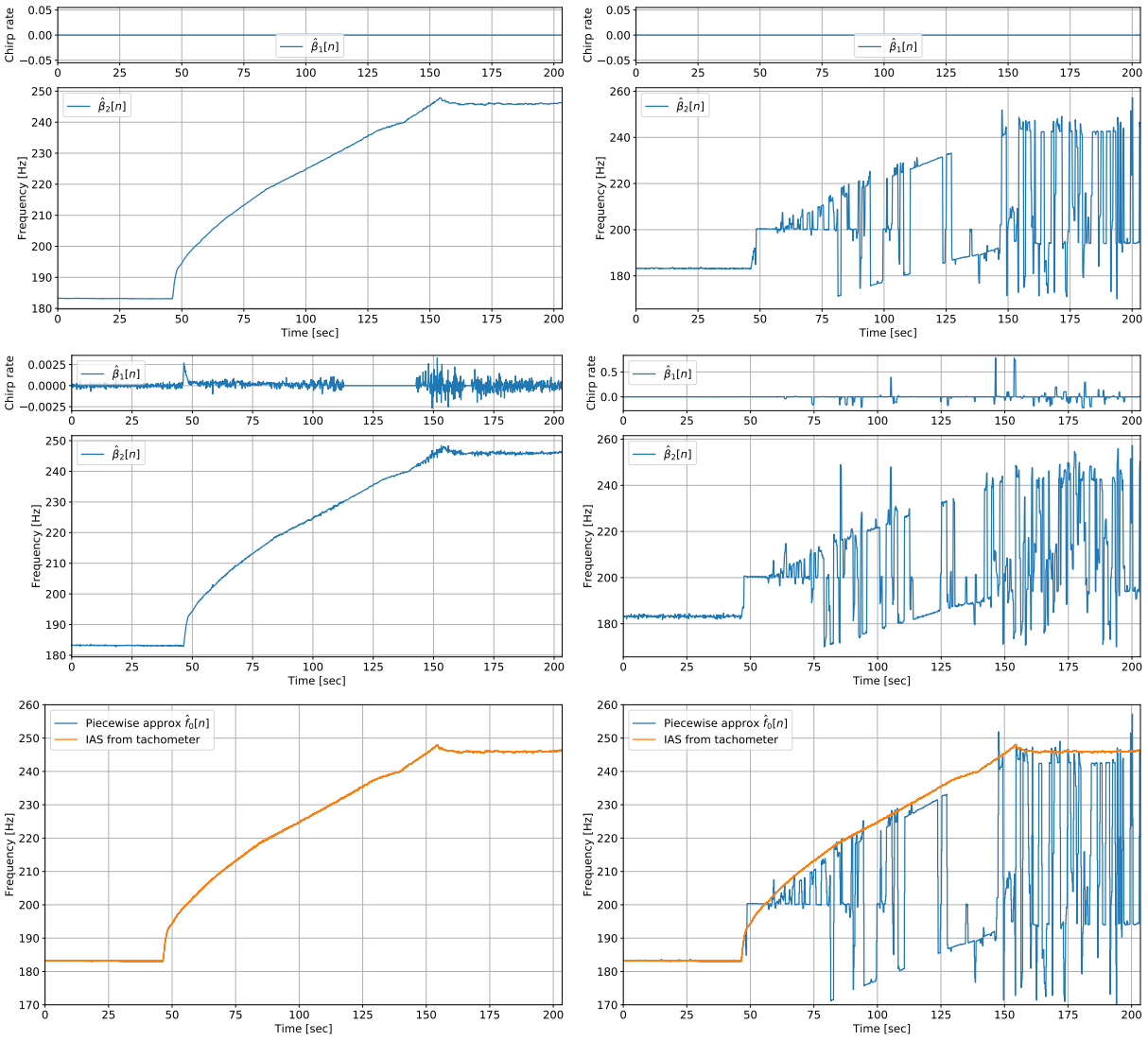


Figure 10: Left-column ACC1, right-column ACC2, rows: i) phase estimation without linear approximation, ii) phase with linear approximation $\{\beta_i[n]\}_{i=1}^{2}$, iii) IAS estimation after linear approximation.

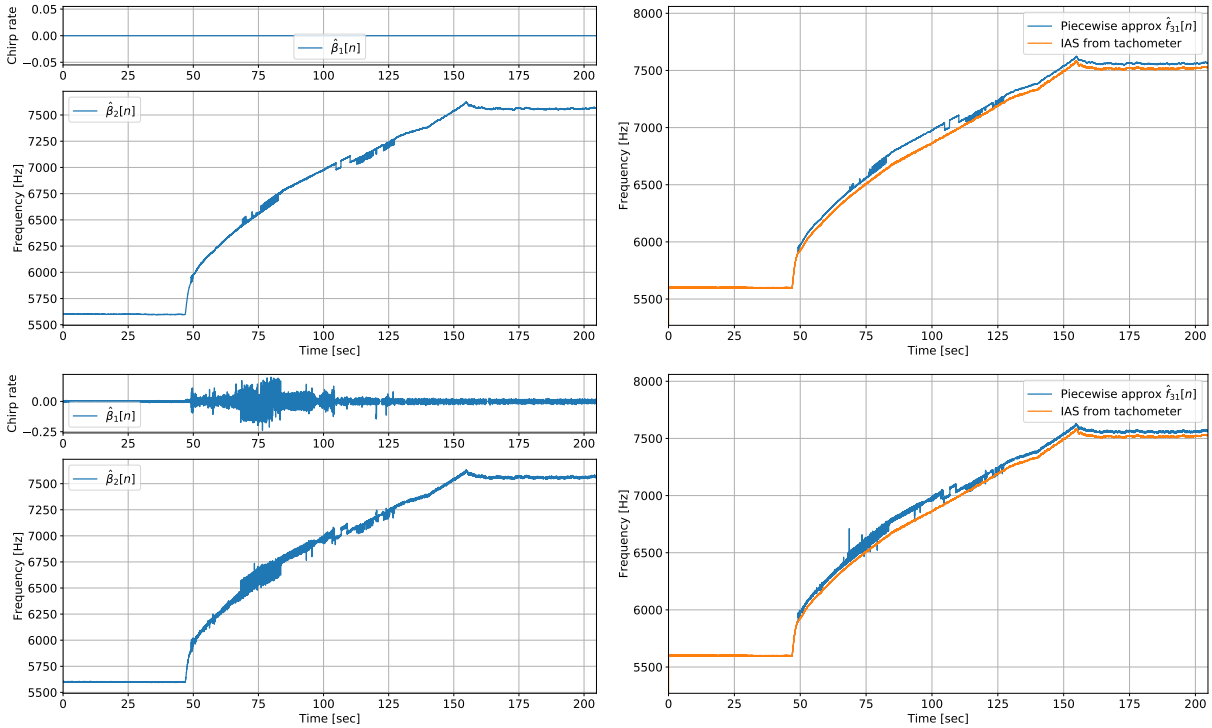


Figure 11: ACC2: IAS estimation approximation taking as fundamental frequency $f_{31}[n]$, top-row: without linear approximation, bottom-row: with linear approximation.

then envelope spectrum, the state of the art approach) SK filtering, then angular re-sampling, the envelope spectrum (valid only for small variations on the IAS), and finally the proposed approach short-angle based the STSK, angular re-sampling the short-angle SK based filter and the envelope spectrum obtained by means of the Welch's method of periodograms (filtered periodograms).

The results of the three considered scenarios are shown in the Fig. 12, where it is made a zoom close to the fault frequency 7.68evp (events per revolution). For the naive approach there is no highlight of the energy in the fault frequencies, for the traditional state of the art method the fault is highlighted, yet, the spectrum is noisy due to the SK is conceived for stationary signals a condition that is not fulfilled, given the significant variation of the IAS. Finally with the proposed STSK method the failure is highlighted and now appears some side-bands corresponding to half of the cage failure frequency of the other bearing failure in L1 a failure that only is made present with the proposed approach. The method is not sensible to the parameters, due to the only parameters necessaries are the window for the short angle SK (a larger window than the one of the SK itself), and a window for the SK computation itself (SK computation is based on the STFT). The first window is the next power of 2 of the period of the smallest expected frequency (131072 samples), and the window for the computation of the SK is 4096 samples, given a sapling frequency of 256 samples per revolution.

4 Conclusion

In the present work, a REB failure identification methodology under variable IAS is proposed, two tasks are addressed, the IAS failure estimation and the impulsive behaviour filtering for failure identification. Due to the fact that IAS introduces non-stationarities assuming that a signal in a short time window is highly stationary (cyclo-stationary), short-time/angle approaches are proposed for the two tasks at hand. For the proposed STNLS IAS estimation technique a robustness test is done building a database of contaminated signals with AWGN and Pink Noise and measuring the MSE between the theoretical IAS and its estimation for different levels of Noise. Also, to corroborate the methodology in a real scenario, the IAS is successfully retrieved from two records from an aircraft engine under a failure in REB. The second contribution is a short-angle SK filtering approach (STSK) to highlight the impulsive behaviour that has the information related to the failure. The proposed STSK is tested in a numerical signal and the aircraft engine, successfully highlighting the fault frequencies, and

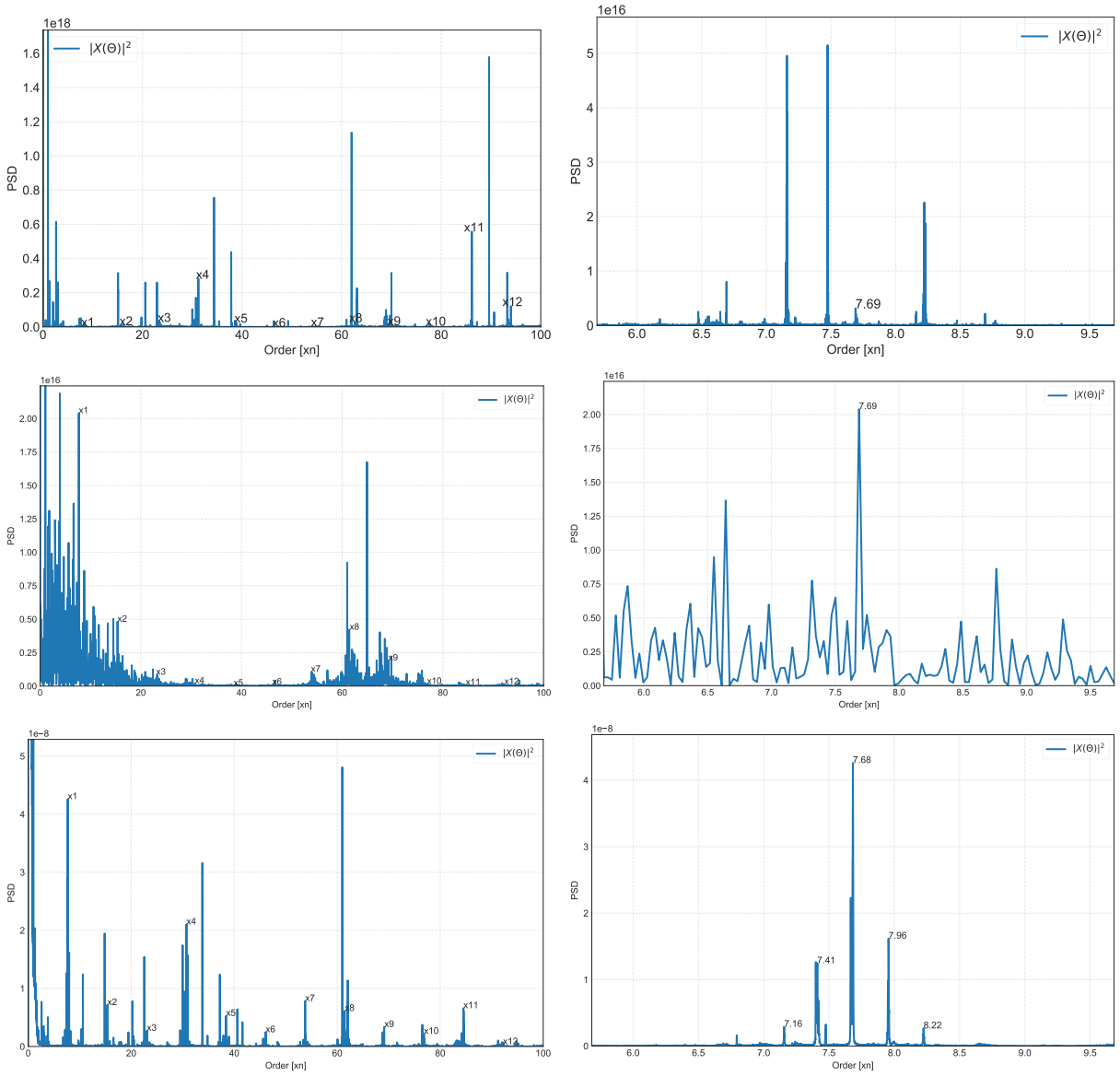


Figure 12: Top-row: baseline one, angular domain transform, SK filtering, then envelope spectrum, Middle-row: SK filtering, angular domain transform, then envelope spectrum, Bottom-row: **proposal** angular domain transform, then, envelope spectrum convolved with a time-frequency dependent SK based filter.

outperforming the traditional methodologies of the state of the art. As future work it will be interesting to study another parametric approaches for IAS estimation like the particle filter and polynomial Fourier transform. The algorithms are publicly available in https://github.com/efsierraa.

Acknowledgements

This work was performed within the framework of the Labex CeLyA of Université de Lyon, operated by the French National Research Agency (ANR-10-LABX-0060/ANR-11-IDEX-0007), the Convocatoria 647 de 2014 de Colciencias-Colfuturo, and IDEX-LYON "aide á la mobilité internationale des doctorants".

Références

[1] Randall, R.B., Antoni, J. : Rolling element bearing diagnostics - a tutorial. Mechanical Systems and Signal Processing 25(2), 485–520 (2011)

[2] Zimroz, R., Urbanek, J., Barszcz, T., Bartelmus, W., Millioz, F., Martin, N. : Measurement of instantaneous shaft speed by advanced vibration signal processing-application to wind turbine gearbox. Metrology and Measurement Systems 18(4), 701–712 (2011)

[3] Leclère, Q., André, H., Antoni, J. : A multi-order probabilistic approach for instantaneous angular speed tracking debriefing of the cmmno 14 diagnosis contest. Mechanical Systems and Signal Processing 81, 375–386 (2016)

[4] Antoni, J., Griffaton, J., André, H., Avendaño-Valencia, L.D., Bonnardot, F., Cardona-Morales, O., Castellanos-Dominguez, G., Daga, A.P., Leclère, Q., Vicuña, C.M., et al. : Feedback on the surveillance 8 challenge : Vibration-based diagnosis of a safran aircraft engine. Mechanical Systems and Signal Processing 97, 112–144 (2017)

[5] Fyfe, K., Munck, E. : Analysis of computed order tracking. Mechanical Systems and Signal Processing 11(2), 187–205 (1997)

[6] Antoni, J. : The spectral kurtosis : a useful tool for characterising non-stationary signals. Mechanical systems and signal processing 20(2), 282–307 (2006)

[7] Randall, R., Sawalhi, N., Coats, M. : A comparison of methods for separation of deterministic and random signals. International Journal of Condition Monitoring 1(1), 11–19 (2011)

[8] Borghesani, P., Ricci, R., Chatterton, S., Pennacchi, P. : A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. Mechanical Systems and Signal Processing 38(1), 23–35 (2013)

[9] Abboud, D., Baudin, S., Antoni, J., Rémond, D., Eltabach, M., Sauvage, O. : The spectral analysis of cyclo-non-stationary signals. Mechanical Systems and Signal Processing 75, 280–300 (2016)

[10] Christensen, M.G., Jensen, J.R. : Pitch estimation for non-stationary speech. In : 2014 48th Asilomar Conference on Signals, Systems and Computers. pp. 1400–1404. IEEE (2014)

[11] Christensen, M.G., Jakobsson, A. : Multi-pitch estimation. Synthesis Lectures on Speech & Audio Processing 5(1), 1–160 (2009)

[12] Randall, R.B., Antoni, J., Chobsaard, S. : the Relationship Between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals. Mechanical Systems and Signal Processing 15(5), 945–962 (sep 2001), http://linkinghub.elsevier.com/retrieve/pii/S0888327001914153

[13] Picinbono, B. : On instantaneous amplitude and phase of signals. IEEE Transactions on Signal Processing 45(3), 552–560 (1997)

Investigation of the influence of the operating parameters on the magnetic encoder geometric error compensation.

Halil Ibrahim Cakar [1], Ilyes Khelf [2], Hugo André [1]

[1] Univ Lyon, IUT de Roanne, LASPI, 42334 Roanne Cedex, France
[2] ENGIE GREEN, 92400, Courbevoie, France
{hugo.andre}@univ-st-etienne.fr

Abstract

The encoder is the only sensor needed to perform Instantaneous angular speed (IAS) analysis, an alternative technique used to monitor gears, bearings or other electro-mechanical elements. The encoder is subject to an intrinsic defect called Geometric Error (GE). Although it has various origins, GE can be simplified as being related to the variable angular size of every encoder segments forming the theoretically uniform pattern. As a result, GE introduces a cyclic perturbation observed on the spectrum of the estimated IAS. These perturbations exhibit a first order cyclostationary behaviour which replicates themselves in each revolution of the shaft. Since the impacted frequency channels can also be studied to monitor the health status of the shaft line, GE should be corrected for a better IAS estimation.

In this study, a rotation domain averaging based algorithm is developed to compensate the GE of the estimated IAS signals. The GE signature of a given signal is estimated and is used to compensate the GE of the other signals as well as itself. The term cross-correction is introduced to mention the correction of signals with each other's GE signature. The quality of the correction is analysed and is shown that it depends on several operating conditions. In other words, signals obtained for certain operating conditions are shown to be better at correcting GE than signals obtained for different operating conditions.

The developed algorithm is tested on a 2-MW wind turbine campaign which is instrumented with a magnetic encoder. These observations makes it possible to qualify the properties of the best GE corrector signals and dress an optimized correction algorithm suitable for any database. Since there were several interventions on the wind-turbine like re-installation of the encoder, gearbox change and gear defect, it is also possible to observe the influences of these interventions on the GE compensation. The results of this work are expected to be useful for gearbox operators as it represents a probable solution for early fault detection especially in demanding operating conditions.

1 Introduction

In the specific domain of varying speed rotating machine diagnosis, Instantaneous Angular Speed (IAS) monitoring is one of the very interesting alternative to classical vibration measurement system. It has been tried to detect chatter in milling [2], bearing defects on wind turbines [3] or on truck wheels [1]... This technology first presents the advantage to be easy to install, since only one encoder is sensible to speed variations induced by components installed far away from it: in [4], the encoder installed on the high speed shaft of a turbine was shown able to monitor the low speed shaft characteristic frequencies, beyond the mechanical coupling and the three stage gear box. Moreover, the elapse time acquisition technique present the benefit to be intrinsically made in the angular domain, and therefore easily deal with macroscopic speed variations. However, this technique does not benefit from the same feedback than classical vibration monitoring, and plenty of questions need to be tackled before it can compete with it on an industrial scale. It has been shown able to detect different kinds of bearing problems, and mechanical models have been built up to justify the small speed variations induced by such defects [6]. The peculiar quantification error, specifically linked to the elapse time technique, along with the inability of IAS monitoring to be protected from aliasing phenomenon has been thoroughly studied in [7] and [8]. Geometric Error (GE) is another major limitation for IAS, since it is present on both acquisition techniques (ADC based and Elapse Time). This error is linked to several aspects: imperfection of the encoder

gratings, the interpolation process unproperly tuned, or even the imprecise installation of the encoder. the common feature of all these aspects is their cyclic signature. Leclere et al proposed a technique to extract the from the IAS signal using rotation domain averaging [8], based on rotation domain averaging and used for steady speed signals. Bruand proposed two extra methods to adapt the GE estimation and correction to varying speed scenario [9]. The first is a data-driven approach based on a local weighted least squares method, while the latter is a model-based approach. The main drawback of all these approaches is linked to the fact that it is yet impossible to distinguish the GE from actual speed variations whose frequency is synchronous with the GE. In other words, if unbalance or gear mesh frequency happen to be coincident with the GE, they might disappear along with GE. This is troublesome if one is aiming at monitoring the unbalance of the gear-mesh health status. This paper comes within the scope of monitoring a defect whose frequency is synchronous with the shaft carrying the encoder. The main idea is to correct every measurement using the GE estimated using always the same signal, and to assume that the GE will not change from one measurement to the next. An evolution of the synchronous content would therefore be considered as actual speed variation, and therefore be assigned to the monitored element.

This paper aims at designing which signal(s) should be used to correct all consecutive measurement: to find the measurement that will rule the all... or to find the operating conditions that has an influence over the GE.

In order to avoid any terminological conflict, the measurement whose GE is used to correct all the measurements is called as the *Corrector Signal*. When one of the measurement is selected as the corrector signal, the other measurements are named as the *Signal to be Corrected*. When a measurement is corrected with the GE signature of the corrector signal, this measurement is called as the *Corrected Signal*. This terminology will be used throughout this manuscript. In the subsequent chapters the details are presented.

In the first part, the paper presents the method employed to estimate, to correct, and then to designate the what signals should be used to correct every others. The results obtained using a magnetic encoder on a long term measurement campaign are presented in the second part.

2 Method

This section first explains how is estimated the GE of an encoder using an elapse time measurement. Then, the correction process used on the corrected signal is presented. In the 3rd section, a quality indicator is presented to assess the quality of the processed corrected signal using the corrector signal. Finally, the maps used to synthesise the results will be introduced to efficiently show the influence of the operating conditions on the GE correction.

2.1 Geometric Error assessment

In the first step the GE signatures of both *Corrector Signal* and *Corrected Signal* are estimated. The method used in this paper is based on the one proposed by Leclere et al and quickly adapted to deal with non stationary speed [8]. Let first introduce the Rotation Domain Averaging (RDA) applied directly on the Elapse Time signal $\tau[j], j \in [1 : N \cdot R]$ with R the resolution of the encoder, N the number of revolution in the measurement. The RDA result is defined on one revolution: $\overline{\tau[j]}, j \in [1 : R]$, such as:

$$\overline{\tau[j]} = \frac{1}{N} \sum_{k=1}^{N} \tau[j + k \cdot R] \tag{1}$$

Let then assume that the resulting RDA $\overline{\tau[j]}$ is only due to the GE on one hand, and to the macroscopic speed trend on the other hand. The latter is deduced from the original measurement with any kind of Low pass Filter (LF), whose cut-off frequency is lower than 1 per revolution. The RDA of the macroscopic trend is noted $\overline{\omega_{LF}}(j), j \in [1 : R]$. Then, the GE can be deduced from:

$$\overline{\tau[j]} = \frac{\Delta\theta}{\overline{\omega_{LF}}[j]} \rightarrow \Delta\theta[j] = \overline{\omega_{LF}}[j] \cdot \overline{\tau[j]} \tag{2}$$

In the special but ordinary case where the speed trend is steady once it is observed on the RDA (see figure 1 the left bottom plot) which means the average speed trend is steady and that every originally speed variation

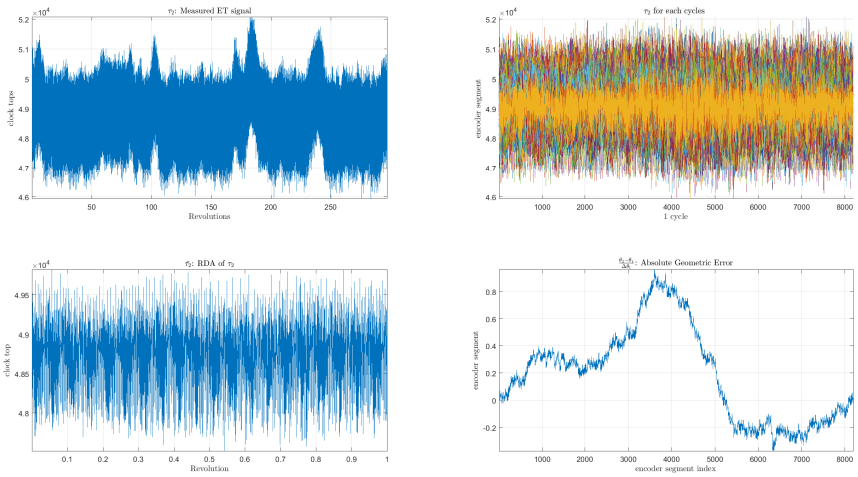


Figure 1 – GE Estimation of one of the measurements

rightfully compensates themselves, it is even possible to skip the low frequency filter and use the simplification proposed here under:

$$\Delta\theta[j] = \frac{\overline{\tau[j]}}{\sum_{i=1}^{N} \overline{\tau[i]}} \tag{3}$$

This steady speed assumption does not need the speed to be equal from one revolution to the next. However, it assumes that once the speed is averaged on every revolution, the resulting speed (1 revolution long) is steady. In the general case, and in this paper, the formulation proposed in Eq. 2 is preferred.

2.2 Cross Correction of the Geometric Error

The discrete integration of $\Delta\theta$ obtained in the previous section gives access to the angle θ between the launch of the counter clock and each consecutive encoder segment. These angles are not uniformly divided (that's the main reason of the paper!) and therefore influence the observed elapse time. From now on, these irregularly sampled angles are noted $\theta_2, j \in [1 : N \cdot R]$. As already mentioned, these errors are expected not to change from one revolution to the other, and from one measurement to the other.

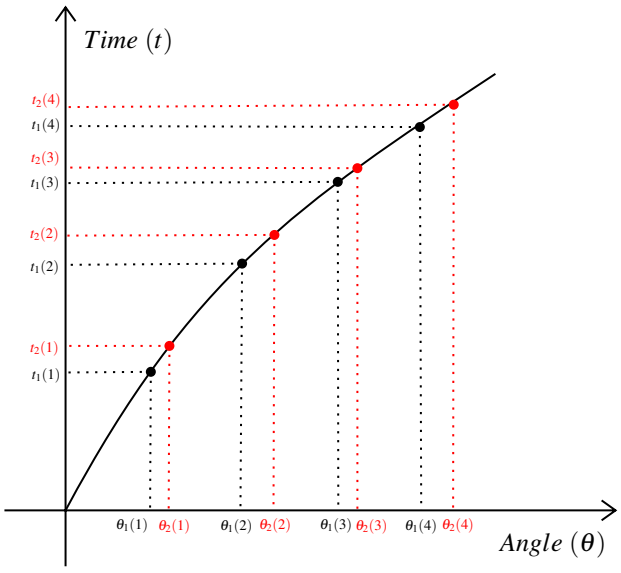


Figure 2 – Angle versus Time

The original elapse time measurement, obtained with irregularly sampled angles, are noted $t_2[j], j \in [1 : N \cdot R]$ once they are cumulated. A classical interpolation process is used to estimate the cumulated elapse time values $t_1[j], j \in [1 : N \cdot R]$ corresponding to the regularly sampled $\theta_1[j], j \in [1 : N \cdot R]$. In this section as in Fig. 2, presenting the correction process on a Time-Angle scheme, the index 2 stands for GE-Noised signal when the index 1 stands for the GE-corrected signal. the spline interpolation method has been chosen to estimate the GE-corrected values.

In this paper, one signal is used to correct an other. The *corrector signal* is used to obtain the angles $\theta_2[i] = \theta_2[j] + k \cdot N.\ \theta_1[i]$. The *signal to be corrected* is cumulated in $t_2[i] = \sum_{k=1}^{i} \tau_2[k]$, interpolated to give $t_1[i]$ and finally differentiated to estimate the elapse time we would get without GE: $\tau_1 i = t_1[i] - t_1[i-1], i \in [2 : N \cdot R]$. $\tau_1 i$ is the *corrected signal*.

2.3 Correction quality assessment

In order to asses the ability of one signal to accurately remove the GE from any other signal, the estimation of the GE removal quality is needed. The GE being concentrated on integer orders, due to its first order cyclostationnary behaviour, the energy contained by the integer orders appears as the ideal indicator.

$$L_1 = \frac{1}{N} \sum_{m=1}^{N/2} \mathbf{F}_x^{\theta}(m) \tag{4}$$

with $\mathbf{F}_x^{\theta}(m)$ the DFT of the signal x computed for the order frequency m. This equation sums up the integer orders of the amplitude spectrum of the finite length signal. This corresponds to the sum of the TSA amplitude spectrum, and this is therefore not only representative of the GE, but also from every actual synchronous speed variations.

Fortunately, the order frequency of the most energetic synchronous speed variation are known. Unbalance and blade pass frequency is energetic up to the sixth order; gears frequencies are inducing energetic peaks on integer multiples of 93... Therefore we propose to remove these frequencies, and others, from the quality indicator. Equation 5 presents an example of such an indicator. More frequencies have been removed, but this paper will not bring out of the shadows these details for the sake of simplicity.

$$I = \frac{1}{N} \left(\sum_{m=1}^{N/2} \mathbf{F}_x^{\theta}(m) - \sum_{m=1}^{9} \mathbf{F}_x^{\theta}(m) - \sum_{m=1}^{44} \mathbf{F}_x^{\theta}(93 \cdot m) \right) \tag{5}$$

2.4 Operating parameter analysis

This part presents the *cross-comparison* of several measurements obtained on one machine under various operating conditions, in order to investigate the quality of the correction process, and to designate the best operating to conditions to estimate GE. *Cross-comparison* is applied on a group of M measurement, where every measurement is used to successively correct every others. Once the *corrected signal* is corrected with the GE signature of the *corrector signal*, the GE correction efficiency is observed using the quality index presented in the previous section. Then, every measurement are arranged according to an operating parameter. For example, if the average power is the analysed operating parameter, the measurement are arranged in ascending order of mean power: the first signals are those acquired while the machine is not producing any current while the lasts were acquired at nominal power.

The result of the computation/arrangement is a matrix $I_{[M \times M]}$ which columns designate the *corrector signal* while its rows designate the *corrected signal*. Therefore, $I_{i,j}$ is presenting the quality index of the i^{th} measurement GE corrected using the j^{th}. Finally, Every rows are normalized with their corresponding *auto-corrected* values. Diagonal terms present the quality index obtained when the measurements are *auto-corrected*, *ie* they are corrected using their own GE estimation. Hence, the *cross-correction* matrices are obtained as in eq. 6.

$$I = \begin{pmatrix} 1 & \frac{I_{12}}{I_{11}} & \frac{I_{13}}{I_{11}} & \dots & \frac{I_{1m}}{I_{11}} \\ \frac{I_{21}}{I_{22}} & 1 & \frac{I_{2m}}{I_{22}} & \dots & \frac{I_{2m}}{I_{22}} \\ \dots & \dots & \dots & \dots & \dots \\ \frac{I_{m1}}{I_{mm}} & \frac{I_{m2}}{I_{mm}} & \frac{I_{m3}}{I_{mm}} & \dots & 1 \end{pmatrix} \tag{6}$$

The results presented hereafter will plot such cross correction matrices in greyscale colormaps. In this study, the number of measurements in each cross-comparison group is limited to $M = 300$ to save computation time.

3 Experimental results

3.1 Campaign presentation

A long term study is being carried over a MM92 wind turbine. The wind turbine set-up is presented in Figure 3 for the reader to realize the easiness of the involved instrumentation in regards with the kinematic complexity of the turbine line shafting. The speed transmission is made of one star epicyclical gear train and one parallel stage mounted in serial configuration to obtain a global speed ratio approximately equal to 119. The IAS signal can be computed from the generator optical encoder, which is the high quality incremental encoder used by the converter to correctly synchronize the asynchronous generator; but also from a lower quality magnetic encoder, installed in a retrofit operation on the low speed shaft directly carrying the rotor hub. It has been decided to equip the low speed shaft since the most expensive shaft line elements are kinematically and physically closer from it. This study will focus on this 20480 pulses per rev magnetic encoder installed beyond the slip ring, on the low speed shaft.

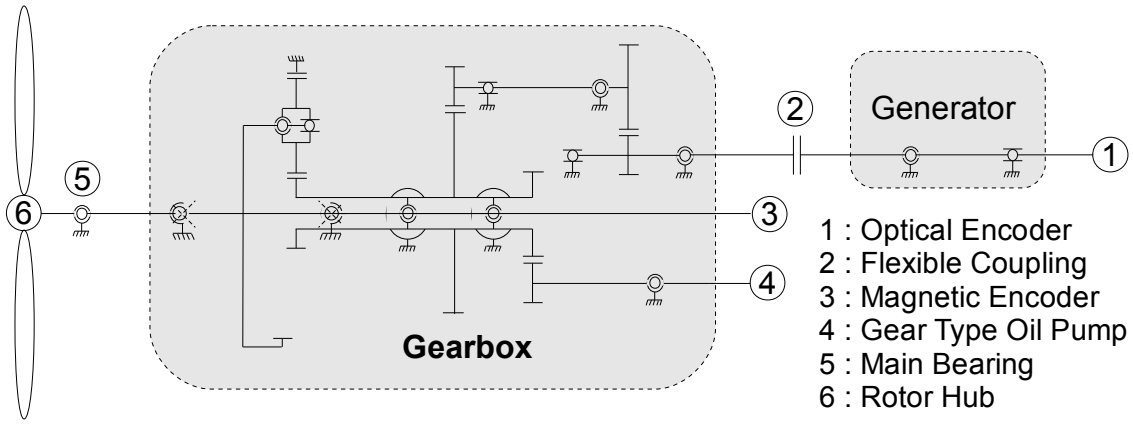


Figure 3 – Kinematic Scheme of the wind turbine set-up.

The measurement have been acquired between august 2017 and October 2018 on the same machine. Since the machine was suffering a major gear crack on the epicyclic ring, the gear box and the magnetic encoder were changed on the 25th of September. Nothing but the new gearbox running in was expected to be seen later on.

The acquisition card, FPGA type, is embedded in an industrial PC directly installed in the nacelle and reads the elapse time signal with a 120MHz counter clock. The signals are 300 revolutions long, and are synchronized with 1hz process data describing the operating conditions:

- date
- wind speed / direction,
- active/reactive/actual power,
- nacelle orientation,

- inside/outside temperature

All these operating parameters were analysed through the process detailed in the previous part. This paper will only present the most interesting results.

3.2 Resulting maps

This section successively presents the parameters that have been seen to play a role in the GE correction quality, starting from the most influential.

3.2.1 Influence of the Date

The 300 measurements on the cross-comparison map shown in Fig. 4 have been selected randomly before and after the encoder removal date. The only condition was set on the minimum gearbox speed, which must be greater than $800rpm$ to avoid start/stop measurement. Those measurement have been successively cross-corrected according to every available operating parameter. And the date criteria appeared as the the most informative one. The diagonal of the plot presents the lowest values: 1. This reminds that the auto-corrected values are used as references.

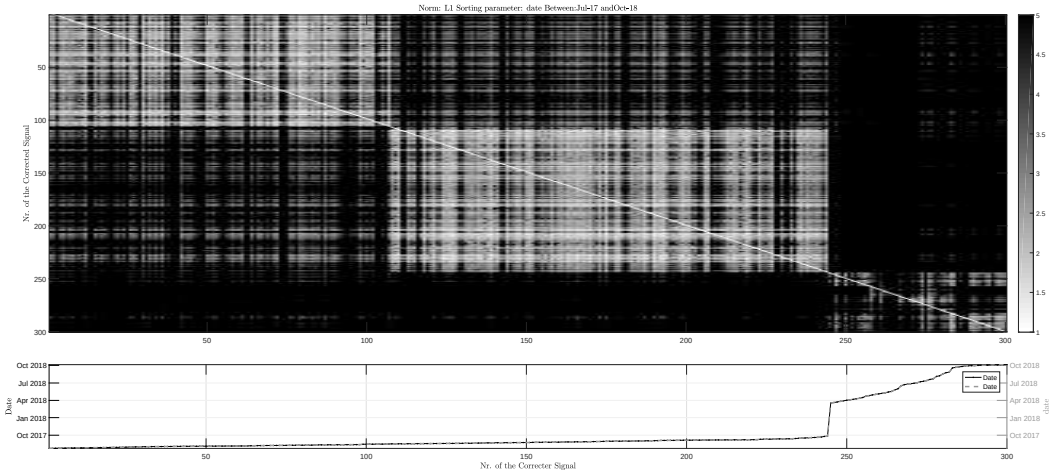


Figure 4 – Cross correction map sorted by date. The bottom plot reads the date of each measurement.

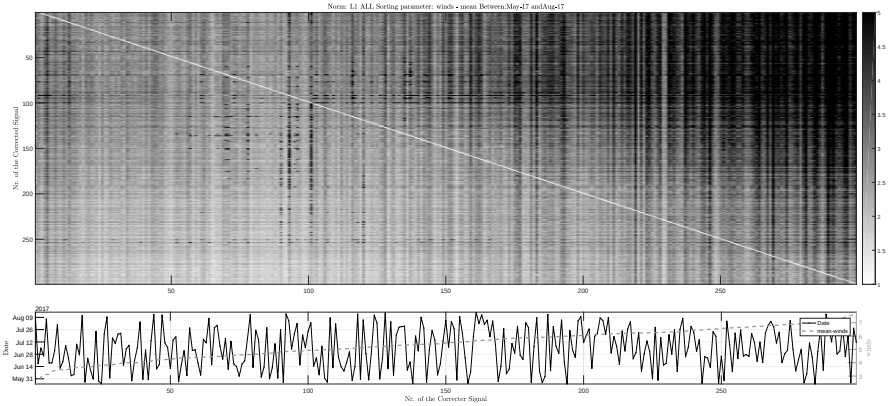
Unlike the maps obtained using other parameters, this map clearly exhibits vast light and dark regions. Lighter regions means every signal within it correct and are corrected in a better way between themselves. The top left region is limited by the measurement 105^{th}, which has been acquired just before encoder removal, on 25^{th} September, 2017. If the measurements obtained before this date are corrected by measurement obtained after: the quality of the correction is lower. This observaiton confirms that the removal of the encoder, the orientation and the geometric position between the encoder ring and the reader head modifie the GE.

Moreover, still looking at figure 4, one may pay attention to the bottom right region, starting on the 244^{th} measurement. There is no clear explanation to this sudden change of behaviour. The history of this machine was checked by its owners: it was noticed that there was only one distant reset on the machine but no encoder removal. A distant reset is the routine action performed remotely by the supervising operators when a minor alarm stops the machine. Such an alarm is sometimes reseted without further analysis if it does not repeat. This coincidence is interesting, though not demonstrative.

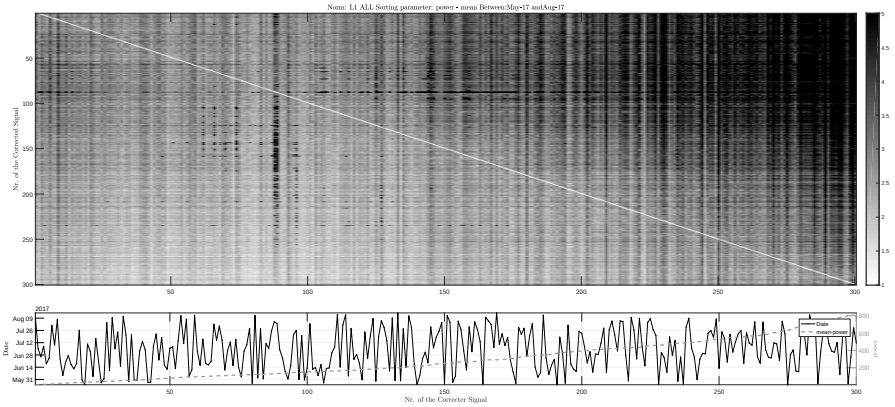
3.2.2 Influence of the Power and Wind Speed

Let's now investigate the next most influential parameter. 300 new measurements are selected using the same conditions as before, plus a date criteria: measurement must be acquired between May and August (the

bottom right region observed in the previous section). Among the different re-sorted matrices, the matrices sorted according to their mean power and the mean wind speed were observed to be the best classifiers. The sorted maps demonstrate the similar behavior as observed on figure 5a and figure 5b, which is fairly normal: actual power and wind speed are linearly correlated.



(a) Mean Wind Speed



(b) Mean Power

Figure 5 – (a)Cross correction map sorted by wind speed. The bottom plot reads the date and the average wind speed of each measurement. (b) Cross correction map sorted by actual power. The bottom plot reads the date and the average actual power of each measurement.

3.2.3 Influence of the Nacelle

Let's now investigate the next most influential parameter. 300 new measurements are selected using the same conditions as before, plus a power criteria: measurement must be acquired between 0 and 300 watts. Among the different re-sorted matrices, the matrices sorted according to their mean nacelle direction were observed to be the best classifier. Figure 6 present an interesting *cross shape*, showing that low values are well corrected by low and by high nacelle direction values. Indeed, low and high nacelle direction values both correspond to the north direction, while 180 degrees correspond to the south direction. The most reasonable explanation is that the earth magnetic field influences the magnetic encoder GE. An alternative explanation was linked to the topographic environment of the turbine, but the monitored machine was actually shown to stand in the middle of a naked plain, far away from any forest, hills, or other turbine's wake effect.

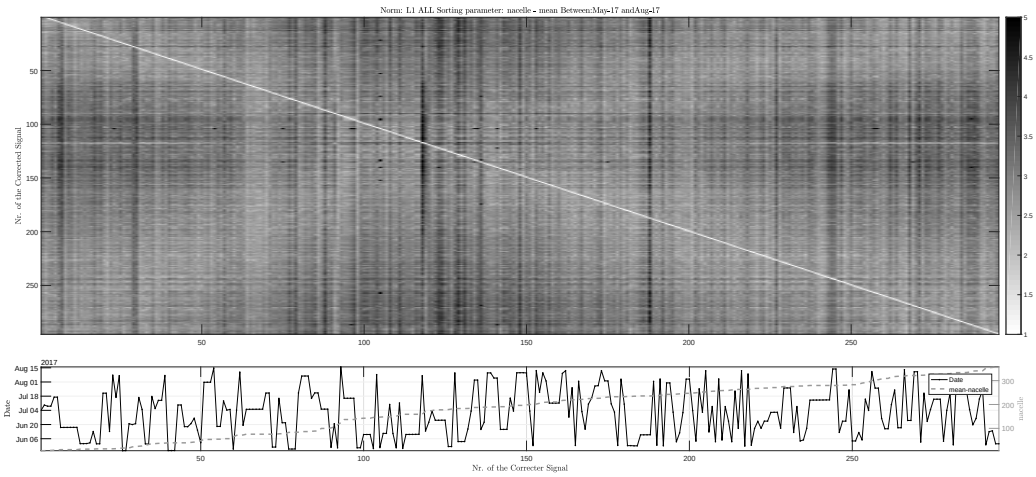


Figure 6 – Map demonstrates the influence of the nacelle values on GE correction efficieny. *The map is sorted by mean nacelle values.*

4 Conclusion and perspective

Instantaneous angular speed monitoring is often based on the use of an angular sensor, such as an encoder, whose gratings are often perturbed by an unequal distribution. This problem might be due to the manufacturing, the installation of even to the calibration of the electronics used to interpolate the signal. Anyhow, this kind of defect is called Geometric Error, and is interpreted as speed variations whose frequencies are the orders of the shaft carrying the encoder. Among the various techniques existing to remove GE, none spare the actual synchronous speed vibration. therefore, how to monitor a gear whose amplitude is biased by GE ? The solution proposed in this paper is to correct every signals GE obtained on a machine with one reference signal. if something evolves: it's the synchronous component! Finally, one signal is not yet enough to characterise the GE. low speed measurement appear as a good corrector, but the GE need to be reseted when the encoder ring or sensor is reinstalled. Moreover, the orientation of the sensor regarding the earth magnetic field has to be taken into account. The latest issue might be tackled by a model correcting the influence of the magnetic field on the sensor, knowing the orientation of the gratings during the reference and the corrected measurement.

4.1 Acknowledgements

the authors would like to thank ENGIE GREEN for their long term support. Instantaneous Angular Speed is a box full of both good surprises barriers to be broken down. For the latter, they indisputably deserve a good part of the merit.

References

[1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, D. Rémond, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, Mechanical Systems and Signal Processing, Volume 24, Issue 7, October 2010, Pages 1998-2011, ISSN 0888-3270, http://dx.doi.org/10.1016/j.ymssp.2010.05.005

[2] M. Lamraoui, M. Thomas, M. El Badaoui, F. Girardin, Indicators for monitoring chatter in milling based on instantaneous angular speeds, Mechanical Systems and Signal Processing, Volume 44, Issues 1-2, 2014, Pages 72-85, ISSN 0888-3270

[3] H. André, A. Bourdon, D. Rémond, On the use of the Instantaneous Angular Speed measurement in non-stationary mechanism monitoring. Proceedings of the ASME 2011 International Design Engineering Technical Conferences & 23rd Biennial Conference on Mechanical Vibration and Noise, IDETC/CIE 2011, Washington, DC, USA

[4] H. Andre,I. Khelf, Q. Leclere, Harmonic product spectrum revisited and adapted for rotating machine monitoring based on IAS., Surveillance 9, 2017, Fes, Morocco

[5] C.J. Stander, P.S. Heyns, Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions, Mechanical Systems and Signal Processing, Volume 19, Issue 4, 2005, Pages 817-835, ISSN 0888-3270

[6] JL. Gomez, A. Bourdon, H. André, Didier Rémond, Modelling deep groove ball bearing localized defects inducing instantaneous angular speed variations, Tribology International, Volume 98, 2016, pp 270-281

[7] H. André, F. Girardin, A. Bourdon, J. Antoni, D. Rémond, Precision of the IAS monitoring system based on the elapsed time method in the spectral domain, Mechanical Systems and Signal Processing, Volume 44, Issues 1-2, 20 February 2014, Pages 14-30, ISSN 0888-3270, http://dx.doi.org/10.1016/j.ymssp.2013.06.020

[8] Q. Leclere, F. Girardin, D. Rémond. An Analysis of Instantaneous Angular Speed Measurement Errors. Surveillance 7 International Conference, Oct 2013, CHARTRES, France. pp.1-11. hal-00905103

[9] G. Bruand, F. Chatelain, P. Granjon, N. Martin, C. Duret, et al., Estimating the Rotational Synchronous Component from Instantaneous Angular Speed Signals in Variable Speed Conditions. International Conference on Condition Monitoring of Machinery in Non-Stationary Operations (CMMN0'2018), Jun 2018, Santander, Spain. hal-01761532

Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners

Q H Pham[1], M Gagnon[2], J Antoni[3], S A Tahan[4], C Monette[5]

[1]Phd Student at ÉTS Montréal and INSA de Lyon
[2]Institut de recherche d'Hydro-Québec, Varenne, Québec, Canada
[3]Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, Villeurbanne, France
[4] École de technologie supérieure (ÉTS), Montréal, Québec, Canada
[5]Andritz Hydro Canada LTD, Point-Claire, Québec, Canada

Email: quang-hung.pham.1@ens.etsmtl.ca

Abstract

The lack of experimental information can lead to an inaccurate prediction of hydroelectric turbine runners fatigue damage. Therefore, to recover this information, this research aim is the use of existing data measured by strain gauge to interpolate the unknown or not observed information about runner strain over the complete range of the steady-operating conditions for hydroelectric turbine. At steady-operating conditions, a strain signal, measured on the runner, can be separated into three principal components: static, periodic and stochastic. This paper presents the first step of our research that extracts and interpolates the periodic part at steady-operating states. A case study is used to compare two different kriging interpolation methods: the Spatial Kriging Method (based on 2D semivariogram) and the Spatio-Temporal Kriging Method (based on 3D semivariogram). The interpolation results are compared and validated with the experimental values.

1 Introduction

Strain levels play an important role in the fatigue reliability evaluation of hydroelectric turbine runners. Due to high costs and downtime required for the instrumentation, the experimental strain measurements on turbine runners are often limited (short measurement length, limited number of operation conditions, limited number of measured locations...), which could lead to an inaccurate evaluation of fatigue damage. Over the past years, many researches have tried to improve the limitation of experimental measurements. Poirier et al. (2016) regenerated strain signal over a long period by extrapolating a few minutes collected from the experimental measures [1]. Firas (2012) realized spatial interpolations between different locations on turbine runners to evaluate the distribution of damage and uncertainty propagation [4]. Gagnon et al. (2012) developed a model of turbine runner reliability by considering the High Cycle Fatigue (HCF) onset as the limit for the fatigue evaluation [2]. This model highlights the role of HCF, which is mainly linked to hydroelectric turbine Steady-Operating Conditions (noted as SOC). Currently, we are unable experimentally to obtain the dynamic strain in runner blades over all the possible operating conditions. Therefore, our research aims to estimate the missing experimental data by interpolating existing measured data. The idea is to develop an interpolation tool between different SOC available experimental data (Figure 1). This tool should help to make more accurate maintenance plan, leading to time and cost reduction.

The initial parameters used for interpolation are from measurements obtained by the strain gauges installed on a Francis turbine runner blade. Strain signals are complex and contain several physical phenomena. Some of these phenomena are hidden by others. Therefore, the interpolation of the complete signal is difficult. The proposition is to interpolate each phenomenon independently. We propose that the strain signals at SOC are separated into three principal components: static, periodic and stochastic. This paper presents a method that extracts and interpolates the periodic part. The periodic phenomenon hidden in the signal is linked with the

synchronous rotation speed of the turbine and is extracted by using the synchronous average operator to obtain the first order cyclostationary components. In this study, the interpolation method chosen is kriging. Kriging is a well-known spatial interpolation method that is commonly used in many domains, especially geostatistics. In general, the strain signal on turbine runner varies based on time and operating conditions of the turbine, but the influences of these two "dimensions" during interpolation phase are unclear. Thus, two kinds of kriging process were studied, spatial kriging and spatio-temporal one, for the interpolations of the periodic component. This initial interpolation between different operating conditions provides the basis for interpolations of the other components of the signal.

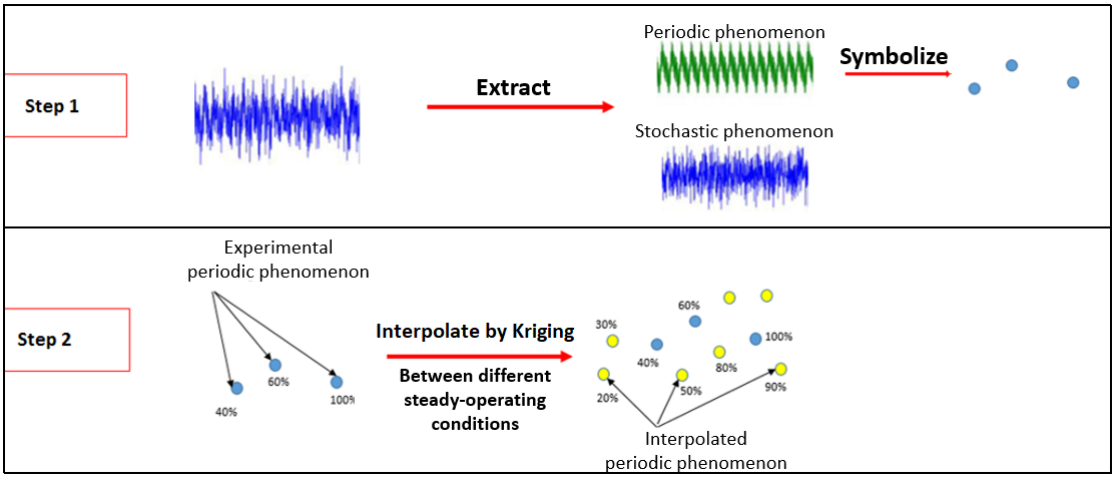


Figure 1 – Proposed interpolation tool

The paper is organized as follow: First, in Section 2, a description of the synchronous averaging, which extracts the periodic phenomenon in the signal, is presented. Then, in Section 3, the interpolation process is presented with spatial and spatio-temporal kriging. Finally, the interpolation results are shown and discussed in Section 4 and 5.

2 Synchronous Average - First order of cyclostationary

The strain signal in the case of the turbine consists of static, periodic and stochastic parts. The periodic parts in the strain signal measured on runner are typical phenomena related to hydroelectric turbine runners rotation (especially in the steady states where the speed of rotation is synchronized to the electrical network) while the stochastic phenomena correspond to the asynchronous parts (residue parts, stochastic vibrations, noises...). The separation will facilitate the interpolation because the relation between the phenomenon assigned to different operating conditions is less complex than the relation between whole signal where some phenomena are hidden by others. This separation leads to cyclostationnarity process, which will make possible the formalization of principles to extract hidden periodic signals. For the cyclostationary domain, the dynamic part of an experimental strain signal $S[n]$ measured on runner can be decomposed into three components: the cyclostationary moment of order 1 (noted as $CS1[n]$) linking to the rotation speed synchronous, the order 2 (noted as $CS2[n]$) referring to the periodic fluctuation of energy, and the residue part $R[n]$ [1].

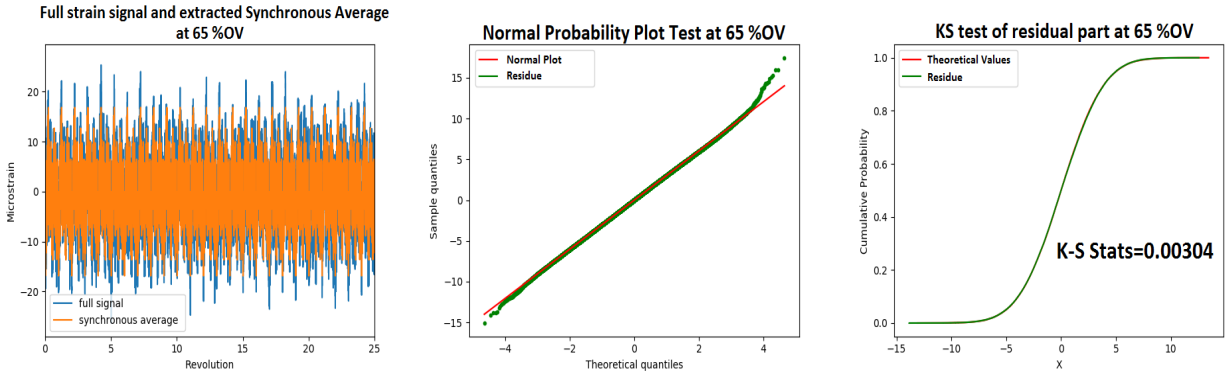
$$S[n] = CS1[n] + CS2[n] + R[n] \tag{1}$$

Synchronous Averaging is one of the extraction tools which allow the extraction of the $CS1$ (periodic part) hidden in the signal. The synchronous average (noted as SA), hidden in a signal $S[n]$ of finite cycles K, is extracted by using the equation (2).

$$CS1[n] = \frac{1}{K}\sum_{k=0}^{K-1} S[m+kN] \tag{2}$$

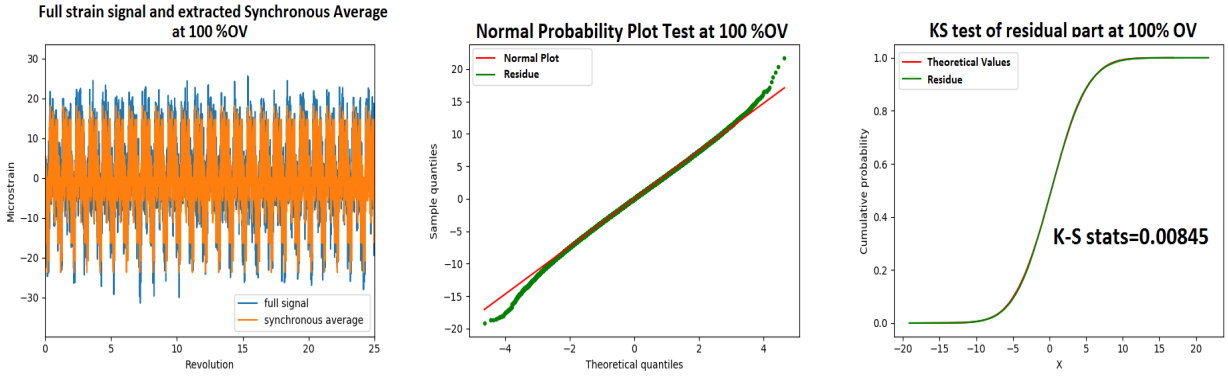
where K is number of cycles, N=L/K is length of cycle (L is full length of signal), m= mod[n,N]= n-[n/N]N.

This formula is applied for experimental strain signal to extract information at a cyclic frequency (including its harmonics), which relates to the synchronous rotation of the runner at steady-operating conditions. The periodic part can be directly extracted in the time domain. However, in the case of the rotating machinery like the hydroelectric turbine, the cyclostationnarity model is often based on the study of the angular domain rather than time. Therefore, an angular sampling of time signal (proposed by Bonnardot [5]) before the extraction is required in order to improve the quality of the synchronous average extraction. Figures 2 and 3 show the synchronous average at some operating conditions of the turbine and the statistic tests of residual part (after subtracting the average). Two statistic tests: the normal probability plot and the Kolmogorov-Smirnov test (noted as KS test) have been applied to verify the quality of the synchronous average extraction. Small KS statistic value and good fit between sample and theoretical quantiles verified that the residual part is almost aleatory. These results mean that the periodic part was almost fully extracted from the strain signal. The word "almost" is mentioned because there are slight divergences at extremity observed in Normal Probability graph. Higher order cyclostationary extraction could be applied to obtain a more aleatory residue. However, by these statistic test results, the synchronous average operator is an appropriate method for the initial extraction of periodic phenomena hidden in runner strain signal.



(a) (b) (c)

Figure 2 – Results of SA extraction process at 65 %OV. (a) Extracted Synchronous Average ; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.



(a) (b) (c)

Figure 3 – Results of SA extraction process at 100 %OV. (a) Extracted Synchronous Average; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.

3 Interpolation process

In this case study, the interpolation of periodic phenomenon is performed between different levels of guide vane opening (noted as %OV). The guide vanes is a part of the turbine that controls the flow rate as a function

of vanes opening. We consider that the interpolation space in this study is different steady-operating conditions (different %OVs) of the hydroelectric turbine. The proposed interpolation process consists of three steps as shown in Figure 4.

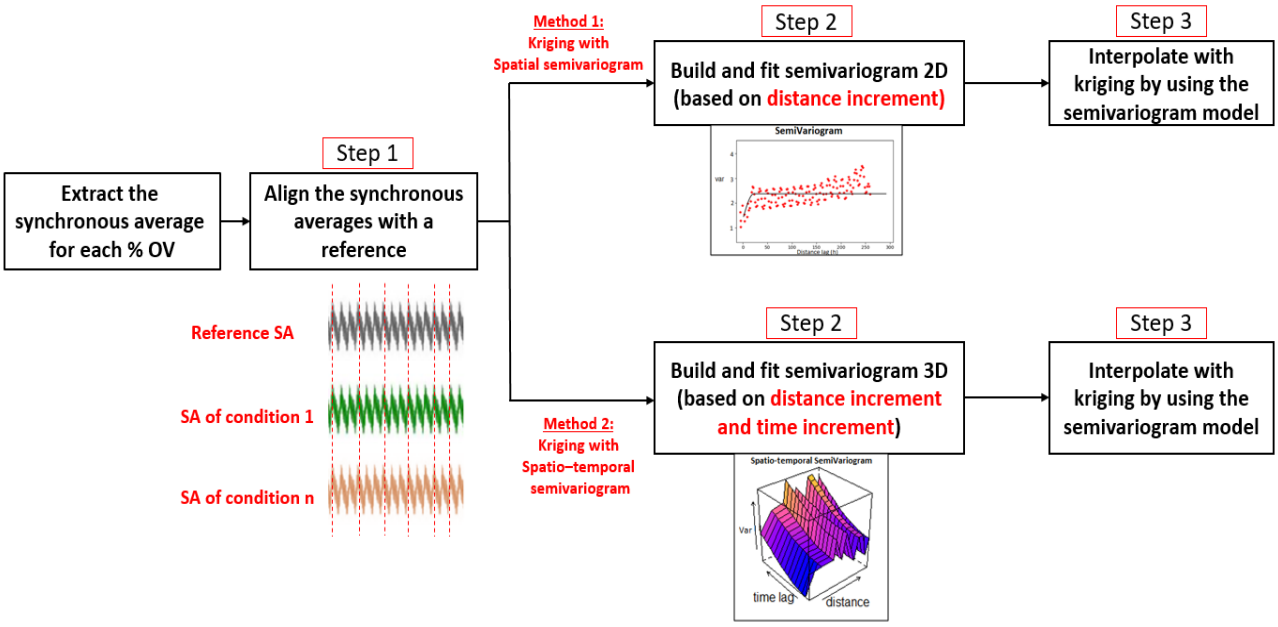


Figure 4 – Proposed interpolation process

Before applying kriging interpolation, the first step is to align the synchronous averages (noted as SA) of measured and known operating conditions (considered as known locations in the interpolation process). The SAs figure in the polar domain (figure 5) shows that there are small lags between values at 0° runner rotation of different %OVs. These lags do not affect the measurement results, they can cause however a systematic error for the interpolation process. For aligning these SAs, the first step is to choose a SA reference from which the other SAs can be aligned (e.g., the reference selected in this case study is the SA at 50 %OV (See figure 5)). The correlation coefficients, between value at 0° rotation of the reference with values around 0° rotation of other SA (from other conditions), is then determined until the best correlation for each condition is found. The final step is to re-arrange all SA by following the new values at 0° rotation corresponding to the best correlation found in the previous step (figure 5).

Next steps of interpolation process relate to the application of kriging interpolation. Kriging allows the estimations of missing values at given locations (from known observed data) by determining the "linear regression weights" which minimize the error variance. The error variance (also called the estimation variance [6]) is theoretically defined as a variance of the difference between the experimental strain value S and the interpolated strain value S^* at the same location u (Equation 3). The location u can be a spatial location, time location or both depending on the choice of kriging method.

$$\sigma_e(u)^2 = Var[S(u) - S^*(u)] = Var[S(u)] + Var[S^*(u)] - 2Cov[S(u), S^*(u)] \tag{3}$$

Each type of kriging has a different way for minimizing this estimation variance. In this study, the Ordinary Kriging (noted as OK) is applied for the interpolation process. According to interpolation results between different points on runner blade in [4], the OK gave the smallest estimation variances between three traditional linear kriging methods: Simple Kriging, Ordinary Kriging and Universal Kriging. The estimation function and the modeling equations for error variance for OK (detail in [4] and [6]) are expressed in equations (4) and (5).

$$S^*_{OK}(u) = \sum_{i=1}^{n} \lambda_i S(u_i) \tag{4}$$

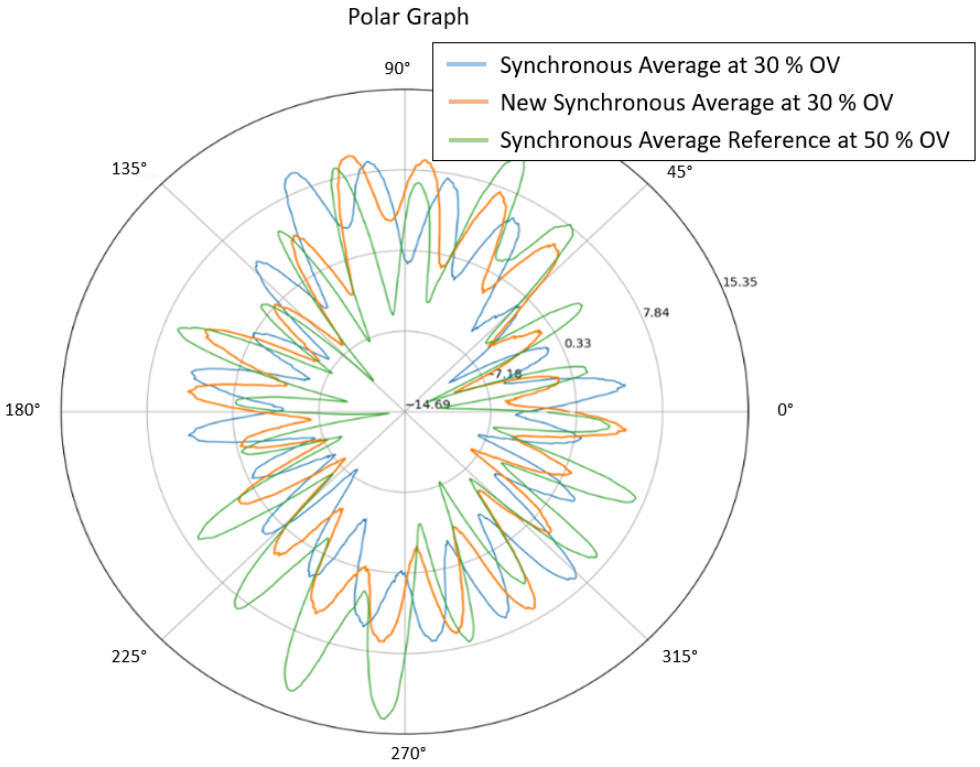


Figure 5 – Synchronous Average in polar field

$$\begin{cases} L_{OK}(\lambda_i,\mu) &= \sigma_e(u)^2 + 2\mu(\sum_{i=1}^{n}\lambda_i - 1) \\ \sigma_{OK}^2 &= \sum_{i=1}^{n}\lambda_i\gamma(u,u_i) - \mu \end{cases} \tag{5}$$

where the first equation of (5) is Lagrange function applied in OK to optimize the mean squared error, the Lagrange multiplier μ associates with the constraint $\sum_{i=1}^{n}\lambda_i = 1$, which is the OK constraint [6]. The second equation of (5) is the estimation variance obtained after minimizing the Lagrange function. Therefore, in OK, to minimize the estimation variance σ_{OK}^2, it is necessary to choose not only the weights λ_i, but also the Lagrange multiplier μ. γ is the semivariogram value (detail in the following paragraphs).

The minimization of estimation variance requires a model for the covariance between value of different locations (both interpolation and known locations). According to equation (5), these covariances are replaced by using the semivariogram values γ. In this paper, two types of semivariogram are applied: the Spatial Semivariogram and the Spatio-temporal Semivariogram. These two methods show the influences of two different types of information (operating conditions and temporal/angular information) on the kriging interpolation process.

Kriging interpolation with spatial semivariogram

To verify the intrinsic hypothesis that the finite variance of increment [$S(u)$ - $S(u+h)$] does not depend on location u, the variogram can be defined by the following function:

$$Var[S(u+h)-S(u)] = 2\gamma(h) = E[S(u+h)-S(u)]^2 \tag{6}$$

where h is the distance increment, γ is the semivariogram.

According to the equation (6), the spatial semivariogram can be defined as the half-variance between pairs of values and depends only on the distance increment h of these two locations. The spatial semivariogram is used for modeling spatial variability in kriging [6]. In this research, the spatial semivariogram is calculated by determining the dispersion of two different observations of a strain data set.

$$\hat{\gamma}(h) = \frac{1}{2N(h)}\sum_{(i,j)\in N(h)}[S(u_i)-S(u_j)]^2 \tag{7}$$

where $N(h)$ is number of paired points separated by a distance increment h, $S(u_i)$ and $S(u_j)$ are respective experimental strain values at positions u_i et u_j. The positions u_i or u_j are defined for each kriging method. The $\hat{\gamma}$ means the experimental semivariogram.

To use the semivariogram, a scatter plot (called as experimental semivariogram) has to be built using equation (7) with the measured data and then fitted with a numerical model by using the least square method. The experimental semivariogram cannot be directly used for kriging interpolation because it sometimes consists of "negative variance". So, the kriging interpolation is performed by using the numerical model of semivariogram for the covariances. The numerical model of semivariogram can be defined by sill, range and nugget effect parameter (if it is necessary). The range parameter is the distance from which there are no longer correlations (null covariance) in the data set. The sill parameter defines the average variance where the experimental variogram stabilizes and it is reached at the range level. The nugget effect parameter is sometimes added to the semivariogram model to represent a very short range variability in the data set and also the error of the experimental measurement. This nugget parameter is added into the sill parameter to obtain the total sill of the model. Some semivariogram models are presented in the following equations:

$$\begin{cases} \gamma_{exponential}(h) &= \text{sill.}[1 - exp(-\frac{3h}{\text{range}})] \\ \gamma_{Gaussian}(h) &= \text{sill.}[1 - exp(-\frac{(3h)^2}{\text{range}^2})] \\ \gamma_{hole-effect}(h) &= \text{sill.}[1 - cos(\frac{h}{\text{range}}.\pi)] \end{cases} \tag{8}$$

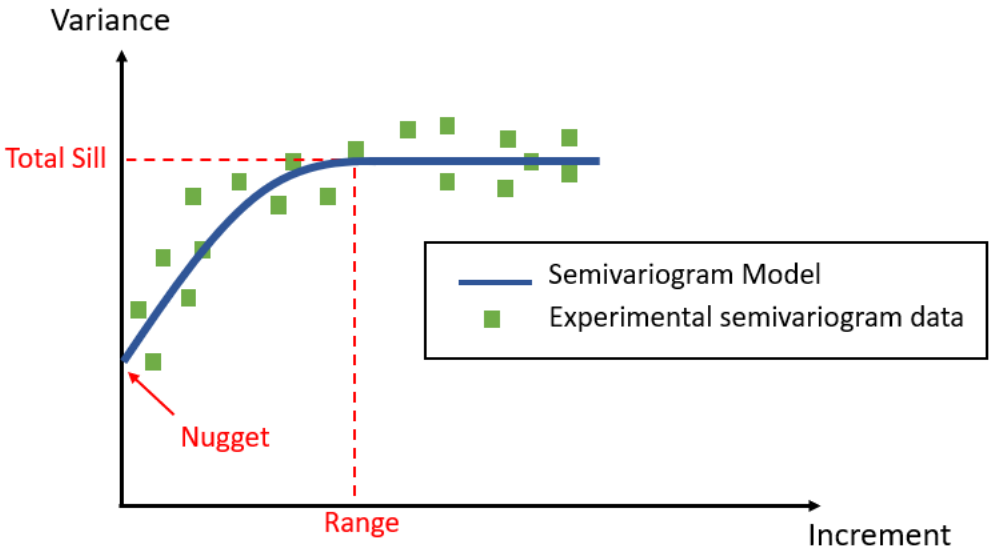


Figure 6 – Semivariogram model

For the case of spatial semivariogram, two different approaches are compared. For the first, an interpolation location u is defined by only one coordinate %OV. The influence of angular/temporal information is ignored. Thus, the interpolation is accomplished between different %OV for each degree of rotation. An extracted synchronous average of each %OV contains points values of a complete rotation of runner (0° to 360°). Therefore, this approach requires many semivariograms (number of semivariograms is equal to number of SA point values that is about 1000 points), leading to many interpolations to execute. To reduce the calculation time, only three simple models are used to fit the experimental semivariogram: Spherical, Exponential and Gaussian. Moreover, because of experimental limitations, the number of operating-conditions measured is also limited, which can affect interpolation results.

For the second approach of spatial semivariogram, the angular information θ (or time information) of measurement is taken into account as a "second spatial coordinate". An interpolation location is now defined by two coordinates, %OV and rotation angle of runner ($u_i = (\%OV_i, \theta_i)$). The distance between two locations in

this interpolation space is calculated as the Euclidian distance. However, these two coordinates do not have the same range to the spatial dimension, which can lead to an error of kriging interpolation. Thus, a normalization is required before building the semivariogram. Contrary to the first approach, this second one has more points in the scatter plot of semivariogram and requires only one semivariogram for interpolation. Therefore, to fit the experimental semivariogram, some nested variogram models are used in order to improve the performance of kriging interpolation. The nested model is a linear combination of several simple models like a combo presented in the result part: nugget model combines with a long-range Gaussian model and a short-range hole effect model (figure 7b). Since kriging uses the semivariogram model for covariances, the nested model helps to avoid the lack of information during the interpolation process.

Kriging interpolation with spatio-temporal semivariogram

In the second interpolation method, the spatio-temporal semivariogram is built with two independent dimensions: spatial and temporal/angular (respectively %OV and rotation angle of runner in this research). The half variance between two "locations" depends not only on distance increment h, but also on time/angle increment a (see equation 9). This type of semivariogram is often used in geostatistics for the case that the temporal dimension does not have the same range to the spatial one. Therefore, using this spatio-temporal semivariogram helps to independently observe the influences and the contributions of time (or angular field) to the interpolation process.

$$\hat{\gamma}(h,a) = \frac{1}{2N(h,a)} \sum_{(i,j)\in N(h,a)} [S(\%\mathrm{OV}_i, \theta_i) - S(\%\mathrm{OV}_j, \theta_j)]^2 \tag{9}$$

The experimental semivariogram is now a 3D scatter plot with an axis of half variance and two axes of distance and time increment (figure 10). For modeling covariance, semivariogram model is now presented by combining the spatial semivariogram and the temporal/angular semivariogram (there is also the joint semivariogram in some models[7].) with different mathematical operators. To fit the 3D experimental semivariogram, each semivariogram is first independently fitted by a simple model (the definition of model parameters, like sill and range, for the temporal/angular semivariogram model is similar to the spatial one presented in the first method) and they are then combined by using the spatio-temporal semivariogram model function. Separable model function [7], which is used in this case study, are presented in equation (10). This Separable model and the other spatio-temporal semivariogram models are available in code library Gstat [7]. This library is used to execute the calculation and the graphic observation of the spatio-temporal semivariogram in this study.

$$\begin{cases} C_{separable}(h,a) &= C_s(h).C_t(a) \\ \gamma_{separable}(h,a) &= \text{sill}.(\bar{\gamma}_s(h) + \bar{\gamma}_t(a) - \bar{\gamma}_s(h)\bar{\gamma}_t(a)) \end{cases} \tag{10}$$

where $C_s(h)$ and $C_t(a)$ are respectively the spatial and temporal covariance, sill is the overall sill parameter, $\bar{\gamma}_s$ and $\bar{\gamma}_t$ are respectively spatial and temporal semivariogram [7].

4 Interpolation results

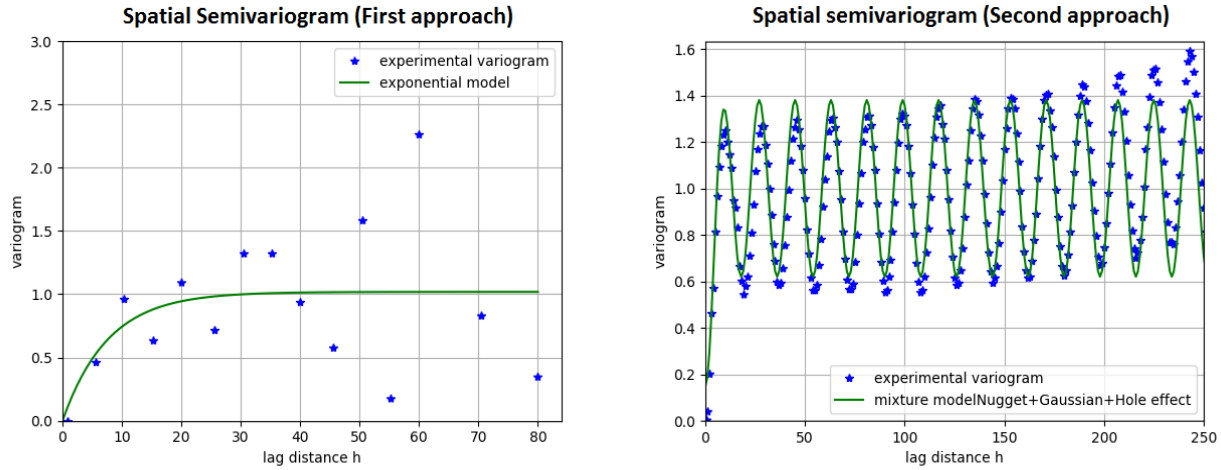
To deal with the limited experimental data, the interpolation is performed from a part of known observed operating conditions and the remaining ones are used for comparison. Input and Output data is shown in the Table 1. As mentioned in the previous section, the semivariogram model directly influences on the kriging interpolation quality. Therefore, only results of the best model are shown for each method. Using cross-validation technique helped to find out these best models. The principle of cross validation technique is to temporary remove one or several data and then to re-interpolate these data from the remaining data by kriging, this process is repeated several times (100-1000 times) to find at the end a set of correlation coefficients between real and interpolated data. Moreover, in the case of spatio-temporal semivariogram, the Mean Absolute Error (MSE) has been also calculated to verify the interpolation quality of model [7]. In the following, only two interpolation results at 20 %OV and 55 %OV are presented.

For the Spatial Semivariogram method, the interpolation results of the first approach are presented in figures 8a and 9a. These results show a low quality of interpolation because of many disturbances throughout

Input: SA(θ)	Semivariogram model		Output: SA(θ)
	Spatial Semivariogram	Spatio-temporal Semivariogram	
At 30 %OV, 40 %OV, 50 %OV, 60 %OV, 70 %OV, 90 %OV	First Approach: Exponential model Second Approach: Nugget effect + long-range gaussian model + short-range hole effect model	Separable model: +) Gaussian model (with Nugget effect) for the **spatial semivariogram** +) Gaussian model (with Nugget effect) for the **temporal semivariogram**	At 20 %OV, 45 %OV, 55 %OV, 80 %OV, 100 %OV

Table 1 – Information of input, ouput and semivariogram model used in this study paper

the angular axis (temporal axis). This observation highlights the importance of angular/temporal information in the kriging interpolation process. The high frequency noises observed in these figures confirm that the angular/temporal information must be considered in this interpolation study. Moreover, the limited experimental data for the semivariogram might also contribute to this. However, the interpolated signals are smoother with the second approach (figures 8b and 9b). This can be explained via a large number of information points taken by the nested semivariogram model (figure 7b).



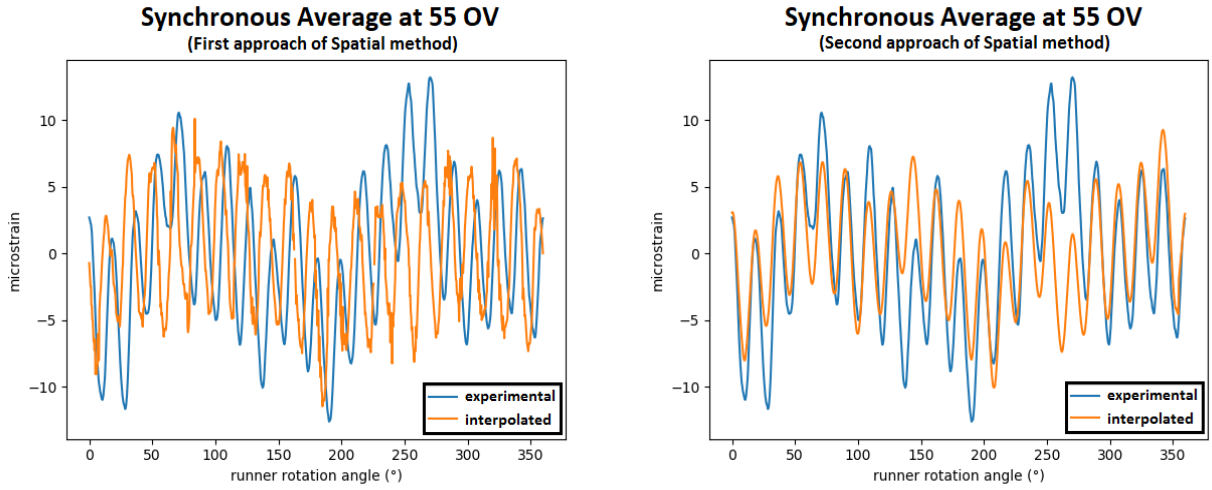
(a) (b)

Figure 7 – Experimental semivariogram and fitted model in the case of spatial method: (a) First approach; (b) Second approach

The interpolation results of the spatio-temporal semivariogram method are presented in figure 11. By graphically comparing to the spatial semivariogram method, the form and the amplitude level of the interpolated signal in this case have better correlations with the experimental signal. However, these results are not enough to confirm which process is better for the synchronous average interpolation, a proper validation is required.

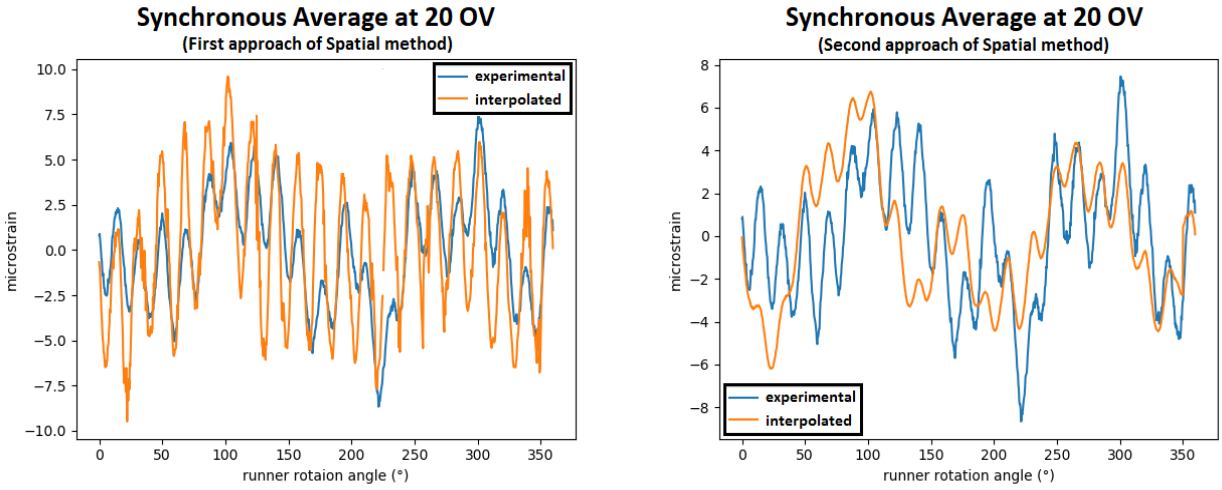
5 Discussion

In this paper, using the interpolation tool, a representation of the periodic phenomenon, varying through time and operating conditions, are generated. It is unable to foresee all variations of this phenomenon. With this kind of data, the generation of an interpolated signal that perfectly matches the experimental signal is not a requirement of this study. To verify the quality of interpolation process, the load spectrum, which is mainly used in the fatigue evaluation process, is presented. This graph, based on the Rainflow algorithm, allows the representation of the strain cycles contained in the signal as a function of cumulative number of cycles for fatigue. Figure 12 presents load spectrum curves of two kriging methods. The load spectrum curve found



(a) (b)

Figure 8 – Experimental and interpolated data at 55 %OV in the case of spatial method: (a) First approach; (b) Second approach

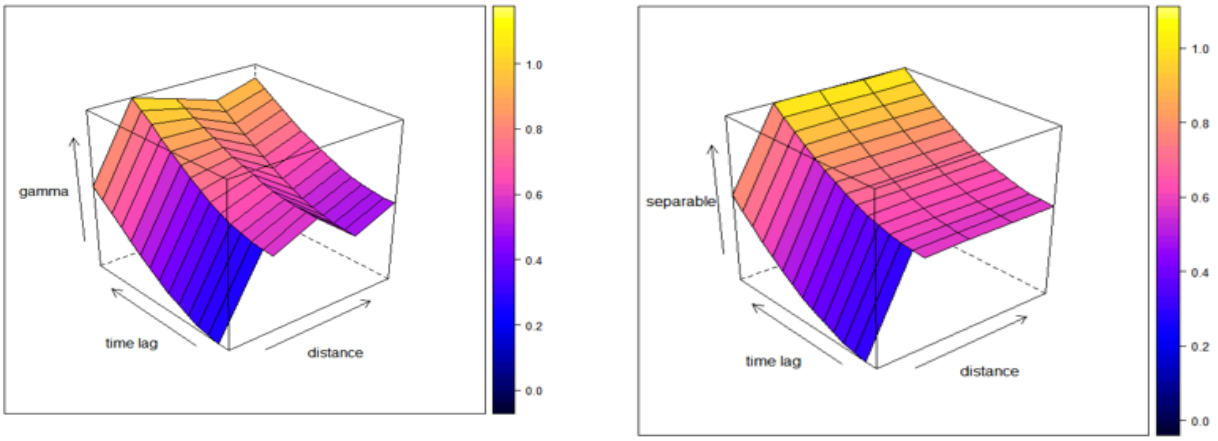


(a) (b)

Figure 9 – Experimental and interpolated data at 20 %OV in the case of spatial method: (a) First approach; (b) Second approach

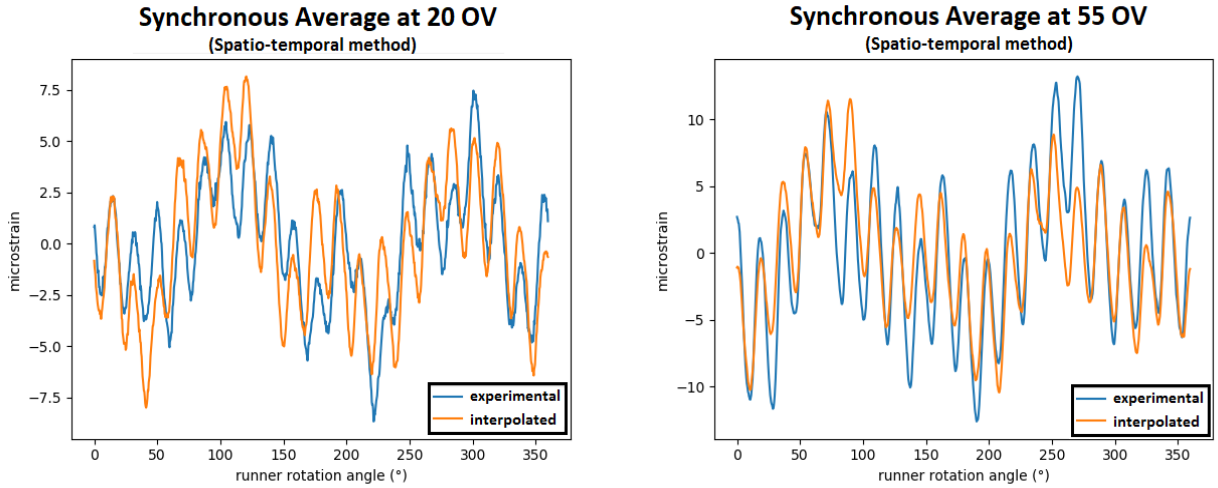
for the spatio-temporal method is clearly closer to the experimental value than for the spatial one. It can be observed from these load spectrum performances that the spatio-temporal semivariogram is better adapted this type of data.

The spatio-temporal semivariogram seems better but it is unable to conclude that this method satisfies our requirements. We need not only a unique interpolation result, but also a confidence interval. In this paper, the conditional stochastic simulation is used to generate a first assessment of the confidence interval. The conditional stochastic simulation is a statistic simulation process based on kriging and experimental data. The simulations are commonly used to correct the smoothing effect of kriging method [4][6]. The figure 13 presents 1200 conditional stochastic simulations at 20 %OV based on the Separable spatio-temporal model (see Table 1). The interpolation value and the experimental value mostly fall into the set of conditional simulations.



(a) (b)

Figure 10 – (a) Experimental spatio-temporal semivariogram; (b) Separable model used to fit the experimental semivariogram

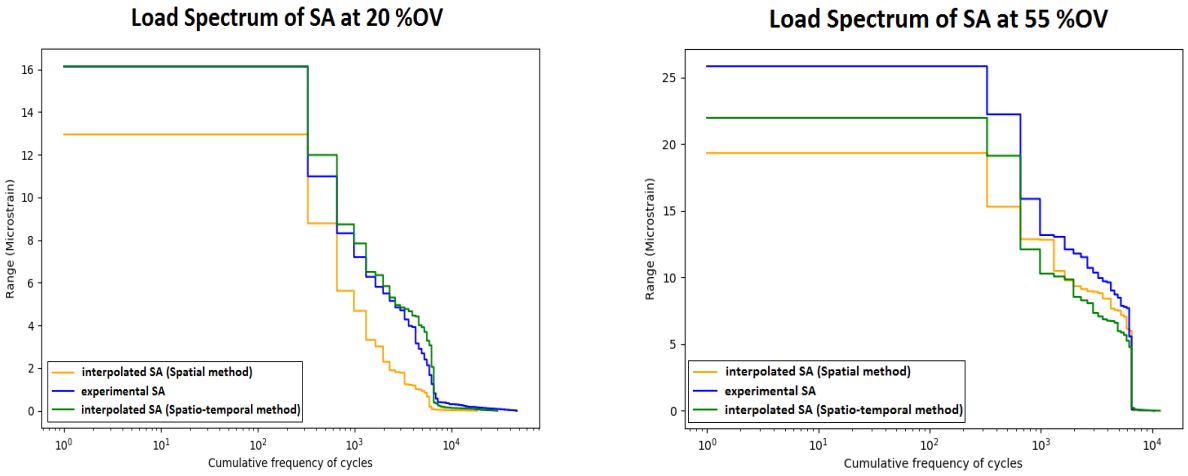


(a) (b)

Figure 11 – Experimental and interpolated data in the case spatio-temporal method: (a) 20 %OV; (b) 55 %OV

6 Conclusion

The periodic phenomenon has been extracted as the first order of cyclostationarity from the runner strain signal by applying the synchronous average operator. Even if the periodic part does not give all the needed information to assess the runner fatigue, the results in this paper set the first bases to build an interpolation tool for the whole signal. The synchronous average signals are considered as a function depending on the different %OV and the runner rotation angle in the interpolation process. The spatio-temporal kriging has a good interpolation performance for the synchronous average. Although the spatio-temporal kriging process might not be the most appropriate interpolator for other phenomena in signal, it can be a good start nonetheless.



(a) (b)

Figure 12 – Load spectrum comparison between original SA and interpolated SA of two kriging methods: (a) 20 %OV; (b) 55 %OV

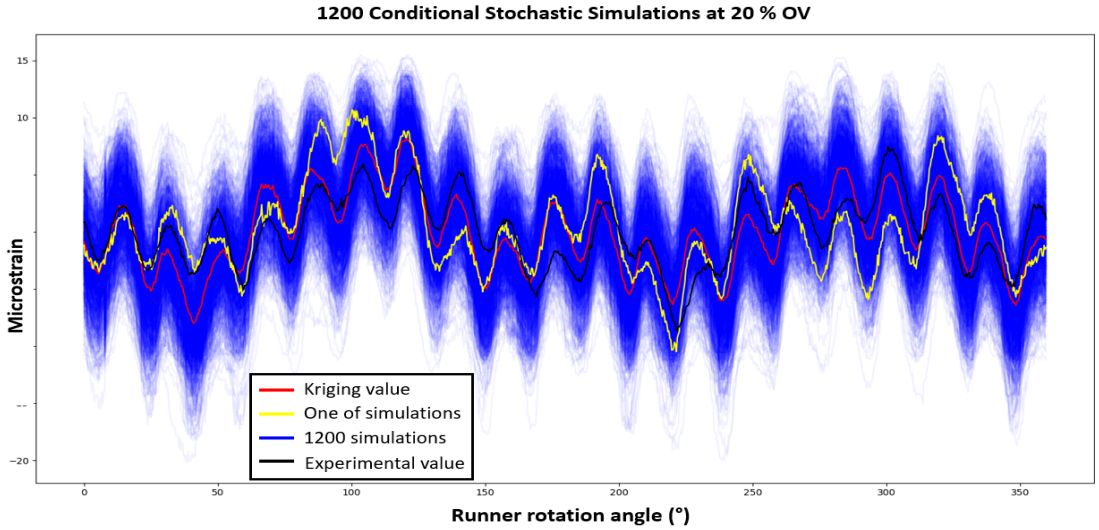


Figure 13 – Confidence interval by 1200 conditional stochastic simulations at 20 %OV

Acknowledgements

The author would like to thank the Hydro-Québec Research Institute (IREQ), the Andritz Hydro Canada LTD, the Insa de Lyon, the École de technologie supérieure Montréal (ÉTS) and the Mitacs-Accélération program for their support and financial contribution.

References

[1] M. Poirier , M. Gagnon, A. Tahan, A. Coutu, J. Chamberland-Lauzon, *Extrapolation of dynamic load behaviour on hydroelectric turbine blades with cyclostationary modelling*, Mechanical Systems and Signal Processing, 2017, Vol. 82, p. 193-205.

[2] M. Gagnon, A. Tahan, P. Bocher, D. Thibault, *The role of high cycle fatigue (HCF) onset in Francis runner reliability*, In IOP Conference Series: Earth and Environmental Science, 2012, IOP Publishing,p. 22005.

[3] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, *Cyclostationary modelling of rotating machine vibration signals*, Mechanical Systems and Signal Processing, 2004, Vol. 18, p. 1285-1314.

[4] B. Firas, *Modélisation de la propagation des incertitudes des mesures sur l'aube d'une turbine hydraulique par Krigeage et simulations stochastiques*, Research paper for École de technologie supérieure Montréal,Québec,Canada, 2014.

[5] F. Bonnardot, *Comparaison entre les analyses angulaire et temporelle des signaux vibratoires de machines tournantes. Etude du concept de cyclostationnarité floue*, Traitement du signal et de l'image. Institut National Polytechnique de Grenoble - INPG, 2004. Français. (tel-00103067)

[6] C.V. Deutsch, A.G. Journel, *GSLIB: Geostatistical Software Library and User's guide* , Oxford University Press, 1995, 8e edition.

[7] B. Gräler, E. Pebesma, G. Heuvelink, *Spatio-temporal interpolation using Gstat*, The R Journal, 2016, Vol. 8:1, p. 204 - 218.

[8] J. Antoni, *Cyclostationarity by examples*, Mechanical Systems and Signal Processing 23, 2009, p. 987 - 1036.

Adaptive tacho information estimation for surveillance of rotatory machine under nonstationary conditions

Yi Wang[1,2,3] *, Bingchang Hou[3], Baoping Tang[1,3], Yi Qin[1,3]

[1] The State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, P.R. China

[2] Industrial IOT and Networked Control Key Laboratory of the Ministry of Education, Chongqing University of Posts and Telecommunications, Chongqing 400065,China
[3] College of Mechanical Engineering, Chongqing University, Chongqing 400044, P.R. China

*Corresponding author, E-mail: wycqdx@cqu.edu.cn

Abstract

Rolling bearing faults are the leading causes of downtime in rotary machines. In recent years, numerous and various vibration-based approaches have been put forwarded for rolling bearing fault detection. In the vibration-based techniques, order tracking-based methods are considered as very effective techniques. In the current reported order tracking methods, auxiliary devices are essential to obtain the instantaneous angular speed (IAS) of the machine. Aiming at this shortcoming, estimating IAS from vibration signals has been studied and some tacho-less order tracking (TLOT) techniques have been put forwarded. However, the effectiveness of the current available TLOT algorithms rely on the manually selection of the initial parameters for IAS estimation, which bring about user-friendliness. In order to tackle the aforementioned obstacles, a novel adaptive tacho information estimation method based on nonlinear mode decomposition (NMD) is proposed. In the proposed method, the nonlinear mode decomposition (NMD) method is improved and its computational burden is reduced. And then, the tacho information is adaptively estimated. The vibration signal collected from an aircraft engine is used for signal analysis and the effectiveness of the proposed is successfully validated.

1 Introduction

Rolling element bearing (REB) is especially important part of a rotatory machine, it usually works with variable load and speed, hence, failure is more possible happened here [2][3]. The failure of REB not only decreases machine working efficiency, but may also causes enormous lost in some extreme circumstances. In this way, it is of vital significance to detect the REB failure.

Vibration signal analysis is a kind of classic tool to detect machine fault [4], for the reason that vibration signal usually owns sufficient operating state information, it is viable to detect fault from it. However, most of the now existing methods are based on the assumption of stationary operating condition, which are impractical for accurate fault detection. To obtain a better fault detection results, the order tracking (OT) method [5][6] which is very suitable for speed variation condition is proposed and been widely studied. OT method transfers the non-stationary vibration signal in time domain into cyclostationary vibration signal in angular domain by resampling the non-stationary vibration signal with uniform angular increment, and fault can be accurately estimated from the order spectrum. Before resampling, the instantaneous angular speed (IAS) is needed. The conventional method uses auxiliary equipment [7], such as tachometer, to acquire the IAS information.

However, the auxiliary equipment brings extra cost and is not convenient in all conditions, and the tacho-less method is rapidly needed and catch considerable researchers' attention [8]. Zhao et al. [9] proposed a generalized demodulation based tacho-less OT method, this method combined the advantages of tacho-less OT method and envelope order spectrum, and could detect the REB fault under variation speed

condition. To tackle with the inaccuracy of phase information estimation for OT method in large speed variation condition, Schmidt et al. [10] put forward a tacho-less OT methodology which is based on probabilistic approach, and the effectiveness is validated by both simulated and experimental signals. Sound signal also includes the REB healthy state information, Lu et al. [11] proposed a sound analysis-based method for bearing fault detection under speed fluctuation condition, the bearing fault can be uncovered from envelope spectrum of resampled signal successfully. In our previous study [12], a tacho-less OT tracking method which acquires better ISA information from adjacent vibration signal is proposed and can discover compound fault on wind turbines. Hu et al. [13] proposed an adaptive tacho-less OT method which is based on enhanced empirical wavelet transform (EEWM), this method introduces EEWM to analyse the characteristics of resampled signal and is relatively robust to noise.

Though tacho-less OT method shows great advantage to detect REB fault under speed fluctuation condition and researchers have make a lot progress about it, because of the unclear TFR of analysed signal caused by background noise and irrelevant components in the OT method, expertise knowledge and must be needed to extract the IAS information for resampling, which cannot ensure the detection accuracy and is inconvenient for real industry application. Hence, an adaptive IAS calculation strategy is urgently needed for tacho-less OT method. The Empirical Mode Decomposition (EMD) based fault detection method [14] is a kind of adaptive method and extracts mon-component via numerical approximation. However, because the phenomenon of spectral mixing among different modes caused by spectral overlaps between different components, biased IAS information for resampling and inaccuracy fault detection results may be obtained in some circumstance. Therefore, a more reliable adaptive IAS estimation strategy for tacho-less OT method in REB fault detection needs to be developed.

Aiming at the abovementioned requirement in the current existed tacho-less OT method for REB fault detection under speed variation condition, this paper proposed a new adaptive IAS information estimation strategy which is on the basis of Nonlinear Mode Decomposition (NMD) [15]. NMD is a hybrid product which combined the TFA and surrogate test, it shows great application prospect and could adaptively decompose a mixed signal into a set of mono-component which possesses clearly physical meanings. Our main works in this paper can be summarised as follows. First of all, the computation efficiency of NMD is meliorated to fit the high sampling frequency of the analysed vibration signal. Then, an adaptive IAS information estimation strategy is raised with the utilization of NMD. Last but not least, a noise-robust adaptive tacho-less OT method for REB fault detection under speed variation conditions is proposed, in which the expertise knowledge for IAS information extraction is not need and a broad industry application can be seen.

The remaining paper is organized as follows. In section 2, the theory of improved NMD is introduced and the detailed implemented procedures of our proposed method are given. In section 3, the validity verification of proposed method is given by the field test data experiments. And a conclusion is drawn in section 4.

2 The theory of meliorated NMD and implementation of the adaptive tacho-less OT method

In order to adaptively calculate the IAS information for resampling in a tacho-less OT method, the NMD theory is introduced in our study and elaborated in this section. And then, the realization process of the corresponding tacho-less OT method for REB fault detection is given.

2.1 The improved NMD theory for adaptive extraction of IAS information

As mentioned before, the EMD based IAS information extraction strategy is not robust in some cases, and an advanced technique is urgent to developed for tacho-less REB fault detection. The NMD is a newly developed method integrating with the advantage of parameterized TFA and surrogate test, it could decompose a complex signal into a series of physically meaningful oscillations. Therefore, it's possible to adaptively extract IAS information using NMD.

The whole frequency range in TFR of analysed signal is searched in original NMD method, for a high sampling vibration signal, this is not necessary and takes long time. Hence, the NMD method must make

some improvement in calculation speed for adaptive IAS information extraction. The detailed illustration of our improvement on NMD method is given in our former works [16], and the basic thinking is as follows. For the reason that the interested harmonics won't exist in the entire frequency range, and frequency searching range can be restricted to a special area decided by some equations which takes the peaks of TFR into consideration. With calculation efficiency of NMD method improved, the procedures of adaptive IAS information extraction strategy can be drawn like this.

Step 1: obtain the TFR of the REB vibration signal via Short-Time Fourier Transform (STFT).

Step 2: using Fourier transform surrogates test method to distinguish the extracted reference component from the noise.

Step 3: extract the subharmonics of the reference component, and investigate the consistency of the subharmonics by time-shifted surrogate.

Step 4: confirm the fundamental harmonic and find its possible higher order harmonics through consistency test.

Step 5: obtain the fundamental signal and its harmonics for further analysis.

Repeat step 3 to extract all of its harmonic signals from the TFR plane. Continue to perform the above steps on the residuals of analysed signal until a stopping criterion is satisfied. For details of the NMD method, please refer to [15].

2.2 The realization process of NMD based tacho-less OT method for REB fault detection

The meliorated NMD method are shown in last subsection, and the realization process of NMD based tacho-less OT method for REB fault detection can be summarized as follows.

Procedure 1: Utilize the meliorated NMD method to adaptively calculate the IAS information from the vibration signal.

Procedure 2: Resample the analysed signal with the obtained IAS information. In this way, this non-linear and non-stationary signal is transformed into cyclostationary signal in angular domain.

Procedure 3: Figure out the order spectrum with conventional means, such as Hilbert demodulation and spectral kurtosis.

Procedure 4: Recognize the fault type from order spectrum by analysing fault characteristic order.

3 The performance of the proposed method for aircraft engine tacho information estimation

3.1 An Overview of the Investigated aircraft engine

In this section, a difficult industrial case data is used to validate the effectiveness of the proposed method. The signals are acquired during a ground test campaign on a civil aircraft engine with two damaged bearings. The data is provided by Safran contest, Conference Surveillance 8, October 20–21, 2015, Roanne, France [1].

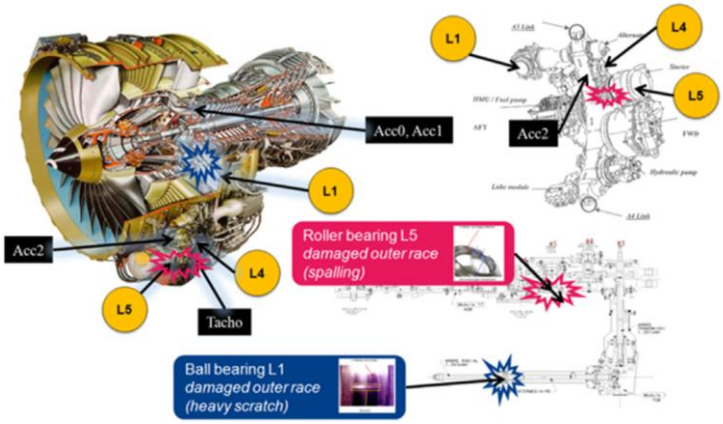


Figure 1: General overview of the engine and the accessory gearbox. [1]

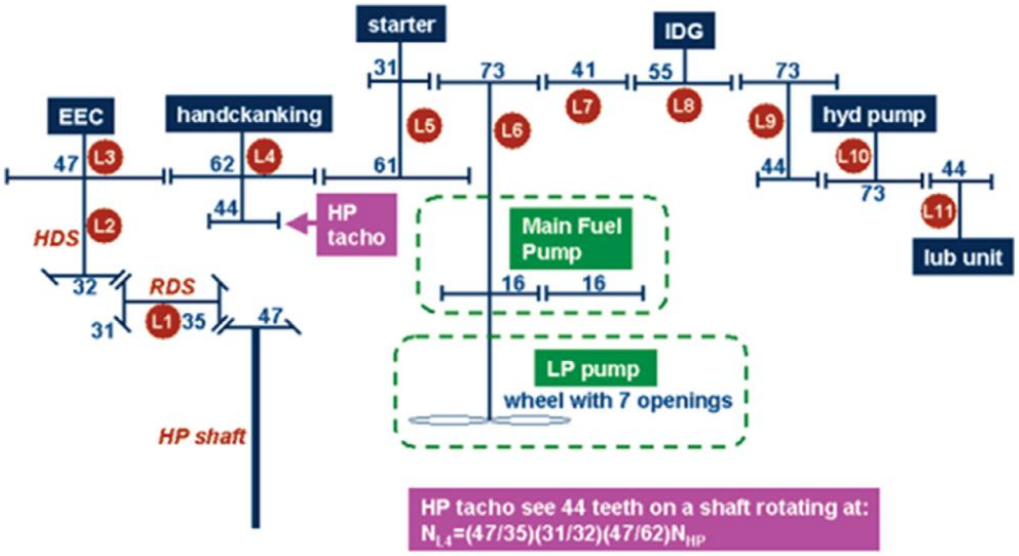


Figure 2: kinematics of the aircraft engine gearbox. [1]

An overview of the investigated engine and its accessory gearbox is demonstrated in Figure.1, and the kinematics of the accessory gearbox is depicted in Figure.2. The data used in section is collected by three sensors, named "Acc1", "Acc2" and "Tacho" in Figure.1, respectively. The sampling rate during signal collection process is 50 kHz and the sampling length is 200 s. The sensor "Tacho" is used to collect the rotating speed signal from shaft L4. One of the accelerometers is located on the intermediate case near the radial drive shaft and the other one is on the flange of the accessory gearbox in the vicinity of shaft L5.

3.2 The experimental results obtained by the proposed method

The partially enlarged drawing of the TFD of the original noise signal is shown in Figure 3, in which some frequency components are apparent. To conduct tacho-less order tracking on aircraft engines, it is very essential to extract one certain harmonic which is mono-component of the fundamental or higher harmonics of the rotating speed signal. In the conventional tacho-less order tracking techniques, the beginning of the interested harmonic is manually selected according to the fluctuation trend of IRF ridges in TFD. As a result, the current techniques are not applicable when the expertise knowledge is not available. To address the shortcomings encountered by the conventional techniques, the proposed method based on NMD is applied and its performance is demonstrated as follows.

Firstly, the dominant component with much higher energy in TFR is extracted as a reference component $x_r(t)$ by ridge detection, and the corresponding instantaneous amplitude $A(t)$, phase $\phi(t)$ and frequency $f(t)$ are reconstructed. And then, Fourier transform surrogate test against null hypnosis of noise is conducted to check whether the extracted signal is a true component. Totally 40 surrogates are created, the significance D_s of each surrogates and significance D_0 of the extracted component as depicted in Figure 4. All of surrogates with $D_s > D_0$, it indicates that the extracted component is true, therefore the null hypothesis of noise is rejected and continue the decomposition.

Further, the extracted dominant component $x_r(t)$ is assumed to be the fundamental one, and subsequently, time-shifted surrogate test against null hypothesis is conducted to investigate the independence between the extracted component $x_r(t)$ and its subharmonic candidates. The surrogate signals with a number N_s=20 are generated. The consistence $\rho_d^i(1,1,0)$ of the time-shifted surrogates and $\rho_0^i(1,1,0)$ of the candidate subharmonic with zero time shift $VT_0 = 0$, are calculated as shown in Figure 5. For the candidates subharmonic of $x_r(t)$, all of the its time-shifted surrogates' consistency values are lower than the threshold, i.e. $\rho_d^{1/2} < \rho_{\min}$, which is equal to 0.25 [15]. In this circumstance, all of the candidate subharmonics are identified as false, because it does not pass the time-shifted surrogate test. Therefore, the extracted component $x_r(t)$ is taken as the fundamental harmonic of the rotating speed signal and regarded reference component for nonlinear mode extraction a step further. Similarly, based on the time-shifted surrogate test, the higher order harmonics are investigated and are also tested against the null hypnosis of noise. As a

consequence, the second order harmonic passed the surrogate test and confirmed as a true component.

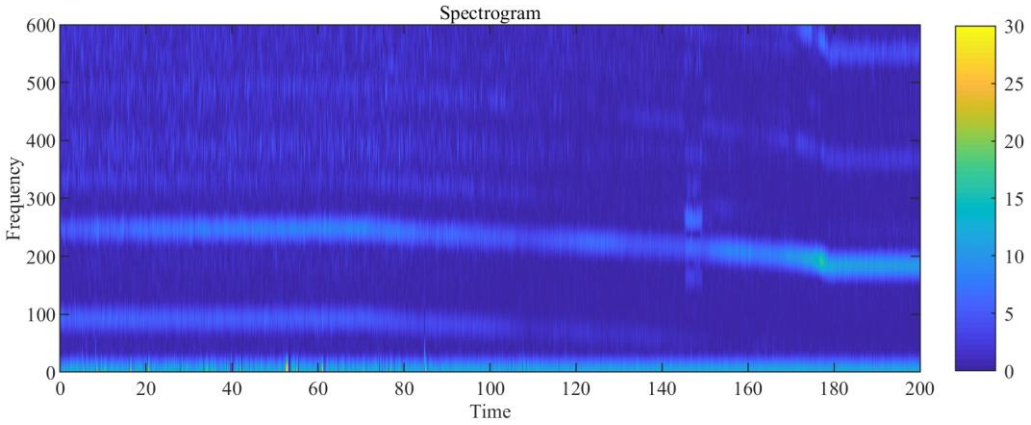


Figure 3: The partially enlarged drawing of the TFD of the original signal.

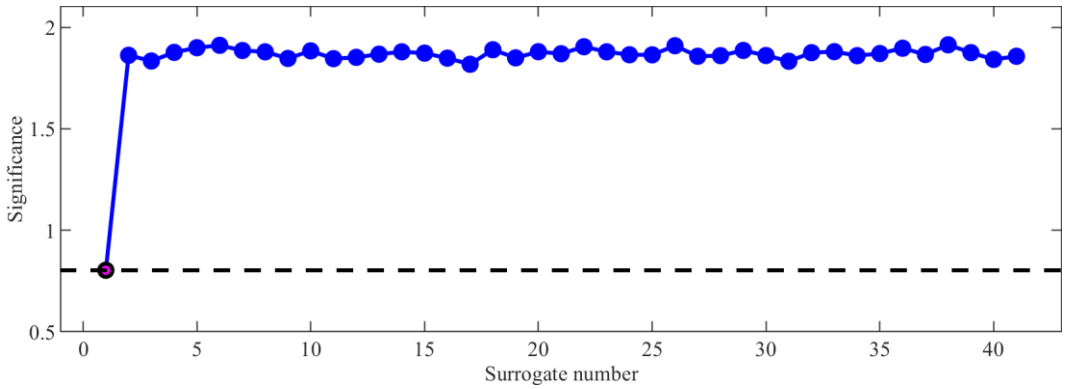


Figure 4: The Fourier transform surrogate test.

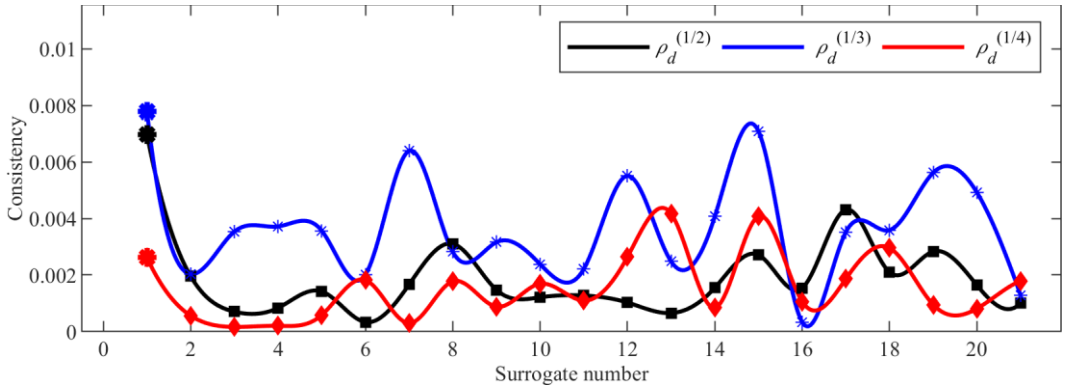


Figure 5: The consistency of the candidate sub-harmonics obtained by time-shifted surrogate test.

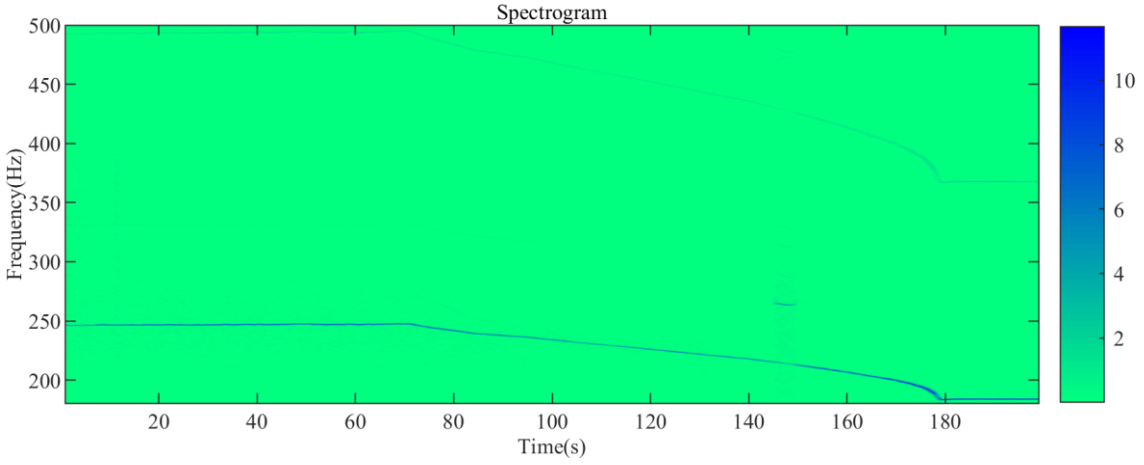


Figure 6: The extracted component in TFD domain.

The extracted rotating fundamental signal and its second order harmonic signal in TFD domain are depicted in Figure 6. Furthermore, the extracted fundamental component of the rotating speed signal is presented in Figure 7, while the extracted second order harmonic is depicted in Figure 8. On the basis of the extracted mono-component of the fundamental signal, the instantaneous frequency of the aircraft engine is calculated as shown in Figure 9. When compared with the instantaneous frequency estimation result reported in [1], the instantaneous frequency ridge is adaptively obtained by the proposed method successfully characterized the true fluctuation trend of the rotating speed.

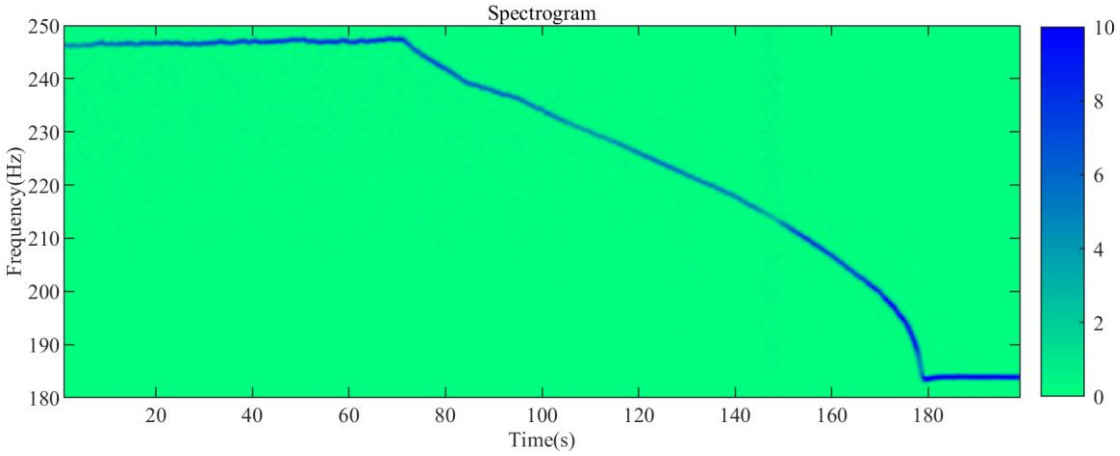


Figure 7: The extracted fundamental component of the rotating speed signal.

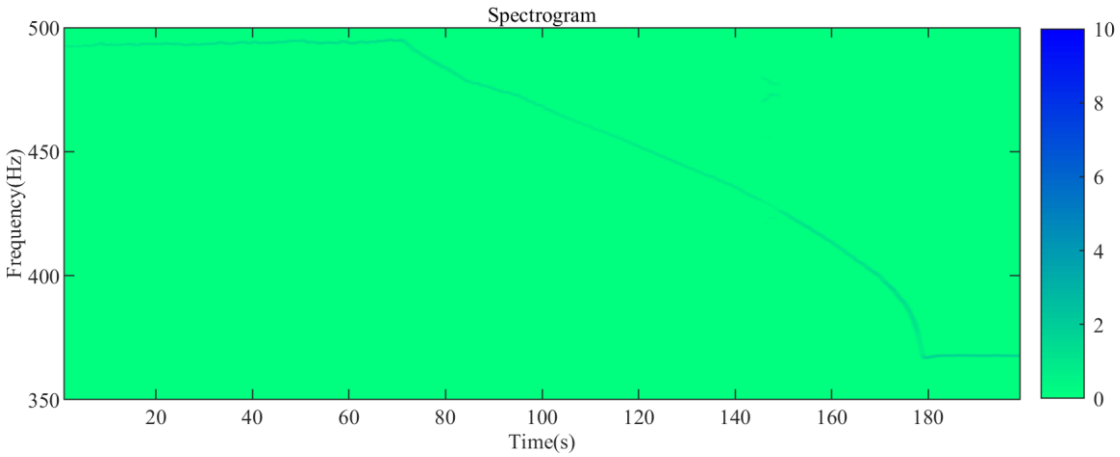


Figure 8: The extracted second harmonic of the rotating speed signal.

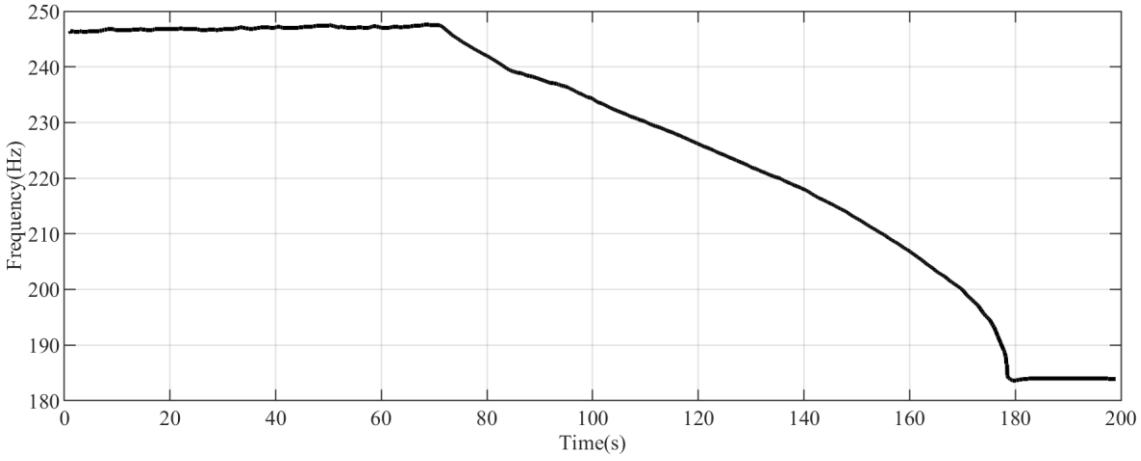


Figure 9: The estimated instantaneous frequency of the rotating speed signal.

4 Conclusions

To overcome the shortcomings of the conventional techniques for tacho information estimation under non-stationary operating conditions, a novel method based on improved NMD method is proposed in this paper. The merit of NMD method, which can adaptively separate mono-component from non-stationary signals and determine the fundamental harmonic, is inherited in our proposed method. The original NMD method is improved and the computational burden is reduced to make it applicable for aircraft engine vibration signal processing. On this basis, the shaft speed signal is adaptively extracted by improved NMD method and the tachometer information of the entire drivetrain is obtained. The effectiveness and improved features of the proposed are demonstrated by real aircraft engine vibration signals. The validation results indicate that the proposed method is more flexible for tacho information estimation, and provide a promising tool for fault diagnosis of rotating machines operating under nonstationary conditions.

5 Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No.51805050), the Fundamental Research Funds for the Central Universities (Project No.2018CDXYJX0019) and the Foundation from Industrial IOT and Networked Control Key Laboratory of the Ministry of Education, China (Project No.2018FF03), which are highly appreciated by the authors. Special thanks should be expressed to the bearing faults dataset of a civil aircraft engine supplied by SAFRAN Contest, Conference Surveillance 8, October 20–21, 2015, Roanne, France.

References

[1] J. Antoni, J. Griffaton, H. André, et al., *Feedback on the Surveillance 8 challenge: Vibration-based diagnosis of a Safran aircraft engine*, Mechanical Systems and Signal Processing, 2017 vol. 97, pp. 112-144.

[2]Idriss El-Thalji, Erkki Jantunen, *A summary of fault modelling and predictive health monitoring of rolling element bearings*, Mechanical Systems and Signal Processing, 2015 Aug. vol. 60-61, pp.252-272.

[3]Robert B. Randall, Jérôme Antoni, *Rolling element bearing diagnostics—A tutorial*, Mechanical Systems and Signal Processing, 2011 Feb. vol. 25, pp.485-520.

[4]Tianyang Wang, Qinkai Han, Fulei Chu, Zhipeng Feng, *Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review*, Mechanical Systems and Signal Processing, 2019 Jul. vol. 126, pp.662-685.

[5]Yu Guo, Ting-Wei Liu, Jing Na, Rong-Fong Fung, *Envelope order tracking for fault detection in rolling element bearings*, Journal of Sound and Vibration, 2012 Dec. vol.331, pp.5644-5654.

[6]P. Borghesani, R. Ricci, S. Chatterton, P. Pennacchi, *A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions*, Mechanical Systems and Signal Processing, 2013 Jul. vol. 38, pp.23-35.

[7]K.S. Wang, P.S. Heyns, *The combined use of order tracking techniques for enhanced Fourier analysis of order components*, Mechanical Systems and Signal Processing, 2011 Apr. vol 25, pp.803-811.

[8]Ming Zhao, Jing Lin, Xiufeng Wang, Yaguo Lei, Junyi Cao, *A tacho-less order tracking technique for large speed variations*, Mechanical Systems and Signal Processing, 2013 Oct. vol. 40, pp.76-90.

[9]M. Zhao, J. Lin, X. Xu, and Y. Lei, *Tacholess envelope order analysis and its application to fault detection of rolling element bearings with varying speeds*, Sensors, 2013 Aug. vol. 13, pp.10856–10875.

[10]S. Schmidt, P.S. Heyns, J.P. de Villiers, *A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process*, Mechanical Systems and Signal Processing, 2018 Feb. vol. 100, pp.630-646.

[11]Siliang Lu, Xiaoxian Wang, Qingbo He, Fang Liu, Yongbin Liu, *Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals*, Journal of Sound and Vibration, 2016 Dec. vol 385, pp.16-32.

[12]Bingchang Hou, Yi Wang, Baoping Tang, Yi Qin et al., *A tacholess order tracking method for wind turbine planetary gearbox fault detection*, Measurement, 2019 May. vol. 138, pp.266-277.

[13]Yue Hu, Xiaotong Tu, Fucai Li, Hongguang Li, Guang Meng, *An adaptive and tacholess order analysis method based on enhanced empirical wavelet transform for fault detection of bearings with varying speeds*, Journal of Sound and Vibration, 2017 Nov. vol. 409, pp.241-255.
[14]W. Teng, F. Wang, K. Zhang, Y. Liu, X. Ding, *Pitting fault detection of a wind turbine gearbox using empirical mode decomposition*, Strojniški vestnik-J. Mech. Eng, 2014 Jan. vol.60, pp.12-20.
[15]D. Iatsenko, P.V.E. McClintock, A. Stefanovska, *Nonlinear mode decomposition: a noise-robust, adaptive decomposition method*, Phys. Rev. E., 2015 Oct. vol. 92.
[16]Y. Wang, B. Tang, L. Meng and B. Hou, *Adaptive Estimation of Instantaneous Angular Speed for Wind Turbine Planetary Gearbox Fault Detection*, IEEE Access, 2019 Dec. vol. 7, pp.49974-49984.

Diagnostics and Dynamic models

Characterization of a Bouc-Wen model-based damper model for automobile comfort simulation

Hanwei GAO[1,2], Louis JÉZÉQUEL[1], Eric CABROL[2], Bernard VITRY[2]

[1]LTDS, Ecole Centrale de Lyon, UMR5513, 69134, Ecully Cedex, France
hanwei.gao@ec-lyon.fr
[2]Technocentre Renault, 78280, Guyancourt, France

Abstract

Ride comfort is considered as one focus for a chassis system. To ensure a satisfying comfort performance of a vehicle in development, a detailed damper characterization needs to be pre-defined in the early phase of project with the help of simulation results in virtual proving ground. However, the current damper model integrated in whole vehicle simulations is sometimes difficult to fit to test results due to its over-simplifications especially in low speed excitation regimes. Thus this article proposes an enhanced shock absorber model to improve simulation predictions without increasing substantially calculation costs. The parameterized model is mainly based on a modified Bouc-Wen model considering its capability of reproducing highly nonlinear hysteretic phenomena. In order to identify the parameters, a multi-objective optimization using NSGA-II algorithm has been applied. The excitations on test bench have been separated into several groups according to their force-velocity curve shapes. Finally an optimum set which represents the compromises between the objectives is obtained and forms a Pareto front. Comparative examples in virtual proving ground show that the correction quality is well improved for chassis' comfort prediction using the proposed model. This example demonstrates the effectiveness of the modelling and its potential in comfort improvement with the help of design of experiments.

1 Introduction

Ride comfort is one of the focuses of vehicle suspension design. To ensure that the comfort performance has been satisfied, the technical specification has been defined at the early stage of new vehicle projects and then validated after the chassis system is designed. With the development of numerical technologies, the prediction of ride comfort at an earlier phase of project is becoming important because several specifications need to converge at the same time and the unsatisfied comfort specification may cause the retreat of the designing choices during validation stage.

One main difference between simulations for ride comfort and simulations for other specifications such as chassis durability or dynamic behaviour is that the low velocity regime of the shock absorbers is more important since it plays a critical role to filter small road oscillations [1] . In order to improve the quality of prediction by comfort simulations, a detailed damper model which is capable of reproducing the physical damping forces around zero velocity is searched. In virtual proving ground simulation, several damper models are proposed.

The classical way is to define a one-to-one look-up table between force and velocity according to the measurements. A single curve with several slopes is easy to define at the early development stage and this model requires the least calculation resources. So it is commonly used in durability test track simulations, where the acceleration peaks with high velocity will cause more damage to chassis components. However, the single curve model may be over-simplified in low velocity regime where hysteresis force tends to 'open' the damping curve. As a result, more details should be extracted from the measurements and included in the model.

The second way is to construct a semi-physical damper model through a combination of several sub-models such as spring, damper or friction parts. These parts can be expressed in analytical expressions by partial-derivative equations, which cost larger but still acceptable calculation resources. The main difficulty of application is to find the parameters of these sub-models which cannot be derived directly from physical construction of components. Different identification methods have been developed based on the experimental data ([2–5]) . In [4], it has been shown that this model is suitable for real-time hardware-in-the-loop simulations.

The third way is to construct a physical damper model which aims at simulating the phenomena in real components, such as the circulation of oil. The physical parameters such as the piston diameter and valves' opening pressures need to be identified basing on real shock absorbers. However, the influence of these parameters on damping curves may be indirect and thus it is hard to be predefined in early development phases. This kind of model is usually provided by shock absorber suppliers as a black-box model. In [6, 7], a model established in AMESim has been proposed and it can be integrated into virtual proving ground by co-simulation. It is theoretically the most precise but takes the longest time to solve.

An alternative way is to identify a data-driven model without concerning its physical characters [8]. The method is efficient when the physical modelling is too complex to be identified or some specific characters may be neglected because of its complexity. However, as in virtual ground simulation, it is preferable to keep some physical relations into shock absorbers as a reference for further designing.

In this paper, the semi-physical method is selected as it is the compromise between the single-curve model and the physical model in terms of configuration process and calculation costs. Another advantage of the functional model is that it is easy to reconstruct each sub-model in MSC ADAMS/Car environment without introducing a third software for co-simulation, which makes it more user-friendly to deploy the method in vehicle projects. Before being integrated into a full-vehicle simulation, the damper model is first identified by a multi-objective optimization algorithm, which is also used in [9, 10] for a Bouc-Wen model. However, based on the measurements it is shown that a set of parameters is not capable of satisfying all the excitations with different frequencies or velocities for a given shock absorber. By contrast, by separating the objectives into several groups, a set of non-dominated optimums can be obtained and then the model parameters can be decided according to the type of frequency needed.

Part 2 of this paper summarizes the friction models, especially Bouc-Wen one. Part 3 introduces the Bouc-Wen based damper model and its identification process using a multi-objective optimization plan. Part 4 applies the proposed model into a full vehicle comfort simulation, which indicates that the new model can produce better test correlation than a single damper curve.

2 Hysteresis models

A vehicle suspension shock absorber is based on a mono-tube or bi-tube technology. One of the difficulties is its non-linear hysteresis response which is non-negligible in low velocity regime. Different models have been proposed to model the hysteresis phenomenon which is a combination of the effect of oil viscosity and interior friction. In Renault, the measurement procedure of a shock absorber is realized by a series of sinus excitations with different amplitudes and frequencies. Figures in Table 1 have shown the strong non-linearity character of a measured shock absorber. The force-velocity curve may be very different with the same damping velocity but different excitation frequencies. To model this phenomenon, several sub-models are introduced.

2.1 Coulomb friction model

This is the basic friction model. The friction force changes direction with the velocity, but its absolute value is a constant. The mathematical equation can be expressed as

$$f = -sign(v)F_{Coulomb} \tag{1}$$

where the signum function $sign(x) = \begin{cases} 1 & x > 1 \\ -1 & x < -1 \end{cases}$, v is the damping velocity, $F_{Coulomb}$ is the constant friction force. However, the discontinuity at the zero velocity may cause numeric problems in simulation [4].

2.2 Dahl model

The Dahl friction model can solve the discontinuous problem of Coulomb model by introducing two more variables σ_0 and i. It was first proposed by P.R Dahl in 1969 [11]. The equations of the model are

$$\begin{aligned} f &= \sigma_0 u \\ \dot{u} &= v \cdot sign(1 - sign(v)\tfrac{\sigma_0 u}{F_c})|1 - sign(v)\tfrac{\sigma_0 u}{F_c}|^i \end{aligned} \tag{2}$$

where u is a state variable, F_c is the maximum friction force which equals to the constant of Coulomb model [12]. σ_0 defines the stiffness of model and it can be translated to the slope of the curve when $F = 0$. When $\sigma_0 \to \infty$ the model will approach the Coulomb model. i models the shape of the curve at the beginning of the velocity change. The normalized friction force under different σ_0 and i are showed in Figure 1.

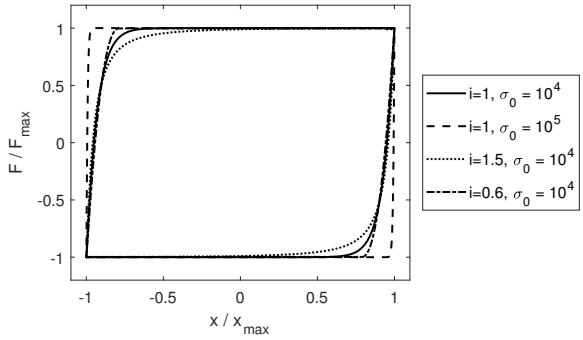


Figure 1 – Comparison of different parameters settings of i and σ_0 of Dahl model

2.3 Bouc-Wen model

The Bouc-Wen model gives a smooth description of the process from the elastic deformation to sliding between the contact surfaces. It was proposed by Bouc in 1967 and then developed by Wen. The equations of Bouc-Wen model [13] are

$$\begin{aligned} F &= cv + kx + f - f_0 \\ f &= \alpha z \\ \dot{z} &= \delta v - \beta v|z|^n - \gamma z|v||z|^{n-1} \end{aligned} \tag{3}$$

where F is the total damping force and f is the hysteresis force. x is the relative displacement of two ends of the shock absorber and v is the damping velocity. c is the viscous coefficient. k is the stiffness of the damper and f_0 is preload. Each part of the total damping force is described in Figure 2. In the hysteresis sub-model αz, z is a state variable. α defines the maximum hysteresis force which is equivalent to F_c in Dahl model. δ defines the linear stiffness of the hysteresis part. β, γ define its non-linear stiffness and damping character. n has an influence on the smoothness of the system. Figure 3-5 show the influence of each parameters.

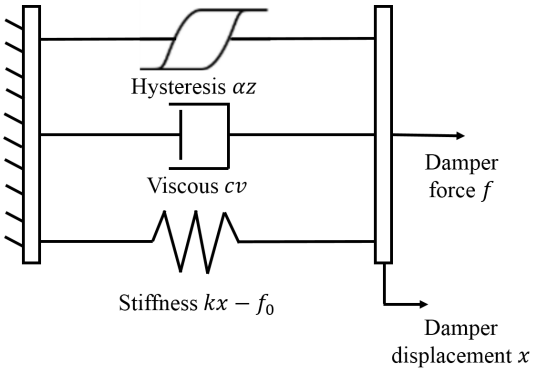


Figure 2 – Construction of the Original Bouc Wen model

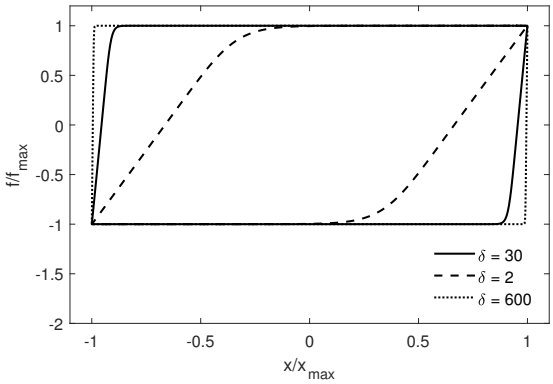


Figure 3 – Influence of δ on Damper force-displacement curve

Bouc-Wen model can reproduce different shapes under various excitation frequencies (see Figure 6). In contrast, there are more parameters to be identified than Dahl model:$c, k, \alpha, f_0, \delta, \beta, n$.

In addition, Ref [14] has extended the model to better correct non-symmetric hysteretic responses around zero velocity. It is called "shifted hysteresis".The hysteresis part of (3) can be rewritten as the function of v

$$\dot{z} = (\delta - (\beta + \gamma sign(zv))|z|^n)v \tag{4}$$

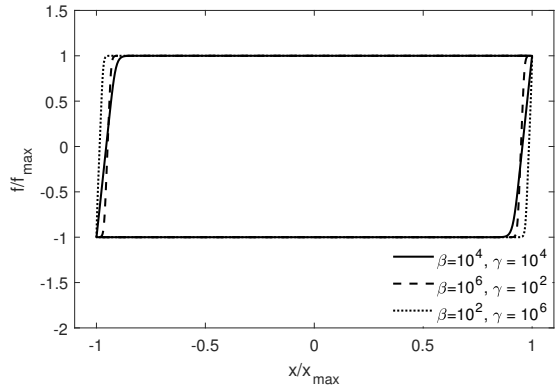


Figure 4 – Influence of β and γ on Damper force-displacement curve

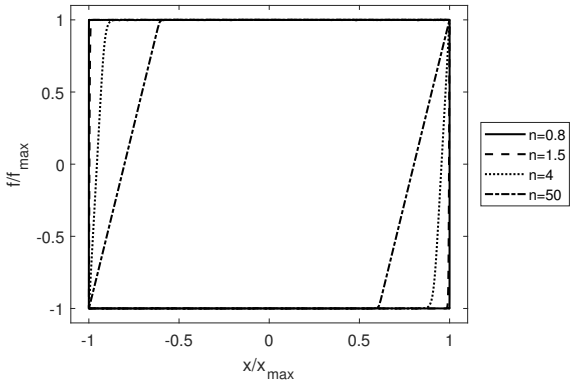


Figure 5 – Influence of n on Damper force-displacement curve

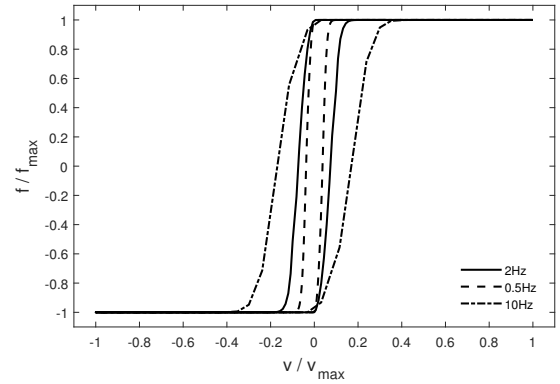


Figure 6 – Hysteresis curve under different excitation frequencies

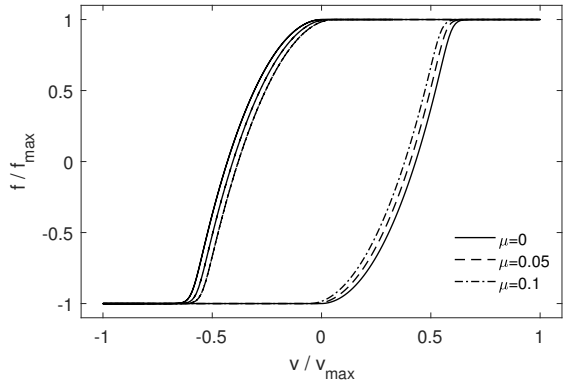


Figure 7 – Influence of μ on Damper force-displacement curve

then v is replaced by a shifted velocity which is defined as

$$v \leftarrow (v - \mu \cdot sign(v)) \tag{5}$$

where μ is one additional parameter to be identified. The main effect is to modify the hysteresis curve around zero velocity without changing that of higher velocity (see Figure 7).

3 Parameter identification process

3.1 Model description

In Equation (3) and Figure 2, the original Bouc-Wen model is composed of several sub-models: spring part $(kx + f_0)$, viscous part (cx) and hysteresis part (αz). In this article, a modified Bouc-Wen model where the spring part and viscous part have been redefined is proposed. It serves as a template of which the parameters are to be identified basing on measurements. The construction of the model is shown in Figure 8. The whole modified Bouc-Wen model has been firstly established in Simulink for identification before being integrated into the vehicle model in virtual proving ground simulation environment in MSC Adams/Car.

3.1.1 Viscous model

The viscous part normally takes the majority of total damper force, and it is defined as a function of damping velocity. In this model, the linear relation has been replaced by a continuous function created from a look-up table which groups the maximum damping velocity measured at compression and rebound phases with the maximum force. Figure 9 shows an example of a test curve, where a strong non-linearity is observed. The

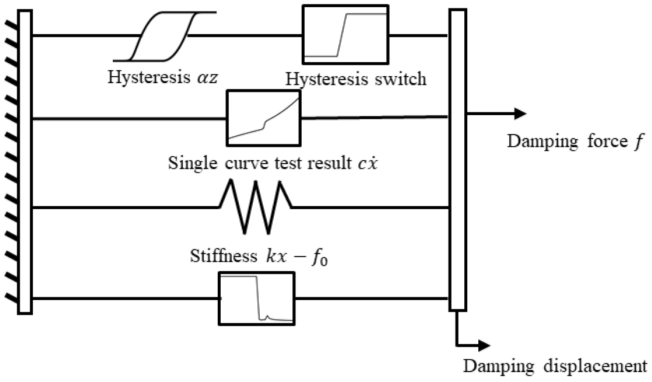


Figure 8 – Construction of a modified Bouc-Wen model

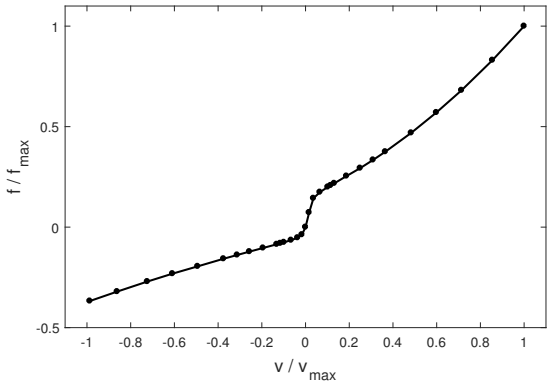


Figure 9 – Example of a single damping curve

viscous model serves as a basic line for the damper force which is previously used as the damper model in virtual proving ground simulation. By linear interpolation, the look-up table can translate the test data with an acceptable accuracy and good calculation efficiency. However, it considers only the peaks and troughs in the measured curves and thus the transformation information between the two is lost. The following sub-models aim to further exploit the test results.

3.1.2 Spring model

When the shock absorber is under excitation of very low velocity (for example, $< 0.01m/s$), the force of gas compression becomes dominant compared to viscous force. The shock absorber behaves like a spring. The spring model is to model the gas compression character of a shock absorber where the force depends on the damping displacement. In the original Bouc-Wen model, this force is linear to the displacement with a slope of k and a preload f_0. Ref [4] states that sometimes the linear model fails to perfectly represent the gas force and the quadratic gas force model can solve this problem. The equation of the spring model thus can be expressed as

$$F_{gas} = Ax^2 + Bx + C \tag{6}$$

where A, B, C are the spring model parameters. In the test progress, C is the preload after installation. A and B can be found with quasi-static excitations. But in some practical cases, the quasi-static test is absent because the test bench is not capable of producing a very low velocity excitation. The gas force is then considered as linear $(A = 0)$ with a slope calculated by two measurements of displacements. In this example, A still holds to zero as it is neglectable for small damping displacements and B and C are considered with a certain tolerance, thus are also involved into the group of parameters to be identified.

3.1.3 Hysteresis model

The Hysteresis model is the identification focus of a Bouc Wen model. From the measurement curve in Table 1 and in other test results, hysteresis phenomenon is more significant in rebound phase than in compression phase. In compression phase, the curve keeps 'closed' when velocity is larger than $0.2m/s$, which is difficult to be reproduced by an original Bouc-Wen hysteresis sub-model. An additional switch has been added after the hysteresis model which aims to forcedly 'switch off' the hysteresis for large compression velocities. The modified hysteresis curve is showed in Figure 10.

3.1.4 Compensation curve

The compensation part aims to cancel the double-counted contribution of hysteresis model to the maximum damping force because the saturation viscous part has already involved in the viscous model. The hysteresis model is first characterized individually under several sinus excitations in order to identify the maximum Bouc-Wen force under excitation velocities. Then compensation curve is defined by another look-up table.

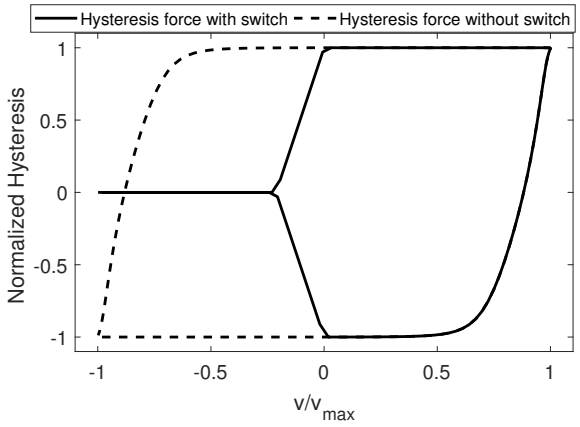


Figure 10 – The hysteresis force-velocity curve after passing the switch

3.2 Identification process

3.2.1 Optimization objectives

The model identification is to find the set of parameters which can produce the most similar damper character curves. The process can also be regarded as a multi-objective optimization plan aiming to minimize the quadratic errors between the model and the tests.

The algorithm schema of identification plan is showed in Figure 11. The same sinus signals are inputs to a physical shock absorber as well as the numeric model. The outputs of two models are differenced to obtain error signals, which are then post-treated to single indicators. The quadratic error of correction to i-th excitation (e_i) be defined as

$$e_i(\mathbf{x}) = \frac{1}{N}\sum_{j=1}^{N}(\varepsilon_i(\mathbf{x},t_j))^2 = \frac{1}{N}\sum_{j=1}^{N}(F_i(\mathbf{x},t_j) - \hat{F}_i(t_j))^2 \tag{7}$$

where $F(\mathbf{x},t)$ is the force sample at time t obtained from the numeric model with a vector of parameters $\mathbf{x}$. $\hat{F}_i(t_j)$ is the force curve from the test. $t_1, t_2..t_N$ are the sample points on $F(\mathbf{x},t)$ or $F_s(t)$. N is the total number of sample points. The optimization objective is to minimize $e(x)$. However, it should be noted that $e_i(x)$ contains not only the correction error but also disturbances such as the measurement noises on the test bench, so there is no meaning to pursue a model with zero correction errors which will result in an over-fitted problem.

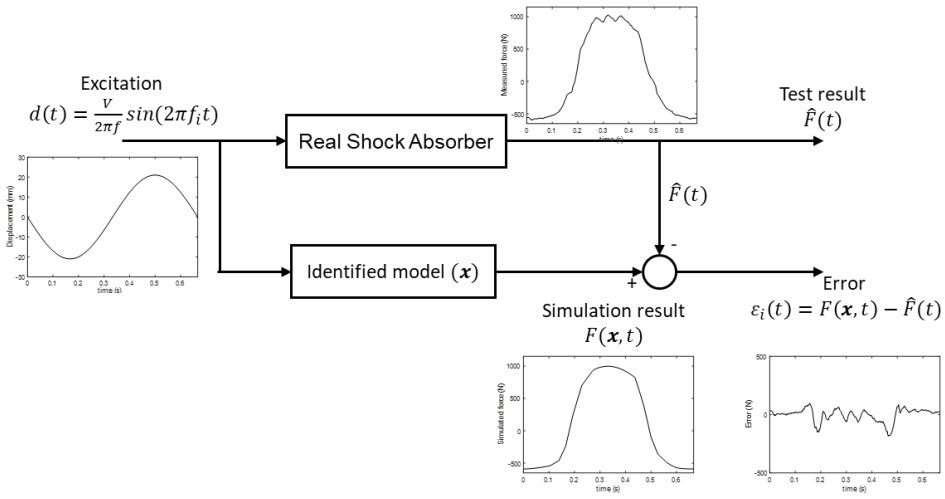


Figure 11 – The hysteresis force-velocity curve after passing the switch

A classical objective of identification is to minimize the sum of all the correction errors as shown in (8).

$$e_{total}(\mathbf{x}) = \sum_{i=1}^{M} e_i(\mathbf{x}) \tag{8}$$

where M is the total number of test excitations. The first attempt to find the best correction of the sum of all the excitations is not satisfying due to the damper's highly dissimilarity under different excitations in Table 1. The reason is that by summing all the errors the optimization has already a preference to a set of excitations. It is difficult to define a normalization factor of each objective at the beginning of the optimization which leads to satisfying results. Furthermore, the simple sum of all the objectives may result in an incomplete exploration of the designing space [9].

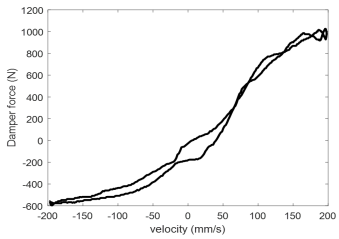
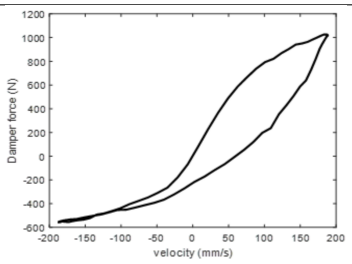
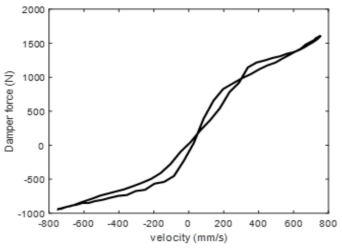
Group	LVLF	LVHF	HVHF
Velocity	0.1-0.4m/s	0.1-0.3m/s	0.5-1.2m/s
Frequency	1.5Hz	12Hz	12Hz
Nb. Signal	6	3	4
Curve example			

Table 1 – Repartion of Excitaion signals into several groups

Another approach is to separate the correction errors into several objectives when it is hard to find one dominating parameter set which minimizes all the errors. Thus, the test curves are attributed into several groups according to their curve similarities as well as the frequency regimes. Three groups have been defined for the shock absorber studied in Table 1: low velocity low frequency (LVLF), low velocity high frequency (LVHF) and high velocity high frequency (HVHF).

1. LVLF Group involves 6 excitations with maximum velocities smaller than 0.4m/s at the excitation frequency of 1.5Hz. In this group small hysteresis phenomenon is observed with a relatively closed curve and by the same time some irregular crossovers of the force may take place, which may be caused by the small measurement noise.

2. LVHF Group involves 3 excitations with maximum velocities smaller than 0.3m/s and excitation frequency is 12Hz. The shape of the test curve contains more hysteresis forces comparing to the same velocity excitations with 1.5 Hz. Here the highly dissymmetric hysteresis is also observed in the compression regime ($v < 0m/s$).

3. HVHF Group involves 4 excitations with maximum velocity between 0.5 and 1.2m/s at 12Hz. There is less hysteresis around zero velocity comparing to HVHF group and even less at higher velocities. So the shape of the curves is approaching a single curve model.

Then the objectives are defined as the average of the normalized mean quadratic errors in each of the groups. As the damper curves in each group are similar to each other so the sub-objectives are compatible.

$$\begin{aligned} e_{LVLF}(\mathbf{x}) &= \tfrac{1}{6}\textstyle\sum_{i=1}^{6} e_i(\mathbf{x})/e_i(\mathbf{x_0}) \\ e_{LVHF}(\mathbf{x}) &= \tfrac{1}{3}\textstyle\sum_{i=7}^{9} e_i(\mathbf{x})/e_i(\mathbf{x_0}) \\ e_{HVHF}(\mathbf{x}) &= \tfrac{1}{4}\textstyle\sum_{i=10}^{13} e_i(\mathbf{x})/e_i(\mathbf{x_0}) \end{aligned} \tag{9}$$

where $e_i(x)$ is the correction error defined in (7). It has been normalized by the correction results of the starting parameter sets x_0.

3.2.2 Optimization algorithm

To solve the multi-objective problem, a set of solutions which forms a Pareto front is to be searched. The Pareto front consists the non-dominated optimums where there exists no other solution that can improve one objective without degrading at least one another. A vector of parameters x^* can be defined as one Pareto optimum mathematically if and only if

$$\begin{aligned} &\forall x \in \varepsilon, \forall j \in \{1,...,m\}, f_i(x) < f_j(x^*) \\ &\Rightarrow \exists i \neq j, f_i(x) > f_j(x^*) \end{aligned} \tag{10}$$

With iterations the solution can converge to a potential Pareto front and all the points included in the front are optimized solutions. The Pareto front will collapse to one single point if it can minimize all the objectives and, in this case, the multi-objective optimization is equivalent to a mono-objective one.

The genetic optimization algorithm NSGA-II has been applied after the objectives and the design space are defined. NSGA-II can be viewed as a stochastic search method which explores the design space and selects the parent generation according to the fitness ranks of each existed solution to generate a new generation.

The optimization starts with an initial design of experiments to establish a relationship between the design space and the objectives by response surfaces. These response surfaces serve as the prediction functions from which a set of potential optimums is proposed before being validated by simulation. For the next iteration the former solutions will pass to a potentially better generations by selection, crossover and mutation operations [15]. The optimum sets proposed by the algorithm should finally be converged to a Pareto front. Figure 12 resumes the procedure of this optimization plan.

4 Identification results

4.1 Design space

In this chapter the identification model has been carried out on the shock absorber described in Table 1 as an example. All the designing parameters have been arranged in the vector **x**

$$\mathbf{x} = \{A, B, C, \alpha, \beta, \gamma, n, \mu\} \tag{11}$$

The initial, minimum and maximum values of these parameters are defined in Table 2. In this example, the gas pression forces are modelled linearly. Thus A remains to be zero.

	Initial value	Minimum Value	Maximum Value
A	0	0	0
B	-0.25	-0.5	-0.1
C	-150	-300	0
α	120	50	320
δ	30	10	600
$log(\beta)$	5.78	4	6
$log(\gamma)$	4.3	2	6
n	1.5	1.5	8
μ	0	0	0.001

Table 2 – Definition of design space

So the mathematical expression of the multi-objective optimization problem can be expressed as
Minimize:

$$F(x) = \{e_{LVLF}(x), e_{LVHF}(x), e_{HVHF}(x)\} \tag{12}$$

Under the constraints:

$$x \in DesignSpace \tag{13}$$

$F(x)$ is the target function used in generic optimization algorithm. Its compositions are the mean quadratic error of each group which defined in (9).

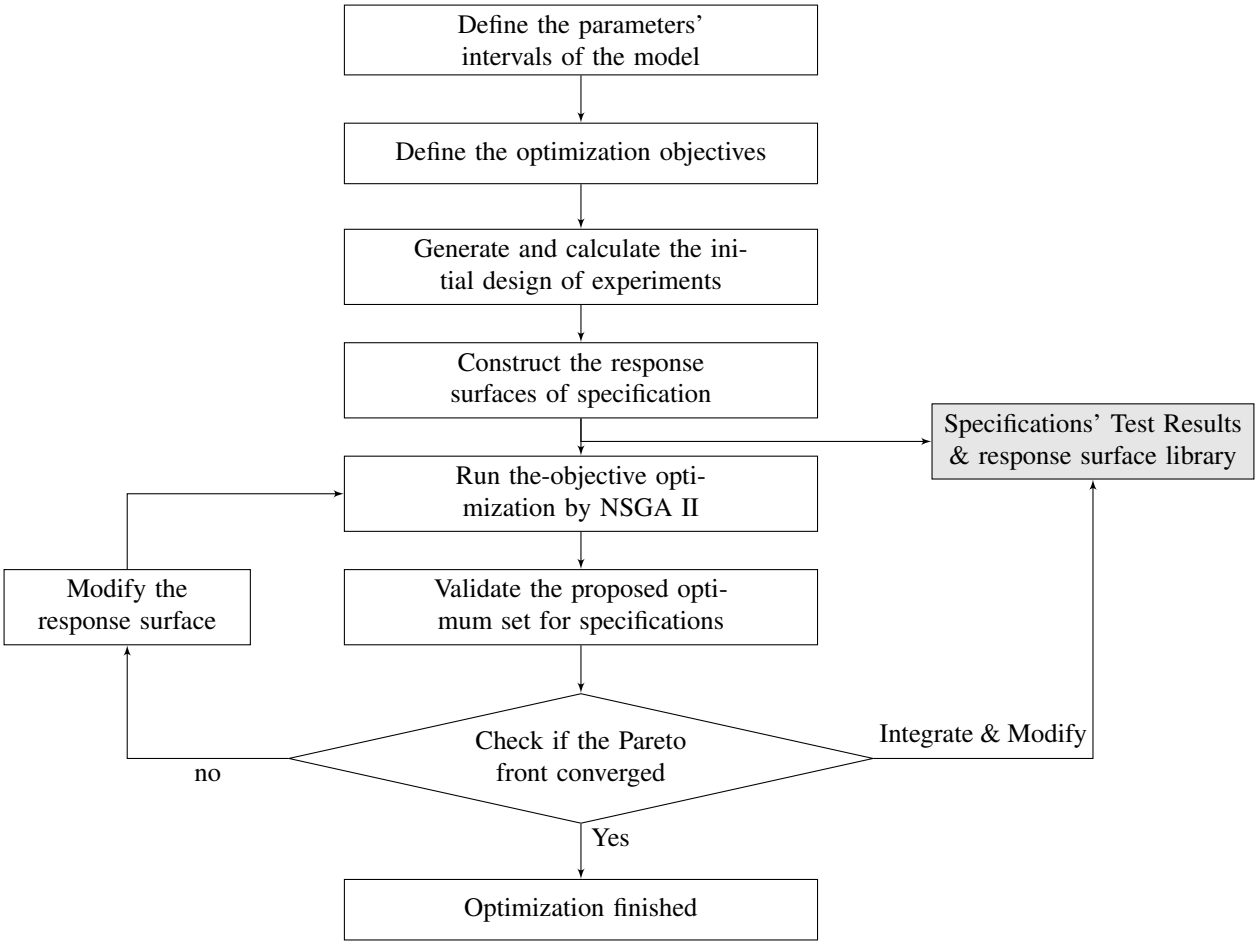


Figure 12 – Optimization process without robustness notions

4.2 Pareto fronts and surfaces

After 9 iterations, the correction objectives of non-dominated solutions are shown two-by-two in the Figure 13 and 14. The figure shows the correction objectives of three groups are incompatible especially between two different frequencies. Due to the complexity of the real shock absorber, the better correction to the low frequency signals may fit worse to high frequency ones and vice versa.

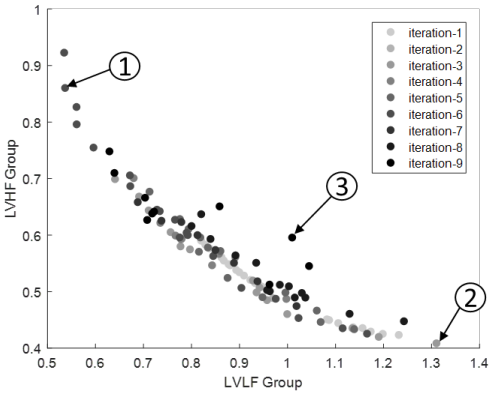


Figure 13 – Non-dominated solutions in LVLF-LVHF plane

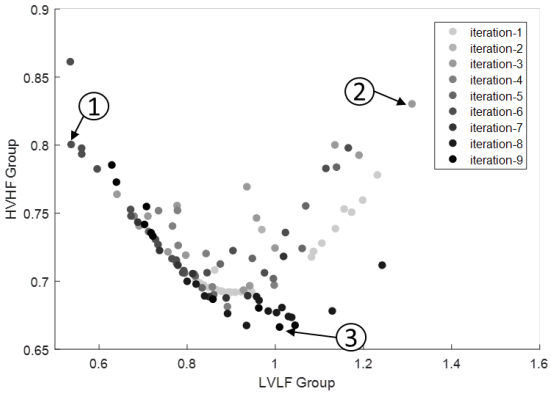


Figure 14 – Non-dominated solutions in LVLF-HVHF plane

Figure 15 shows the projection of the 3D Pareto surface on the LVHF-HVHF plan. The darker colour means better correction to LVLF group. This figure explains the reason why there are points far from the optimum front in Figure 13 and 14: The solution presented on one 2D Pareto front may not be involved in another 2D

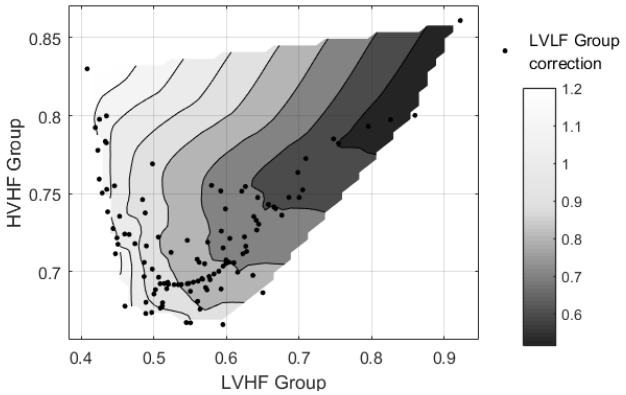


Figure 15 – Projection of Pareto surface on in LVHF-HVHF plot

front, and thus it is hard to reach an optimised correction for all the three objectives at the same time.

Table 3 shows the selected parameters of three points numbered in Figure 13 and 14, which are on different Pareto front but with different orientations. Table 4 shows one typical curve in each group for the three selected points in Table 3. With the parameters of point 1, the excitation of low frequency is well fitted and thus they are more applicable to the comfort test track. Point 2 and Point 3 are oriented to high frequency so they have worse corrections to low frequency comparing to Point 1.

	Point 1	Point 2	Point 3
A	0	0	0
B	-0.1	-0.1	-0.1
C	-150.5	-117	-146
α	85	265	179
δ	50	600	600
$log(\beta)$	4	4.33	5.6
$log(\gamma)$	2.65	2	4.26
n	5	7.5	8
μ	0	0.002	0
e_1	0.53	1.31	1.01
e_2	0.86	0.40	0.59
e_3	0.80	0.83	0.66

Table 3 – Resume of three points on Pareto Front

4.3 Post-treatment of Pareto front

Figure 16 shows the choice of α and n for the 15 best corrections to each group on Pareto front. Table 1 shows that the hysteresis force is relatively small for low frequencies, so the NSGA II proposed a smaller α comparing that to LVHF group. n has also a great impact on identification results and a smaller n is preferred for the LVLF group. It can be noticed that the interaction of parameters of Bouc-Wen model is not negligible which explains that several solutions with a smaller n also fit well with HVHF groups. These regimes can be served as reference data for the further robust analysis.

5 Comfort simulation

To validate the identified shock model, a comfort simulation in virtual proving ground has been realized. The full vehicle assembly has been posed on a virtual 4-cylinder test bench which provides 4 vertical degrees of freedom [16]. A simultaneous sweep frequency sinus signal (0.5Hz-6Hz) is defined as the displacement of 4

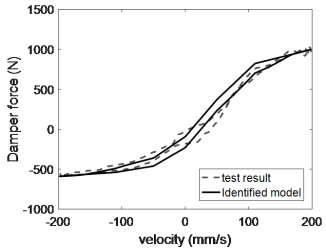
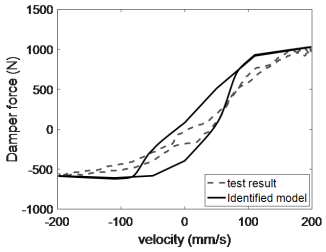
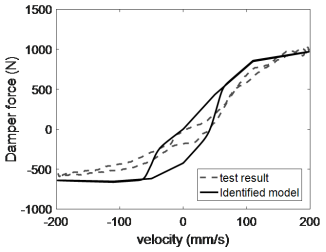
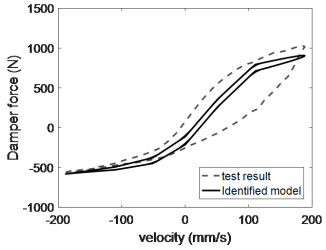
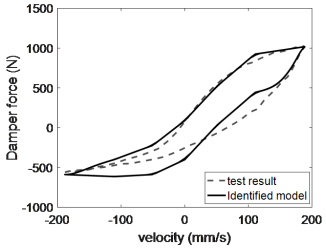
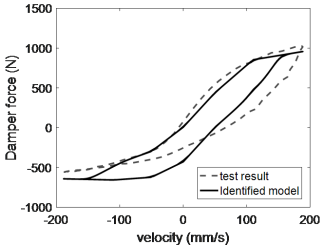
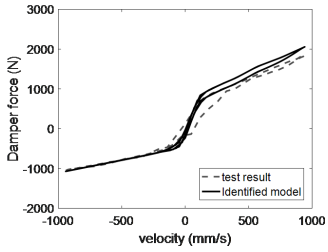
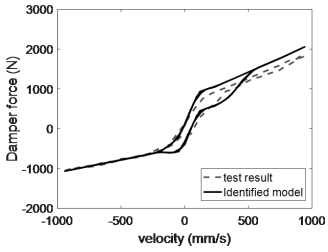
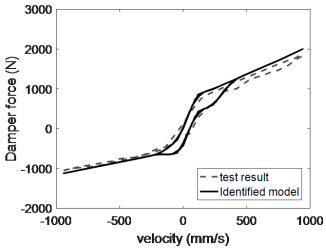
	Point 1	Point 2	Point 3
LVLF			
LVHF			
LVLF			

Table 4 – Correction of test results from choice of different orientations (Point 1: Best correction for low frequency. Point 2 and 3: Best corrections for high frequency)

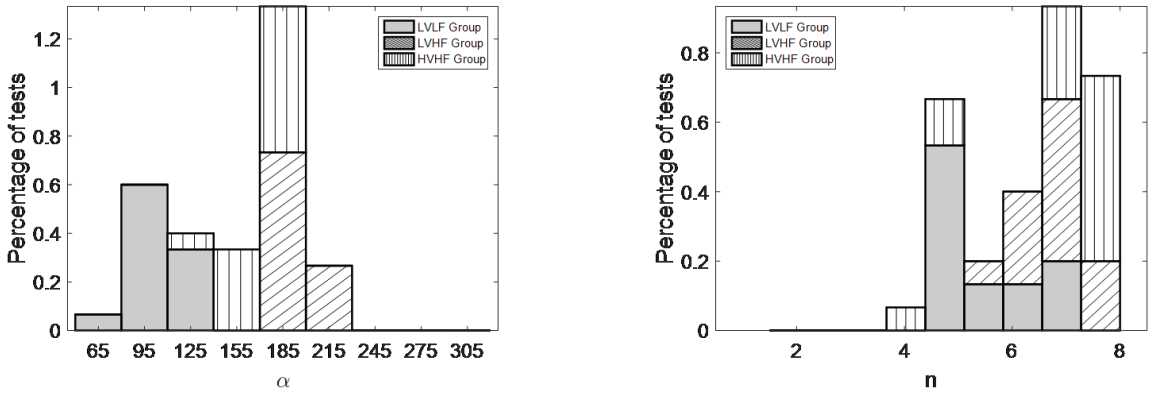


Figure 16 – Resumes of α (left) and n (right) of 15 best corrections in each group

cylinders (see Figure 17, 18).Two transfer functions of displacement from front or rear car body to the cylinder displacement have been calculated and serves as a criteria for ride comfort specification.

The best correction sets of low velocity regime and low frequency group of the modified Bouc-Wen model have been chosen for front and rear suspensions and are integrated into the full vehicle simulation. At the same time, the comparative simulations are launched with single curve models. The transfer function can be seen in Figure 19. With the modified Bouc Wen model, there is no great correction to the amplitude around the proper frequencies of front and rear suspensions. However, at the beginning phase of the signal, a richer model containing the hysteresis force can improve the quality of simulation especially for the rear suspension. The damper force at the first 20 seconds has shown that the shapes are very different for small excitations but after that the damper force from two models becomes almost identical.

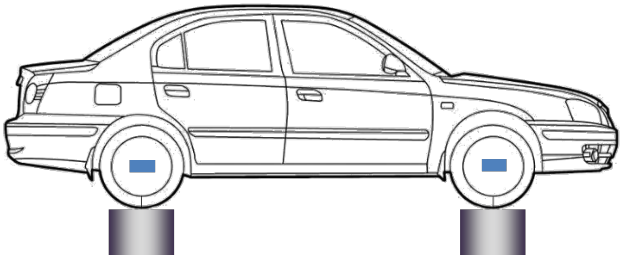


Figure 17 – The full vehicle comfort test bench

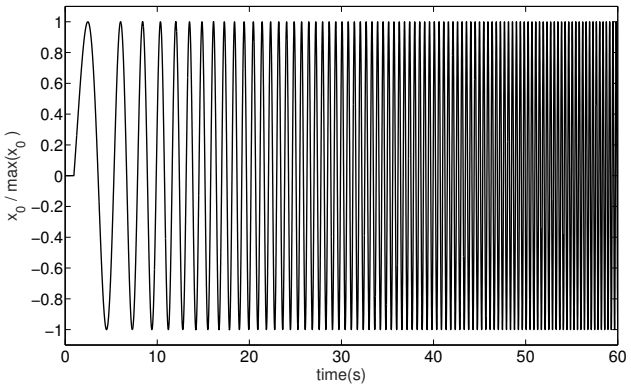


Figure 18 – One part of excitation signal

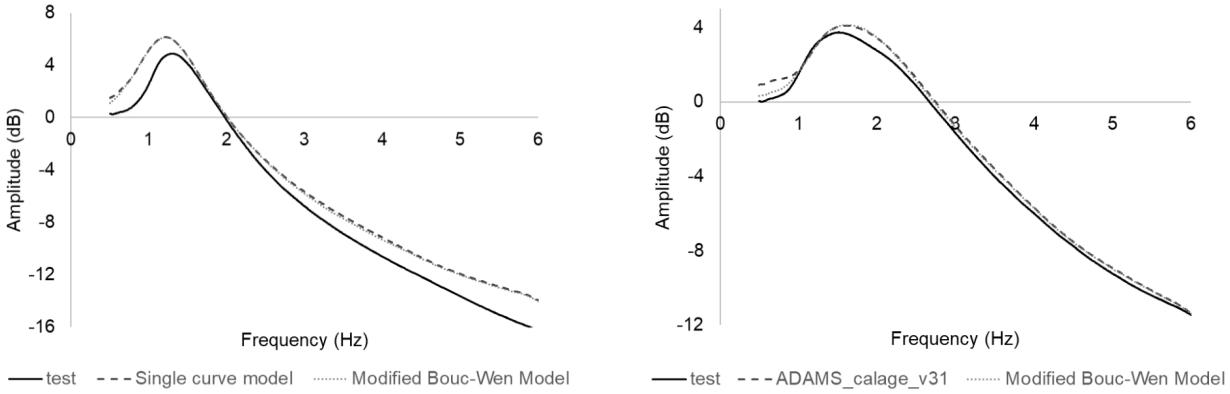


Figure 19 – Comparison of transfer function for front (left) and rear (right) suspensions

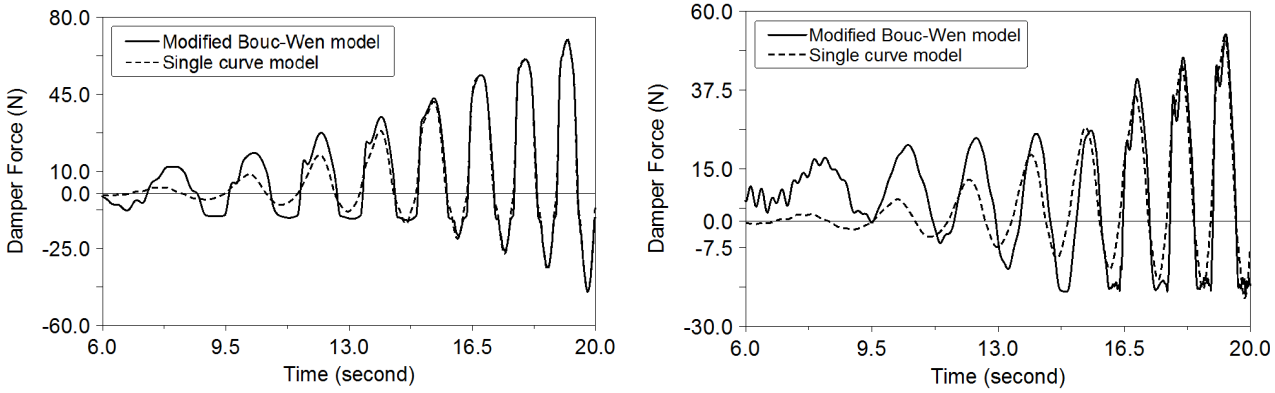


Figure 20 – Comparison of transfer function for front (left) and rear (right) suspensions

6 Conclusion and perceptive

In this article, a damper model based on Bouc-Wen model and its identification method are proposed. By a multi-objective optimization process, the model can reproduce the test curves with different velocities and frequencies. A comfort simulation in virtual ground orienting to low frequencies has been realized and it showed the impact of hysteresis model to the ride comfort prediction.

The next step will be the robust tests of parameters to the simulation results according to the analysis in Section 4.3 because measurements have shown that the curves may be slightly different for the two shock absorbers from the same technical definition.

References

References

[1] J. Arbiol, J. A. Muñoz, X. Armengo, and E. Aramburu, "Full vehicle durability analysis by means of the idiada virtual proving ground," in *Proceedings of the FISITA 2012 World Automotive Congress*. Springer, 2013, pp. 337–347.

[2] H. Wee, Y. Y. Kim, H. Jung, and G. N. Lee, "Nonlinear rate-dependent stick-slip phenomena: modeling and parameter estimation," *International Journal of Solids and Structures*, vol. 38, no. 8, pp. 1415–1431, 2001.

[3] G. Wang, G. Chen, and F. Bai, "Modeling and identification of asymmetric bouc–wen hysteresis for piezoelectric actuator via a novel differential evolution algorithm," *Sensors and Actuators A: Physical*, vol. 235, pp. 105–118, 2015.

[4] "A new method for damper characterization and real-time capable modeling for ride comfort."

[5] G. Yao, F. Yap, G. Chen, W. Li, and S. Yeo, "Mr damper and its application for semi-active control of vehicle suspension system," *Mechatronics*, vol. 12, no. 7, pp. 963–973, 2002.

[6] M. Sadeghi Reineh, "Physical modeling and simulation analysis of an advanced automotive racing shock absorber using the 1d simulation tool amesim," 2012.

[7] "Improved functional modelling in comfort analyses for hydraulic suspension testing."

[8] H.-B. Yun, S. F. Masri, R. W. Wolfe, and G. Benzoni, "Data-driven methodologies for change detection in large-scale nonlinear dampers with noisy measurements," *Journal of Sound and Vibration*, vol. 322, no. 1-2, pp. 336–357, 2009.

[9] B. Loyer, "Conception fonctionnelle robuste par optimisation multicritère de systèmes de suspension automobile passifs et semi-actifs," Ph.D. dissertation, Ecully, Ecole centrale de Lyon, 2009.

[10] M.-M. Chatillon, "Méthodologie de conception robuste appliquée aux trains de véhicules de tourisme," Ph.D. dissertation, Ecully, Ecole centrale de Lyon, 2005.

[11] P. R. Dahl, "A solid friction model," Aerospace Corp El Segundo Ca, Tech. Rep., 1968.

[12] A. Al Majid, "Dissipation de l'énergie en mécanique vibratoire: opérateur d'hystérésis, phénomène métrique," Ph.D. dissertation, Lyon, INSA, 2002.

[13] B. Spencer Jr, S. Dyke, M. Sain, and J. Carlson, "Phenomenological model for magnetorheological dampers," *Journal of engineering mechanics*, vol. 123, no. 3, pp. 230–238, 1997.

[14] N. Kwok, Q. Ha, M. Nguyen, J. Li, and B. Samali, "Bouc–wen model parameter identification for a mr fluid damper using computationally efficient ga," *ISA transactions*, vol. 46, no. 2, pp. 167–179, 2007.

[15] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: Nsga-ii," in *International conference on parallel problem solving from nature*. Springer, 2000, pp. 849–858.

[16] P. G, "Déploiement et validation du confort," Renault, Rapport dâExpertise, 2018.

Modal identification of machining robots in service

Asia MAAMAR[1], Vincent GAGNOL[1], Thien-Phu LE[2], Laurent SABOURIN[1]
[1] Université Clermont Auvergne, CNRS, SIGMA Clermont, Institut Pascal, 63170 Aubière cedex, France
asia.maamar@sigma-clermont.fr
vincent.gagnol@sigma-clermont.fr
laurent.sabourin@sigma-clermont.fr
[2] LMEE, Université d'Évry Val-d'Essonne, 40 rue du Pelvoux, Évry cedex, 91020, France
thienphu.le@univ-evry.fr

Abstract

The requirement of high performance in the industry has led to the introduction of industrial robots in the machining field. Machining robots have become a potential and promising alternative to standard machine tools because of their important workspace and their high flexibility in the machining of complex parts. However, their lack of precision and rigidity is still a limit for precision tasks.

Therefore, the modal identification of a machining robot is important for an accurate knowledge of its dynamic behavior. Usually, the characterization is carried out through an experimental modal analysis performed at rest. However, the excitation, artificially, created by a hammer or a shaker is not representative of the real cutting force applied in machining. Unfortunately, the dynamic behavior of a machining robot in rest differs significantly from that identified in service.

In this paper, an experimental modal analysis of an ABB IRB 6660 robot is firstly investigated. Then, modal parameters are identified during a machining operation through an operational modal identification. A significant variability of modal parameters identified at rest from those identified in service is observed, which highlights the need to identify robot modal parameters under operational machining conditions.

1. Introduction

Industrial robots have an important role in the industry. They are used in a large area of applications such as welding, assembling and painting. Due to their significant advantages of high flexibility, large workspace, more accessibility, high productivity, and relatively low cost compared to a CNC machine tool (MOCN), industrial robots have been introduced in the machining field. However, many factors are degrading the accuracy of the machining operation performed. One of the main problem is the low performance of the robot in terms of stiffness that strongly affects the machining stability and the quality of the workpiece during machining operations. Therefore, it is interesting to evaluate the robot structure modal parameters. As a part of the proposed study, we focus on the identification of modal parameters of a poly-articulated industrial robot ABB IRB 6660 (located at SIGMA Clermont) equipped with a HSM Spindle (36000 rpm, 15.2kW).

Despite the huge amount of work present in the modal identification domain, the modal identification of machining robots is still nowadays considered as an open issue [1-2]. Most of the scientific works presented in literature are related to the modal parameters identification through an Experimental Modal Analysis (EMA) under an artificial excitation using an impact hammer or shaker

tests, at rest [3]. However, the dynamic behavior of machining robots at rest is not the same as that observed in service, due to numerous differences such as the command influence and the machining interaction with the workpiece. Thereby giving rise to the need to identify this dynamic behavior in machining conditions, using an Operational Modal Analysis (OMA) approach.

The objective of this work is to identify modal parameters of a machining robot in order to point out the influence of the task position in the robot workspace concerning the modal behavior of the structure. The paper is organized as follows: in section 2, an experimental modal analysis of the ABB IRB 6660 robot is conducted in different configurations of the robots in its workspace. The identified modal parameters vary significantly from a configuration to another, at rest, which introduces the need to identify modal parameters continuously in operational conditions. For an accurate knowledge of the dynamic behavior of the considered robot, an operational modal analysis is carried out in section 3 during a milling test. The identification is investigated using the Transmissibility Function Based (TFB) method. Finally, section 4 concludes this paper.

2. Experimental modal analysis

2.1 Experimental setup and measurements

An experimental Modal Analysis (EMA) of the ABB IRB 6660 machining robot is conducted in different positions in its workspace, in order to observe the evolution of the modal behavior of the robot at rest. Four positions (P_1, P_2, P_3, P_4) along the *Y* direction, as shown in Figure 1, corresponding to four robot arm extension configurations are considered.

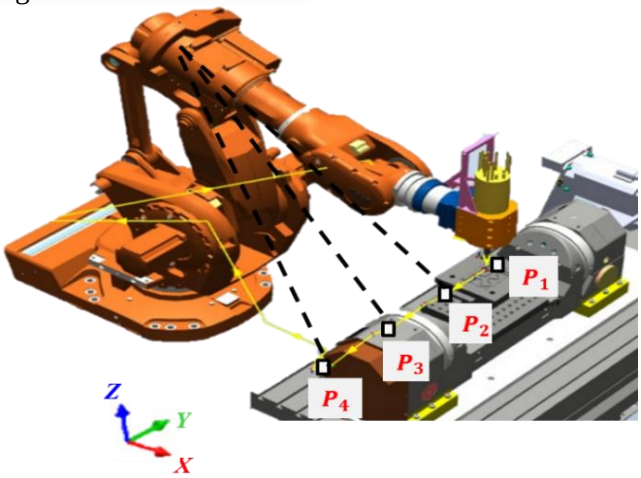


Figure 1: Robot Tool Center Point (TCP) positions along the *Y* direction in the robot's workspace

The EMA is based on an experimental identification of the Frequency Response Functions (FRFs), where the responses are measured with accelerometers and the excitation is performed using an impact hammer. Exciting forces are measured with force transducers. Two tri-axis PCB accelerometers, with a sensitivity of 99.9 mV/g and 101.2 mV/g respectively, were mounted on the spindle head in the two directions *X* and *Y*, as shown in Figure 3. Model identification was performed by applying impact hammer shocks in the two directions *X* and *Y*.

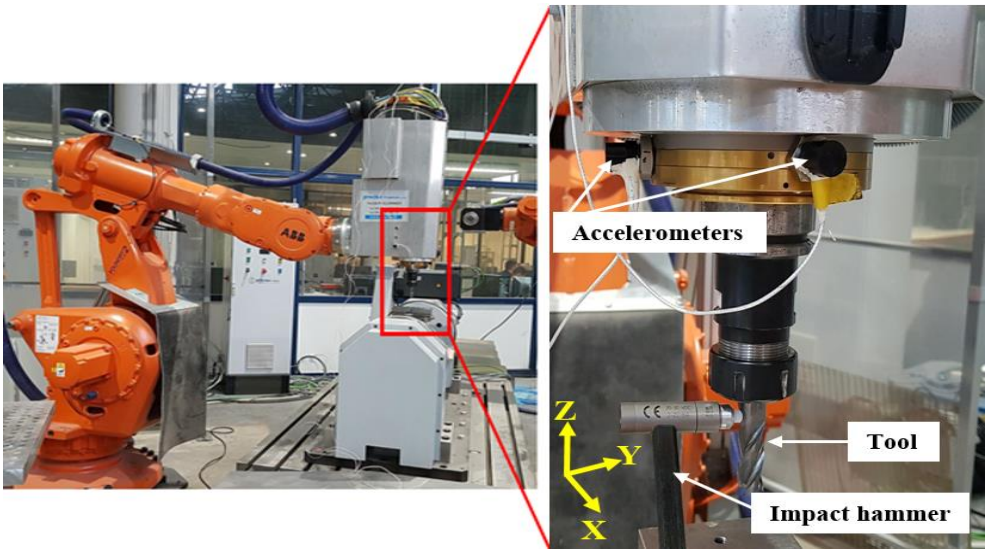


Figure 2: Experimental protocol for an impact testing on an ABB IRB 6660 machining robot

Two FRFs H_{xx} and H_{yy} were measured from the two experimental tests.

$$\begin{cases} H_{xx}(\omega) = \dfrac{X(\omega)}{F_x(\omega)} \\ H_{yy}(\omega) = \dfrac{Y(\omega)}{F_y(\omega)} \end{cases} \quad (1)$$

2.2 Results and discussion

The PolyMAX method is then investigated in order to estimate modal parameters from the two measured FRFs H_{xx} and H_{vv}. Natural frequencies f and damping ratios ξ identified from the measured FRF H_{xx} for each configuration of the robot in position 1 from are given in Table 1.

Mode	1st position (P_1)		2nd position (P_2)		3rd position (P_3)		4th position (P_4)	
	f (Hz)	ξ (%)	f (Hz)	ξ (%)	f (Hz)	ξ (%)	f (Hz)	ξ (%)
1	11.09	1.70	11.06	1.47	11.09	1.81	10.30	2.01
2	16.98	1.25	18.17	1.26	18.29	0.72	----	----
3	23.54	1.44	23.44	0.91	23.39	0.32	24.15	0.31
4	43.30	3.69	45.28	4.35	46.33	4.26	46.50	3.58
5	61.09	3.40	62.58	4.65	62.95	4.73	62.94	4.85
6	136.76	3.11	137.61	3.12	138.54	2.90	139.48	2.51
7	155.08	2.24	155.15	2.13	155.57	2.31	156.4	2.21
8	177.71	4.40	179.35	4.12	179.28	3.75	178.89	3.87
9	----	----	----	----	----	----	210.82	0.75
10	213.09	2.26	213.39	2.29	217.58	2.30	221.85	0.21
11	285.25	5.16	280.23	0.55	282.50	0.97	283.87	1.04
12	359.37	0.57	359.54	0.49	359.56	0.59	360.50	0.70
13	407.62	1.01	403.56	0.71	403.56	0.69	403.53	0.66
14	457.66	0.59	449.48	2.52	440.86	0.62	447.54	1.01
15	511.65	0.96	509.99	0.86	510.2	1.08	510.55	1.22
16	553.82	0.99	545.64	1.16	542.47	1.40	----	----

Table 1: Identified modal parameters of the robot ABB IRB 6660 in position 1 through an EMA

Results show that by changing the robot position, different structural modes are excited and identifiable. In order to quantify the variation of the identified natural frequency as a function of the

configuration and the position of the robot in its workspace, the relative standard deviation is calculated as follows:

$$\varsigma = \frac{\sigma * 100}{\mu} \tag{2}$$

Where σ is the standard deviation and μ is the mean value of the identified frequencies.

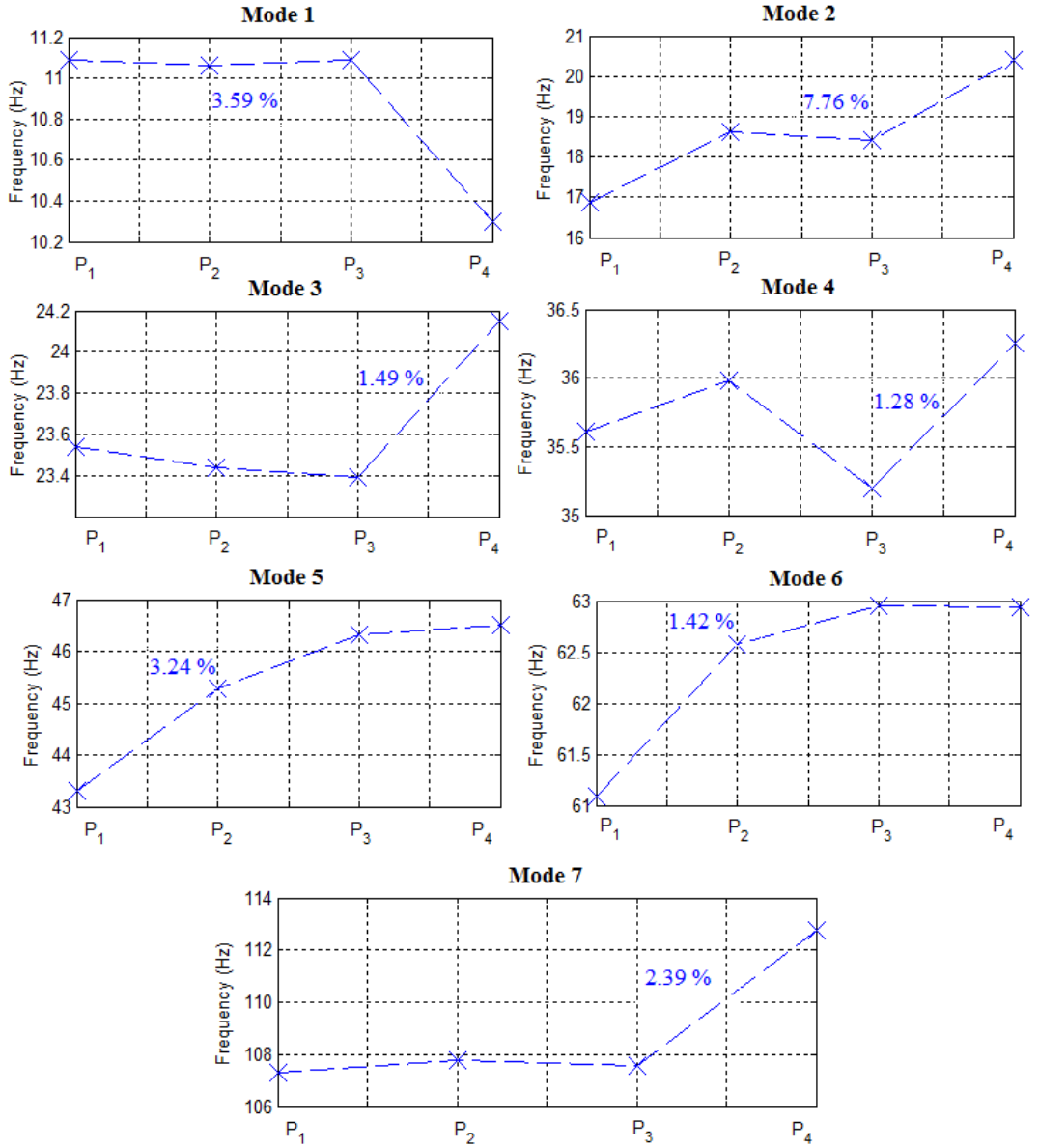


Figure 3: Evolution of the natural frequencies of the robot ABB IRB 6660 in function of its position

The figure 3 shows the evolution of the natural frequencies, corresponding to the first seven robot modes as a function of its position in its workspace. These modes are retained because they are systematically present in the four considered positions of the robot.

Results show that the identified natural frequencies of the considered modes vary between 1.28% and 7.76% in function of the robot position, at rest. This demonstrates that experimental modal analysis is not sufficient for an accurate knowledge of the dynamic behavior of the machining robot, and leads to the necessity of in-service modal parameter identification. Nevertheless, EMA provides an idea about the modal model of the robot structure. For this reason, EMA is carried out as a preparation step before performing an operational modal analysis of the machining robot ABB IRB 6660.

3. Operational modal analysis

3.1 Experimental setup

An operational modal analysis is carried out, when robot is in the first position (P_1), in order to identify modal parameters of the robot in service and compare results with those identifies through impact tests . Four different machining operations in the *X* and *Y* directions (pass 1, pass 2, pass 3 and pass 4), as illustrated in Figure 4, are performed with the ABB IRB 6660 robot, using a three-tooth, 16 mm end-mill cutting tool. Machining operations are carried out on a square aluminum (2017) piece of 100 mm per side, at a constant spindle speed of 10000 rpm. The axial and the radial depth of cut are equal to 3 and 12 mm, respectively. The feed rate is set to 0.3 mm/rev, resulting in chip loads of 0.1 mm/tooth.

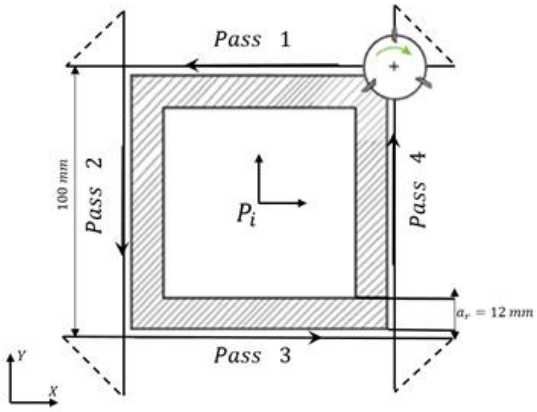


Figure 4: Machining operation when robot is in position 1 and configuration 1

Two PCB-type tri-axis accelerometers, with a sensitivity of 99.9 mV/g and 101.2 mV/g respectively, are mounted on the spindle head in the *X* and *Y* directions. Four each pass, two acceleration signals are recorded simultaneously using the LMS TEST.Lab acquisition system. The rotational spindle speed is equal to 10000 rpm, so harmonic components will be multiple of 166,66 Hz. It is clear in Figure 5, the first harmonic frequency is around 166 Hz.

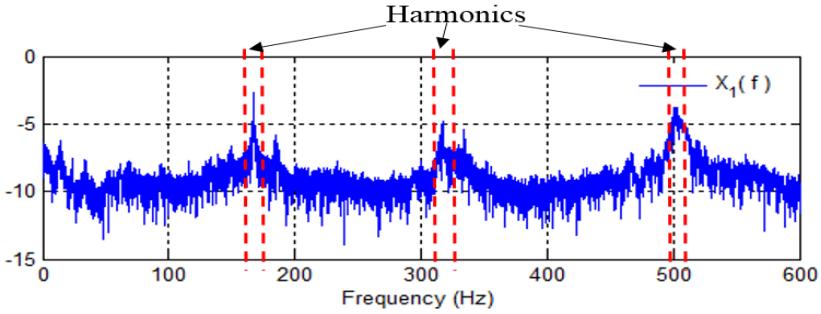


Figure 5: FFT of the measured acceleration in the x direction in pass 1 and presence of harmonics associated to the machining process

2.2 Operational modal identification by the Transmissibility Function Based (TFB) method

The operational modal analysis method based on transmissibility functions, proposed by Devriendt et al. [5], is the only OMA technique able to identify modal parameters in presence of preponderant harmonic components during machining process [6-7].

A transmissibility function is defined as the ratio between the motion response $X_i^k(\omega)$ and the reference motion response $X_j^k(\omega)$ under a single force located at *k*.

$$T_{ij}^{k}(\omega) = \frac{X_i^k(\omega)}{X_j^k(\omega)} \tag{3}$$

From the measured acceleration signals during the cutting path, four transmissibility functions are calculated:

$$\begin{cases} T_{xy}^{1}(\omega) = \dfrac{X_1(\omega)}{Y_1(\omega)} \\ T_{xy}^{2}(\omega) = \dfrac{X_2(\omega)}{Y_2(\omega)} \\ T_{xy}^{3}(\omega) = \dfrac{X_3(\omega)}{Y_3(\omega)} \\ T_{xy}^{4}(\omega) = \dfrac{X_4(\omega)}{Y_4(\omega)} \end{cases} \tag{4}$$

Transmissibility functions cross each other at the resonant frequencies, corresponding to the poles of the rational transmissibility functions ΔT_{ij}^{-1}.

$$\Delta^{-1} T_{ij}^{kl}(\omega) = \frac{1}{T_{ij}^{k}(\omega) - T_{ij}^{l}(\omega)} \tag{5}$$

When considering the TFB method, modal parameters are obtained from the transmissibility functions. In this case, six rational transmissibility functions ΔT_{ij}^{-1} are calculated:

$$\begin{cases} \Delta T_{12}^{-1}(\omega) = \dfrac{1}{T_{xy}^{1}(\omega) - T_{xy}^{2}(\omega)} \\ \Delta T_{23}^{-1}(\omega) = \dfrac{1}{T_{xy}^{2}(\omega) - T_{xy}^{3}(\omega)} \\ \Delta T_{34}^{-1}(\omega) = \dfrac{1}{T_{xy}^{3}(\omega) - T_{xy}^{4}(\omega)} \\ \Delta T_{14}^{-1}(\omega) = \dfrac{1}{T_{xy}^{1}(\omega) - T_{xy}^{4}(\omega)} \\ \Delta T_{13}^{-1}(\omega) = \dfrac{1}{T_{xy}^{1}(\omega) - T_{xy}^{3}(\omega)} \\ \Delta T_{24}^{-1}(\omega) = \dfrac{1}{T_{xy}^{2}(\omega) - T_{xy}^{4}(\omega)} \end{cases} \tag{6}$$

Theoretically, not all of the zeros of $\Delta^{-1} T_{ij}^{kl}$ are the system's poles. In Figure 6, the six rational transmissibility function are illustrated. The red dashed lines indicate the location of the harmonic frequencies. It's clear that in ΔT_{ij}^{-1} most of the harmonics are reduced or even eliminated. Hence, the rational transmissibility function contains harmonic components in addition to the system poles.

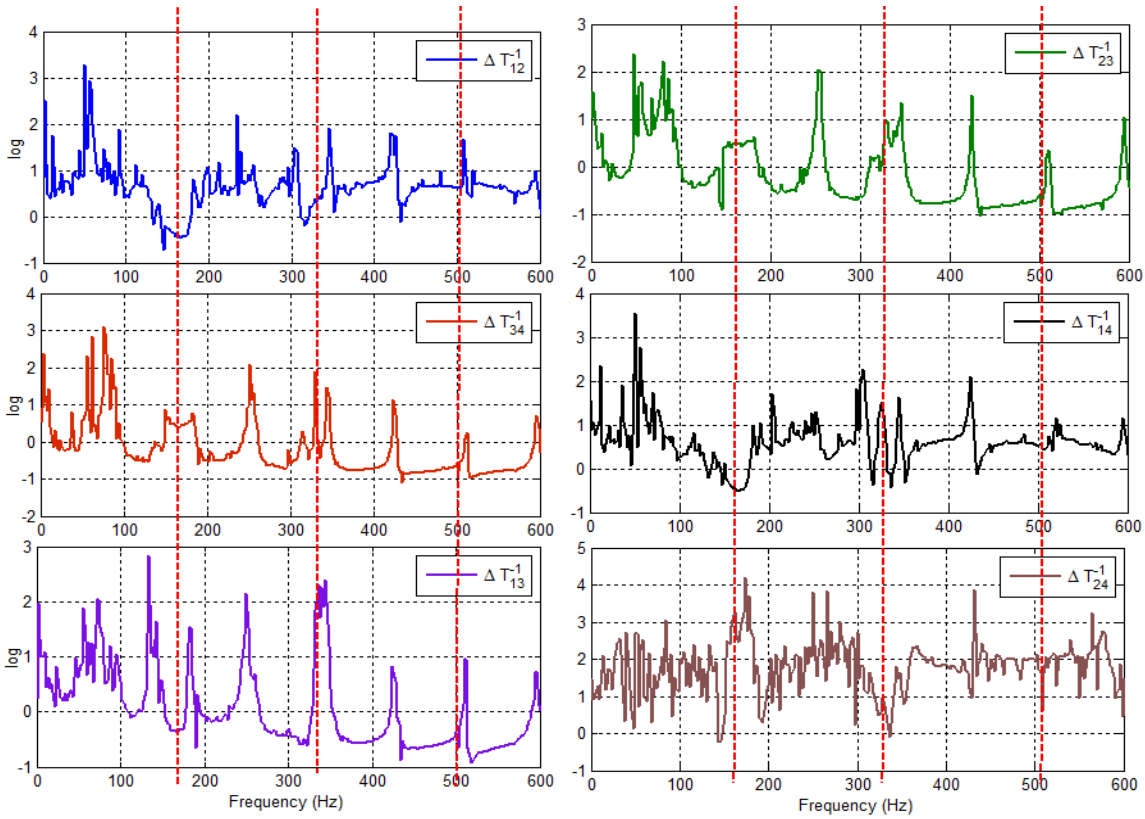


Figure 6: Rational transmissibility function ΔT_{ij}^{-1} and presence of harmonic components

It is therefore necessary to distinguish the physical poles of the system from spurious ones. For this reason, this two steps identification procedure is proposed:

a. Step 1: Stabilization diagram

Stabilization diagram shows the stability of the poles as a function of ΔT_{ij}^{-1} and as a function of increasing model order were used to distinguish the spurious modes from the physical poles. The identification of the modal parameters is performed by applying the frequency domain estimator "PolyMAX" to ΔT_{ij}^{-1}.Then, the modes corresponding to the order "*n*" are compared to the lower order modes "*n-1*". If the modal parameters variation doesn't exceed the defined tolerances (variation of 5% in frequency and 5% in damping ratio corresponding to the identified pole), the mode is considered stable and indicated by the letter (s), otherwise it is unstable and indicated by the letter (o).
The sum $\Delta^{-1}T$ of the six measured functions ΔT_{ij}^{-1} is plotted when the stability diagram is generated, in order to take the robot behavior in its entirety.

$$\Delta^{-1}T = \Delta^{-1}T_{12} + \Delta^{-1}T_{23} + \Delta^{-1}T_{34} + \Delta^{-1}T_{14}\Delta^{-1}T_{13} + \Delta^{-1}T_{24} \quad (7)$$

The figure 7 shows the stabilization diagram, established for an order varying from 33 to 72.

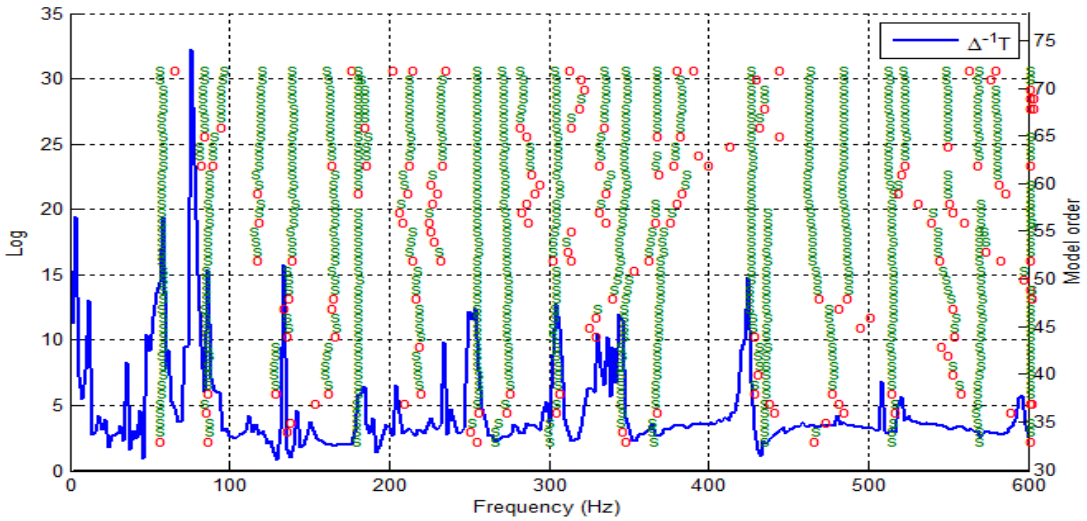


Figure 7: Stabilization diagram (s: stable pole, o: unstable pole)

The stabilization diagram is considered in order to identify the physical poles and to separate the digital poles. However, some modes are identified and verify the frequency and damping stability criteria, although they are not structural modes. This is largely due to the fact that the ΔT_{ij}^{-1} functions contain modes that are not related to the robot's modal behavior.

b. Step 2: Selection of the system's poles by means of singular value decomposition of the transmissibility matrix

To select only the correct system poles, the following transmissibility matrix **T** is considered, and a singular value decomposition is performed.

$$\boldsymbol{T} = \begin{bmatrix} T_{xy}^{1} & T_{xy}^{2} \\ T_{xy}^{3} & T_{xy}^{4} \end{bmatrix} \tag{8}$$

At the system poles λ_r, the rank of the transmissibility matrix is equal to one. Consequently, $\sigma_1(s) > \sigma_2(s) \geq 0$. The peaks of $\frac{1}{\sigma_2}$ as a function of frequency indicate the system's poles. Thus, from this curve, we obtain the information on the location of the robot's natural frequencies.

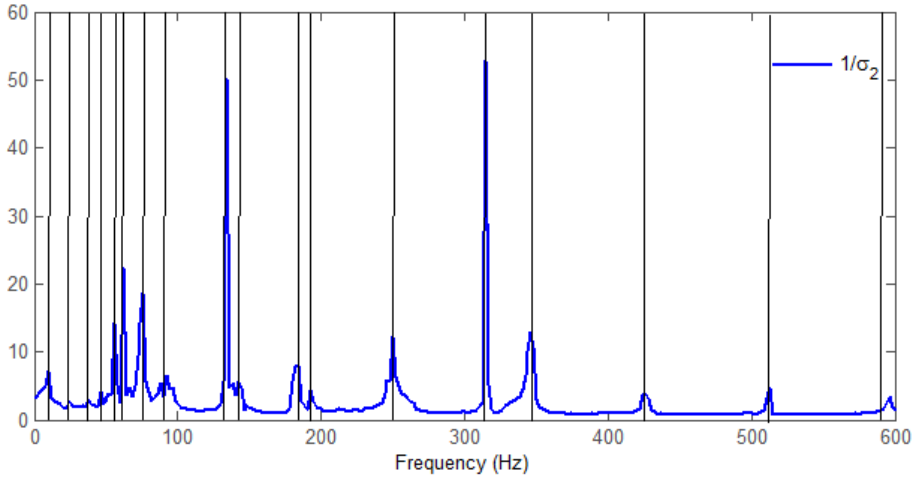


Figure 8: Selection of the system's poles by means of singular value decomposition

When applying the singular value decomposition of the proposed transmissibility matrix, all of the harmonics are eliminated and only the correct poles are selected, as shown in Figure 8. The modal parameters identified through an EMA and an OMA approach (TFB), when robot is in position 1, are illustrated in Table 2.

	AME		AMO	
Mode	Frequency (Hz)	Damping ratio (%)	Frequency (Hz)	Damping ratio (%)
1	11.09	1.70	10.29	2.54
2	16.98	1.25	----	----
3	23.54	1.44	22.69	0.29
4	----	----	38.04	0.24
5	43.30	3.69	49.96	0.45
6	----	----	56.98	1.24
7	61.09	3.40	63.67	3.58
8	----	----	75.50	0.48
9	----	----	93.67	0.91
10	136.76	3.11	134.04	0.25
11	155.08	2.24	141.24	2.55
12	177.71	4.40	182.37	0.98
13	213.09	2.26	190.27	0.45
14	----	----	252.05	0.31
15	285.25	5.16	----	----
16	----	----	316.48	0.53
17	359.37	0.57	346.35	0.43
18	407.62	1.01	----	----
19	457.66	0.59	424.82	0.31
20	511.65	0.96	512.25	0.19
21	553.82	0.99	595.55	0.18

Table 2: Identified modal parameters of the robot ABB IRB 6660 in position 1 and configuration 1 through an EMA/OMA

Firstly, an evolution of the identified modal parameters from the rest state to the machining state, can be noticed, due to the spindle rotation and the significant changes in the robot dynamic behavior. Also, structural modes which are not identified through experimental modal analysis, appear through the operational modal identification analysis. This proves that the energy delivered by the impact hammer is not sufficient to excite all of the structural poles of the machining robot ABB IRB 6660. Generally, the damping ratios, in service, are strongly reduced compared to those identified at rest. This is due to the tool/piece interaction that makes the robot structure more rigid. These results illustrate the importance of the modal parameters identification, in machining conditions. Although, the EMA is of great value in order to obtain a modal model as a reference for the validation of the modal parameters obtained, in service.

Conclusions

The evolution of the machining robot performance as a function of its position and the orientation of the tool center point is important in order to ensure stability during machining operations. This paper aims at characterizing the dynamics change of the robot dynamical behavior through several points of the workspace. Especially the evolution of its modal parameters for different configurations of the workspace is analyzed.

In this paper, the modal identification of a machining robot is proposed. This analysis is done, principally, in two stages. Firstly, the identification is conducted when robot is at rest, in different configuration in its workspace. Frequency analysis showed a small change in owns way their four configurations depending on the position. Results make it possible to evaluate the evolution of its modal behavior at rest and introduce the need to identify modal parameters of the robots in service. Thereafter, the operational modal analysis of the machining robots with the transmissibility function based method (TFB) was studied. The TFB method is adequate for machining conditions because of its ability to distinguish structural poles from spurious ones [7]. Modal parameters identified from an OMA are

different from those identified through an EMA. This is because the conditions of the machining robot at rest are not the same from its real conditions during machining, due to numerous differences such as the command influence and the machining interaction with the workpiece.
In perspective, an analysis of the measured vibration responses will be made in order to highlight the dynamic behavior in different work configurations of the robot, under machining process. Monitoring the evolution of the robot modal parameters in its workspace, in service, is one of the main insights of this future work.

Acknowledgments

This work has been sponsored by the French government research programme Investissements d'avenir through the RobotEx Equipment of Excellence (ANR-10-EQPX-44) and the LabEx IMobS3 (ANR-10-LABX-16-01), by the European Union through the Programme Regional Competitiveness and Employment 2007-2013, European Regional Development Fund (ERDF – Auvergne region) and by « Région Auvergne ».

References

[1] V.Gagnol, T-P.Le and P.Ray, *Modal identification of spindle-tool unit in high-speed machining*, Mechanical Systems and Signal Processing, 2011.

[2] M.Leonesio, E.Villagrossi and all, *Vibration analysis of robotic milling tasks*, 11th CIRP Conference on Intelligent Computation in Manufacturing Engineering, 2018.

[3] S.Mejri , V. Gagnol , T-P.Le, L.Sabourin , P.Ray, P.Paultre, *Experimental protocol for the dynamic modeling of machining robots*, 21ème Congrès Français de Mécanique, Août 2013.

[4] R.Cousturier, L.Sabourin, G.Gogu and Y.Mezouar, Identification et modélisation du comportement en rigidité d'un robot à structure à parallélogrammes: application aux opérations d'usinage, MUGV 2014.

[5] C.Devriendt, P.Guillaume, *Identification of modal parameters from transmissibility measurements*, Journal of sound and vibration, 314(1), 343-356.

[6] A. Maamar, M. Abdelghani, V. Gagnol , T-P.Le, L.Sabourin, *Operational Modal Identifcation methods in presence of harmonic excitations*. Applied Acoustics, 2019.

[7] A. Maamar, V. Gagnol , T-P.Le, L.Sabourin, *Modal identification of a machine tool structure during machining operations*, The International Journal of Advanced Manufacturing Technology, 2019.

Demodulating of the 3-D tip clearance of turbine blades using BP neural network optimized by genetic algorithm

Hongcheng Liu[1], Xiaodong Zhang[1,2],Yiwei Xiong[1], Siying Xie[1],

[1]School of Mechanical Engineering - Xi'an Jiaotong University
[2]Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System – Xi'an Jiaotong University
No.28 Xianning West Road, 710049, Xi'an, Shaanxi, P.R. China
{liuhongcheng}@stu.xjtu.edu.cn
{xdzhang}@mail.xjtu.edu.cn

Abstract

The 3-D tip clearance refers to the three-dimensional space between the probe end face and turbine blade tip surface, which contains more abundant fault information than the traditional one-dimensional tip clearance. In this paper, an approach for demodulating of the three-dimensional (3-D) tip clearance of turbine blades is presented using BP neural network optimized by genetic algorithm (GA). Through the static calibration experiments, the ratios of outer circle voltage to inner circle voltage of three units from the optical fiber probe are used as the training and test data for demodulation. The maximum errors of the radial displacement z_0, the axial angle α and the circumferential angle β demodulated by the demodulation model based on BP neural network optimized by GA are 0.1321mm, 0.9875° and 0.6456°, respectively. The mean square errors of the radial displacement z_0, the axial angle α and the circumferential angle β are 0.0006mm, 0.0528° and 0.0266°, respectively. The experiment results show that this demodulation method have an excellent precision, which can fulfil the requirements of the 3-D tip clearance detection and provide a basic guarantee for the fault diagnosis of turbine blades.

1 Introduction

As we all know, the health monitoring and fault diagnosis of the aero-engine have always been research hotspots and that of the aero-engine turbine blades are very important, which have been studied by many researchers. The monitoring of tip clearance of turbine blades is often used for both the active control and fault diagnosis, and it's an effective and significant monitoring method. Nevertheless, the traditional tip clearance of turbine blades refers to the radial displacement, which contains limited fault information of turbine blades. Xie et al. have presented the 3-D tip clearance of turbine blades which refers to the three-dimensional space between the probe end face and the surface of the turbine blade tip [1]. Teng et al. have utilized the 3-D tip clearance as the fault information carrier of turbine blades and have analysed the response characteristics of its characteristic parameters to a typical crack fault of the high pressure turbine blade by a finite element method [2]. Xiong et al. have researched on the response characteristics of the 3-D tip clearance of turbine blades to axial displacement of turbine rotor under the crack failure [3]. As a new fault feature carrier, the 3-D tip clearance contains more fault information than the traditional tip clearance and can reflect the health status of aero-engine turbine blades more comprehensively.

In order to monitor the 3-D tip clearance of aero-engine turbine blades, Xie et al. [4] have designed an optical fiber probe with three two–circle coaxial bundles based on intensity modulation and have researched its output characters. Based on this optical fiber probe, Zhang et al. [5] have built an optical fiber measurement system for the 3-D tip clearance of the aero-engine turbine blades and successfully demodulated the 3-D tip clearance from the output signals from the optical fiber probe using BP neural network. Xie et al. [6] have proposed a new demodulation technique for the 3-D tip clearance based on the ratio of the difference in the signal intensities between any two sensing units of the optical fiber probe. This demodulation method requires high consistency for the three sensing units of the optical fiber probe and an additional fast search algorithms must be needed.

The signals acquired from the optical fiber probe with three two-circle coaxial optical fiber bundles are modulated simultaneously by the distance and inclination angles between the optical fiber probe end face and

the reflective surface. Therefore, it is difficult to demodulate the 3-D tip clearance from the optical fiber probe output signals. In this study, an approach for demodulating of the 3-D tip clearance of turbine blades is proposed based on BP neural network optimized by GA. This method combines the fitting ability of BP neural network and the global search ability of GA, and can improve the demodulation accuracy of the 3-D tip clearance, which provides a possibility for the subsequent fault diagnosis of turbine blades using the 3-D tip clearance.

2 Demodulation method

2.1 Demodulation method based on BP neural network

The back propagation (BP) neural network algorithm is a multi-layer feedforward network trained according to error back propagation algorithm and is one of the most widely applied neural network models [7]. In theory, any input and output function can be fitted when the appropriate structure and parameters of the neural network are chosen. Similarly, demodulation of the 3-D tip clearance is equivalent to fitting the input and output function of the optical fiber probe, which can be achieved through BP neural network. This study utilized an optical fiber probe with three two-circle coaxial bundles, of which the output signals are the ratios of the outer circle voltage to inner circle voltage corresponding to the three units respectively [8]. In order to demodulate the 3-D tip clearance from the output signals of the optical fiber probe, a three-layer BP neural network was used in this study. The input of the demodulation model based on BP neural network is the three ratios from the optical fiber probe and the output is the value of the 3-D tip clearance. The number of BP neural network input units is $n=3$ and the number of output units is $m=3$. The number of hidden units s_1 is estimated according to the formula [9]: $s_1=\sqrt{n+m}+a$, where a is an integer between 1 and 10. The value of s_1 is between 4 and 13, which is taken as 10 in this model according to the demodulation effect. The activation function between input layer and hidden layer is chosen as 'tansig' (Hyperbolic tangent sigmoid transfer function). The activation function between hidden layer and output layer is chosen as 'purelin' (Linear transfer function). Figure 1 shows the structure of the three-layer BP neural network used in this study. r_1, r_2, r_3 refer to the three ratios corresponding to the three units of the optical fiber probe, respectively. z_0, α, β refers to the radial displacement, the axial angle and the circumferential angle, respectively.

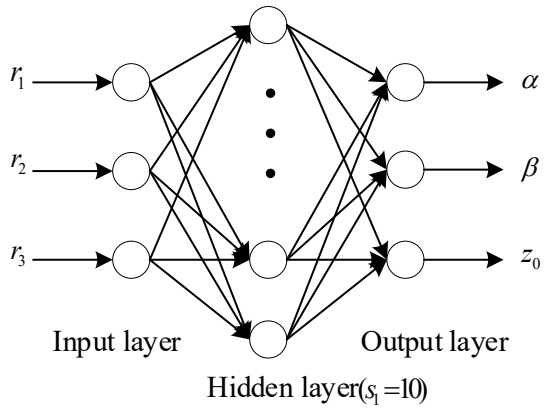


Figure 1: The BP neural network structure

The initial weights and thresholds are selected randomly in the traditional BP neural network, which seriously limits the precision and convergence speed of the network training and makes the network easily fall into local extremum. In order to improve the convergence speed of traditional BP neural network, GA was used to optimize the weights and thresholds of BP neural network in this study.

2.2 Genetic algorithm

Genetic algorithm is a computational model that simulates the natural evolution of Darwin's biological evolution theory and the biological evolution process of genetics. It is a method to search for optimal solutions by simulating natural evolutionary processes. The main steps of GA are as follows:

(1) Encode the parameters that need to be optimized to form a chromosome.

(2) Initialize population and related parameters such as crossover probability, mutation probability and population size.

(3) Calculate the fitness of each individual in the existing population.

(4) Select the candidate individuals based on the fitness using roulette wheel selection. As usual, the individual with high fitness is more likely to be selected than the individual with low fitness.

(5) Generate new individuals according to pre-set crossover probability.

(6) Generate new individuals according to pre-set mutation probability.

(7) Repeat generational process until a termination condition has been reached.

Because the search of GA only relays on the fitness not the gradient information of objective function, GA is fit for the problems that are complex and nonlinear. Therefore, GA is used to optimize the weights and thresholds of BP neural network in this study and it can improve the training accuracy and convergence speed of BP neural network.

2.3 Demodulation method based on BP neural network optimized by GA

There are some typical limitations of the BP neural network: The error is not sensitive to the change of the weights. The gradient descent algorithm is generally very slow. The number of iterations is large. The convergence speed is slow and the output of the neural network is easily trapped in the local minimum. As a global optimization algorithm, GA can compensate the deficiency of BP neural network to improve the accuracy of demodulation model. After determining the structure of BP neural network by the input and output, the initial weights and thresholds of BP neural network are encoded to generate the individuals and the length of individuals of the GA can be determined. Subsequently, by means of the selection, crossover and mutation, the best fitness value corresponding to the individual is obtained. Next, the BP neural network obtains the optimal initial weights and thresholds provided by the GA and predicts the demodulation output after the network has been trained [10-12].

As shown in Figure 2, the process of optimizing BP neural network by GA mainly includes [13]:

(1) Determine the structure and parameters of BP neural network.

(2) Optimize the initial weights and thresholds of BP neural network by GA.

(3) Continue the training of BP neural network after obtaining the optimal initial weights and thresholds.

(4) Predict the demodulation output using BP neural network.

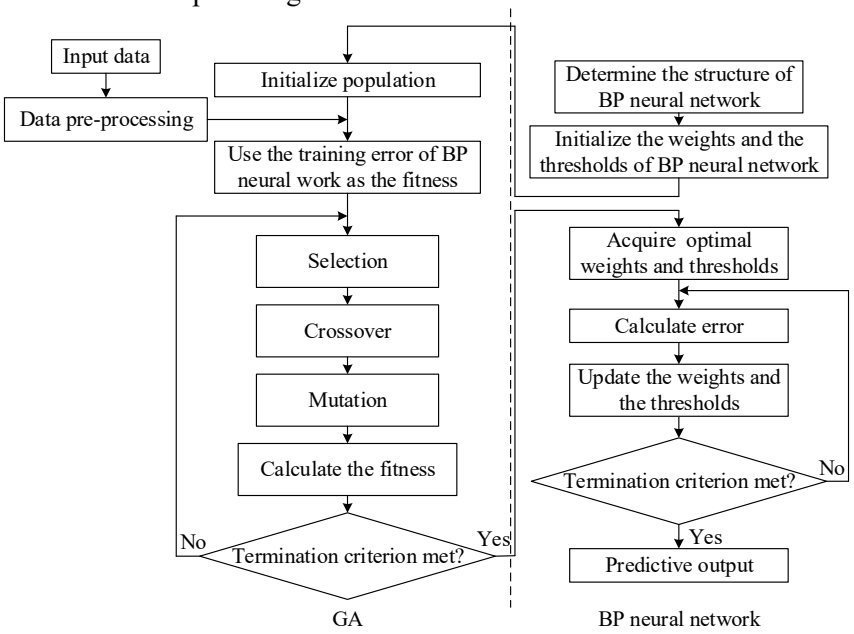


Figure 2 Flow chart of BP neural network optimized by GA

The data obtained through the calibration experiments, which contain the ratios from the output of the optical fiber probe and the 3-D tip clearance corresponding to the ratios, are used to build the demodulation model based on BP neural network optimized by GA. In other word, the three ratios (r_1, r_2, r_3) obtained by

the optical fiber probe and the calibrated 3-D tip clearance (z_0, α, β) set by adjusting the calibration table are used as the input and output of the demodulation model. The BP neural network is composed of three input units, three output units and one hidden layer with 10 units. The population size, crossover probability and mutation probability of GA are 20, 0.8, and 0.1, respectively.

3 Experiments and analysis

3.1 Calibration experiments

As shown in Figure 3, the 3-D tip clearance consists of the radial displacement z_0, the axial angle α and the circumferential angle β, which are the vertical distance between illuminating fiber of Unit0 and the reflective surface, the intersection angle at x-axis direction between Unit0 and the reflective surface, and the intersection angle at the y-axis direction between Unit0 and the reflective surface [6]. The calibration table consists of a stage for adjusting radial displacement and two stages for adjusting angle. Different 3-D tip clearance between the probe end face and the reflective surface can be simulated through adjusting the radial displacement and two inclination angles of the calibration table. The range of the radial displacement of the calibration table is from 0mm to 10mm with the accuracy of 0.01mm. The range of the inclination angle of the calibration table in both direction is from -15° to +15° with the accuracy of 0.1°.

The steps of calibration experiments are as follows:

(1) Determining calibration points

According to the variation range of the 3-D tip clearance of the turbine blade under typical failure [14], the calibration range of the radial displacement is from 1.4mm to 2.5mm at interval of 0.02mm and the calibration range of the inclination angle in both direction is from -0.4° to 2.6° at interval of 0.02°. The total number of the calibration points are 14336.

(2) Calibrating the required points

According to the 3-D tip clearance need to be calibrated, the radial displacement stage and both the two inclination angle stages of the calibration table are adjusted manually. The output signals, three ratios of outer circle voltage to inner circle voltage with respect to the 3-D tip clearance is recorded.

(3) Pre-processing the calibration data

In order to prepare for training BP neural network and building the demodulation model of the 3-D tip clearance, the acquired calibration data are pre-processed based on the basic rules of the output from the optical fiber probe.

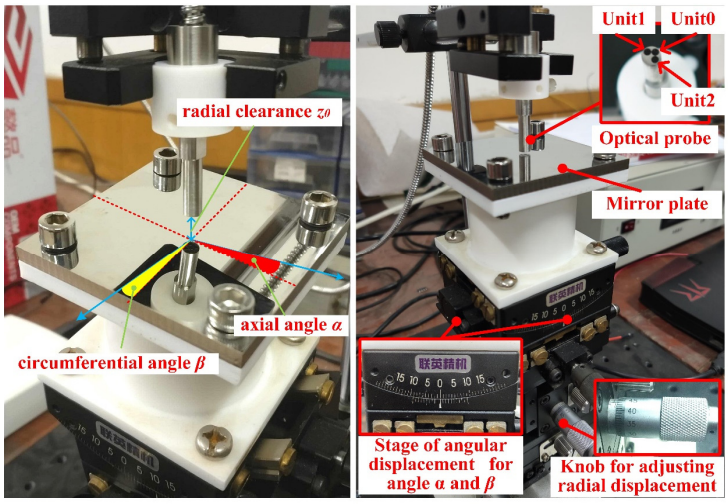
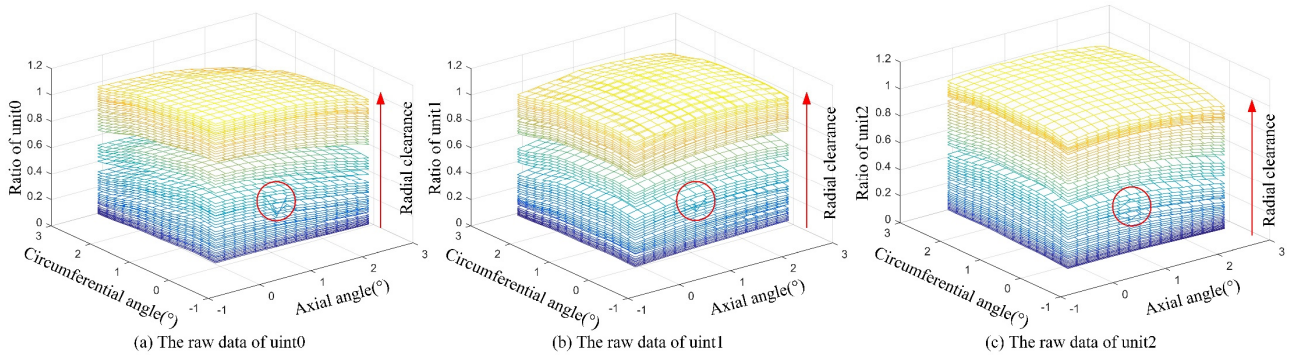


Figure 3 Calibration table for the 3-D tip clearance

3.2 Data pre-processing

The raw data of this study obtained according to the above calibration experiments have 14336 sets. Figure 4 (a), (b) and (c) show the raw data in the three-dimensional coordinate system of unit0, uint1 and unit2, respectively. The x-axis, y-axis and z-axis represent the axial angle, circumferential angle and the ratio

of the outer circle voltage to inner circle voltage, respectively. In the picture, each layer represents the data measured at the same radial displacement and the radial displacement of the upper layer is larger than that of the lower layer. As indicated by the red circle marks, the calibration data are erroneously fluctuant due to the instability of the hardware circuitry of the detection system. These calibration data, which can't correctly reflect the input and output relationship of the optical fiber probe, will affect the demodulation accuracy of the neural network. Therefore, a simple processing method was utilized to deal with it. According to the experiment data, the ratios of the outer circle voltage to inner circle voltage of each unit increase with the increase of the radial displacement under the same axial angle and circumferential angle. Therefore, for each unit, the ratios of the outer circle voltage to the inner circle voltage and that of the adjacent radial displacement are compared in turn and the data that doesn't satisfy the aforementioned rules are eliminated. After the pre-process, 13157 sets of data are left, and 11000 sets of data are selected as the training data and the left are used as the test data.

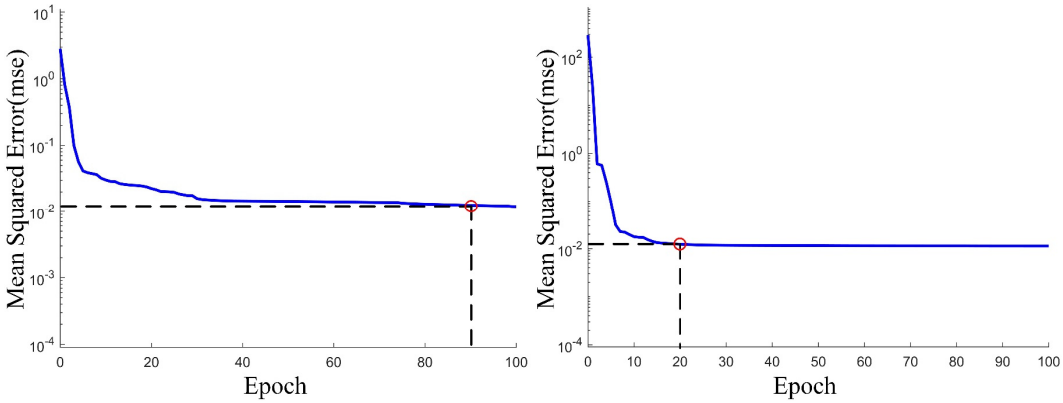


(a) The raw data of uint0 (b) The raw data of uint1 (c) The raw data of unit2

Figure 4 The raw data collected by the calibration experiments

3.3 Analysis of demodulation results

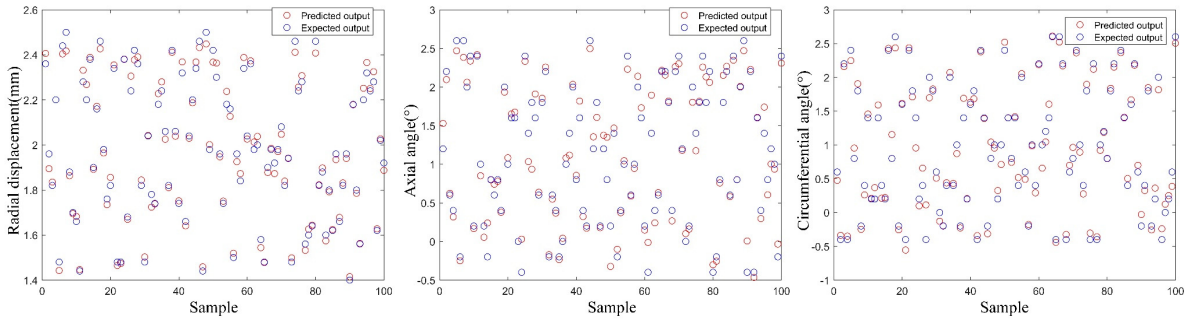
It can be seen from Figure 5 that when the neural network training target error is the same, the epochs of the traditional BP neural network are 90, and the epochs of the BP neural network optimized by GA are 20. Obviously, the epochs of BP neural network optimized by GA are significantly less than that of traditional BP neural network, and the convergence speed of BP neural network optimized by GA is fast.



(a) BP neural network (b) BP neural network optimized by GA

Figure 5 Neural network training error curve

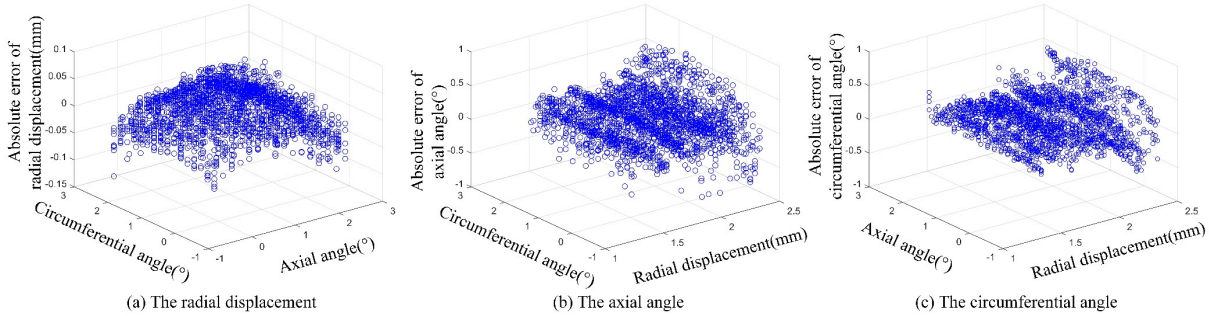
Figure 6 shows a comparison between the predicted output of the 3-D tip clearance demodulated based on BP neural network optimized by GA and the expected output corresponding to that. According to the comparison results, it's obvious that the demodulation model based on BP neural network optimized by GA can demodulate the 3-D tip clearance from the output signals of the optical fiber probe and fulfil the requirements of the 3-D tip clearance detection system. The detailed analysis is presented as follows combined with the maximum error and the mean square error.



(a) The radial displacement (b) The axial angle (c) The circumferential angle

Figure 6 Comparison of the predicted output and the expected output of the 3-D tip clearance

Demodulated based on the traditional BP neural network, the maximum error of the radial displacement, axial angle and circumferential angle are 0.047mm, 0.49° and 2.32°, respectively. The mean square error of the radial displacement, the axial angle and circumferential angle are 0.010mm, 0.13° and 0.36°, respectively [5]. As shown in Figure 7, the maximum error of the radial displacement, axial angle and circumferential angle using the demodulation model based on BP neural network optimized by GA are 0.1321mm, 0.9875° and 0.6456°, respectively. The mean square error of the radial displacement, axial angle and circumferential angle are 0.0006mm, 0.0528° and 0.0266°, respectively. The comparison of the demodulation accuracy between BP neural network and BP neural network optimized by GA (GA-BP) is shown in Table 1.



(a) The radial displacement (b) The axial angle (c) The circumferential angle

Figure 7 Demodulation error distribution of the 3-D tip clearance

Algorithm	The maximum error			The mean square error		
	z_0 (mm)	α (°)	β (°)	z_0 (mm)	α (°)	β (°)
BP	0.047	0.49	2.32	0.010	0.13	0.36
GA-BP	0.1321	0.9875	0.6456	0.0006	0.0528	0.0266

Table 1 The accuracy of BP and GA-BP

From the error analysis above, the demodulation model based on BP neural network optimized by GA has greatly improved the demodulation accuracy of circumferential angle and the maximum demodulation error is reduced from 2.32° to 0.6456°. In terms of the mean square error, the 3-D tip clearance demodulation accuracy based on BP neural network optimized by GA is an order of magnitude higher than the traditional BP neural network. The reasons why the maximum error of demodulating the 3-D tip clearance is relatively large contain the presence of dark current in the photoelectric conversion circuit, human error caused by the manual adjustment of the calibration table, and fluctuations of the light source caused by power fluctuations. In order to reduce the maximum demodulation error and make the training data of neural network more reliable, the static calibration system needs to be improved from the above aspects in the future.

In summary, the demodulation model based on BP neural network optimized by GA has a better performance and a higher accuracy than the demodulation model based on traditional BP neural network.

4 Conclusion

In this paper, an approach for demodulation of the 3-D tip clearance of turbine blades has been presented based on BP neural network optimized by GA. The large amount of data collected by the calibration

experiments are used as the training data and test data of the neural network after data pre-processing. The mean square error of the 3-D tip clearance demodulated based on BP neural network optimized by GA are (0.0006mm, 0.0528°, 0.0266°), which is an order of magnitude lower than that based on traditional BP neural network. The demodulation results indicate that the demodulation method can fulfil the basic requirements of the optical fiber detection system and provide the basis for the subsequent real-time detection of the 3-D tip clearance on the rotor laboratory bench. The maximum error of the 3-D tip clearance demodulated based on BP neural network optimized by GA are (0.1321mm, 0.9875°, 0.6456°).The main reason of the relatively large errors is that the hardware circuit of the static calibration system is not stable enough. The subsequent improvement of that can greatly improve the stability and accuracy of the calibration data, which can further reduce the maximum error of the 3-D tip clearance demodulation.

Acknowledgments

The project is supported by National Natural Science Foundation of China (NSFC) (51575436).

References

[1]S.Y.Xie, X.D.Zhang, *Design and modeling of three-dimensional tip-clearance optical probe based on two-circle reflective coaxial fiber bundle*[C]// Sensors. IEEE, 2017.

[2]F.Teng, X.D.Zhang, S.Y.Xie, *Response Characteristics Analysis of 3-D Blade Tip Clearance to Typical Fault Information*[J]. Aeroengine, 2017.

[3]Y.W.Xiong, X.D.Zhang, and S.Y.Xie, *Response characteristics of three-dimensional blade tip clearance of turbine under crack fault to rotor axial displacement*. Proceedings of the 12th National Conference on Vibration Theory and Application, Nanning, China, 2017.

[4]S.Y.Xie, X.D.Zhang, *Design and modeling of three-dimensional tip-clearance optical probe based on two-circle reflective coaxial fiber bundle*[J]. Proceedings of IEEE Sensors.

[5]X.D.Zhang, B.Wu, S.Y.Xie, *Optical fiber measurement system for 3-D variation of turbine blade tip clearance* [J]. Optics and Precision Engineering, 2018, 26(7): 1578-1587.

[6]S.Y.Xie, X.D.Zhang, Y.W.Xiong, and H.C.Liu, *Demodulation technique for 3-D tip clearance measurements based on output signals from optical fiber probe with three two-circle coaxial optical fiber bundles*, Opt. Express 27, 12600-12615 (2019)

[7]J.Li, J.H.Cheng, J.Y.Shi, et al. *Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement*[M]// Advances in Computer Science and Information Engineering. Springer Berlin Heidelberg, 2012.

[8]S.Y.Xie, X.D.Zhang, *Three-Dimensional Behavior of Displacement Sensor of Two-Circle Reflective Coaxial Fiber Bundle*[J]. Journal of Vibration, Measurement & Diagnosis, 2017(1).

[9]Z.X.Ge, Z.Q.Sun, *The Theory of Neural Network and The Implementation of MATLAB R2007* [M]. Electronic Industry Press, 2007.

[10]C.Li, L.Li, *Application of BP Neural Network Based on Genetic Algorithms Optimization in Prediction of Postgraduate Entrance Examination*[C]// 2016 3rd International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2016.

[11]C.J.Tang, Y.G.He, L.F.Yuan, *A Fault Diagnosis Method of Switch Current Based on Genetic Algorithm to Optimize the BP Neural Network*[J]. Lecture Notes in Electrical Engineering, 2011, 99:943-950.

[12]B.U.Islam, Z.Baharudin, M.Q.Raza, et al. *Optimization of neural network architecture using genetic algorithm for load forecasting*[M]. 2014.

[13]J.H.Holland, *Genetic algorithms and classifier systems: foundations and future directions, Genetic algorithms and their applications*, Proceedings of the second international conference on genetic algorithms[C].Beijing, china, 1997:626-630 .

[14]Y.W.Xiong, X.D.Zhang, and S.Y.Xie, *Effect of thermal fatigue on three-dimensional blade tip clearance of aero-engine*, presented at 14th Conference on Test and Measurement Technology of Aero-engine, Xiamen, China, 5–7 Nov. 2018.

Research on the Variation Mechanism of the 3-D Tip Clearance of a Cracked Blade under Multi-parameters in the Aero-engine Acceleration Process

Yiwei XIONG[1], Xiaodong ZHANG[1, 2], Siying XIE[1], Hongcheng LIU[1]

[1]School of Mechanical Engineering, Xi'an Jiaotong University, No.28 Xianning West Road, 710049, Xi'an, Shaanxi, P.R. China
[2]Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, No.28 Xianning West Road, 710049, Xi'an, Shaanxi, P.R. China
xiong1w@stu.xjtu.edu.cn; xdzhang@mail.xjtu.edu.cn

Abstract

The 3-D tip clearance has some advantages over the traditional radial tip clearance in the fault diagnosis of the turbine blade crack. The research on the variation mechanism of the 3-D tip clearance is of great significance, but previous researches only focused on the steady state condition of the aero-engine, lacking considerations of the aero-engine acceleration process. In this study, a numerical model of high pressure turbine blisk was established, and the centrifugal load, thermal load and aerodynamic load of the aero-engine, varying with time, were considered in this model. Besides, the cracks with different length and location were added to the trailing edge of the blade, and the 3-D tip clearance of cracked blades was calculated. The results show that there are some obvious distinctions in the 3-D tip clearance between a normal blade and cracked blades, which can reflect blade crack information accurately and effectively. The results also indicate that the 3-D tip clearance is promising in fault diagnosis of the turbine blade crack.

1 Introduction

The aero-engine is known as the heart of an airplane, providing power for the flight. The turbine is one of key components of the aero-engine, whose health status directly affects the safety and stability of the airplane. One crucial parameter to monitor the health status of the turbine is the blade tip clearance, which has a significant influence on the performance of the turbine [1]. The efficiency of the turbine can be obviously improved by decreasing the tip clearance, and the fuel consumption of the aero-engine can be reduced as well [2]. However, if the value of the tip clearance is too small, the catastrophic rubbing fault may happen to the turbine. Therefore, the tip clearance can reflect the operational status information of the turbine.

The tip clearance usually refers to the radial distance between the blade tip surface and the inner surface of the casing. Due to the blades are subjected to large centrifugal loads, thermal loads and aerodynamic loads when the turbine is running, the blades are prone to failure, and the crack is the most typical fault of the turbine blades. The crack will cause the deformation of the turbine blade, which will further lead to the changes in the tip clearance. Thus, blade crack information can be obtained by monitoring the tip clearance.

For the purpose of diagnosing the blade crack fault through the tip clearance, it's very crucial to research the variation mechanism of the blade tip clearance. Many researches have been done by numerous scholars and experts. Lattime and Steinetz of NASA Glenn Research Center [3] have pointed out that the loads affecting the tip clearance of the high pressure turbine include engine loads and flight loads. Kypuros [4] and Harish [5] have estimated the tip clearance value of the turbine blade by establishing simplified mathematical models of the tip clearance, which take into account the radial deformation of the turbine blade, disk and casing. Chapman [6] has established a universal and realistic high pressure turbine tip clearance model, which has been integrated with a gas turbine engine simulation system to build a test platform for investigating engine performance by adjusting the tip clearance.

At the same time, many researchers have studied on the fault mechanism and diagnosis method of the crack. Poursaeidi [7] and Wassim [8] have investigated the causes of the blade crack initiation through the mechanical, metallography and chemical analysis, and the stress and strain values of a cracked blade have been obtained by performing a finite element analysis. In regard to the diagnosis method of the blade crack, the blade tip timing technique has been extensively studied [9-12]. Zhang [13] has proposed an approach for the blade crack diagnosis, combining with the blade tip time technique and the tip clearance information.

However, the previous researches on the variation mechanism of the tip clearance only concerned about the one-dimensional radial tip clearance. In fact, the deformation of a blade becomes more complicated when there is a crack on the blade, which will lead to three-dimensional spatial characteristics of the tip clearance, so the tip clearance is three-dimensional actually, such as Figure 1. Besides the radial clearance, there are two more angles which are called axial angle and circumferential angle respectively. Zhang and Tei proposed the 3-D tip clearance and researched on the effect of blade crack on the 3-D tip clearance [14]. But the variation of the 3-D tip clearance was studied only in constant turbine conditions.

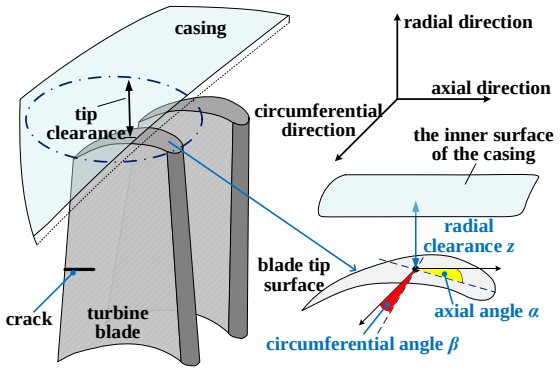


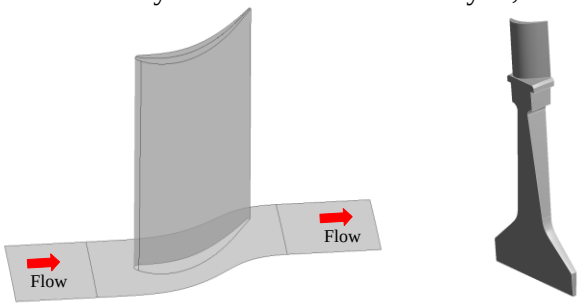
Figure 1: The 3-D tip clearance of a cracked turbine blade

Actually, in the aero-engine acceleration process, the rapidly increasing of turbine loads can easily lead to the crack propagation. Therefore, it's necessary to research the variation mechanism of a cracked blade during the acceleration process. In this study, first, numerical models of the fluid flow passage with a blade and the high pressure turbine blisk with the blade crack were established. Second, multi-loads including the time-varying centrifugal load, thermal load and aerodynamic load were applied to the model to compute the deformation of the turbine blisk, and then the 3-D tip clearance was calculate. Finally, the variation rule of the 3-D tip clearance of cracked blades in the aero-engine acceleration process was obtained, and the effect of the crack length and location on the 3-D tip clearance was analysed.

2 Numerical method

2.1 Geometrical model and grid

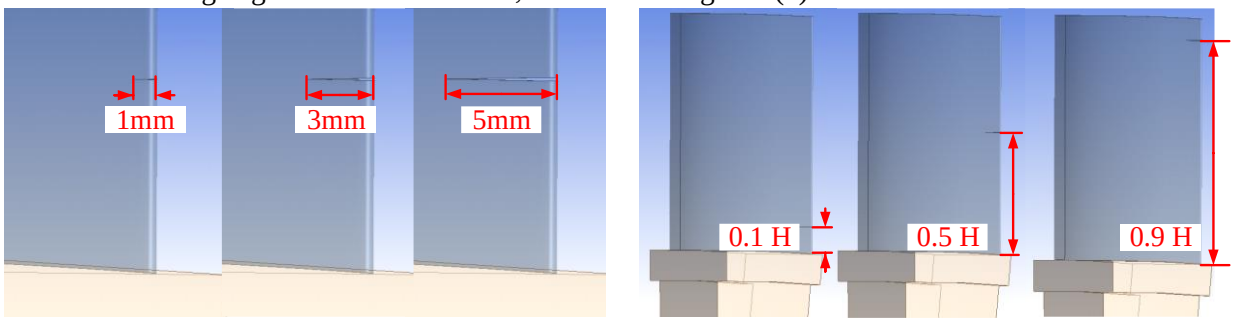
The simplified 3-D geometries of the flow passage with a blade and the blisk are shown in Figure 2. Ansys BladeGen was used to create the computational domain of the fluid flow analysis, which included only one rotor blade shown as Figure 2(a). The turbine blisk consisted of 60 blades, but the sector with only one blade was created to analyse the 3-D tip clearance in order to reduce the amount of the calculation. The geometry of the blisk was used in both the thermal analysis and the structural analysis, shown as Figure 2(b).



(a) The flow passage with one blade (b) The turbine blisk

Figure 2: The 3-D geometries of the flow passage with one blade and the turbine blisk

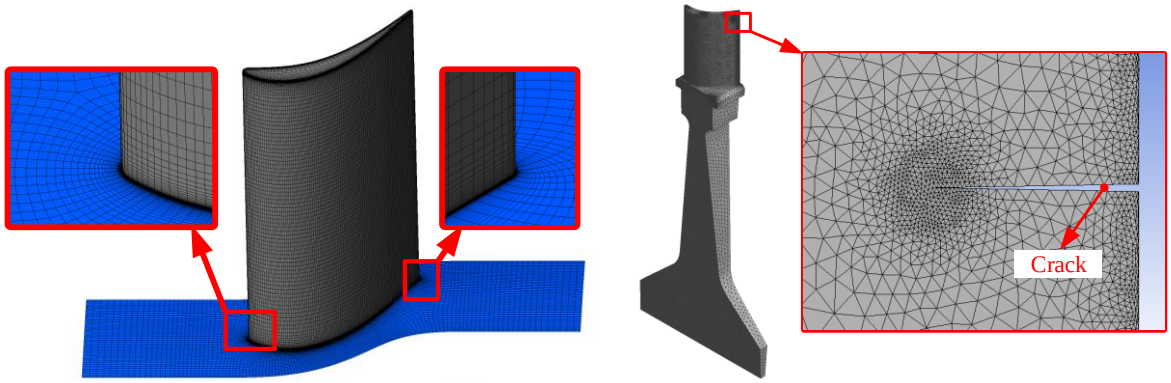
In order to investigate the variation mechanism of the 3-D tip clearance of cracked blades, several blade cracks with different length and location were added to the trailing edge of the turbine blade in this study, as shown in Figure 3. Three cracks with the length of 1mm, 3mm and 5mm were added to the turbine blade at the same location of 0.1H from the blade root, where H is the span of the turbine blade, as shown in Figure 3(a). Moreover, three cracks at different locations of 0.1H, 0.5H and 0.9H with the same length of 5mm were added at the trailing edge of the turbine blade, as shown in Figure 3(b).



(a) Cracks with different length (b) Cracks at different location

Figure 3: Different cracks at the trailing edge of the turbine blade

In this study, ANSYS TurboGrid was used to generate the structured hexahedral grid for the computational domain of the fluid flow analysis. A total of 120 layers of grids were generated along the span-wise direction and 50 layers of grids were inserted to the blade tip region, and the total number of elements was more than 2×10^6. Furthermore, in order to accurately obtain the thermal and flow characteristics of the computational domain, the grid within the boundary layers was refined to ensure y^+ was equal to 1 at the blade surface. Figure 4(a) presents the computational grids of the fluid domain with a blade.



(a) The computational grids of the fluid domain (b) The grid of the blisk and the crack front

Figure 4: The computational grids of the fluid domain and the turbine blisk

As for the thermal analysis and the structural analysis, the tetrahedral grid was generated for the geometry of the turbine blisk as shown in Figure 2(b). The grid around the crack front was refined to obtain accurate calculation results, as shown in Figure 4(b). Furthermore, the fracture mesh was inserted to generate the crack grid and pre-meshed crack was used in this study. The fracture mesh is only supported in the structural analysis, with which the mechanical properties of the cracked blades can be simulated more accurately.

2.2 Grid independence of the fluid flow analysis

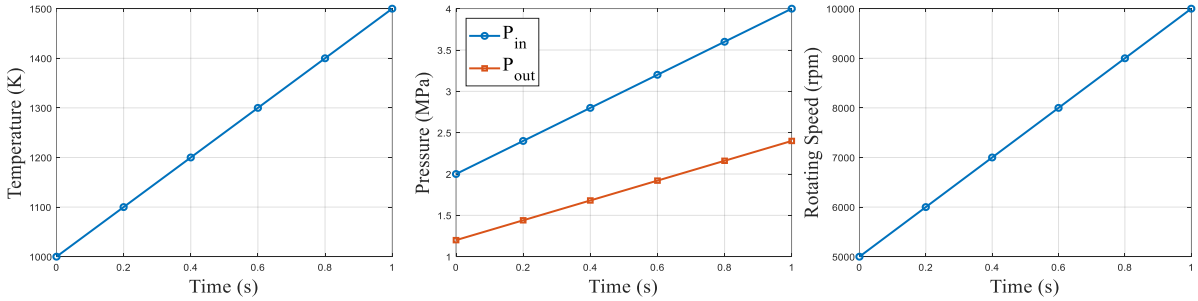
In order to ensure the reliability of the results of the fluid flow analysis and determine the appropriate number of elements to be used, a grid independence test was performed. Table 1 lists the detailed comparison of several grids with different number of nodes. The heat transfer coefficient is defined as $h = q\left(T_w - T_{aw}\right)$, where q is the heat flux of the wall surface. T_w and T_{aw} are the constant wall temperature under the isothermal boundary conditions and the adiabatic wall temperature under the adiabatic boundary conditions, respectively. The area-averaged heat transfer coefficient on the blade surface and its relative error was calculated as shown in Table 1. The relative error between the heat transfer coefficient of the No. 4 grid and that of the No. 3 grid is less than 1%, thus the No.4 grid was used for the computational domain in the fluid flow analysis.

No.	Number of nodes	Heat transfer coefficient ($W/(m^2K)$)	Relative error (%)
1	926250	5510.3362	-
2	1226890	5166.788	6.2346
3	1602110	5097.8388	1.3345
4	2096280	5112.9162	0.2958
5	2758200	5135.254	0.4369
6	3612500	5148.3241	0.2545

Table 1: Area-averaged heat transfer coefficient on the blade surface

2.3 Boundary conditions

The time-varying loads in the aero-engine acceleration process need to be determined to analyse the variation mechanism of the 3-D tip clearance of a cracked blade. As shown in Figure 5(a) and (b), the inlet total temperature, inlet total pressure and outlet average static pressure of the turbine were defined with regard to the fluid flow analysis. Figure 5(c) shows the rotating speed of the turbine rotor. To reduce the amount of calculation, the duration of the acceleration process was reduced to one second, and all of the parameters were assumed to vary with time linearly.



(a) The inlet temperature (b) The inlet and outlet pressure (c) The rotating speed

Figure 5: Time-varying loads in the aero-engine acceleration process

The surface temperature and pressure distributions of the blade obtained from the fluid flow analysis were imported into the thermal analysis and the structural analysis as boundary conditions, respectively. The body temperature distribution of the blisk, as the result of thermal analysis, was also imported into structural analysis to calculate the deformation of the blisk. Because only a sector of turbine blisk was used in the thermal and structural analysis, the cyclic symmetry boundary was defined on the sector of the turbine blisk.

2.4 Monitoring locations on the tip surface

After the structural analysis of the turbine blisk, the deformation at the blade tip surface needs to be measured to calculate the 3-D tip clearance. Four monitoring locations were chosen on the blade tip surface and each one consisted of three monitoring points, arranged in an isosceles right triangle as shown in Figure 6. Monitoring locations 1~3 locate at the leading edge, mid-chord and trailing edge of the blade tip respectively, and the distance between two monitoring points in the y axis and z axis direction is 2mm. Monitoring location 4 also locates at the trailing edge, but the distance between two measuring points is 0.8mm because of the thin thickness of the trailing edge. On each monitoring point, a vertex is created so that the deformations of the monitoring point in x, y and z directions can be measured.

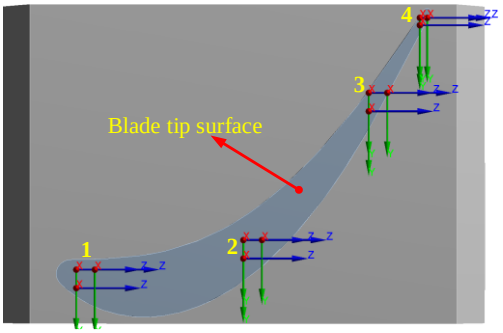
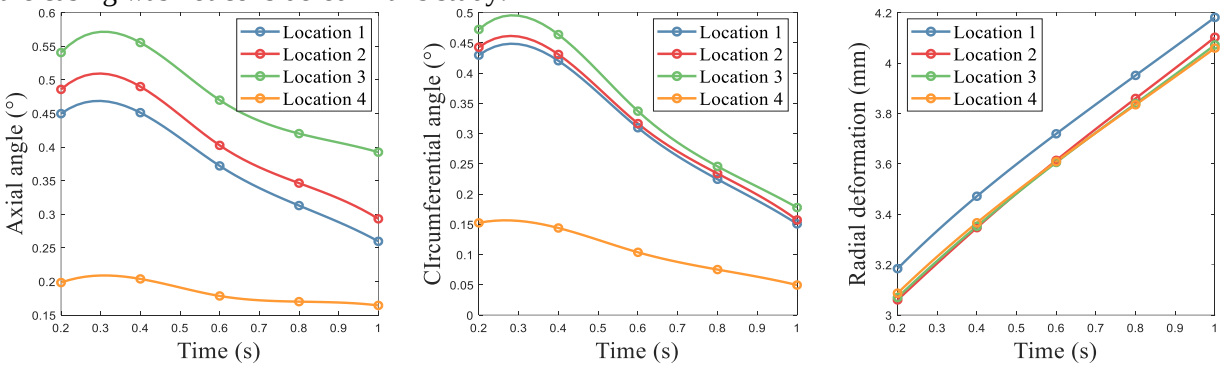


Figure 6: Monitoring points on the blade tip surface

3 Results and discussion

3.1 The 3-D tip clearance of a cracked blade in the acceleration process

A cracked blade with the length of 5mm at the location of 0.1H from the blade root was analysed. The blade tip deformations at four monitoring locations (shown in Figure 6) were measured, and the 3-D tip clearance in the aero-engine acceleration process was calculated as shown in Figure 7. It should be noted that the radial tip clearance was represented by the radial deformations of the turbine blisk because the deformation of the casing was not considered in this study.



(a) Axial angle (b) Circumferential angle (c) Radial deformation

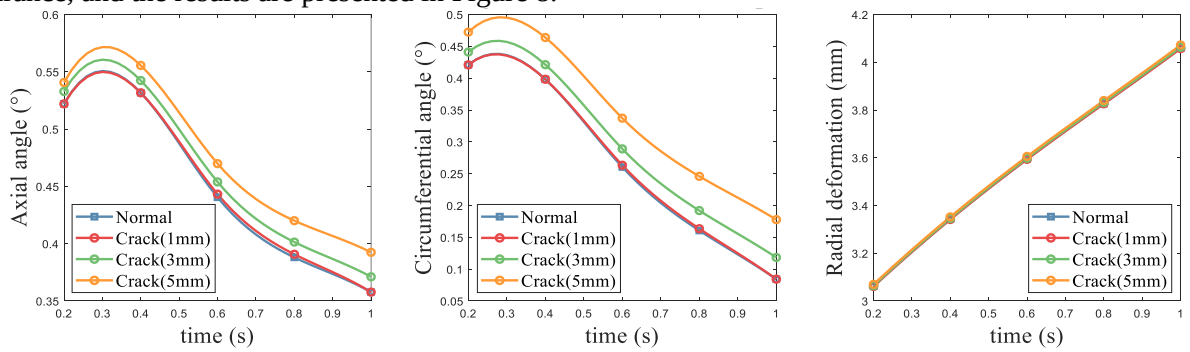
Figure 7: The 3-D tip clearance of a cracked blade in the acceleration process

In the aero-engine acceleration process, both the axial angle and circumferential angle decrease with time. Though the radial deformation of the blisk increases with time, the actual radial tip clearance still decreases because the casing is not as large in deformation as the blisk. At the same time, comparing the axial angle and circumferential angle at the monitoring locations 1, 2, and 3, the maximums of the two angles appear at the trailing edge (monitoring location 3), and the values of the two angles get smaller when the monitoring locations are away from the trailing edge of the blade. Therefore, the closer the monitoring points are to the blade crack, the more sensitive the 3-D tip clearance is to the crack failure.

However, the monitoring location 4 are much closer to the trailing edge than the monitoring location 3, but the axial angle and circumferential angle are much smaller at the monitoring location 4. So comparing with the 3-D tip clearance at the monitoring location 4, the 3-D tip clearance at the monitoring location 3 is more sensitive to the crack because the distance between the two monitoring points of the monitoring location 3 is larger. Thus, the distance between the two monitoring points have a significant effect on the sensitivity of the 3-D tip clearance to the blade crack, and the distance of 2mm is better than that of 0.8mm in this study.

3.2 The effect of the crack length on the 3-D tip clearance

Three cracks with different length at the same location (shown in Figure 3(a)) were analysed to study the effect of the crack length on the 3-D tip clearance. From the aforementioned analysis, the 3-D tip clearance of the monitoring location 3 is more sensitive to the crack fault than the 3-D tip clearance of the other monitoring locations, so the blade tip deformations at the monitoring location 3 are obtained to calculate the 3-D tip clearance, and the results are presented in Figure 8.



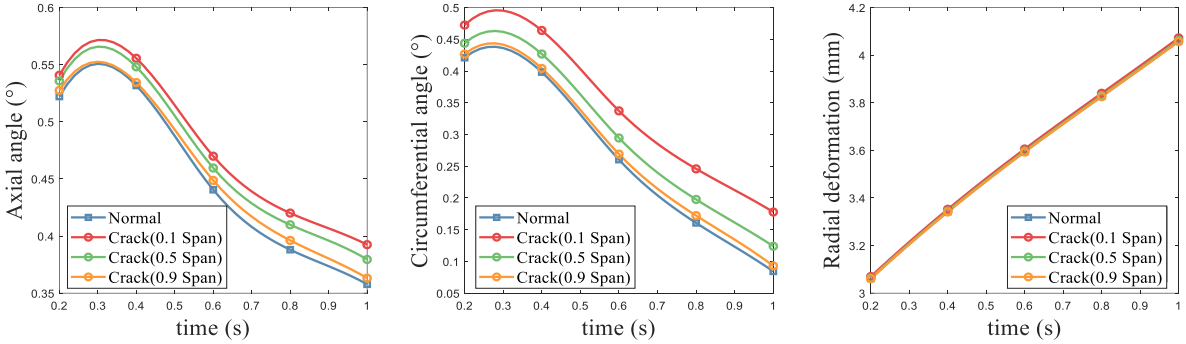
(a) Axial angle (b) Circumferential angle (c) Radial deformation

Figure 8: The 3-D tip clearance of the fault blades with different crack length at the monitoring location 3

The radial deformation of the normal blade and the cracked blades are almost the same, but there are some differences between the normal blade and the cracked blades in the axial angle and circumferential angle. Comparing the normal blade with the cracked blade with the crack length of 1mm, the axial angle and circumferential angle of the two blades are nearly the same, but with the increase of the crack length, the axial angle and circumferential angle of the cracked blades also increase obviously. The crack has a great effect on the stiffness of the blade, and the longer the crack length is, the smaller the blade stiffness becomes. Therefore, in the identical acceleration process, the blade with a longer crack has a larger deformation, and also has larger axial angle and circumferential angle.

3.3 The effect of the crack location on the 3-D tip clearance

In order to study the effect of the crack location on the 3-D tip clearance, three cracks at different location with the same length (shown in Figure 3(b)) were analysed. As mentioned above, the sensitivity of the 3-D tip clearance to the crack is the highest at the monitoring location 3, thus the 3-D tip clearance of the normal and cracked blades at the monitoring location 3 are presented in Figure 9.



(a) Axial angle (b) Circumferential angle (c) Radial deformation

Figure 9: The 3-D tip clearance of the fault blades with different crack location at the monitoring location 3

It is extremely hard to tell the differences between the normal blade and the cracked blades from the radial deformation, but the cracked blades at different location have different axial angle and circumferential angle. As shown in Figure 9, the axial angle and circumferential angle of the normal blade are the smallest compared with the axial angle and circumferential angle of the other cracked blades. Besides, the axial angel and the circumferential angle get larger when the location of the crack gets closer to the blade root.

Under the thermal load, centrifugal load and aerodynamic load, the bending and torsional deformations will happen to the turbine blade, and the most severe bending and torsional deformations occur at the location of the crack because of the low stiffness at the crack. At the same time, the deformation of the part above the crack of the blade has an effect on the axial angle and circumferential angle. Therefore, the closer the crack is to the blade root, the larger the deformation of the blade, resulting in the increase of the axial angle and circumferential angle. Thus the maximums of the axial angle and circumferential angle occur at the location of 0.1H from the blade root in this study.

Both Figure 8 and Figure 9 show the radial deformation of the normal blade and the cracked blades are almost the same, but the axial angle and circumferential angle are obviously different. So the 3-D tip clearance indeed contains more abundant fault information than traditional radial tip clearance, which can be used to fault diagnosis of the turbine blade crack in the further.

4 Conclusions

In this study, the numerical models of the flow passage with one rotor blade and the turbine blisk was established to research the variation mechanism of the 3-D tip clearance in the aero-engine acceleration process, and the thermal load, centrifugal load and aerodynamic load were added to this model. A normal blade and different cracked blades were analysed and four monitoring locations were chosen to study the effects of the crack length and location on the 3-D tip clearance. The results show the monitoring location which is closer to the trailing edge of the blade, is more sensitive to the crack fault, and the distance between two monitoring

points shouldn't be too small. Both the length and the location of the crack have an effect on the 3-D tip clearance. The longer the crack and the closer the crack is to the blade root, the larger the axial angel and the circumferential angle, however, the radial deformations of normal and cracked blades are almost the same in the identical acceleration process. The 3-D tip clearance contains abundant fault information and can be used to the fault diagnosis of the blade crack, which will be further investigated in the future.

Acknowledgments

This study is supported by the National Natural Science Foundation of China (Grant No.51575436).

References

[1]Y.C. Nho, Y.J. Lee, J.S. Kwak, *Effects of Tip Shape on the Gas Turbine Blade Tip Heat Transfer*[J]. Journal of Thermophysics and Heat Transfer, 2012,26(2):305-312.

[2]S.B. Lattime, B.M. Steinetz, *High-pressure-turbine Clearance Control Systems: Current Practices and Future Directions*[J]. Journal of Propulsion and Power, 2004,20(2):302-311.

[3]S.B. Lattime, B.M. Steinetz, *Turbine Engine Clearance Control Systems: Current Practices and Future Directions*, 2003[C].

[4]J.A. Kypuros, K.J. Melcher, *A Reduced Model for Prediction of Thermal and Rotational Effects on Turbine Tip Clearance*[R].2003.

[5]H. Agarwal, S. Akkaram, S. Shetye, et al, *Reduced Order Clearance Models for Gas Turbine Applications*: Aiaa/asme/asce/ahs/asc Structures, Structural Dynamics, & Materials Conference Aiaa/asme/ahs Adaptive Structures Conference 10t, 2013[C].

[6]J.W. Chapman, J. Kratz, T. Guo, et al, *Integrated Turbine Tip Clearance and Gas Turbine Engine Simulation*[R].2016.

[7]E. Poursaeidi, M. Aieneravaie, M.R. Mohammadi, *Failure Analysis of a Second Stage Blade in a Gas Turbine Engine*[J]. Engineering Failure Analysis, 2008,15(8):1111-1129.

[8]W. Maktouf, K. Sai, *An Investigation of Premature Fatigue Failures of Gas Turbine Blade*[J]. Engineering Failure Analysis, 2015,47(A):89-101.

[9]C.P. Lawson, P.C. Ivey, *Tubomachinery Blade Vibration Amplitude Measurement Through Tip Timing With Capacitance Tip Clearance Probes*[J]. Sensors & Actuators A: Physical, 2005,118(1):14-24.

[10]S. Madhavan, R. Jain, C. Sujatha, et al, *Vibration Based Damage Detection of Rotor Blades in a Gas Turbine Engine*[J]. Engineering Failure Analysis, 2014,46:26-39.

[11]R.G. Du Toit, D.H. Diamond, P.S. Heyns, *A Stochastic Hybrid Blade Tip Timing Approach for the Identification and Classification of Turbomachine Blade Damage*[J]. Mechanical Systems and Signal Processing, 2019,121:389-411.

[12]Z. Chen, Y. Yang, Y. Xie, et al, *Non-contact Crack Detection of High-speed Blades Based on Principal Component Analysis and Euclidian Angles Using Optical-fiber Sensors*[J]. Sensors and Actuators A: Physical, 2013,201:66-72.

[13]J.W. Zhang, L.B. Zhang, L.X. Duan. *A Blade Defect Diagnosis Method by Fusing Blade Tip Timing and Tip Clearance Information*[J]. Sensors, 2018,18(7):2166.

[14]F. Teng, X.D. Zhang, S.Y. Xie, *Research on Variation Mechanism of Three-dimensional Blade Tip Clearance of Aero-engine*: 2016 13th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), 2016[C].

Modeling and identification of mechanical systems

The Dynamics of Helicopters with Nonlinearities on the Fuselage

Cesar A. B. da SILVA[1]

[1]School of Mechanical Engineering, UFU - Federal University of Uberlândia, Brazil
cabdasr@hotmail.com

Leonardo SANCHES[2*], Guilhem MICHON [2+]

[2]Université de Toulouse, ICA, CNRS, ISAE-Supaero, Toulouse, France
[*] leonardo.sanches@isae-supaero.fr, [+] guilhem.michon@isae-supaero.fr

Abstract

The present paper aims at evaluating the influence of nonlinearities on the dynamic response of an helicopter on the ground. Indeed, the helicopter on the ground might be faced to resonance and instability conditions that could lead to high vibration levels. Nonetheless, under the effects of nonlinearities, the dynamical behaviour of the aircraft could have substantial changes in which quasi-periodic and/or chaotic motions can occur. Poincaré section and bifurcation diagrams are evaluated for different helicopter configurations in order to highlight the dynamical behavior of helicopters with nonlinearities in the fuselage.

1 Introduction

Concerning the helicopter dynamics, one aims avoiding any rise of vibration level during operational conditions. For this purpose, the comprehension of the dynamics of the mechanical components and how they interact with the environment are necessaries. Beyond these aspects, the interaction of the helicopter with the terrain during take-off and landing can cause multiples resonances which might lead to ground resonance with fatal consequences [13, 2].

In order to suppress the ground resonance, among other existent means, elastomeric dampers can play a significant role in the stability of the aircraft. Therefore, the modeling of the elastomeric lag dampers have received increasing attention, specially concerning its nonlinear characteristics [1, 6, 4]. For example, Gandhi and Chopra [1] place an additional nonlinear spring in series with the linear, parallel spring and dashpot in order to represent the elastomeric dampers. Other possibility is to consider the inclusion of pneumatic elements on the aircraft suspension for absorbing and dissipating the vibrating energy [13].

Nonetheless, these dissipative elements and some structural ones introduce nonlinearities that might substantially affect the dynamics of the helicopter [5].

It is known from the literature that rotating machines under nonlinear operating characteristics could attain other than the periodic motion, predictable for linear systems, such as the quasi-periodic or chaotic motions. Recently, Varney and Green observed the presence of quasi-periodic and chaos on the rotor dynamics by assuming rotor-stator contact [14]. Indeed this phenomenon is observed since precisely manufactured bearings are used and thus reduced clearances are imposed for improving the performance of the rotating machines.

Under the hypothesis of nonlinearities for the ground resonance modelling purposes, this paper aims verifying, their effects on the helicopter response. if at certain operating conditions or design properties, the appearance of non-periodic and/or chaotic motion are reached. The helicopter considered contains nonlinear spring stiffness on the fuselage displacements (longitudinal and lateral). The four-bladed rotor consider rigid structures having only lead-lag oscillations. Parametric analyses combined with examination of the Poincaré maps and bifurcation diagrams, the nonlinear dynamical behavior of the helicopter was assessed.

2 Mechanical Modeling

The physical model of the helicopter is based on that presented by [10]. The hypotheses assumed are sufficient for computing accurately (through a linear model) the ground resonance instability of an helicopter with four ($N_b = 4$) articulated bladed rotor.

The aircraft is considered having two translation movements, along x and y-directions (i.e.: $x(t)$ and $y(t)$, respectively) for the fuselage and the angular displacement of each blade ($\varphi_k(t)$,k=1..4). Moreover, it is supposed that the fuselage and blades structures are sufficiently rigid in which they can be modeled as rigid body. No aerodynamic effects are take into account and the rotor speed is assumed constant Ω.

The blades are connected to rotor hub through a hinge axis. At each blade hinge, a set of spring (K_{bk}) and damper (C_{bk}) elements is placed. The landing skids are represented by a set of nonlinear elastic and linear viscous damping forces in both directions of the fuselage movements. Indeed, the nonlinear elastic function can represent the different sources of nonlinearities, e.g.: suspensions forces, lading skid structure or the soil restitution force. In the nonlinear elastic force acting on the fuselage, the k_{fx} and k_{fy} are the coefficients of the linear terms, while k_{nlx} and k_{nly} are the coefficients of the cubic ones. Others polynomial orders are not considered in the present work. The damping coefficients are c_{fx} and c_{fy} along the x and y directions, respectively. The Figure 1 sketches the mechanical model adopted for the helicopter.

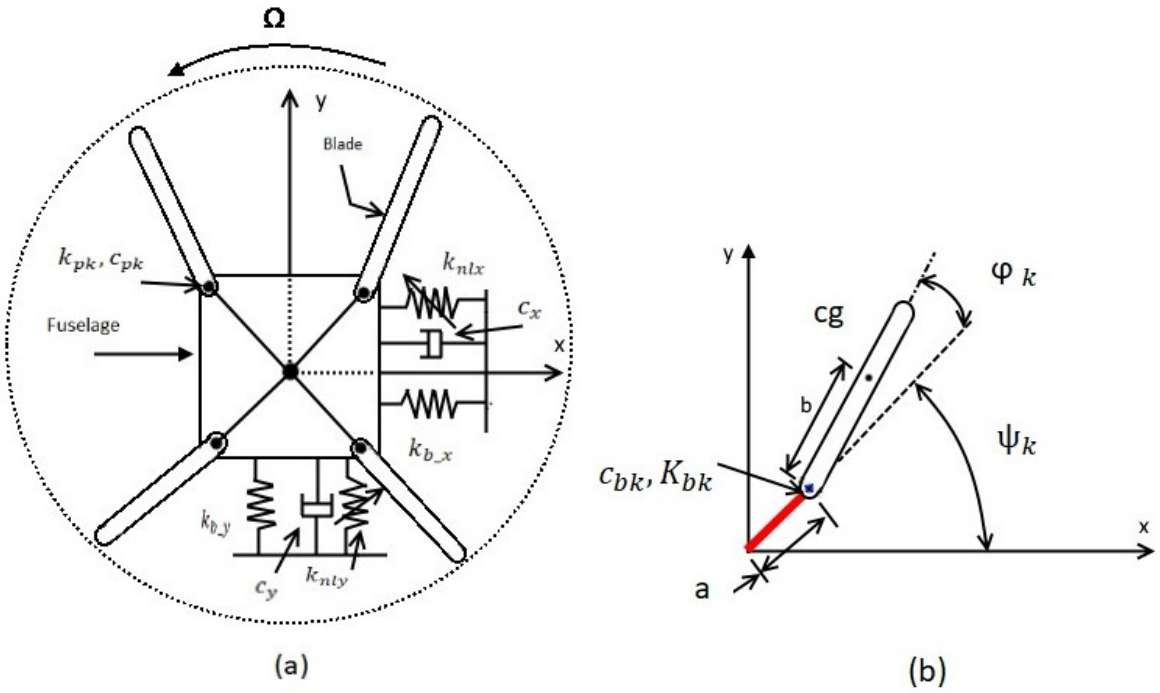


Figure 1 – Sketch of Helicopter Mechanical Model

The equations of motion are obtained through the Lagrange equation [10, 9, 15] applied at the kinetic and potential energies expressions and by considering the virtual work of the non-conservative forces on the system. Under the hypothesis of small angular displacements $\varphi_k(t)$, the trigonometric terms can be reduced to linear expressions when small perturbation theory is considered [9]. Thus, the linear matrix equation, Eq. 1, can be easily determined as:

$$\mathbf{M}(t)\ddot{\mathbf{u}} + \mathbf{G}(t)\ddot{\mathbf{u}} + \mathbf{K}(t)\mathbf{u} = \mathbf{F_{nl}} \tag{1}$$

with $\mathbf{M}(t)$, $\mathbf{G}(t)$ and $\mathbf{K}(t)$ are the mass, damping / gyroscopic and stiffness matrix. The vector $\mathbf{u}$ corresponds to $[x(t), y(t), \varphi_1(t), \varphi_2(t), \varphi_3(t), \varphi_4(t)]^T$ which are the helicopter degrees of freedom. The $\mathbf{F_{nl}}$ represents the vector of nonlinear terms of the helicopter model.

It is important to note the time dependence of the matrices since the blades angular displacements are given on a rotating frame while fuselage movements are described in the inertial one. Using the Multi-blade coordinate transformation as showed in [12, 11], such time dependent matrix might be reduced to one with constant coefficients. The new vector of generalized coordinates $\mathbf{q}(t)$ is obtained through the following relation:

$$\mathbf{q}(t) = \mathbf{T}(\mathbf{t})\ \mathbf{u}(t) \tag{2}$$

with,

$$[\mathbf{T}(\mathbf{t})] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & \cos(\Omega t) & \sin(\Omega t) \\ 0 & 0 & 1 & -1 & -\sin(\Omega t) & \cos(\Omega t) \\ 0 & 0 & 1 & 1 & -\cos(\Omega t) & -\sin(\Omega t) \\ 0 & 0 & 1 & -1 & \sin(\Omega t) & -\cos(\Omega t) \end{bmatrix} \tag{3}$$

Once the Eq.3 has been replaced in Eq.1, after some mathematical manipulations, the helicopter dynamics is given by:

$$\mathbf{M_C}\ddot{\mathbf{q}} + \mathbf{G_C}\dot{\mathbf{q}} + \mathbf{K_C}\mathbf{q} = \mathbf{F_{nl}} \tag{4}$$

where,

$$\mathbf{M_C} = \begin{bmatrix} \mathrm{m}_f + N_b m_b & 0 & 0 & 0 & -2\,\mathrm{b}\,\mathrm{m}_b & 0 \\ 0 & \mathrm{m}_f + N_b m_b & 0 & 0 & 0 & 2\,\mathrm{b}\,\mathrm{m}_b \\ 0 & 0 & \mathrm{N}_b\,(m_b b^2 + I_{zb}) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{N}_b\,(m_b b^2 + I_{zb}) & 0 & 0 \\ -2\,\mathrm{b}\,\mathrm{m}_b & 0 & 0 & 0 & \mathrm{N}_b/2(m_b b^2 + I_{zb}) & 0 \\ 0 & 2\,\mathrm{b}\,\mathrm{m}_b & 0 & 0 & 0 & \mathrm{N}_b/2(m_b b^2 + I_{zb}) \end{bmatrix} \tag{5}$$

$$\mathbf{G_C} = \begin{bmatrix} \mathrm{c}_x & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{c}_y & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{N}_b c_b & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{N}_b c_b & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{N}_b/2\,c_b & -\mathrm{N}_b\,(m_b b^2 + I_{zb})\,\Omega \\ 0 & 0 & 0 & 0 & \mathrm{N}_b\,(m_b b^2 + I_{zb})\,\Omega & \mathrm{N}_b/2\,c_b \end{bmatrix} \tag{6}$$

$$\mathbf{K_C} = \begin{bmatrix} \mathrm{k}_{fx} & 0 & 0 & 0 & 0 & 0 \\ 0 & \mathrm{k}_{fy} & 0 & 0 & 0 & 0 \\ 0 & 0 & \mathrm{N}_b\,(k_b + \Omega^2 a b m_b) & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathrm{N}_b\,(k_b + \Omega^2 a b m_b) & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathrm{k}_\tau & -\,\mathrm{c}_b\,\Omega N_b/2 \\ 0 & 0 & 0 & 0 & \mathrm{c}_b\,\Omega N_b/2 & \mathrm{k}_\tau \end{bmatrix} \tag{7}$$

with $k_\tau = N_b/2(k_b + \Omega^2(a b m_b - I_{zb} - b^2 m_b))$ and $\mathbf{q} = [x(t), y(t), \tau_d(t), \tau_0(t), \tau_{1c}(t), \tau_{1s}(t)]^T$. The variables $\tau_d(t)$, $\tau_0(t)$, $\tau_{1c}(t)$, and $\tau_{1s}(t)$ are related to rotor modes of vibration and they replace the blade displacements on generalized coordinate vector. Through this coordinate transformation, the Eq. 1 becomes time-invariant (*c.f.* Eq.4).

3 Linear Stability Analysis

Several authors have addressed to the stability analysis of the ground resonance previously by considering linear dynamical models [15, 4]. Indeed, the stability analysis of the ground resonance will contribute for further comprehension of the behavior of the nonlinear dynamical system. Therefore, present section aims at verifying the stability of the linear LTI system by neglecting the nonlinear efforts $\mathbf{F}_{nl}$ from Eq. 4. The mechanical data for the stability analysis are given in Tab. 1.

It is important to note that the natural frequencies and damping factors written in Tab. 1 are defined with respect to the helicopter dynamical properties at rest with null rotor speed. Thus,

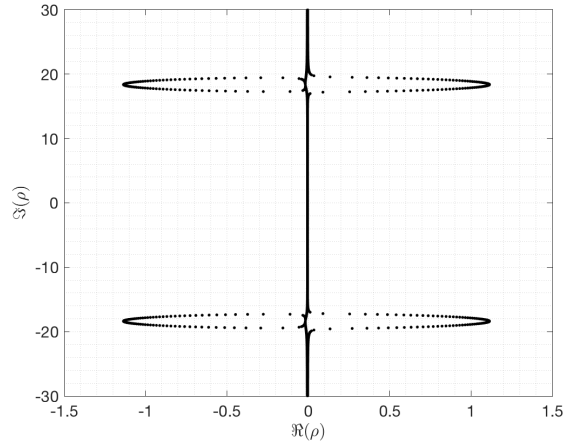
$$\omega_x = \sqrt{k_{fx}/(m_f + N_b m_b)} \qquad \omega_y = \sqrt{k_{fy}/(m_f + N_b m_b)} \qquad \omega_b = \sqrt{k_b/(I_{zb} + b^2 m_b)} \tag{8}$$

$$c_x = 2\xi_x \omega_x (m_f + N_b m_b) \qquad c_y = 2\xi_y \omega_y (m_f + N_b m_b) \qquad c_b = 2\xi_b \omega_b (I_{zb} + b^2 m_b) \tag{9}$$

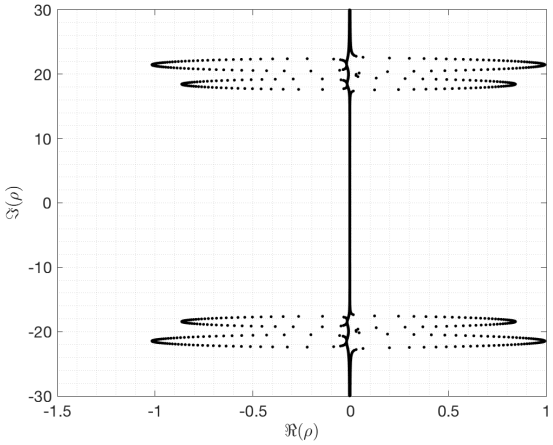
Property	Value	Units
m_f	2902	kg
m_b	31.9	kg
Iz_b	259	kg m^2
a	0.2	m
b	2.5	m
ω_x	6π	rad/s
ω_{bk} ($k = 1 \ldots 4$)	3π	rad/s
$\xi_x = \xi_y = \xi_{bk}$	0.1	%

Table 1 – Helicopter's Data

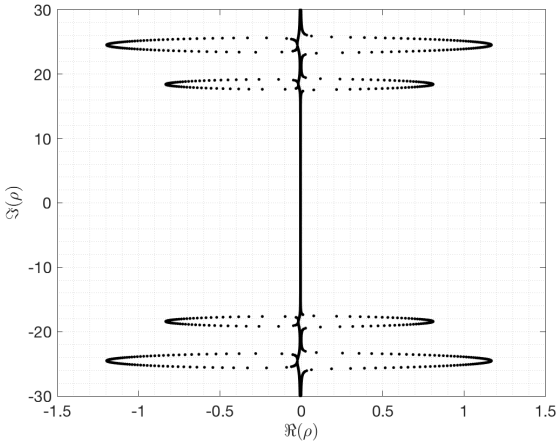
The stability boundaries are verified through the eigenvalues (ρ) of the linear system investigated for several rotor speed values (i.e., $0 \leq \Omega \leq 10$ Hz) and for three fuselage configurations, i.e: $\omega_y = [6, 7, 8]$ Hz. The results are illustrated in Figs. 2a to 2c for $\omega_y = 6$, 7 and 8 Hz, respectively.



(a) $\omega_y = 6\pi$



(b) $\omega_y = 7\pi$



(c) $\omega_y = 8\pi$

Figure 2 – Eigenvalue Evolution with respect to rotor speed and fuselage configurations.

One notice from each graph the changes of the natural frequencies and exponential decay constant of the dynamical system once the parameter (rotor speed) varies. For the three fuselage configurations analyzed, clearly it can be pointed out the existence of unstable regions (i.e., $\mathscr{R}(\rho) > 0$) for certain range of rotor speed values. These unstable regions are associated with the natural frequencies of the fuselage at rest. As the natural frequencies of the fuselage along x and y directions become dissimilar (*c.f.* Fig. 2b-2c), the instability

regions evolve leading to clearly two distinct zones: each one is related to each direction of fuselage oscillation. Nonetheless, when the fuselage frequencies are equals (i.e., $\omega_x = \omega_y$), the instability regions are then superposed. Specially concerning this last case, it is observed that two pairs of eigenvalues become unstable simultaneously, indicating the existence of double hopf bifurcation. Otherwise, only one pair of eigenvalue becomes unstable and therefore a single hopf bifurcation happens.

In order to determine the critical rotating speeds for each fuselage configuration, Figure 3 shows the maximum real part of the eigenvalues with respect to the rotor speed. The boundaries of instabilities are determined by inspecting when positive values are attained for the real part of the eigenvalues. The instability region for

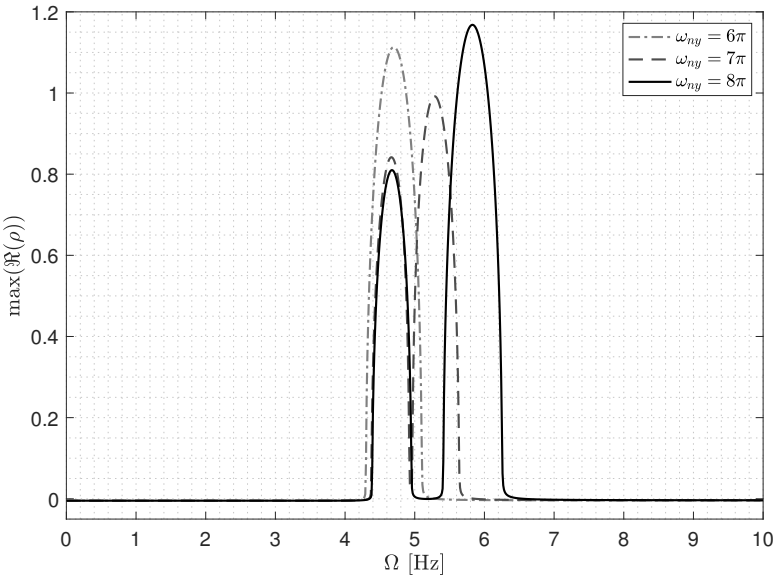


Figure 3 – Instability boundaries for different fuselage configurations

$\omega_y = 6\pi$ are given for rotor speed values within $4.25 < \Omega < 5.3$ Hz. When the fuselage natural frequency is changed to $\omega_y = 7\pi$, the critical speeds are within $4.33 < \Omega < 5.87$ Hz. Finally, when $\omega_y = 8\pi$, there are two unstable regions identified, i.e., $4.35 < \Omega < 5.08$ Hz and $5.23 < \Omega < 6.56$ Hz.

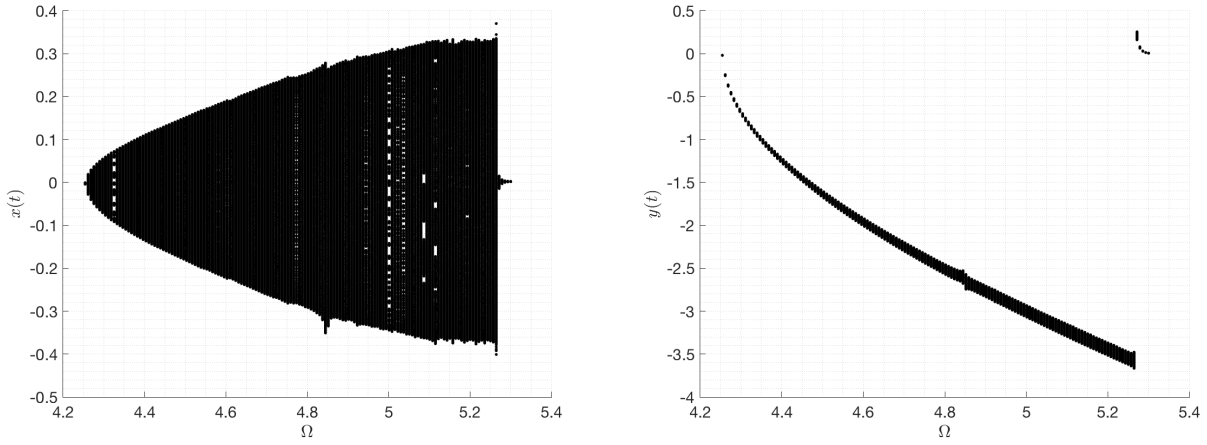
4 Bifurcation Diagrams

The present section aims at analyzing the effects of the nonlinear efforts on the dynamic response of the helicopter. Indeed, the restoring forces from the pneumatic components presented in some helicopter suspensions can be expressed through nonlinear functions [13, 7, 6]. In addition, for some cases, the interactions between the soil and aircraft might cause nonlinear efforts [2]. For this purpose, one assumes a pure cubic polynomial function acting along the fuselage displacements with the following coefficients $k_{nlx} = k_{nly} = 1x10^5 N/m^3$.

Since hopf bifurcation points on the helicopter dynamics were evidenced from the linear stability analysis (*c.f.* section 3), the literature shows that hopf bifurcation might leads to periodic motion which characterizes the limit cycle oscillations. Nonetheless, the periodic motion can evolve under control parameter variation and becomes into non-periodic one. It can further evolve from the non-periodic motion to a chaotic motion [3]. Therefore, Poincaré sections and bifurcation diagrams are evaluated from the time history data obtained through the numerical integration of the nonlinear equations (*c.f.,* Equation 4). The initial condition is the same for all simulations and it considers a shift of $0.1m$ for the fuselage displacements from the equilibrium position, while others displacements and speeds are nulls. The Poincaré section is considered into the plane $\tau_c = 0$. Henon algorithm is used for precisely obtain the points intercepting the Poincaré section and, therefore, used to obtain the bifurcation diagrams [8].

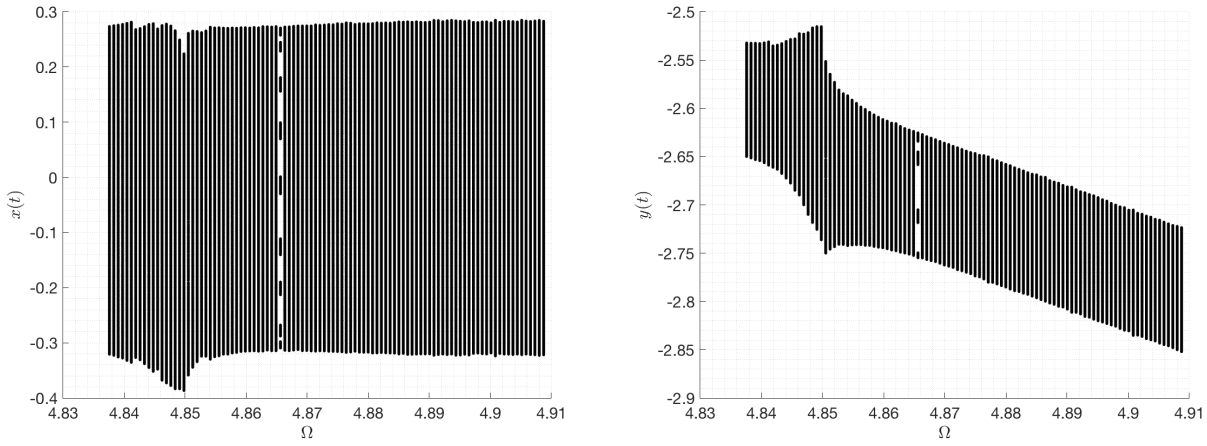
Figures 4 and 5 describe the bifurcation diagram for $x(t)$ and $y(t)$ with respect to Ω for the helicopter with identical natural frequencies of the fuselage (i.e., ω_x and ω_y are 6π rad/s). Since the Poincaré sections are evaluated in the plane $\tau_c = 0$, and thus for the bifurcation diagrams, one observes that $y(t)$ reaches higher amplitude level than $x(t)$, accordingly to Figures 4a) and 4b). This fact might be explained since the variables

$x(t)$ and $\tau_c(t)$ are in-phase motion, while $y(t)$ is quarter phase with $\tau_c(t)$. Moreover, concerning Figure 5 at $\Omega = 4.865$ Hz, the values of $x(t)$ and $y(t)$ are concentrated at some points instead of being uniformly distributed, as it can be evidenced for the whole rotor speed conditions analyzed. A closer investigation is carried out for comprehending the differences between the two cited cases. Therefore, the Poincaré section and the phase subspace for $\Omega = 4.85$ Hz and $\Omega = 4.865$ Hz are represented in Figures 6 and 7. Clearly, by comparing the Poincaré sections for two speed conditions, the Figure 6a represents a quasi-periodic behavior while at Figure 7a a periodic signal with nine harmonics is observed.



(a) (b)

Figure 4 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 6\pi$: a) $x(t)$ and b) $y(t)$

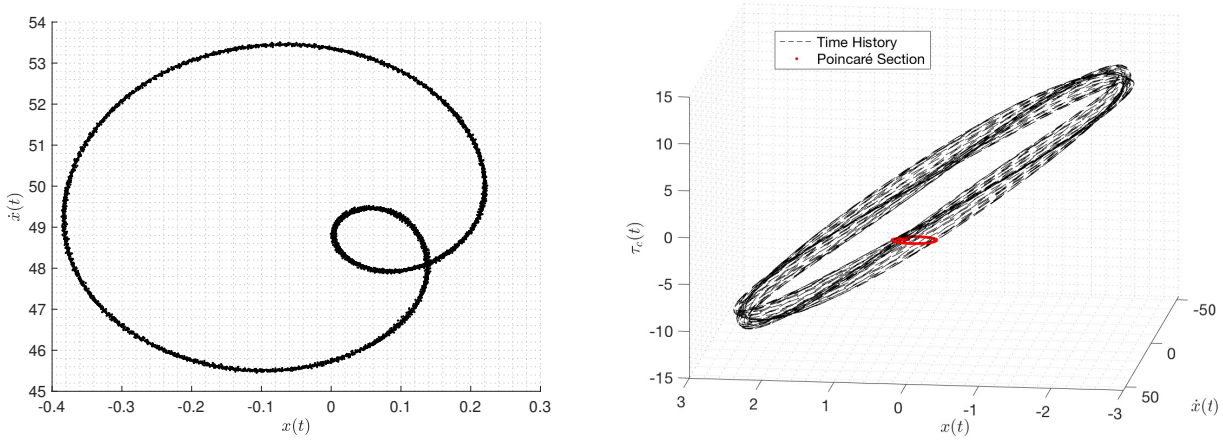


(a) (b)

Figure 5 – Detailed Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 6\pi$: a) $x(t)$ and b) $y(t)$

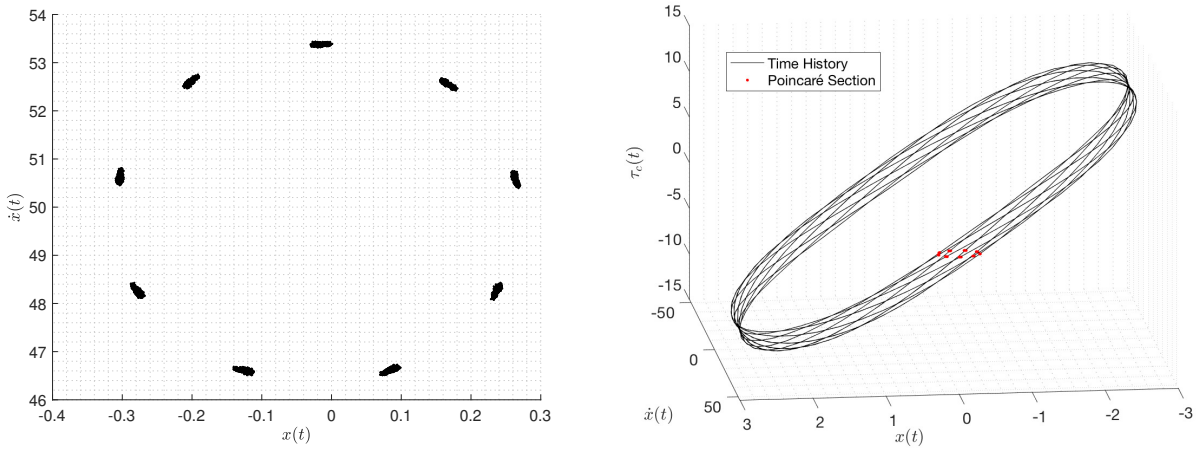
The helicopter is now considered to have an asymmetry between the frequencies of the fuselage i.e., the oscillations along x-direction is at $\omega_x = 6\pi$ rad/s while in y-direction it is at $\omega_y = 7\pi$ rad/s. The influence of the nonlinearity in the fuselage on the helicopter is evidenced through the bifurcation diagram in Figure 8. One observes that, as Ω increases, the bifurcation amplitude levels for x and y are also augmenting. At $\Omega = 5.4$ there is no movements noted, since the helicopter are nearly stable (*c.f.* Figure 3). Moreover, the amplitudes of the bifurcation section get increased for rotor speeds higher than 5.66 Hz. Figure 9 compares the evolution of the Poincaré sections for different rotor speed values in order to highlight the increment of the amplitude level on the bifurcation diagram. A closer inspection on the time response is done through Figure 10 in which a slight modulation of the signal for $\Omega = 5.59$ Hz is observed; while for $\Omega = 5.82$ Hz the signal modulation has significantly changed.

Finally, the helicopter with high asymmetry level on the fuselage frequencies is investigated, i.e., $\omega_x = 6\pi$



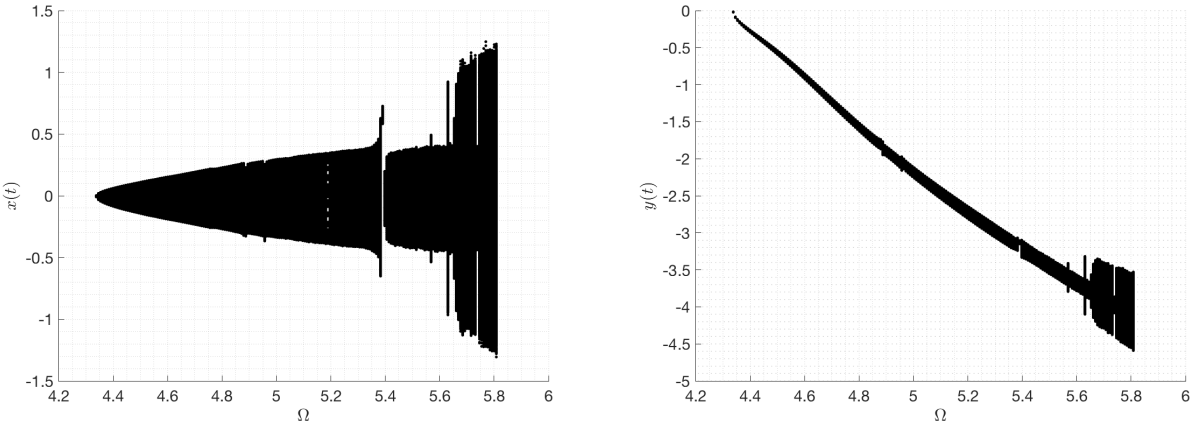
(a) Poincaré Section (b) Phase Subspace

Figure 6 – Poincaré section and phase subspace at $\Omega = 4.85$ Hz of an helicopter with $\omega_x = 6\pi$ and $\omega_y = 6\pi$



(a) Poincaré Section (b) 3D- Phase Subspace

Figure 7 – Poincaré section and phase subspace at $\Omega = 4.865$ Hz of an helicopter with $\omega_x = 6\pi$ and $\omega_y = 6\pi$



(a) (b)

Figure 8 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 7\pi$: a) $x(t)$ and b) $y(t)$

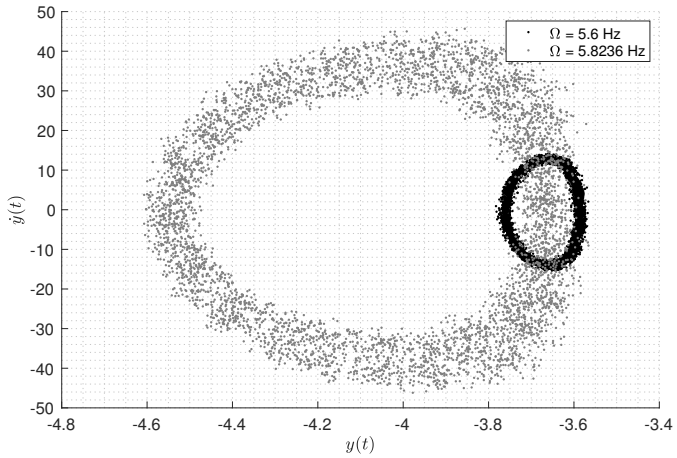
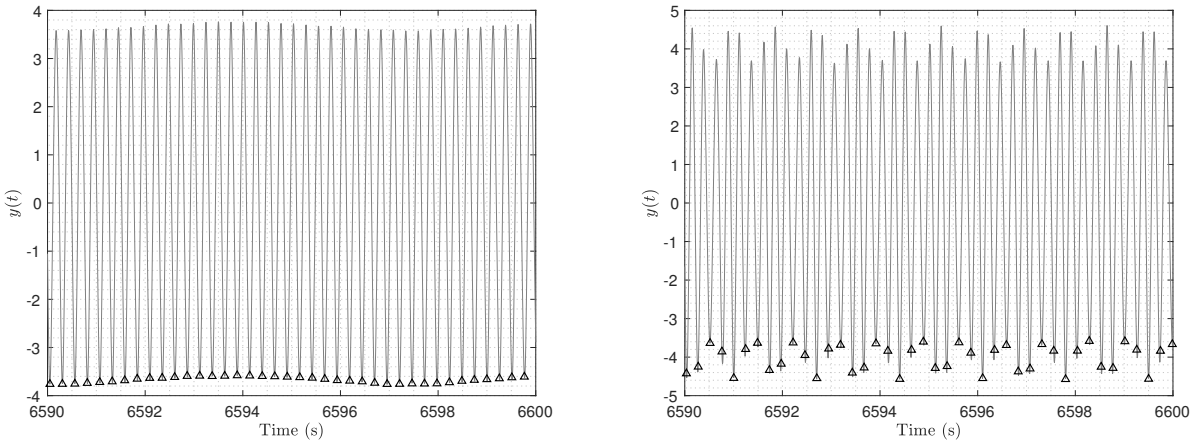


Figure 9 – Comparison of Poincaré Sections for Helicopter with $\omega_x = 6\pi$ and $\omega_y = 7\pi$

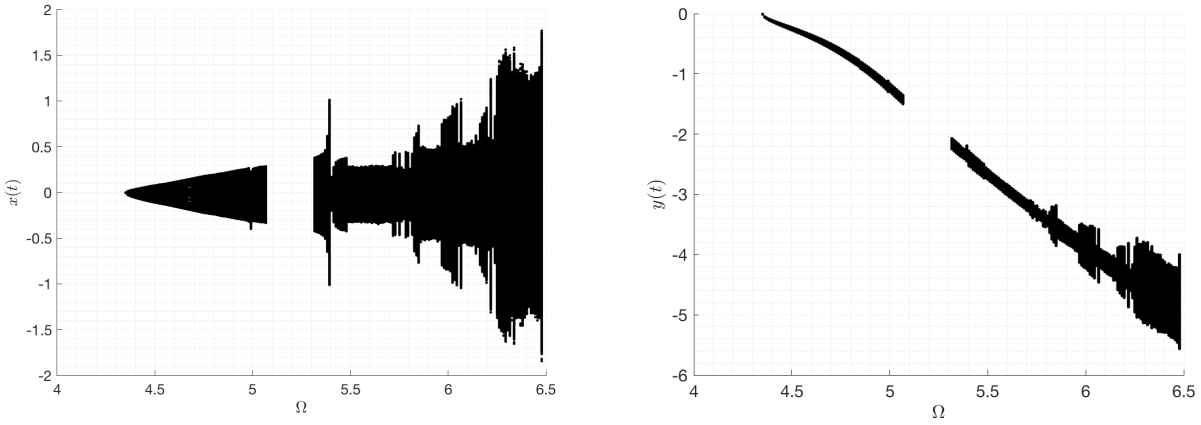


(a) $\Omega = 5.59$ Hz (b) $\Omega = 5.82$ Hz

Figure 10 – Temporal Responses of the helicopter with $\omega_x = 6\pi$ and $\omega_y = 7\pi$ - Δ Poincaré section points

rad/s and $\omega_y = 8\pi$ rad/s. The bifurcation diagram for such dynamical system under the nonlinear efforts are presented in Figure 11. The reader should note that for $5.08 \leq \Omega \leq 5.23$ Hz there is no amplitudes on the bifurcation diagram since the helicopter is stable and thus the response converges to zero (static equilib-

rium). Moreover, similar to the previous helicopter configuration analyzed (i.e., helicopter with $\omega_x = 6\pi$ rad/s and $omega_y = 7\pi$ rad/s), there exist some regions in the bifurcation diagram in which the amplitudes growth abruptly. A closer inspection of some of these regions is carried out. For this purpose, the Poincaré sections were investigated for $\Omega = 5.9,\ 6.0,\ 6.34$ and 6.48 Hz. Beyond the amplitude level that changed as function of the rotor speed, one observe that for $\Omega = 6.34$ Hz the points on the Poincaré section are distributed along likely two circumferences.



(a) (b)

Figure 11 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 8\pi$: a) $x(t)$ and b) $y(t)$

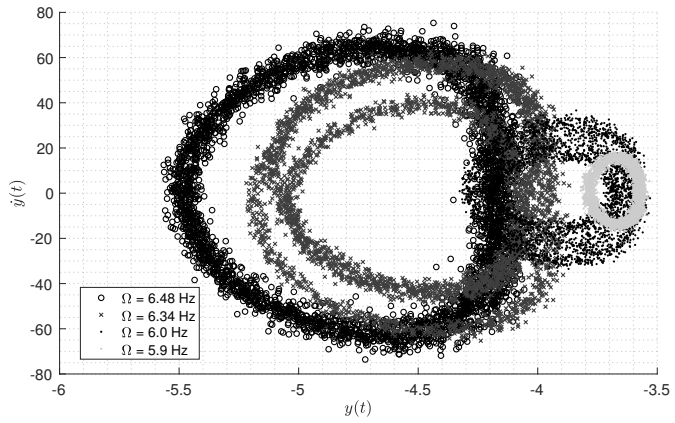


Figure 12 – Comparison of Poincaré Sections of a Helicopter with $\omega_x = 6\pi$ and $\omega_y = 8\pi$

5 Conclusions

The present paper aims at analyzing the effects of the nonlinear efforts on the dynamic response of the helicopter on the ground. The nonlinearities might change the behavior of the dynamical system, in which other than periodic motion is attained. Through the analysis, this paper investigated if non-periodic or chaotic motions are observed for the helicopter.

For this purpose, Poincaré sections and bifurcations diagrams were done for different helicopter configurations. Three different sets of fuselage's natural frequencies were evaluated, accordingly to the value of ω_y adopted.

The bifurcation diagrams highlight that for some rotor speed values, an abruptly rise in amplitude was observed for asymmetric fuselage configuration. For some exception cases where periodic motion was attained, non-periodic motion occurs for several rotor speeds values.

References

[1] GANDHI, F., AND CHOPRA, I. Elastomeric lag damper effects on flap-lag stability in forward flight. In *35th Structures, Structural Dynamics, and Materials Conference* (1994), p. 1311.

[2] GUALDI, S., MASARATI, P., MORANDINI, M., AND GHIRINGHELLI, G. L. A multibody approach to the analysis of helicopter-terrain interaction. In *Proceedings of 28th European Rotorcraft Forum, Bristol, UK* (2002).

[3] HILBORN, R. C., ET AL. *Chaos and nonlinear dynamics: an introduction for scientists and engineers*. Oxford University Press on Demand, 2000.

[4] KIM, S. J., AND YUN, C. Y. Performance comparison between piezoelectric and elastomeric lag dampers on ground resonance stability of helicopter. *Journal of Intelligent Material Systems and Structures 12*, 4 (2001), 215–222.

[5] KING, R. L. Nonlinear inplane flexbeam stiffness provides rotor system stability without lag dampers. *Journal of the American Helicopter Society 46*, 4 (2001), 283–289.

[6] KUNZ, L. Influence of elastomeric damper modeling on the dynamic response of helicopter rotors. *AIAA journal 35*, 2 (1997), 349–354.

[7] LIU, Y., WANG, J., AND TONG, Y. Kinetic analysis of elastomeric lag damper for helicopter rotors. *IOP Conference Series: Materials Science and Engineering 311*, 1 (2018), 012006.

[8] PALANIYANDI, P. On computing poincaré map by hénon method. *Chaos, Solitons & Fractals 39*, 4 (2009), 1877–1882.

[9] SANCHES, L., MICHON, G., BERLIOZ, A., AND ALAZARD, D. Parametrically excited helicopter ground resonance dynamics with high blade asymmetries. *Journal of Sound and Vibration 331*, 16 (2012), 3897–3913.

[10] SANCHES, L., MICHON, G., BERLIOZ, A., AND ALAZARD, D. Helicopter ground resonance phenomenon with blade stiffness dissimilarities: experimental and theoretical developments. *Journal of Vibration and Acoustics 135*, 5 (2013), 051028.

[11] SKJOLDAN, P., AND HANSEN, M. H. On the similarity of the coleman and lyapunov–floquet transformations for modal analysis of bladed rotor structures. *Journal of Sound and Vibration 327*, 3-5 (2009), 424–439.

[12] STOL, K., MOLL, H.-G., BIR, G., AND NAMIK, H. A comparison of multi-blade coordinate transformation and direct periodic techniques for wind turbine control design. In *47th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition* (2009), p. 479.

[13] STOLZ, B., BRÖDERMANN, T., CASTIELLO, E., ENGLBERGER, G., ERNE, D., GASSER, J., HAYOZ, E., MÜLLER, S., MUHLEBACH, L., LÖW, T., SCHEUER, D., VANDEVENTER, L., BJELONIC, M., GÜNTHER, F., KOLVENBACH, H., HOPFLINGER, M., AND HUTTER, M. An adaptive landing gear for extending the operational range of helicopters. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Oct 2018), pp. 1757–1763.

[14] VARNEY, P., AND GREEN, I. Nonlinear phenomena, bifurcations, and routes to chaos in an asymmetrically supported rotorâstator contact system. *Journal of Sound and Vibration 336* (2015), 207 – 226.

[15] WANG, J., AND CHOPRA, I. Dynamics of helicopters in ground resonance with and without blade dissimilarities. In *Dynamics Specialists Conference* (1992), p. 2108.

Comparison of pseudo-static and response spectrum seismic analyses of motor-driven pump units: is 1.5 security coefficient of pseudo-static method relevant?

Sylvie AUDEBERT[1,2], Damien ROUSSEU[3]

[1]EDF R&D, EDF Lab Paris-Saclay, ERMES Department, F-91140 Palaiseau, France
[2]IMSIA, UMR EDF-CEA-ENSTA-CNRS 9219, F-91140 Palaiseau, France
sylvie.audebert@edf.fr
EDF DT, F-69363 Lyon, France
damien.rousseu@edf.fr

Abstract

In the framework of the seismic verification of plant equipments, the determination of the seismic loads applied to motor-driven pump anchorages is optimised. A rough justification is usually first performed using the 1 degree-of-freedom pseudo-static analysis, including a 1.5 multi-mode factor. The question is asked about the opportunity to decrease the multi-mode factor value, by comparison to response spectrum analysis, here considered as the reference method. Comparative seismic analyses are performed on more and more complex dynamical systems and excitations; seismic responses of a thin square plate, motor, pump, motor-driven pump unit connected or not to suction and delivery pipes are thus successively determined, under 1D and 3D excitations. Two different motor-driven pump units are studied: flexible with vertical axis and stiff with horizontal axis. The quantities of interest are the shearing and tearing loads, deduced from seismic loads at anchorage points.

1 Introduction

The motor-driven pump units are designed so that they can resist without damage to seismic excitations: stability, integrity and functionality must thus be saved during and after the earthquake. In the case of the seismic verification of an installed motor-driven pump unit, since the soil excitation levels considered during decennial visits in nuclear industry are higher and higher, the objective is to perform more realistic simulations of resulting loads applied to anchorages, compared with those carried out for design purpose. Two ways are so followed: optimise the excitation loads and optimize the determination of the equipment response.

The purpose of the paper concerns the influence of the methods used to determine the resulting inertial seismic loads at equipment anchorages, typically the equivalent static method by comparison with the response spectrum analysis, here considered as the reference method. Two different motor-driven pump units are studied: flexible with vertical axis and stiff with horizontal axis. The quantities of interest are the shearing and tearing loads, deduced from seismic resulting loads at anchorage points.

Principles of the two seismic equivalent static and response spectrum analyses are presented, with their application on motor-driven pump units. The one-degree-of-freedom pseudo-static method is usually applied to quickly design the motor-driven pump units with no needs to elaborate a finite element model; a multi-mode factor is then associated to ensure conservatism. Using a finite element model, linear response spectrum analysis is widely used to design and justify buildings and equipments regarding seismic risk. It allows the probable maximum response of scalar quantities of interest (acceleration, displacement, stress, force, moment) due to seismic excitation, which is represented by directional floor response spectra.

Comparative seismic analyses are performed on more and more complex dynamical systems and excitations; seismic responses of a thin square plate, motor, pump, motor-driven pump unit connected or not to suction and delivery pipes are thus successively determined, under 1D and 3D excitations. Recommendations are then given about the relevancy of the 1.5 multi-mode factor value for motor-driven pump units.

2 Theoretical backgrounds

2.1 Types of seismic analyses

Seismic analyses used in the design of nuclear safety-related structures are normally conducted using linear, elastic methods. In some cases, nonlinear or inelastic seismic analyses may be conducted to obtain more realistic results. Two types of linear elastic methods are commonly used: equivalent static and dynamical methods. Among dynamical methods are response spectrum and linear time history analyses, with the seismic input motion respectively represented by floor response spectrum, and floor acceleration, velocity and displacement, functions of time.

2.2 The pseudo-static method

2.2.1 Literature review

Principle

The pseudo-static method (or Static Coefficient Method SCM, or Equivalent Static Method ESM, or Equivalent Static Lateral Force Method), is a simplified seismic analysis, that represents the effect on a system, structure, component SSC or equipment, of a seismic input motion by an equivalent static force F, determined by applying a uniform acceleration A_{max} to the mass m of the SSC [1]:

$$F = \alpha m A_{max} \quad (1)$$

The acceleration can be applied either at the SSC gravity center, as a punctual force, or on a finite SSC element model, represented by its mass matrix.

The dynamic amplification factor α (or multi-mode factor or Equivalent-Static Load Factor (ESLF) [2] is applied to take into account the multi-frequency input motion and the multi-modal SSC characteristic, to prevent from possible unfavourable dynamic combinations.

Multi-mode factor

A 1.5 multi-mode factor have been currently used for practical application of the pseudo-static method since 1976. NRC has been recommended the 1.5 value since 1981 [3]. Number of studies have been performed in order to justify [4][5] or reduce this value.

Application domain

Geometry: in IEEE, USNRC and ASCE codes, the pseudo-static method is only recommended for structures that can be simply modelled (regular horizontal and vertical geometry, equal distribution of mass and stiffness, symmetry so that torsional movement are avoided).

Dynamical behavior: the system is assumed to respond on its fundamental eigenmode [1]. The method is applicable if its vibrational behavior is not affected by modes, in every principal directions, with eigenfrequency greater than the fundamental one [1]. The method is recommended for systems whose vibrational behavior is not far from a cantilever or clamped-free beam behavior [6].

Conservatism

The conservatism of the pseudo-static method, with 1.5 multi-mode factor, is evaluated by comparison with dynamical seismic analysis methods, generally the response spectrum method.

Non conservatism can be observe in case of:

- dynamical systems with more than 2 resonancies in the amplification domain of the seismic excitation spectrum [6];
- dynamical systems with local eigenmodes not far from global modes, whose eigenfrequencies are near the peak excitation frequency; typically, not use the method if the ratio between local and global eigenfrequencies is between 0.5 and 3 [7].

2.2.2 Practical application to nuclear safety-related pump units

Comprehensive methodology for nuclear safety-related equipements

For each direction, the spectral accelerations are determined from floor response spectra, at support elevations. The same input seismic motion is applied to all the supports. The reduced damping value is generally 5%.

The spectral accelerations to be used are peak spectral acceleration if the modal SSC characteristics are unknown, or ZPA in case of seismic rigid equipment, or spectral acceleration at fundamental SSC frequency in case of seismic flexible equipment.

Equivalent static force is applied the SSC gravity center (the equivalent static method is named 1 degree-of-freedom pseudo-static method in this case). The α multi-mode factor value is generally taken as 1.5. Total response is obtained using quadratic or 100-40-40 Newmark directional combinations.

Determination of quantities of interest of nuclear safety-related pump units

The quantities of interest are the shearing and tearing loads, deduced from seismic loads at anchorage points.

The three directional components of seismic inertial loads induced at the SSC gravity center are first determined using Eq. (1). The seismic effort torsor $(F_X, F_Y, F_Z, M_X(O), M_Y(O), M_Z(O))$ at the geometrical center O of the anchorages can then be deduced. After distribution of torsor components on bolts, under the assumption of equi-distribution of loads at anchorages, total seismic shearing and tearing loads can thus be calculated, depending on the number and location of bolts.

2.3 The linear elastic response spectrum analysis

2.3.1 Principle

Based on a finite element SSC model, linear response spectrum analysis allows the probable maximum response of scalar quantities of interest (acceleration, displacement, stress, force, moment) due to seismic excitation, which is represented by directional floor response spectra. It is based on the combination of individual modal responses. To ensure an adequate representation of the equipment dynamical response, all the eigenmodes with frequencies less than the zero-period acceleration (ZPA) frequency (and no more) should be included. The residual rigid response should be systematically addressed and combined quadratically with the modal response combination. Acceptable procedures for combining modal responses include the complete quadratic combination (CQC) method and others that account for the correlation between closely spaced modes. In case of seismically stiff dynamical system, the response spectrum result is but composed of the residual rigid response. When using 3D individual earthquake components (two horizontal and one vertical directions), the directional responses should be combined at the last step either by the SRSS or the Newmark's method.

2.3.2 Application to pump shearing and tearing load determination

The resulting of the nodal reactions is calculated for each anchorage and each direction: F_X, F_Y and F_Z.

Total shearing load F_{Htotal} can be deduced using:

$$F_{Htotal} = \sqrt{F_X{}^2 + F_Y{}^2} \tag{2}$$

Total tearing load simply is:

$$F_{Ztotal} = F_Z \tag{3}$$

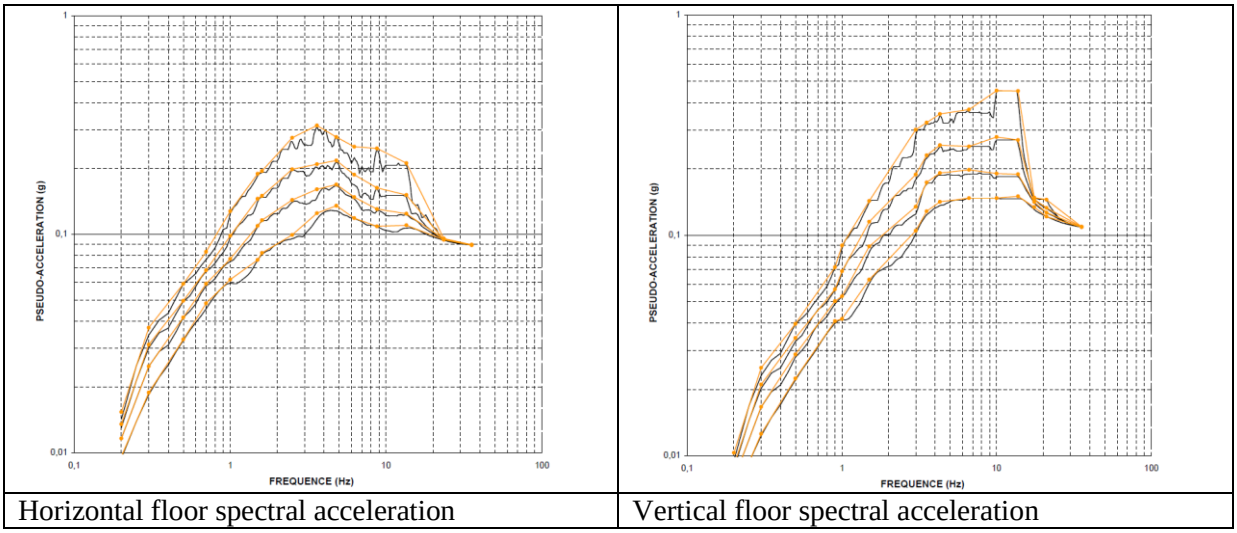
2.4 Comparison methodology

To validate the methodology of equivalent static and response spectrum comparison, more and more complex dynamical systems and excitations are considered. Comparative seismic resulting anchorage reactions are presented within 2 steps: nodal reaction torsor, then shearing and tearing loads. Only force components are compared: moments relatively to the center of anchorages issued from response spectrum simulations are not used for comparison because these moments are not provided by the equivalent static method.

3 Application to the dynamical pump component and unit models

3.1 Seismic excitation

The spectral accelerations in the three directions are issued from building responses to seismic ground motion, at the floor where the pump units are located. The zero-period acceleration is 35.5 Hz; reduced damping value is 5%.



Horizontal floor spectral acceleration	Vertical floor spectral acceleration

Figure 1: Horizontal and vertical floor spectral accelerations

3.2 The finite element pump unit models

Two different motor-driven pump units are considered:
- a seismically stiff pump unit, with horizontal axis;
- a seismically flexible pump unit, with vertical axis.

Components are simply represented, including suction and delivery pipes after their first supports, so that the first eigenmodes can be represented with satisfactory accuracy, in comparison with experimental modal characteristics. The connections between components are represented either thanks sticked surfaces or stiffness elements; their values are updated so that they fit the pump eigenmodes in the bandwidth of interest. The corresponding finite element meshes are illustrated on Figure 2 and Figure 3.

3.2.1 The horizontal stiff pump unit

Components of the horizontal-axis pump unit are the pump, bearing, coupling, motor, mounted on a metallic frame, solidary with a concrete slab: the whole system length is about 1 meter.
Boundary conditions are clamping at 6-screw pump locations and 4-screw motor locations for models without frame, or clamping at 4-screw locations under the frame. The seismic loads at anchorages are determined as the resultant force on the application 0.07 m-diameter discs for screws on motor and pump (Figure 2).

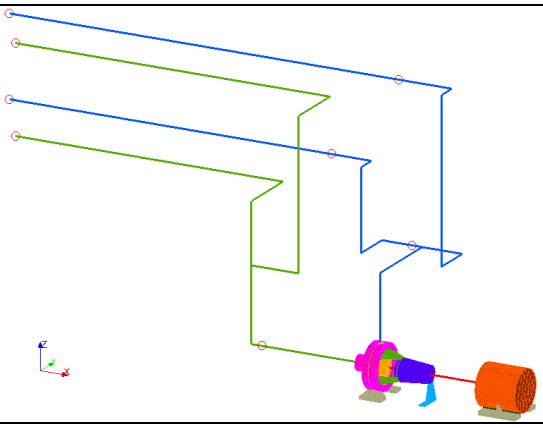


Fig.2.a: Model without frame

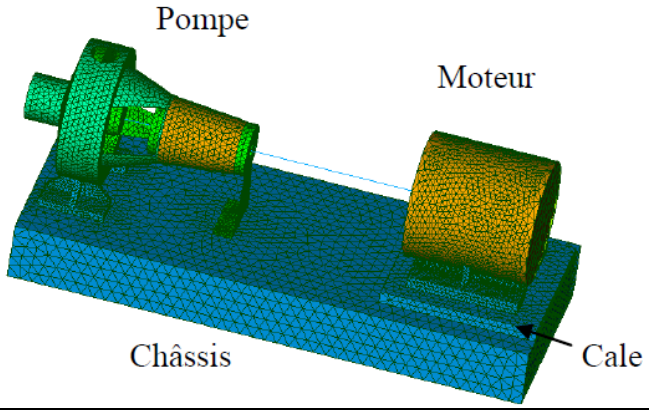


Fig.2.b: Motor-pump unit with frame

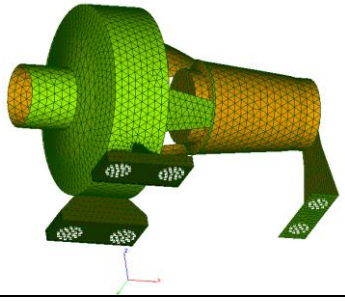


Fig. 2.c: Application discs for screws under pump

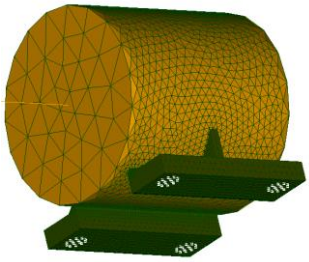


Fig. 2.d: Application discs for screws under motor

Figure 2: The horizontal pump unit: model with its pipes (without frame); model with frame (without pipes) ; pump model; motor model

3.2.2 The vertical flexible pump unit

The vertical-axis pump unit is composed of the pump, bearing support, motor at high part, mounted on three concrete studs on low part. The base of the three studs is clamped. The seismic loads at anchorages are determined as the resultant force on the higher stud faces.

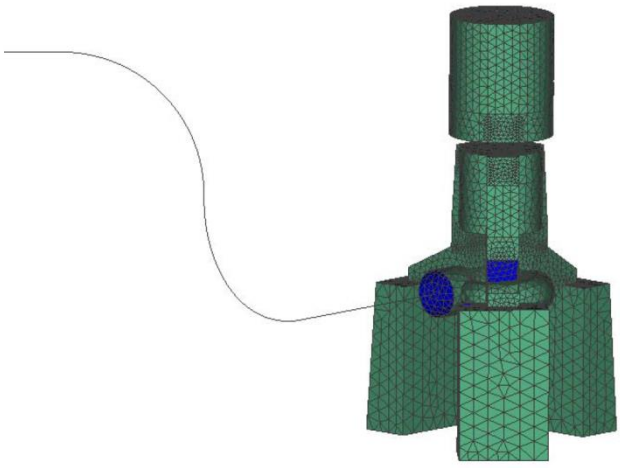


Fig.3.a: The vertical pump unit model and its pipe

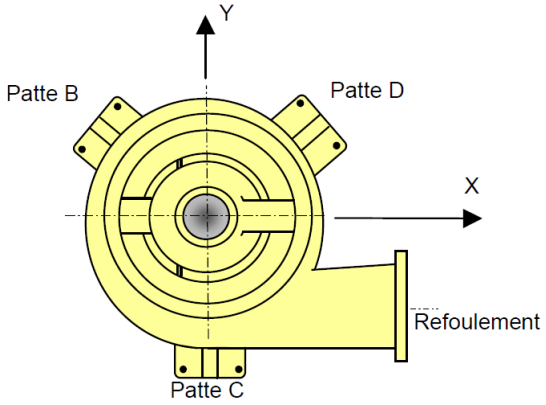


Fig.3.b : The three brackets (top view)

Figure 3: The vertical pump unit

3.2.3 Modal characteristics

The modal characteristics of the two pump units are presented in Table 1. The first eigenmodes of the horizontal pump unit only concern the pipe, conformely to its stiff behavior; the first eigenfrequency concerned with the pump unit itself is 65.9 Hz, largely beyond the zero-period acceleration. Concerning the vertical flexible pump unit, two pump eigenmodes are present in the amplification area of the floor seismic

excitation. Model parameters could be updated so that these two numerical flexion eigenmodes well represent the measured corresponding modes (0.9 MAC criterion about and 2.5% frequency gap); cumulative modal mass is less than 40% of the total mass in each direction, because the studs do not participate to the movement.

	Horizontal stiff pump unit		Vertical flexible pump unit			
Mode number	Num. eigenfrequency (Hz)	Characte-risation	Num. eigenfrequency (Hz)	Exp. eigenfrequency (Hz)	MAC	Characterisation
1	4.2	Pipe	14.5	14.3		Pipe
2	15.9	Pipe	**18.4**	**18.3**	**0.91**	**1st pump flexion**
3	16.3	Pipe	**21.5**	**21.6**	**0.86**	**2nd pump flexion**
4	19.4	Pipe	23.7	23.5		Pipe
5	23.7	Pipe	29.1	29.1		Pipe
6	24.4	Pipe				
7	25.8	Pipe				
8	27.9	Pipe				
9	31.7	Pipe				

Table 1: Modal characteristics of the pump units in [0 Hz; 35.5 Hz] frequency bandwidth

4 Comparative resulting seismic loads at anchorages

4.1 Horizontal stiff pump unit

4.1.1 Motor

In Table 2 are reported reaction force components and shearing and tearing loads of right back, right front, left back and left front anchorages, issued from response spectrum and pseudo-static analyses, respectively for 1D X-horizontal and 3D seismic excitation. In the right column is summarized the ratio between pseudo-static and response spectrum results, concerning the most loaded anchorage (from pseudo-static results). In case of the 3D-excitation, SRSS quadratic directional combination is applied.

		Response spectrum					Pseudo-static		Ratio
		Right back	Right front	Left back	Left front	Resul-tant		Resul-tant	
1D excitation along Y	F_X (N)	-85.7	98.0	88.2	**-98.5**	0	0	0	0
	F_Y (N)	**-97.8**	-91.0	-97.2	-91.7	-377.7	141	565.5	**1.4**
	F_Z (N)	**-170.0**	-167.1	**170.0**	167.1	0	0	0	0
3D excitation SRSS combination	F_H (N)	181.9	**182.1**	182.8	182.3			199.9	1.1
	F_Z (N)	233.7	225.2	233.6	**224.9**		605		2.6

Table 2: Horizontal stiff pump unit motor
1D horizontal Y-axis: reaction force components
3D excitation: shearing and tearing loads

Due to the motor stiff behavior, it is checked that, for a 1D-excitation, the resulting load issued from the 4 nodal reactions at anchorages, as the response spectrum result in the excitation direction, is equal to the product of the pseudo-acceleration applied (0.21 g) by the motor mass (183 kg), that is 377 N. As it can be theoretically proved, the application of the 1.5 multi-mode factor is not required for the evaluation of this quantity of interest.

The reaction component values are quasi-uniform across the anchorages, that illustrate a relative geometrical motor symmetry.

Orders of magnitude of reaction component values, in the directions orthogonal to the seismic excitation direction and relative to the seismic excitation direction, are the same; for a horizontal along X or Y

excitation, vertical reaction component values at anchorages are even greater than the horizontal ones. It can be checked that the resulting 4-anchorage-reaction components is zero, in the directions orthogonal to the excitation direction.

Comparison of resulting shearing and tearing loads shows that taking into account reaction force components, generated in directions orthogonal to the seismic excitation direction, which are calculated by response spectrum method and cannot be considered in pseudo-static method, induces a decrease of the margin observed in 1D excitation results (ratio 1.1 for shearing loads, less than 1.5 multi-mode factor).

		Right back		Right front		Left back		Left front	
		Spec. (ESM)	ESM/Spec	Spec. (ESM)	ESM/Spec	Spec. (ESM)	ESM/Spec	Spec. (ESM)	ESM/Spec
3D excitation SRSS combination	F_Z (N)	233.7 (383.3)	1.6	225.2 (140.9)	**0.6**	233.6 (604.6)	2.6	224.9 (362.2)	1.6

Table 3: Horizontal stiff pump unit motor
3D excitation: tearing loads
Comparative pseudo-static ESM results at each anchorage

Furthermore, it can be observed that pseudo-static method does not systematically lead to conservative values, if we consider the comparative tearing loads at each anchorage, and not only the most loaded anchorage. Table 3 shows thus that, under 3-D excitation, variability of tearing load values issued from pseudo-static method, relatively to the four anchorages, is high and not coherent with the quasi-symmetry of the motor system; at right front motor anchorage, the pseudo-static tearing load value (140.9 N) is even lower than the reference one (225.2 N).

4.1.2 Motor-pump unit with frame and pipes

Similar comparative analyses are performed on the full motor-pump unit model, including frame and suction and delivery pipes. In Table 4 are reported reaction force components, and shearing and tearing loads of the four application discs for screws, located at the inferior frame face, on right and left sides, under the motor and the pump. These quantities of interest are issued from response spectrum and pseudo-static analyses, respectively for 1D X-horizontal and 3D seismic excitation. Loads resulting from response spectrum analysis are not signed, due to combination of modal responses.

		Response spectrum				Pseudo-static	Ratio
		Motor right	Pump Right	Motor Left	Pump Left		
1D excitation along X	F_X (N)	265.0	**333.4**	265.4	301.1	596	1,8
	F_Y (N)	120.2	71.8	**121.5**	61.7	0	0
	F_Z (N)	**606.5**	339.2	581.1	307.8	0	0
3D excitation SRSS combination	F_H (N)	468.4	457.4	**473.9**	470.1	1489	3.1
	F_Z (N)	837.0	502.5	**809.5**	563.1	3086	3.8

Table 4: Horizontal stiff motor-pump unit with frame and pipes
1D horizontal X-axis: reaction force components
3D excitation: shearing and tearing loads

It can be shown that shearing and tearing loads determined via pseudo-static analysis overestimate the response spectrum results (3.1 and 3.8 factors respectively) relative to the most loaded anchorage.

4.2 Vertical flexible pump unit

In Table 5 are reported reaction force components, and shearing and tearing loads, determined as the resultant force on the higher stud faces issued for response spectrum evaluation; they are compared with the pseudo-static corresponding results, assuming an equal distribution of the loads on the 3 studs.

F components (kN)		Response spectrum			Pseudo-static	Ratio		
		Bracket A	Bracket B	Bracket C		Bracket A	Bracket B	Bracket C
1D excitation along X	F_X	2.72	1.56	1.84	3.98	1.5	2.5	2.2
	F_Y	1.82	1.93	1.44	0	0	0	0
	F_H	3.27	2.48	2.34	3.98	1.2	1.6	1.7
	F_Z	7.51	4.60	5.31	0	0	0	0
1D excitation along Z	F_X	0.38	0.45	0.47	0	0	0	0
	F_Y	24	69	28	0	0	0	0
	F_H	0.45	0.82	0.47	0	0	0	0
	F_Z	1.01	1.99	1.12	4.63	4.6	2.3	4.1
3D excitation SRSS combination	F_H	4.11	4.28	3.51	5.63	1.4	1.3	1.6
	F_Z	8.87	9.05	8.24	33.8	3.5	3.0	4.1

Table 5: Vertical flexible pump unit
1D horizontal X-axis: reaction force components
3D excitation: shearing and tearing loads

Concerning 1D seismic excitation, the 1-dof pseudo-static method overestimates the shearing loads F_H with a 1.2 to 1.7 margin, and the shearing loads F_Z with a 2.3 to 4.6 margin, compared with the response spectrum method, depending on the stud considered. Concerning the 3D seismic excitation, the margin varies from 1.3 to 2.1 for the shearing loads and from 2.7 to 4.1 for the tearing loads, depending on the stub considered and the directional combination (quadratic or Newmark).

5 Comments – Conclusion

A series of comparative seismic analyses, based on 1-dof pseudo-static and response spectrum methods, have been performed in order to determine resulting loads at anchorages, on:
- a stiff squared thin plate (not reported here) ;
- a horizontal stiff motor-pump unit and components;
- a vertical flexible motor-pump unit.

Considering the response spectrum method as the reference method, these quantitative results have permitted to determine the domain of pertinent applicability of 1-dof pseudo-static method, including 1.5 multi-mode factor, for more and more complex excitations and dynamical systems.

5.1 Multi-mode factor nature of 1 dof pseudo-static method

The multi-mode factor has been historically introduced to take into account effects due to multi-frequential excitation and multi-modal dynamical system studied (dynamical cumulative effects possibly defavourable): its 1.5 value is justified on an academic multi-dof example [4]. In case of seismically stiff system, the use of the Equivalent-Static Load Factor is to be evaluated regarding the multi-directional excitation, due to the fact that loads generated in a direction orthogonal to the excitation one cannot be reached by the 1-dof pseudo-static method: the factor is thus proposed to compensate this lack of information.

5.2 Conservatism of the 1 dof pseudo-static method

It has been checked that, for a 1D mono-supported excitation, the resulting load component in that direction, issued from the nodal reactions at anchorages of a flexible multi-modal system, as the response spectrum result in the excitation direction, is less than the product of the pseudo-acceleration applied by the system mass (equal in case of stiff system). As it can be theoretically proved, the application of the 1.5 multi-mode factor is not required for the evaluation of this – and only for this - quantity of interest; in particular, displacement, stress, strain, acceleration quantities are not concerned.
For pump units designed relatively to the most loaded anchorages, the pseudo-static analysis overestimates resulting total shearing and tearing loads. On the studied examples, margin relative to tearing loads is greater than 1.5, but margin relative to shearing loads can be lower than 1.5 (see Table 2, 1D-excitation along horizontal Y axis).
Several effects can be pointed out as an explanation of discrepancies between the two seismic analysis methods:
- assumptions on geometry and dynamical behaviour (§2.2.1) of the motor-driven pump units are not satisfied, for a justified application of the pseudo-static method; representation of this type of equipment by a 1 dof system is not reliable;
- reaction load components induced in directions orthogonal to the seismic excitation direction cannot be obtained using the pseudo-static method.

5.3 Recommendations

Based on comparisons with the reference response spectrum method, if a finite element model of the pump unit cannot be elaborated, it is recommended not to reduce the 1.5 multi-mode factor for the application of the 1 dof pseudo-static method for the early determination of the loads at anchorages.
Nevertheless, if a finite element model can be available, it is highly recommended to apply the response spectrum method instead of the pseudo-static method. More reliable results and consistency can then be obtained with the response spectrum response of piping.

5.4 Perspectives

Considering the time-history method is more representative than the response spectrum method, further comparisons will be performed between the 1 dof pseudo-static and the reference time-history analyses, in order to possibly reduce the 1.5 multi-mode factor, in case of unavailability of motor-pump unit finite element model.

References

[1] Encyclopedia of Earthquake Engineering, DOI 10.1007/978-3-642-36197-5_169-1, Springer-Verlag Berlin Heidelberg, 2013.
[2] W.H. White, A.K. Adediran, O. Gürbüz, *Multi-mode factors for distributive systems*, SMiRT19, Toronto, Canada, August 2007, K11-5.

[3] U.S. NUCLEAR REGULATORY COMMISSION STANDARD REVIEW PLAN, NUREG-0800, 3.7.2 *Seismic System Analysis*, Revision 4, September 2013, 3.7.2-9.
[4] P. Maurel, *Calcul sismique, Non conservatisme du calcul au 'pic' du spectre*, SEIS-COMB001-A-06-2010, juin 2010, http://docplayer.fr/9165236-Calcul-sismique-non-conservatisme-du-calcul-au-pic-du-spectre.html.
[5] D. Nichoff, O. Gürbüz, *Multi-mode factor for cantilevered structures with variable mass and stiffness*, Transactions, SMiRT 19, Toronto, Canada, August 2007.
[6] R. P. Kennedy, R. D. Campbell, D. A. Wesley, H. Karnil, A. Gantayat, R. Vasudevan, NUREG/CR-1706 UCRL-15216, *Subsystem Response Review, Seismic Safety Margins Research Program, Engineering Decision Analysis Company*, Inc, Lawrence Livermore Laboratory, Prepared for U.S. Nuclear Regulatory Commission Subsystem Response Review Seismic Safety Margins Research Program, July 1981.
[6] Y.M. Parulekar, G.R. Reddy, K.K. Vaze, H.S. Kushwaha, *Comparative study of detailed dynamic analysis and equivalent static analysis of structures*, Transactions of the 15th International Conference on Structural Mechanics in Reactor Technology SMiRT-15, Seoul, Korea, August 15-20, 1999, K06-1, VIII-pp. 217-224.
[7] American Society of Civil Engineers, *Seismic Analysis of Safety-Related Nuclear Structures*, ASCE STANDARD, ASCE/SEI 4-15, Published by the American Society of Civil Engineers, updated 10/16/2014.

Characterization of the damping added by a foam on a plate by an inverse vibration problem

Meryem LE DEUNF[1,2], Charles PÉZERAT[1], Frédéric ABLITZER[1] and Serge PUVILLAND[2]
[1]Laboratoire de l'Université du Mans, LAUM - UMR 6613 CNRS, Le Mans Université,
Avenue Olivier Messiaen, 72085 LE MANS CEDEX 9, France
[2] CEVAA, 2, rue Joseph Fourier Technopôle du madrillet,76800 St Etienne du Rouvray, France
m.ledeunf@cevaa.com

Abstract

The industrial solution used today to improve the acoustic performance of a structure is often ensured by the addition of damping treatments such as elastomer, PVC or bituminous. In the transportation, these materials increase the costs and mass of the vehicle and have a negative environmental impact. Unlike theses heavy materials, it is commonly known in the professional standards of automotive designers that foams in vehicle trim provide damping to the structure. It has been shown that some impregnated PU foam coatings provide significant and equivalent damping to conventional bituminous materials used in the automotive industry. This observation makes it possible to extend the function of the trim to vibration damping, in order to mutualize the both problem (acoustic and vibration) in one treatment. To understand and quantify this dissipation mechanism involved by adding a porous material to a supporting structure [1], it is proposed to treat the problem from the angle of an experimental quantification obtained by an inverse problem. The proposed approach is based on the use of the Force Analysis Technique (FAT) method [2] [3], where the first objective is to locate and quantify the forces applied to a vibrating structure. In this case, the FAT is designed to define the damping provided by the foam to the vehicle trim [4] [5].
Keywords : foam, damping, FAT method

1 Introduction

The transport industry is constrained by vehicle weight reduction and the resulting impact on performance. One of the solutions available is the additional damping treatments such as elastomer, PVC or bituminous. But, they increase the weight of the vehicle and reduce the benefits of lightening solutions (use of thermoplastic shell instead of steel panel, reduction in steel thicknesses used for the structure). In previous studies, some impregnated PU foam coatings provide significant damping to the structure. The viscoelastic damping of these materials in tension-compression is low and does not explain these performances. The origin of this damping can be generated by the dry friction at the interface between the foam and the structure. The objective of the study is also to dissociate the part of the damping generated by the viscoelasticity of the foam and generated by the relative displacement of the material with respect to the supporting structure.

To measure the dry friction damping at the interface, the Force Analysis Technique (FAT) method [6] [4] is used to solve the equation of plate motion. With this local method, it is possible to find the damping of the foam-plate structure in the medium and high frequencies. The Corrected Force Analysis Technique (CFAT) method [7] [5] is used in high frequencies to improve measurements. A combination of these two methods increases the frequency range studied. The system studied is a viscoelastic material (PU foam) placed on a steel plate. Using a LASER vibrometer, the foam-plate contact area is swept to measure the vibration field. The results obtained for the loss factor and the stiffness term are given for with and without foam on the plate.

2 Theory

2.1 Equation of motion of a plate

The equation of motion for a thin isotropic plate in harmonic regime for bending stress is

$$D\left(\frac{\partial^4 w}{\partial x^4}+\frac{\partial^4 w}{\partial y^4}+2\frac{\partial^4 w}{\partial x^2\partial y^2}\right)-\rho h\omega^2 w=F(x,y,\omega), \tag{1}$$

where D is the bending stiffness, ρ the density, h the thickness, ω the angular frequency, $w(x,y)$ the transverse displacement field and $F(x,y,\omega)$ the distribution of the external forces exerted on the plate. Flexural rigidity

$$D=\frac{E(1+j\eta)h^3}{12(1-\nu^2)}, \tag{2}$$

introduces the Poisson coefficient *nu* and the Young's modulus complex $E(1+j\eta)$ where η is the loss factor, which is the damping factor of the system. The equation of motion (1) is called local, because it is valid at any point in the structure and independent of boundary conditions. Considering an area where no force is applied ($F(x,y,\omega)=0$), the equation (1) becomes

$$\frac{D}{\rho h\omega^2}\left(\frac{\partial^4 w}{\partial x^4}+\frac{\partial^4 w}{\partial y^4}+2\frac{\partial^4 w}{\partial x^2\partial y^2}\right)=w. \tag{3}$$

Knowing the displacement field $w(x,y)$ of the structure and its spatial derivatives $\frac{\partial^4 w}{\partial x^4}$, $\frac{\partial^4 w}{\partial y^4}$ and $\frac{\partial y^4 w}{\partial x^2\partial y^2}$, it is possible to identify the term $\frac{D}{\rho h\omega^2}$, which may vary with the frequency. The real and imaginary parts of this term give the stiffness and damping of the structure. The characteristics of the structure can be determined by a measurement of the displacement field and the estimation of spatial derivatives, however the measurements made are noisy. The FAT and CFAT methods are used to reduce measurement noise and retrieve essential information.

2.2 CFAT and FAT methods

The CFAT method regularizes the inverse resolution using the natural filter of discretization by finite differences in the equation of motion. The CFAT method initially allows to find the force distribution on a known structure using the measured displacement field, in the case it is not the force distribution, but the characteristics of the structure which are studied. This method consists in introducing correction coefficients into the equation of motion, in order to benefit from the filtering effect of the finite difference scheme while correcting the bias it introduces into the resolution. The equation of discretized motion corrected

$$\frac{D}{\rho h\omega^2}\left(\tilde{\mu}^4\delta_{ij}^{4x}+2\tilde{\nu}^4\delta_{ij}^{2x2y}+\tilde{\mu}^4\delta_{ij}^{4y}\right)=w_{ij} \tag{4}$$

presents schemas with finite differences

$$\delta_{ij}^{4x}=\frac{1}{\Delta x^4}(w_{i+2,j}-4w_{i+1,j}+6w_{i,j}-4w_{i-1,j}+w_{i-2,j}),$$
$$\delta_{ij}^{4y}=\frac{1}{\Delta y^4}(w_{i+2,j}-4w_{i+1,j}+6w_{i,j}-4w_{i-1,j}+w_{i-2,j}),$$
$$\delta_{ij}^{2x2y}=\frac{1}{\Delta x^2\Delta y^2}(w_{i+1,j+1}-2w_{i+1,j}+w_{i+1,j}-2w_{i,j+1}+4w_{i,j}-2w_{i,j-1}+w_{i-1,j}+2w_{i-1,j}+w_{i-1,j-1}),$$

and corrective coefficients

$$\tilde{\mu}^4=\frac{\Delta^4 k_f^4}{4[1-\cos(k_f\Delta)]^2},$$
$$\tilde{\nu}^4=\frac{\Delta^4 k_f^4}{8[1-\cos\left(\frac{k_f\Delta}{\sqrt{2}}\right)]^2}-\tilde{\mu}^4.$$

The corrective terms require knowing the number of natural bending waves of the plate

$$k_f^4 = \frac{\rho h}{D}\omega^2. \tag{5}$$

This wave number k_f is unknown because it depends on the characteristics of the plate. Characteristics are identified iteratively, with a first iteration without correction (i.e. by imposing $\tilde{\mu}^4 = \tilde{\nu}^4 = 1$), to provide a first initial value of k_f. At each iteration the value of $\frac{D}{\rho h}$ is identified, which provides a new estimate of k_f. According to the system, the number of iterations is not the same, in our case 10 iterations are necessary to ensure the convergence of the inverse problem. The advantage of the CFAT method is that the regularization is done automatically during the inverse resolution. It is not necessary to calibrate a regulation parameter, unlike the RIFF method. However, it has a certain validity range given by

$$\begin{array}{l|l} f^{min}(\Delta) = \frac{\pi}{8\Delta^2}\sqrt{\frac{Eh^2}{12\rho(1-\nu^2)}} & f^{min}(0.0133) = 3.985 \cdot 10^3 \text{ Hz}, \\ f^{max}(\Delta) = \frac{\pi}{2\Delta^2}\sqrt{\frac{Eh^2}{12\rho(1-\nu^2)}} & f^{max}(0.0133) = 1.594 \cdot 10^4 \text{ Hz}, \end{array}$$

where Δ represents the spatial discretization between two points. To identify the characteristics of the plate over the entire frequency range, it is necessary to use in combination with the RIFF method for low frequencies.

The FAT method is used to regularize errors due to measurement noise using a low-pass filter in wave number. First of all, it is necessary to window the signal to soften the discontinuities at the limits and avoid the negative effects of the filter (Gibbs phenomenon). This is done using a Tukey window

$$\Psi_{i,j}^{2D} = \Psi^{1D}(x_{i,j} - x_{1,1} - 2\Delta_x) \cdot \Psi^{1D}(y_{i,j} - y_{1,1} - 2\Delta_y), \tag{6}$$

where

$$\Psi^{1D}(x) = \begin{cases} 0.5\left(1 - \cos\left(\frac{\pi x}{\alpha}\right)\right) & \text{si } 0 \leq x < \alpha, \\ 1 & \text{si } \alpha \leq x < L - \alpha, \\ 0.5\left(1 - \cos\left(\frac{\pi(x-L+2\alpha)}{\alpha}\right)\right) & \text{si } L - \alpha < x \leq L, \\ 0 & \text{sinon .} \end{cases} \tag{7}$$

with $L = L_x - 4\Delta_x$ and

$$\alpha = \begin{cases} \lambda_c & \text{si } L \geq 2\lambda_c, \\ \frac{L}{2} & \textit{sinon.} \end{cases}$$

The filter, which removes high wave numbers, is weighted by a Hanning window to keep the local aspect of the method. This filter eliminates the amplification of errors associated with the inverse problem. The spatial response of the filter

$$h(x,y) = \begin{cases} \frac{k_c^2}{4\pi^2 xy N_f}\left(1 + \cos\left(\frac{k_c x}{2}\right)\right)\left(1 + \cos\left(\frac{k_c y}{2}\right)\right)\sin(k_c x)\sin(k_c y) & \text{si } x \text{ et } y\varepsilon\left[\frac{-2\pi}{k_c}, \frac{2\pi}{k_c}\right], \\ 0 & \text{sinon ,} \end{cases} \tag{8}$$

introduces a normalization parameter N_f, chosen so that $\iint h(x,y)dxdy = 1$ and a cut-off wave number k_c. The latter is generally chosen in proportion to the natural waves number

$$k_c = a \cdot k_f, \tag{9}$$

where the regulation parameter a is generally set at 4 for very good measurements and 1 for very noisy measurements. In our case, the combination of the FAT method with the CFAT method avoids using a regulation parameter. Here, the cut-off wave number k_c is chosen equal to the bending wave number of the plate k_f determined with the CFAT method.

3 Experimental validation

3.1 Set-up

To characterize the dry friction damping between the foam and the plate, the temperature (19° C) chosen is lower than the glass transition (55° C) of the foam to have the least viscoelastic damping effect. The measurement is made on a structure composed of a steel plate (700x700x2 mm) suspended at the four ends by elastics, to approach conditions at the free limits and a foam plate (40x40 cm) simply placed on the plate. A scanning laser vibrometer is used to measure the displacement field in the study area (56xx56 cm). The vibrometer is located below the steel plate, at a distance of 132 cm . The excitation, of impulse type, is applied using an automatic impactor with an amplitude of 5 N and repeated 3 times for each measurement point. Figure 1 shows a schema of the experimental set-up.

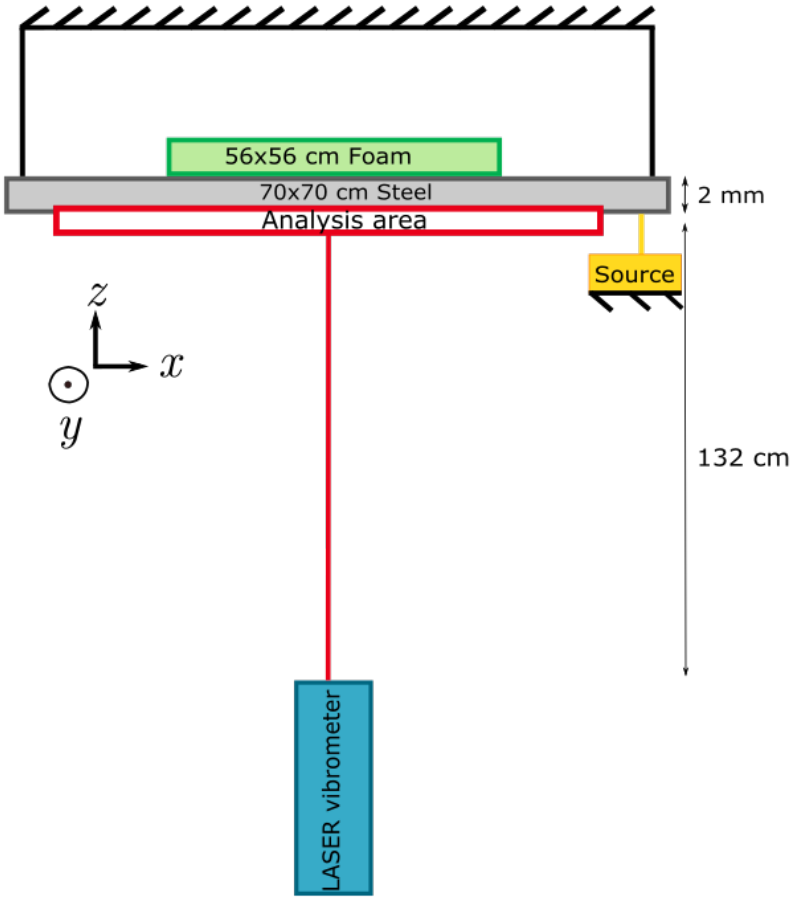


Figure 1: Experimental set-up

The measurement mesh has 33x33 = 1089 points, with a spatial pitch of about 0.13 cm. The studied frequency range is up to 10 kHz and the frequency step is 0.78 Hz. A wide frequency range is observed to measure the effects of foam at low and high frequencies.

The foams used are of the thermoplastic elastomer type, they are normally used for the sound insulation of vehicles. In this case the study extends to the use of these foams for the vibration damping of a steel structure. Different foam thicknesses are tested, but only one is compared with the blade plate. Friction at the interface between the foam and the plate is a dry friction, which may result in a non-linear response of the structure. In this study, it is not the damping at the interface that is identified, but the global damping of the foam-plate structure.

3.2 Results

Two measurements are made to obtain the displacement of the plate without foam, then with 11 mm thick foams. The characteristics obtained for the structure are the stiffness $\Re\left(\frac{D}{\rho h}\right)$ (Figure 2) and shock absorption

$\eta = \frac{\Im\left(\frac{D}{\rho h}\right)}{\Re\left(\frac{D}{\rho h}\right)}$ (Figure 3) for a frequency range from 400 Hz to 10 kHz.

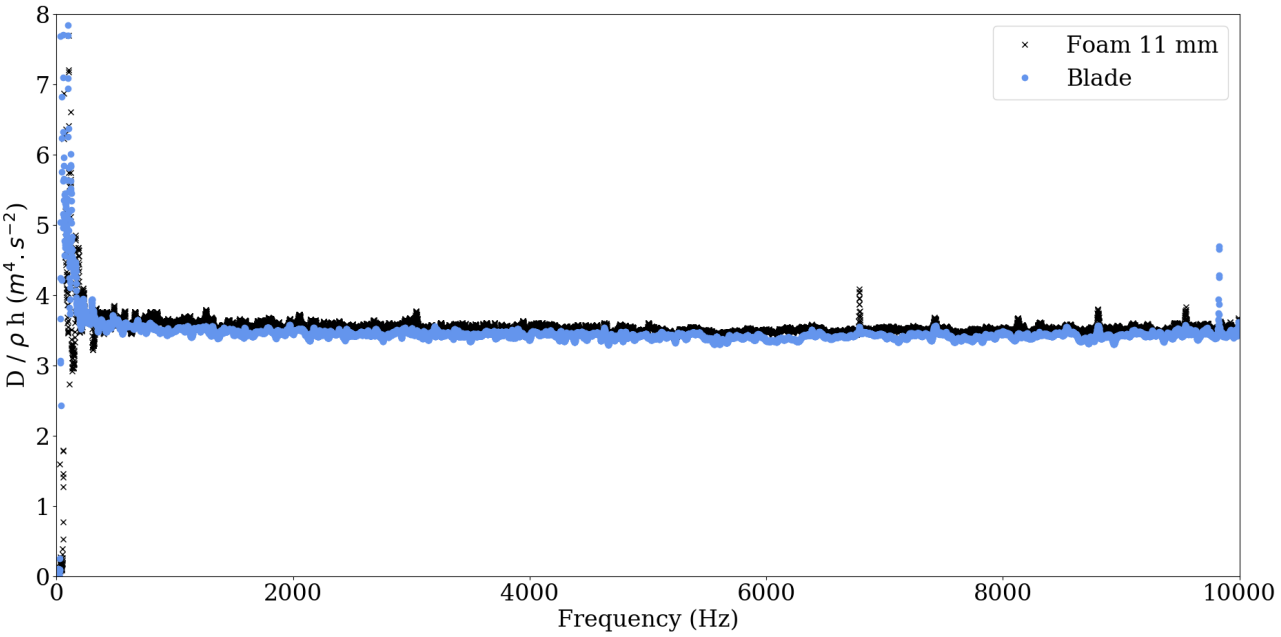


Figure 2: Stiffness $\frac{D}{\rho h}$ of the structure with only the plate and different foam thicknesses

The stiffness for the two configurations varies. The average value of the stiffness is $3,528\ \mathrm{m}^4 \cdot \mathrm{s}^{-2}$ for the configuration with the 11 mm foam and $3,454\ \mathrm{m}^4 \cdot \mathrm{s}^{-2}$ for the bare plate. The difference in stiffness between the configuration with the foam (98 g) of 11 mm and the bare plate is 7%.

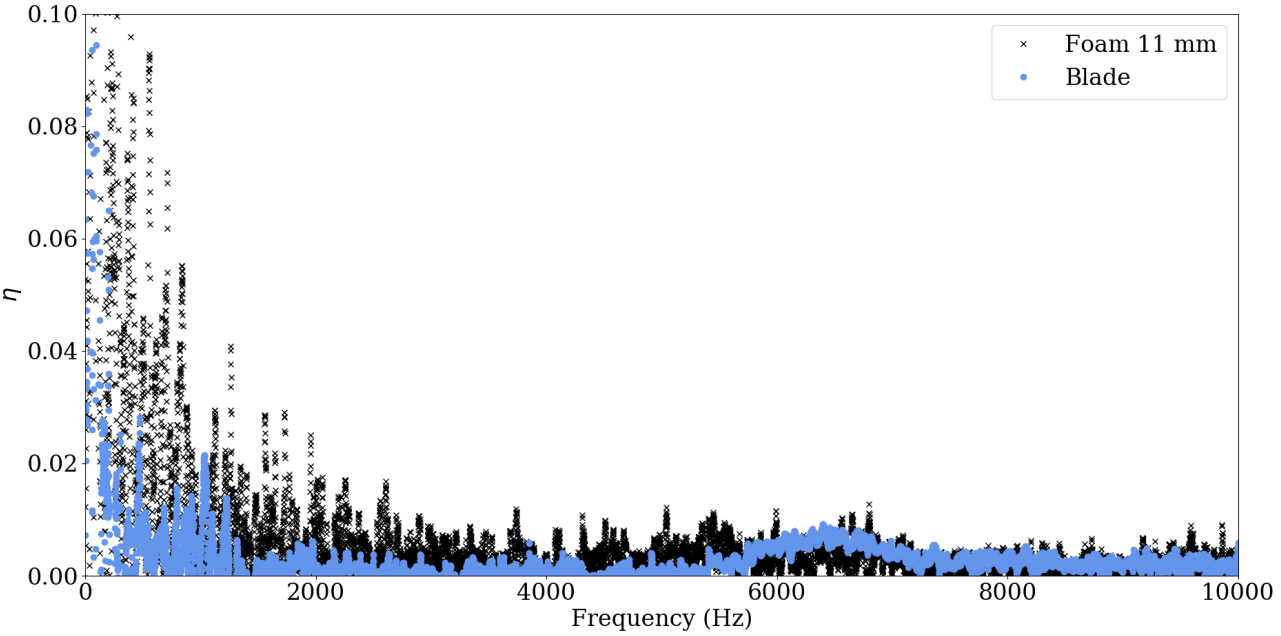


Figure 3: Loss factor η of the structure with only the plate and different foam thicknesses

The damping increases with the foam, mostly at low frequencies. Around 7000 Hz, there is an increase in the damping for the bare plate. This increase may be related to the vibration mode of the hooks used for suspension. Further study would reveal the cause of this increase in absorption.

4 Conclusion

The loss factor obtained after laser vibrometer measurement validates the method (FAT-CFAT) used to determine the damping of foam placed on a steel plate. In addition, the combination of the FAT and CFAT method allows the frequency validity range to be extended compared to only used the CFAT method. So, it isn't necessary to use an adjustment parameter for the FAT-CFAT method. The results obtained make it possible to validate both the experimental procedure and the FAT-CFAT method. In order to determine the origin of the resultant damping, a numerical study is in progress to characterize the damping of the vibration by the viscoelastic effect of the foam. To decide on the origin of the damping, whether viscoelastic or by dry friction, the test is compared to the numerical calculation. In order to increase the measurement dynamics, several modifications of the test method are planned. To limit the damping effect of the free suspension, the use of a system with supported boundary conditions is considered. In order to increase the damping of the system, a decrease in the rigidity of the support plate will be studied.

References

[1] Z. Abbadi, N. Merlette, N. Roy, *Response Computation of Structures with Viscoelastic Damping Materials using a Modal Approach - Description of the Method and Application on a Car Door Model Treated with High Damping Foam*, 5th Symposium of Automobile and Railroad Comfort, Le Mans, France, Nov. 19-20, 2008.

[2] C. Faure, *Identification de source vibratoire sur structure mal connue*, Thèse Sci. - Université du Maine, 2016.

[3] T. Wassereau, *Caractérisation de matériaux composites par problème inverse vibratoire* , Thèse Sci - Université du Maine, 2016.

[4] F. Ablitzer, C. Pézerat, J.-M. Genevaux, J. Begue, *Identification of stiffness and damping properties of plates by using the local equation of motion*, Journal of Sound and Vibration, Volume 333, Issue 9, 2454 - 2468 p, 2014.

[5] Q. Leclère, F. Ablitzer, C. Pézerat, *Practical implementation of the Corrected Force Analysis Technique to identify the structural parameter and load distributions*, Journal of Sound and Vibration, Volume 351, Issue 1, 106-118 p, 2015.

[6] C. Pézrat, J.-L. Guyader, *Force Analysis and Technique: Reconstruction and of Force and Distribution on and Plates*, ACUSTICA acta acustica, 322-332, 2000.

[7] Q. Leclère, C. Pézrat, *Vibration source identification using corrected finite difference schemes*, Journal of Sound and Vibration, 1366-1377, 2012.

Coupled bending torsional vibrations of non-ideal energy source rotors going through critical speeds

Emna SGHAIER[1], Jean-Luc DION[1], Nicolas PEYRET[1], Adeline BOURDON[2],Didier REMOND[2]

[1]Laboratoire QUARTZ EA 7393 - SUPMECA Paris, 3 rue Fernand Hainaut, 93400 Saint Ouen, France
emna.sghaier@supmeca.fr, jean-luc.dion@supmeca.fr, nicolas.peyret@supmeca.fr

[2]University of Lyon, LaMCoS, INSA-Lyon, CNRS UMR5259, F-69621, France
emna.sghaier@insa-lyon.fr, adeline.bourdon@insa-lyon.fr, didier.remond@insa-lyon.fr

Abstract

With the increasing number of rotors working at very high angular velocity, it becomes crucial to understand the dynamic behavior of rotating machineries when going through critical speeds. Models assuming constant velocity speed are not valid in this case of study since crossing the critical speeds implies necessarily non-stationary working conditions. The present work offers a new finite element model for rotors working at non-stationary regime. The rotational speed is introduced to the unknowns of the dynamic problem and six degrees of freedom are considered on each node. A main focus is given to the study of the coupling between the torsional and flexural degrees of freedom. This coupling is introduced by the intrinsic gyroscopic effect as well as the mass unbalance terms. It results in torsional vibrations containing frequency components of twice the excitation frequency of the mass unbalance as well as frequency components reflecting a modulation with the first bending natural frequency. We show that when crossing the critical speed, an additional frequency component of four times the lateral excitaion frequency appears. The coupling is observed through the analytical equations of motion and confirmed by the numerical simulation.

1 Introduction:

The majority of studies carried on rotordynamics focus either in the lateral behavior of rotating machineries or in the torsional behavior in seperate ways [1, 2, 3]. Fewer studies has been performed for the exploration of coupled bending-torsional behavior. However, those studies for coupled behavior are usually made under some simplifying assupmtions. The mutual influence between transverse and torsional behavior may occur due to several reasons. The gyroscopic effect is the intrinsic source of coupling for rotating machineries as well as the mass unbalance. The coupling between lateral and torsional vibrations in rotors may also arise due to rotor cracks. Another important source of flexural-lateral coupling in rotors is the presence of a gears. One of the early studies on this topic was made by Bernasconi. If the coupling means a mutual influence between the lateral and torsional behavior, Bernasconi, in his paper [4], explored only the torsional vibretions induced by transverse ones. Rao et al [5] explored the effect of the presence of gears on the bending vibrations on the case of permanent regime. XYShen et al. [6] studied the coupled behavior of flexible rotor with six degrees of freedom on each node but also limited the study to the stationnary regime for a given speed of rotation Ω. Al-Bedour [7] Studied the particular case of Jeffcot rotor with no gyroscopic effects and explored the coupling induced by the mass unbalance. R.Sukkar [8] studied an unbalanced Jeffcott rotor but this time in the presence of axial load at stationnary and non-stationnary operating conditions. The aim of this paper is to present a innovative fully coupled model for the study of non-ideal energy source rotors at non-stationnary regime. The speed of rotation of the rotor is considered as an unknown of the dynamic problem and is included in such way that it combines at the same time the nominal rigid body rotation Ω and the torsional deformation θ_t as following:

$$\theta_z = \Omega + \theta_t \tag{1}$$

This way of introducing the degree of freedom θ_z gives more freedom for the simulation of the rotor behavior under non-stationnary regime and offers a more realistic way for observing the phenomena related to the non-ideal energy source rotors, mainly, the sommerfeld effect [9]. The latter is a phenomena that reflects energy exchanges between the rotational direction and the lateral one and can be observed only if the speed of rotation is included to the unknowns of the dynamic problem.
In this paper, the lateral-torsional coupling is observed through the analitycal equations as well as the numerical results.

2 New model for rotordynamics

In this section, we consider a basic rotor made of a shaft, a disk and linear bearings. The excitation of the rotor is due to the mass unbalance.

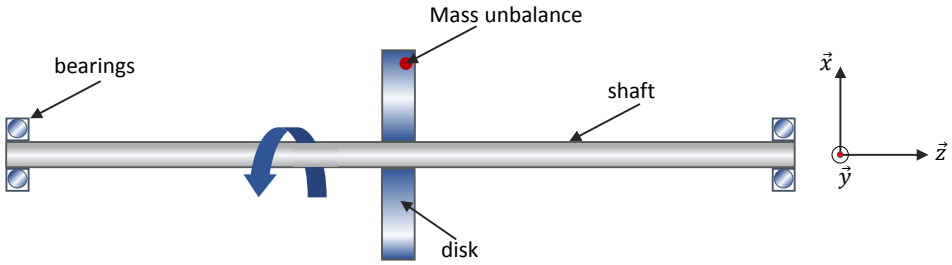


Figure 1: Illustration of the rotor

The finite element method is used to write the equation of motion over a shaft element under the considerd assumptions. Six degrees of freedom are considered on each node. The displacement vector includes the three translations and the three rotation leading to an elementary displacement vector of the following form:

$$\{\delta^e\} = (u_1; v_1; w_1; \theta_{x_1}; \theta_{y_1}; \theta_{z_1}; u_1; v_2; w_2; \theta_{x_2}; \theta_{y_2}; \theta_{z_2})_{\{1,12\}} \tag{2}$$

The energitic approach is adopted to calculate the kinetic and strain energy of the different rotor components. The overall equations of motion are obtained using the Lagrange dynamics and are put in the following form:

$$\begin{aligned}\left(\sum_e [M_s^e] + [M_D] + [M_u]\right)\left\{\ddot{\delta}\right\} + \left(\sum_e [C_s^e] + [G_D]\right)\left\{\dot{\delta}\right\} + \left(\sum_e [K_s^e] + [K_D]\right)\{\delta\} = \\ \{F_{ext}\} + \sum_e [S_s^e] + \sum_e \{F_{se}^{coup}\} + \{F_D^{coup}\} + \{F_{Nlu}\}\end{aligned} \tag{3}$$

$$[S_s^e] = -\ddot{\theta}_{z_2}[A_1]\{\delta^e\} - \dot{\theta}_{z_2}[Gyr]\left\{\dot{\delta}^e\right\} + \left(\ddot{\theta}_{z_2} - \ddot{\theta}_{z_1}\right)[A_2]\{\delta^e\} + \left(\dot{\theta}_{z_2} - \dot{\theta}_{z_1}\right)[Gyr^*]\left\{\dot{\delta}^e\right\} \tag{4}$$

$$\begin{aligned}\left\{F_{s_e}^{coup}\right\} = -\left\{N_3^g(l)\right\}\left(\left\{\ddot{\delta}^e\right\}^t\left[A_1^g\right]\{\delta^e\}\right) - \left\{N_3^g(l)\right\}\left(\left\{\dot{\delta}^e\right\}^t\left[A_1^g\right]\left\{\dot{\delta}^e\right\}\right) \\ + \left\{\frac{\partial N_3^g}{dz}\right\}\left(\left\{\ddot{\delta}^e\right\}^t\left[A_2^g\right]\{\delta^e\}\right) + \left\{\frac{\partial N_3^g}{dz}\right\}\left(\left\{\dot{\delta}^e\right\}^t\left[A_2^g\right]\left\{\dot{\delta}^e\right\}\right)\end{aligned} \tag{5}$$

$$[Gyr] = \left[A_1^g\right] - \left[A_1^g\right]^t \quad ; \quad [Gyr^*] = \left[A_2^g\right] - \left[A_2^g\right]^t \tag{6}$$

$$\left[A_1^g\right] = -2\frac{\rho I_p}{l}\int_0^l \left\{\frac{\partial N_2^g}{\partial z}\right\}\left\{\frac{\partial N_1^g}{\partial z}\right\}^t \mathrm{d}z \quad ; \quad \left[A_2^g\right] = -2\frac{\rho I_p}{l}\int_0^l\int_0^z \left\{\frac{\partial N_2^g}{\partial z}\right\}\left\{\frac{\partial N_1^g}{\partial z}\right\}^t \mathrm{d}z \tag{7}$$

$$\begin{cases} u(z,t) &= \{N_1(z)\}^t \{\delta_u^e(t)\} \quad ; \quad w(z,t) &= \{N_3(z)\}^t \{\delta_w^e(t)\} \\ v(z,t) &= \{N_2(z)\}^t \{\delta_v^e(t)\} \quad ; \quad \theta_z(z,t) &= \{N_3(z)\}^t \{\delta_{\theta_z}^e(t)\} \end{cases} \tag{8}$$

Where $[M_s^e]$, $[C_s^e]$, $[K_s^e]$ and $[Gyr]$ are the classical mass, stiffness, damping and Gyroscopic effect matrix. Matrix $[A_1]$ is the 'stiffness matrix' resulting from the assumption of the non-stationnary regime.$\{F_{ext}\}$ is the vector of external efforts and $\{p^e\}$ is the effort applied by the neighbouring elements on the considered one. Matrices $[A_2]$ and $[Gyr^*]$ are related to the gyroscopic effect under non-stationnary regime and would vanish if the torsional deformation is neglected in the study. Finally, the vector $\left\{F_{se}^{coup}\right\}$ derives also from the gyroscopic effect. More details about the vectors and matrices used in the previous equations is presented in [10] where it was explained that the gyroscpic effect terms are taking this form due to the considered assumptions of non-ideal energy source, non-stationnary regime as well as the introduction of torsion in the study. The analytical formulation shows coupling between the lateral and torsional displacements induced by both the mass unbalance and gyroscopic effect. This coupling will be later explored through numerical simulations.

2.1 Numerical results

We consider the following rotor made of a shaft, a disk, linear bearings and elastic coupling (see fig.2). The rotor is excited with a mass unbalance situated on the disk.

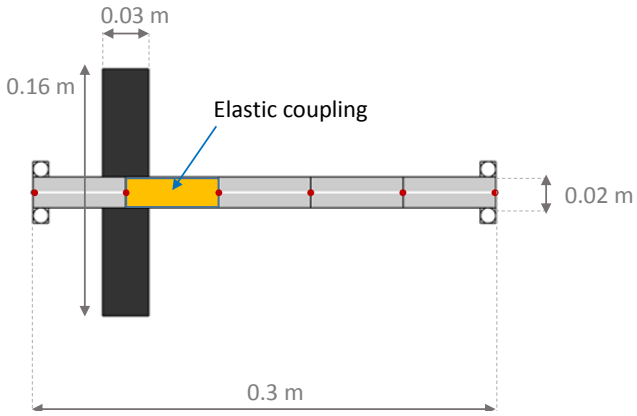


Figure 2: studied rotor

The campbell diagram of the rotor shows two crtical speeds in the speed range between $0 rpm$ and $45000 rpm$. The first critical speed $\Omega_{cr_1} = 2038\, rpm$ and the second one is at $\Omega_{cr_2} = 20388\, rpm$ (see fig.3).

A linear torque is induced to the rotor. As we can see on the results for the evolution of the instantanious angular speed as a function of time (see fig.4), the sommerfeld effect takes place when the rotor crosses the second critical speed. In this case, the energy induced to the rotor is no longer used to increase its speed but is communicated to the transverse vibrations[11, 12, 13]. The sommerfeld phenomena is usually indesirable because it causes high lateral vibrations magnitude if not enough damping is present in the structure. It is a manifestation of the ineraction between the lateral and rotational direction resulting from the lateral-torsional coupling.The sommerfeld effect couldn't be observed if the assumption of non-ideal energy source wouldn't have been made for the modeling. If not taken into consideration, simulations may lead to an under-estimation of the lateral vibration.

To see more clearly the coupling between the flexural and torsional displacements, we perfom time-frequency anylsis to the torsional deformation on the soft element as shown in figure 5. The torsional vibration signal contains frequencies corresponding to the frequency of excitation relative to the mass unbalance f_u as well as the bisynchronous frequency $2f_u$. Also frequencies of $f_u - f_{b_1}$ and $f_u + f_{b_1}$ are contained in the torsional displacement signal, such as f_{b_1} is the transverse natural frequency at Ω_{cr_2} wich is given by $f_{b_1} = 154\, Hz$. Finally,

only when crossing the second critical speed, a frequency of $4\,f_u$ is observed in the time-frequency analysis. This highlights the possibility of the transverse vibrations to induce torsional ones. The behavior of the torsional vibrations induced by the lateral ones is slightly different when crossing a critical speed in the presence of the sommerfeld effect as an extra frequency component of $4\,f_u$ contibutes to the composition of the torsional signal.

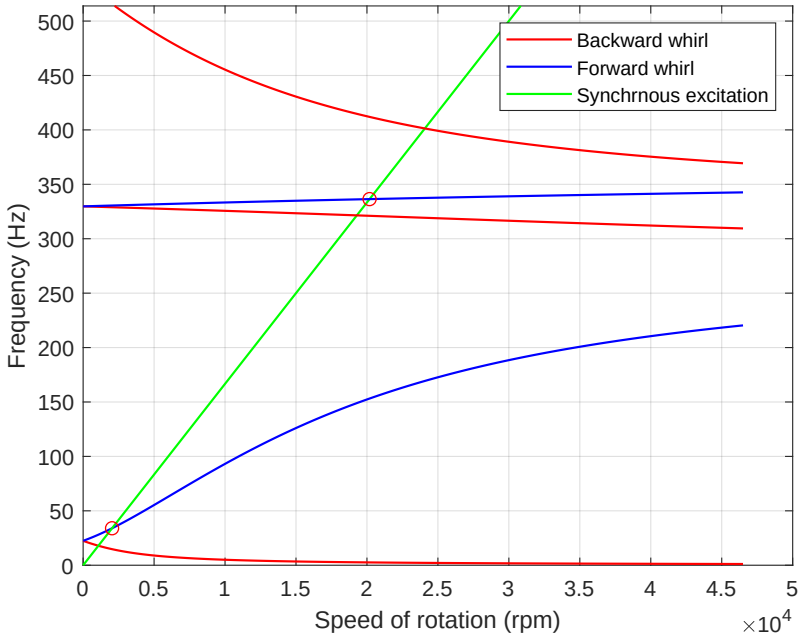


Figure 3: Campbell diagram

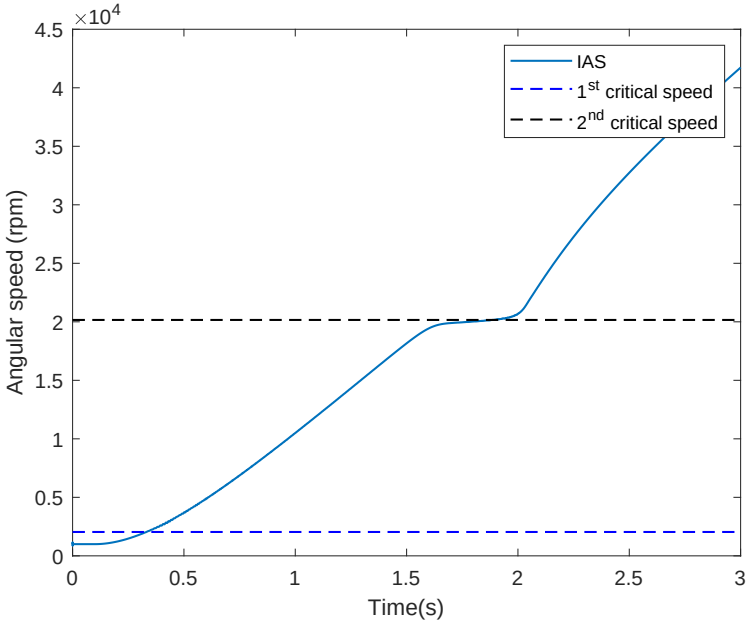


Figure 4: Angular velocity as a function of time

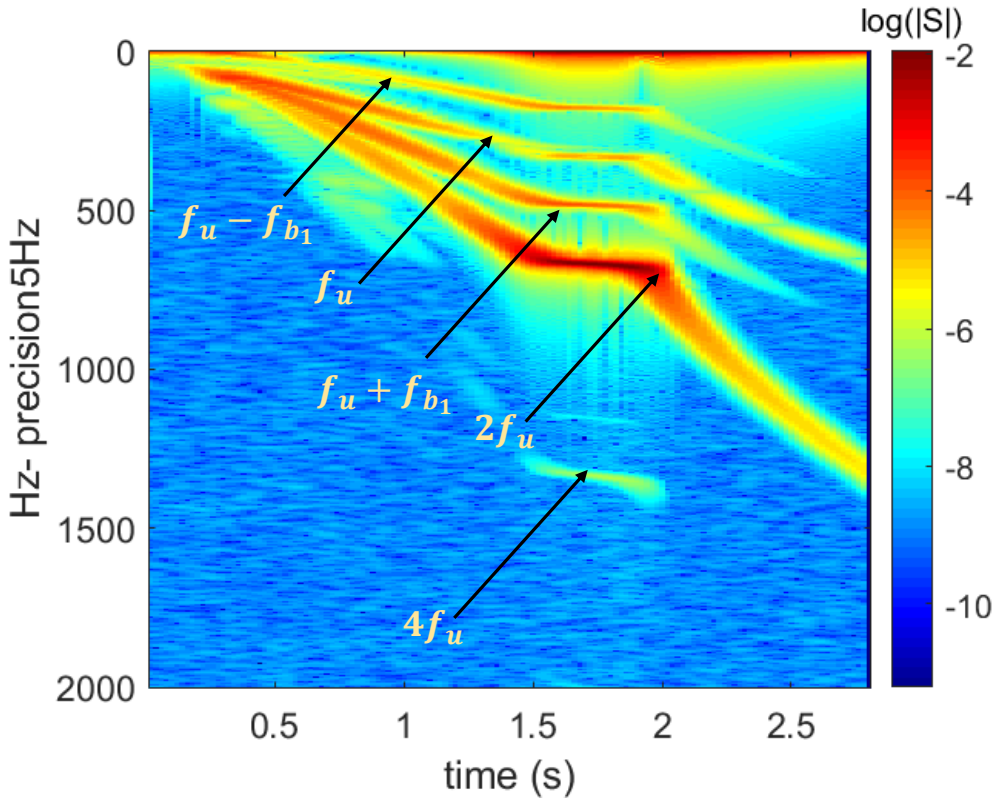


Figure 5: Time-frequency analysis of torsional vibrations on the second element

3 Conclusion

A new finite element model with six degrees of freedom per node is developed for the dynamic response analysis of a rotor systems operating at non-stationnary regime. The proposed new dynamic model is build under the assumption of non-ideal energy source and is a fully coupled lateral-torsional model. The coupling is introduced by both the intrinsic gyroscopic effect as well as the mass unbalance. The time-frequency analysis show that the torsional behaviour is slightly different when crossing a critical speed in the presence of the sommerfeld effect.

The model is build under the less constraining assumptios which extends its use to large case studies. It can be easily extended to the study of multiple rotors connected between each others by elastic coupling.

References

[1] M. Lalanne, G. Ferraris, Rotordynamics prediction in engineering, Vol. 2, Wiley, 1998.

[2] M. I. Friswell, J. E. Penny, A. W. Lees, S. D. Garvey, Dynamics of rotating machines, Cambridge University Press, 2010.

[3] D. Childs, D. W. Childs, Turbomachinery rotordynamics: phenomena, modeling, and analysis, John Wiley & Sons, 1993.

[4] O. Bernasconi, Bisynchronous Torsional Vibrations in Rotating Shafts, Journal of Applied Mechanics 54 (4) (1987) 893. doi:10.1115/1.3173135.

[5] J. S. Rao, T. N. Shiau, J. R. Chang, Theoretical analysis of lateral response due to torsional excitation of geared rotors, Mechanism and Machine Theory 33 (6) (1998) 761–783. doi:10.1016/S0094-114X(97)00056-6.

[6] X. Y. Shen, J. H. Jia, M. Zhao, J. P. Jing, Coupled torsional-lateral vibration of the unbalanced rotor system with external excitations, Journal of Strain Analysis for Engineering Design 42 (6) (2007) 423–431. doi:10.1243/03093247JSA304.

[7] B. O. Al-Bedoor, Modeling the coupled torsional and lateral vibrations of unbalanced rotors, Computer Methods in Applied Mechanics and Engineering 190 (45) (2001) 5999–6008. doi:10.1016/S0045-7825(01)00209-2.

[8] R. Sukkar, A. S. Yigit, Analysis of fully coupled torsional and lateral vibrations of unbalanced rotors subject to axial loads, Kuwait Journal of Science and Engineering 35 (2) (2008) 143–170.

[9] J. M. Balthazar, D. T. Mook, H. I. Weber, R. M. Brasil, A. Fenili, D. Belato, J. Felix, An overview on non-ideal vibrations, Meccanica 38 (6) (2003) 613–621.

[10] E. Sghaier, A. Bourdon, D. Remond, J.-L. Dion, N. Peyret, Non-stationary operating conditions of rotating machines: Assumptions and their consequences, in: International Conference on Condition Monitoring of Machinery in Non-Stationary Operation, Springer, 2018, pp. 11–19.

[11] A. Samantaray, S. Dasgupta, R. Bhattacharyya, Sommerfeld effect in rotationally symmetric planar dynamical systems, International Journal of Engineering Science 48 (1) (2010) 21–36.

[12] A. Bisoi, A. Samantaray, R. Bhattacharyya, Sommerfeld effect in a two-disk rotor dynamic system at various unbalance conditions, Meccanica 53 (4-5) (2018) 681–701.

[13] M. Karthikeyan, A. Bisoi, A. Samantaray, R. Bhattacharyya, Sommerfeld effect characterization in rotors with non-ideal drive from ideal drive response and power balance, Mechanism and Machine Theory 91 (2015) 269–288.

NAFID - A Grid Tool for output only modal analysis

Thanh-Trung DO, Viet-Hung VU, Zhaoheng LIU

Department of Mechanical Engineering
École de technologie supérieure, Montreal, QC, Canada
thanh-trung.do.1@ens.etsmtl.ca, viet-hung.vu@etsmtl.ca, zhaoheng.liu@etsmtl.ca

Abstract

In this paper, we propose a technique to enhance and facilitate the output only modal analysis of systems and structures by using the vector autoregressive (VAR) model. As we have witnessed, the VAR model with its robustness, accuracy, and noise - excitation resistance is beneficial for output only modal analysis. However, the VAR model and other parameters models have to deal with the variation of the model orders such as the frequency stabilization diagram. A grid technique is introduced to classify the natural frequencies and damping ratios in order to automatically evaluate its stabilization. The combination of the grid technique and the stabilization diagram will allow to users to have a better perspective of the modal parameters and a more accurate modes. The method is implemented and built in Matlab as the NAFID-tool which is users friendly and interactive. Examples on simulations of a MDOF system and on a real structure the applicability of the technique are illustrated to prove the efficiency of this technique.

1 Introduction

Natural frequencies, damping ratios and mode shapes, called modal parameters, are three important properties of mechanical systems and structures. Together with mathematical model, they allows us to analyze and predict dynamic behavior of systems under external excitations. For a mechanical system with several degrees of freedom, analytical model derived based on dynamic principles can be used to compute these parameters. However, for complex systems, modal parameters can be obtained using experiments. Although the Finite Element method (FEM) can be used for this purpose, however, as pointed out in [4], modal parameters for real-systems estimated by FEM is not accurate enough.

In practice, the well-known technique, called experimental modal analysis (EMA), is often used for this purpose [5]. By using the EMA technique, both excitation forces (input) and response (output) are used to identify modal parameters. In many situations, the excitation forces are unknown or cannot be measured. This leads to another technique for modal parameter identification, called operational modal analysis (OMA) [2]. In the OMA method, the modal parameters are extracted from the measured response and the excitation forces are modeled as white noise with zero mean.

The time domain has been found to be more suitable for the OMA method [7]. The AR model for single output and the VAR model for multi output can be used to estimate modal parameters from the measured response. The VAR model is proved very robust to identify natural frequencies even it can detect closed modes. This is because response of the system is measured simultaneously by many sensors.

When using VAR model, the selection of model order is the crucial because the size of the state matrix used to compute frequencies increases when the model order increases. The criteria proposed in [3] may be used for this purpose. In [9], a new method based on the concept of optimal model order was proposed for automatically classifying the modes and identifying the modal parameters.

In this study, the grid techniques is proposed to identify natural frequencies and damping ratios using measured response only without excitation forces. Basically, the method is developed based on the vector autoregressive (VAR) model in which the parameter model is obtained using multivariate least-square method. From the stabilization diagram, stable modal parameters are detected using the grid technique. Based on the proposed method, a new program, called NAFID-tool, has been implemented in Matlab. This tool can be used to identify modal parameters

This paper is organized as follows. In Section 2, the VAR model is presented in order to establish the state matrix and to compute modal parameters. In Section 3, the grid technique is addressed for identifying stable frequencies and damping ratios. Modal parameter identification of some mechanical systems is presented in Section 4. Finally, conclusions are given in Section 5.

2 Theoretical background

2.1 Vector autoregressive model

In the case of operational modal analysis (OMA), we assume that the excitation is unknown and may be modeled by Gaussian white noise. Using m sensors, output response of a mechanical system is measured at m predefined locations with constant sampling time Δt. The measured output data including n data points can be expressed by the following matrix

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & y_{2,3} & \cdots & y_{2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y_{m,1} & y_{m,2} & y_{m,3} & \cdots & y_{m,n} \end{bmatrix} = \begin{bmatrix} y_{[1]}, y_{[2]}, y_{[3]}, \ldots, y_{[n]} \end{bmatrix} \text{ where } y_{[i]} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{m,i} \end{bmatrix} \tag{1}$$

where $Y \in \mathbb{R}^{m \times n}$ is called as data matrix with $n \gg m$ and $y_{[i]} \in \mathbb{R}^{m \times 1}$ is the column i of Y, for $i = 1, \ldots, n$.

Based on the vector autoregressive model of order p, denoted by VAR(p), as presented in [7, 1, 10], dynamical model of the considered mechanical system may be expressed as

$$y_{[t+k]} = A_1 y_{[t+k-1]} + A_2 y_{[t+k-2]} + \cdots + A_p y_{[t+k-p]} + e_k \tag{2}$$

or equivalence to

$$y_{[t+k]} = \begin{bmatrix} A_1, A_2, \ldots, A_p \end{bmatrix} \begin{bmatrix} y_{[t+k-1]} \\ y_{[t+k-2]} \\ \vdots \\ y_{[t+k-p]} \end{bmatrix} + e_k \tag{3}$$

where $A_j \in \mathbb{R}^{m \times m}$ for $j = 1, \ldots, p$ are the autoregressive matrices, $y_{[t+k]}, y_{[t+k-j]} \in \mathbb{R}^{m \times 1}$ are vectors of the current and past response, respectively, and $e_k \in \mathbb{R}^{m \times 1}$ is a residual vector.

2.2 Evaluating parameter matrix

If we consider N consecutive values of the responses, Eq. (3) may be expanded for $k = 0, \ldots, N$. Therefore, the relationship between the current response and the previous (past) response is written in compact form as

$$B = \Phi_{\mathrm{A}} R + E \tag{4}$$

where $B \in \mathbb{R}^{m \times N}$ is the matrix of N responses, $\Phi_{\mathrm{A}} \in \mathbb{R}^{m \times pm}$ is the parameter matrix of the system, $R \in \mathbb{R}^{pm \times N}$ is the regression matrix of the output, and $E \in \mathbb{R}^{m \times N}$ is the model error matrix of the system. These matrices are defined as

$$B = \begin{bmatrix} y_{[t]}, y_{[t+1]}, \ldots, y_{[t+N]} \end{bmatrix} \tag{5}$$

$$\Phi_{\mathrm{A}} = \begin{bmatrix} A_1, A_2, \ldots, A_p \end{bmatrix} \tag{6}$$

$$E = \begin{bmatrix} e_0, e_1, \ldots, e_N \end{bmatrix} \tag{7}$$

$$R = \begin{bmatrix} y_{[t-1]} & y_{[t]} & \cdots & y_{[t+N-1]} \\ y_{[t-2]} & y_{[t-1]} & \cdots & y_{[t+N-2]} \\ \vdots & \vdots & \vdots & \vdots \\ y_{[t-p]} & y_{[t+1-p]} & \cdots & y_{[t+N-p]} \end{bmatrix} \tag{8}$$

From Eq. (4), the parameter matrix Φ_A may be obtained using the multivariate least-square method. This task is equivalent to compute the right pseudo-inverse of R as [5]

$$\Phi_A = BR^T \left(RR^T\right)^{-1} \tag{9}$$

In order to avoid computing the inverse matrix, solution of the linear least-square problem can be obtained using the robust techniques such as QR factorization, singular value decomposition (SVD), and LU factorization. For example in [8], the authors proposed the use of QR factorization to derive the parameter matrix.

2.3 Modal parameters

Once the parameter matrix Φ_A is determined, the state matrix of the discrete system, denoted by $\Phi \in \mathbb{R}^{pm\times pm}$, is established as [7]

$$\Phi = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I & 0 \end{bmatrix} \tag{10}$$

where $I \in \mathbb{R}^{m\times m}$ is the identity matrix. Because the state matrix represents the dynamics of the real system, it can be used to find natural frequencies and damping ratios. Assume that Φ can be decomposed as: $\Phi = VUV^{-1}$ where $U \in \mathbb{R}^{pm\times pm}$ is the diagonal matrix of eigenvalues and $V \in \mathbb{R}^{pm\times pm}$ is the matrix of eigenvectors. In Matlab, the matrix V and U can be found using the following function: $[V,U] = \text{eig}(\Phi)$. Consequently, each complex eigenvalue $U_{r,r}$ of the discrete system corresponds to one eigenvalue λ_r of the mechanical system

$$\lambda_r = \frac{\ln(U_{r,r})}{\Delta t} \tag{11}$$

Therefore, natural frequency ω_r (rad/s) or f_r (Hz) and damping ratio ζ_r are computed from complex conjugate pairs of λ_r as follows

$$\omega_r = \sqrt{(\text{real}(\lambda_r))^2 + (\text{imag}(\lambda_r))^2}\ (\text{rad/s}) \quad \Rightarrow f_r = \frac{\omega_r}{2\pi}(\text{Hz}) \tag{12}$$

$$\zeta_r = -\frac{\text{real}(\lambda_r)}{\omega_r} \tag{13}$$

When the model order p increases, there are more computational frequencies and damping ratios to be found from Eq. (10) to Eq. (13). That leads to more difficulty in classifying and identifying natural frequencies. The relationship between model orders and frequencies/damping ratios is described by the stabilization diagram. Stable frequencies/damping ratios corresponding to 'real' natural frequencies/damping ratios of the mechanical systems can be identified from this diagram. Other frequencies and damping ratios are unstable. They are due to the error of the identification model or measured data.

3 Modal parameter identifications using grid techniques

In this section, the grid technique is proposed to identify natural frequencies and damping ratios using the VAR model presented in the previous section. In addition, based on this technique, a new tool, called NAFID-tool (natural frequency identification), was successfully implemented in Matlab in order to identify stable frequencies and damping ratios from stabilization diagrams.

3.1 Grid technique

Assume that for each model order p, there are n_p natural frequencies (Hz), denoted by a vector $f^{(p)} \in \mathbb{R}^{1\times n_p}$ and n_p damping ratios (%), denoted by a vector $\zeta^{(p)} \in \mathbb{R}^{1\times n_p}$, to be found using Eq. (10) to Eq. (13) as.

$$f^{(p)} = \left[f_1^{(p)}, \quad f_2^{(p)}, \quad \dots, \quad f_{n_p}^{(p)}\right] \tag{14}$$

$$\zeta^{(p)} = \left[\zeta_1^{(p)}, \quad \zeta_2^{(p)}, \quad \dots, \quad \zeta_{n_p}^{(p)}\right] \tag{15}$$

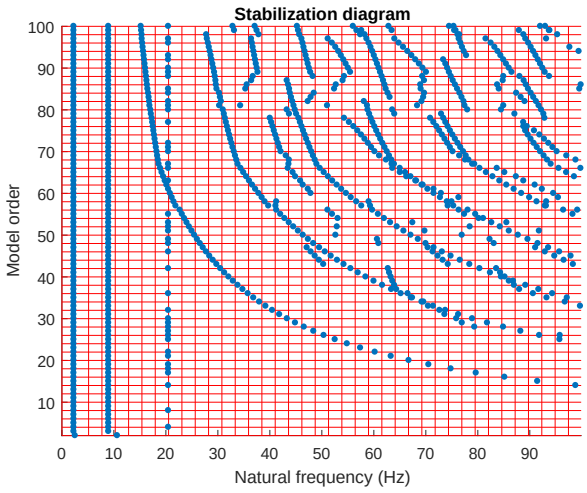


Figure 1 – Illustration of the grid technique.

When the model order varies in the interval $[p_1, p_2, \ldots, p_{\mathrm{m}}]$ the number of frequencies/damping ratios is $[n_{p_1}, n_{p_2}, \ldots, n_{p_{\mathrm{m}}}]$ in which $n_{p_1} < n_{p_2} < \ldots < n_{p_{\mathrm{m}}}$. In the proposed grid technique, whole frequencies and damping ratios of the system for the model order from p_1 to p_{m} are stored in two special matrices (called cell arrays in Matlab) as follows

$$F = \begin{bmatrix} f^{(p_1)} \\ f^{(p_2)} \\ \vdots \\ f^{(p_{\mathrm{m}})} \end{bmatrix}, \quad Z = \begin{bmatrix} \zeta^{(p_1)} \\ \zeta^{(p_2)} \\ \vdots \\ \zeta^{(p_{\mathrm{m}})} \end{bmatrix} \tag{16}$$

where F and Z are the frequency matrix and damping-ratio matrix of the system, respectively. Based on these matrices, stabilization diagrams can plot easily. The grid algorithm proposed here allows us to identify stable frequencies and damping ratios from the stabilization diagrams. Basically, the proposed method includes seven steps as follows:

Step 1 Define the model order range $[p_1, p_{\mathrm{m}}]$, the frequency range $[f_{\min}, f_{\max}]$ and the damping ratio range $[\zeta_{\min}, \zeta_{\max}]$.

Step 2 Compute all frequencies and damping ratios for $p = [p_1, p_{\mathrm{m}}]$ in the ranges $[f_{\min}, f_{\max}]$ and $[\zeta_{\min}, \zeta_{\max}]$

$$f_{\min} \le F \le f_{\max} \quad \text{and} \quad \zeta_{\min} \le Z \le \zeta_{\max} \tag{17}$$

Step 3 Define the frequency resolution Δf and make a virtual grid around the frequency range where the number of grid points is defined by

$$N_{\mathrm{f}} = \frac{f_{\max} - f_{\min}}{\Delta f} \tag{18}$$

Step 4 Define the number of repeating frequency in the interval $[\Delta f, 2\Delta f]$, denoted by N_{rf}, where:

$$1 \le N_{\mathrm{rf}} \le p_{\mathrm{m}} \tag{19}$$

Set $k = 1$

Step 5 Classify frequencies and damping ratios in the interval as follows

$$[I, \overline{F}] = \texttt{FIND\_STABLE\_FREQUENCIES}\left(f_{\min} + (k-1)\Delta f < F \le f_{\min} + k\Delta f\right) \tag{20}$$

$$\overline{Z} = \texttt{GET\_DAMPING\_RATIOS}(Z, I) \tag{21}$$

$$\overline{N} = \texttt{COUNT\_STABLE\_FREQUENCIES}(\overline{F}) \tag{22}$$

Step 6 If $\overline{N} \ge N_{\mathrm{rf}}$ then $\overline{F}$, $\overline{Z}$ and the index matrix I (index of $\overline{F}$ in F) is saved to a file.

Step 7 If $k \leq N_{\mathrm{f}}$ then $k = k+1$ and return **Step 4**. Otherwise, algorithm stops.

In the **Step 1** of the grid technique, the frequency and damping-ratio ranges of interest need to be determined from the user in order to eliminate frequencies and damping ratios out of ranges in the **Step 2**. However, if these ranges are unknown the following values can be used

$$f_{\min} = 0, f_{\max} = \frac{1}{2\Delta t}, \zeta_{\min} = 0, \zeta_{\max} = 100. \tag{23}$$

where $\frac{1}{2\Delta t}$ is the Nyquist frequency.

It can point out that main parameters of the grid technique are the frequency resolution Δf and the number of repeating frequency N_{rf}. By changing these parameter appropriately, the stable frequencies and damping ratios are identified quickly from the stabilization diagrams.

Figure 1 presents a simple example of the proposed technique to detect stable frequencies of the 2-DOF system under the harmonic excitation. Once all frequencies of the system are determined (the model order is increased from 2 to 100), a grid with red color is established over the frequency range from 0 to 100 (Hz). The frequency resolution used to make the grid is 2.5 (Hz) for better illustration. It can see clearly that stable frequencies with blue color lie between two straight red lines. They are detected using **Step 4** of the grid technique. We can conclude that the system may have two natural frequencies and one harmonic excitation associated with near zero damping ratio (see Figure 2).

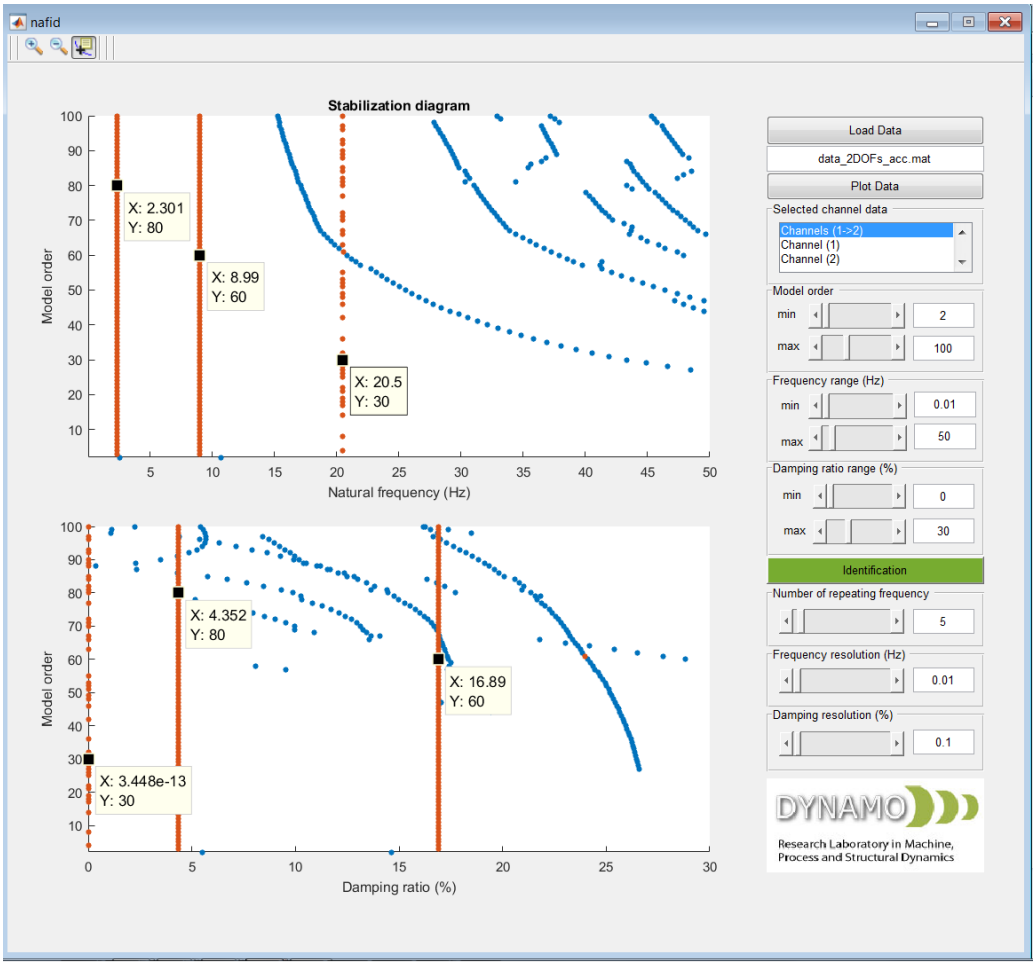


Figure 2 – Interface of the NAFID-tool implemented in Matlab.

3.2 About the NAFID-tool

The NAFID-tool was implemented in Matlab as shown in Figure 2 using the the proposed grid technique. This tool can be used to identify natural frequencies and damping ratios of mechanical systems or structures using the VAR model and the grid technique. On the left is the stabilization diagrams and on the right is the input parameters which are set by the user. The input of the NAFID-tool is the a file (*.mat) including: the sampling time and the measured responds. For example, three stable frequencies as well as three damping ratios of the system presented in Section 3.1 are identified and displayed by orange color. To obtain this result, the parameters of the program are set as follows: the frequency range is $[0,50]$ (Hz) and the resolution is 0.01 (Hz) while the damping-ratio rang is $[0,30](\%)$ and the number of repeating frequency is 5.

In the next section, modal parameter identification of several mechanical systems is presented to illustrate the efficiency of the grid technique. All results are obtained using the NAFID-tool.

4 Examples

4.1 Modal parameter identification of the 3-DOF system

In the first example, the measured data for the 3DOF system is taken from the Signal Processing Toolbox (Matlab, 2018b) [6] using the command 'load modaldata'. The measured data is shown in 3 in which the sampling rate is 4 (kHz).

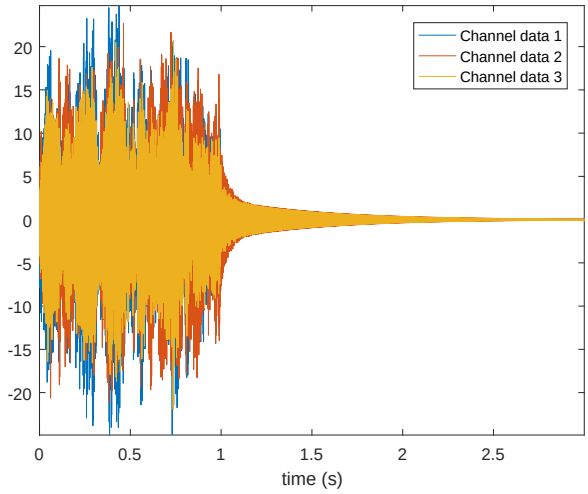


Figure 3 – Time response of 3 channels.

In order to estimate natural frequencies and damping ratios of this system, the following parameters are used in the NAFID-tool: $f_{\max} = 0$ and $f_{\max} = 2000$ (Hz); $\zeta_{\min} = 0.0$ and $\zeta_{\max} = 2(\%)$; $N_{\text{rf}} = 6$. Consequently, stable frequencies are identified as shown in Figure 4 by orange lines. The natural frequencies of the system are $f_1 = 373$, $f_2 = 852.3$ and $f_3 = 1369$ (Hz). The obtained results are almost identical to those obtained using the function 'modalsd' implemented in Matlab. In addition, damping ratios are also identified but very small, less than $0.6(\%)$.

4.2 Modal parameter identification of the 6-DOF system

The mechanical system with 6 degrees of freedom [4] is shown in Figure 5. Using the Lagrange formulation with the generalized coordinates $q = [q_1, q_2, q_3, q_4, q_5, q_6]^{\mathsf{T}}$, equations of motion of the system is given by

$$M\ddot{q} + C\dot{q} + Kq = f \tag{24}$$

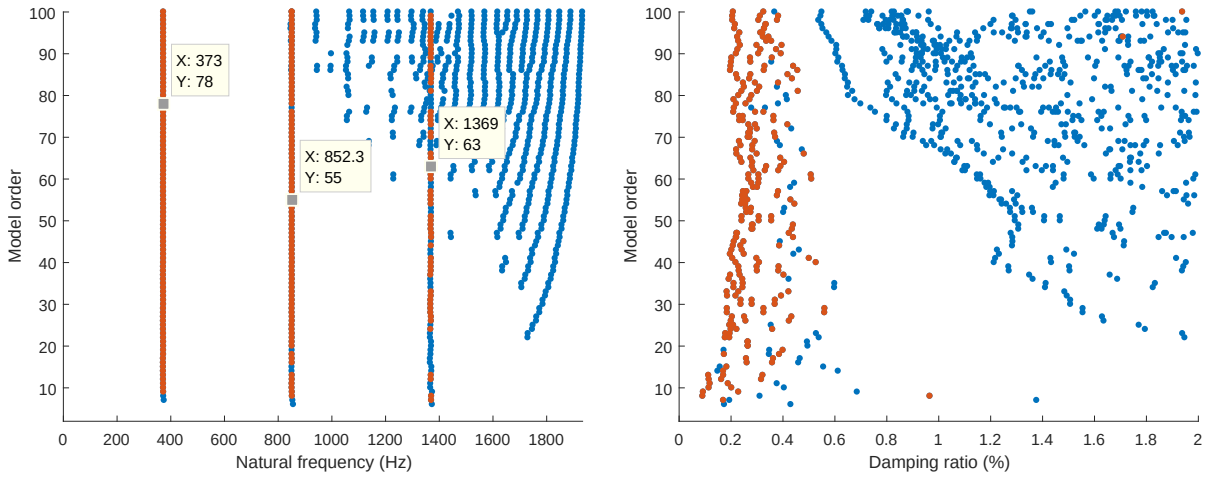


Figure 4 – Modal parameters of the 3DOF system using the NAFID-tool

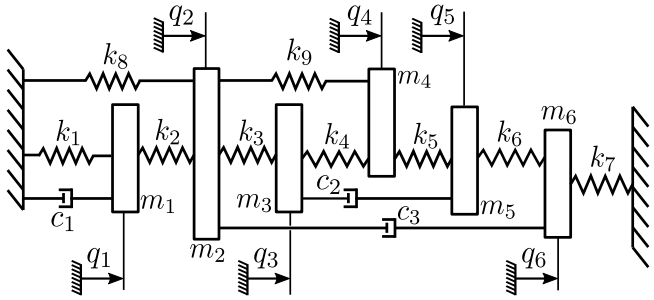


Figure 5 – The mechanical system with 6 DOFs.

where

$$M = \begin{bmatrix} m_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & m_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & m_6 \end{bmatrix}, \; C = \begin{bmatrix} c_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_3 & 0 & 0 & 0 & -c_3 \\ 0 & 0 & c_2 & 0 & -c_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c_2 & 0 & c_2 & 0 \\ 0 & -c_3 & 0 & 0 & 0 & c_3 \end{bmatrix} \tag{25}$$

$$K = \begin{bmatrix} k_1 + k2 & -k_2 & 0 & 0 & 0 & 0 \\ -k_2 & k_2 + k_3 + k_8 + k_9 & -k_3 & -k_9 & 0 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 & 0 & 0 \\ 0 & -k_9 & -k_4 & k_4 + k_5 + k_9 & -k_5 & 0 \\ 0 & 0 & 0 & -k_5 & k_5 + k_6 & -k_6 \\ 0 & 0 & 0 & 0 & -k_6 & k_6 + k_7 \end{bmatrix}, \tag{26}$$

For numerical simulation, physical parameters of the system are given by: $m_1 = m_2 = m_5 = 2\,(\text{kg})$, $m_3 = m_4 = m_6 = 1\,(\text{kg})$, $k_5 = k_8 = k_9 = 2.0\,10^6\,(\text{N/m})$, $k_1 = k_2 = k_3 = k_4 = k_6 = k_7 = 10^6\,(\text{N/m})$ and $c_1 = c_2 = c_3 = 10^3\,(\text{Ns/m})$. The system is excited by initial condition with $q = 0$ and $\dot{q} = [0, 10, 0, 0, 0, 0]^{\mathsf{T}}$ and vector of external forces is set to zero $(f = 0)$. Using function ODE45 in Matlab, displacement vector is plotted in Figure 6 in which the time step is 0.0001 (s).

Identified frequencies and damping ratios of the system are shown in 7. The parameters used to identify stable frequencies and damping ratios of the system are: $f_{\text{min}} = 0$ and $f_{\text{max}} = 500$ (Hz); $\zeta_{\text{min}} = 0.0$ and $\zeta_{\text{max}} = 70(\%)$; $N_{\text{rf}} = 5$ and $\Delta f = 0.1$ (Hz).

In addition, the theoretical and identified frequencies/damping ratios are shown in Table 1. It can see very good agreement between theoretical and identified modal parameters.

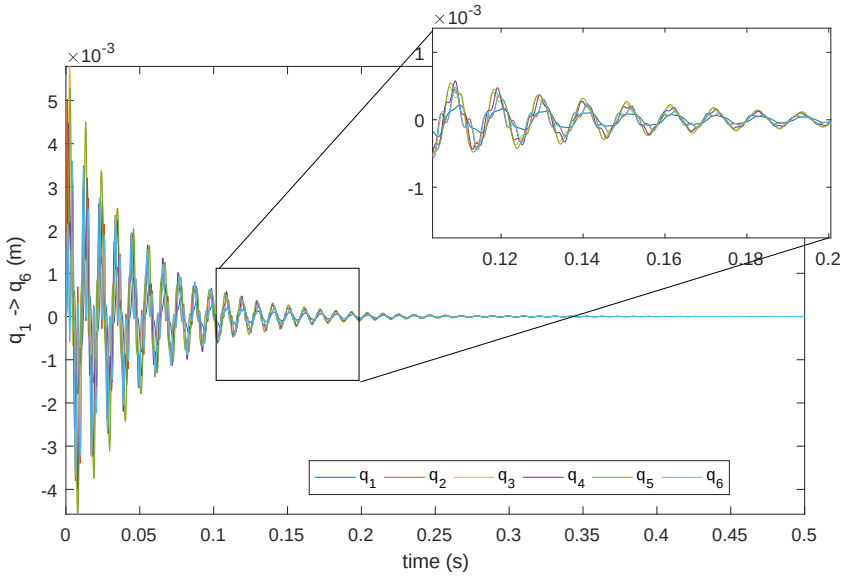


Figure 6 – Displacement vector is calculated using ODE45.

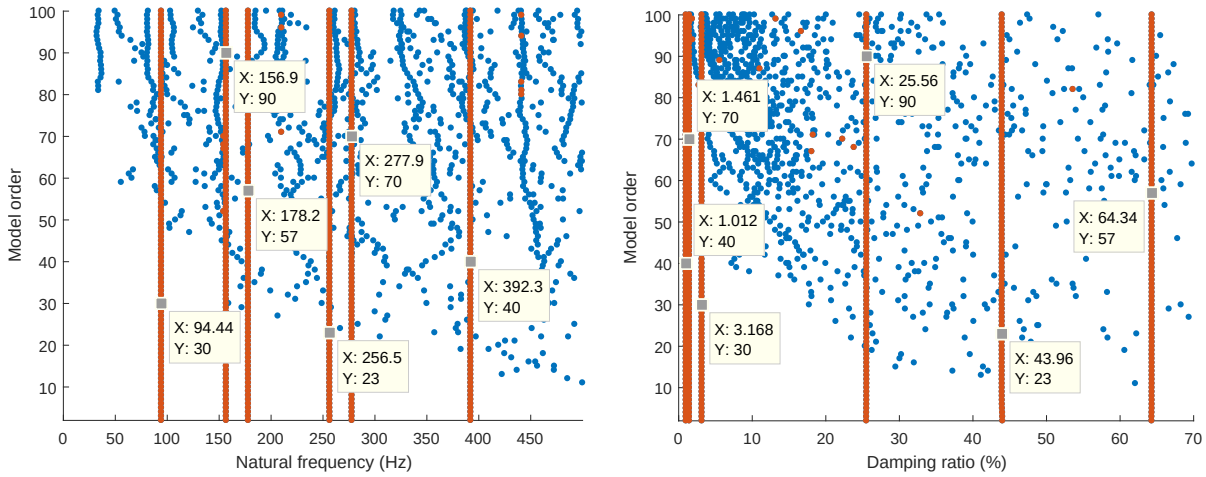


Figure 7 – Modal parameters of the 6DOF-mechanical system using the NAFID-tool

5 Conclusion

The grid technique was proposed in this work to identify modal parameters of mechanical systems using the vector autoregressive model for operational modal analysis. Based on the measured output response and the model order the state matrix was first established and then frequencies and damping ratios were computed using Matlab. Natural frequencies and damping ratios were classified quickly from the stabilization diagram using the grid technique. The NAFID-tool was successfully implemented in Matlab based on this technique. Results obtained based on numerical simulations show that the efficiency of the proposed method. The presented technique can further develop in the future for identifying modal parameters in real applications.

Acknowledgements

The authors would like to acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for its financial support.

Mode	1	1	3	4	5	6
Theoretical frequency (Hz)	94.4415	156.8693	178.1672	256.5395	277.9077	392.2243
Identified frequency (Hz)	94.44	156.9	178.2	256.5	277.9	392.3
Theoretical damping ratio (%)	3.1679	25.5645	64.3426	43.9668	1.4607	1.0039
Identified damping ratio (%)	3.168	25.56	64.34	43.96	1.461	1.012

Table 1 – Natural frequencies and damping ratios of the system using analytical method and the NAFID-tool

References

[1] M. M. Abdel Wahab and G. De Roeck. An effective method for selecting physical modes by vector autoregressive models. *Mechanical Systems and Signal Processing*, 13(3):449–474, 1999.

[2] R. Brincker and C. Ventura. *Introduction to Operational Modal Analysis*. Wiley, 2015.

[3] S. D. Fassois. Parametric identification of vibrating structures. In S. G. Braun, D. J. Ewins, and S. S. Rao, editors, *Encyclopedia of Vibration*, pages 673–685. Academic Press, 2001.

[4] M. I. Friswell and J. E. Mottershead. *Finite Element Model Updating in Structural Dynamics*. Springer, 1995.

[5] N. M. M. Maia and J. M. M. Silva, editors. *Theoretical and Experimental Modal Analysis*. Research Studies Press, 1997.

[6] MathWorks. Signal processing toolbox user's guide, 2018.

[7] S. M. Pandit. *Modal and Spectrum Analysis: Data Dependent Systems in State Space*. Wiley-Interscience, 1991.

[8] V.-H. Vu, Z. Liu, M. Thomas, W. Li, and B. Hazel. Output-only identification of modal shape coupling in a flexible robot by vector autoregressive modeling. *Mechanism and Machine Theory*, 97:141–154, 2016.

[9] V.-H. Vu, M. Thomas, F. Lafleur, and L. Marcouiller. Towards an automatic spectral and modal identification from operational modal analysis. *Journal of Sound and Vibration*, 332(1):213–227, 2013.

[10] V.-H. Vu, M. Thomas, A. Lakis, and L. Marcouiller. Operational modal analysis by updating autoregressive model. *Mechanical Systems and Signal Processing*, 25(3):1028–1044, 2011.

Study of the static and dynamic behavior of PU foams: from the sample to the automotive seat.

Corentin Blanchard[1,2], Thomas Weisser[2], Romain Barbeau[1],
Anne-Isabelle Mallet-Da Costa[1], Evelyne Aubry[2]
[1]Faurecia Automotive Seating, CS 10001, 91152 Etampes, France
[2]IRIMAS, Université de Haute-Alsace, 2 rue des Frères Lumière, 68093 Mulhouse Cedex
corentin.blanchard@uha.fr

Abstract

This study concerns the vibrational comfort of automotive seats. It is commonly characterized by the seat transmissibility, obtained by computing the ratio of the acceleration at the seat surface to the one at the seat base. This curve, observed in the frequency domain, depicts the seat performance in terms of vibrations filtration. The transmissibility is computed on a loaded structure, which means that an initial compression is present. Since the complete seat is a complex system, the first part of the paper presents static and dynamic experiments and simulations on a foam sample. The experiments are then reproduced using a finite element model. In the case of the static compression tests, the stress-strain curve is chosen as validation criterion. The simulation shows that for strains lower than 75%, the model is accurate. In the case of the dynamic experiment, the measured and simulated transmissibilities are compared. While the resonance and cut-off frequencies are close to the experimental results, the gain at the resonance is overestimated. The second section presents static and dynamic measurements performed on a seat cushion loaded by a rigid mass. The results show that the trim has an impact on the measured displacements and pressure distributions. In dynamics, it also has an impact on the resonance peak of the transmissibility. Finally, the simulation process proposed to validate the complete seat model is presented.

1 Introduction

The automotive sector faces new evolutions caused by the arrival of new features such as partial or full automated driving. These evolutions allow to have new reflections on the various postures that could be adopted by the occupants. This leads car manufacturers and suppliers to consider different seating positions. The driving position, where the occupant sits while holding the steering wheel, remains the most common. However, other positions are studied, such as the working position, where the occupant does not touch the steering wheel while the backrest is slightly reclined, or the relax position, where it is almost lying on the seat, are some examples.

Given these new reflections, the seat design remains constrained, either externally (by regulations or norms) or internally (depending on the features that need to be implemented). The present study focuses on the vibrational comfort, which is an internal constraint. The seat vibrational comfort is observed by measuring its transmissibility. It is obtained by computing the ratio of the transmissibility at the seat surface to the one at the seat base. The measurement has to be performed when an occupant sits on the seat. Three different classes of occupants can be considered : rigid masses, manikins and human subjects.

It is important to see this measurement as two separate steps: first, the occupant is placed on the seat. At the end of this step, the seat is deformed and a steady state is reached, corresponding to the equilibrium between the weight of the occupant and the reaction forces in the seat. The second step is the dynamic measurement itself, which consists in measuring the acceleration at the surface and at the base of the seat to derive the corresponding transmissibility.

The goal of this study is to perform such a process using a finite element model, which raises questions about the foam modeling. Indeed, transmissibility measurements using a rigid mass performed by Barbeau [6] show that the foam has a strong influence on the transmissibility. The polyurethane foam belongs to the viscoelastic materials, and its static and dynamic behaviors are non-linear. The validation must therefore be conducted at each step. To validate the results from the static step, it is possible to use different kind of measurements. For

instance, Zhang [4] uses force-displacement curves, Siefert et al. [3] compare the final displacement on the cushion and backrest, and Verver [9] works with pressure maps measurements. Regarding the dynamic step, the transmissibility is the criteria chosen in the articles from Zhang and Siefert et al.

Several authors have chosen to conduct deeper investigation of the foam behavior, and have therefore worked on foam samples. Krishan [7] and Verver [9] use tabulated stress-strain curves to model foam samples and use the data to conduct parametric studies. Siefert et al. [3] and Grujicic [10] have used analytical models and implemented them in complete seat models. Martinez-Agirre [2] or Lee and Kim [5] also propose dynamic models which take the initial deformation as an additional parameter, but this concerns other viscoelastic materials (rubber and thin films).

The goal of this study is to investigate the possible bridges between the material characterization at the sample scale and the experimental measurements at the seat scale. The first section regards the study of foam samples. It will be split into two steps: static and dynamic characterization. For each step, experimental validation is presented. The second section regards the experimental measurements conducted on the seat and introduces the simulation process to be used.

2 Analyses on a foam sample

The complete seat is considered as a complex system due to the high number of subcomponents involved. To properly simulate its global behavior, it is important to have accurate models for each of its subcomponents. This first section aims at studying the static and dynamic behaviors of the polyurethane foam used for the seat cushion and backrest. To do so, the analyses are carried on academic foam samples. The goal is to extract values that can be used as input data for static and dynamic simulations using a finite element model.

2.1 Static analysis

2.1.1 Measurements

The static characterization of the foam sample is done by measuring its stress-strain curves using an Instron 33R4204 machine following the protocol described in the ISO 3386-1 norm [11].

Three samples were used, each of them measuring $95mm \times 95mm \times 40mm$ and weighting $14g$. The choice of their size was dictated by the fact that they were cut out of the seat cushion and by the length/height ratio specified in the norm (equal or greater than 2). Five compression-decompression cycles are performed at $100mm/min$ up to 80% of deformation. The results are presented in figure 1.

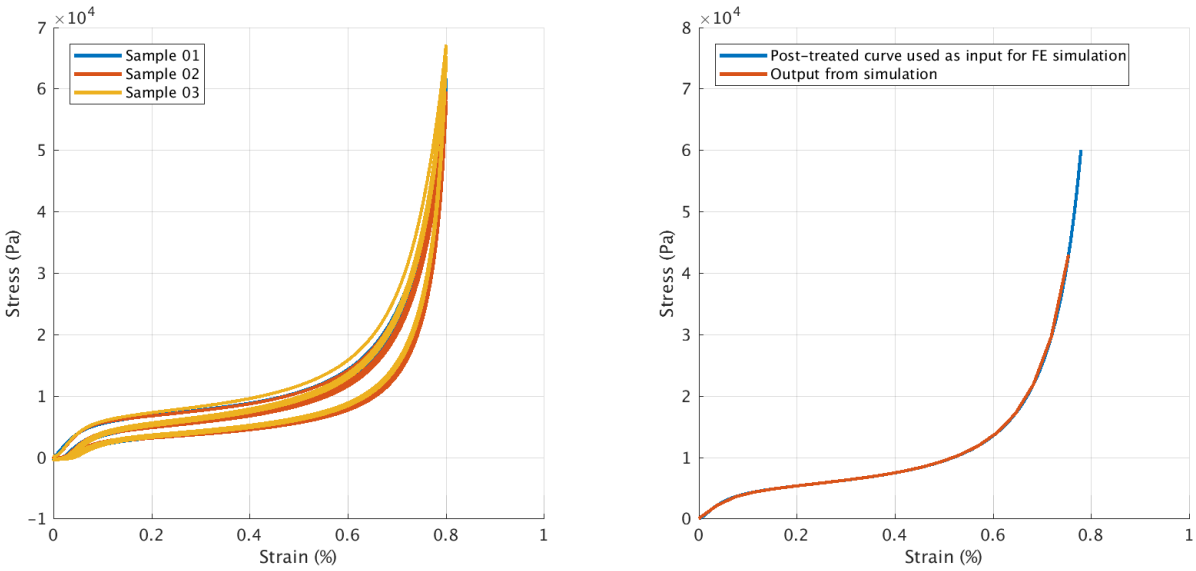


Figure 1: Left: Raw stress-strain curves measured on three foam samples. Right: Post-processed curve used as input for the finite element simulation (blue) and simulation result (red).

Although the curves associated to the first cycle are different depending on the sample, it is possible to observe the typical behavior of the foam in the three cases. First, a linear-elastic behavior for strains less than 5%, then a plateau region from 5% to 60%, and finally a densification region above 60%. It is also possible to see that, although the loading curves are different with regards to the cycle, the unloading curve always follows the same path. Finally, after the first cycle, the stress-strain curves are shifted, which is called the Mullins effect [1]. Each curve represents the static behavior of the foam sample and can be used as input data for the finite element model which is presented below.

2.1.2 Finite Element simulation

The use of the tabulated stress-strain curve for the foam static behavior is validated by conducting a simulation that reproduces the test presented above. The model consists in a foam sample made of 3D elements (ca. 4000 nodes and 3000 elements). The associated material model uses a stress-strain curve as input [8]. The last cycle of the experimental curve is extracted and only the compression part is used. An offset is applied to remove the Mullins effect from the curve and to make it start at the origin, see figure 1.

The compression is performed by prescribing a displacement at the top surface of the foam sample. The top surface is compressed at the same speed than the experiment up to 75%. The goal is to ensure that the stress-stain curve is accurately interpreted by the solver. The results are presented in figure 1. The comparison of the experimental and simulated stress-strain curves shows that the simulated curve follows the same path as the experimental curve.

2.2 Dynamic analysis

The static characterization of the foam is followed by its dynamic characterization. In this study, the transmissibility curve of a foam sample is presented.

2.2.1 Measurements

The dynamic characterization of the foam sample, performed by Barbeau [6], is made using its transmissibility curve. The sample is placed on a flat surface and is compressed by adding additionnal masses until reaching 30% of compression. This value is chosen based on Faurecia's know-how and corresponds to a commonly measured value when using an automotive seat. Then, a swept sine from 3 to $19Hz$ with a displacement amplitude of $\pm 0.5mm$ is applied at the base of the sample using a vibration shaker. The acceleration at the base $a_{base}(\omega)$ and at the surface $a_{top}(\omega)$ are measured. The ratio $T(\omega) = a_{top}(\omega)/a_{base}(\omega)$ is then computed and represented in figure 2.

Three parameters can be extracted from the transmissibility curve:

- The resonance frequency f_r: the frequency at which the transmissibility reaches its highest value,
- The gain at the resonance G_r: the value of the transmissibility at the resonance frequency,
- The cut-off frequency f_c: the frequency after which the transmissibility is lower than 1, thus characterizing the beginning of the filtration zone.

For the considered foam sample, the resonance frequency is $f_r = 5.81Hz$, the gain at the resonance is $G_r = 5.05$ and the cut-off frequency is $f_c = 8.95Hz$ (see table 1). The goal is then to compare these experimental results with the simulated transmissibility.

2.2.2 Finite Element simulation

Similarly to the static analysis, a finite element model of this experiment is built. The model consists in two parts, representing the foam sample and the rigid mass respectively. The mass density of the rigid part is chosen so that the part has the same mass as the experimental masses used to compress the foam at 30%.

The simulation process is divided into two steps. The first step is a simulation of the sinking of the mass. In that case, the material model used for the foam part takes the stress-strain curve as input. The gravity is applied to the model and the vertical displacement of the rigid mass is monitored. At the end of the simulation, the

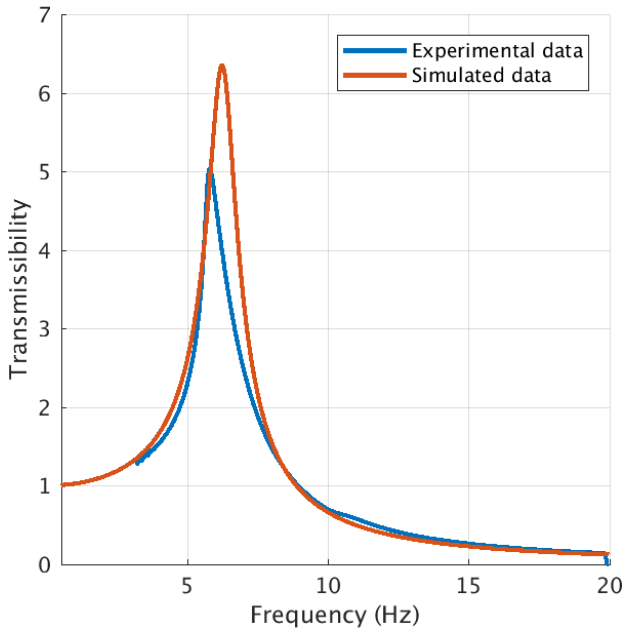


Figure 2: Experimental transmissibility curve (blue) and simulated curve with an equivalent elasticity modulus (red).

sample is compressed at 30%. The final state of the model (i.e. the nodal coordinates and the material stresses tensors) is exported and used as input for the dynamic simulation.

The second step is the dynamic simulation. In that case, the material model used for the foam needs to be changed. Indeed, when considering initial stresses with a material defined by its stress-strain curve, the solver computes an equivalent linear elasticity modulus taken as the tangent of the curve. Since the sample is compressed at 30%, this corresponds to the plateau zone. In that case, using the tangent modulus would result in a non-physical behavior.

To counter this problem, the transmissibility curve is approximated using the transmissibility of a 1-DOF mass-spring-damper system. Indeed, it is assumed that the excitation amplitude applied for the transmissibility measurement is low enough to excite the linear behavior of the material. For such a system, the transmissibility is defined as:

$$T(\omega) = \frac{k + jc\omega}{-m\omega^2 + jc\omega + k}, \tag{1}$$

where m, k and c are the mass, stiffness and damping coefficients of the equivalent model. Since m is known, the stiffness can be found using $k \approx m(2\pi f_r)^2$. The damping coefficient c is found using the graphical method of the $-3dB$ bandwidth. Here, $m = 8.3kg$ and the identified values are $k = 11000Nm^{-1}$ and $c = 47.1Nm^{-1}s$. The elasticity modulus is then found using the fact that $k = \frac{ES}{l}$, with S the sample top surface area and l the sample height. The associated modulus is then $E = 44000Pa$. This value is used as input for the linear-elastic material model used for the foam.

The transmissibility is computed using the modal superposition method. A modal damping of ζ defined as $\zeta = \frac{c}{2\sqrt{km}} = 0.08$ is used. The results are presented in figure 2. The comparison of both curves shows that the resonance frequency f_r and the cut-off frequency f_c are well caught. However, the gain at the resonance is overestimated. Table 1 summarizes the parameters derived from both transmissibility curves. This comparison shows the limitations of the asumption of linear behavior. Another way to see this is to notice that the resonance peak is inclined towards the left, indicating a softening behavior which is not captured with a linear 1-DOF hypothesis.

This concludes the analyses on the foam sample. The goal is then to apply this methodology to the automotive seat. The following section presents the different measurements performed on the automotive seat as well as the simulation process which is considered.

	Experimental	Simulated	Relative difference (ref. experimental)
f_r (Hz)	5.81	6.22	7%
G_r (.)	5.05	6.35	25.7%
f_c (Hz)	8.95	8.88	−0.8%

Table 1: Comparison of the parameters derived from the experimental and simulated transmissibilities.

3 Analyses on the automotive seat

As in the previous section, the analysis on the complete seat is divided into two steps: the static analysis, where the deformation of the seat under the load due to an occupant is studied, and then the dynamic analysis where the vibrational behavior of the seat is studied.

The seat chosen for the present study is a serial production seat used from a C-segment production car. It can be divided into 4 main subcomponents: a frame made of metal parts that corresponds to the main structure of the seat; two suspension mats, that are an assembly of metal wires and plastic parts, which are used for static comfort; the foam pads and the seat cover. The complete seat is represented in figure 3.

Figure 3: Exploded view of the complete seat finite element model.

3.1 Static analysis

As said above, before studying the seat vibrational behavior, a static analysis has to be performed. Its goal is to obtain the deformation of the seat under the load due to a given occupant. Three occupants can be distinguished: rigid masses, dummies or human subjects. Here, a rigid mass has been chosen to ensure the repeatability of the experimental results. The one used for the present stydy is called a lead buttock. It is made of a part that has the shape of the human thighs and buttocks connected to an arm. The whole structure is rigid and connected to the jig with a revolute joint.

To compare experimental and simulated data, a common point is needed. This point is called the Hip point (H-point) and corresponds to the pivot between the torso and the upper legs for a 50th percentile male occupant. During the design process, the H-point position is defined and the seat is defined using it as a reference. This point will be used to ensure that the rigid mass is accurately placed on the seat.

3.1.1 Measurements

The first measurement is the sinking of the lead buttock into the seat. To do so, the H-point of the lead buttock has to be placed at the H-point of the seat. Due to the geometry of the arm, this is not possible when the backrest is mounted. To ensure that both H-points are coincident, the backrest is therefore removed. Finally, the measurements are made according to two setups: first, when the seat is untrimmed (i.e. with no fabric cover on the foam cushion) and with the fabric cover.

The lead buttock is first placed so that its H-point matches seat one. The angle of the arm has to be nil to ensure that the angle of the lead buttock is 18°. This value corresponds to the femur angle of a seated human subject (15°) and a 3° tilt applied to compensate for thigh flexibility since the lead buttock is rigid. The sinking of the lead buttock is computed by measuring the distance between a point on the suspension mat with and without the occupant. This point is chosen in order to have the same X-coordinate than the H-point.

Values of the H-point as well as the sinking for each setup (untrimmed and trimmed seat) are presented in table 2, along with the nominal values taken as reference. The results show that the sinking of the lead buttock is lower when the seat is untrimmed. This shows that when the seat is untrimmed, the lead buttock sinks mainly in the foam (hence the lower value of the Z-coordinate), whereas when the seat is trimmed, the deformation is transferred to the suspension mat (hence the higher mat deformation).

	Nominal values (CAD)	**Trimmed seat**	**Untrimmed seat**
H-point X-coordinate (*mm*)	1256.6	1259.8	1259.4
H-point Z-coordinate (*mm*)	315	297.4	293
Mat deformation	NA	11.3	10.9

Table 2: Comparison between theoretical and measured values for the H-point and the mat deformation according to the seat setup.

The second static measurement available is the pressure map. Using a pressure sensor mat, it is possible to obtain the pressure distribution over the contact surface. The measurements have been performed on the trimmed and untrimmed seat following an internal Faurecia standard [13] with the lead buttock. The results are presented in figure 4. Both measurements present the same pattern, with a main contact surface reprensenting the bottom of the thighs, and three additional surfaces associated to the contact between the lead buttock and the seat bolsters.

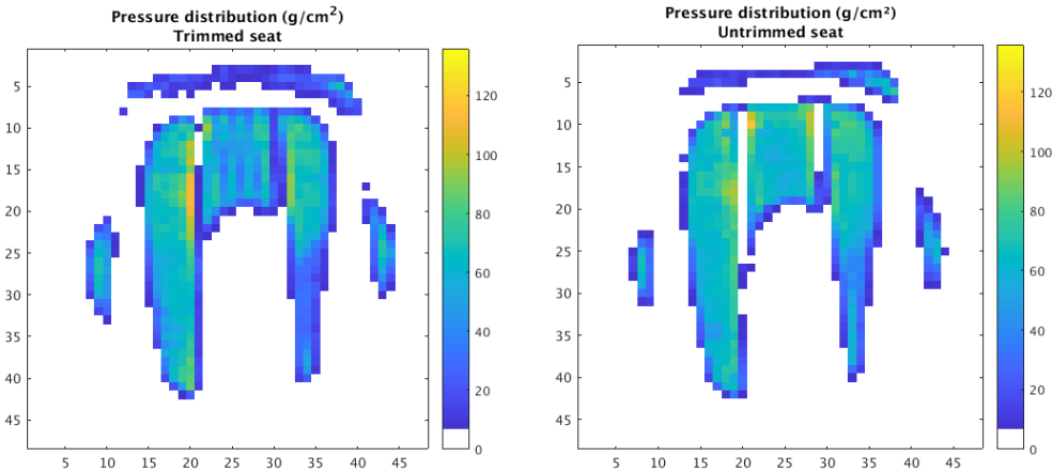


Figure 4: Pressure maps obtained depending on the seat setup.

Four parameters can be extracted from these measurements: the contact surface area, and the minimum, average and peak pressure. The values obtained for the two measurements are presented in table 3. Regarding the contact surface, the area measured for the untrimmed seat is 3.5% lower than the one obtained for the trimmed seat. Regarding the pressure, while the minimum value remains the same (i.e. the lowest value that can be measured by the sensors), the average and peak pressures are 13.2% and 2.5% higher respectively. This shows the impact of the trim, which helps to distribute the applied pressure over a larger area, thus resulting in

a lower average pressure.

	Trimmed seat	Untrimmed seat	Relative difference
Contact surface area (cm^2)	1033.87	956.45	-7.5%
Minimum pressure ($g.cm^{-2}$)	6.82	6.80	-0.3%
Average pressure ($g.cm^{-2}$)	42.94	48.62	13.2%
Maximum pressure ($g.cm^{-2}$)	111.93	114.7	2.5%

Table 3: Comparison between theoretical and measured values for the H-point and the mat deformation according to the seat setup.

It is interesing to see that the maximum pressure is $114.7 g.cm^{-2}$. When this value is multiplied by the gravity acceleration ($g = 9.81 ms^{-2}$), it is possible to obtain the maximum applied stress. Here, $\sigma_{max,exp} = 11.25 kPa$. This value is lower than the maximum stress obtained in the simulation ($\sigma_{max,sim} = 43 kPa$, see figure 1) and shows therefore that the tabulted curve can be used for simulations on the seat.

After the static characterization of the seat, the dynamic characterization is performed. Like the study on the foam sample, the transmissibility will be used.

3.2 Dynamic analysis on the seat cushion

3.2.1 Measurements

Following the static measurements on the seat cushion using the lead buttock, the dynamic analysis is performed. The seat and the lead buttock are kept in the same position as the static measurements and the transmissibility of the system is measured.

The experimental protocol is defined by an internal standard [12]. The excitation signal is a white noise with a bandwidth from 2 to $30 Hz$ and with an RMS value of $0.5 ms^{-2}$. The seat is first excited during $2 min$ until the steady state is considered as reached, and then the measurement is performed during $6 min$. Two accelerometers are placed on the test bench to measure the acceleration at the base $a_{base}(\omega)$ and at the lead buttock H-point $a_{H-point}(\omega)$. The transmissibility is then computed as $T(\omega) = a_{H-point}(\omega)/a_{base}(\omega)$. Once again, the measurements have been performed on the trimmed and untrimmed seat. In each case, the measurements have been done twice. The resulting curves are presented in figure 5 and the derived parameters are listed in table 4.

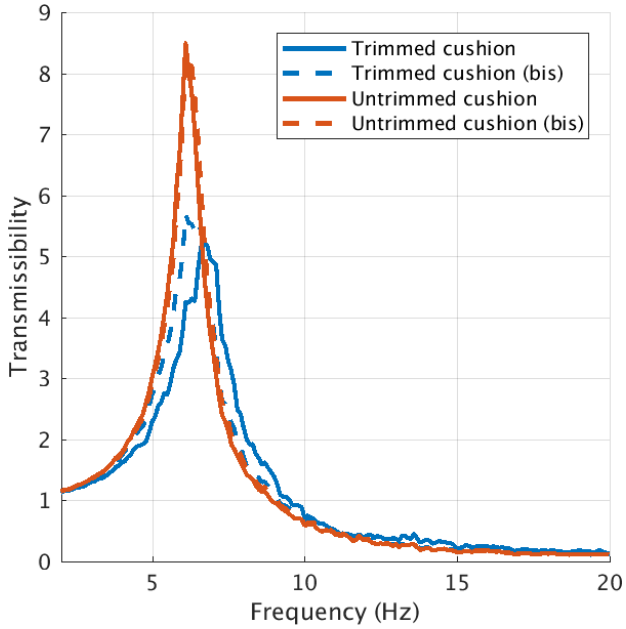


Figure 5: Transmissibilites obtained for the trimmed and untrimmed seat cushion with a lead buttock.

These measurements show the influence of the trim on the transmissibility curve. Indeed, when the seat cushion is trimmed, the gain at the resonance G_r decreases while the cut-off frequency f_c increases. The impact

	f_r (Hz)	G_r (.)	f_c (Hz)
Trimmed cushion	6.80	5.19	9.51
Trimmed cushion (bis)	6.10	5.69	9.15
Untrimmed cushion	6.10	8.50	8.90
Untrimmed cushion (bis)	6.10	8.28	9.08

Table 4: Values of the derived parameters depending on the seat setup.

of the trim on the resonance frequency f_r is harder to evaluate because of the strong difference between the two measurements for the trimmed cushion. When the seat is trimmed, damping is added and results in a lower but wider resonance peak.

3.3 Simulations

These experimental results have to be compared with numerical simulations to ensure that the modelling choices retained for the seat are valid. The simulation process is similar to the one applied to the foam sample.

The static analysis is performed by computing the sinking of the lead buttock on the seat cushion. The lead buttock is placed above the seat and the gravity is applied to the model. The material law used for the foam is the one from figure 1. The first goal of this simulation is to validate the model using experimental data: the location of the H-point, the arm angle, the mat deformation and the pressure maps are the four criteria that can be used. The comparison can be done with the results presented in tables 2 and 3 and in figure 4. The second goal of this simulation is to extract the final state of the model (deformed geometry and internal stresses).

This final state is then used as input for the dynamic simulation. Like the dynamic simulations done with the foam sample, a different material law than the one used for the static simulations needs to be used. The choice of the material model for the foam will be critical since the results of the dynamic simulations on the foam sample show that the use of an equivalent linear elasticity modulus overestimates the gain at the resonance. The results from this simulation have to be compared with values from figure 5 and 1.

4 Conclusions

In the first part of this study, the static and dynamic behaviors of foam samples are studied. The goal is to develop finite element models that match the physical behavior of the foam. The foam static behavior is obtained by performing compression-decompression tests. This allows to obtain the stress-strain curve of the material, which is then used for the finite element simulation. The results show that the model fits the experimental data up to 75% of deformation. The foam sample is then dynamically characterized by measuring its transmissibility. By assuming that the system behaves like a 1-DOF system, it is possible to extract linear-elastic equivalent parameters that will be used for the finite element simulation. The comparison shows that the resonance frequency and the cut-off frequency are well found, but the gain at the resonance is overestimated.

In the second part of this study, static and dynamic measurements performed on an automotive seat are performed. In statics, sinking and pressure distribution are performed when it is loaded with a rigid mass. When the seat is untrimmed, the rigid mat causes less suspension mat deformation and results in a higher average pressure on a smaller contact surface area, thus showing the role of the trim. In dynamics, transmissbilites of both setups are measured. The impact of the trim is confirmed since the resonance peak is lower and wider when the seat is trimmed.

While the use of a tabulated stress-strain curve allows to properly simulate the static behavior of the foam, it is shown that the modelling choices for its dynamic behavior could be improved. The asumption of a linear behavior of the elasticity modulus over the studied frequency range should then be rethought. A possible solution would be to use an enhanced model. The choice of such a model should be guided by additional measurements on the foam samples to characterize its material properties. This could also allow to study the influence of external factors such as the excitation frequency.

References

[1] L. Mullins, Effect of stretching on the properties of rubber, Rubber Chemistry and Technology, 21(2) (1948) 281–300.

[2] M. Martinez-Agirre, S. Illescas, M. J. Elejabarrieta, Characterisation and modelling of prestrained viscoelastic films, International Journal of Adhesion & Adhesives, 50 (2014) 183–190.

[3] A. Siefert, S. Pankoke, H.-P. Wölfel, Virtual optimisation of car passenger seats: Simulation of static and dynamic effects on drivers' seating comfort, International Journal of Industrial Ergonomics, 38 (2008) 410–424.

[4] X. Zhang, Y. Qiu, M. Griffin, Developing a simplified finite element model of a car seat with occupant for predicting vibration transmissibility in the vertical direction, Ergonomics, 58 (2015) 1–12.

[5] J.-H. Lee, K.-J. Kim Characterization of Complex Modulus of Viscoelastic Materials Subject to Static Compression, Mechanics of Time-Dependent Materials, 5 (2001) 255–271.

[6] R. Barbeau, Characterization and modeling of automotive seat dynamics: toward the robust optimization of vibrational comfort. Thèse, Université de Haute-Alsace, 2018.

[7] K. Krishan, Occupant-Seat Contact Pressure Characteristics of Polyurethane Foam Seats Using Explicit Finite-Element Analyses. Master's Thesis, Concordia University, 2017.

[8] LSTC, LS-DYNA Keyword User's Manual, Volumes I-III, Online, Accessed on Mai 28th 2019, Livermore, 2014.

[9] M. M. Verver, J. van Hoof, C. W. J. Oomens, J. S. H. M. Wismans, F. P. T. Baaijens, A finite element model of the human buttocks for prediction of seat pressure distributions, Computer Methods in Biomechanics and Biomedical Engineering, 7 (2004) 193–203.

[10] M. Grujicic, B. Pandurangan, G. Arakere, W. C. Bell, T. He, X.Xie, Seat-cushion and soft-tissue material modeling and a finite element investigation of the seating comfort for passenger-vehicle occupants, Materials and Design, 30 (2009) 42773–4285.

[11] International Organization for Standardization (ISO), ISO 3386-1 Determination of stress-strain characteristic in compression, Online, Accessed on Mai 29th 2019, 1998.

[12] GRP S RDT 1005, Evaluation de la réponse dynamique de l'assise d'un siège, Faurecia Standard, 2002.

[13] RBR I DSS 9450, Pressure mat Comfort Test Xsensor Procedure - Human Subject, Faurecia Standard, 2011.

Condition Monitoring

Edge computing for advanced vibration signal processing

Timothy VERSTRAETEN[1]*, Felipe GOMEZ MARULANDA[1]*, Cédric PEETERS[2], Pieter-Jan DAEMS[2], Ann NOWÉ[1], Jan HELSEN[2]

[1]Artificial Intelligence Lab Brussels
[2]Acoustics & Vibrations Research Group
Vrije Universiteit Brussel, Pleinlaan 2, 1050, Elsene

Abstract

Today, Industry 4.0 is being introduced. Machines are equipped with internet connection and increasingly sensorized using Industrial Internet of Things (IIoT) sensors. Especially the emergence of 5G is a game changer in this regard. It becomes possible to send data at high speeds to cloud computing data-centers. However, streaming all data is deemed to be unnecessary. It is more advantageous to use the additionally available bandwidth to drastically increase the number of connected sensors. Thus, on-board processing of the data directly at the edge is necessary. This paper illustrates this edge computing concept using data of wind turbines. Different fault indicators are calculated directly on an embedded system. In addition to statistical features also more complex signal processing pipelines combined with machine learning approaches are used. An example of a more advanced technique is the spectral coherence approach. This is one of the most promising approaches for bearing fault diagnostics to extract the optimal envelopes. This approach requires a significant amount of computational power. Today, different Advanced Risc Machine (ARM) processors are available in embedded architectures. Moreover, CPU based single board computers are available. Embedded GPUs allow dedicated machine learning algorithm processing. In this paper an NVIDIA Jetson device combining multiple ARM cores with a GPU is used. The edge computing concept is validated by processing pipelines on vibration and SCADA data originating from operational wind turbines using such architectures. Both healthy and faulty data sets are processed.

1 Introduction

In the context of Industry 4.0 efforts there is a continuously decreasing cost for sensors. As such the range of machines and other systems that are equipped with on-board instrumentation has increased substantially and will increase even more in the years to come. For those industries where the cost of downtime is high there is a strong interest and economic opportunity to move towards predictive maintenance. Therefore, more and more companies show interest to acquiring more data from their product for condition monitoring and design validation purposes. Continuous data collection allows to gain insights in product usage and thus forms the basis for design improvements from better understanding asset behavior in the field. Adequate processing algorithms are needed to perform usage evaluation and failure prediction to extract useful information from these sensors. Typically these algorithms use acceleration or current signals sampled at high frequency. The wide adoption of the Internet has brought broad coverage and continuous data connections at many locations all over the world. However, for many industrial applications the local connectivity can still be problematic due the limited bandwidth of wired or mobile connections. As such, streaming high frequency data is still unfeasible. Local processing is thus necessary and will become more important with increasing data volumes.

*These authors contributed equally.

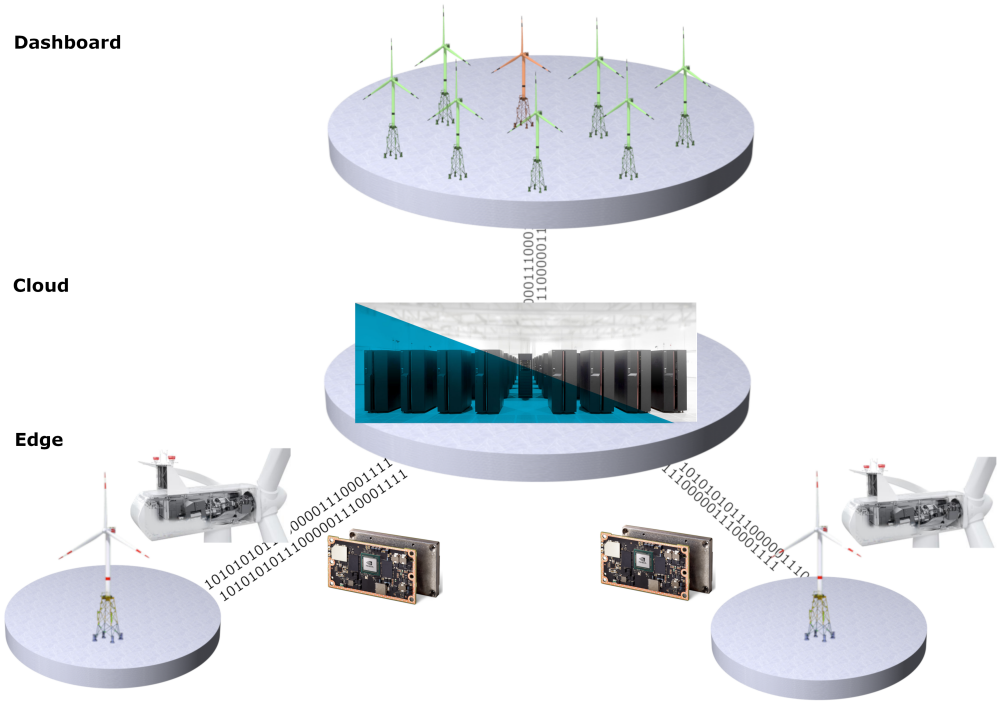


Figure 1 – Streaming data in the context of an edge-cloud balanced architecture for condition monitoring.

To overcome the streaming limitations, today, high frequency samples are typically taken for a short period of time in burst mode. Such bursts are done at intermittent periods in time. However, this means that for machines operating at non-constant speed and load chances are high that data-points are taken at different loading conditions. Due to the continuously changing nature of the system excitation, its response will also permanently change. These changes can have an influence on the resulting monitoring feature values. If the data samples—acquired at the intermittent moments—are spread too much over different loading conditions, then trending becomes challenging. Particularly since today's innovative industrial machines typically target a wider operational range, their speed and loads are varying continuously. To allow extraction of high quality condition indicators it becomes therefore interesting to explore, not only continuous data collection, but also continuous processing. This paper targets the latter.

Ideally we can instrument all machines in the fleet. The collected data should then allow the extraction of directly actionable insights for machine designers on the one hand machine and for owner-operators on the other hand. The one will use the insights to improve the design, whereas the other uses alarms to perform predictive maintenance. To allow the instrumentation of many machines it is necessary to have integrated processing algorithms capable of automatically processing the monitoring data. Edge computing can play an important rule to allow the extraction of health and design information from a large number of machines in a fleet where it might be unfeasible to transfer all data to a central location. Extensive research about the detection of failure in rotating machinery is available in literature today. More recently, machine learning is used more and more for condition monitoring. This offers opportunities towards automation. Leaning algorithms can enhance vibration signal processing methods to make them autonomous and more repetitive. This paper targets such methods by combining advanced signal processing techniques with anomaly detection and feature fusion based on data-driven techniques.

In this paper we target the assessment of the feasibility to use advanced edge devices for overcoming the limitations linked to intermittent data gap. We develop an integrated approach combining advanced signal processing methods with anomaly detection and a Bayesian regression approach to deal with vibration data in the new digital context. We target maximal computation close to where the sensor data is measured. Thus maximally leveraging processing power of the embedded ARM cores and GPUs. Devices of this architecture are plenty. In this paper, we use the NVIDIA Jetson TX2 embedded computing board, which features an ARM

for general-purpose computations, and a GPU for specialized fast matrix-vector computations that are apparent in many machine learning techniques. This device combines low cost with high computational density. To assess the feasibility of using this device in a condition monitoring context, we use data of healthy and failed wind turbines from the multi-megawatt range.

2 Methodology

Our goal is to maximally automate the condition monitoring process such that calculations can be done maximally at the edge. Those approaches that cannot be calculated at the edge we will calculate centrally a the cloud level. As such we generate a balance between edge and cloud, as illustrated in Figure 1. For those approaches that cannot be calculated at the edge level, we target to only use high quality data.

We focus on the wind turbine drivetrain system. Different monitoring pipelines are calculated in parallel on the edge device. In this paper we assess the possibility to calculate processing pipelines of different complexity on the embedded processors. A multitude of features is generated. These range from statistical indicators on raw sensor data, that are less computationally intensive, to complex sequences of filters. Anomalies in these features are afterwards annotated using machine learning. To optimize the usage of the calculation power of the edge device, analysis methods of different types are coupled to their most optimal processor type. All signal processing calculations are done on the ARM processors. Multiple ARM processors in parallel allow to calculate features on multiple channels at the same time. The anomaly detection using Bayesian approaches is then done using the GPU processor. This allows to exploit the fast matrix-vector computations.

Since the focus of this paper is on the assessment of the edge computation aspect, we only discuss the pipelines used in this paper in a high-level overview. For details on the different methods the reader is for each sub-block referred to our prior publications or relevant literature. The following paragraphs discuss these processing pipeline blocks.

Statistical indicators
Statistical values of different nature can be calculated on acceleration data to detect changes in vibration behaviour of the system over time. We use the following indicators:

1. RMS: This gives an indication of the overall energy level present, $x_{RMS} = \sqrt{\frac{1}{N}\sum_n x^2(n)}$, with $x(n)$ the sampled signal.
2. Crest factor: Max peak value over RMS, $CF = \frac{|x_{peak}|}{x_{RMS}}$.
3. Kurtosis: A measure for the dispersion of the signal's distribution, $\kappa = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^4}{(\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2)^2} - 3$.
4. Moors kurtosis: An alternative implementation of kurtosis based on quantiles [6], $\kappa_{Moors} = \frac{(E_7-E_5)+(E_3-E_1)}{E_6-E_2}$.
5. Peak-to-Peak: A straightforward indicator that quantifies the distance between the maximum and minimum acceleration, $x_{P2P} = x_{max} - x_{min}$.
6. Peak Energy Index: $PEI = \sqrt{\frac{1}{N_p}\sum_{n=1}^{N} x_p^2(n)}$, where N_p is the number of peaks exceeding a threshold equal to $\mu_x + 2\sigma_x$, with μ_x the mean and σ_x the standard deviation.

Speed compensation
Complex processing pipelines contain different processing steps that are chained together. For wind turbines a first step is always correction for speed fluctuations due to the stochastic nature of the wind. Typically this is achieved by converting the acceleration signals acquired in the time domain to the angular domain by means of angular re-sampling methods. Accurate speed measurements are necessary to achieve this step. Different methods are available in literature. We opt to use the Multi-Order Probabilistic Approach (MOPA). This method is based on interpreting the short time Fourier transform (STFT) of the vibration signal as a probability density function of the instantaneous angular speed. Consequently if the STFT has a high amplitude at frequency f,

then the probability that the shaft frequency is equal to f/H_i with H_i being the excitation order. For details on the specifics of the method the reader is referred to [4, 7].

Cepstrum liftering

In case bearing damage signature extraction is targeted, a second step consists in the removal of the dominating effect of the harmonics originating from the gears. Typically cepstrum liftering is used to achieve this goal. In this paper we opt to use the automated cepstrum editing procedure (ACEP). The cepstrum allows to concentrate the energy of periodically spaced spectral components into a smaller number of impulses. These are referred to as rahmonics. Removing these peaks decreases the amplitude of the periodic signal content. Similar to filtering in the frequency domain the term liftering is used in the cepstral domain. Randall & Sawalhi showed that the real cepstrum can be used to edit the log amplitude spectrum which contains the discrete harmonics [8]. Recombining this modified amplitude spectrum with the original phase creates an edited version of the time signal with significantly less pronounced harmonic content. In our case we use an automated cepstrum editing procedure. In this procedure first a long-pass lifter is applied to the cepstrum. The corresponding content will be read to the signal after the editing procedure. This in order to prevent liftering of this content which is dominated by the structural behaviour of the system. First noise reduction is achieved by a wavelet de-noising approach. Then a comb lifter allows the removal of selected distinct peaks in the cepstrum. Finally the signal is transferred back to the time domain. In the resulting signal the stochastic content is dominant.

Filtering

Once the data is cleaned and disturbances are removed, filtering is done prior to calculation of the statistical indicators. A variety of frequency ranges and filter types can be chosen. For the purpose of this paper the frequency range up to the Nyquist frequency is divided in 4 different bands. For each of these bands the same statistical indicators as for the raw data are calculated.

Cyclic spectral coherence

In parallel to the pipelines calculating statistical features, enveloping is done to detect bearing faults. Even after reducing the influence of harmonic disturbances using cepstrum liftering techniques the detection of bearing faults remains a challenge. The bearing fault impulsive excitation signatures are small compared to potential other disturbing energy sources. In complex gearboxes with many stages this can complicate detection. Enveloping techniques linked to band pass filtering allow to highlight the fault. However, accurate knowledge about the most optimal frequency band for envelope extraction is needed. The use of cyclic spectral coherence to identify this frequency band improves detection potential by identifying the frequency ranges in which resonances can amplify the signatures [2]. Calculation of the spectral coherence maps and envelopes are done both directly at the edge.

Anomaly detection

The features that result from the processing pipelines discussed above are treated as time series data. This allows them to be trended in order to accurately capture failure-driven anomalies. To automate and objectify this process, we use linear Bayesian Ridge regression [5] to model the features under healthy conditions using the machine operational parameters as inputs. Bayesian Ridge Regression is a probabilistic approach to regression with regularization. Essentially, it fits the linear parameters (i.e., slopes and intercept) and inherent noise to the observed data, while maintaining the uncertainty over the parameters. This type of regression fully captures all stochastic components in the modeling process, and allows the identification of anomalies that are not due to this stochasticity. The models are trained during a healthy period and thus predict expected feature response for healthy conditions. Outliers exceeding a three standard deviations range around the expected feature behaviour are classified as anomalies and thus unhealthy behavior.

3 Experimental case

To be able to keep up with the continuous data streams we opt to not transmit all data to a central cloud processing platform but balance processing between cloud and edge. Figure 1 gives a schematic overview of our architecture. In this paper we will make use of embedded ARM and GPU technology and use the commercially available NVIDIA Jetson TX2 product as edge device. This features a quad-core ARM Cortex-57 MP Core CPU. In addition also a 256-Core NVIDIA Pascal GPU is available.

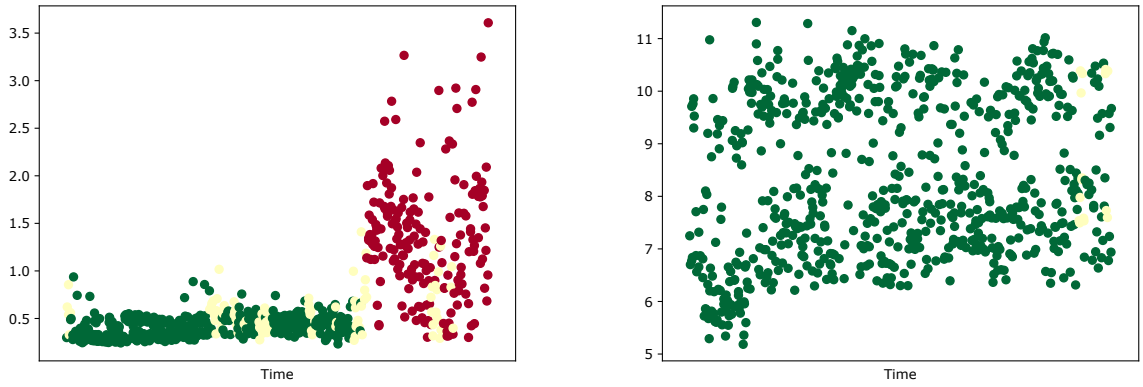
Signal processing pipelines exploit a Python architecture. These are deployed on the quad-core ARM processor. The anomaly detection models are implemented in TensorFlow [1], which allows for execution at the edge on the GPU. TensorFlow is a library that automatically organizes operations in a computational graph, such that many repetitive simple operations can be executed in parallel on the arithmetic units of the GPU. Such a framework is suitable, as Bayesian Ridge regression requires matrix multiplications during prediction, and thus can be efficiently executed on a GPU. Training of the models is performed in the cloud, as it is too computationally intensive.

To asses the processing ability of the Jetson and identify its limitations we opt to use vibration monitoring data collected from offshore wind turbines and process the analysis pipelines discussed above. Based on the insights gained in the speed and processing capabilities of the ARM processor and GPU, the balance between edge and cloud computing is identified. Computations that are feasible to be performed at the edge are done there, whereas the other processing is done in the cloud. To limit bandwidth usage of the data connections only high quality data is sent to the central cloud. The indicators calculated at the edge allow to determine whether it is useful to transfer the data to the central cloud.

Industrial CMS data is taken as starting point. In this way the analysis is done on a representative dataset. Data-sampling rates are above 25kHz. Each data block is approximately 10 seconds. Data is collected at intermittent moments in time. One accelerometer channel at the planetary and one at the high speed stage are processed. First, the processing pipelines using signal processing techniques, discussed in the previous section, are calculated. The six statistical indicators are calculated on the raw sensor data as a baseline. 240 processing pipelines provide an extensive feature set. The final step in each pipeline is the calculation of a statistical indicator to generate a summary value. To allow data-trending, intermittent data samples over a multi-year period are processed. After the trending step anomaly detection is executed on the GPU of the Jetson for fast prediction. Model training, comparison, and visualization of the resulting features is done on the VUB AVRG cloud platform.

Accelerometers are mounted on the gearbox at the planetary and high speed stage. We target the detection of deterioration on the planetary gear stage, which can lead to more severe damage. Constructed health indicators of both the planetary stage and high-speed stage are compared to illustrate deviations in nominal behavior. Figure 2 shows an example, comparing a peak-to-peak feature for the planetary stage and the corresponding feature for the high speed stage over multiple years of data. Based on these indicators it is clear that the fault is in the planetary stage. The indicators clearly show a strong anomaly score towards the end, whereas before some outliers already start to pop up. The indicators for the high speed stage show no anomalous behaviour. This shows the ability to locate the fault in the gearbox system.

Based on this analysis it is possible to perform calculations for failure detection, taking into account a large quantity of indicators calculated in parallel at the edge. For the moment calculations are far from real-time so only intermittent measurements are possible. As such more computationally intensive calculations using more detailed processing methods, such as for example the Kurtogram [3], need to be performed at the cloud level anyhow. However, there is definitely potential to use this technology for continuous condition monitoring if enough calculation cores are made available on the device.



(a) Planetary stage

(b) High-speed stage

Figure 2 – Comparison of peak to peak based health indicator between two affected and unaffected channels.

4 Conclusion

This paper investigated the potential to use combined embedded ARM and GPU processing architectures for edge computing in the context of condition monitoring. Both signal processing and machine learning approaches were calculate locally on the device. The NVIDIA Jetson TX2 was used as testing device. Using real-life data it was shown that failure detection can be achieved by edge computing. Complex signal processing pipelines, comprising of amongst others speed compensation, cepstrum liftering and enhanced enveloping were calculated on the device. These was complemented with Bayersian feature fusion using Tensor Flow on the embedded GPU after model training in the cloud.

As such it is shown that the edge device can be used to monitor a gearbox using typical measurements from CMS devices used in industry today. In addition to these local calculations the computationally more intensive calculations able to detect this failure type earlier will be performed in the cloud. In future research this balance will be further optimized.

Acknowledgements

The authors would like to acknowledge FWO (Fonds Weten- schappelijk Onderzoek - Vlaanderen) for their support through the SB grant of Timothy Verstraeten (1S47617N), the agency for Innovation by Science and Technology in Belgium for supporting the SBO HYMOP project

References

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] J. Antoni. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and vibration*, 304(3-5):497–529, 2007.

[3] J. Antoni. Fast computation of the kurtogram for the detection of transient faults. *Mechanical Systems and Signal Processing*, 21(1):108–124, 2007.

[4] Q. Leclere, H. André, and J. Antoni. A multi-order probabilistic approach for instantaneous angular speed tracking debriefing of the cmmno 14 diagnosis contest. *Mechanical Systems and Signal Processing*, 81:375–386, 2016.

[5] D. J. C. MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.

[6] J. Moors. A quantile alternative for kurtosis. *The statistician*, pages 25–32, 1988.

[7] C. Peeters, Q. Leclere, J. Antoni, P. Guillaume, and J. Helsen. Vibration-based angular speed estimation for multi-stage wind turbine gearboxes. *Journal of Physics: Conference Series*, 842(1):012053, 2017.

[8] R. B. Randall and N. Sawalhi. A new method for separating discrete components from a signal. *Sound and Vibration*, 45(5):6, 2011.

Remote diagnosis by analyzing heterogeneous data

Joris GUERRY, François CHANEL, Nicolas TARDIEU

EDF R&D CHATOU Lab, 6 quai Watier, Chatou, France
{firstname.NAME}@edf.fr

Abstract

Based on a modernization project of its information system, EDF's hydropower sector is now able to merge several heterogeneous databases to produce new applications. EDF Research and Development produced a new software to assess the impact of maintenance actions on performance of equipments. In practice, the tool allows, on the one hand, to display several time series of sensors between two dates, and on the other hand, to superimpose on the same graph the relevant maintenance acts, linked to the displayed sensors. We use the new efficient structuring of sensor nomenclatures and maintenance data to sort by relevance the maintenance events to be displayed first, based on expert knowledge.

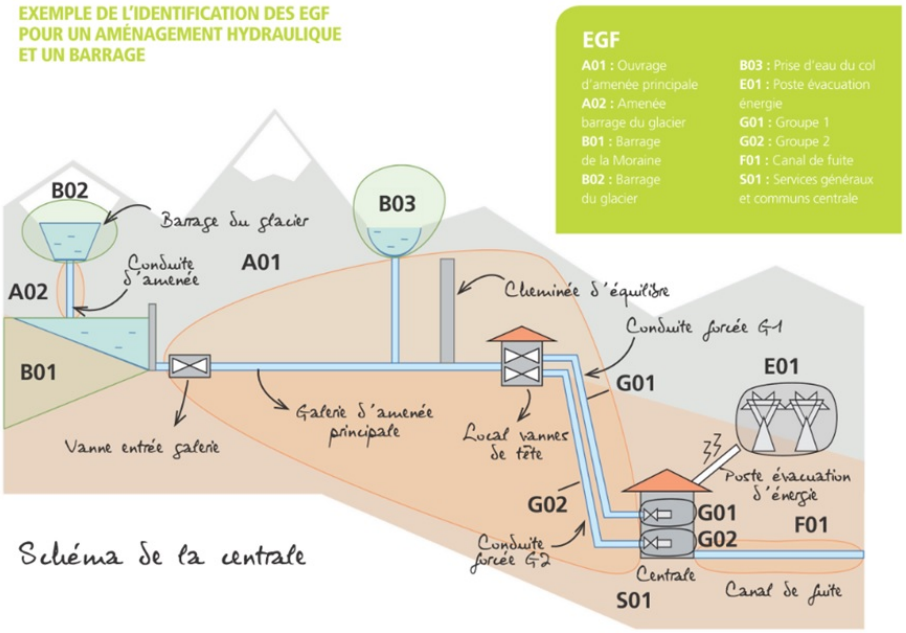


Figure 1: Schematic diagram of a hydraulic installation and name of the main elements (in French)

1 Technical context and need

1.1 Separation of process, maintenance and alert data

Historically, in the Production and Hydraulic Engineering Division of Electricité de France (EDF), process, maintenance and operation alert data were recorded and consulted in totally separate IT tools.

The data capitalized in these databases were therefore exploited separately, and no advantage was taken of any cross-checks and cross-references of information between databases, which could have led to a better understanding of events, a better anticipation of operational fortuitous events, and a better monitoring of the effect of the maintenance applied. One of the objectives of a data cross-referencing tool is to remove these information silos to allow enriched analyses by consulting heterogeneous databases.

1.2 Separation of operations, monitoring and maintenance units

As EDF's hydraulic division is large, it is segmented into different specialist units: some units specialize in the operation of hydraulic installations, others in the programming, implementation and operational monitoring of maintenance operations, and others in the e-monitoring of hydraulic production facilities.

While these different units cooperate effectively on a daily basis on many issues, each one carries a lot of information that is specific to its business lines, and the sharing of information between units cannot be perfect and complete on an ongoing basis. In particular, **some scheduled maintenance events may not be known to the e-monitoring units**, while these operations then generate monitored signals which, if not placed in the context of ongoing maintenance, may lead to a misinterpretation of abnormal situation and lead to a monitoring alert. These situations of imperfect interpretation then require additional communication time between operations, maintenance and monitoring engineers, whereas a centralizing digital tool could have made relevant information immediately accessible to everyone and avoided any incorrect interpretation of the signals read and operations in progress.

1.3 Standardization of equipment names (ECSH codes)

In recent years, EDF's hydraulic division has carried out a major project to standardize and homogenize the various tools of its information system, which has made it possible to standardize the model for forming the names of hydraulic power plant components (at least for high-power hydraulic sites). The template of names obtained is called ECSH codification (for EDF Coding System Hydraulic).

This standardization was an essential prerequisite for the project to create a heterogeneous data cross-referencing tool, because it was essential that an object present in the operating databases (as part of the time series that are measured and recorded via the various sensors with which it is equipped) as well as in the maintenance databases (as part of the various maintenance operations that have been carried out to ensure its reliability over time) bears the same name within these different tools: this makes it possible to identify which data are relevant to be processed together. This coincidence of names was therefore the first prism used to group heterogeneous data together.

Figure 1 shows the ECSH names used to designate the main structures of a hydropower plant : these names then serve as roots for the names of all the sub-components of each major structure. Figure 2 shows the decomposition of a tag in our hydraulic data historian. This tag is the name of a sensor time series, and it is directly derived from the ECSH code name of the component on which said sensor is placed. Figure 3 shows the decomposition of a component name associated to a maintenance event (dated text data) in our hydraulic Computerized Maintenance Management System (CMMS) named Gem'Hydro. As this is actually the same component in both cases (the component on which the sensor is placed and whose data are recorded, and the component on which a maintenance operation has been performed and then recorded), one can see that the parsing of the 2 names reveals the same construction by tree structure, which will later allow to group and compare these data.

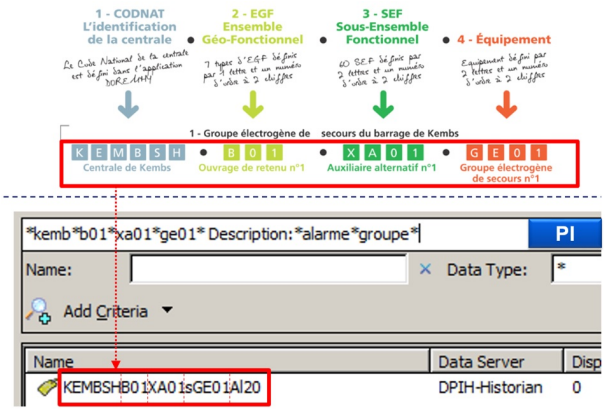
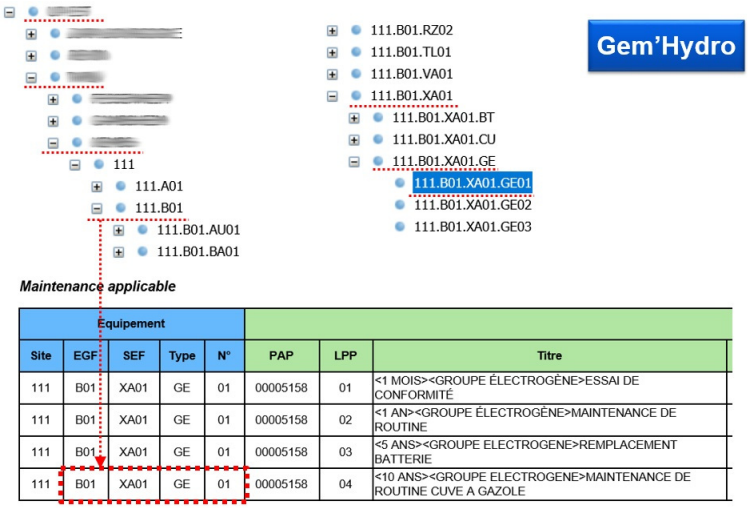


Figure 2: Parsing of the name of a PI tag linked to the equipment tree in which the monitored equipment is located



Equipement							
Site	EGF	SEF	Type	N°	PAP	LPP	Titre
111	B01	XA01	GE	01	00005158	01	<1 MOIS><GROUPE ÉLECTROGÈNE>ESSAI DE CONFORMITÉ
111	B01	XA01	GE	01	00005158	02	<1 AN><GROUPE ÉLECTROGÈNE>MAINTENANCE DE ROUTINE
111	B01	XA01	GE	01	00005158	03	<5 ANS><GROUPE ELECTROGENE>REMPLACEMENT BATTERIE
111	B01	XA01	GE	01	00005158	04	<10 ANS><GROUPE ELECTROGENE>MAINTENANCE DE ROUTINE CUVE A GAZOLE

Figure 3: Parsing of the name of a maintenance event in Gem'Hydro related to the equipment tree in which the maintained equipment is located

1.4 Target audience and expected benefits

The first audience targeted by this tool is Engineers performing e-monitoring of installations in CReEX (regional e-Operating Centres) and Engineers of the operator support teams.

The expected benefits of the tool include:

- The centralization of heterogeneous information (here: time series and structured textual data) in a single tool, thus facilitating their cross-analysis. The searchable data will be enriched as decisions are made to release other databases into the tool.
- The possibility of making visual comparisons before / after maintenance, in terms of sensor signals evolutions, and thus to assess qualitatively and quantitatively the effects produced by maintenance operations
- The possibility of directly seeing the quantitative effects of operating recommendations
- **Allowing a better circulation and exploitation of relevant information for the monitoring of installations, which also limits the risk of false alarms, limiting unnecessary solicitations and exchanges between monitoring, operation and maintenance units by sharing the same context.**

2 A new cross-referencing module

This tool is currently being developed as a Proof of Concept, in order to demonstrate the interest for the operational divisions in developing a data cross-referencing tool that allows them to make full use of their very large data assets (time series, highly structured document base, etc.). As such, it is still subject to several restrictions: only data from a few hydraulic sites are dumped there during the current test phase, the current IT architecture is not yet optimized to support a large number of concurrent requests. Depending on feedback and economic evaluation (effectively saved engineer time and avoided costs), additional developments may be considered for a truly robust industrial deployment.

In order to reduce EDF workload, we merged our development into an existing visualisation tool from another EDF entity, SuperViz'Orte, named after the first site testing it. Our data cross-referencing software then took the form of an additional module added to this more general tool for consulting and analyzing hydraulic time series. This module is shown in Figure 6.

2.1 Time series of our sensors: PI

The database used to capitalize the time series of hydraulic production facilities is the PI OSIsoft data historian, a robust and widely proven industrial solution. The PI OSIsoft Application Programming Interface

(API) allows to extract the PI tags corresponding to a set of parameters, which then allows the user to precisely choose the time series to be displayed over the period of his choice. The native services of the PI suite then allow you to directly request either the raw time series (extracted at the real time step of measurement and recording), or the average values calculated on the fly (limiting the total volume of data transmitted, when the exploitation of the data at a time step as narrow as that of the raw data is of no interest).

2.2 Act of maintenance database : Gem'Hydro

The CMMS tool used by the hydraulic division is an Oracle database called Gem'Hydro. This database imposes a strong structure in the recording of the various work orders: this structure has therefore made it possible to develop fairly fine weights for the selection of relevant maintenance tasks to be proposed for display, in addition to the process time series graphs.

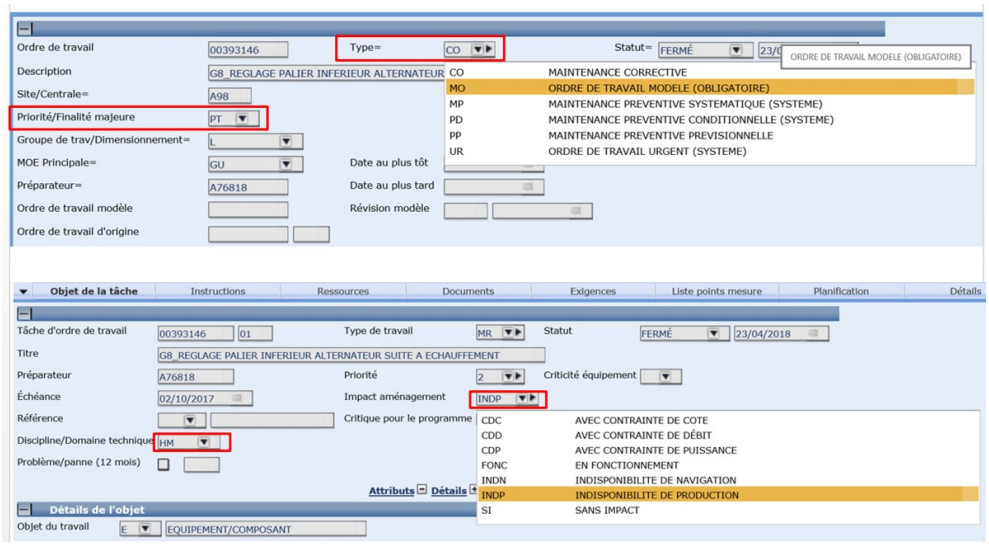


Figure 4: View of the different fields that apply to maintenance tasks registered in Gem'Hydro

Figure 4 shows a view of the different fields that must be filled in when recording a maintenance order or task. These different fields allow you to specify:

- The preventive or corrective nature of the maintenance action carried out
- The general purpose of the intervention (hydraulic safety, maintenance of the assets, increase in performance, etc.)
- The impact on production (impossible production, reduced river navigation, etc.)
- The general field of competence concerned (civil engineering, electrical engineering, control command, etc.)

In addition, the tool allows the recording of a comment in free entry. Operators sometimes comment extensively on the intervention performed and the salient information to be retained like shown on Figure 5. This input is to take into account with precaution as mistakes or abbreviations can be used.

The feedback from this database after several years of existence is double-edged: its robust design and the fairly exhaustive mandatory entry lead to a very good level of reliability of the information contained in most fields (only a few of the available fields had to be discarded due to a lack of satisfactory reliability). On the other hand, the relative "heaviness" of the associated interface and its not very ergonomic nature lead to a minimal input by the maintenance units: only strictly mandatory information (major work) is included, whereas ideally one would like to see the slightest small maintenance intervention entered completely.

Finally, since the system name tree structure through which maintenance tasks are recorded is the same as the one around which process PI tag names are constructed (both following the templates imposed by ECSH coding), it is relatively easy to associate the right maintenance acts with the right process time series curves.

Historique des activités								
Date de fin de réalisation	Objet du travail	Liste d'équipements	N° TOT	Type d'OT	Libellé de tâche	Commentaire de fin de tâche	Références documentaires du compte rendu	
01/06/2018	A98.G04.GU04.PB01		00365469-01	MP	<S><PIVOTERIE>CONTRÔLE PRESSION POMPE INJECTION_ME	ras		
01/06/2018	A98.G04.TU04.JS01		00450208-01	CO	REQUALIFICATION SUITE REMPLACEMENT JOINT A FUITES G4	Requalifs. ok => T°C et débit joint à fuite ok.	H-41554603-2018-000143	REQUALIF. G04 T°C + JOINT A FUITE_01062018
31/05/2018	A98.G04.AL04.AT04		00365448-01	MP	<R><ALT>VISITE X-MA-08-14 CTRL ROTOR ET INDEX_EL	Contrôle rotor réalisé, résistances de connexions, équi-répartition et index de polarisation sur le stator. R.A.S Présence d'un trou en partie baesse de la masse polaire sur le pôle 3 Suite au vernissage du rotor par OMEXOM ezn 2016 - aucune trasse d'extrusion de matière sur l'ensemble des pôles	H-41554603-2018-000141	20183005 G04 CONTRÔLE ROTOR
31/05/2018	A98.G04.AL04.AT04		00445255-01	MP	<ALTERNATEUR>MAINTENANCE LOURDE TÊTE D'EXCITATION_EL	Entretien lourd réalisé le 30/05/2018 suite à travaux EIM joint à fuite, contrôle rotor...: Aspirateur nettoyé, filtre remplacé, R.A.S. Isolement après remontage 1.5Gohm		
31/05/2018	A98.G04.TU04.TB04		00462280-01	CO	REMPLACEMENT DU JOINT À FUITE DU G4	Remplacement du joint fuite effectué	H-41554404-2018-000003	RT GMH - 2018 - G4 - REMPLACEMENT DU JOINT À FUITE
						Entretien léger à 500h + nettoyage des têtes de bobines pour essayer de faire remonter l'isolement.		

Figure 5: Comment fields associated with the maintenance tasks performed

2.3 Prioritization of maintenance operation reports

The tool is designed so that the user first chooses the production site and the sensor data (time series) he wants to display, as well as the start and end dates of the display. This forces him to specify which local group of hydraulic production units, site, production group and components he wants to display the sensor data of. The first step is to display the time series graphs, as shown in the top of Figure 6.

Once these first choices have been made, all the maintenance reports corresponding to this local grouping of units are displayed below the time series (see array in 6). However, the order in which they are proposed and displayed to the user is calculated according to a weighting that takes many criteria into account. **This weighting is intended to promote the immediate display of the most relevant reports, related to the theme explored by the user and the selected options.**

- Very high priority is given to reports on the same equipment as those whose sensor times series graphs are displayed, as well as those that are time-stamped between the start and end dates required for the display of sensor data
- The presence of filters (editable on the right of Figure 6) also makes it possible to modify the weighting of the reports: these filters include the attributes of the maintenance tasks in the Gem'Hydro database (presented in Figure 4), i.e. their selection will reinforce the priority of the reports presenting these same selected attributes. Several attributes have been checked by default (as per instructions from the operational units). It is also possible to add keyword search which would still prioritize acts of maintenance without rejecting any.

Then, the user is led to make a selection of reports relating the maintenance events that seem most relevant to the theme he is trying to explore (as long as the priority weighing of the reports is effective, and the selection criteria expressed are relevant, these are the reports that have been proposed to him first). The selected reports are then displayed as vertical bars in the time series graph window, as shown in Figure 6.

This cross-display therefore makes it possible to directly visualize whether the selected maintenance events correlate with the displayed time series. In the case studied in Figure 6, we displayed concomitantly:

- Time series of cold air temperatures of the generator in group 1
- Maintenance reports on the refrigeration system of the same alternator

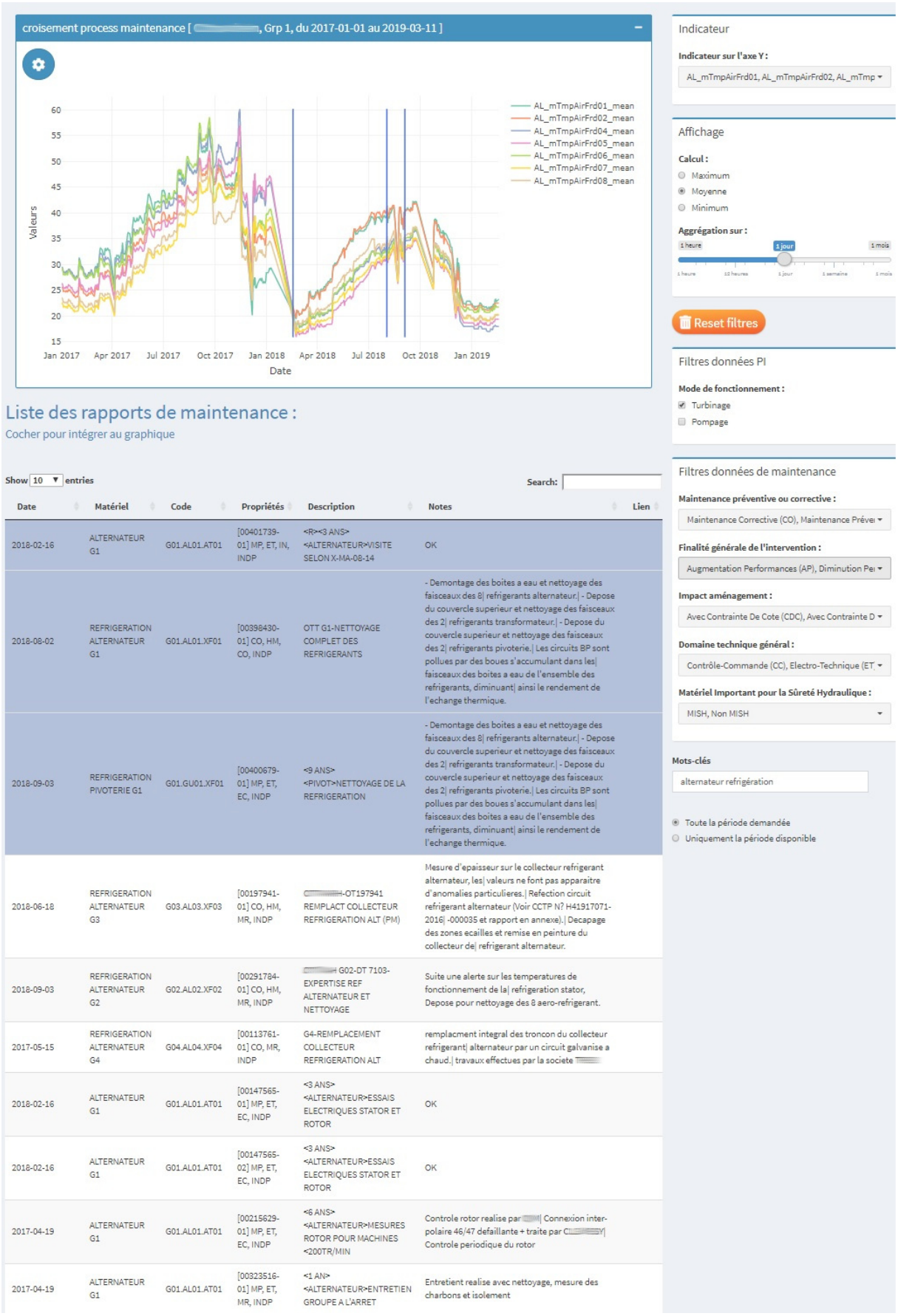


Figure 6: SuperViz'Orte cross-reference module screenshot (in French). The context restricts time series to alternator cold air temperatures and maintenance operations related to the alternator refrigeration system.

It can be seen directly on the Figure 7 that the maintenance operations carried out have the expected effect on the evolution of the component: at each maintenance operation on the refrigeration system, the alternator's cold air temperatures drop afterwards. Looking at the global Figure 6 might give the impression that the intervention causes an increase in cold air temperatures. In reality, the graphs display is partially misleading: since no operating data are recorded during the shutdown during which maintenance operations are carried out, straight lines are drawn between the actual data points, before and after maintenance, which may lead to the impression that there is a gradual decrease that stops as soon as the intervention on the refrigeration system is carried out, whereas in reality it is these operations that cause a drop in generator cold air temperatures, observed at the end of the intervention by measuring temperatures during the resumption of production (temperatures which then actually rise again and will lead to other acts of maintenance).

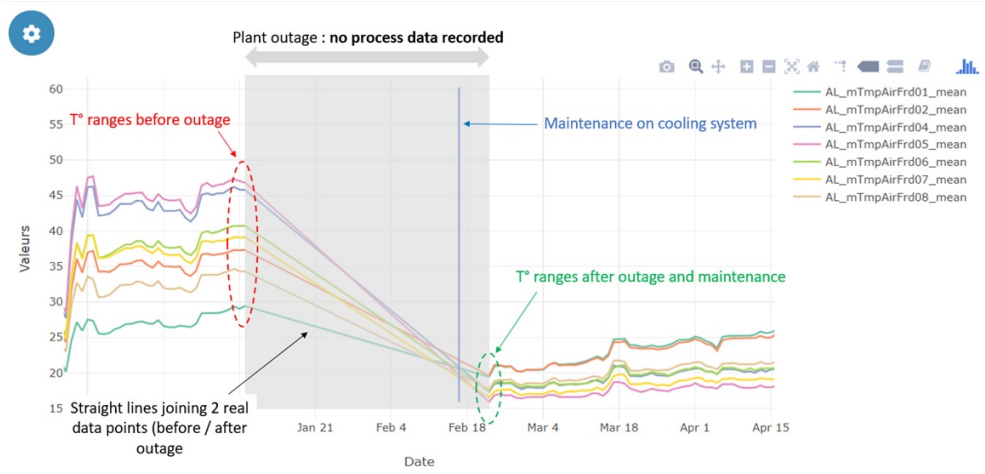


Figure 7: Focus on the first maintenance intervention: the plant outage and the maintenance performed are indeed at the origin of the temperature drops

3 Conclusion and perspectives

The tool has already been tested among the maintenance and monitoring units of EDF's hydraulic fleet. Feedback is considered positive: several relevant use cases (as shown in this article) have already been identified, showing that this application will be of real interest to these units in the future.

It is considered to add other different databases to the tool, so that it can probe a wider range of "events" to be displayed together with operational data: in particular, it is thought to add data on hazard and alert detections issued and capitalized by the operating and monitoring units. The expected benefits would be numerous: to note if the alerts are followed by actions (and to improve their consideration), to note if they have sufficiently anticipated the occurrence of possible operational hazards, ...

The weighing of relevance by which maintenance reports are prioritized is currently of relatively modest complexity. We could try to probe the documentary collection (in particular thanks to the free entry comment field) with more elaborate text mining tools, making it possible, for example, via the entry of a query expressed in French sentences and summarizing the general theme explored by the user, to propose a more robust order of relevance of the maintenance reports.

For the moment the interface only works in one direction: the user chooses which sensor data to display, then which maintenance reports to display among those proposed. One could consider an inverse operation, which would fully automate the selection of reports: according to the complete request expressed by the user, as well as the detection of salient events among the graphs which would then be exploited as ground truth (growth inversion, sudden changes in growth rate, etc.), propose to display the maintenance reports relevant for the context studied and time-stamped in the vicinity of these salient events of the operating data.

Acknowledgement

We thank our colleagues from EDF Hydro, and especially the CReEX UP Alpes who provided insight and expertise that greatly assisted the real heart of the developed module.

We thank Jérôme Boudon from EDF CIH for his assistance with the PI software part.

We would also like to show our gratitude to the "Usine Data Analytics pour les Producteurs" who gave us access to the Gem'Hydro database and provided us with their best R Shiny developers to merge our module into their visualization tool SuperViz'orte. A great thanks to Antoine De Caminel, Da Lan Nguyen, Laura Rouhier and Yann Renaud.

It was a real pleasure to work with all of them.

Tool Condition Monitoring Method in Milling Process Using Wavelet Transform and Long Short-Term Memory

Fatemeh Aghazadeh[1], Antoine Tahan[2] and Marc Thomas[3]

[1] Departement de Génie mécanique, École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec), Canada
fatemeh.aghazadehkouzekonani.1@ens.etsmtl.ca
[2] Departement de Génie mécanique, École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec), Canada
antoine.tahan@etsmtl.ca
[3] Departement de Génie mécanique, École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec), Canada
marc.thomas@etsmtl.ca

Abstract

Industrial automation is a promising move to fulfill today's competitive manufacturing industry demands by lowering operation costs, increasing productivity and quality. Monitoring the production process is one of the important steps toward total autonomy of manufacturing plants, which reduces routine checks, enables proactive maintenance and reduces repair costs. This research investigates tool wear as one of the most common faults in milling process during cutting of the D2 high speed steel as a hard to cut material using Carbide Walter End Mill Protostar tool. Vibration signal is chosen to represent the system status due to its applicability in industry. Signals are transformed into time-frequency domain using Wavelet Transform method to reveal both time domain and frequency domain features of the signal simultaneously. In order to model the complex and non-linear relations between tool wear and vibration signals under varying cutting parameters, a deep learning based algorithm, Long Short-Term Memory (LSTM) Artificial neural networks (ANNs) is employed. Deep learning algorithms are getting lots of attention recently within the diagnosis and prognosis community because of their exceptional performance in exploiting information in big data to solve complex problems. LSTM network is a type of recurrent ANNs that have some internal cells that act as long-term or short-term memory units, which is most suitable for sequential data and time series like vibration signals in our analysis. After designing the system, performance of the monitoring method is validated using experimentally acquired data with K2X10 Huron high speed CNC machine in LIPPS and Dynamo labs of ETS.

Keywords

Deep Learning, Tool Wear, Wavelet Transform, Condition Monitoring, Time-Frequency Transformation, Machining Process

1. Introduction

Machining processes are key components of industrial manufacturing, which requires higher productivity, parts quality, workers safety and lower operational costs. Therefore, there is growing demand to make the machining operation autonomous. Along with other initiatives in automation, online monitoring of machining process is beneficial to assure the production safety and quality. Tool wear is one of the most common and costly defects of the machining process, which is caused by excessive, contact forces and friction between cutting tool and workpiece material, high

temperatures in the cutting surfaces and pressure of the chips on the tool. It may deteriorates the surface finish or causes damage or breakage to the tool, workpiece or machining center if is not detected and fixed on time [1]. Therefore, designing a reliable and robust online automatic tool condition monitoring (TCM) system is in high demand to actively monitor the cutting process and provides actionable reports of tool condition status.

TCM systems can be divided into two main sub-categories: direct and indirect methods. Direct methods involve a procedure to directly measure actual value of faults with a laser, optical or ultra-sonic sensor. This could be costly and causes interruption in the manufacturing process for the measurements. Indirect methods in contrast employs physical parameters of the system such as force, vibration, etc. to indirectly reflect the status of the system [2]. Indirect methods can be used to fulfill TCM requirements as an alternative to indirect methods with accurate results and acceptable cost[3]. Another advantage of this approach is that the same sensor can be used for multiple monitoring purposes.

Force, vibration, acoustic emission, current and power signals are among the applicable and most common signals for TCM application in the literature. Li et al. studied TCM using force signals to reveal tool faults in turning[4]. Fourteen time-domain features of the signal are extracted and fed into a v-support vector regression model to developed flank wear prediction model. Force signal shows high accuracy to represent tool wear variations, however, it is also dependent on other operating conditions and relatively costly for industrial application [3]. Vibration sensors are practical in industrial environments and can represents the tool wear with appropriate performance. Harun et al. studied tool condition during deep twist drilling process using time and frequency domain fault features and compared vibration and force signals in this application. They concluded that both sensors are acceptable for this application, however they recommended vibration signal [5]. Acoustic emission is another efficient signal for TCM which is highly used in the literature [6] . Power and current sensors are also cost effective and applicable for industrial environment. In and study, current signals of the spindle of the milling machine is used to investigate tool wear. S-transform is used to transform the signals to time-frequency domain [7]. Sensor fusion is another approach to increase the accuracy and reliability of the monitoring. In an study, Segreto et al. combined information of the force, acoustic emission and vibration signals for tool condition monitoring of the turning process [8].

In the next step, signals are processed to magnify the effect of monitoring variables and reducing the effect of the noise in the signals. Time, frequency and time-frequency domain analysis are most common methods for signal processing in TCM [9]. Time-frequency analysis is appropriate for this application as it investigates both time variant and frequency dependent characteristics of the signal simultaneously, although it has higher computational costs [10]. In a study s-transform as a powerful time-frequency transformation method is used by Rehorn et al. to generate a feature called selective regional correlation, for machining condition monitoring [11]. In another study, a comparative analysis is conducted among common time-frequency transformation methods for the purposes of TCM in milling operation [12].

The relations between extracted features of the signals and tool wear is non-linear and complex, especially under varying cutting parameters such as depth of cut and feed rate which makes the monitoring task difficult. Therefore, a solid algorithm is necessary to accomplish the decision-making requirements. Machine learning algorithms such as artificial neural networks (ANNs), support vector machine (SVM) and Bayesian networks are common in the literature to fulfil this need. ANNs method is employed by Patra et al. to investigate tool wear of the micro drilling process [13]. In another study, a sound based system is developed using discrete wavelet transform (DWT) and SVM algorithms in face milling operation for TCM [14]. Tobon-Mejia employed Baysian network for the prediction of remaining useful life (RUL) of the tool in machining process [15].

Recently, deep learning algorithms draw attention of researchers in different fields due to their promising capabilities to solve complex challenges [16]. Deep learning refers to machine learning algorithms with deep multiple layers which enable them to learn highly complex patterns from even low-processed to raw signals [17]. In the era in which sensors are continuously producing enormous amounts of data, such techniques are in need to make the most information out of this data. These algorithms are less dependent on applications and frameworks and they are most efficient to outperform other methods when the relationship between the input data and desired outputs are complex [18]. Despite this potential, they are relatively new in the field of machinery fault monitoring. In an study, Jing et al. developed a Convolutional neural network based algorithm for gearbox condition monitoring [19]. Zhao et al. conducted a study to investigate the researches using deep learning methods in machine health monitoring [20]. Further research is crucial to examine deep learning algorithms applicability with different signals and levels of signal processing in TCM applications.

In this study, a TCM system is proposed using LSTM ANNs as a powerful and state of the art deep learning algorithm. Vibration signals from ETS experimental dataset are used to develop the monitoring system. Signals are processed using Wavelet method to transform them to time-frequency domain. Afterwards, the frequency bands energies calculated in the previous step is fed to the LSTM network as the features to construct the monitoring system. The algorithm accuracy is compared with a baseline Multi-Layer Perceptron (MLP) ANNs. This paper is organized as follows: Section 2 represents the formulation and backgrounds of the techniques of the paper. The proposed methodology is elaborated in Section 3. Results and discussion are presented in Section 4 and Section 5 is devoted to conclusion.

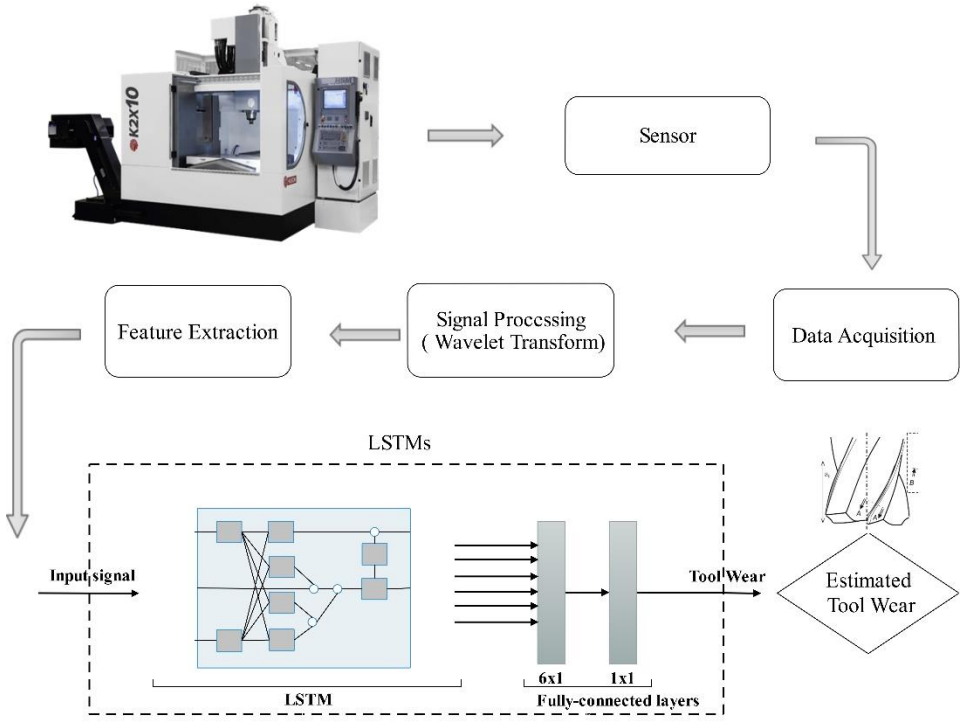


Figure 1.The monitoring system framework

2. Background of methods

2.1. Wavelet Transform

Wavelet transform is one of the widely used algorithms for fault diagnosis and health condition monitoring. In wavelet transform, wavelets are used as the basis instead of sinusoidal functions that are used in fast Fourier transforms which is the main difference between wavelet transform (WT) and Fast Fourier Transform (FFT). It is famous for transient signal analysis as well as time-frequency localization because it introduces a scale variable in addition to the time variable in the inner product transform. It has a better time localization but a lower frequency resolution for higher frequency components. In contrast, for lower frequency components, the frequency resolution is higher while the time localization is worse. Following equation describes the formulation of the continuous wavelet transform [10].

$$WT_x(t,a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u)\psi\frac{(u-t)}{a}\,du$$

where wavelet $\psi\frac{(u-t)}{a}$ is derived by dilating and translating the wavelet basis \$ $\psi(t)$, and $\frac{1}{\sqrt{a}}$ is a normalization factor to maintain energy conservation and $a > 0$.

2.2. LSTM Neural Networks

LSTM ANNS have recently demonstrated a great success in many machine-learning tasks, such as regression, prediction, etc. While conventional machine learning models can only map from input data to outputs, LSTM is capable of building multi-directional connections and it is effective at capturing long-term temporal dependences and keeps a memory of previous inputs to in the network's internal state, which makes it ideal for sequential data. The following equations is the hidden layer function that give the update for a layer of memory cells [20][21]:

$$\begin{aligned}
g_l^{(t)} &= \theta(W_l^{gx} h_{l-1}^{(t)} + W_l^{gh} h_{l-1}^{(t-1)} + b_l^g) \\
i_l^{(t)} &= \sigma(W_l^{ix} h_{l-1}^{(t)} + W_l^{ih} h_{l-1}^{(t-1)} + b_l^i) \\
f_l^{(t)} &= \sigma(W_l^{fx} h_{l-1}^{(t)} + W_l^{fh} h_l^{(t-1)} + b_l^f) \\
o_l^{(t)} &= \sigma(W_l^{ox} h_{l-1}^{(t)} + W_l^{oh} h_{l-1}^{(t-1)} + b_l^o) \\
s_l^{(t)} &= g_l^{(t)} \odot i_l^{(i)} + s_l{}^{(t-1)} \odot f_l{}^{(t)} \\
h_l^{(t)} &= \theta(s_l^{(t)}) \odot o_l^{(t)}
\end{aligned}$$

where σ is an element-wise application of the sigmoid function, θ is the $tanh$ function, and $\odot$ is the element-wise product. g is the input node with a $tanh$ activation function and i , o and f are the input, output and forget gates, respectively.

3. Proposed Methodology

The proposed methodology of this paper is elaborated in this section. In the signal acquisition step, an accelerometer is used to capture vibration data of the machine spindle for further processing. The framework of the monitoring system of this research is presented in Figure1.

Signals are processed after the acquisition step to extract informative fault indicators and remove noise. Time-frequency analysis is used for this step because of its superior potential in revealing the time variant characteristics of the signals in frequency domain using Morlet wavelet transform method.

In the next step, a set of features are extracted from the wavelet transform to describe the fault properly. The signal energy in different frequency bands are used as the monitoring features. Therefore, minimum pre-processing is implemented to explore the capability of LSTMs in eliminating unnecessary information and magnifying relevant features. In contrast to other hand-crafted feature learning models, deep learning methods are capable to provide an effective prediction tool for fault detections by learning robust feature representations directly from input signals.

A deep LSTMs model is proposed in this paper to accurately predict the faults in machining process. A simple yet effective architecture as shown in Figure 1 is considered due to the constraints of tool condition monitoring system. The keras deep learning library is employed [22] with tensorflow as the back-end [23] to implement the proposed model. The proposed architecture of the paper involves an LSTM with four neurons in the first hidden layer. then the output is fed into two fully-connected layers. The fully-connected layers are responsible to compute the softmax activation with a matrix multiplication followed by a bias in order to produce the prediction value. Mean Absolute Error (MAE) is chosen as the loss function. The model is fit during 2000 training epochs.

4. Results and discussion

4.1. ETS Experimental Dataset

A set of experiments are performed to measure tool flank wear during machining of hard to cut materials. K2X10 Huron high speed CNC machine of the LIPPS laboratory at ETS is used to perform the experimental tests. A tri-axial accelerometer was mounted on the spindle of the machine with a sensitivity of 100mV/g for measuring acceleration.

D2 high speed tool steel is selected as the workpiece material with hardness of 60-62 HRC due to its high wear resistance in order to investigate tool wear in machining hard material with dimension of 200 × 54 × 4. Carbide Walter End Mill Protostar H50 Ultra tool with 6 teeth is selected as the cutting tool with 50 degrees of helix angle. Different cutting speeds of 2500 rpm and 6000 rpm and feed rates of 0.12 mm/tooth and 0.05 mm/tooth with 4 mm depth of cut and tool wear were measured at different intervals which results in 63 cases with different tool wears and cutting conditions. Figure 2 demonstrates this experimental setup.

4.2. Tool wear estimation using vibration signals from ETS dataset

The monitoring system is developed as per the methodology described in the previous section. Also another system without the LSTM layer with just fully connected layers of ANN is developed as the baseline. Fully connected layers can be considered as multi-layer perceptron ANNs which are widely used in this application as a common ANN technique. Data is divided into two categories, training and testing with 70% and 30% of the data respectively. For evaluating the performance of monitoring systems, average accuracy in percentage (the differences between predicted and actual tool wear value divided by average of tool wears) and RMSE are calculated as representative of the performance from the Scikit-learn machine learning performance analysis toolboxes.

Figure 2. Experimental set up

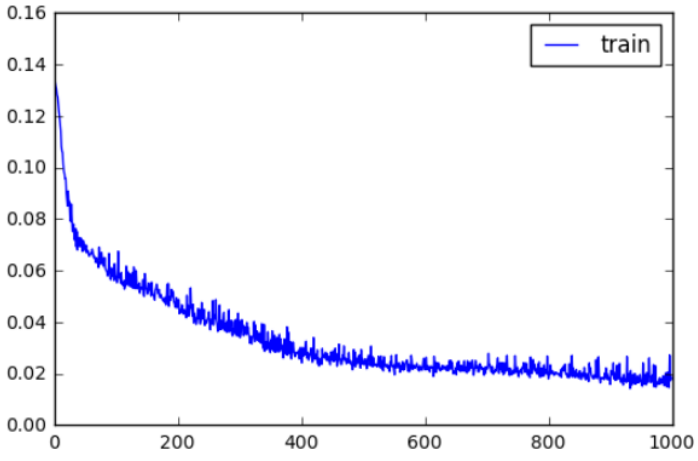


Figure 3. Loss function during training process

Figure 3 reports the loss values of the LSTM training method, which shows it is converging close to zero during the epochs of the training step. Table 1 presents the results of tool wear estimation using test dataset for two different algorithms.

Table 1. Comparison Between regression results

Regression Algorithms	Average Accuracy %	RMSE Test	RMSE Train
LSTM ANNs	92.37	0.00015	0.0001
MLP ANNs	82.21	0.00264	0.00139

Based on the results, LSTM has higher accuracy (92.4%) and lower root mean square error (RMSE) which are acceptable for most industrial applications. Figure 4 illustrates the predicted versus actual tool wears using the LSTM based algorithm for two tools from the no wear (VB=0) state up to the high tool wears. It is observed based on the diagrams and table that LSTM has a promising performance in this application.

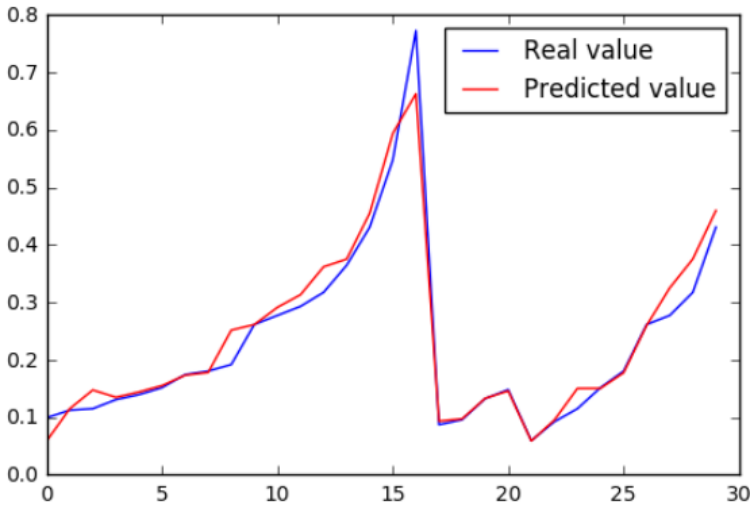


Figure 4. Estimated and real tool wear values using vibration signals

5. Conclusions

A robust tool condition monitoring method is proposed and validated in this research with ability to tolerate changing cutting parameters. Spindle vibration signals from the ETS dataset are used as the fault indicator. Wavelet transform time-frequency transformation method is employed for the signal processing step due to its great applicability to process signals and reveal rich information in both time and frequency domain simultaneously and its proven performance in this application. A deep LSTM based ANNs method is also implemented as the last step to model the complex relationships between extracted features and tool wear.

Time frequency step of the research revealed information on both time domain and frequency domain characteristics of the signals and the study confirms its performance and effectiveness in tool wear monitoring. Table 1 report the comparative results of the LSTM ANNS based proposed methodology of the paper versus MLP ANNs which is one of the most common and widely used ANNs in the Literature. Based on the results, LSTM outperforms MLP with above 10% in accuracy and it has a significantly lower RMSE for both training and test results. So it proves the applicability of LSTM for tool wear estimation.

As the next steps of this study, the results will be validated with other common sensors in this application, especially more economic and applicable sensors such as power and current sensors. In addition, sensor fusion techniques will be investigated at different levels of analysis to increase accuracy and robustness of the system.

References

[1] Kunpeng Zhu and Birgit Vogel-Heuser. Sparse representation and its applications in micro-milling condition monitoring: noise separation and tool condition monitoring. The International Journal of Advanced Manufacturing Technology, 70 (1-4):185–199, 2014.

[2] A Siddhpura and R Paurobally. A review of flank wear prediction methods for tool condition monitoring in a turning process. The International Journal of Advanced Manufacturing Technology, 65(1-4):371–393, 2013.

[3] Jose Vicente Abellan-Nebot and Fernando Romero Subiron.´ A review of machining monitoring systems based on artificial intelligence process models. The International Journal of Advanced Manufacturing Technology, 47(1-4): 237–257, 2010.

[4] Ning Li, Yongjie Chen, Dongdong Kong, and Shenglin Tan. Force-based tool condition monitoring for turning process using v-support vector regression. The International Journal of Advanced Manufacturing Technology, 91(1-4): 351–361, 2017.

[5] MHS Harun, MF Ghazali, and AR Yusoff. Analysis of tri-axial force and vibration sensors for detection of failure criterion in deep twist drilling process. The International Journal of Advanced Manufacturing Technology, 89(9-12): 3535–3545, 2017.

[6] Javad Soltani Rad, Youmin Zhang, and Chevy Chen. A novel local time-frequency domain feature extraction method for tool condition monitoring using s-transform and genetic algorithm. IFAC Proceedings Volumes, 47(3):3516–3521, 2014a.

[7] Javad Soltani Rad, Ensieh Hosseini, Youmin Zhang, and Chevy Chen. Online tool wear monitoring and estimation using power signals and stransform. In Control and Fault-Tolerant Systems (SysTol), 2013 Conference on, pages 234– 238. IEEE, 2013.

[8] T Segreto, A Simeone, and R Teti. Multiple sensor monitoring in nickel alloy turning for tool wear assessment via sensor fusion. Procedia CIRP, 12:85–90, 2013.

[9] Adam G Rehorn, Jin Jiang, and Peter E Orban. Stateof-the-art methods and results in tool condition monitoring: a review. The International Journal of Advanced Manufacturing Technology, 26(7): 693–710, 2005.

[10] Zhipeng Feng, Ming Liang, and Fulei Chu. Recent advances in time–frequency analysis methods for machinery fault diagnosis: A review with application examples. Mechanical Systems and Signal Processing, 38(1):165–205, 2013.

[11] Adam G Rehorn, Ervin Sejdic, and Jin Jiang. Fault di-´ agnosis in machine tools using selective regional correlation. Mechanical Systems and Signal Processing, 20(5):1221–1238, 2006.

[12] Javad Soltani Rad, Youmin Zhang, Fatemeh Aghazadeh, and Zezhong Chevy Chen. A study on tool wear monitoring using time-frequency transformation techniques. In Innovative Design and Manufacturing (ICIDM), Proceedings of the 2014 International Conference on, pages 342– 347. IEEE, 2014b.

[13] K Patra, AK Jha, Tibor Szalay, J Ranjan, and Laszl´ o´ Monostori. Artificial neural network based tool condition monitoring in micro mechanical peck drilling using thrust force signals. Precision Engineering, 48:279–291, 2017.

[14] CK Madhusudana, Hemantha Kumar, and S Narendranath. Face milling tool condition monitoring using sound signal. International Journal of System Assurance Engineering and Management, 8 (2):1643–1653, 2017.

[15] DA Tobon-Mejia, Kamal Medjaher, and Noureddine Zerhouni. Cnc machine tool's wear diagnostic and prognostic by using dynamic bayesian networks. Mechanical Systems and Signal Processing, 28:167–182, 2012.

[16] Jurgen Schmidhuber.¨ Deep learning in neural networks: An overview. Neural networks, 61:85–117, 2015.

[17] Li Deng. A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA Transactions on Signal and Information Processing, 3, 2014.

[18] Feng Jia, Yaguo Lei, Jing Lin, Xin Zhou, and Na Lu. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mechanical Systems and Signal Processing, 72:303–315, 2016.

[19] Luyang Jing, Ming Zhao, Pin Li, and Xiaoqiang Xu. A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. Measurement, 111:1–10, 2017.

[20] Rui Zhao, Ruqiang Yan, Zhenghua Chen, Kezhi Mao, Peng Wang, and Robert X Gao. Deep learning and its applications to machine health monitoring: A survey. arXiv preprint arXiv:1612.07640, 2016.

[21] Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. arXiv preprint arXiv:1511.03677, 2015.

[22] Franc¸ois Chollet et al. Keras, 2015.

[23] Mart´ın Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467, 2016.

Signal processing

Autonomous Embedded Vibroacoustic Measurements: an efficient tool for railway monitoring

Romain AUGEZ[1], Aurélien CLOIX[1], Christian CLERC[1]

[1] VIBRATEC, 28 chemin du Petit Bois, 69130, ECULLY (France)
Romain.augez@vibratec.fr

Abstract

Efficient maintenance and monitoring are key points for rolling stock or railway network operators. A good knowledge of the structural health and a capability to predict evolution are main assets to ensure a high level of performance. Vibrations and dynamic forces borne by the wheel-rail contact contain the main information to reach this objective. Therefore, mechanical health features can be deduced from the signal measured on the bogie or on the rail using appropriate data processing algorithms. Moreover, the improvements in electronics and communication technologies make it possible to integrate measurement and data processing functions in a compact embedded system.

VibraTec's wheel-rail contact experience leads to consider an indirect measurement: a sensor mounted on the bogie to monitor the rail and/or a sensor mounted on the rail to monitor the rolling stock. Upon this assumption, VibraTec developed a new tool able to detect defaults and assess their evolution. To reach the objectives, the device had to be able to measure vibrations, acoustics, train speed and location, then to process and send the data. A key point is to apply dedicated algorithms developed to identify and quantify defaults from different origins operating in time domain or in frequency domain.

1 Introduction

In order to increase maintenance efficiency in the rail industry, several condition monitoring solutions have been developed over the past few years. Rolling stock or railway network operators need real-time accurate information about the structural health of their equipment. Early damage detection is a powerful tool for maintenance management: avoid customer complaints, save equipment from hard cracks, limit repair and operation times, and consequently, reduce global maintenance cost.

This paper focuses on the rail corrugation monitoring system developed in the scope of the MAVICO project. This prototype solution dedicated to railway maintenance management teams was developed in order to prevent rail corrugation effects on the track and on rolling stock equipment.

The development of communication tools and electronics components, the increase in computation capabilities, the enhancement of MEMS sensors make it possible to design an efficient Proof Of Concept (POC) for corrugation analysis. This embedded system provides timely information about the state of the rail infrastructure.

This paper begins by introducing the theoretical background of the methodology: from rail roughness to bogie (or axle box) acceleration. Then, the embedded monitoring system and its calibration process are presented. Finally, results deduced from measurements carried out on a tramway bogie running on an urban network are given. The measurement campaign was performed in partnership with Keolis Lyon, involved in the MAVICO project. The analysis focused on rail corrugation, quantified by its roughness and wavelength.

2 Theoretical background

2.1 From corrugation to bogie acceleration

During the rolling of a railway vehicle, the vertical acceleration on the bogie or the axle box may be considered, in first order, as proportional to the amplitude level of the surface defects on the rail. This consideration is true if the wheels do not have any high-amplitude default (new or recently reprofiled wheels).

The physical phenomena involving the bogie acceleration are shown on the following equations:

Step 1: Dynamic force at rail/wheel contact,

$$F(f) = Z(f).[A_w(f)+ A_r(f) + A_c(f)]^{-1} \quad (1)$$

With:
F(f), the dynamic force at contact point,
Z(f) the rail vertical defect,
A_w, A_r, A_c respectively the admittance of the wheel, the rail, and the contact,
f, the frequency dependence.

Step 2: Bogie (axle box) acceleration

$$\Upsilon^2(f)= F(f).A_{cb}(f) \quad (2)$$

With:
$\Upsilon(f)$ the acceleration on the axle box,
$A_{cb}(f)$ the transfer function between the rail and the bogie (or axle box)

2.2 Step 1 : from Rail vertical defect to Dynamic force

The vertical dynamic force spectrum at the contact generated by the defect is deduced from the product of the vertical defect spectrum of the rail *F(f)* with the transfer function $[A_w(f) + A_r(f) + A_c(f)]^{-1}$ with:
the vertical receptance of the path seen from the contact $A_r(f)$,
the vertical receptance of the vehicle seen from the contact $A_w(f)$,
the receptance of the contact $A_c(f)$, defined as the inverse of the Hertz stiffness.

N.B. The receptance corresponds to the dynamic flexibility; it is defined as the ratio between the vibration amplitude *Z(f)* at the excitation point and the applied dynamic force *F(f)*.

These three receptances are available by calculation or/and by measurement; VibraTec has developed GroundVib software which calculates them, using the track and material properties. This calculation approach makes it possible to take into account the impact of track design in the transfer function between the default and the acceleration on the vehicle.

This step is presented in section 3 of this paper.

2.3 Step 2: from dynamic force to bogie acceleration

The level of acceleration on the bogie is deduced from the contact force spectrum using the transfer function $A_{cb}(f)$. This transfer function can be calculated from the ratio between:

- the roughness Z (1/ λ) measured by the corrugation measurement trolley (e.g. Figure 1) on a portion of the track, and
- the vertical acceleration recorded on the bogie ϒ(f), in running conditions at the speed V of the train on the same portion of track.

Figure 1: rail corrugation analysis trolley

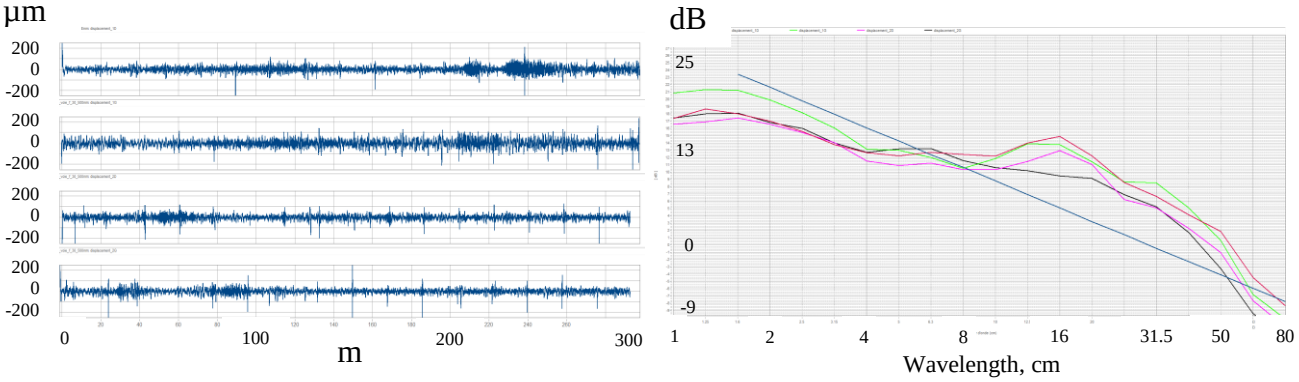


Figure 2: Left side: roughness in µm measured on 4 track lines.
Right side: Associated roughness spectra (1/m) of a portion of track

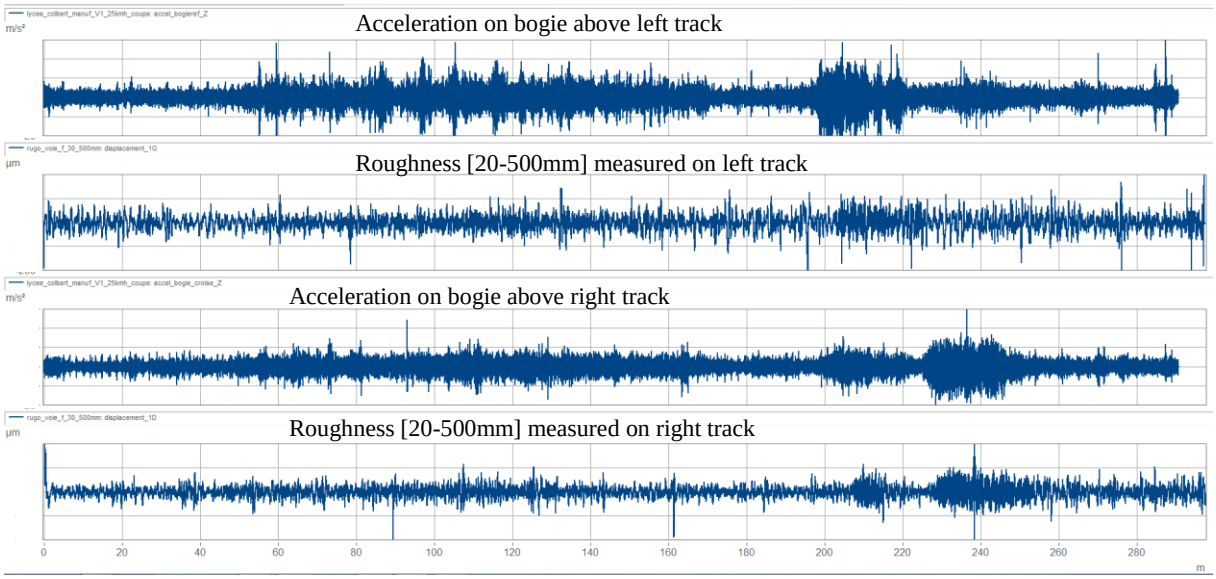


Figure 3: Accelerations measured on a bogie and roughness on the same portion of track at 25 km/h

A post-processing of these two measurements makes it possible to compute the transfer function $\Upsilon^2(f)/ Z^2(f)$, by a conversion of the roughness abscissa '1 / λ' to a frequency abscissa 'f' using the train speed ‘v’.

3 Practice: From bogie acceleration spectrum to rail corrugation

The methodology for determining the track roughness from the acceleration measured on the bogie is shown in the equation (3).

Part of the input is the transfer function $[\Upsilon^2(f)/ Z^2(f)]$ established in paragraph 2 of the preceding process. The reference transfer function, estimated on one or more reference sections, is inverted to obtain the transition function between the bogie acceleration and the track roughness: $[\Upsilon 0^2(f)/ Z0^2(f)]^{-1}$.The roughness PSD is then directly computed using the product of the acceleration with the transition function.

$$Z^2(f) = [\Upsilon_0^2(f)/ Z_0^2(f)]^{-1}. \Upsilon^2(f) \quad (3)$$

With:
$\Upsilon^2(f)$ the bogie acceleration PSD
$Z^2(f)$ the rail roughness PSD

Results can be evaluated by comparing the calculated roughness with the direct roughness measured by the trolley on a section of track which was not used in the transfer function definition. Figure 4 presents this kind of comparison for a 50m section.

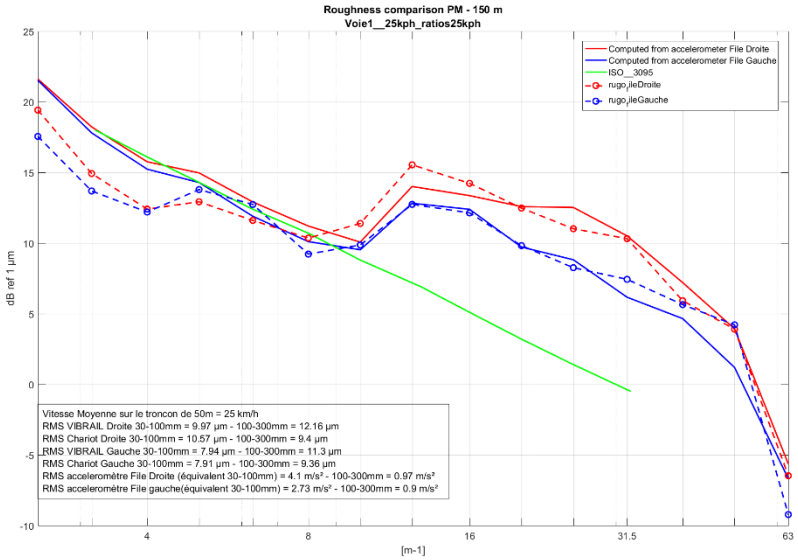


Figure 4 : 1 Third octave spectra (in 1/m) measured (with trolley in dashed lines) and computed (with inverse method in solid lines). In green, the ISO 3095 standard

4 Developed tools

A dedicated tool has been developed to monitor the bogie acceleration, the acoustic pressure in the bogie area, the speed and localisation of the train on the network.

The device contains:

-A data logger that records analogic channels and stores high-speed (acoustic pressure, vibration, and speed tachometer), and low-speed sampling rate signals such as GPS data,

-A 4G router that transmits the data from the train to VibraTec servers,

-A battery to ensure the transfer of data during train electrical power outages.

The sensors used are a mix between common sensors used for vibration investigations and new MEM's technology. The global approach for the complete processing, from the sensors' raw data to the corrugation defect alert is shown in the Figure 5.

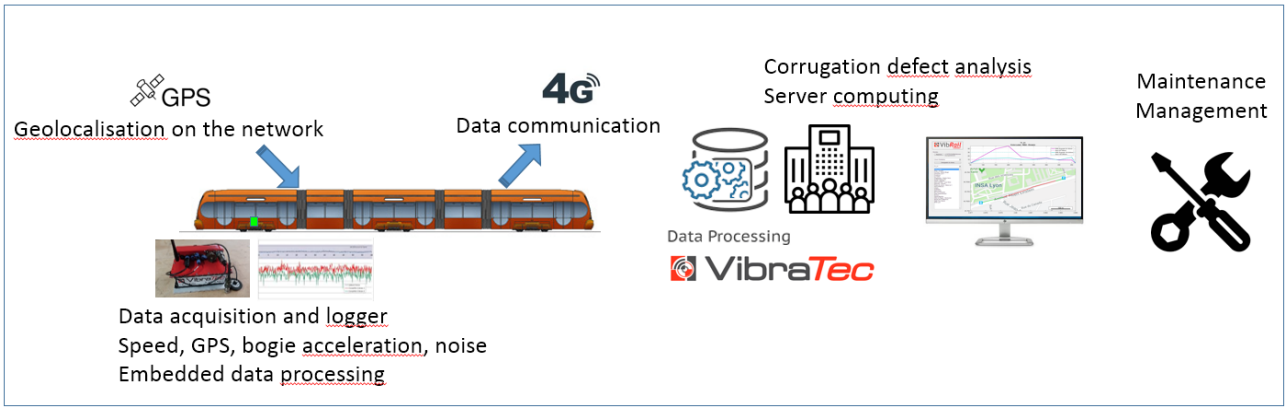


Figure 5: Global approach – from raw data to corrugation analysis

Figure 6 presents the VibRail Proof Of Concept. The tool is installed and fixed directly, wireless, on the tramway bogie, without impacting passengers, e.g. Figure 8.

Figure 6: VibRail Proof Of Concept n°2

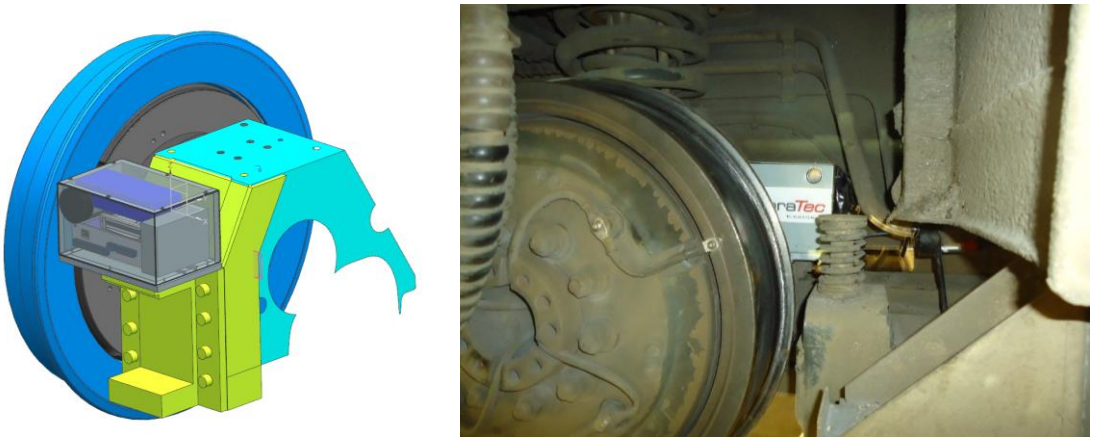


Figure 7: localisation of VibRail Concept n°1 on a tramway bogie

Examples of raw data are presented in Figure 8 and Figure 9.

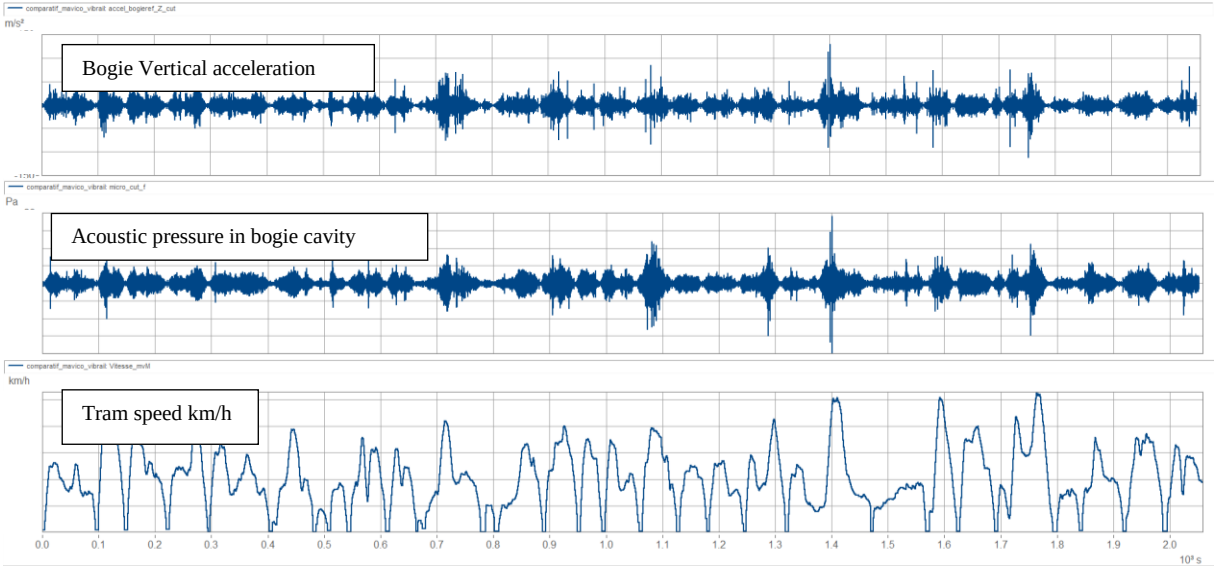


Figure 8: raw data measured by the device on the Lyon tramway network T1 Line

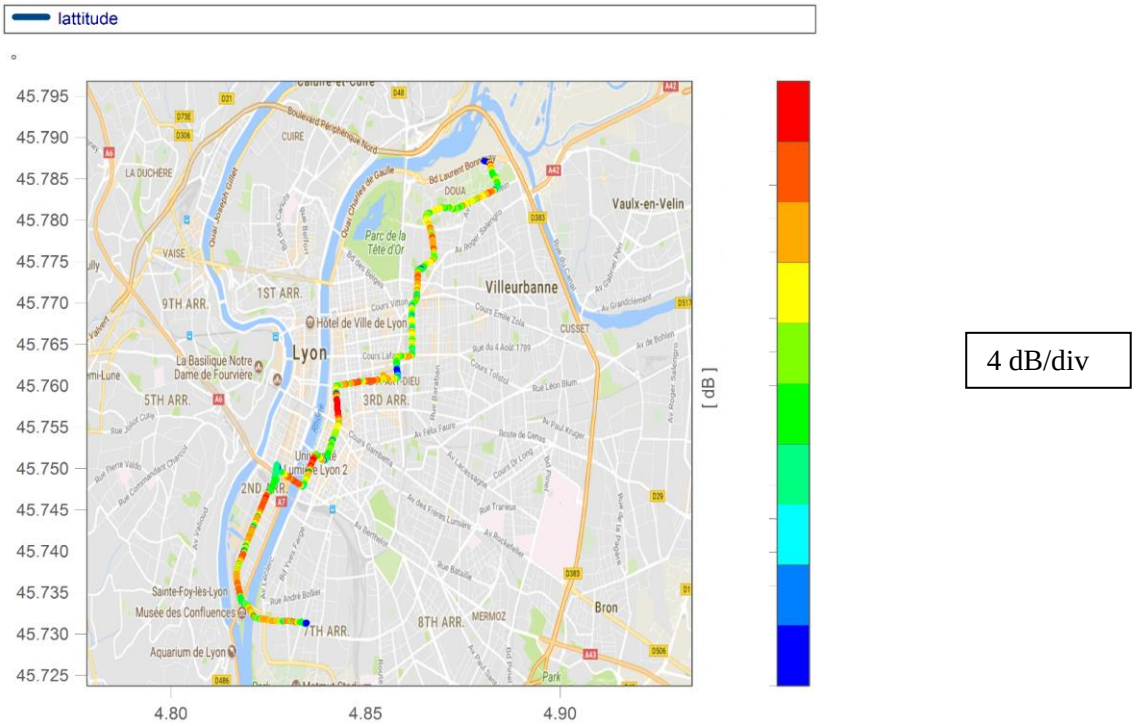


Figure 9: Lyon tramway network T1 Line. GSP track, with a color scale corresponding to RMS acoustic pressure under the bogie in dB

NB: The acoustic pressure under the bogie was monitored to detect screeching noise and send an alarm to the tramway operator to avoid disturbing the neighborhood. The principle of detection based on wheel mode resonance is not developed in this paper.

5 Measurement results

In the scope of the MAVICO research project, 3 online measurement campaigns were carried out, spaced about 6 months apart, on the full tramway network in Lyon.

Each measurement campaign was used to determine the roughness of the network tracks. Having 3 campaigns spaced in time made it possible to assess the evolution of the roughness from both operation and grinding campaigns.

5.1 Identification and quantification of Rail roughness

The wavelengths taken into account on tramway tracks are in the [30-300]mm range. The relation between excited frequencies and train speeds is presented in Table 1.

Speed	20 km/h	40 km/h	50 km/h
Frequency band	18.5 Hz – 185 Hz	37 Hz – 370 Hz	46.3 Hz – 463 Hz

Table 1: relation between excited frequencies and train speed for wavelengths from 30mm to 300mm

The people who are in charge of track surveys select the areas to be ground by riding the network aboard the tram and listening to the emergence of noise related to corrugation. A first part of the data processing was to analyse and compare the online measurement data to human perception, in order to establish a correlation between human detection and accelerations measured on the bogie. Figure 10 presents an example of this correlation.

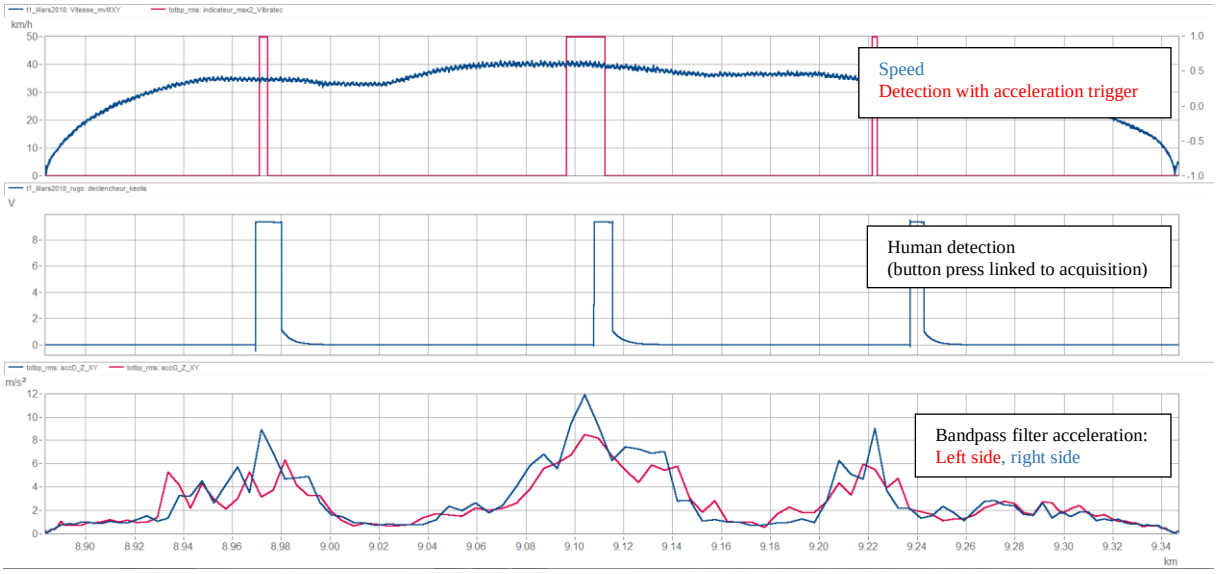


Figure 10: correlation between bandpass filtered acceleration and human detection on an interstation

In this example, a trigger was activated when the bandpass filtered RMS acceleration reached 8m/s². The correlation with human detection is satisfying and the defect localisation is more accurate than the human detection thanks to the GPS signal acquisition. One of the conclusions of this correlation with human perception is that the wavelengths involved are in the [30-100]mm range.

The second phase was to compute the absolute value of the roughness thanks to the measured accelerations and the transfer functions (equation (3)). For this purpose, algorithms have been developed to:

- locate the train on the network: line number, train direction,
- split the signals according to the interstations,
- calculate the roughness RMS level for each interstation. This level is computed versus the distance, with a sampling of 20m,
- plot this criterion on GPS maps and generate reports that allow the customer to communicate with the team in charge of grinding and/or to evaluate the actual roughness of the network.

An example of a GPS map is presented in Figure 11.

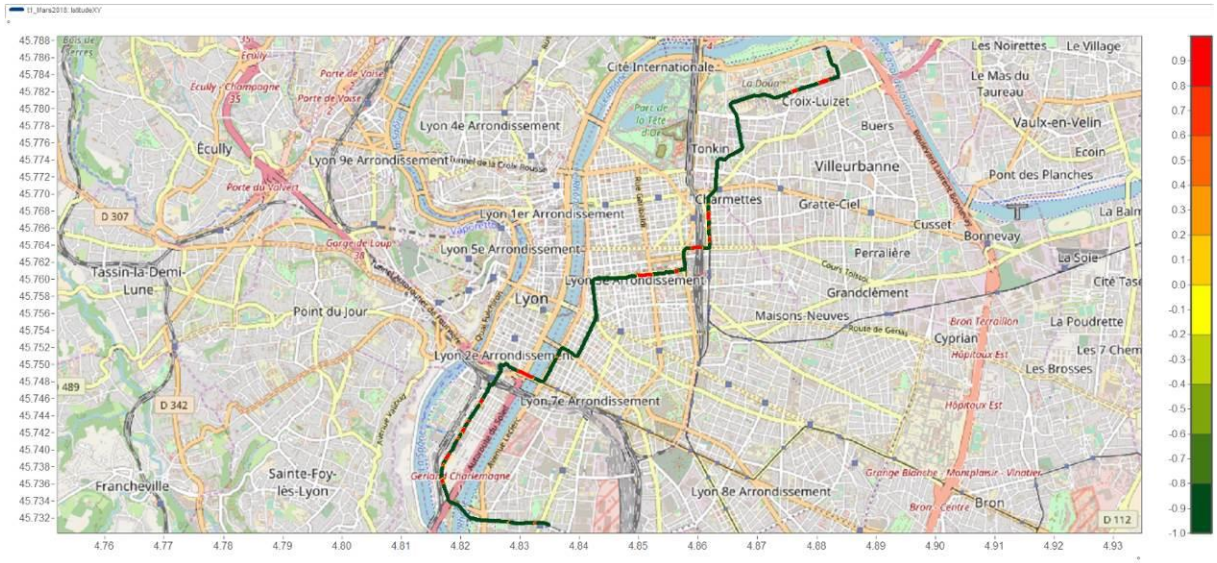


Figure 11: Example of a GPS map of the roughness level (Lyon tramway network T1 Line).

The criteria used to determine the zones to grind can be defined from the literature. In the [30-150]mm wavelength range, a criterion of 10 µm RMS can be implemented [1]. This limit can be tuned according to the corrugation expert's sensitivity, and also on the grinding budget.

Figure 12 presents the RMS roughness levels in the [30-150]mm wavelength range calculated related to the distance (for example, the starting point 0 km is the beginning of the line). The RMS level was only computed when the speed of tram was sufficient to have a good acceleration signal/noise ratio. Below this speed, the RMS was set at zero.

The red line represents the limit/criterion tuned in accordance with Keolis' corrugation expert, to fit with their detection zone (presented below).

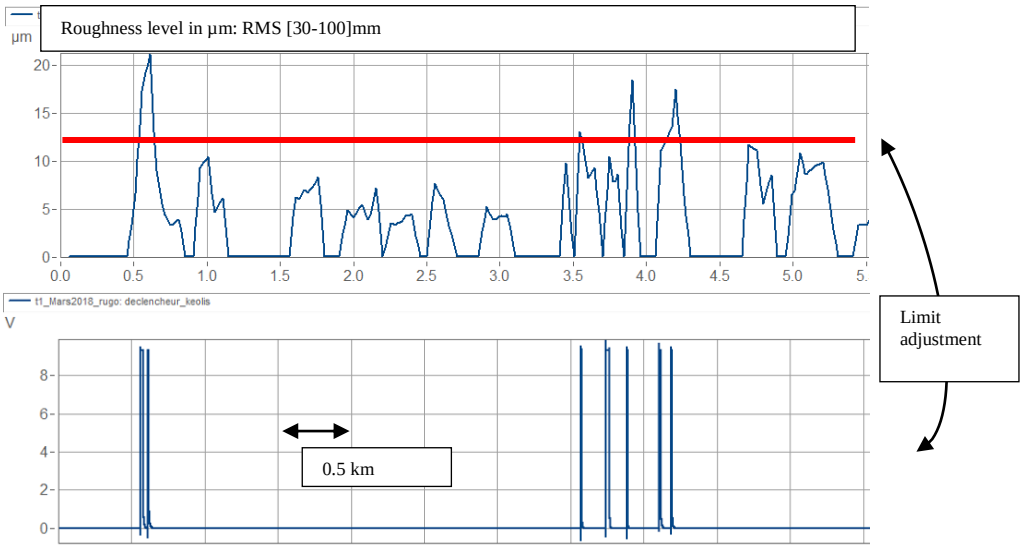


Figure 12: Top: Rail roughness - RMS level in the [30-150]mm wavelength. Bottom: Correlation with subjective detection

5.2 Interest of regular monitoring of rail roughness

Usually, grinding campaigns are performed each year. However, there is a great interest in monitoring the rail roughness more often: impact of grinding, impact of exploitation.

Figure 13 presents the rail roughness measured on the same interstation at different times. The initial roughness is presented in blue (time t_0). The roughness just after a grinding campaign (time t_1 = 8 months after t_0) is presented in red. The third roughness, 4 months after the grinding campaign, is presented in green (t_2 = 4 months after t_1).

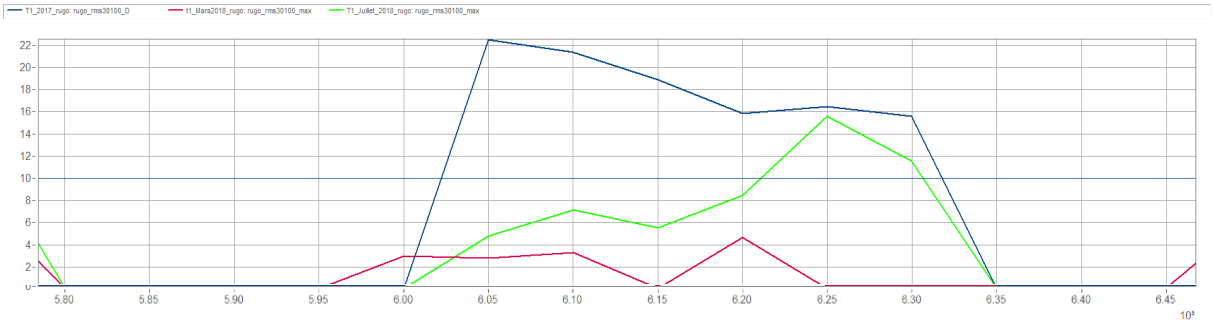


Figure 13: rail roughness - RMS level in the [30-150]mm wavelength. Status July 2017, Status March 2018, Status July 2018

In this example, the grinding campaign reduced the roughness from 22µm to 4µm RMS [30-150]mm, in the first 100m of the interstation.

The comparison of the red and green curves shows that the roughness has increased by a factor 2 in 4 months.

6 Summary/Conclusions

In the scope of the MAVICO project, a POC of an embedded monitoring system has been designed in order to detect corrugation defect on tracks. The developed algorithms are based on an inverse approach, using bogie acceleration measurements to determine the state of corrugation. This system has been assessed on the complete Keolis Lyon tramway network. This tests have confirmed the interest in the roughness level as a first indicator of corrugation defects. Moreover, steady corrugation controls are interesting for track monitoring and for early damage detection.

On dedicated track areas used for validation, the estimation of the roughness RMS level based on the indirect method is satisfying, in comparison to the classical direct measurement method using a trolley. The system described in this paper can estimate the roughness level of a complete tramway network of more than 100km in one day. In comparison, such measurements using the direct method would take several weeks.

The corrugation threshold alert has been tuned using the correlation with subjective human detection of corrugation defects. This corrugation alert described in this paper is a powerful indicator for track maintenance management.

Through a new collaborative project (MEEQUAI), an improvement of the rail-to-bogie transfer function, based on a calculation/measurement hybrid approach, is currently being studied.

Acknowledgment

The results presented in this paper have been obtained within the scope of the MAVICO project. The authors wish to thank Keolis Lyon and Sytral for their involvement in the project. MAVICO is a collaborative project supported by the ADEME through the framework programme “PIA-FER15-Transport ferroviaire”.

References

[1] S. Grassie, *A practical methodology to prioritise reprofiling sites for corrugation removal– 11th international conference on contact mechanics and wear of rail/wheel systems (CM2018)*, Delft, The netherlands, 2018 September 24-27, p 329.

Blind vibration filtering using envelope linear prediction for fault detection without knowledge of machine kinematics

Cédric Peeters[1], Jérôme Antoni[2], Jan Helsen[1]

[1]Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
cedric.peeters@vub.be

[2]LVA, INSA de Lyon,
F-69621, Villeurbanne cedex, France

Abstract

The central idea behind this paper is to propose a means to filter out vibration signals of interest from a fault detection perspective without actually having knowledge about the kinematics of the machine. In other words, this paper investigates blind filters that do not require a-priori knowledge about the fault frequencies, e.g. of a bearing or gear. This kind of approach opens the door for the condition monitoring of complex machines where insufficient information is available about the inner components or where replacements have been carried out that changed characteristic frequencies and that were not logged. This feat is achieved by employing the squared envelope as a metric for the blind filter. The main assumption of the proposed method is that when a fault occurs, it introduces a second-order cyclostationary (CS2) component in the vibration signal which manifests itself in the squared envelope (SE) as a harmonic sine modulation at its corresponding fault frequency. This modulation correspondingly also increases the sparsity of the envelope spectrum. To avoid interfering influences of CS1 components, the signal is typically pre-whitened, e.g. through linear prediction filtering, cepstrum editing, etc. The paper investigates the minimization of the relative prediction error of the linear prediction of the squared envelope for use in the iterative updating procedure of the blind filter.

Keywords

Blind filtering, vibrations, linear prediction, sparsity, envelope, fault detection

1 Introduction

Complex machines nowadays can consist of dozens of bearings and gears, with modern examples being the gearboxes of wind turbines and helicopters. These gearboxes typically have one or multiple planetary gear stages in combination with parallel gear stages. Not all kinematic information about the system might be available to the machine operator, or the information might be inaccurate due to reparations with new components. This issue constitutes the need for a method capable of tracking the condition of these components without the need for a-priori knowledge about the kinematics.

One of the most popular approaches for fault detection is to look at the cyclostationary behavior of the vibration signal [1, 2, 3, 4, 5, 6, 7, 8, 9]. Inspired by this fact, this paper investigates the possibility to utilize the cyclostationary content of a signal in a blind manner. Therefore, instead of just looking at the statistics of the time waveform, the squared envelope of the signal is employed as a means to gain more information about potential defects. From experience it is known that most mechanical faults of bearings or gears induce some form of cyclostationary behavior in the observed vibration signals [4, 10, 11]. This cyclostationary behavior alters the modulation signature of the signal. For example a repetitive impulse train (similar to a bearing fault) introduces harmonics at the repetition frequency of the impulses into the envelope. This means that the envelope signal becomes more predictable and can thus be fitted with an autoregressive model. This property is thus exploited to find a filter that minimizes the relative prediction error of the squared envelope since it is assumed that a good fit corresponds to a mechanical fault and not to normal behavior.

An important remark about the proposed approach is that the blind filtering methodology described in this paper cannot be categorized as blind deconvolution, blind signal separation, or denoising. The proposed approach namely does not attempt to deconvolve the signal in order to recover the source signal (e.g. impulses), nor does it attempt to separate signals from a mixture or remove noise from the signal without distorting it. In fact, it actually does distort the signal such that the squared envelope is as predictable as possible. This is a fairly new concept since from this perspective the algorithm does not care about restoring the signal or recovering the source signals. Instead the algorithm just tries to maximize the figure of merit and thus enhance the envelope spectrum. It is important to take into account this distinction when inspecting filtering results since the results might not correspond to what is expected.

This paper attempts to highlight the utility of blind filtering based on the signal envelope and the versatility of the Rayleigh quotient regarding the indicator choice for the blind filtering step. First, the theoretical background is explained in Section 2. The indicator choice and the derivation of the Rayleigh quotients are described. Next, the method is validated on simulated signals in Section 3 and experimental data of a gearbox data set in Section 4. The results show that the proposed approach is capable of extracting a cyclostationary fault signature and that the prediction error measure of the envelope in itself can be used as a tracking parameter.

2 Methodology

The idea of the proposed methodology is to exploit the predictability of a fault modulation signature by trying to fit it with a linear prediction (LP) filter or auto-regressive (AR) filter. The prediction error of an auto-regressive all-poles model of the squared envelope serves then as the metric of interest. The better the AR model can fit the actual envelope, the more predictable and thus the less noisy it is. This means that if there is a signal component present with e.g. a clean sinusoidal amplitude modulation, the AR model is then capable of predicting future samples accurately which in turn corresponds to a low prediction error. This does indicate again the need for prewhitening the signal to make sure the AR model does not try to fit the envelope of deterministic components in the signal.

2.1 Blind filtering

The concept of blind filtering is to find a filter that maximizes a certain criterion of the signal starting from a noisy measured signal $\mathbf{x}$:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} \tag{1}$$

where $\mathbf{s}$ is the estimated input, $\mathbf{h}$ is the inverse filter, and $*$ refers to the convolution operation. It should be noted that vectors and matrices are set in bold font to illustrate the difference with scalars. The convolution is expressed as:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \tag{2}$$

$$\begin{bmatrix} s_{N-1} \\ \vdots \\ x_{L-1} \end{bmatrix} = \begin{bmatrix} x_{N-1} & \dots & x_0 \\ \vdots & \ddots & \vdots \\ x_{L-1} & \dots & x_{L-N-2} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

with L and N the number of samples of $\mathbf{s}$ and $\mathbf{h}$ respectively.

Now the squared envelope $\varepsilon_{\mathbf{x}}$ can be defined as follows:

$$\varepsilon_{\mathbf{x}} = |\mathbf{s}|^2 = |\mathbf{X}\mathbf{h}|^2 \tag{3}$$

It can also be written as:

$$\varepsilon_{\mathbf{x}} = \begin{bmatrix} s_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{L-N+1} \end{bmatrix}^H \mathbf{X}\mathbf{h} = diag(\mathbf{s}^{\mathbf{H}})\mathbf{X}\mathbf{h} \tag{4}$$

with $\mathbf{s}^{\mathbf{H}}$ being the Hermitian transpose of $\mathbf{s}$, and $diag(\mathbf{s}^{\mathbf{H}})$ being a diagonal matrix with the values of the vector $\mathbf{s}^{\mathbf{H}}$ on its diagonal.

Figure 1: Linear prediction coding.

2.2 Derivation of LP-envelope filter

The relative prediction error of the AR model is closely related to the spectral flatness as the AR model also maximizes the spectral flatness of the squared envelope prediction error [12]. The relative prediction error (RPE) of the AR model of the squared envelope is given by:

$$RPE = \frac{\sigma_e^2}{\sigma_{SE}^2} \tag{5}$$

with σ_e being the prediction error of the squared envelope, and σ_{SE} being the standard deviation of the squared envelope. The autoregressive coefficients can be obtained by fitting a linear prediction model on the squared envelope. The standard LPC representation of a signal $x(n)$ for a model of order N is:

$$x(n) = \sum_{i=1}^{N} a_i x(n-i) + e(n) \tag{6}$$

with a_i the autoregressive coefficients, and where the error $e(n)$ can be obtained by:

$$e(n) = \sum_{i=0}^{N} a_i x(n-i) \; with \; a_0 = 1. \tag{7}$$

Figure 1 illustrates the straightforward process. The autoregressive filtering necessary to obtain the prediction error $\mathbf{e}$ of the squared envelope ε_x can be written in matrix notation as follows:

$$\mathbf{e} = \mathbf{A} diag(\mathbf{s^H}) \mathbf{s} \tag{8}$$

with $\mathbf{A}$ a band matrix containing the autoregressive coefficients as follows:

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & \dots & a_N & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_N & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_0 & a_1 & \dots & a_N \\ & & & & \ddots & \ddots & \vdots \\ & & & & & a_0 & a_1 \\ 0 & & & \dots & & 0 & a_0 \end{bmatrix} \tag{9}$$

Since the intent is actually to minimize the relative prediction error in Eq. 5 and not maximize it, the ratio to be maximized is actually the inverted RPE. Equation 5 can thus be written in the following manner:

$$\boxed{\frac{\sigma_{SE}^2}{\sigma_e^2} = \frac{{\varepsilon_\mathbf{x}}^H \varepsilon_\mathbf{x}}{\mathbf{e}^H \mathbf{e}} = \frac{\mathbf{h^H X^H} diag(\mathbf{s}) diag(\mathbf{s^H}) \mathbf{Xh}}{\mathbf{h^H X^H} diag(\mathbf{s}) \mathbf{A}^H \mathbf{A} diag(\mathbf{s^H}) \mathbf{Xh}} = \frac{\mathbf{h^H R}_{XW_1X} \mathbf{h}}{\mathbf{h^H R}_{XW_2X} \mathbf{h}}} \tag{10}$$

The generalized Rayleigh quotient [13] can be recognized in Eq. 10 and can be maximized using an iterative maximization of the eigenvalues:

$$\lambda = \frac{\mathbf{h^H R}_{XW_1X} \mathbf{h}}{\mathbf{h^H R}_{XW_2X} \mathbf{h}} \tag{11}$$

The Rayleigh quotient has the interesting property that its maximal value with respect to $\mathbf{h}$ is equivalent to its largest eigenvalue λ and corresponding eigenvector. Thus, maximizing the Rayleigh quotient allows finding the maximal values of the corresponding indicator and filter. In order to obtain real eigenvalues however, the

correlation matrices $\mathbf{R}_{XW_1X}$ and $\mathbf{R}_{XW_2X}$ need to be Hermitian, and $\mathbf{R}_{XW_2X}$ needs to be positive semidefinite. If these conditions are met, the Rayleigh quotient offers an efficient means to calculate iteratively the filter coefficients. Only the largest eigenvalue and corresponding eigenvector need to be computed in each iteration, which can be achieved efficiently by using algorithms such as the power method [14].

The generalized eigenvalue problem to be solved can be formulated as such:

$$\mathbf{R}_{XW_1X}\mathbf{h} = \mathbf{R}_{XW_2X}\mathbf{h}\lambda \tag{12}$$

The iterative algorithm used to minimize the prediction error consists out of four basic steps:

1. Assume an initial guess for $\mathbf{h}$
2. Estimate $\mathbf{R}_{XW_1X}$ and $\mathbf{R}_{XW_2X}$ based on $\mathbf{h}$ and $\mathbf{X}$ using Eq. 10
3. Solve Eq. 12 to find λ_{max} and a new filter $\mathbf{h}$ that corresponds to a higher value of the used criterion
4. Return to step 2 using the new $\mathbf{h}$ until convergence is reached or the maximum number of iterations

The name of the proposed method is abbreviated to LPE (Linear Prediction of Envelope) to keep the text concise.

3 Simulation

To validate the proposed approach, a straightforward simulated case is first considered. To add some point of reference, the performance of the proposed LPE method is compared to Minimum Entropy Deconvolution (MED), which is a commonly used blind deconvolution filtering technique to maximize the kurtosis. An outer race bearing fault signal is simulated with a normalized sample rate of 1 Hz and duration of 20000 samples. White Gaussian noise is added at varying degrees of signal-to-noise ratio. The parameters used for the simulation of the bearing fault signal:

- the outer race bearing fault frequency, $f_{BPFO} = 0.31Hz$
- the resonance frequency excited by the fault, $f_{IRF} = 0.25Hz$
- the damping ratio, $\zeta = 0.1$
- the jitter, $J = 1\%$

The signal-to-noise ratio of the fault signal is varied linearly from -30 dB to + 15 dB. The overall variance of the signal is kept constant however. The filter length is chosen to be 15 and 150 samples respectively for the proposed LPE method and MED filtering. The max number of iterations is set at 100.

Figure 2 shows the evolution of the indicators after blind filtering using the proposed method and MED. It can be observed that the relative prediction error starts to decrease gradually around -17 dB SNR. In this case the proposed method outperforms MED mainly because the generated bearing fault signal does not immediately give rise to a large increase in the kurtosis of the signal. This is due to the fault frequency being relatively high and therefore causing the exponentially decaying impulse responses to smear over. This is not an ideal scenario for kurtosis maximization. However, when inspecting the envelope, the fault frequency modulation shows up faster because it does not suffer as much from this smearing. Instead it detects still the envelope fluctuation at exactly the fault frequency, albeit less pronounced than it would be without smearing.

Figure 3 shows a color map in grayscale of the evolution of the normalized squared envelope spectra. The fault frequency of 0.31 Hz can be clearly distinguished after approximately -17 dB and 2 dB SNR for respectively the proposed and MED method. An alternative 3D view of this color map is provided in Figs. 4 and 5.

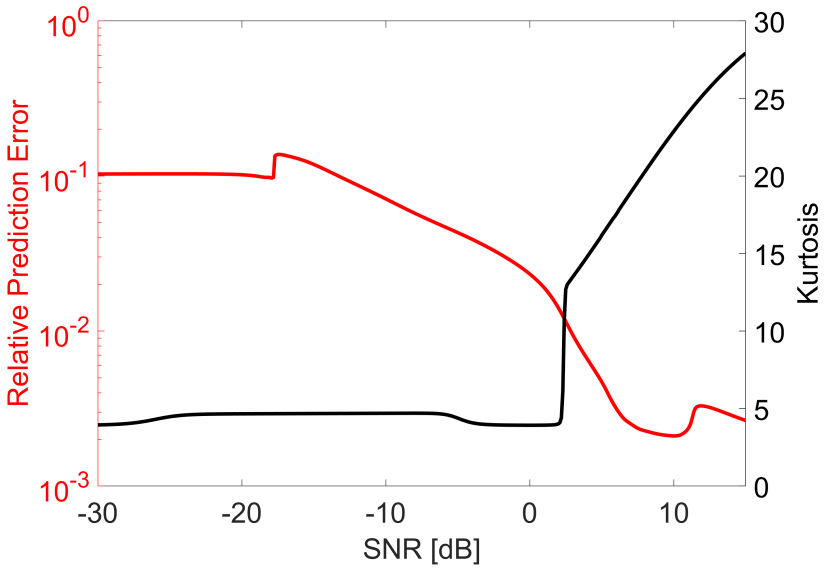


Figure 2: Trending of Relative Prediction Error (RPE) versus kurtosis after blind filtering using the proposed LPE approach and MED respectively.

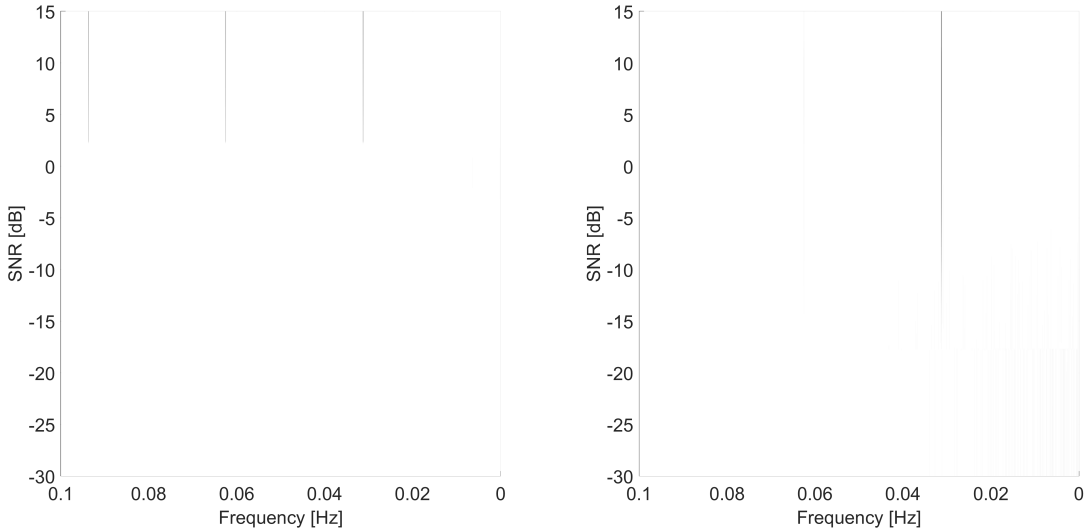


Figure 3: Color maps (in grayscale) of the squared envelope spectra after blind filtering using the MED (left) and proposed approach (right) respectively. The black lines appearing at approximately -17 dB and 2 dB SNR for respectively the proposed and MED method correspond to the 0.31 Hz fault frequency.

4 Experimental application

To verify whether the proposed method also works for more complex signals, the approach is tested on an experimental vibration signal. The real-world case chosen for this purpose is the IMS bearing prognostic dataset [15]. This dataset contains an outer race bearing fault in measurement campaign 2. The BPFO is approximately 236 Hz. The record chosen for filtering is n 690 since it is known that there is already damage distinguishable using other signal processing methods but the damage is not distinguishable directly from the raw vibration signal. The result after filtering using the LPE method is displayed in Fig. 6. While the original input signal does not show any clear signs of damage, the filtered signal evidently exhibits a strong modulation at exactly the BPFO of 236.1 Hz and its second harmonic at 472 Hz. This is a favorable result since no a-priori knowledge of the fault frequency was provided to method. It should also be noted that in testing the MED filtering method failed to extract any clear indication of the fault for the same signal.

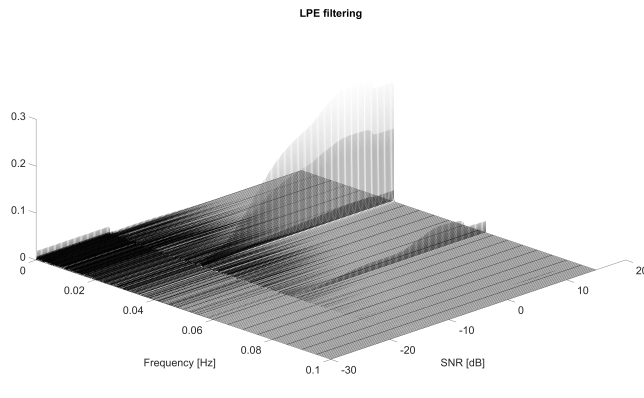


Figure 4: Squared envelope spectrum waterfall plot after filtering with the proposed method based on the relative prediction error.

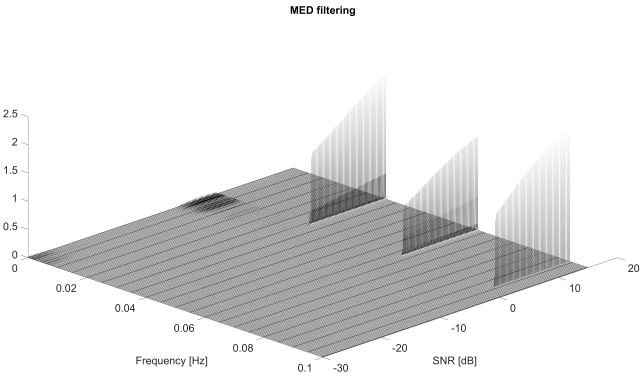


Figure 5: Squared envelope spectrum waterfall plot after filtering with the MED method based on kurtosis.

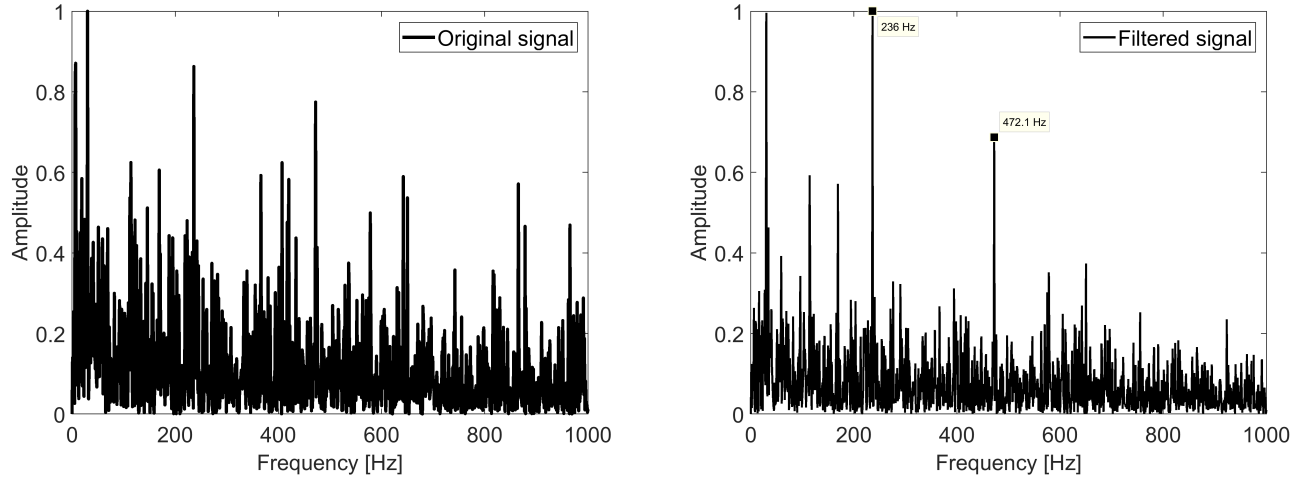


Figure 6: Experimental case: (Left) The envelope spectrum of the raw input signal. (Right) The envelope spectrum after filtering with the proposed LPE method.

5 Conclusions

This paper describes a novel fault detection approach to blind filtering of vibration signals. Instead of looking to maximize a statistic of the time waveform, it proposes to utilize the squared envelope and more in particular the predictability of it. Most fault signatures of bearing or gear faults induce structured and thus predictable second-order cyclostationary behavior. The main assumption is thus that an increase in the envelope predictability is linked to the emergence of a fault. This knowledge is then used to derive a blind filtering approach employing a generalized Rayleigh quotient iteratively to optimize the filter coefficients. The proposed method is validated on both simulation data as experimental data. Both scenarios prove that the proposed approach can work in an efficient and accurate manner. Employing the envelope instead of the time waveform, as done by MED filtering, also alleviates some of the limitations the latter can have. An example of such a limitation is the fact that MED filtering assumes that the fault signal has a high kurtosis value due to its impulsiveness. In reality this is not always the case. Additionally, MED filtering has the tendency to deconvolve a single high amplitude peak due to the fact that kurtosis is maximized in such a scenario. This is unlikely for a rotating component where impulses are generated at every rotation. Therefore the use of the envelope signal is a logical way to assess the presence of faults in rotating machinery.

While there is no need to input the fault frequency in the method, there is still one parameter that can influence the outcome, namely the filter length. A general recommendation for the filter length is to try to keep it as short as possible. Not only does this reduce the computation time, but it also means that there are less filter coefficients that need to be optimized and thus less coefficients that vary. The updating procedure also becomes

more stable and fluctuates less due to numerical inaccuracies (e.g. high condition numbers of the matrices).

The results presented in this paper are promising as they prove that detailed information about the system of interest is not always necessary to still perform health tracking. It also opens the door for more and different approaches using blind filtering based on the envelope and its properties instead of using the raw time waveform.

References

[1] Zimroz R and Bartelmus W 2009 Gearbox condition estimation using cyclo-stationary properties of vibration signal *Key Engineering Materials* vol 413 (Trans Tech Publ) pp 471–478

[2] Antoni J, Daniere J and Guillet F 2002 *Journal of sound and vibration* **257** 815–837

[3] Randall R B 2011 *Vibration-based condition monitoring: industrial, aerospace and automotive applications* (John Wiley & Sons)

[4] Antoni J, Bonnardot F, Raad A and El Badaoui M 2004 *Mechanical systems and signal processing* **18** 1285–1314

[5] McCormick A and Nandi A 1998 *Mechanical systems and signal processing* **12** 225–242

[6] Antoniadis I and Glossiotis G 2001 *Journal of sound and vibration* **248** 829–845

[7] Capdessus C, Sidahmed M and Lacoume J 2000 *Mechanical systems and signal processing* **14** 371–385

[8] Borghesani P, Pennacchi P, Ricci R and Chatterton S 2013 *Mechanical Systems and Signal Processing* **40** 38–55

[9] Antoni J, Xin G and Hamzaoui N 2017 *Mechanical Systems and Signal Processing* **92** 248–277

[10] Randall R B, Antoni J and Chobsaard S 2001 *Mechanical systems and signal processing* **15** 945–962

[11] Randall R B and Antoni J 2011 *Mechanical Systems and Signal Processing* **25** 485–520

[12] Stoica P, Moses R L *et al.* 2005

[13] Hom R A and Johnson C R 1985 *Cambridge University Express*

[14] Mises R and Pollaczek-Geiringer H 1929 *ZAMM-Journal of Applied Mathematics and Mechanics/Zeitschrift für Angewandte Mathematik und Mechanik* **9** 152–164

[15] Lee J, Qiu H, Yu G and Lin J 2007 *Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames*

Vibration Representation in Time and Phase Domains, Applications to Aircraft Engines

Julien GRIFFATON[1], Jose L. GOMEZ[1]

[1]Safran Aircraft Engines, Rond-point René RAVAUD, 77550 Moissy-Cramayel, France

Abstract

While in operation, aircraft engines may be subjected to various severe mechanical events such as bird ingestion, blade separation, ice accretion, shaft unbalance, compressor stall or rotor-stator interaction. During the development of a new engine application, such phenomena are simulated in test cells in order to make sure that the engine will still operate safely. During such tests, accelerometers and strain gages are mounted on the engine cases and near the shaft bearings in order to measure loads and vibrations and in order to understand the engine behaviour. The engine dynamic behaviour during those phenomena may be fleeting or sustained, cyclic or asynchronous, transient or stationary. Usual analysis are performed in various domains such as time, frequencies and orders of a rotating shaft speed, depending on the nature of the vibration behaviour.

This paper describes a new kind of vibrations representation that considerably facilitate the interpretation of fleeting or sustained events when both time and phase location of a shaft are meaningful. In a first section of the paper, the representation is described with a simple simulation of a rotating shaft operating with variable angular speed. In a second section, the representation is applied on a real aircraft engine during tests with various engine behaviours. The last section presents a simulation of those different behaviours with a simple lumped-mass model.

1 Introduction

Aircraft engines may be subjected to various severe mechanical events such as bird ingestions, blade separation, ice accretion, shaft unbalance, compressor stall or rotor-stator interaction [1].

During the development of a new engine application, such phenomena are simulated in test cells in order to make sure that the engine will still operate safely.

Those phenomena may be transient or stationary, synchronous, almost synchronous, or not correlated with the rotating speed of a shaft. They may be caused by internal activity of the engine or by external aggression. In order to better understand the behaviour of the engine during those tests, the engine is equipped with acceleration or displacement sensors, strain gauges, tachometers, temperature and pressure sensors and high speed cameras. Those measurements are recorded with a sampling frequency in the range from 10 Hz to 100 kHz, depending on the sensors capabilities and on the frequency content that is expected to happen during the test.

The measurements are then analysed with different kinds of data processing tools in order to extract from the noise the relevant information that will facilitate the understanding of the phenomena. According to the type of events, several kinds of data analysis are appropriate. Frequency analysis, order analysis [2], wavelet analysis and image processing are standard tools that may be very helpful. Another very popular tool is simply the analysis of the time history of different measurements when they are synchronized. Such analyses may be performed after some data processing such as non-causal filtering in order to enhance the information hidden in the signal. Time history analysis can reveal correlation or causality between phenomena.

In this paper we present a new representation of time history where we add the dimension of angular position of a shaft in the display of the signals.

In a first section, the representation is described with a simple simulation of a rotating shaft operating with variable angular speed.

In a second section, the representation is applied on a real aircraft engine during tests with various engine behaviours. The relevance of the new representation compared to more usual representation is discussed.

The last section presents a simulation of those different behaviours with a simple model.

For sake of confidentiality, plots units are normalized. Also, experimental and model parameters are not shared throughout the paper.

2 Description of the time/phase representation with a simple simulation

In order to explain the transformation of a standard time history representation to a time/phase domain representation, we will use a simple simulation of a vibration synchronous with a shaft rotating speed. A vibration signal was simulated with the following hypotheses:

- A rotating speed N of the shaft that decreases linearly, and with an inversion of the direction of rotation,
- A vibration signal that is synchronous with the rotating speed,
- The amplitude of the vibration that is proportional to the square of the rotating speed.

A standard time history representation of this simulation is illustrated in figure 1. As can be seen on figure 1(b), the sign of the rotating speed is meaningful for the representation of the shaft angular position. A convention shall be used for the sign of the rotating speed, for instance it could be positive when the shaft rotates in the clockwise direction in the ‘forward looking aft position‘.

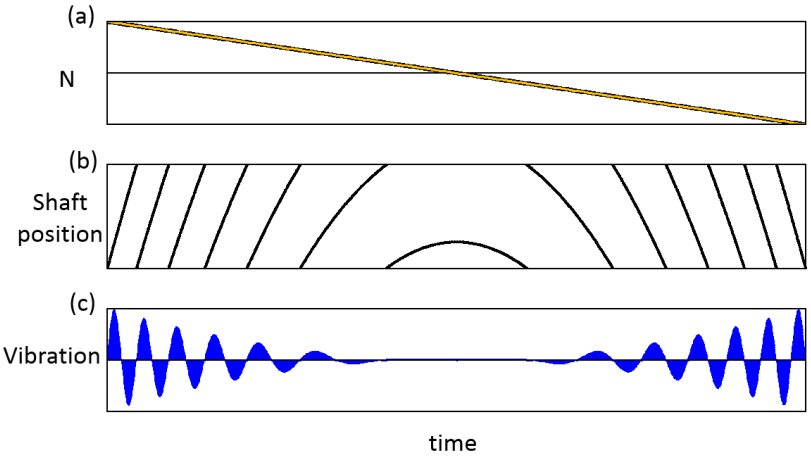


Figure 1: Standard time history representation with a simple simulation. Time history of the rotating speed N of a shaft (a), time history of the shaft angular position (b), and time history of a simulated vibration signal (c).

In a standard time history representation, the amplitude of the vibrations is plotted on the y-axis.

The representation in the time/phase domain of the simulated signal is shown in figure 2(a).

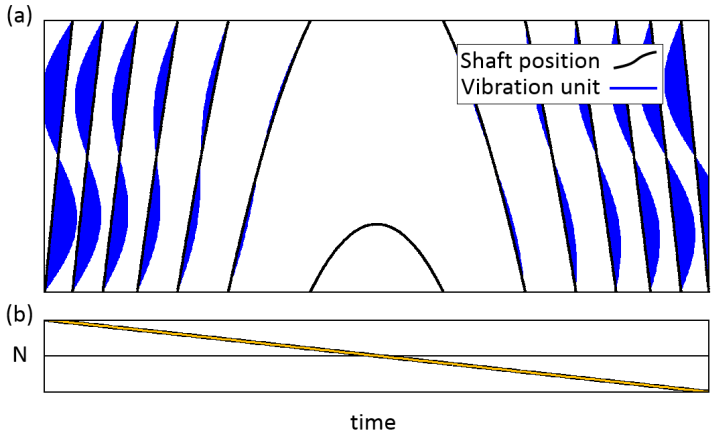


Figure 2: Representation in the time/phase domain (a) and time history of the shaft rotating speed (b).

This representation is composed of two signals:

- A plot of the time history of the shaft angular position,
- A plot of the vibration amplitude around the shaft angular position plot.

In such representation, the vibration is plotted in the direction of the x-axis. The zeros of the vibration amplitude are located on the curve of the shaft angular position.

A convention for the sign of the amplitude must be applied. As illustrated in figures 1 and 2, the convention used in this paper is to plot the positive values at the right of the shaft angular position, which is a natural convention. Another convention must be applied for the scale of the vibrations amplitude since the x-axis is also used for the representation of time. The scale of the vibration amplitude can be arbitrarily chosen, but it should be selected according to the range of amplitude and to the time length between two cycles for visual convenience.

Representation in the time/phase domain is therefore not a transformation of the signal, it is simply a different presentation of the time history of a signal. The aim of this kind of representation is to ease the identification and physical understanding of angular or temporal periodic phenomena present in the vibration signal.

In the next section, this representation will be applied to a variety of phenomena experienced by civil aircraft engines during engineering tests.

3 Application of the representation to several aircraft engine events

A first illustration in figure 3 is the installation of an oil unbalance. This event occurred while the rotating speed was almost stabilized as it can be seen on figure 3(c) and 3(d). The normalized rotating speed is plotted between 0 and max(N) while the detail is plotted between min(N) and max(N). In this case, the vibration signal is the displacement of the shaft in a given radial direction. This signal has been filtered with Vold-Kalman filters, described in [3], around 1 N and around 0.8 N.

The tracked signal around 1 N in figure 3(a) is representative of the displacement caused by a rigid unbalance of the shaft. This kind of unbalance is synchronous with the shaft. When the rotating speed is at steady state, the displacement caused by such unbalance also rotates synchronously with the shaft.

The tracked signal around 0.8 N in figure 3(b) is representative of an oil unbalance. Oil unbalance can occur when oil is present inside the shaft when a wave of oil is formed within. This wave causes an unbalance that is rotating in the direction of the shaft rotation but more slowly, hence the application of a filter around 0.8 N to track the displacement of the shaft caused by this type of unbalance. During this event, the rigid unbalance remained constant while the oil unbalance progressively increased in amplitude.

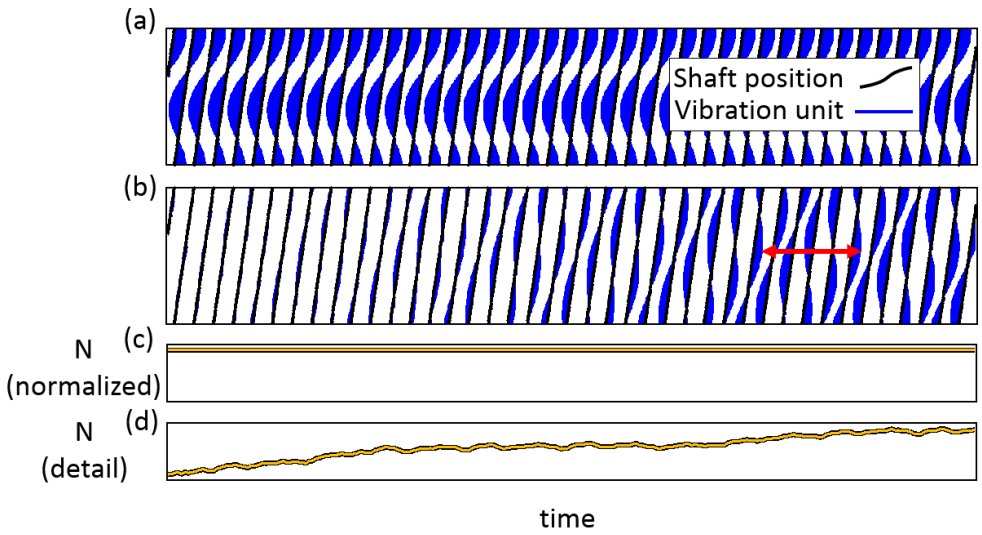


Figure 3: Application to a constant rigid unbalance and the progressive increase of an oil unbalance. Representation in the time/phase domain of the vibration amplitude tracked on the first order 1 N of the rotating speed (a) and of the vibration amplitude tracked around 0.8 N (b). Time history of the normalized rotating speed (c) and of the detail of the rotating speed (d). The red arrow shows that it takes 5 rotations of the shaft for the oil unbalance to be repetitive.

Another kind of test is ice accretion on the main shaft of aircraft engines. During ice accretion, several successive ice sheddings can occur. Those ice sheddings are fleeting events that change the unbalance of the shaft. Figure 4(a) shows an ice shedding and the high frequency content in the vibration during this event. Figure 4(b) is the vibration signal filtered around 1 N representative of the unbalance. In this case, the amplitude of the unbalance increases during the shedding. Figure 5 shows a different behaviour of an ice shedding. In this case, an unbalance was present prior to the shedding. After the shedding, the unbalance is higher and its angular location is at the opposite direction as can be seen in figure 5(b). Figure 4(d) and figure 5(d) show that the torsional mode of the shaft is excited by the sheddings.

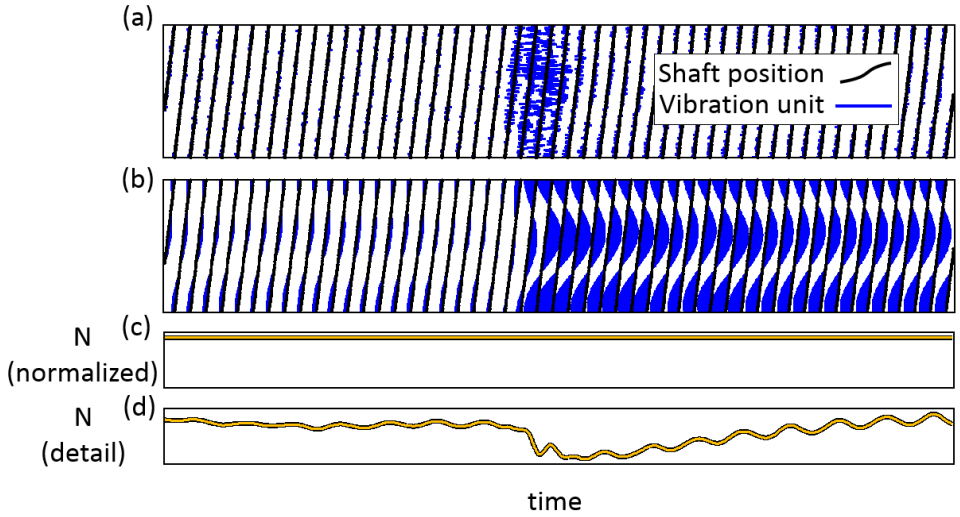


Figure 4: Application to ice shedding, fleeting event and sustained unbalance. Representation in the time/phase domain of the vibration amplitude (a) and of the vibration amplitude tracked around 1 N (b). Time history of the normalized rotating speed (c) and of the detail of the rotating speed (d). The unbalance increases in amplitude during the shedding while preserving its phase location.

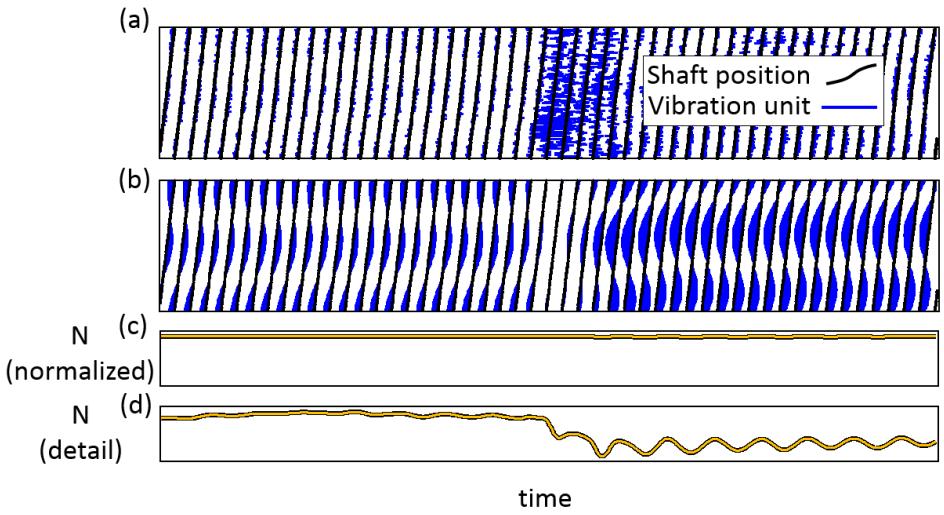


Figure 5: Application to ice shedding, fleeting event and change in unbalance amplitude and position. Representation in the time/phase domain of the vibration amplitude (a) and of the vibration amplitude tracked around 1 N (b). Time history of the normalized rotating speed (c) and of the detail of the rotating speed (d). The unbalance decreases and increases in amplitude during the shedding while its phase location is almost at the opposite direction.

The ingestion of several birds and the stall of a compressor are illustrated in figures 6 and 7. During those events, high frequency content appears but are not sustained. Time/phase representation is not really relevant in this cases where no obvious correlation of the frequency content and the phase location is observed.

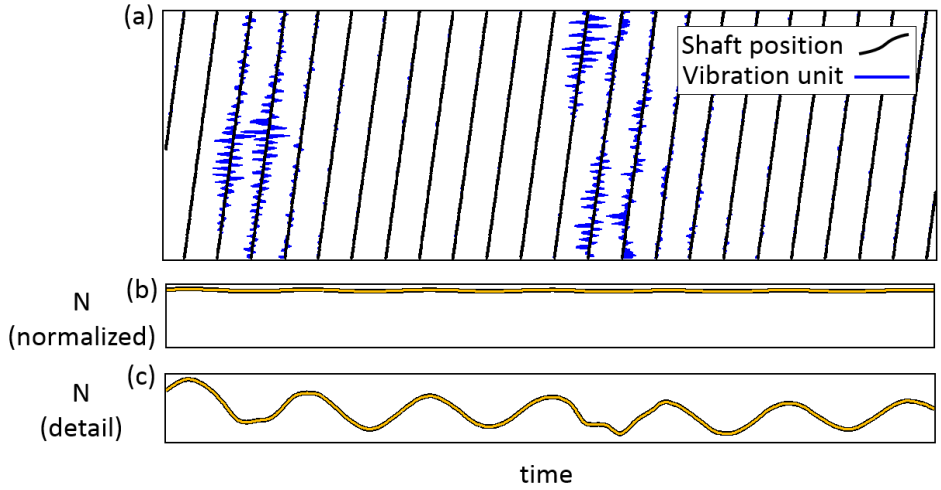


Figure 6: Application to the ingestions of several birds. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

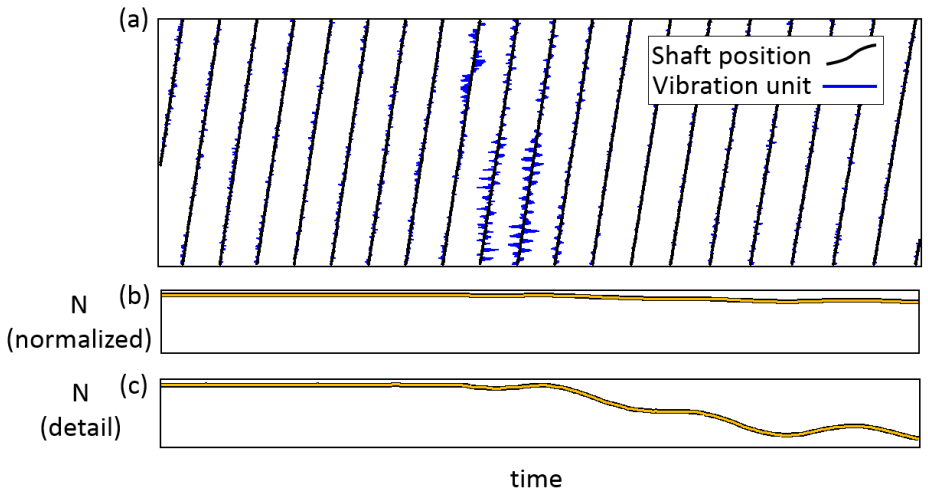


Figure 7: Application to a compressor stall. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

Another severe event is the separation of a blade from the shaft. In this case, the unbalance can be huge and a solution is to change the dynamic response of the shaft in order to limit the loads in the engine and on the aircraft. During the deceleration of the shaft after a fan blade separation, the rotating speed crosses a rotor mode. Figure 8 shows the vibration representative of a load on a static bearing. During the crossing of the mode, the load increases and the phase delay between the load and the angular position of the shaft decreases.

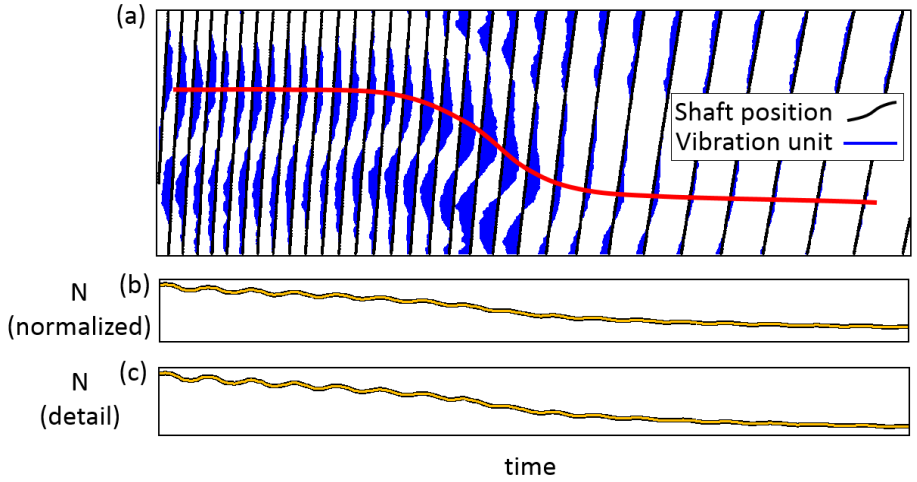


Figure 8: Application to the crossing of the rotating speed and a mode of the shaft. Representation in the time/phase domain of the vibration amplitude of a strain gauge mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c). The red curve follows the maximal positive amplitude at each cycle.

The representation in the time/phase domain is particularly appropriate for the analysis of this type of event because the representation has an immediate physical sense that cannot be easily understood with other kinds of phase representation. For instance it is not necessary to understand if the convention for the phase between vibrations and a reference on the shaft is ‘phase lag’ or ‘phase lead’. In figure 7, it is obvious that the phase of the vibration is slowly accelerating, compared to the rotating speed of the shaft, during the crossing of the mode, and this is what is expected during a deceleration.

Figure 9 shows another behaviour for which the time/phase representation is particularly suitable. During this event, an interaction between the rotor and a static case occurred. The torsional mode of the shaft was

excited, and the vibrations recorded on the static case shows that this interaction is characterized by an excitation of a sub-harmonic of the rotating speed. The time/phase representation allows to identify at a glance that the interaction occurs every three rotations of the shaft, and that the interactions occur when the shaft is approximately at the same angular position, meaning that the interactions between the rotor and the stator were very likely to appear repetitively in the same angular areas.

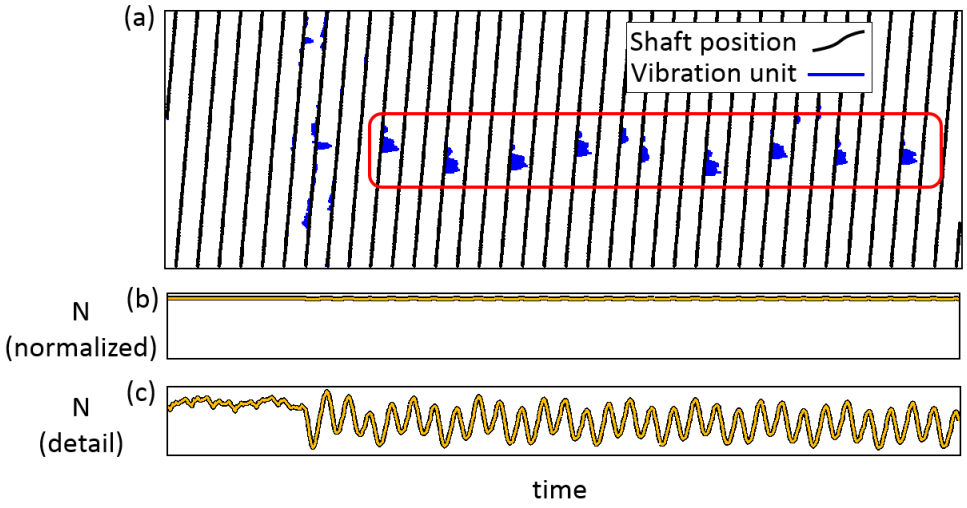


Figure 9: Repetitive shocks on a static case when a rotor interacts with a stator. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a case of the engine (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

4 Representation of several events with a simple model

Several events were simulated with a simple lumped-mass model [4]. Such simulations are typically used for the understanding of the dynamic behaviours of rotating shafts during the design of new aircraft engine types.

The lumped parameter model simulates the low speed shaft of an aircraft engine subjected to different external load conditions simulating the experimental cases shown on the previous section. Figure 10 shows a schema of the system which considers 3 degrees of freedom. The mass of the whole shaft is lumped to one rigid node resting on flexible supports with viscous damping.

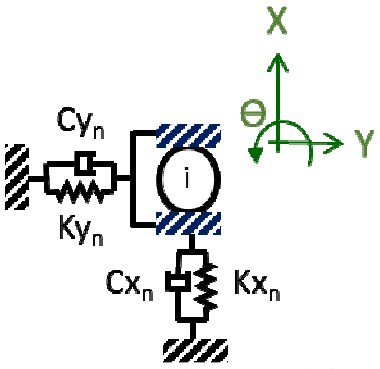


Figure 10: Lumped mass model Schema.

The general system of equations is:

$$[M].\{\ddot{X}\} + [C].\{\dot{X}\} + [K].\{X\} = \{F_{ext}(t)\} \quad (1)$$

Matrices M, C and K represent the mass, damping and stiffness of the system and vector X contains the generalized displacements of the node I on the x, y and Θ directions. The modelling approach is completely

described in reference [4]. The resolution is performed in the angular domain as described in the given reference and the resolution is performed with Matlab's® ode15s solver.

Figure 11 shows the result of the measurements and of the simulation of the application of a rigid and constant unbalance and the application of a progressively increasing oil unbalance. The results from the simulation and from the measurements are very similar.

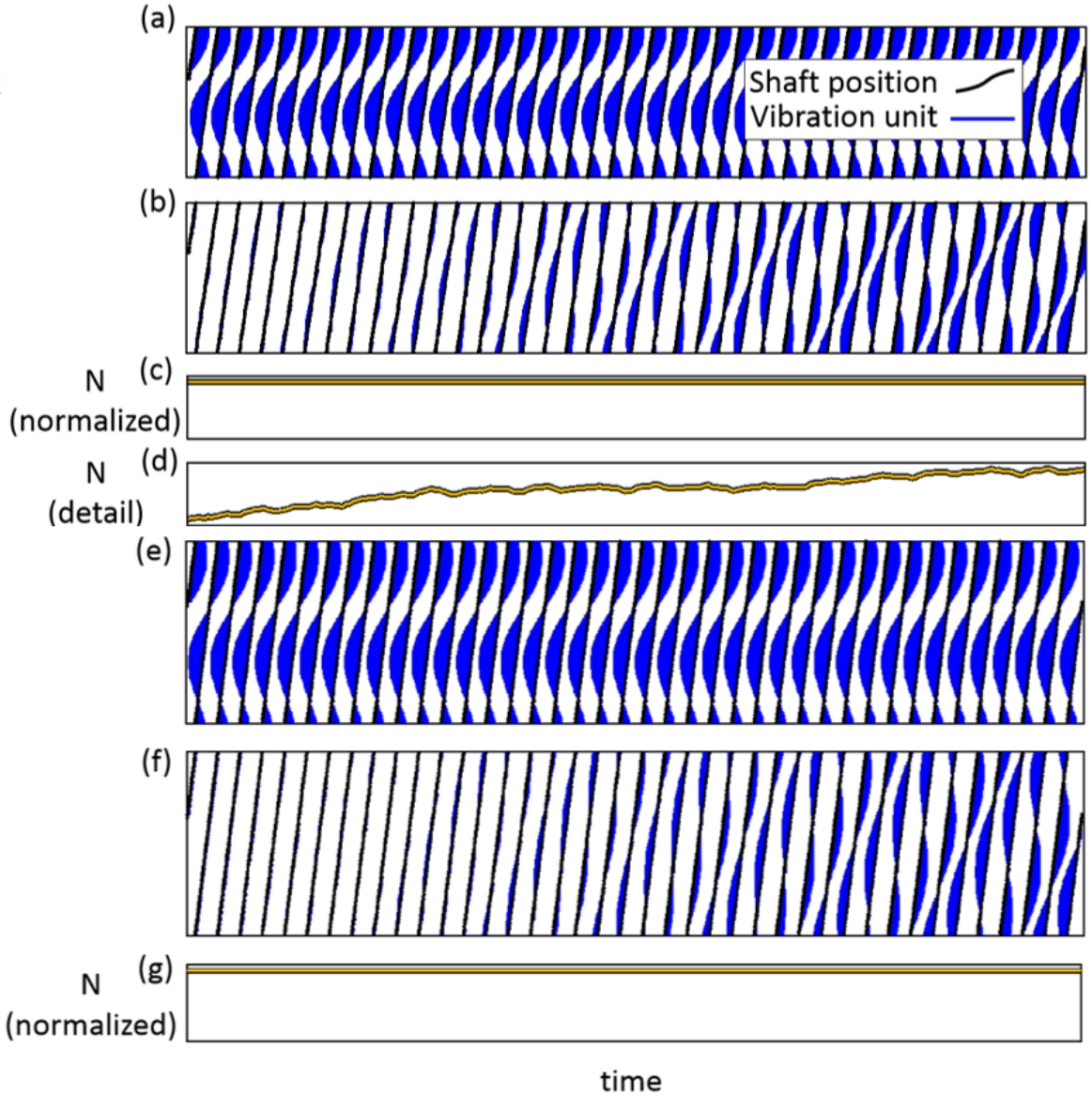


Figure 11: Application to a constant rigid unbalance and the progressive increase of an oil unbalance. Representation in the time/phase domain of the vibration amplitude tracked on the first order 1 N of the rotating speed (measurement (a), simulation (e)) and of the vibration amplitude tracked around 0.8 N (measurement (b), simulation (f)). Time history of the normalized rotating speed (measurement (c), simulation (g)) and of the detail of the rotating speed for the measurement (d).

Figure 12 shows the results of a simulation of the crossing of a rotating speed with a mode of the rotor. The model shows the changes in amplitude and phase that are similar to the changes observed during the tests in figure 8.

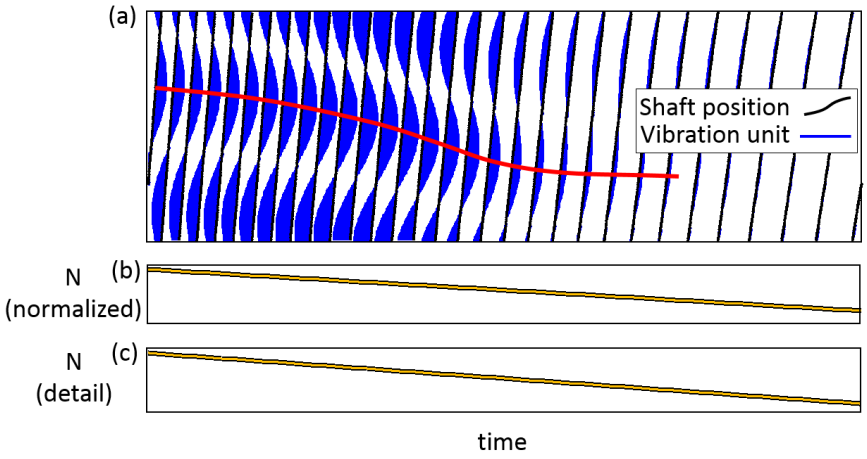


Figure 12: Simulation of the response to a crossing of the rotating speed and a mode of the shaft. Representation in the time/phase domain of the vibration amplitude of a strain gauge mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c). The red curve follows the maximal positive amplitude at each cycle.

For the simulation of repetitive shocks between a rotor and a stator, the contact between the rotor and the stator has been simulated by a sudden change of the stiffness and damping in the vertical direction when the clearance is consumed due to unbalance as shown in Figure 13. The stator has been given a sinusoidal displacement law with a frequency close to a third of the rotating speed of the shaft.

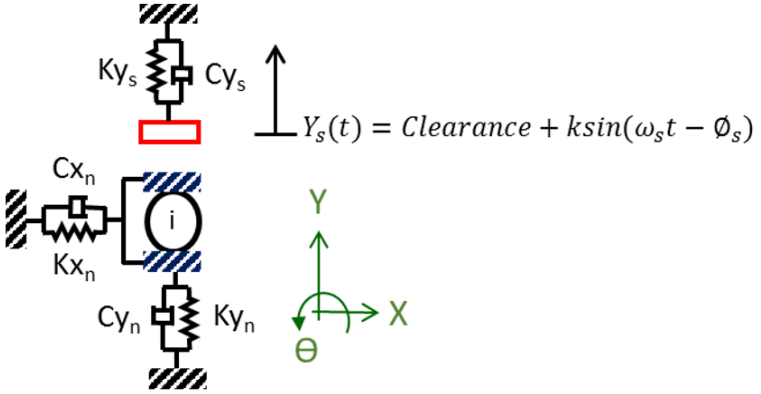


Figure 13: Rotor-stator contact model schema.

Figure 14 shows the results of the simulation of repetitive shocks between a rotor and a stator. The results are also similar to the records in figure 9.

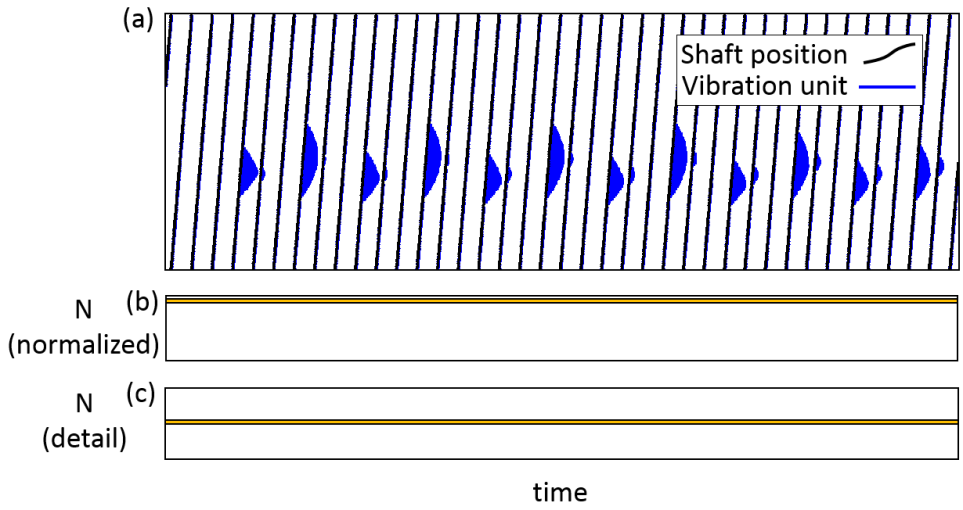


Figure 14: Simulation of the response to repetitive shocks. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a case of the engine (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

The results obtained with the very simple model and the comparison of those results with those obtained from tests are similar. The time/phase representation of both model and tests results facilitates this comparison.

The model can reproduce almost all dynamic behaviours observed during the tests with the time/phase representation. The only limitation is related to the impossibility to reproduce the torsional mode because there is only a single degree of freedom on the angle direction. Most influencing factors can be identified while changing the parameters of the model (mass, stiffness damping, external loads and rotating speed) without further expensive tests on real aircraft engines.

5 Conclusion

A new representation of vibration signals in the time/phase domain has been described. This representation has been applied on data recorded during several tests of aircraft engines, and on results obtained from a very simple model. The benefits of this representation are the following:

- The representation is a small change of the standard time history representation, it is therefore easy to become familiar with the new representation,
- The representation facilitates the search for correlations between the vibrations and the angular location of the shaft because the angular location is on the y-axis of the graphs,
- The time axis is preserved, correlations can be identified and causal analysis can be performed.

References

[1] F. F. Ehrich, *Observed Rotordynamic Phenomena in Aircraft Gas Turbine Development*, ASME 2015. International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Volume 8: 27th Conference on Mechanical Vibration and Noise.

[2] R. Potter, M. Gribler, *Computed Order Tracking Obsoletes Older Methods*, Journal of passenger cars, 1989, Vol. 98, section 6, pp.1099-1103.

[3] C. Feldbauer., R. Holdrich., *Realization of a Vold-Kalman Tracking Filter – A Least Squares Problem,* Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), Verona, Italy, December 7-9, 2000

[4] A. Bourdon et al., Introducing angularly periodic disturbances in dynamic models of rotating systems under non-stationary conditions, Mechanical Systems and Signal Processing, 2014, vol.44, pp.60-71.

Fault diagnosis and prognosis for rolling bearings

Fourier-Bessel series expansion based blind deconvolution method for bearing fault detection

E. Soave[1], **G. D'Elia**[1], **G. Dalpiaz**[1]
[1]University of Ferrara, Department of Mechanical Engineering,
via Saragat 1, 44122 Ferrara, Italy
e-mail: elia.soave@unife.it

Abstract

In the last few years blind deconvolution techniques proved to be useful in order to extract impulsive patterns related to bearing fault from noisy vibration signals. Recently, a novel blind deconvolution method based on the generalized Rayleigh quotient has been proposed and an iterative algorithm related to the maximization of the cyclostationarity of the source has been defined. This paper presents a new condition indicator that exploit the Fourier-Bessel series expansion for the computation of a new cyclostationarity index that drives the maximization problem for the extraction of the excitation source. The main target of this work is to compare the results obtained through the exploitation of the Fourier-Bessel transform with respect to the classic Fourier transform in term of lower number of cyclic frequencies required for the algorithm. The comparison between the application of the two different methods involves both simulated and real signal taking into account a bearing fault. The results prove the capability of the new indicator to extract the impulsive source without the need of a set of cyclic frequencies but only with the first one, with a strong reduction of the computational time.

Keyword: Cyclostationarity, Fourier-Bessel transform, blind deconvolution, bearing fault detection.

1 Introduction

Nowadays, bearing fault detection and identification play a fundamental role in the diagnostics of rotating machines being this component one of the most critical in mechanical systems. However, the detection of bearing faults, particularly in the first stage of the damaging process, may be challenging due to the fact that the impulsive pattern related to the fault is very weak and consequently masked by the background noise and other interference. For this purpose, blind deconvolution (BD) techniques proved to be able to extract the characteristic impulsive pattern of a faulty bearing direct from a noisy observation, under the hypothesis of linear time-invariant system.

The first BD method, called Minimum Entropy Deconvolution (MED) [1], exploits an iterative algorithm in order to extract the impulsive source through the maximization of the kurtosis directly from a noisy vibration signal convolved with an unknown impulse response function. Over the years, the MED algorithm has been exploited for machine diagnosis by many authors both for tooth faults [2, 3] and bearing faults [4] in combination with other signal processing techniques. However, from the machine diagnosis point of view kurtosis can not be considered a a useful indicator. In fact, this statistics is sensitive to signal with a single peak rather than to signal peakedness according to a given periodicity, e.g. represented by a train of impulses.

In order to overcome this problem and exploit the BD for the fault detection in rotating machines, McDonald at al. [5] proposed a new BD method based on a criterion called correlated kurtosis that is sensitive to signal composed by a train of periodic peaks. This algorithm, called Maximum Correlated Kurtosis Deconvolution (MCKD), is based on an iterative process that aims to estimate the source having the maximum correlated kurtosis.

In the same direction, McDonald and Zhao [6] proposed a new BD method called Multipoint Optimal Minimum Entropy Deconvolution Adjusted (MOMEDA). This direct algorithm promotes the estimation of a

pattern of periodic impulses by considering a target vector **t**, defined as a train of equispaced impulses. The main issue of MOMEDA is related to the definition of **t** that allows the extraction of periodic impulses.

Over the years, cyclostationarity proved to be effective for machine diagnosis. In particular, several works [7, 8] prove that the vibration signal related to a bearing fault is not strictly periodic due to a random "jitter" around the fault period, but it can be seen as a second order cyclostationary signal. The correlated kurtosis can be seen as a cyclostationary criterion being it related to the autocorrelation function of the instantaneous power of the signal [9]. However, it has been defined empirically, without an explicit formulation regarding it cyclostationary nature, thus there is the need to define a pure cyclostationary criterion.

Recently, in this direction, Buzzoni et al. [10] proposed a new BD method based on the maximization of the cyclostationarity of the researched signal through an iterative algorithm that solves an eigenvalue problem starting from a generalized Rayleigh quotient. This algorithm aims to extract the source related to the maximum value of the Indicator of Ciclostationarity (ICS). This indicator, proposed for the first time by Raad et al. [11], exploits the Fourier transform in order to extract the hidden periodicity inside the acquired signal. However, the Fourier series expansion requires an high number of terms for the description of an impulsive pattern. For this reason, the algorithm needs to take account of a certain number of harmonics of the cyclic frequency related to the investigated periodicity in order to reach a sufficient reconstruction quality.

The target of this work is the definition of a new BD indicator based on the maximization of the ciclostationary behavior of the analyzed signal. This indicator exploits the Fourier-Bessel series expansion instead of the classical Fourier transform. The choice of this particular series expansion is related to the fact that, unlike the sinusoidal function based Fourier transform, it is based on the Bessel functions that quickly decay according to a specific law [12, 13]. For this reason, the proposed method requires a set of cyclic frequencies with a lower number of harmonics, strongly reducing the computational time of the iterative algorithm.

The BD method modified through the exploitation of a new indicator is verified through the analysis of both simulated signal and a real test case in order to demonstrate its effectiveness for the fault detection and the condition monitoring of bearings. The comparison between the proposed method and the other methods based on the ciclostationarity maximization aims to point out the improvement given by the proposed algorithm, in particular in term of computational time reduction.

2 Overview about blind deconvolution methods through generalized Rayleigh quotient optimization

The target of BD is the extraction of an input signal, typically related to the excitation given by a fault, directly from a noisy observation under the hypothesis of linear time invariant system.

The vibration signal, from now called **x**, can be considered as composed by three different parts: an impulsive pattern $\mathbf{s}_0$ related to a local fault, a periodic component **p** and a Gaussian background noise **n**, all convolved with their respective Impulse Response Function (IRF), viz:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n \tag{1}$$

where $\mathbf{g}_s$, $\mathbf{g}_p$ and $\mathbf{g}_n$ are the IRFs related to $\mathbf{s}_0$, **p** and **n**, respectively and * refers to the convolution operation.

BD aims to estimate an inverse filter **h** (assumed to be a FIR filter) in order to deconvolve the excitation $\mathbf{s}_0$ from **x** minimizing the other contributions, such that:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} = (\mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n) * \mathbf{h} \approx \mathbf{s}_0 \tag{2}$$

In the first part of eq. 2 the convolution can be expressed for discrete signals in matrix form, as follow:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \tag{3a}$$

$$\begin{bmatrix} s[0] \\ \vdots \\ s[L-1] \end{bmatrix} = \begin{bmatrix} x[N-1] & \dots & x[0] \\ \vdots & \ddots & \vdots \\ x[L-1] & \dots & x[L-N-2] \end{bmatrix} \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} \tag{3b}$$

where L and N are the number of samples of **s** and **h**, respectively. The main problem is related to the fact that the IRFs described in eq. 2 are not available. A possible solution can be achieved considering an arbitrary

criterion based on a priori assumption, e.g. assuming that a statistical property is maximized by the researched source. Thus, BD aims to recover the source that maximizes a certain statistical property, e.g. kurtosis [1], correlated kurtosis [5] or multi D-Norm [6], through iterative or direct algorithms.

Although cyclostationarity plays a fundamental role in the diagnostics of rolling element bearings the CYCBD method [10] can be considered as the first BD method that exploits the cyclostationary signature of faulty bearings.

2.1 Blind deconvolution algorithms driven by cyclostationarity maximization based on generalized Rayleigh quotient

Recently, the method proposed in Ref. [10] demonstrated that the generalized Rayleigh quotient can be exploited in order to define a new optimization problem. Starting from this statement a new BD method based on the cyclostationarity maximization has been defined.

A generical process can be defined as cyclostationary if its statistical properties exhibit a periodic behaviour. It has been demonstrated by several authors [7, 8] that the vibration signature related to a faulty bearing shows a cyclostationary behavior, thus this property became pivotal for bearing fault detection and identification. In this scenario, the frequency related to the periodic behavior of any statistic of the process is called cyclic frequency and, for discrete signal, can be expressed as follow:

$$\alpha = \frac{k}{T} \tag{4}$$

where k is the sample index and T the cyclic period, e.g. the characteristic period of a bearing fault. The BD algorithm descried in this section, called CYCBD, is based on the maximization of the second-order ICS that describes the presence of periodic fluctuations of the energy flow of the signal.

First of all, it is necessary to remember the definition of second-order ICS:

$$ICS_2 = \frac{\sum\limits_{k>0} |c_s^k|^2}{|c_s^0|^2} \tag{5}$$

with

$$c_s^k = \frac{1}{L-N+1} \sum_{n=N-1}^{L-1} |s[n]|^2 e^{-j2\pi \frac{k}{T} n} \tag{6a}$$

$$c_s^0 = \frac{||\mathbf{s}||^2}{L-N+1} \tag{6b}$$

or, in matrix form:

$$c_s^k = \frac{\mathbf{E}^H |\mathbf{s}|^2}{L-N+1} \tag{7a}$$

$$c_s^0 = \frac{\mathbf{s}^H \mathbf{s}}{L-N+1} \tag{7b}$$

where

$$|\mathbf{s}|^2 = [|s[N-1]|^2, \ldots, |s[L-1]|^2]^T \tag{8a}$$

$$\mathbf{E} = \begin{bmatrix} e^{-j2\pi \frac{1}{T}(N-1)} & \ldots & e^{-j2\pi \frac{K}{T}(N-1)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi \frac{1}{T}(L-1)} & \ldots & e^{-j2\pi \frac{K}{T}(L-1)} \end{bmatrix} \tag{8b}$$

Now it is possible to rewrite eq. 5 by substituting eq. 7a and 7b as:

$$ICS_2 = \frac{|\mathbf{s}|^{2H} \mathbf{E}\mathbf{E}^H |\mathbf{s}|^2}{|\mathbf{s}^H \mathbf{s}|^2} \tag{9}$$

Finally, ICS_2 can be rewritten by substituting eq. 3a into eq. 9, viz:

$$ICS_2 = \frac{\mathbf{h}^H \mathbf{X}^H \mathbf{W} \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XWX} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XX} \mathbf{h}} \tag{10}$$

where $\mathbf{R}_{XWX}$ and $\mathbf{R}_{XX}$ are the weighted correlation matrix and the correlation matrix, respectively, and $\mathbf{W}$ is the weighting matrix, expressed as:

$$\mathbf{W} = diag\Big(\frac{\mathbf{E}\mathbf{E}^H |\mathbf{s}|^2}{(L-N+1)\mathbf{s}^H \mathbf{s}}\Big)(L-N+1) = \begin{bmatrix} \ddots & & 0 \\ & \frac{\mathbf{E}\mathbf{E}^H |\mathbf{s}|^2}{(L-N+1)} & \\ 0 & & \ddots \end{bmatrix} \frac{(L-N+1)}{\sum_{n=N-1}^{L-1} s[n]^2} \tag{11}$$

It should be noticed that eq. 10 is a generalized Rayleigh quotient. It has been demonstrated [10] that the maximization of the ICS_2 in eq. 10 with respect to the filter coefficients $\mathbf{h}$ is equivalent to the eigenvector related to the maximum eigenvalue λ of the following generalized eigenvalue problem:

$$\mathbf{R}_{XWX} \mathbf{h} = \mathbf{R}_{XX} \mathbf{h} \lambda \tag{12}$$

thus, the maximum value of λ correspond to the maximum value of ICS_2.

The maximization of ICS_2 and the extraction of the associated filter $\mathbf{h}$ is reached through an iterative process. The iteration starts with an initialization of the inverse filter $\mathbf{h}$ in order to compute the weighting matrix $\mathbf{W}$ and solve the eigenvalue problem described in eq. 12, obtaining the filter $\mathbf{h}$ related to the maximum value of λ. The iterative process restarts from the evaluation of $\mathbf{W}$ and goes on until convergence. A suitable solution for the computation of a guess of $\mathbf{h}$ is to estimate the filter coefficient through an auto-regressive (AR) model filter, e.g. by means of the Yule-Walker equations, according to Ref. [10]. This model permits to attenuate all the components related to deterministic sources in the signal spectra thus it is possible to obtain a flat spectral density, typical of a impulsive patterns.

Once the iterative process is ended, the maximum value of ICS_2 is available and the source obtained by substituting the related value of $\mathbf{h}$ into eq. 3a represents the final deconvolved source.

3 Fourier-Bessel series expansion based blind deconvolution method

3.1 Fourier-Bessel series expansion

Let us consider a generic discrete signal x(n); the Fourier-Bessel series expansion can be written in the discrete form as follows:

$$x(n) = \sum_{i=1}^{L} C_i J_0\Big(\frac{\beta_i n}{L}\Big), n = 0, 1, ..., L-1 \tag{13}$$

where L is the length of x, J_0 is the zero-order Bessel function and C_i are the Fourier-Bessel series coefficients, defined as:

$$C_i = \frac{2}{L^2 [J_1(\beta_i)]^2} \sum_{n=0}^{L-1} n x(n) J_0\Big(\frac{\beta_i n}{L}\Big) \tag{14}$$

where J_1 is the first order Bessel function and β_i are the positive roots of $J_0 = 0$.

According to Schroeder [14], the positive roots of the zero order Bessel function are related to the frequency domain by the following:

$$\beta_i \approx \frac{2\pi f_i L}{f_s} \tag{15}$$

where f_s is the sampling frequency of x(n).

It is possible to compare the Fourier-Bessel coefficients C_i with the Fourier coefficients for the description of the signal x(n). However, the Bessel functions decay within the range defined by L according to eq. 13, instead of the sinusoidal functions on which is based the Fourier series. For this reason, the Fourier-Bessel transform requires a lower number of coefficients, with respect to the classic Fourier transform, in order to obtain high quality in the reconstruction of the signal.

This work aims to exploit this characteristic of the Fourier-Bessel transform in order to define a new BD criterion that permits the reduction of the computational time of the iterative algorithm due to the lower number of cyclic frequency harmonics required for the analysis.

3.2 Proposed method

The CYCBD algorithm can be reformulated as follows. First of all, let us remember the second order indicator of ciclostationarity defined in eq. 5. Unlike the formula described in eq. 6, the numerator of the ICS_2 can be rewritten exploiting the Fourier-Bessel series expansion described in eq. 14 as follows:

$$c_F^k = \frac{2}{(L-N+1)^2[J_1(\beta_k)]^2} \sum_{n=N-1}^{L-1} n|s(n)|^2 J_0\Big(\frac{\beta_k n}{L-N+1}\Big) \tag{16}$$

where the roots of the zero-order Bessel function can be expressed as:

$$\beta_k = \frac{2\pi \frac{k}{T_s}(L-N+1)}{f_s} \tag{17}$$

where the term k/T_s represents the cyclic frequency, according to eq. 4.

Eq. 16 can be expressed in matrix form as follows:

$$c_F^k = \frac{2}{(L-N+1)^2} \frac{\mathbf{J}_{0_n}^H |\mathbf{s}|^2}{\mathbf{J}_1 \mathbf{J}_1^H} \tag{18}$$

where

$$|\mathbf{s}|^2 = [|s[N-1]|^2, \ldots, |s[L-1]|^2]^T \tag{19a}$$

$$\mathbf{J}_{0_n} = \begin{bmatrix} J_0\Big(\frac{\beta_1(N-1)}{L-N+1}\Big)(N-1) & \ldots & J_0\Big(\frac{\beta_K(N-1)}{L-N+1}\Big)(N-1) \\ \vdots & \ddots & \vdots \\ J_0\Big(\frac{\beta_1(L-1)}{L-N+1}\Big)(L-1) & \ldots & J_0\Big(\frac{\beta_K(L-1)}{L-N+1}\Big)(L-1) \end{bmatrix} \tag{19b}$$

$$\mathbf{J}_1 = [J_1(\beta_1) \ldots J_1(\beta_k) \ldots J_1(\beta_K)] \tag{19c}$$

Starting from eq. 7b and eq. 18, eq. 5 can be rewritten as:

$$ICS_{2F} = \frac{4}{(L-N+1)^2 |\mathbf{J}_1 \mathbf{J}_1^H|^2} \frac{|\mathbf{s}|^{2H} \mathbf{J}_{0_n} \mathbf{J}_{0_n}^H |\mathbf{s}|^2}{|\mathbf{s}^H \mathbf{s}|^2} \tag{20}$$

Analogously to the CYCBD method, it is possible to note that all the periodic components of $|\mathbf{s}|^2$ are comprising into the following signal:

$$P[\mathbf{s}] = \frac{4}{(L-N+1)^2} \sum_{k>0} \frac{J_{0_n}^k (J_{0_n}^k |s|^2)}{[J_1(\beta_k)]^2} = \frac{4}{(L-N+1)^2} \frac{\mathbf{J}_{0_n} \mathbf{J}_{0_n}^H |\mathbf{s}|^2}{|\mathbf{J}_1 \mathbf{J}_1^H|^2} \tag{21}$$

Substituting eq. 3a and eq. 21 into eq. 20 it is possible to write the ICS_2 as a generalized Rayleigh quotient:

$$ICS_{2F} = \frac{\mathbf{h}^H \mathbf{X}^H \mathbf{W} \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XWX} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XX} \mathbf{h}} \tag{22}$$

where the weighting matrix can be expressed as:

$$\mathbf{W} = diag\Big(\frac{P[|\mathbf{s}|^2]}{\mathbf{s}^H \mathbf{s}}\Big)(L-N+1) = \begin{bmatrix} \ddots & & 0 \\ & P[|\mathbf{s}|^2] & \\ 0 & & \ddots \end{bmatrix} \frac{1}{\sum_{n=N-1}^{L-1} s[n]^2} \tag{23}$$

Eq. 22 is the base of the proposed BD method, called Fourier-Bessel Blind Deconvolution (FBBD). The maximum value of the proposed criterion can be found by solving the same iterative process based on the maximization of the Rayleigh quotient described in section 2.1.

4 Application to simulated signal

The FBBD method proposed in the previous section has been validated taking into account a simulated signal that reproduces a characteristic cyclostationary signal. During the validation the method is compared with the CYCBD method described in section 2 in order to demonstrate the improvement given by the Fourier-Bessel transform, specially in term of lower computational time due to the lower number of cyclic frequencies required by the algorithm.

4.1 Simulated signal

The simulated signal used for the validation of the proposed method has been created according to eq. 1, neglecting the periodic pattern **p** as follows:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{n} * \mathbf{g}_n \tag{24}$$

where the IRFs $\mathbf{g}_s$ and $\mathbf{g}_n$ have been defined according to the model of the response of a damped single degree of freedom (SDOF) system to a unit impulse in the time domain. For continuous signal, this IRF can be defined in term of displacement as:

$$\mathbf{x}_{SDOF} = Ae^{-\zeta\omega_n t}sin(\omega_d t) \tag{25}$$

where A is the response amplitude, ζ is the damping coefficient, ω_n is the resonance frequency and ω_d is given by the following:

$$\omega_d = \omega_n\sqrt{1-\zeta^2} \tag{26}$$

Starting from eq. 25 the IRFs can be defined in term of acceleration by taking the second derivative with respect to the time.

Table 1: Parameters used for the IRFs.

	$\mathbf{g}_s$	$\mathbf{g}_n$
A	1.810^{-10}	1.210^{-10}
ζ	0.005	0.045
ω_n(rad/s)	18.84	62.83

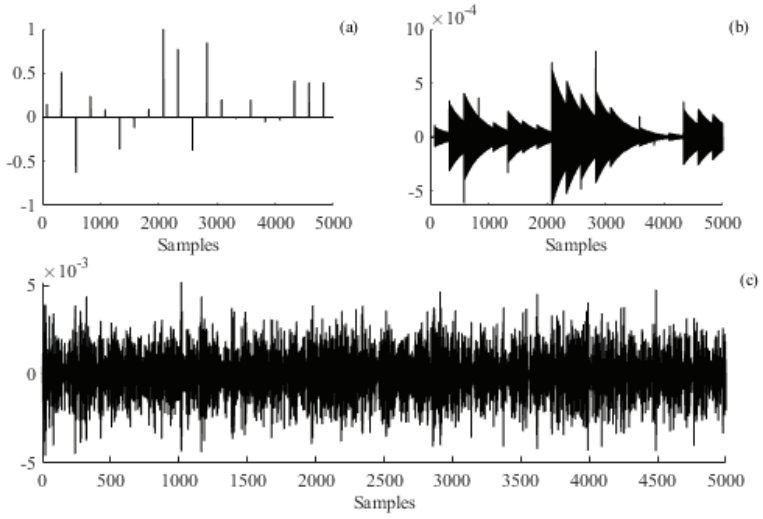


Figure 1: Simulated signal used for the validation: impulsive pattern with Gaussian distributed amplitudes (a), impulsive pattern convolved with its IRF (b) and overall signal (c).

The cyclostationary signal used for the validation is composed by a pattern of periodic impulses (with 250 samples period) with amplitudes distributed according to a Gaussian law and an additive Gaussian background noise (SNR = -17dB). The IRFs $\mathbf{g}_s$ and $\mathbf{g}_n$ are described by the parameters collected in table 1.

Figure 1 depicts the overall simulated signal and the patterns that compose it. It is possible to note that in the overall signal the periodic impulses are strongly masked by the background noise. For this reason the extraction of the cyclostationary source may represent a challenging test for the proposed method thus this validation may permit to demonstrate the effectiveness of the FBBD algorithm.

4.2 Results

The simulated signal has been analyzed through FBBD and CYCBD method in order to highlight the improvement given by the proposed algorithm, in particular in term of lower number of cyclic frequencies required. Due to the inability of BD methods to find the real amplitude of the researched pattern, all the sources obtained from the analysis has been normalized by their respective maximum value.

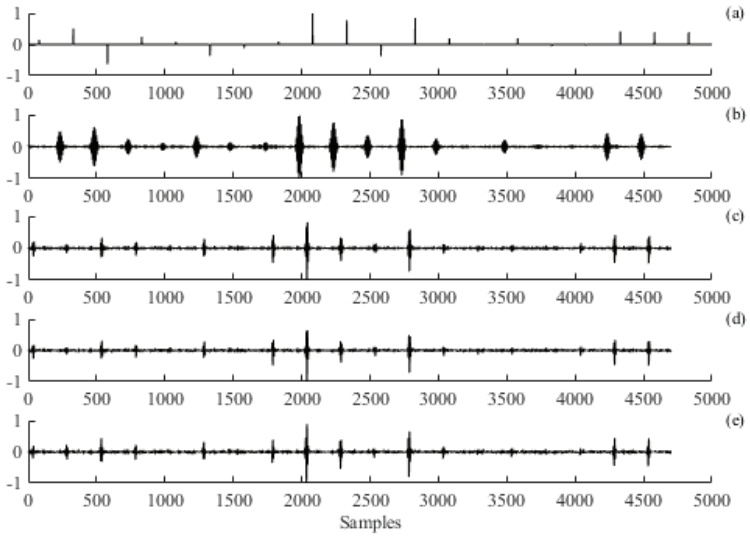


Figure 2: Comparison between the target source (a) and the patterns estimated through the FBBD method at increasing number of cyclic frequencies (from b to e).

Figure 2 reports the results obtained through the application of the FBBD method considering an increasing number of cyclic frequency harmonics, i.e. 1, 10, 30, 50 (from top to bottom). It clearly shows that the FBBD is able to extract the impulsive pattern with the same period of the target source and with the correct relative amplitudes. It has to be underlined that the deconvolved source presents a delay with respect the reference signal, due to the filtering operation inside the BD algorithm. However, this delay does not influence the results thus it is not corrected in the following analysis.

The main issue pointed out by the comparison between the estimated sources and the target pattern is related to the sign of the recovered impulses. In fact, it has to be noticed that the FBBD algorithm is not able to recover the correct sign of the impulses but it extracts a source composed by symmetric peaks. This particular behaviour do not depend on the number of cyclic frequencies used for the analysis. In fact, comparing the sources in figure 2(c to e) it is possible to see no changing in the reconstructed signal amplitudes with increasing number of cyclic frequency harmonics. This result agrees with the consideration explained in section 3 about the low number of term required by the Fourier-Bessel transform for the reconstruction of signal with sufficient quality.

Figure 3 depicts the sources estimated after the application of the CYCBD algorithm at increasing number of considered cyclic frequency harmonics, i.e. 10, 30, 50, 100 (from b to e). There is a significant difference between the behaviour of the CYCBD and the FBBD for different number of considered harmonics. For low number of cyclic frequency harmonics, i.e. 10 as shown in figure 3(b), the estimated source is not able to reproduce the real sign of the impulses as well as for the FBBD method. However, considering an higher

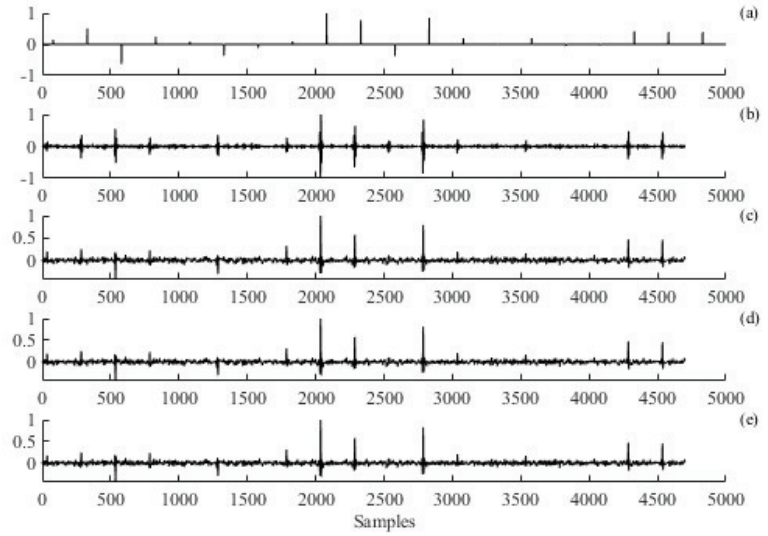


Figure 3: Comparison between the target source (a) and the patterns estimated through the CYCBD method at increasing number of cyclic frequencies (from b to e).

number of cyclic frequency harmonics, i.e. 30, 50 and 100 as shown in figure 3(c-e), the algorithm permits to extract the correct sign of the peaks, reproducing the target source with higher quality with respect to the FBBD algorithm.

Moreover, comparing figure 3(b-e) it has to be noticed that the quality of the reconstruction reaches an acceptable level considering at least 30 cyclic frequency harmonics (figure 3(c)). Further increasing the number of harmonics (figure 3(d-e)), the deconvolved source does not seem to present any improvement from this point of view.

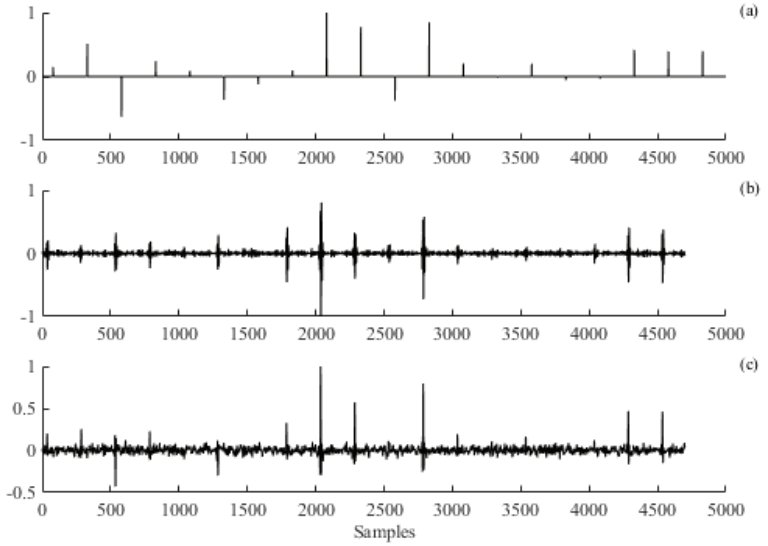


Figure 4: Comparison between the target source (a) and the sources estimated through the FBBD method (b) and the CYCBD method (c).

Figure 4 compares the impulsive patterns recovered using both FBBD and CYCBD. All two methods has been applied using the minimum number of cyclic frequency harmonics identified by the previous analysis, i.e. 10 and 30 respectively.

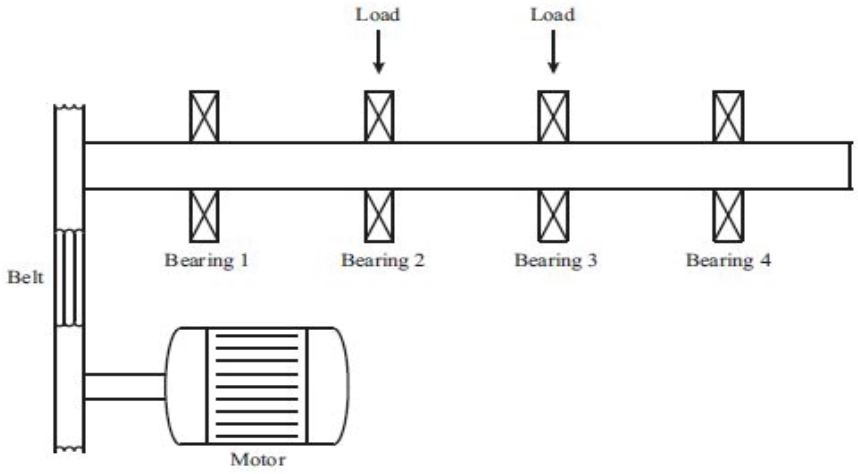


Figure 5: Scheme of the experimental test rig.

The comparison clearly underlines the main difference between the results obtained applying the two methods, represented by the gap in term of reconstruction of the correct sign of the source. In fact, as already seen in figure 2 the FBBD extracts a zero mean impulsive pattern, independently from the real sign of the peaks. However, this issue may be considered irrelevant for diagnostics purposes where the main attention is pointed out on the frequency of the impulses instead of their amplitudes. On the other hand, the FBBD seems to permit the identification of the lower amplitude impulses better tha the CYCBD. For examples, comparing figures 4(b-c) it can be seen that the impulses close to 2500 and 4000 samples are clearly identified only by the FBBD, due to the higher amplitude of the noise in the source deconvolved by the CYCBD that masks these peaks.

These results, combined with the lower number of cyclic frequencies required (70% less), demonstrates the clear improvement given by the FBBD method for the extraction of hidden ciclostationary pattern directly from noisy observations such as the characteristic vibration signal related to faulty bearings.

5 Application to real signal

The second part of the experimental application of the proposed method regards the early detection and identification of a bearing fault during a run-to-failure test.

5.1 Experimental setup

This part of the analysis has been carried out by using the vibration signals contained into a data set provided by the Center for Intelligent Maintenance Systems of the University of Cincinnati [15]. As shown by the scheme in figure 5, the test rig is composed by four bearings type Rexnord ZA-2115 mounted on the same shaft. During the test the rotational speed has been fixed at 2000rpm and a radial load of 27.7kHz has been applied to bearings 2 and 3. The vibration signals have been acquired through two accelerometers model PCB 253B33 mounted in radial direction. Each bearing has been monitored continuously acquiring 1s of signal each 10min with a sampling frequency of 20.48kHz. After 7 days the test has been stopped and an outer race fault has been detected in bearing 1.

5.2 Results

The target of this section is to demonstrate that the proposed criterion is able to monitor the progressive evolution of the bearing condition during the entire operative life in order to detect the appearance of faults as early as possible. For this scope, the variation of the maximized BD criterion during the endurance test can be observed taking account of three characteristic frequencies: the ball pass frequency outer race (BPFO), the ball pass frequency inner race (BPFI) and the ball spin frequency (BSF).

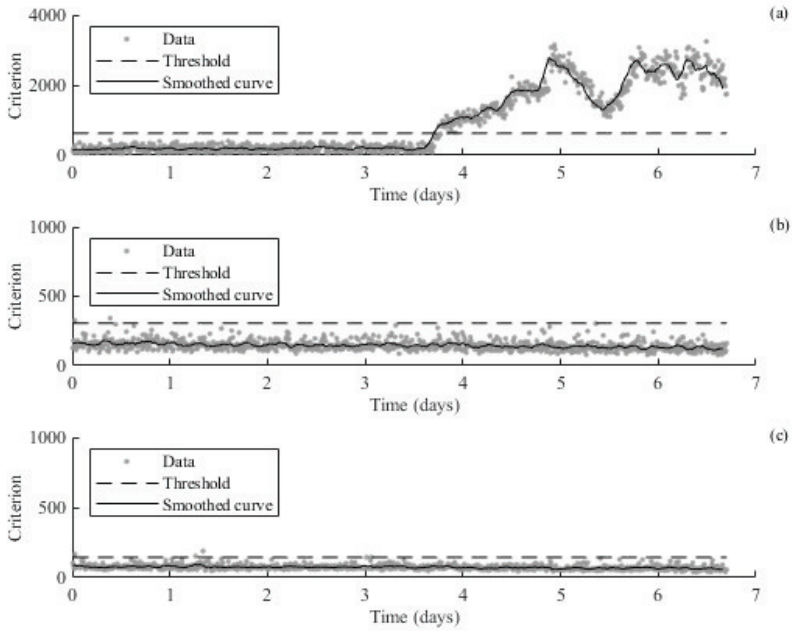


Figure 6: Final values of maximized criterion during the endurance test taking account of BPFO (a), BPFI (b) and BSF (c).

The FBBD method has been applied considering only the fundamental cyclic frequency related to the bearing fault and a 100 samples length FIR filter. Due to the experimental evidence, the analysis has been carried out considering only the faulty bearing 1.

Figure 6 collects the values of the maximized criterion calculated on bearing 1 during the endurance test. The figures depict a constant trend of values in the first stage of the test considering all the characteristic frequencies. However, in figure 6a it is possible to clearly identify an increasing trend starting after 3.7 days. This trend is not monotonically increasing but presents a decreasing trend between 5 and 5.5 days due to the characteristic propagation phenomena of the bearing fault [16].

The identification of the fault appearance can be obtained by designing a statistical threshold starting from the trend of the maximized values during the test. In this direction the values obtained are compared in figure 6 with a statistical threshold computed according to the Tukey's method applied on the values estimated in the first day of test, under the hypothesis that in this time span the system presents healthy condition. It is possible to observe in figure 6 that, accordingly to the experimental evidence on the physical system, only the trend of values related to the BPFO crosses the threshold after 3.8 days.

However, some values related to the BPFI and the BSF cross the respective threshold due to the dispersion of the BD criterion also in the healthy stage. This issue can be overcome by smoothing the values of the maximized criterion in order to make the data interpretation easier and to permit the identification of the time instant related to the fault appearance in a clearer way. The smoothed curves shown in figure 6 have been calculated exploiting the moving average technique implemented by convolving the BD final values series with a 15 samples length rectangular window. From the observation of the smoothed curves it is possible to note that only the trend related to the BPFO crosses the threshold and this behaviour confirms what has been detected experimentally on the studied system.

This application on a real case explains the sensitivity of the proposed BD criterion to the presence of cyclostationary patterns inside vibrational signals. This characteristic makes the FBBD algorithm suitable for the online monitoring of the bearing conditions during the overall operative life. A particular attention has to be pointed out on the number of cyclic frequencies considered; in fact the entire analysis has been carried out with the only cyclic frequencies α related to the specific fault, without consider any harmonics.

These results agree with the analysis proposed in Ref. [10] where the fault has been identified through the application of the CYCBD method also considering the harmonics of the cyclic frequency. The results prove

the effectiveness of the proposed method that is able to identify the fault appearance considering only the first cyclic frequency , although in this way the reconstruction quality of the estimated source may be not sufficient, as demonstrates in the previous section.

6 Final remarks

This paper presents a novel BD method based on the maximization of the cyclostationary behaviour of the signal. The proposed method exploits the Fourier-Bessel series expansion in order to rewrite the Indicator of Ciclostationarity and define a new BD criterion. The choice of the Fuorier-Bessel transform instead of the classic Fourier transform is related to the decay of the Bessel functions within the range defined by the length of the acquired signal. Thus, the impulsive signal can be reconstructed with a lower number of coefficient with respect to the Fourier transform based on sinusoidal functions.

The proposed algorithm, called FBBD, has been validated by means of both simulated and real signals in order to demonstrate the improvement given with respect to the other BD method based on cyclostationarity maximization, called CYCBD, in particular in term of lower computational time.

The application on simulated signal proves the lower number of cyclic frequencies required by the FBBD algorithm with respect to the CYCBD (70% less) according to the theoretical properties of the two different transform. The main issue is related to the inability of the FBBD to deconvolve the correct sign of the impulses; in fact the proposed method extracts an impulsive pattern composed by zero mean impulses, i.e. symmetric with respect to the zero, with the correct period and relative amplitude between the peaks.

The analysis of a run-to-failure test demonstrates the effectiveness of the proposed method for the continuous monitoring of the bearing conditions, in order to detect the fault appearance and to describe the progressive damage of the system. The proposed criterion proves to be very sensitive to the appearance of a cyclostationary source inside the signal, considering only the cyclic frequencies related to the fault, without its harmonics.

All these results demonstrate the clear improvement given by the proposed method that combines the suitability of the cyclostationarity for the description of vibrational signal related to bearing faults with the speed up of the iterative algorithm obtained through the exploitation of the Fourier-Bessel series expansion for the definition of the new BD criterion.

Acknowledgments

For the experimental validation presented in this work the dataset provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati has been used. Part of the analysis has been performed by using the code within Blind Deconvolution based on cyclostationarity maximization pack and provided by Marco Buzzoni.

References

[1] R.A. Wiggins *Minimum entropy deconvolution.* Geoexploration 16, n. 1-2, (1978), pp. 21-35.

[2] H. Endo, R.B. Randall *Enhancement of autoregressive model based gear tooth fault detection technique by the use of minimum entropy deconvolution filter.* Mechanical Systems and Signal Processing 21, (2007), pp. 906-919.

[3] L. Zhang, N. Hu *Fault diagnosis of sun gear based on continuous vibration separation and minimum entropy deconvolution.* Measurement 141, (2019), pp. 332-344.

[4] N. Sawalhi, R.B. Randall, H. Endo *The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis.* Mechanical Systems and Signal Processing 21, n. 6, (2007), pp. 2616-2633.

[5] G. McDonald, Q. Zhao, M. Zuo *Maximum correlated Kurtosis deconvolution and application on gear tooth chip fault detection.* Mechanical Systems and Signal Processing 33, (2012), pp. 237-255.

[6] G. McDonald, Q. Zhao *Multipoint Optimal Minimum Entropy Deconvolution and Convolution Fix: Application to vibration fault detection.* Mechanical Systems and Signal Processing 82, (2017), pp. 461-477.

[7] R.B. Randall, J. Antoni *Rolling element bearing diagnostics-A tutorial.* Mechanical Systems and Signal Processing 25, n. 2, (2011), pp. 485-520.

[8] J. Antoni *Cyclostationarity by examples.* Mechanical Systems and Signal Processing 23, n. 4, (2009), pp. 987-1036.

[9] M. Buzzoni, E. Soave, G. D'Elia, E. Mucchi, G. Dalpiaz *Development of an indicator for the assessment of damage level in rolling element bearings based on blind deconvolution methods.* Shock and Vibration 2018, (2018).

[10] M. Buzzoni, J. Antoni, G. D'Elia *Blind deconvolution based on cyclostationarity maximization and its application to fault identification.* Journal of Sound and Vibration 432, (2018), pp. 569-601.

[11] A. Raad, J. Antoni, M. Sidahmed *Indicators of cyclostationarity: Theory and application to gear fault monitoring.* Mechanical Systems and Signal Processing 22, (2008), pp. 574-587.

[12] R.B. Pachori, P. Sichar *A new technique to reduce cross terms in the Wigner distribution.* Digital Signal Processing 17, (2007), pp. 466-474.

[13] A. Bhattacharyya, L. Singh, R.B. Pachori *Fourier-Bessel series expansion based empirical wavelet transform for analysis of non stationary signals.* Digital Signal Processing 78, (2018), pp. 185-196.

[14] J. Schroeder *Signal processing via Fourier-Bessel series expansion.* Digital Signal Processing 3, (1993), pp. 112-124.

[15] J. Lee, H. Qiu, G. Yu, J.Lin *Rexnord Technical Services, Bearing Data Set, IMS, University of Cincinnati, NASA Ames Prognostics Data Repository.* , 2007. url: http://ti.arc.nasa.gov/project/prognostic-data-repository.

[16] T. Williams, X. Ribadeneira, S. Billington, T.Kurfess *Rolling element bearing diagnostics in run-to-failure lifetime testing.* Mechanical Systems and Signal Processing 15, (2001), pp. 979-993.

Combination of vibration analysis and Acoustic Emission measurements to better characterize damage and mechanical behaviour of aerospace high speed rolling bearing

Yoann HEBRARD[1]

[1] SKF Aerospace, 22 rue Brillat Savarin, 26958 Valence, FRANCE
yoann.hebrard@skf.com

Abstract

Designed to break the paradigm for efficiency, the new generation of engines promises double-digit reductions in fuel burn, as well as an unparalleled single-leap improvement in emissions and lower noise to fulfil societal environmental objectives for a more sustainable future. The end-use consumer and environmental policy requirements for aircrafts of the next generation translate into components with higher temperature and speed. Furthermore, new monitoring technics are needed to closely monitor rolling contact during testing of the next generation of aero engine bearing to check its behavior under the new application condition. Vibration analysis for condition assessment and fault diagnostics is widely used nevertheless interpretation and correlation of collected data is often cumbersome.
That is why combination of vibration analysis and acoustic emission techniques giving different types information in different frequency band can help to understand the behavior of new rolling bearing. This paper presents an experimental testing campaign on bearings with seeded defects. Correlation between low and high frequency signals with different strategy of signal acquisition are presented from signal processing step.
Real time transient analysis with feature extraction was done in parallel with streaming acquisition on both signals. Pattern recognition of individual AE signal is possible and were correlated with more traditional analysis based on signal enveloping vibration analysis. Continuous monitoring on finished duration were done to provide information on no stationary regime and also time of stabilization. Comparison between features extraction is done on damaged and defect free rolling bearing at several rotating speed and loading level.

Keywords: Rolling contact monitoring, vibration monitoring, EHL conditions

1 Introduction

Many studies are dealing about the use of vibration to detect fault in gear box and rolling bearings. Some of them are focusing on the use of Acoustic Emission (AE) and vibration for better characterization of the gear box default type. Based on vibration technology the acquisition of raw signal is done by a partial acquisition of the signal at random. AE technology is more focusing on the detection of transient above a predefined threshold in a narrow bandwidth [1,2]. According the progress of the acquisition system this study proposes a combination of all these types of acquisition with wide band sensors in parallel to offer richer bunch of data that allow us to investigate new method of processing and default characterization. The fault Characterization we propose will be more than a statistic acquisition but a continuous monitoring.

Acoustic emissions (AE) are defined as transient elastic waves generated from a rapid release of strain energy caused by a deformation or damage within or on the surface of a material [4]. This technique is widely used as a non-destructive testing technique for fitness for service evaluation in industrial field. AE is also a powerful tool to characterize and understand damage initiation and propagation. Most of all microscopic mechanisms has been studied and correlated with AE signals as fretting [3]. Many developments in AE technology, mainly developments in AE instrumentation, have occurred in the past ten years.

In this investigation, AE's are defined as the transient elastic waves generated by the interaction of two surfaces in relative motion. The interaction of surface asperities and impingement of the bearing rolling elements over the seeded defect on the outer race will generate AE's. Due to the high frequency content of the AE signatures typical mechanical noise (less than 20kHz) is eliminated.

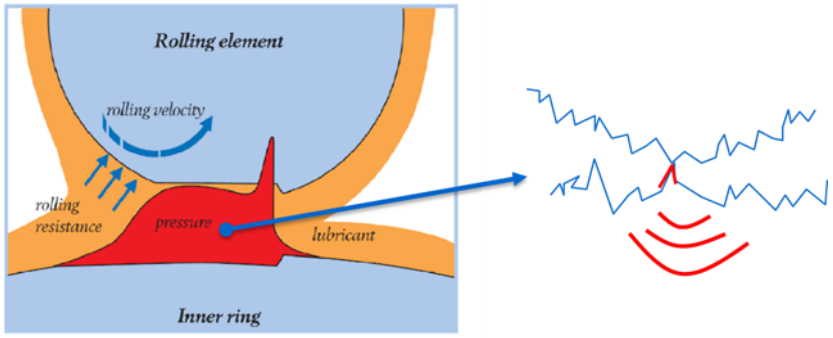


Figure 1. AE signal from asperity in rolling contact

There have been numerous investigations reported on applying AE to bearing defect diagnosis. Roger [5] utilised the AE technique for monitoring slow rotating anti-friction slew bearings on cranes employed for gas production. In addition, successful applications of AE to bearing diagnosis for extremely slow rotational speeds have been reported [6, 7]. Yoshioka and Fujiwara [8, 9] have shown that selected AE parameters identified bearing defects before they appeared in the vibration acceleration range. Hawman et al [10] reinforced Yoshioka's observation and noted that diagnosis of defect bearings was accomplished due to modulation of high frequency AE bursts at the outer race defect frequency. The modulation of AE signatures at bearing defect frequencies has also been observed by other researchers [11, 12, 13]. Morhain et al [14] showed successful application of AE to monitoring split bearings with seeded defects on the inner and outer races.

This paper investigates the relationship between AE signal for a range of defect conditions, offering a more comparative study than is presently available in the public domain. Moreover, comparisons with vibration analysis are presented. The source of AE from seeded defects on bearings, which has not been investigated to date, is presented showing conclusively that the dominant AE source mechanism for defect conditions is asperity contact.

2 Experimental setup

2.1 General

The bearing test rig employed for this study had an operational speed range of 5000 to 15000 rpm with a maximum load capability of 50kN via a hydraulic ram. The test bearing employed was 3 points contact ball bearing. This bearing type was selected as it allowed defects to be seeded onto the races, furthermore, assembly and disassembly of the bearing was accomplished with minimum disruption to the test sequence. Five calibrated dents were done on the bearing inner ring (rotating). Pure axial loading is applied to the rolling bearing. The defects are located along the predicted rolling raceway path inside the hertz contact zone. Dent length against the rolling velocity are around 200µm. The acquisition is performed via a multichannel last generation Mistras AE acquisition system: Express 8.

All capability of this system was used to record the maximum of information. First, we use continuous energy summation without threshold is much accurate than traditional RMS or ASL integration. Then, acquisition of transients based on smart threshold allows us to avoid triggering on continuous signal and can guaranty a significant hits rate whatever the level of background continuous is without any manual change all along the monitoring. Transient is characterized by hit and waveform, it can be feed in Noesis our pattern recognition software for multiparametric evaluation. Express offer also the capability to record streaming

(acquisition of row signal at very rate and with quasi unlimited buffer) to apply more traditional signal processing often use in vibration.

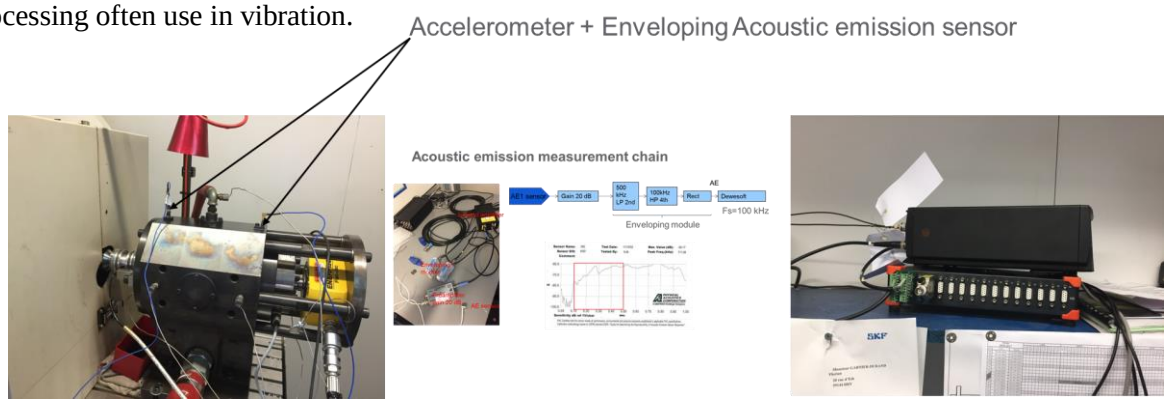


Figure 2 Experimental set-up

In parallel, vibration and AE enveloped signals were acquired on Dewesoft Sirius acquisition station. The next table presents the acquisition characteristics. FFT was computed from the enveloped signal.

Sensor type	Sampling frequency	Enveloped bandwidth
B&K Mistras WD AE sensor	200 kHz 2 gHz	5-40 kHz 100-500 kHz

Table 1 Acquisition characteristics

2.2 Acoustic emission acquisition setting

Mystras Express8 system also offers the capability to record streaming (acquisition of row signal at very rate and with quasi unlimited buffer) to apply more traditional signal processing often use in vibration.

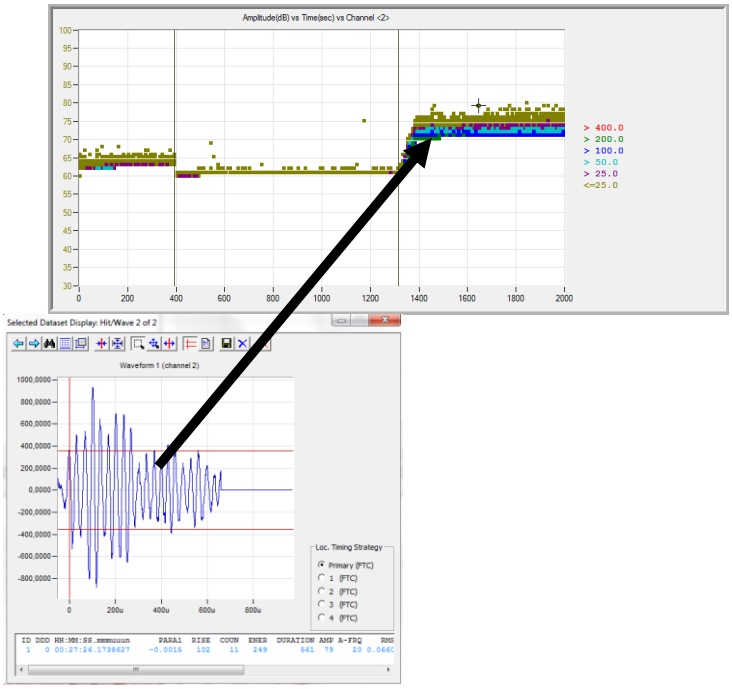


Figure 3 acquisition with smart threshold (above graph) and waveform and hit correspondence (graph below).

The streaming has been also used it parallel without any reduction of performance of traditional AE acquisition. Streaming is a synchronized acquisition the 4 channels without threshold, the windows length can be arbitrary of defined like shown in figure 5.

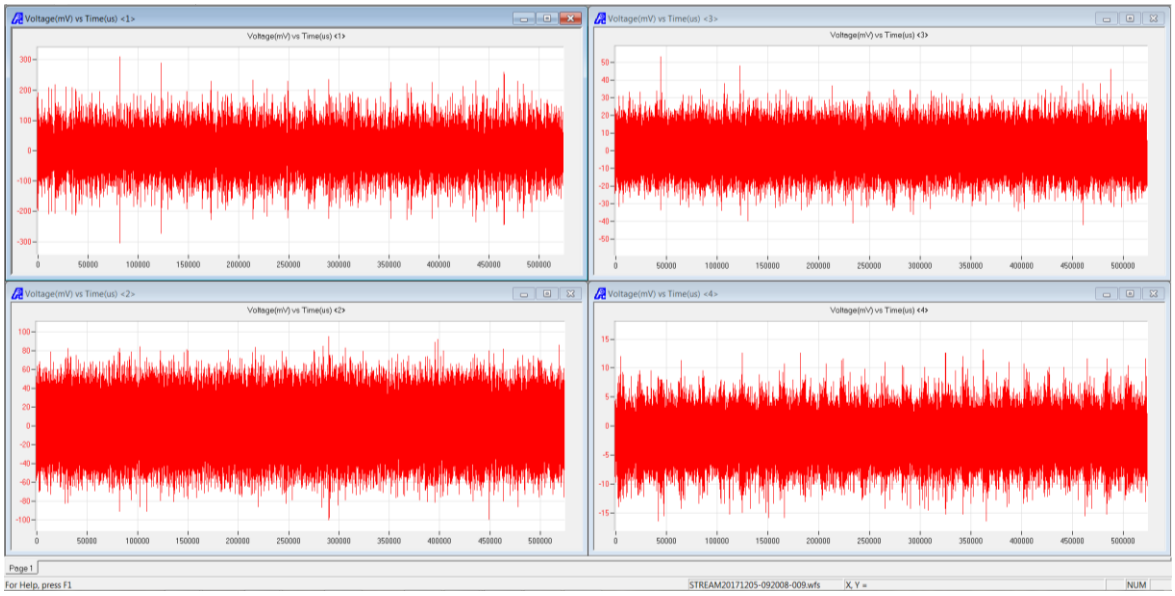
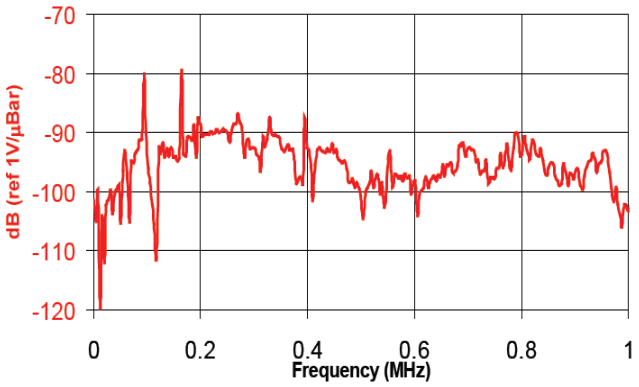
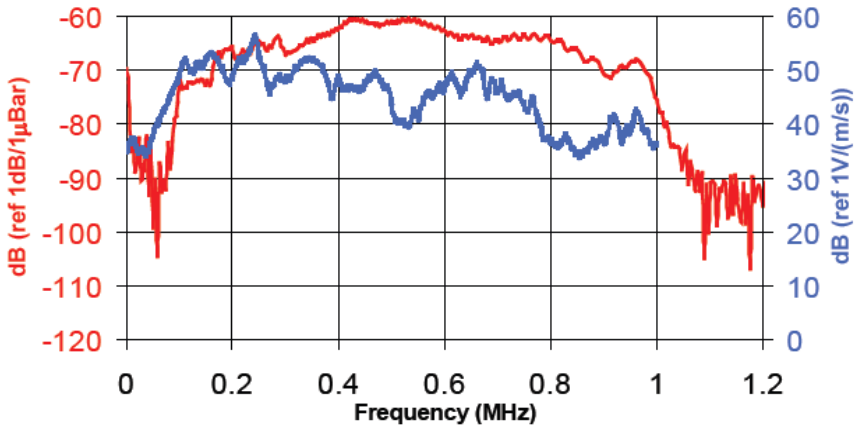


Figure 4 streaming acquisition during 5 sec at 2 Mega sample per second on the four channels.

Three different wideband AE sensors (WD, S9208 and micro 80) are used in parallel with a standard accelerometer Bruel et Kjaer type 4374 (bandwidth 1 to 26KHz 0,5 pC ms-2). The calibration curve of the sensors are given in figure 5.



Frequency response of S9208



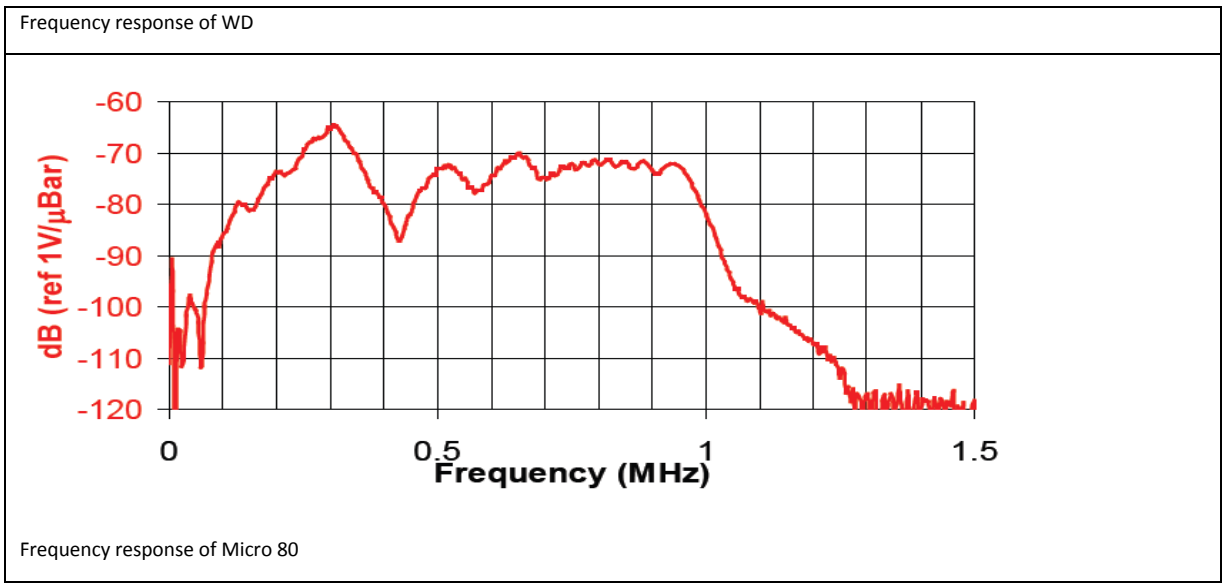


Figure 5 calibration curve of sensors S9203, WD and micro 80.

The sensor must be held in place for the duration of the test. Dry contact between the sensor and the structure does not meet the goal for appropriate wave transmission. For the AE sensors coupling, we used an adhesive tape developed for aerospace industry called “KAPTON” on which the sensor is glued with cyanoacrylate or cement glue. (figure 3). This solution has been tested with success with cyanoacrylate glue for test space telescope for Astrium under high energetic vibration [15]. The reduction of measured amplitude compare to a traditional grease coupling is less than one dB for Hsu Nielsen source (NFEN1330-9).

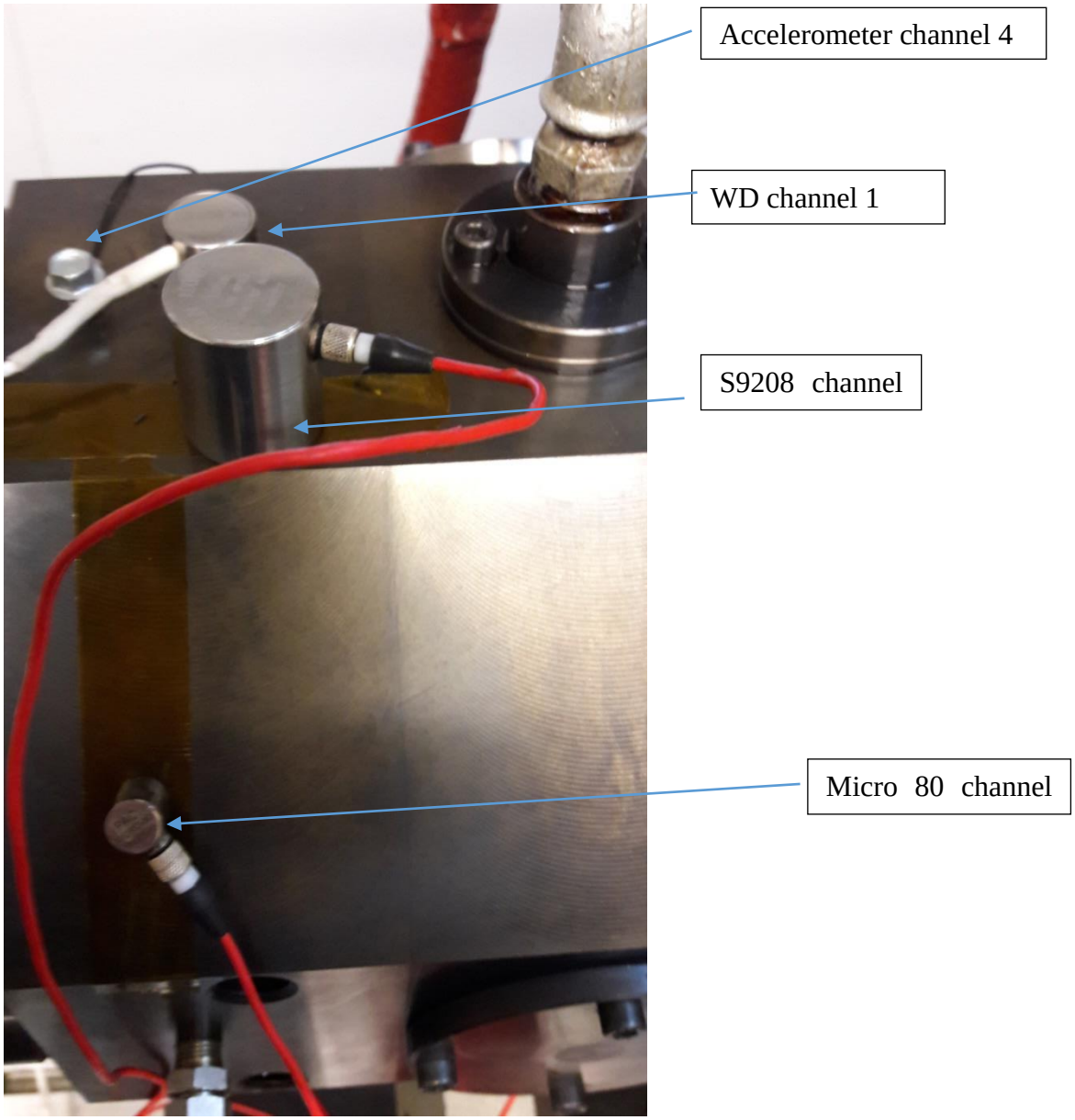


Figure 6 picture of the different sensors used for the monitoring and coupling mode.

3 Introduction

We performed the same loading sequence for all bearings, see table 2.

Speed (Rpm)	Load (kN)
3000	5
6000	5
11500	5
11500	10
11500	**15**
11500	**20**
11500	**25**
6000	**23,5**

Table 2 loading sequence for bearing the colors are used to separate the loading sequences.

The four sensors are compare and the WD (channel 1) and micro 80 (channel 2) are giving better results than accelerometer and S9209.

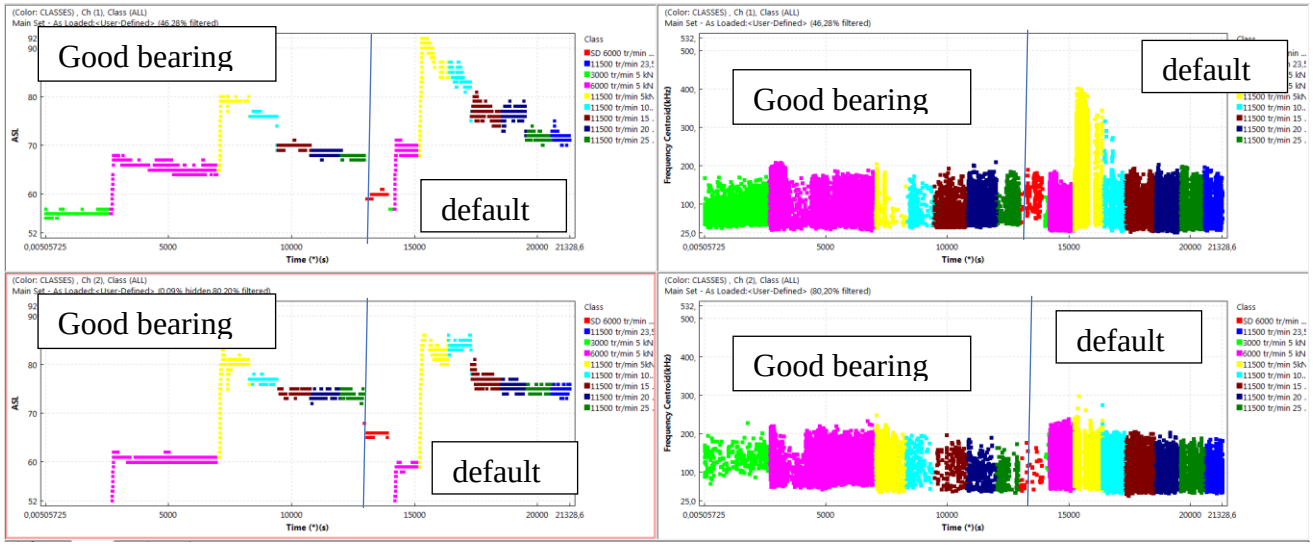


Figure 7 evolution of ASL and frequency centroid versus time with and without default. On the top graph channel 1 (WD) at the bottom channel 2 (Micro 80).

A clear difference of behavior can be seen above 6000 rpm at 5 kN and it enhance at 11 kN on the energy of the AE signal on transients. Also, the center mass of the frequency spectrum (called frequency centroïde) increases a lot for the bearing with default it exhibits the best power of discrimination using our patter recognition software. The stabilization of the AE signal takes at least 2 minutes after the loading condition change.

Considering a more standard acquisition mode for channel 1 (WD), on the bearing with default, it can be seen a new pic at 27 kHz at 11500 tr/min and above speed. For another side, the width of the FFT increases as the load increases (fig. 8).

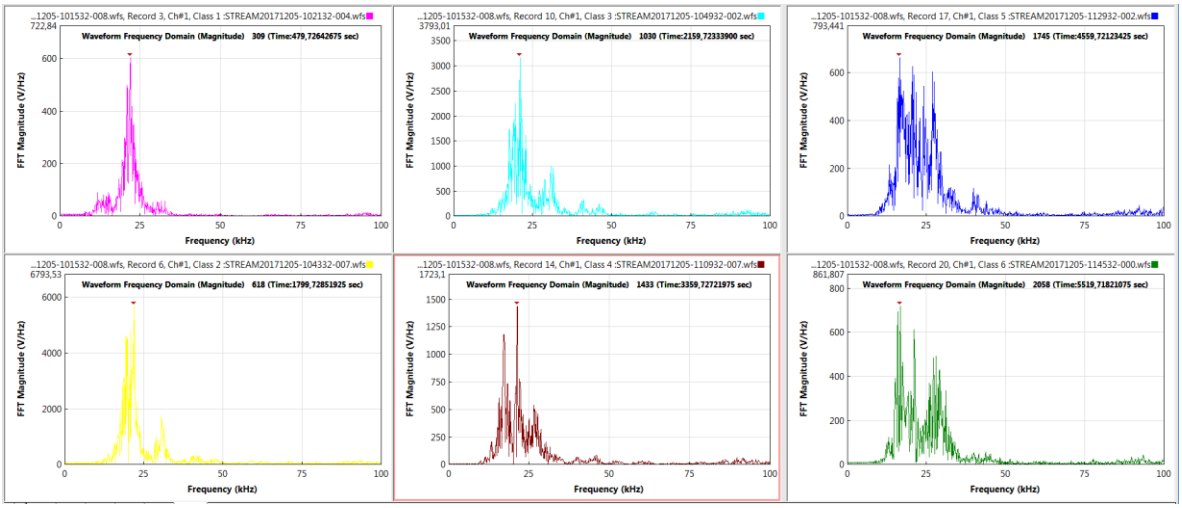


Figure 8 Fourier transform on 5s files on channel 1 for different case of load levels and rotating speed.

Fourier Transform from Enveloped Spectrum from enveloped AE and vibration Signal were computed in real time from the Sirius acquisition station. Bearing and shaft frequencies were identified from the signal.

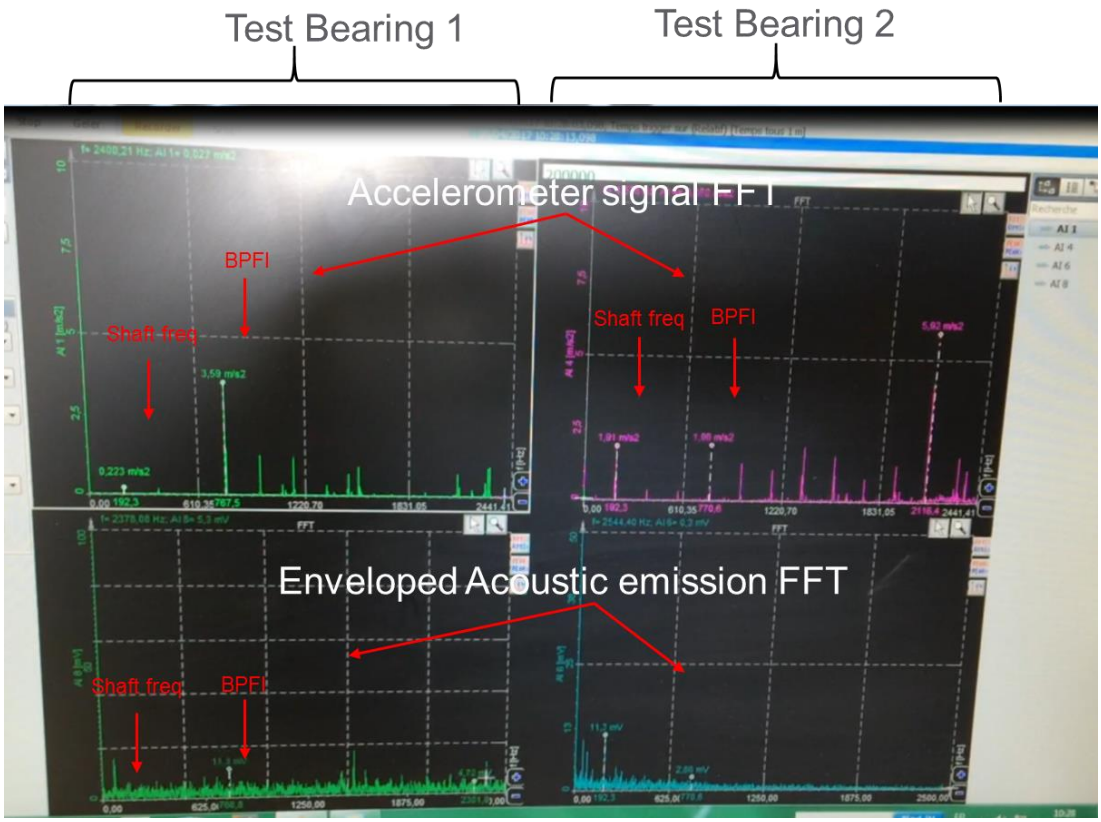


Figure 9 Enveloped spectrum with bearing frequencies identified

4 Discussion

The source of AE for seeded defects is attributed to material protrusions above the surface roughness of the outer race. This was established as the smooth defect could not be distinguished from the no-defect condition. However, for all other defects where the material protruded above the surface roughness, AE transients associated with the defect frequency were observed. As the defect size increased, AE r.ms, maximum amplitude and kurtosis values increased, however, observations of corresponding parameters from vibration measurements were disappointing. Although the vibration r.m.s and maximum amplitude values did show changes with defect condition, the rate of such changes highlighted the greater sensitivity of the AE technique to early defect detection.

Again, unlike vibration measurements, the AE transient bursts could be related to the defect source whilst the frequency spectrum of vibration readings failed in the majority of cases to identify the defect frequency or source. Also evident from this investigation is that AE levels increase with increasing speed and load. It should be noted that further signal processing could be applied to the vibration data in an attempt to enhance defect detection.

Techniques such as demodulation, band pass filtering, etc, could be applied though these were not employed for this particular investigation. The main reason for not applying further signal processing to the vibration data was to allow a direct comparison between the acquired AE and vibration signature. From the results presented two important features were noted:

- firstly, AE was more sensitive than vibration to variation in defect size
- secondly, that no further analysis of the AE response was required in relating the defect source to the AE response, which was not the case for vibration signatures.

The relationship between defect size and AE burst duration is a significant finding. In the longer term, and with further research, this offers opportunities for prognosis. AE burst duration was directly correlated to the seeded defect length (along the race in the direction of the rolling action) whilst the ratio of burst amplitude to the underlying operational noise levels was directly proportional to the seeded defect width.

5 Conclusion

It has been shown that the fundamental source of AE in seeded defect tests was due to material protrusions above the mean surface roughness. Also, AE maximum amplitude has been shown to be more sensitive to the onset and growth of defects than vibration measurements.

A relationship between the AE burst duration and the defect length will be established in further posttreatment.

6 References

[1] Raad. A. Sidahmed, B; Early detection of the gear fault – advantage of the acoustic emission technique 25th European conference, Acoustic emission testing; EWGAE; 2002; Prague; 2; II/125-II/130

[2] Lekou, D.J., Mouzakis, F., Anastasopoulos, A. A., Kourousis, D., Fused Acoustic Emission and Vibration Techniques for Health Monitoring of Wind Turbine Gearboxes and Bearings; Scientific Proceedings of EWEC 2009, pp. 78-82., At Marseille

[3] Boinet, M. Proust, A., Lenain, J.C., Yantio Njankeu Sabeya, G.R., Paris, J.-Y., Denape, J., Analyse par Emission Acoustique de la dégradation d'un contact sollicité en fretting, Journées Francophone de Tribologie Poitiers, 22–23 mai 2007.

[4] Mathews, J. R. Acoustic emission, Gordon and Breach Science Publishers Inc., New York. 1983, ISSN 0730-7152.

[5] Roger, L. M., The application of vibration analysis and acoustic emission source location to on-line condition monitoring of anti-friction bearings. Tribology International, 1979; 51-59.

[6] Mba, D., Bannister, R.H., and Findlay, G.E. Condition monitoring of low-speed rotating machinery using stress waves: Part's I and II. Proceedings of the Instn Mech Engr 1999; 213(E): 153-185

[7] N. Jamaludin, Dr. D. Mba, Dr. R. H. Bannister Condition monitoring of slow-speed

rolling element bearings using stress waves. Journal of Process Mechanical Engineering, I Mech E. Pro. Inst. Mech Eng., 2001, 215(E):, Issue E4, 245-271.

[8] Yoshioka T, Fujiwara T. New acoustic emission source locating system for the study of rolling contact fatigue, Wear, 81(1), 183-186.

[9] Yoshioka T, Fujiwara T. Application of acoustic emission technique to detection of rolling bearing failure, American Society of Mechanical Engineers, Production Engineering Division publication PED, 1984, 14, 55-76.

[10] Hawman, M. W., Galinaitis, W. S, Acoustic Emission monitoring of rolling element bearings, Proceedings of the IEEE, Ultrasonics symposium, 1988, 885-889

[11] Holroyd, T.J. and Randall, N., (1993), Use of Acoustic Emission for Machine Condition Monitoring, British Journal of Non-Destructive Testing, 1993, 35(2), 75-78.

[12] Holroyd, T. Condition monitoring of very slowly rotating machinery using AE techniques. 14th International congress on Condition monitoring and Diagnostic engineering management (COMADEM'2001), Manchester, UK, 4-6 September 2001, 29, ISBN 0080440363

[13] Bagnoli, S., Capitani, R. and Citti, P. Comparison of accelerometer and acoustic emission signals as diagnostic tools in assessing bearing. Proceedings of 2nd International Conference on Condition Monitoring, London, UK, May 1988, 117-125.

[14] Morhain, A, Mba, D, Bearing defect diagnosis and acoustic emission Journal of Engineering Tribology, I Mech E, Vol 217, No. 4, Part J, p 257-272, 2003. ISSN 1350-6501.

[15] Levalllois, F., Sobeczko, A., Proust, A., Marlot, D., Lenain J.C. Non-destructive testing of GAIA frame by means of acoustic emission monitoring during launch simulation tests., 30th European Conference

on Acoustic Emission Testing & 7th International Conference on Acoustic Emission University of Granada, 12-15 September 2012.

Neuroevolution for Bearing Diagnosis

Rita SLEIMAN[1], Amani RAAD[1], Souhayb KASS[2], Jerome ANTONI[2]
[1]Univ Libanaise, Faculté de Génie, Tripoli, LIBAN
[2]Univ Lyon, INSA LYON, LVA, Lyon, France

Abstract

The monitoring of machinery and especially ubiquitous bearings in all means of transport has gained importance for decades in the industry because of the need to increase the reliability of machines and reduce the possible loss of production due to failures caused by the different faults. Many of the available techniques currently require a lot of expertise to apply them successfully. New techniques are required that allow relatively unqualified operators to make reliable decisions without knowing the mechanism of the system and analyzing the data. Reliability must be the most important criterion of the operation. Artificial intelligence is the revolutionary answer in all areas of industrial control. The main goal of this paper is to propose new solutions for bearing diagnosis based on deep neural networks (DNN). However, in general the optimization of the neural network architecture is done by trial and error, and the features reduction problem is solved by using the principal component analysis. In this paper, the application of the neuro-evolution is proposed for bearing diagnosis where the optimization of the neural network topology as well as the features reduction are done by an evolutionary genetic algorithm. An application of the general procedure is proposed for real signals; that shows the superiority of the combination between neural networks and genetic algorithms for bearing diagnosis.

1 Introduction

Rolling element bearing is one of the most critical components used in rotating machinery and many other mechanical equipment [1]. In fact, most of such machines' malfunctions are linked to bearing faults, such as fatigue, corrosion, overload, etc, that may occur unexpectedly if no predictive maintenance is used. This may lead to significant economic loss: high costs of maintenance and loss of revenue [2]. Therefore, bearing state monitoring and fault diagnosis are very important for discovering early bearing faults, assuring efficient and safe operation of all machines containing bearings.

In general, all bearing condition monitoring approaches in the literature can be classified into two categories: statistical-based approaches and pattern recognition-based approaches. In statistical-based approach, various signal processing tools are used, followed by statistical thresholds to detect the presence of a fault as well as to classify the different types of bearing faults [3,4]. For pattern-recognition-based approaches, several machine learning and artificial intelligence techniques, such as Artificial Neural Networks ANN, Support Vector Machine, fuzzy Expert Systems, Random Forest, and many other, have been successfully employed in fault diagnosis [5,6]. More recently, deep learning algorithms, such as deep neural networks, convolutional neural networks and deep belief networks have shown great capabilities in the field of computer vision [7], speech recognition [8] and natural language processing [9], due to their ability to discover hidden patterns in the data by using architectures composed of several non-linear learning layers. These deep learning algorithms were also applied in the field of industrial diagnosis and have been very useful and effective [10,11].

In recent years, a new artificial intelligence approach known as the 'Neuro-evolution' has attracted considerable attention as they proved to be essential in so many applications. The basic idea is that it applies evolutionary algorithms, and more specifically genetic algorithms GA, in order to construct a well suited artificial neural network for a certain application. Earlier successful applications in the field of Neuro-evolution are in reinforcement learning, evolutionary robotics, and artificial life. Sample applications include evolving behaviors for video games such as evolving new content in real time while the game is played [12],

controlling mobile robots such as evolving the neural networks of robots that were 3D-printed and could move around the real world [13], and investigating the evolution of biologically-relevant behaviors such as investigating abstract evolutionary tendencies, like the evolution of modularity or how biological development interacts with evolution [14].
Meanwhile, in order to benefit from such powerful tool of artificial intelligence in the field of industrial diagnosis, researchers in this domain have tried to apply this combination of neural networks and genetic algorithm on bearing fault classification problem: some have used the genetic algorithm for the weights optimization [15], and others for the features selection problem [16].
The main goal of this paper is to apply this concept of Neuro-evolution in industrial automatic diagnosis without any human intervention, especially after the fourth industrial revolution characterized by the fusion of all modern technologies and the concept of digital factories [17]. In details, the optimization of the number of hidden layers and nodes in an artificial neural network is performed using this concept; in order to guarantee ANN architecture with the highest classification accuracy. In addition, feature reduction is obtained using the genetic algorithm instead of the Principal Component Analysis (PCA) [18], since it needs the tuning of some parameters whereas the GA gives good results without any assumptions.
The main benefit behind this concept is to extend the optimization of the topology as well as the features from one dataset to another without any human intervention to find automatically the best classification accuracy; which could not be obtained using neural networks alone. The extracted features are divided into temporal classical and spectral ones. The spectral features are based on the indicators of probability of presence of faults introduced by S. Kass and Al based on the spectral coherence [19]. These indicators are very powerful due to their ability of condensing the whole information initially displayed in three dimensions into a scalar and returning information in terms of a probability of presence of a fault. They also take into consideration uncertainties in the bearing characteristic frequencies, which is crucial in bearing diagnosis. However, a statistical threshold was derived for decision making. In this paper, this statistical threshold is absent and replaced by the techniques of artificial intelligence already mentioned above, and that will be detailed in the next sections.
This paper is divided into seven sections, where section two introduces briefly the artificial neural networks and the third section describes the genetic algorithm. Section four presents in a general way the effectiveness of the combination between neural networks and genetic algorithm, and section five describes the application done in this paper. In section six, the features used as input to the neural network are presented and described and the proposed method is validated using two datasets in section seven. Conclusions are drawn in the last section.

2 Neural Network

Inspired by the human brain, Artificial Neural Networks (ANN) are a family of machine learning models that mimic the structural elegance of the neural system and learn patterns inherent in observations [20]. There are several types of neural networks: Backpropagation networks, Deep Belief networks, Convolutional neural networks, Recurrent networks, Radial Basis Function networks, etc…
One of the most widely used type of neural networks is the Backpropagation network which is a multi-layer perceptron consisting of an input layer with nodes representing input variables to the problem, an output layer with nodes representing the different classes of the corresponding classification problem, and one or more hidden layers containing nodes to help discovering the hidden patterns in the data.
While, in theory, it is possible to apply different types of activation functions for different layers, it is common to apply the same type of an activation function for the hidden layers in the literature [21]. However, it should be a nonlinear function such as the 'logistic sigmoid' function $\sigma(z) = 1/(1+\exp[-z])$, and the 'hyperbolic tangent' function $\tanh(z) = (\exp[z]-\exp[-z]) / (\exp[z]+\exp[-z])$.
In addition, the connections between different layers are weighted: These weights are initially randomly set, and then adjusted between successive training cycles (learning process) in order to increase the classification accuracy. This is generally done by minimizing a cost function using several well-known optimization algorithms: the gradient descent which is the simplest and most popular training algorithm [22], the

stochastic gradient descent which is a lighter algorithm and therefore faster than its all-encompassing cousin [23], the Adaptive Moment, *Adam*, which is based on adaptive estimates of lower-order moments [24], etc.
For K-class classification, it is common to use a cross-entropy cost function defined as follows:

$$E(\omega) = -\frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{K} t_{nk} ln y_{nk} \quad (1)$$

where ω is the weights set, t_{kn} denotes the *kth* element of the target vector t_n and y_{kn} is the *kth* element of the prediction vector y_n for x_n, given a training data set $\{x_n,t_n\}$, n=1…N.

3 Genetic Algorithm

Genetic algorithm (GA) was first introduced by John Holland, from the University of Michigan, in 1975 in its publication '*Adaptation in Natural and Artificial System*' [25]. It is a general-purpose optimization algorithm that is, like neural networks, a bio-inspired artificial intelligence tool. It is a particular class of evolutionary algorithms that are based on the mechanics of natural selection and natural genetics. Unlike other optimization algorithms that improve a single solution at a time, GA uses a strategy of parallel search by working on a population of candidate solutions (also called individuals) to an optimization problem that evolves toward better solutions, enabling extreme exploration and massive parallelization. The basic idea is that over time, evolution will select the most suitable solutions [26]. This is done by evaluating, for each individual in a generation, the fitness function: the higher the fitness, the most probably the individual will be selected.
Technically, when solving an optimization problem using the GA, one must first define:

- Individual (or chromosome): Composed of several genes, it can be binary or real encoded. An individual represents one possible solution of the problem; the collection of multiple chromosomes forms the population which represents a subset of the whole searching space.
- Fitness function: It is the function we tend to maximize; if the problem consists in minimizing a certain function, the latter should be transformed into a fitness function by simply inverting it. The fitness function corresponds to an evaluation of how good the candidate solution is.

Once the fitness function is properly defined, the genetic algorithm generally starts by randomly generating the initial population. It should be large enough so that any solution in the search space can be later engendered [27]. Then, the algorithm loops over an iteration process to make the population evolve. Each iteration consists of the following genetic operators:

- Selection: After evaluating the fitness f(i) for each individual i, Roulette wheel selection is applied in order to give the fittest individuals a higher chance to be selected than weaker ones. This is done by calculating the probability of selection of each individual as follows:

$$p(j) = \frac{f(j)}{\sum_{i=1}^{N} f(i)}; j = 1, \dots, N \quad (2)$$

  where N is the total number of individuals in the population.
- Crossover: Selected individuals should be paired randomly, with a crossover rate Pc, for recombination. The latter is done by exchanging genes between one or more crossover points that are randomly generated.

- Mutation: It randomly alters one or more genes in a chromosome, with a mutation rate Pm, changing it from 1 to 0 and vice versa. It is a powerful operator used to avoid falling into a local optimum.

- Replacement: The new selected, recombined and probably mutated individuals form a new population that replaces the old one.

Commonly, the algorithm terminates when either a maximum number of generations has been produced, or a satisfactory fitness level has been reached for the population. When the algorithm terminates, the individual with the highest fitness is regard as the approximate optimal solution.

4 Combination of neural networks with genetic algorithm

Generally, when constructing an ANN classification model, choosing its various parameters, such as its topology, activation functions, training sample size, etc, may greatly impact the classification results [28]. Basically, one of the most critical tasks in artificial neural network design is choosing the best topology (architecture) that gives the highest classification accuracy. One starts with no prior knowledge as to the number of hidden layers and number of hidden nodes required [20]; Choosing a small number of hidden layers and nodes will lead to an 'underfitting' problem: ANN will not be able to reveal complex and hidden patterns in the data. In contrast, a network with too many hidden nodes tries to model exactly the training dataset following all its noise, and leading to a poor generalization for additional untrained data: this is known as the 'overfitting' problem. In the literature, several researchers have proposed different methodologies for fixing the number of hidden neurons. Most of the methodologies are presented in a review where the authors also proposed a new method to fix the hidden neurons in Elman networks for wind speed prediction in renewable energy systems [29]. In general, the most used technique for finding the optimal architecture is by trial and error. The latter approach has several limitations such as it is time-consuming and the obtained network structure may not be optimal.

In recent years, new topology optimization techniques based on evolutionary algorithms have gained great interest by the researchers in the domain of artificial intelligence. The combination between artificial neural network and genetic algorithm, also known as 'Neuro-evolution', has shown important capabilities and effectiveness in so many fields as it meets the potential of the increasing high performance computation capabilities in our days [30]. Initially, the use of Neuro-evolution was only restricted to the optimization of neural networks weights in order to overcome some limitations of the backpropagation algorithm. Later, this combination was extended to optimize also the ANN architecture. The following flowchart describes the principal steps of the ANN architecture optimization algorithm.

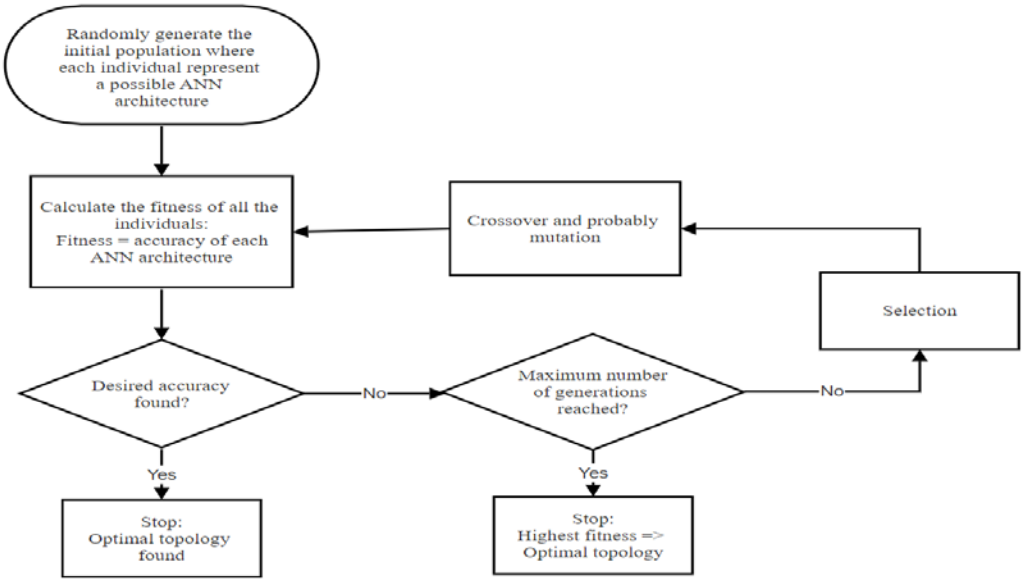
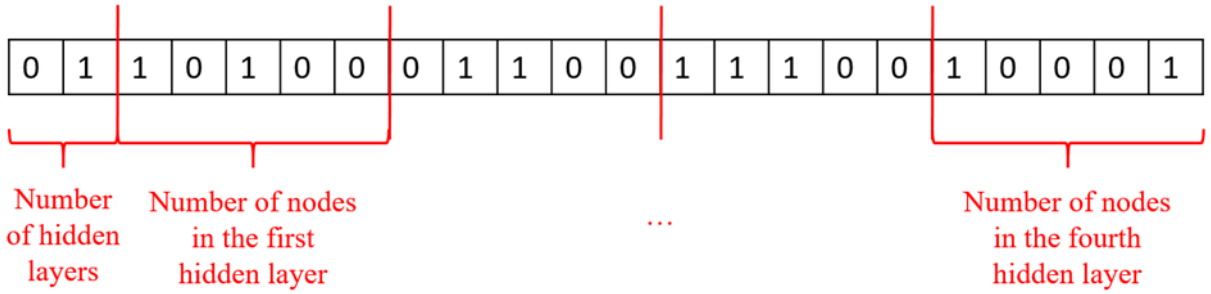


Fig.1 Flowchart of the ANN architecture optimization algorithm

5 Application

As mentioned in section “GA”, in order to apply the GA for any optimization problem, one must first define the encoding way of an individual: the process of representing individual genes. We may have binary or real encoding; The encoding depends mainly on the application. The most common way of encoding is binary strings, which is used in this paper.

For the architecture optimization problem, the chromosome is a binary string composed of 22 bits (genes). The first two bits are reserved for the binary representation of the number of hidden layers, while the other 20 bits are devided into four parts, each of five bits, representing the binary representation of the number of nodes in each hidden layer. This results in a number of hidden layers varying between 1 and 4 layers, and a number of nodes in a layer varying between 1 and 32 nodes. Such chromosome can be illustrated as follows:



On the other hand, such optimization problem tends to maximize the classification accuracy. Therefore, the fitness function must be directly linked to it. The only difference is that a mapping is used in a way to make 90% accuracy worth 10% fitness and 100% accuracy worth 100% fitness. The algorithm that illustrates this idea is as follows:

If accuracy<90%:

Fitness=10%

Else:

*Fitness=((Accuracy-90%)/(100%-90%)) *(100%-10%) + 10%*

The main reason for this mapping is the use of the ‘Roulette wheel’ selection technique that calculate the probability of selection based on the fitness values. Accordingly, mapping is essential in order to ensure that probabilities of selection in a population are not so close in the selection process.

For the features reduction problem, the total algorithm is the same, but with a single difference which is the way of representing an individual. Here, each chromosome in the population is composed of n bits, where n is the total number of features. The value of each bit can be either 1, which indicates the presence of the corresponding feature, or 0, which indicates its absence. In this case, the chromosome can be illustrated as follows:

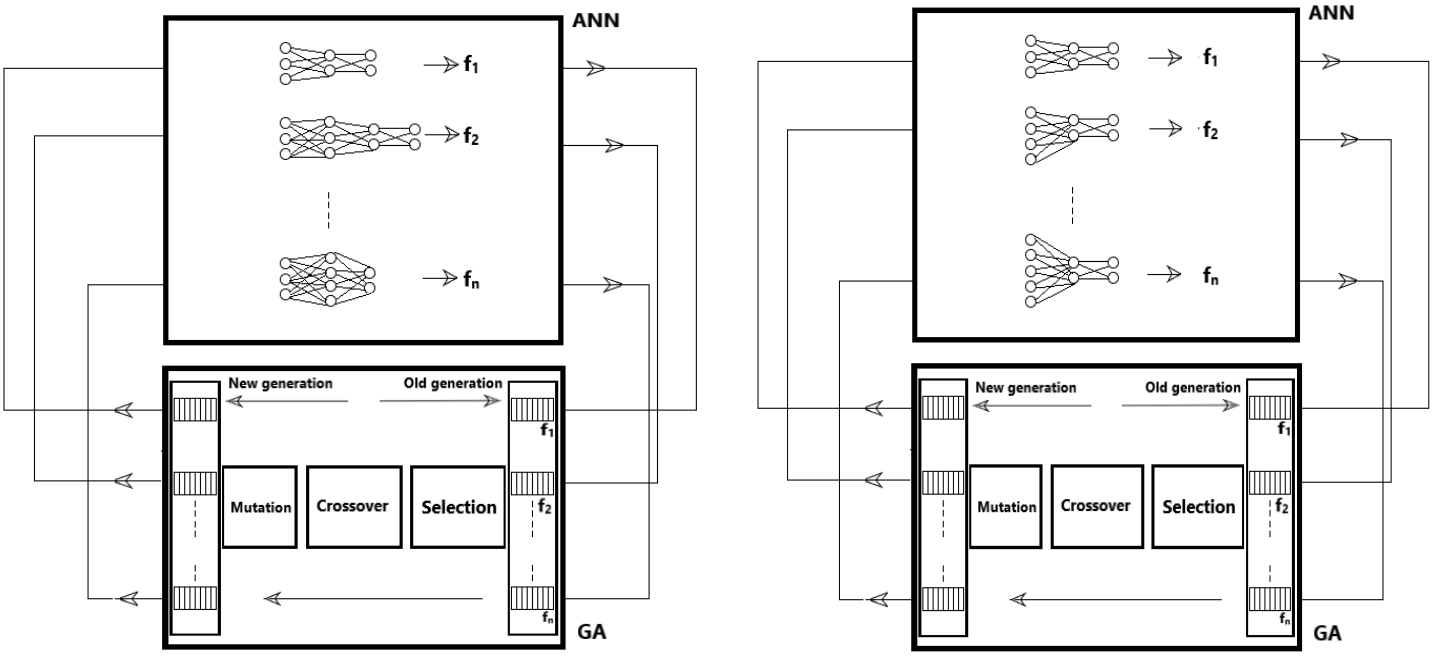


Fig.2 Combination between ANN and GA in two cases of ANN architecture and Feature optimization.

6 Features

6.1 Classical Features

Feature extraction is a very important step in a bearing diagnosis problem. Any feature chosen to be extracted from bearing signals will directly impact the classification results. Therefore, one must properly select the best feature set before moving to the next step of choosing the best classifier.

For detecting the change in bearing signal, traditional statistic features can be used. The advantages of using these features is essentially the ease of implementation and the low computational time. Accordingly, this paper proposes the use of traditional time-domain features presented in the following table:

RMS	$\left(\frac{1}{N}\right)\sum_{i=1}^{N} X_i^2^{\frac{1}{2}}$
Kurtosis	$\frac{1}{N}\sum_{i=1}^{N}\frac{(x_i-\bar{x})^4}{\sigma^4}$
Peak to peak	$x_{max}-x_{min}$
Crest Factor	$\frac{\max\lvert x_i\rvert}{RMS}$
Shape factor	$\frac{RMS}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$
Impulse factor	$\frac{\max\lvert x_i\rvert}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$
Margin factor	$\frac{\max\lvert x_i\rvert}{(\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert^{\frac{1}{2}})^2}$

Tab.1 Temporal features

Along with the above traditional features, extra powerful features for bearing fault classification, based on the second order of cyclostationarity, are used in this paper. These features are deduced from a recently developed indicator that will be detailed in the next section.

6.2 Spectral features

Souhayb and al [19] introduces a new autonomous method of bearing diagnosis in the case where the fault characteristic orders are known, taking into account all factors that may hinder this diagnosis. It is based on the development of new scalar indicators, which can be interpreted in terms of the probability of the presence of a fault. These indicators result from a post-processing of spectral coherence, calculated using the fast version of the spectral coherence algorithm proposed by Antoni et al [31]. It was chosen as a basis for the indicator because it is considered to be the optimal three-dimensional representation in which the bearing failure manifests itself clearly despite extreme operating situations.

In details, the envelope spectrum is calculated, as a first step, by integrating the spectral coherence according to the frequency variable. Therefore, all information initially displayed in three dimensions will be condensed into two dimensions. The next step is to recognize the characteristic peaks of the defects. This recognition must be confirmed by checking the presence of these peaks (theoretical fault frequency and its harmonics) in the envelope spectrum. This step then consists in searching for the maximum amplitude m_i in narrow bands centered on the expected theoretical orders. The search for faults harmonics in bands is intended to compensate for the effects of the sliding phenomenon; The latter is critical since it causes a random deviation from the theoretical orders given by the manufacturer and compromises the automatic tracking of fault harmonics when performed on a single specific order. Therefore, a band B_1 is first defined, centered at the theoretical fault characteristic order α_{c1}, with a deviation tolerance of 100X%, typically between 5% and 10%. The lower and the upper bounds α_1^L and α_1^U of band B_1 are thus defined as:

$$\begin{cases} \alpha_1^L = \alpha_{c1} - X_{\alpha_{c1}} \\ \alpha_1^U = \alpha_{c1} + X_{\alpha_{c1}} \end{cases} \tag{3}$$

After defining the highest peak m_1 in B_1, a measure that represents the probability of the presence of the first harmonic of the fault, PPF_1, is calculated based on the statistical threshold λ_{1-p} already obtained from the histogram of the envelope spectrum with p=0.1. PPF_1 is calculated as follows:

$$PPF_1 = \begin{cases} \dfrac{m_1 - \lambda_{1-p}}{m_1} & m_1 \geq \lambda_{1-p} \\ 0 & m_1 < \lambda_{1-p} \end{cases} \tag{4}$$

The basic idea is that, as with visual inspection, the presence of the characteristic peaks of the defects is confirmed according to the intensity of their amplitudes. If these amplitudes are greater than a statistical threshold, representing background noise, then these peaks are considered symptoms of the defect.

In the case of the non-zero value of PPF_1, the algorithm then searches for the presence of the second harmonic in a new band B_2. In order to properly define B_2, the center of the first band is first corrected to account for the possible mismatch between the actual and the theoretical fault order by defining $\alpha_{C_1}^{corr}$ such that $\alpha_{C_1}^{corr}$=α where the highest peak m_1 was found. Thus, B_2 is centered on α_{C_2}= 2 * $\alpha_{C_1}^{corr}$ and given the same bandwidth as B_1, PPF_2 is calculated. The algorithm terminates when either PPF_i is equal to zero, or a maximum number of harmonics is set. The general formula describing the theoretical fault frequency correction is:

$$\alpha_{C_i} = \frac{i}{i-1} \sum_{n=1}^{i-1} \frac{\alpha_{C_n}^{corr}}{n}, \quad i > 1 \tag{5}$$

Once the PPF_is in all concerned bands have been calculated, an overall indicator PPF is calculated as the mean value.

$$PPF = \frac{1}{n}\sum_{i=1}^{n} PPF_i \quad (6)$$

It should be noted that PPF increases with the severity of the fault. This is because the severity of the fault affects the amplitudes and number of peaks that characterize it.

In [19], defect detection was performed using a non-parametric hypothesis test using the proposed indicator. However, this article follows a different direction in which four of the proposed indicator, each on one of the four fault frequencies: Ball Pass Frequency Inner race (BPFI), Ball Pass Frequency Outer race (BPFO), Ball Spin Frequency (BSF) and Fundamental Train Frequency (FTF), are used as input parameters for the neural network. The four features, PPF_BPFI, PPF_BPFO, PPF_BSF and PPF_FTF, are used to calculate the probability of presence of different types of faults; inner race (PPF_BPFI), outer race (PPF_BPFO) and ball fault (PPF_BSF and PPF_FTF).

7 Experimental Results

7.1 Bearing Data Center

The performance of the proposed algorithm is now evaluated on the bearing signals provided by the CWRU database. The CWRU database has been used in many references and can be considered as a reference to test newly proposed [32].

The datasets are divided into four categories: 48k baseline, 12k drive end fault, 48k drive end fault and 12k fan end fault – according to the sample frequency and the fault's location. The experimental setup consists of a 1.4914 kW, reliance electric motor driving a shaft on which a torque transducer and encoder are mounted. Torque is applied to the shaft via a dynamometer and electronic control system. Four types of vibration signals are collected (normal, ball fault, inner-race fault, and outer-race fault), acquired by accelerometer sensors under different operating loads and speeds. The bearing type is a deep groove ball bearing SKF6205-2RS JEM.

BPFI	BPFO	FTF	BSF
5.415	3.585	0.3983	2.357

Tab.2 Multiplicative factors to calculate four fault frequencies

The table below presents some of the signals from this dataset with some of their corresponding features values (calculated for the four fault frequencies).

Signal	**PPF_BPFI**	**PPF_BPFO**	**PPF_BSF**	**PPF_FTF**	**Kurtosis**	**RMS**	**Peak to peak**	**Margin factor**
Inner fault	0.99	0.53	0.15	0.54	5.38	0.29	3.11	10.42
Outer fault	0.76	0.96	0.04	0.48	6.94	0.31	4.64	14.99
Ball fault	0.45	0.1	0.55	0.89	3.77	2.14	20.36	7.83
Normal	0.18	0.06	0.2	0.03	2.9	0.06	0.66	7.56

Tab. 3 Some of the signals with their features

As indicated in this table, for an inner race faulty signal, the probability of presence of a fault indicator that searches for the presence of BPFI (PPF_BPFI) has a high value=0.99, which insure the presence of the inner fault. The same is for the outer race and inner race faulty signals that have [PPF_BPFO=0.96] and

[PPF_BSF=0.55 & PPF_FTF=0.89] respectively. However, the temporal features values clearly show their ability for detection.

7.2 Second Dataset

In order to validate the proposed algorithm, the latter is evaluated also on another dataset acquired by John Stokes in the University of New South Wales (UNSW) in Australia. The test bench has a gearbox composed of two shafts, one of which is driven by a three-phase motor. The power flows through a hydraulic motor and pump. The two input and output shafts are placed parallel to each other, and are connected to the gearbox by two bearings each. An accelerometer was installed above the defective bearing, and the signals were temporally sampled with a sampling frequency of 48 KHz. The defects installed are localized defects in the form of a small superficial notch placed either on the outer ring, or on the inner ring or on one of the balls. The vibration signals were acquired under different conditions of rotational speed, 3, 6 and 10Hz, and load torque, 25, 50, 75, 100 Nm. The bearing under-test, *Koyo 1205*, has the following fault frequencies:

BPFI	BPFO	FTF	BSF
7.11	4.89	0.41	2.65

Tab.4 Multiplicative factors to calculate four fault frequencies

The table below presents some of the signals from this dataset with some of their corresponding features values (calculated for the four fault frequencies).

Signal	**PPF_BPFI**	**PPF_BPFO**	**PPF_BSF**	**PPF_FTF**	**Kurtosis**	**RMS**	**Peak to peak**	**Margin factor**
Inner fault	0.95	0.08	0.01	0.00	2.89	9.89	65.88	5.83
Outer fault	0.2	0.82	0.14	0.10	3.25	4.17	3.65	1.35
Ball fault	0.16	0.19	0.83	0.18	3.00	5.65	45.5	5.59
Normal	0.055	0.18	0.27	0.57	4.01	4.26	9.25	2.21

Tab. 5 Some of the signals with their features

7.3 Results

The same code was applied on both of the datasets; Even the GA parameters were set the same. These parameters were chosen based on GA logic and concept of starting with a large population in order to be able to obtain a large variety of solutions, and evolving the population over too many generations so we attend the desired solution. Accordingly, the total number of individuals was set to 20, and the maximum number of generations to 30. These numbers may differ from one application to another. In addition, the crossover probability was set to 0.7, which indicates that only 70% of the individuals will be recombined to form the new population. And finally, it is well-known that the mutation probability must not be high, since it will, in that case, negatively influence the main GA concept of evolving the population towards fittest solutions. Consequently, it was set to 0.03 (3%).

For the Bearing Data Center, results have shown that there may be several architectures giving the highest accuracy = 99%. This result is very logic since we are not solving an optimization problem with a well-defined equation and a unique solution. Having several topologies could be very helpful and essential in many applications especially when the data is too large and requires many hidden layers. The architectures found by the GA are formed by two or three hidden layers:

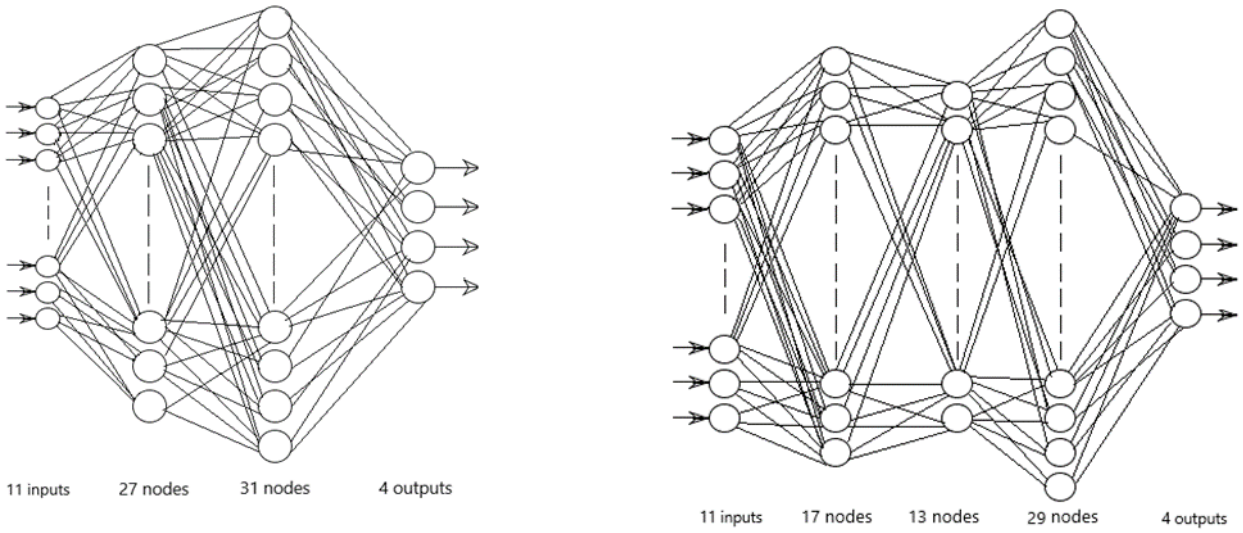


Fig.3 Two possible architectures obtained by the GA

For the features optimization, only six features out of eleven are relevant for the ANN. Those features are: PPF_BPFI, PPF_BPFO, PPF_FTF, Margin Factor, Peak to Peak, RMS. They are able to give the same classification accuracy 98% when fed into the ANN having the best architecture. The results were very convincing: this combination of both temporal features, capable of detecting the presence of a fault, and the spectral features, capable of classifying the different fault types, will surely be enough for getting good classification results. The figure below is an example of an envelope spectrum obtained for an inner race faulty signal. It explains clearly why the spectral indicator was able to classify the different faults by searching on the theoretical fault frequencies taking into consideration the slip phenomenon that may occur.

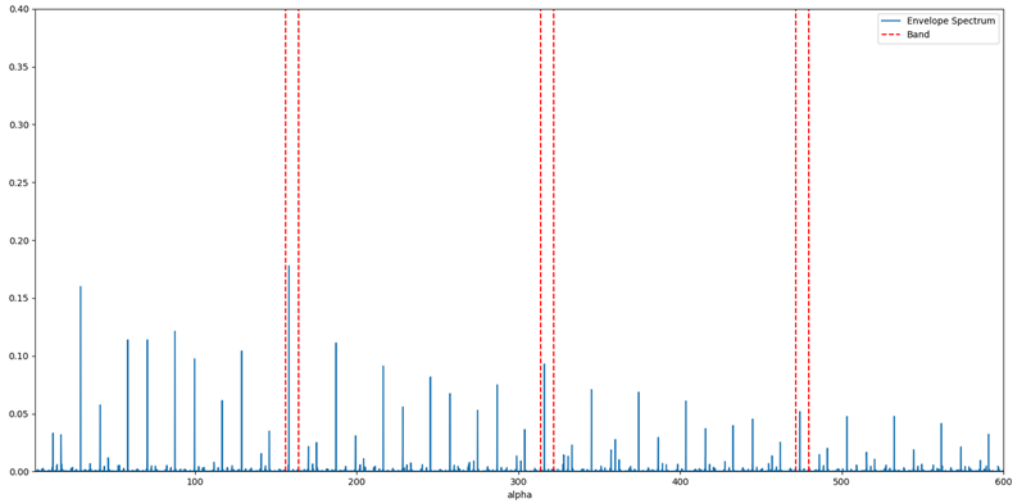


Fig.4 Envelope Spectrum for an inner race faulty signal

Below is the confusion matrix of the resulting classification model, where 0 indicates normal signals, 1 inner race fault, 2 outer race fault and 3 ball fault.

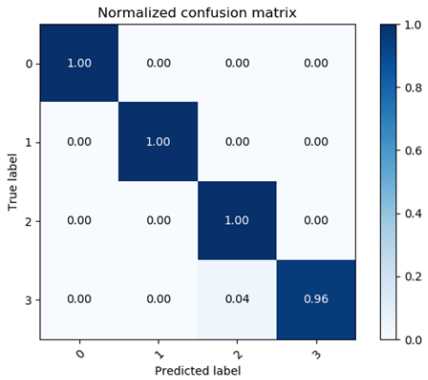


Fig.5 Normalized confusion matrix

Applying the same code on the second dataset also gave important results. The same GA was able to find ANN architectures giving a 100% classification accuracy. These architectures are formed by a single hidden layer containing 31 nodes, or by 2 hidden layers having number of nodes higher than 20. The GA convergence was much faster in this case than in the bearing data center dataset. The main reason is that the envelope spectrum for this database was more clear and without noise which directly influenced the spectral indicators values. This is also the reason of obtaining a higher classification accuracy. Below is an envelope spectrum of an inner race faulty signal from the second database:

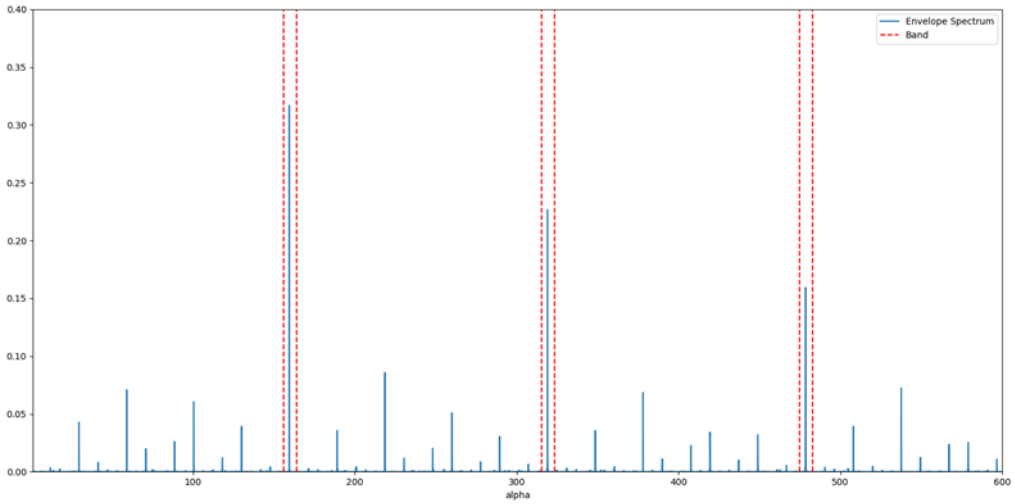


Fig.6 Envelope Spectrum for an inner race faulty signal

For the features optimization problem, here also, the number of features was reduced from 11 to 6: PPF_BPFI, PPF_BPFO, PPF_BSF, Margin Factor, Peak to Peak, RMS, with a single difference of having PPF_BSF instead of PPF_FTF. This result was a little bit confusing, for both of the datasets, since it is well-known that a signal with a ball fault is characterized by the presence of the combination of the ball spin and the fundamental train frequencies. Below is the confusion matrix of the resulting classification model:

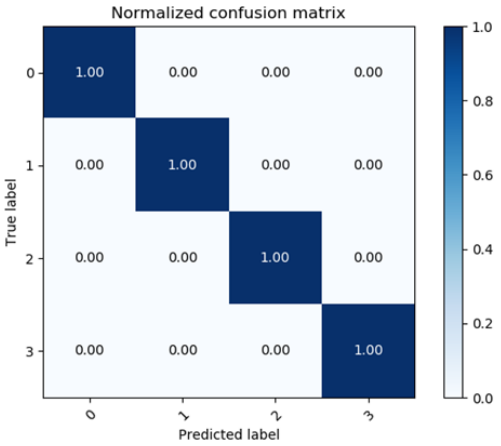


Fig.7 Normalized confusion matrix

8 Conclusion

This study presents an algorithm based on neuro-evolution for bearing fault classification problems. Although it is widely known that Neural Networks alone are used for such applications, the fact that its architecture is fixed makes it not adaptive to change. This study suggests optimizing the neural network architecture and reducing the number of features using the genetic algorithm. This technique was tested on bearing fault detection and classification with two datasets and gave promising results. This concept of applying the neuro-evolution is very powerful and effective in so many fields especially when having a huge dataset to train and a big number of features. Additional work can be done with neuro-evolution like weight optimization in order to overcome some limitations of the classical gradient descent method.

References

[1] K.A. Loparo, M.L. Adams, W. Lin, M.F. Abdel-Magied, N. Afshari, Fault detection and diagnosis of rotating machinery, IEEE Trans. Ind. Electron. 47 (5) (2000) 1005–1014

[2] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, A. Bellini, Fault detection of linear bearings in brushless AC linear motors by vibration analysis, IEEE Trans.Ind. Electron. 58 (5) (2011) 1684–1694

[3] Raad, J. Antoni, M. Sidahmed, Indicators of cyclostationarity: Theory and application to gear fault monitoring, Mech. Syst. Signal Process. 22 (3) (2008) 574–587

[4] A. Klausen, K.G. Robbersmyr, H.R. Karimi, Autonomous bearing fault diagnosis method based on envelope spectrum, IFAC-PapersOnLine 50 (1) (2017) 13378–13383

[5] Widodo, A.; Yang, B.-S. Support vector machine in machine condition monitoring and fault diagnosis. Mech. Syst. Signal Process. 2007, 21, 2560–2574

[6] Kuo, R.J. Intelligent diagnosis for turbine blade faults using artificial neural networks and fuzzy logic. Eng. Appl. Artif. Intell. 1995, 8, 25–34

[7] A. Cruz-Roa, A. Basavanhally, F. Gonzalez, H. Gilmore, M. Feldman, S. Ganesan, N. Shih, J. Tomaszewski, A. Madabhushi, Automatic detection of invasive ductal carcinoma in whole slide images with convolutional neural networks, In: Medical Imaging 2014: Digital Pathology, vol. 9041

[8] J. Huang, B. Kingsbury, Audio-visual deep learning for noise robust speech recognition, In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 7596–7599

[9] R. Sarikaya, G.E. Hinton, A. Deoras, Application of deep belief networks for natural language understanding, IEEE/ACM Trans. Audio Speech Lang. Process. 22 (4) (2014) 778–784
[10] Tao, J.; Liu, Y.; Yang, D. Bearing fault diagnosis based on deep belief network and multisensor information fusion. Shock Vib. 2016, 2016
[11] Janssens, O.; Slavkovikj, V.; Vervisch, B.; Stockman, K.; Loccufier, M.; Verstockt, S.; van de Walle, R.; van Hoecke, S. Convolutional neural network based fault detection for rotating machinery. J. Sound Vib. 2016, 377, 331–345
[12] Togelius, J., Yannakakis, G. N., Stanley, K. O. & Browne, C. Search-based procedural content generation: a taxonomy and survey. IEEE Trans. Comput. Intell. AI Games 3, 172–186 (2011)
[13] Lipson, H. & Pollack, J. B. Automatic design and manufacture of robotic lifeforms. Nature 406, 974–978 (2000)
[14] Clune J, Mouret J-B, Lipson H. 2013 The evolutionaryorigins of modularity. Proc R Soc B 280: 20122863. http://dx.doi.org/10.1098/rspb.2012.2863
[15] J. Montana, David & Davis, Lawrence. (1989). Training Feedforward Neural Networks Using Genetic Algorithms.. 762-767
[16] Fröhlich, Holger & Chapelle, O & Scholkopf, B. (2003). Feature Selection for Support Vector Machines by Means of Genetic Algorithms. Int Conf on Tools with Artificial Intelligence. 142 - 148. 10.1109/TAI.2003.1250182
[17] Trans. Ind. Electron. 2015, 62, 657–667, doi:10.1109/tie.2014.2308133. 2. Jeschke, S.B.C.; Song, H.; Rawat, D.B. Ind
[18] Malhi, A.; Gao, R.X. PCA-based feature selection scheme for machine defect classification. IEEE Trans. Instrum. Meas. 2004, 53, 1517–1525
[19] Kass, Souhayb & Raad, Amani & Antoni, Jerome. (2019). Self-Running Bearing Diagnosis Based on Scalar Indicator using Fast Order Frequency Spectral Coherence. Measurement. 138. 467-484. 10.1016/j.measurement.2019.02.046
[20] Basheer, Imad & Hajmeer, M.N.. (2001). Artificial Neural Networks: Fundamentals, Computing, Design, and Application. Journal of microbiological methods. 43. 3-31. 10.1016/S0167-7012(00)00201-3
[21] Suk, H.-I. “An Introduction to Neural Networks and Deep Learning,” in Deep Learning for Medical Image Analysis, London, Academic Press - Elsevier, 3-24 (2017)
[22] M. T. Hagan and M. B. Menhaj, "Training feedforward networks with the Marquardt algorithm," in IEEE Transactions on Neural Networks, vol. 5, no. 6, pp. 989-993, Nov. 1994
[23] L. Bottou. Stochastic gradient learning in neural networks. In Proceedings of Neuro-Nˆımes. EC2, 1991
[24] Kingma, Diederik & Ba, Jimmy. (2014). Adam: A Method for Stochastic Optimization. International Conference on Learning Representations
[25] John H. Holland. 1992. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence. MIT Press, Cambridge, MA, USA
[26] Sivanandam, S.N and S.N. Deepa. 2007. Introduction to Genetic Algorithms. Springer Berlin Heidelberg.
[27] Xu, Q., Xu, Z.Q. and Wang, T. (2015) A Data-Placement Strategy Based on Genetic Algorithm in Cloud Computing. International Journal of Intelligence Science, 5, 145-157. http://dx.doi.org/10.4236/ijis.2015.53013
[28] Tiwari, K.C.. (2001). Neural network parameters affecting image classification. Defence Science Journal. 51. 263-278. 10.14429/dsj.51.2237
[29] Gnana Sheela, K & Deepa, S N. (2013). Review on Methods to Fix Number of Hidden Neurons in Neural Networks. Mathematical Problems in Engineering. 2013. 10.1155/2013/425740
[30] Stanley, Kenneth & Clune, Jeff & Lehman, Joel & Miikkulainen, Risto. (2019). Designing neural networks through neuroevolution. Nature Machine Intelligence. 1. 10.1038/s42256-018-0006-z
[31] J. Antoni, G. Xin, N. Hamzaoui, Fast computation of the spectral correlation,Mech. Syst. Signal Process. 92 (2017) 248–277
[32] Smith, Wade & Randall, R.B.. (2015). Rolling Element Bearing Diagnostics Using the Case Western Reserve University Data: A Benchmark Study. Mechanical Systems and Signal Processing. 64-65. 10.1016/j.ymssp.2015.04.021

Electrical motors

Long-term vibration monitoring of induction motors in the food industry with low-cost MEMS accelerometers

Agusmian Partogi OMPUSUNGGU[1], Karel JANSSEN[1], Vladimir DZYUBA[1]

[1]Flanders Make vzw, Gaston Geenslaan 8 – B-3001 Leuven, Belgium
agusmian.ompusunggu@flandersmake.be

Abstract

Many production lines in the food industry, that run continuously 24 hours per day, are nowadays equipped with induction motors to drive machines to process raw materials to become final products. As the motor function in the production lines is vital, the failure of the motor can thus interrupt the production process that eventually leads to economic losses, i.e. higher production cost. The statistical analysis on the maintenance record of a specific production line conducted in this study confirms that induction motor breakdowns are the major contributors of the unplanned production downtimes. Furthermore, this case study also shows that the common failure mode of the induction motors is due to the rolling element bearing faults, which is in-line with the findings of many authors in the literature. The main interest of the production line owner is how to minimize the unplanned downtimes such that the productivity can be maintained or increased, and at the same time, the production cost is minimized. In this paper, we present a testimonial story of a setting-up a vibration monitoring system to continuously monitor the condition of motors for the first time in a real production line with low-cost MEMS accelerometers available on the market. Some technical challenges and the state-of-the-art techniques used to compute health indicators from the measured raw vibration signals are presented in this paper. The installed vibration monitoring system has successfully identified a damaged bearing in one of the monitored motors. This finding was also independently validated by a maintenance service company.

1. Introduction

Operations cost efficiency enhancement is the primary decision driver of manufacturing enterprises, in order to stay economically competitive [1]. It not only refers to reducing resource (material, energy, machine and labour) utilization per unit of manufacturing process output, and improving product quality and yield per unit of input but also includes reducing unplanned downtimes or delays of the production process due to technical issues (machine breakdown, unavailable material, blockage of a line). Conventionally, manufacturers tend to ignore a full exploration of the enormous power of data, although these data have huge potential to help increase their operations cost efficiency [2]. The production-related data generated on the shop floor is various, such as production logs, power consumption, maintenance registers, machine performance indicators, labour shifts, material availability, and storage status. A majority of manufacturing enterprises either do not gather these data, or have fragmented data, or simply store these data without any further management and analytics.

With the emerging industrial transition toward Industry 4.0, the penetration of Internet-of-Things (IoT) technologies into manufacturing industry (industrial IoT, or IIoT) enables collecting these various data in a cheaper and more flexible manner [3], hereby unlocking the enormous potential of big data technologies applied to production lines. In this paper, we discuss our experience in assisting a food manufacturing company that runs its production lines 24 hours per day in the transition toward Industry 4.0. This company has a number of production lines to manufacture different types of food. The key performance indicator (KPI) set by the company that need to be achieved within this project is an improved operation cost efficiency and productivity.

The production lines availability is one of the other important aspects that influence the operation cost efficiency and productivity. Currently, the *preventive maintenance* is applied to the production lines to maintain the availability, where the production lines have to be stopped in every pre-determined maintenance window (e.g. every few weeks). To reach the KPI, the maintenance strategy, therefore, needs to be changed from the *preventive maintenance* to the *condition-based maintenance (CBM) /Predictive Maintenance (PdM)*

strategy. The **first critical step** to successfully implement the CBM/PdM strategy is to identify critical assets that have a significant impact on the business. Here, the asset criticality is determined by the duration of production downtimes caused by the failure of a certain asset. The longer the production downtime is, the more critical the asset will be.

Since the beginning of the project, it was not entirely clear yet what production lines/sub-production lines or machines that can be categorised as critical assets. To identify critical assets in an objective manner, the maintenance record and production data of a pilot production plant for more than 6 years (from April 2011 to November 2017) were analysed. The analysis revealed that most of the downtimes in the production plant are caused by the failures of AC electric motors located at different production lines. Further analysis also showed that the majority of the AC electric motor failures is caused by the mechanical damages on the rolling element bearings.

Once the critical assets have been identified, the **second step** is to determine which technologies necessary to monitor the health condition of the critical assets. Vibration based condition monitoring is a well-established approach that has been employed by industries for many years in their maintenance program of rolling element bearings [4]. The common practice of this approach is that, vibration measurements are periodically recorded using portable vibration sensors (i.e. accelerometers) and measurement signals are analysed by an expert to interpret the bearing condition. However, this common practice can lead to serious misinterpretation, where rapidly growing faults, that might occur in rolling element bearings, could be missed. In contrast, a continuous condition monitoring approach offers a more optimal solution in which the bearing condition is continuously tracked. This way total failures can be anticipated in advance thus allowing optimal maintenance action. Despite its advantages, the continuous monitoring program is however not well adopted by industry because of high investment cost, where sensor cost is a major factor. To remedy this gap, cost-effective accelerometers are therefore needed.

The remainder of this paper is organised as follows. Section 2 discusses the market survey for low-cost accelerometers suitable for bearing condition monitoring purposes. Section 3 describes the architecture of the continuous monitoring system installed in a production line. Section 4 discusses the analysis results of the data acquired by the installed monitoring system. Section 5 presents the conclusions drawn from the analysis and proposes a future work.

2. Sensor Selection and Deployment

The high investment cost is one of the bottlenecks for adopting continuous condition-based maintenance strategies in the industry. A major part of these costs is introduced by the sensors. Advancements in the field of MEMS accelerometers have enabled opportunities for low-cost alternatives while maintaining basic-performance requirements for vibration-based condition monitoring purposes.

MEMS accelerometers offer many attractive attributes. They combine the economic benefit with, for example, a compact, a high sensitivity, a good resistance to shocks and acceptable stability over a wide range of temperatures. In the previous study [5], a market survey was carried out and it was concluded that the only MEMS accelerometers available on the market suitable for vibration-based condition monitoring (in particular for bearing faults monitoring) are the ones from Analog Devices, ADXL001-70/ ADXL001-250. The main criteria for selecting such sensor models are because of *i*) the high dynamic range and *ii*) the wide frequency range properties. However, the noise performance over higher frequency ranges of the selected sensor models is relatively low, *i.e.* higher noise density level, which is about **4000 µg/√Hz**.

Recently, the market study has been updated as summarised in Figure 1. It turns out that Analog Devices has released the new generations of MEMS accelerometers for more than one year, namely ADXL1001/ADXL1002 having ultra-low noise density level, which is about **25 µg/√Hz**. These ultra-low noise sensor models are the successor of ADXL001-70/ADXL001-250. Despite the fact that the noise density level of the successors is much lower than that of the predecessors, other potential limitations of the low-cost ADXL1001/ADXL1002 MEMS accelerometers such as the long-term signal drift, bias offset and overall robustness of the sensor to industrial environments, are not clear yet.

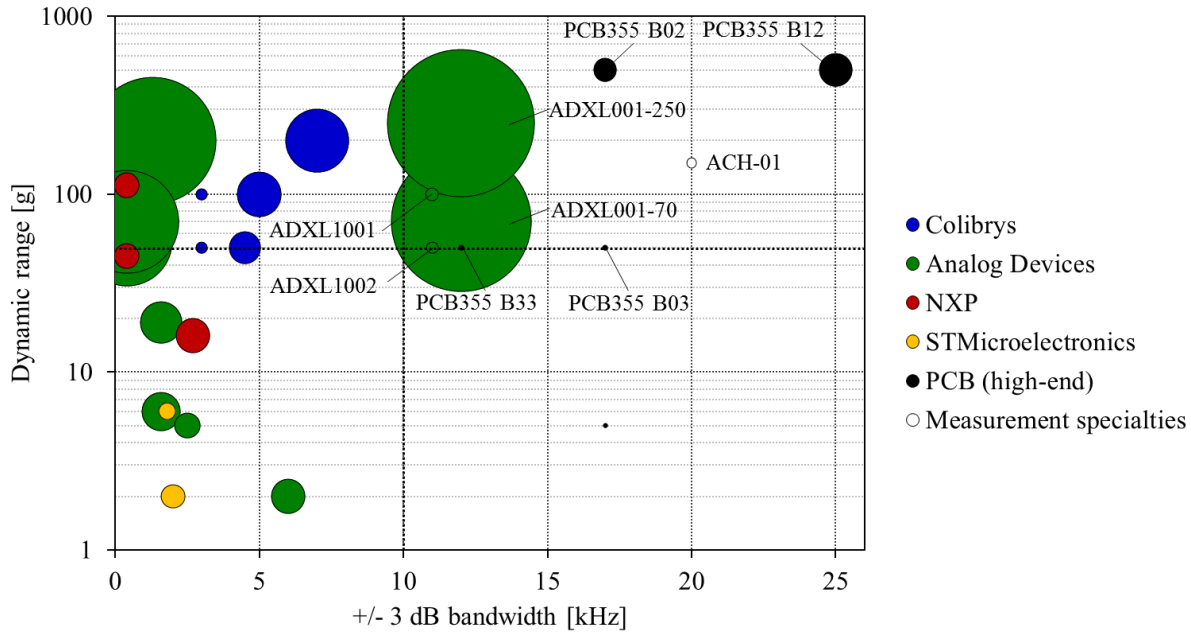


Figure 1: The market overview of analogue MEMS accelerometers updated version of [5]. The dashed lines indicate the minimum requirements set for condition monitoring applications. The diameter of the circles indicates the noise density as specified in the datasheets.

Proper hardware solutions were exploited to cope with the inherent limitations of the low-cost MEMS accelerometer that can affect the monitoring performance. To this end, a printed circuit board (PCB) and a tailored-packaging have been designed and produced. Figure 2 schematically illustrates the sensor deployment process to protect the MEMS sensor and enhance its overall robustness. To preserve the frequency range of the MEMS accelerometer, the packaging should be designed with care. For this purpose, a design criterion for the packaging was imposed, namely, the first packaging resonance frequency should be higher than the maximum frequency range of interest, i.e. 10 kHz. The flowchart of the packaging design is shown in Figure 3.

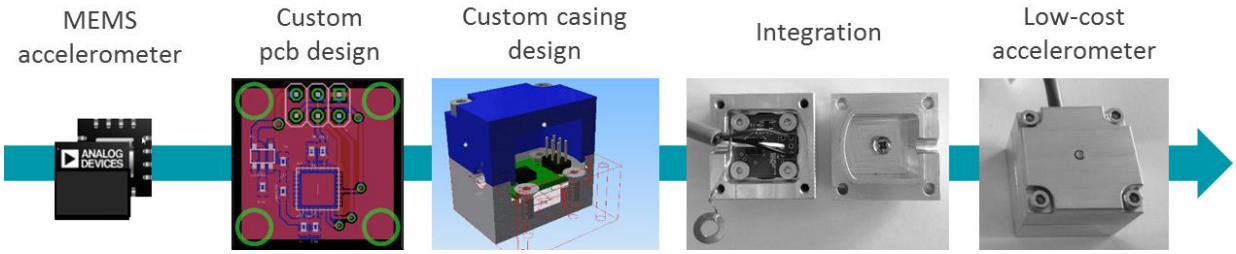


Figure 2: Sensor integration and packaging process.

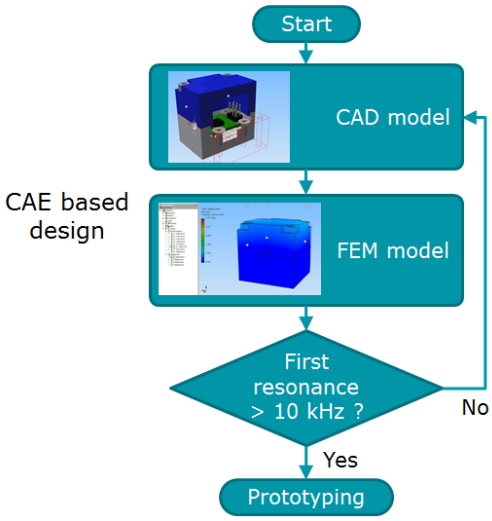


Figure 3: Packaging design flowchart.

3. The Continuous Monitoring System Architecture and The Installation in A Production Line

The asset criticality analysis, as discussed in Section 1, has guided us to consider 4 electric motors which are located in different locations. Hence, four of the selected ultra-low noise sensor model (ADXL1002) have been packaged according to the sensor deployment flow described in Section 2.

Figure 4 illustrates the architecture of the monitoring system, in which each vibration sensor is installed on an individual motor. Each sensor is oriented such that the measuring axis in parallel with the horizontal axis and radial axis of each motor. Note that each sensor is powered by 5 VDC power supply. Sensor #1 and #2 are both installed on an extrusion press in a regular industrial environment, while sensor #3 and #4 are installed on a ventilator and belt motor subject to harsh environmental conditions with the temperature variations between 20 and 120°C and the humidity up to 90%. The latter stressed conditions have a major impact on the lifetime of the motor bearings and are an ideal industrial use case for this study.

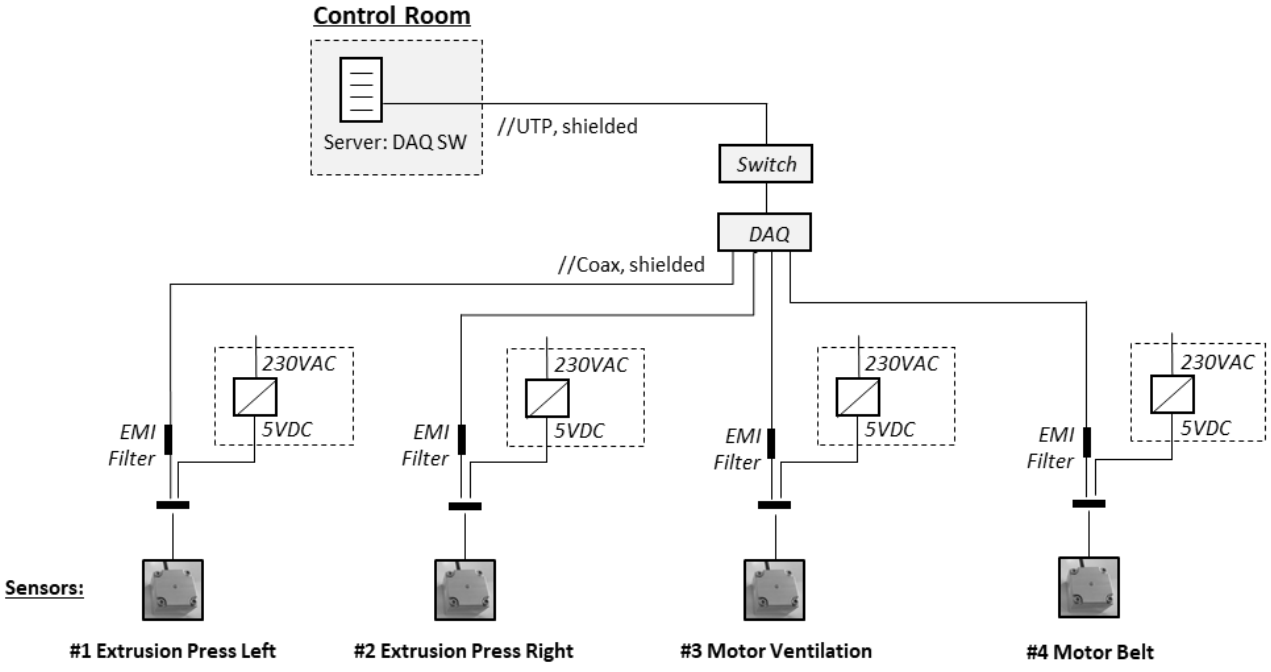


Figure 4: Architecture of the monitoring system installed in a production line.

Each sensor is individually connected to a four-channel data acquisition device (NI CDAQ9191) sampled at a high frequency of 50 kHz. The digital data from the data acquisition device is transferred to a server via an ethernet cable. On the server, a custom data recording program shown in Figure 6, which can be scheduled along the working hours, is run, which stores every half hour a few seconds of data. This monitoring system has been successfully running and generating a dataset of almost one year of production data.

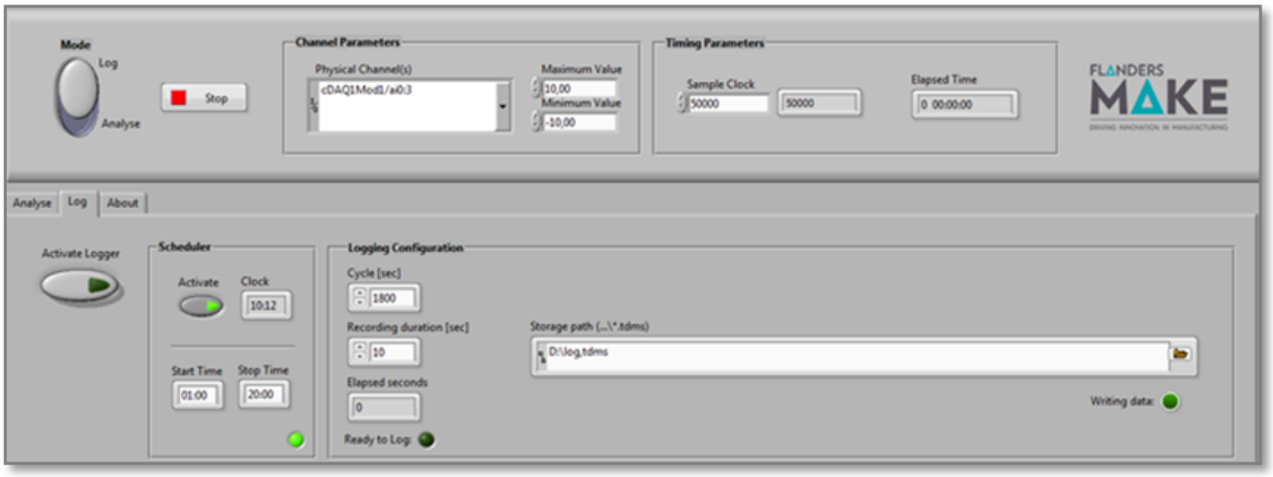


Figure 5: Custom data recording program.

The two sensors installed in the regular industrial environment are still in operation until now. In contrast, after more than one month since the installation, the two sensors installed in the harsh environmental conditions showed an anomaly behaviour, where the DC values of the raw signals have dropped to values around zero. The comparison of the raw signals of a functional and failing sensor is shown in Figure 6. It is not clear yet the reason for the sensor failure. But it seems that one of the electronic components used in the sensor packaging is vulnerable to a long time high temperature.

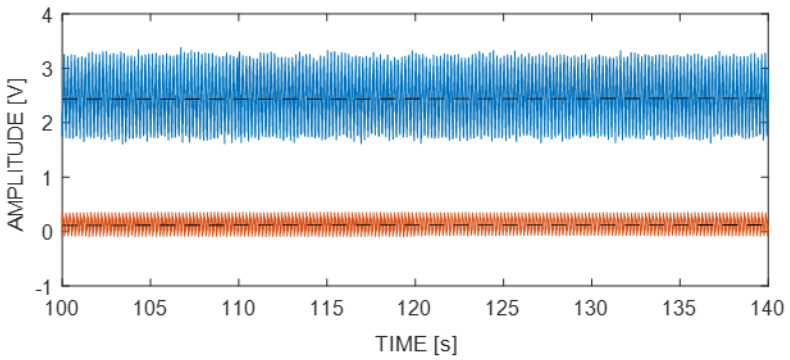


Figure 6: Raw sensor data of a functional (top) and failing (bottom) sensor, sampling at 50 kHz.

4. Data Analysis

Batch data acquisition Sensor data are acquired in batches: in a standard operating mode, a batch is recorded every 30 minutes. If necessary (e.g., if data analysis indicates imminent failure), the acquisition period can be gradually reduced to ensure up-to-date information for decision-making. The length of a batch depends on the rotational speed of the motor: for reliable analysis, a batch should contain a sufficient number of full revolutions. The rotational speed of the monitored motors while executing a typical production order ranges from 1400 to 1500 RPM (approximately 25 Hz), therefore we set the batch length to three seconds. Thus, one sensor generates at least 57 megabytes of uncompressed raw vibration data per day.

Computing health indicators For each monitored motor and each batch, we compute several bearing health indicators, referred to as *features*. Figure 7 illustrates the feature computation algorithm. It requires

three inputs: bearing fault frequencies, the rotational speed of the motor shaft, and the raw vibration signal. The required bearing fault frequencies include the bearing defect frequency (BDF), and the inner & outer ball pass frequencies (BPFI & BPFO), and the ball spin frequency (BSF) for the drive and non-drive ends; their values typically can be found in the manufacturer's catalogue. The shaft speed can either be directly read from the motor controller interface (PLC) or estimated from the vibration data; in this work, we use the latter method. To filter out non-production situations (e.g., maintenance, cleaning, holidays, etc.), we skip the batches where the rotational speed is much lower than the typical values of 1400-1500 RPM; we set the filter threshold to 600 RPM.

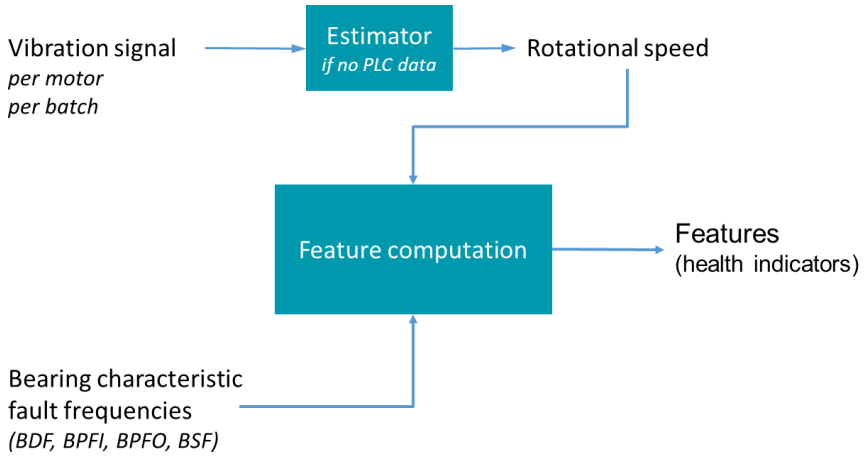


Figure 7: Inputs and steps of the feature computation algorithm.

From a high-level perspective, the main steps of the algorithm are (1) centering the vibration signal, (2) estimating its envelope spectrum, and (3) informally, matching the spectral peaks with the expected fault frequencies: the closer the match, the higher the feature value and hence, the likelihood of a bearing failure. Figure 8 illustrates the third step: the bearing monitored by *Sensor#2* is substantially more likely to fail than the one monitored by *Sensor#1*. For the detailed description of the algorithm and its options, we refer the reader to Ompusunggu *et al.* [5].

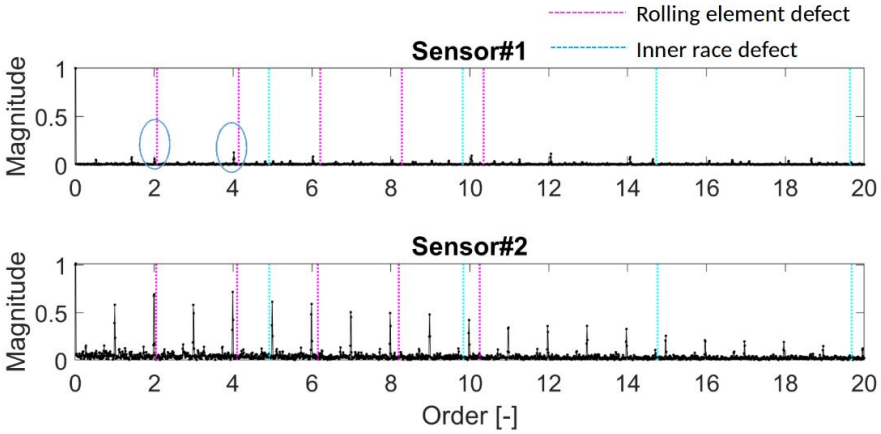
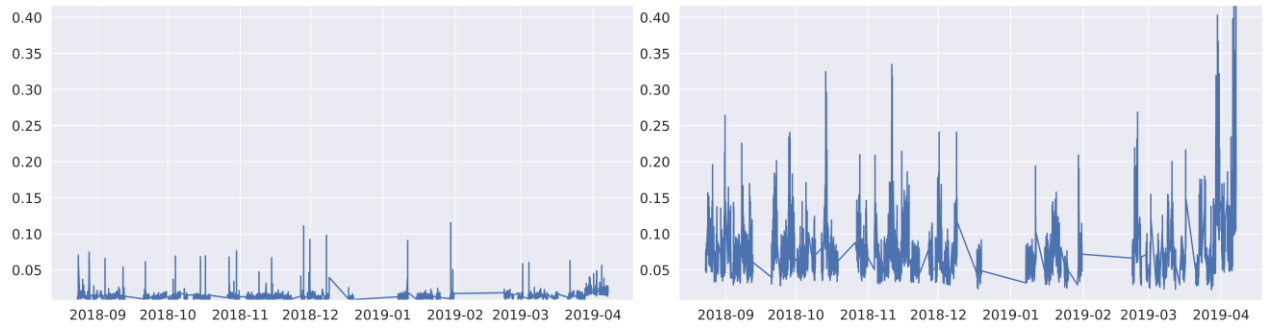


Figure 8: Final step of feature computation: matching *envelope spectral* peaks with bearing fault frequencies. Close matches for *Sensor#2* likely indicate bearing fault.

The algorithm computes 10 features in total: 8 features for each fault frequency listed above, one feature for the shaft, and an aggregate *global* feature. This enables a coarse-grained overview as well as fine-grained analyses. Figure 9 shows the evolution of the global feature over time for the two motors shown in Figure 8, indicating the stable condition of the first motor and the gradual increase of the failure likelihood of the second motor. These findings have been confirmed by an external audit.



(a) Sensor#1 (b) Sensor#2

Figure 9: Evolution of the global bearing health indicator for two motors over eight months (missing values correspond to non-production periods, e.g., maintenance or holidays).

Remaining data analysis challenges include *accounting for operational context* and *calibration*. Relevant contextual data (such as the shaft speed, load, ambient temperature, raw material, etc.) can be read from PLCs and various manufacturing systems, e.g., SCADA or MES. Accounting for these data will reduce the influence of external factors and mitigate spurious feature value fluctuations seen in Figure 9. Calibration requires data from the complete lifecycle of a bearing, from installation to failure. Once such data become available, dimensionless feature values can be used to calculate interpretable health indicators, e.g., fault size or time to failure.

Implementation details The software is implemented in MATLAB with a thin Python wrapper for scripting. The implementation reads the TDMS files output by the NI acquisition device and generates CSV files that can be used directly or uploaded to a database, an IoT platform, or another data processing system.

5. Conclusions & Outlook

In this paper, we have shared our experience in assisting a food manufacturing company in setting up a continuous monitoring system in transition from the current practice *preventive maintenance* toward the condition-based maintenance/predictive maintenance. The maintenance record and production data recorded for more than 6 years have been used and analysed to objectively determine critical assets in a pilot production plant. It turns out from the analysis that AC electric motors located in different production lines are the most critical assets, where rolling element bearings are the main root causes of the motor failures.

The architecture of the monitoring system with a low-cost vibration sensor solution has been proposed and realised in the pilot production line to monitor the health condition of four selected electric motors. The monitoring system has been running successfully and acquiring vibration data, which are stored in a server. The off-line analysis has shown that one of the selected motors is already faulty. The findings have been verified independently by a third party. As a result, a maintenance action for the faulty motor has been planned and will be executed in the coming weeks.

The future work will be to further extend the framework that allows for on-line analysis and decision making.

Acknowledgements

This work was executed within the imec.icon project ELITE, a research project bringing together academic researchers and industry partners. The ELITE project was co-financed by IMEC, Flanders Make and received project support from Flanders Innovation & Entrepreneurship (VLAIO).

The work of sensor selection and deployment was executed within the VIBMON_ICON project, co-financed by Flanders Make and VLAIO.

References

[1] F. Jaehn, "*Sustainable Operations*" European Journal of Operational Research, 2016; 253, 243-264.

[2] E. Auschitzky, M. Hammer and A. Rajagopaul. "How big data can improve manufacturing," McKinsey&Company article, July 2014, http://www.mckinsey.com/business-functions/operations/our-insights/how-big-data-can-improve-manufacturing

[3] J. Wan, S. Tang, Z. Shu, et al. "Software-Defined Industrial Internet of Things in the Context of Industry 4.0," IEEE Sensors Journal, 2016; 16(20): 7373-7380.

[4] R. B. Randall & J. Antoni. Rolling element bearing diagnostics-A tutorial. Mechanical Systems and Signal Processing, Vol. 25, p. 485–520., 2011.

[5] A. P. Ompusunggu, T. Ooijevaar, B. Kilundu, S. Devos, *Automated Bearing Fault Diagnostics with Cost-Effective Vibration Sensor*: Proceedings of the 12th World Congress on Engineering Asset Management and the 13th International Conference on Vibration Engineering and Technology of Machinery, 2019.

Structural Dynamics of Electric Machines Subjected to PWM Excitations

Margaux TOPENOT[1,2], Morvan OUISSE[1], Gael CHEVALLIER[1], Damien VAILLANT[2]

[1]Institut FEMTO-ST, Université de Bourgogne Franche-Comté,
24 Rue de l'Épitaphe, 25000 Besancon, France

[2]Alstom Transport,
7 Avenue du Maréchal de Lattre de Tassigny, 25290 Ornans, France

Abstract

Pulse Width Modulation (PWM) is commonly used for driving asynchronous machines. The mechanical torque is the result of the combination of several multiphysics conversions. The PWM is generally defined on the basis of the electromagnetic performances determined with simple mechanical behavior: typically, rigid bodies are considered to be representative of the mechanical behavior. However, since PWM generates numerous harmonics of current and voltage, it may have an impact on the structural dynamics of the system. In particular, the rotor is subjected to torque oscillations over a wide frequency range that may result in large vibration amplitudes when coincidences occur. These excitations can induce severe damages like fans ruptures. In this work, a finite element model and an reduced model are compared in terms of ability to describe the structural dynamics of the system when the electric machine is driven with PWM. The results are compared and discussed in terms of applicability for the design of electric machines.

1 Introduction

Trains are set into motion through electrical energy. Electrical signals are synthesized by the Pulse Width Modulation (PWM). From a theoretical point of view, train motors can be driven by sine wave signals but practically, due to signal synthesis issues and power limitations, motors are driven thanks to voltage impulsions. This allows to manage easily the variation of the rotating speed. The PWM definition makes the fundamental frequency equal to the sine signal frequency. Unfortunately, PWM generates small-amplitude oscillations which frequencies correspond to the harmonics. They lead to vibration motions that engender spurious mechanical stress. Figure 1 illustrates the difference between PWM and sinus waveform signal. This results in torque oscillations over a wide frequency range and with large amplitudes, sometimes leading to the rupture of several rotating parts due to resonance [1]. The aim of this study is to suggest and compare rotors dynamic modeling able to capture the phenomenon.

In this paper, two models are submitted : the first one consists by coupling magnetic and mechanical finite elements models whereas the second one uses reduced order modelling to get rotor oscillations faster. The workflow of the methodologies can be found in the figure 2 .This study will allow to determine which scale of refinement is necessary to describe the problem with a good accuracy.

2 Design methodologies

2.1 Weakly coupled magnetic and mechanic Finite Elements Model

Firstly, we propose a methodology that couples magnetic and mechanic Finite Elements Models (FEM). The aim of this study is to evaluate the relevance of Maxwell mechanical stress projection on mechanical degrees of freedom (dof), see figure 3. The magnetic model is built with Flux 2D (Altair) whereas the mechanical one is set with Optistruct (Altair).

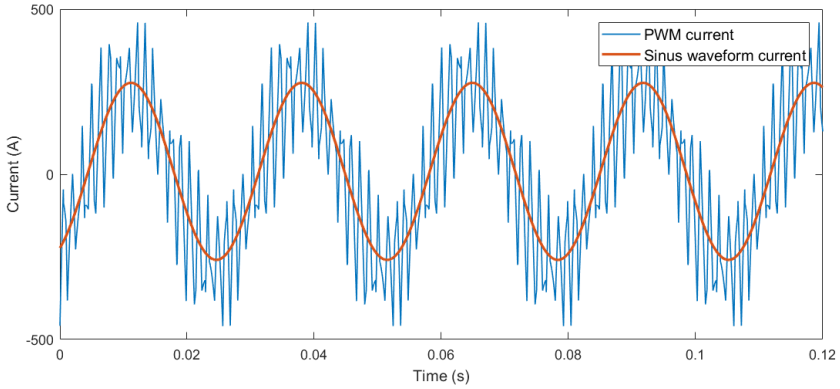


Figure 1: Current waveform with PWM or sinus

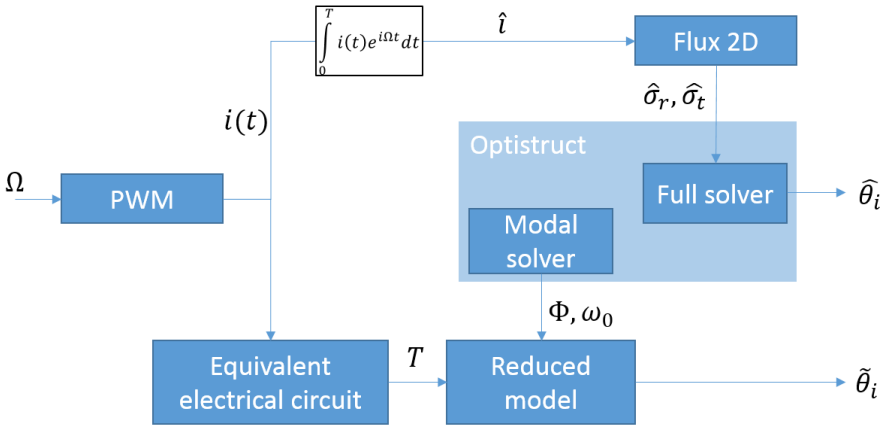


Figure 2: Workflow of methodologies for weakly coupled model and reduced model

Flux2D/Optistruct coupling is introduced in the thesis of Jaafar Hallal [2]. The goal is to develop a multi-physic model between 2D magnetic model and 3D mechanical model in order to evaluate the dynamic behavior coming from magnetic forces to study acoustic noise. Here, the results of this work are used to determine the vibratory fatigue failure that may occur.

The magnetic calculation takes place in two steps : firstly a magnetic harmonic response allows to initialize transient magnetic computation. It is mandatory to define geometry, materials, mesh, and then to create the electrical circuit that corresponds to the motor. Each component of the mesh is assigned to electrical component. Finally, we define the solving scenario: voltage, current, frequency, number of rotation, discretization angular steps. In our case, we simulate a working point that corresponds to a fixed rotating speed. The motor is powered by sinus waveform signal. The magnetic field is governed by the following set of equations.

$$\begin{aligned} \overrightarrow{rot}(\vec{E}) &= -\frac{\partial \vec{B}}{\partial t} \\ div(\vec{B}) &= 0 \\ \overrightarrow{rot}(\vec{H}) &= \vec{J} \\ \vec{J} &= \sigma \vec{E} \\ \vec{B} &= \mu \vec{H} \end{aligned} \tag{1}$$

where $\vec{E}$ is the electrical field (V/m), $\vec{B}$ is the magnetic induction (T), $\vec{H}$ is the magnetic field (A/m), $\vec{J}$ is the current density (A/m2), σ is the conductivity (S) and μ is the permeability (H/m). Flux 2D solves the following equation including magnetic potential vector $\vec{A}$ and electric potential V, see eqn. 2.

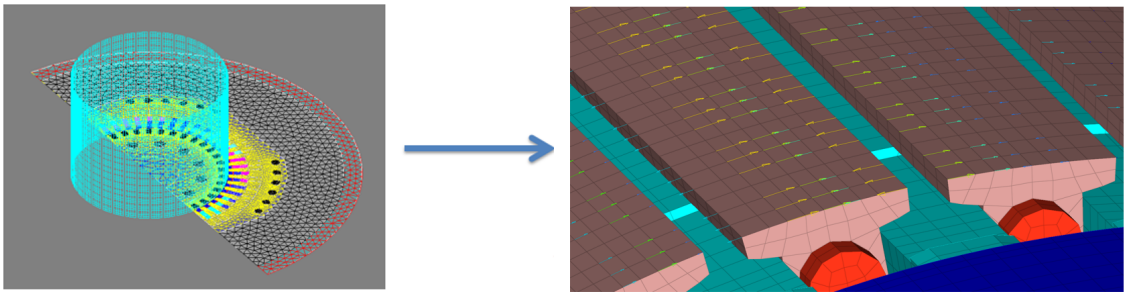


Figure 3: Maxwell pressure projection

$$\overrightarrow{rot}\left(\nu_0[\nu_r]\overrightarrow{rot}(\vec{A})-\overrightarrow{H_c}\right)+[\sigma]\left(\frac{\partial\vec{A}}{\partial t}+\overrightarrow{grad}(V)\right)=0 \tag{2}$$

where $[\nu_r]$ is the magnetic reluctivity, ν_0 is the vacum reluctivity ($\nu_0 = 1/\mu_0 = 1/(4pi10-7)$ m/H), $\vec{A}$ is the magnetic potential vector (Wb/m), $\overrightarrow{H_c}$ is the coercitif field (A/m), $[\sigma]$ is the conductivity tensor (S) and V is the electric potential (V). The link between eqn. 2 and eqn. 1 is given by $\vec{B} = \overrightarrow{rot}(\vec{A})$. It allows to determine the magnetic induction $\vec{B}$ and then the Maxwell pressure (3).

$$\begin{aligned}\sigma_r &= \tfrac{1}{\mu_0}(B_r^2 - B_t^2)\\ \sigma_t &= \tfrac{1}{\mu_0}B_rB_t\end{aligned} \tag{3}$$

where $[B_r]$ is the radial magnetic induction (T), $[B_t]$ is the tangential magnetic induction (T), $[\sigma_r]$ is the radial pressure (N/m2) and $[\sigma_t]$ is the tangential pressure (N/m2)). The mechanical frequency response is driven by the following set of equations.

$$\begin{aligned}-\omega^2\rho\hat{u} &= div(\hat{\sigma})\\ \hat{\sigma} &= H\hat{\varepsilon}\\ \hat{\varepsilon} &= \tfrac{1}{2}\left(grad(\hat{u})+grad^T(\hat{u})\right)\end{aligned} \tag{4}$$

Flux 2D generates loadings for the mechanical model:

$$\hat{\sigma}.n = \begin{vmatrix}\sigma_r\\ \sigma_{t_1}\\ \sigma_{t_2}\end{vmatrix} \text{ on } \Gamma_l \tag{5}$$

where Γ_l is the set of rotor lamination surfaces. Both equations allow to know the displacement u and stress σ in the rotor as a function of the frequency. Rotations can be determined by using a transformation matrix C defined in the following relation :

$$\theta = Cu \tag{6}$$

with θ the rotation of sections, C the transformation matrix built with speed composition relation and u the nodal displacement of the finite element model.

2.2 Mechanical reduced order model

An electromechanical lumped model has already been published by Bruzzese in 2016 [3]. He managed to define electrotechnic model by solving equivalent electrical circuit and apply magnetic torque on a simplified mechanical model composed by masses, springs and dampers. Inertia and stiffness matrices were calculated by a lumped representation.

We offer to use a similar model to represent the electrotechnical part (by solving equivalent equation of the motor), using d-q frame instead of α-β frame, but taking into account all parameters of the inverter and the train (inertial effect). Upstream, algorithms are used to generate PWM signal. Thus, this model allows us to simulate torque oscillations. Besides, we propose to compute mechanics from modal basis and nodal

coordinates of sections coming from the 3D FEM in order to have a more accurate mechanical representation of the real rotor.

The key of such a model is the transformation matrix that allows the reduction. The interest of model reduction is based on the dimension change of the problem. We start with a complete 3D model composed of thousands of translation degrees of freedom and we reduce it in order to obtain a model with only nine rotational degrees of freedom.

Then, we choose to divide the rotor in nine sections as described in figure (4).

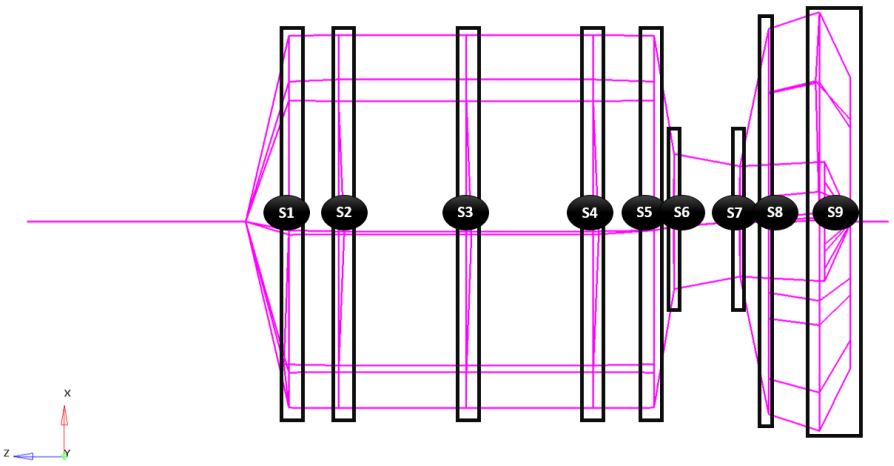


Figure 4: Rotor divided in 9 sections

Next, several matrices are built in order to create a transformation matrix. We define generalised coordinates q as the projection on the modal basis Φ of the displacement u. Let's consider a transformation matrix A between the displacement u and the rotation of sections θ, see eqn. 7.

$$\begin{aligned} u &= \Phi q \\ u &= A\theta \\ \Phi_\theta &= \begin{vmatrix} A^{-1}\Phi \\ C\Phi \end{vmatrix} \end{aligned} \tag{7}$$

where A is defined thanks to the speed composition relation and contains coordinates of nodes in sections (A is the inverse matrix of C defined in the previous paragraph) and with Φ_θ the modal basis expressed in rotation.

The set of equations (7) gives:

$$q = (A\Phi_\theta)^{-1} A\theta = P\theta \tag{8}$$

The classic dynamic equation in translation is given by the following equation.

$$[M]\ddot{u} + [D]\dot{u} + [K]u = F \tag{9}$$

Writted on the modal basis, the dynamic equation in generalized coordinates is :

$$\ddot{q} + [\Delta]\dot{q} + [\Lambda]q = f \tag{10}$$

with $[\Delta] = 2\xi\omega_0[I]$.

By using the transformation (equation 8), the equation in the rotation coordinate frame becomes :

$$[J]\ddot{\theta} + [D_\theta]\dot{\theta} + [K_\theta]\theta = T, \tag{11}$$

with

$$\begin{cases} [J] = P^T P, \\ [K_\theta] = P^T [\Lambda] P, \\ [D_\theta] = P^T (2\xi\omega_0[I]) P. \end{cases} \tag{12}$$

Rotations of rotor sections are obtained by solving the equation (11).

3 Comparison of results

Rotations of rotor sections are obtained with both methods presented above. This paragraph shows the results on the torsion modes, see figure 5. The first table illustrates the validation of the model reduction with PWM excitation, whereas the second table compares the coupled model with sine wave excitation and the reduced model with PWM excitation.

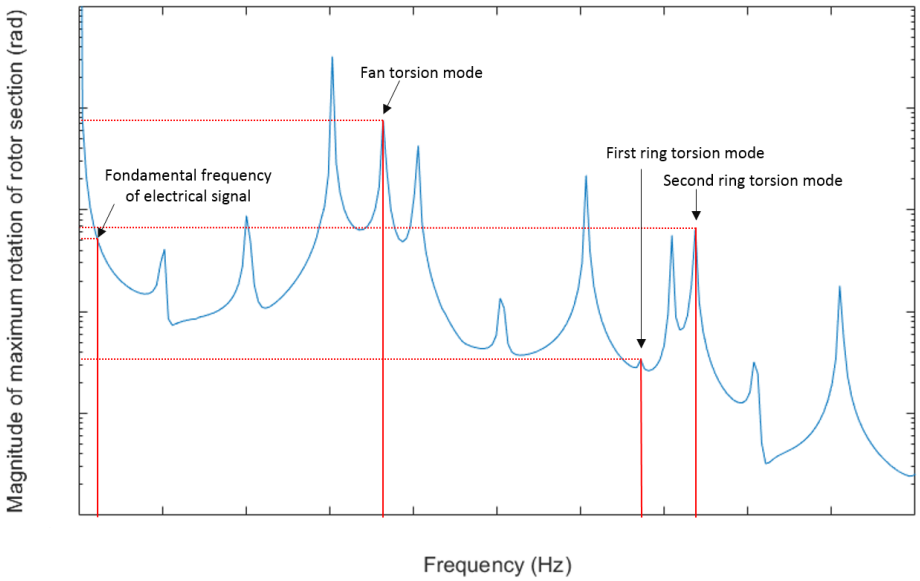


Figure 5: Frequency Response Function

	PWM + FEM	PWM + Reduced model	Relative error (%)
Fondamental frequency	6.648E-7	6.645E-7	0.05
Fan torsion mode	7.03E-6	7.591E-6	7.98
First ring torsion mode	3.363E-8	3.374E-8	0.33
Second ring torsion mode	7.269E-7	6.44E-7	11.4

Table 1: Comparison of results obtained with both models and PWM input - Maximal rotation (rad)

Values of the table 1 are close and permit to justify the validation of the model reduction.

	Sine wave + FEM	PWM + Reduced model	Relative error (%)
Fondamental frequency	1.134E-6	6.645E-7	41.4
Fan torsion mode	1.504E-7	7.591E-6	4947.2
First ring torsion mode	7.779E-9	3.374E-8	333.7
Second ring torsion mode	7.084E-8	6.44E-7	809

Table 2: Comparison of results obtained with both models with sine wave or PWM input - Maximal rotation (rad)

Table 2 shows the difference between the weakly coupled model with sine wave input and reduced model with PWM input. Values are not comparable except for the fondamental frequency of the PWM that corresponds to the sine wave frequency. This illustrates the necessity to use the coupled model to simulate the working point with PWM excitation to validate the approach.

The solving of the equivalent circuit is immediate, the computation of modal basis takes 8 minutes on a calculator with 16 CPU, 3.2GHz and 225692 MB RAM and then the calculation of rotation thanks to reduced model takes 5 seconds. In comparison, the magnetic computation with Flux 2D takes 9 hours in a laptop with 8 CPU, 2.7GHz and 16 Go RAM and then the mechanical computation with Optistruct takes 2 hours on the

calculator. The model reduction permits to pass from 11 hours to 8 minutes of computation with in addition PWM excitation.

4 Conclusion

This paper propose two methods to model the problem of vibratory fatigue that happened on train rotor subjected to PWM excitation. The first one combines two Finite Elements Models in order to solve the multiphysic problem. The projection of magnetic excitation on mechanical mesh allows to determine rotation of rotor sections. Secondly, we propose a model reduction to improve the computation time during design phase. Transformation matrix is proposed to reduce model dimension and also to change the observation in rotation.

As mentioned, the Flux 2D simulation were performed with a sinus waveform instead of PWM, whereas the reduce model included PWM excitations. Values are not comparable. The next step could be to simulate the problem with PWM. Other points to improve are to simulate variable speed and to get stress with both models instead of rotations to be able to design rotor in vibratory fatigue.

References

[1] Bruzzese, Tessarolo, and Santini, "Failure Root-Cause Analysis of End-Ring Torsional Resonances and Bar Breakages in Fabricated-Ca," 2016.

[2] J. Hallal, *Etudes des vibrations d'origine electromagnetique d'une machine electrique : conception optimisee et variabilite du comportement vibratoire*. PhD thesis, 2014.

[3] Bruzzese, "Electromechanical modeling of a railway induction drive prone to cage vibration failures," in *IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society*, IEEE, 2016.

Experimental evidence of MCSA for the diagnosis of ball-bearings

Fabio IMMOVILLI[1], Marco LIPPI[2], Marco COCCONCELLI[3]

DISMI, University of Modena and Reggio Emilia, 42123 Reggio Emilia, Italy
{[1]fabio.immovilli,[2]marco.lippi,[3]marco.cocconcelli}@unimore.it

Abstract

Electrical and mechanical fault diagnosis in induction machines is an extensively investigated field for cost and maintenance savings, as induction motors operating at mains frequency are still the most widespread rotating electric machines employed in industry. Many papers can be found in the literature concerning the general condition monitoring of induction machines. Bearing faults are one of the most common failures in electrical machines: bearing faults that are not detected in time cause malfunction, reduced efficiency, and may even lead to failure of the driven machinery. Online fault detection can be obtained by vibration analysis, but the diagnosis equipment is costly and invasive, requiring dedicated equipment and specific sensors to be installed. Motor current signature analysis (MCSA) is an alternative method that relies on the monitoring of electrical quantities, that are already acquired in the final application, e.g., to implement the control of an electric drive, thus do not require the installation of dedicated transducers. Many research activities were focused on the diagnosis of bearing faults by MCSA. The use of suitable signal processing techniques is required to efficiently extract the fault signatures from raw signals. The use of current and/or voltage signal constitutes a noninvasive method to bring information necessary to diagnose a fault in the system via online monitoring of the electric machine. This paper details the results of a laboratory trial comprising different test sets on the condition monitoring and fault diagnostic of a six-poles induction motor, using a design of experiment (DOE) approach. The manuscript summarizes the results of research that the authors did in the last years, and the aim of the paper is giving a unified point of view on the work done. After the selection of a proper fault-related scalar value, the DOE proves its robustness against setup and working conditions of the motor. Finally, the same scalar value will be proved to be effective as input to long short-memory networks, for automatic condition monitoring of faulted bearings.

1 Introduction

Electrical and mechanical fault diagnosis in induction machines is an extensively investigated field for cost and maintenance savings, as induction motors operating at mains frequency are still the most widespread rotating electric machines in industry, mainly because of their low price, ruggedness and reliability.

Many papers can be found in the literature concerning the general condition monitoring of induction machines [1], [2]. The distribution of failures within the machine sub-assemblies is reported in many reliability survey papers [3]. A rough classification identifies four classes: bearings faults, stator related faults, rotor related faults, other faults (lack of cooling, loose terminal box connection). Bearing faults are one of the most common failures in electrical machines especially in the small-medium power sizes [4]. Bearing faults that are not detected in time cause malfunction, loss of performance, reduced efficiency and may even lead to failure of the driven machinery, [5].

In many situations diagnostics methods based on the analysis of the vibration signals have proved their effectiveness [6], [7]. Among the mechanical problems detected by vibration spectra there are: imbalance, misalignment, loose fitting, bent shafts, and bearing localized faults. On-line fault detection can be obtained by vibration analysis, but the diagnosis equipment is costly and invasive, requiring dedicated equipment and specific sensors to be installed.

Motor current signature analysis (MCSA) is an alternative method that relies on the monitoring of electrical quantities, that are already acquired in the final application, e.g. to implement the control of an electric

drive, thus do not require the installation of dedicated transducers. Many research activities were focused on the diagnosis of bearing faults by MCSA [8]. In many cases mechanical signals cannot be directly acquired, e.g. in harsh environments, remote locations, or because the application is difficult to access. Under such conditions, electric signal measurements would be preferable as they are more immune to external disturbances, [9]. Non-invasive fault diagnosis should ideally detects faults at the early stage, to allow for scheduled maintenance, minimizing system downtime. Under this circumstances, fault signature components feature a very small amplitude that is usually buried in noise and can lead to false positive detection [10]. The use of suitable signal processing techniques is required to efficiently extract the fault signatures from raw signals. The use of current and/or voltage signal constitutes a non-invasive method to bring information necessary to diagnose a fault in the system via on-line monitoring of the electric machine.

Thanks to a dedicated test bench, a laboratory trial comprising different test sets on the condition monitoring and fault diagnostic of a 6-poles induction motor using a design of experiment (DOE) approach was performed. Four different data were acquired (namely vibration data, currents, motor torque and radial force on the motor shaft) and post-processed. This paper focuses on the results of the MCSA, while the analysis of the vibration data is given only as comparison if a reference is needed. The diagnostic techniques used are three: a spectral analysis with an ad-hoc pre-processing, the DOE analysis to qualify the effects of the setup parameters and a machine learning approach based on Long Short-Term Memory (LSTM), particularly suited for the analysis and classification of time-series.

The paper is organized as follows: section 2 reviews the relationship between vibration and current components presented in literature. Section 3 presents the experimental setup with a description on the test bench. Section 4 outlines the two methodologies adopted, specifically the factorial design of experiment and the machine learning technique. The results for test runs under different working conditions and bearing damage are reported in section 5, followed by Conclusion.

2 Currents Based Condition Monitoring

Numerous papers in literature deal with the detection and diagnosis of electro-mechanical faults based on MCSA in induction motors. The link between vibrations and motor current spectral components is still under investigation in the scientific community and is treated in literature according to different approaches.

In the first one, the vibration component causes a rotor eccentricity [14].

The second one links the vibration component to a torque ripple that produces a speed ripple on the electric machine [13]. A unifying approach is presented in [15]. Mechanical vibrations results in a torque ripple that generates in the current a chain of components at frequencies F_{be}:

$$F_{be} = | \, f \pm k \, f_{car} \, | \tag{1}$$

where k is an integer.

Radial bearings consist of two concentric rings containing inner and outer races, separated by rolling elements, Fig. 1. Rolling elements are separated by a cage: a component that maintains a constant angular pitch between adjacent rolling elements, preventing contacts.

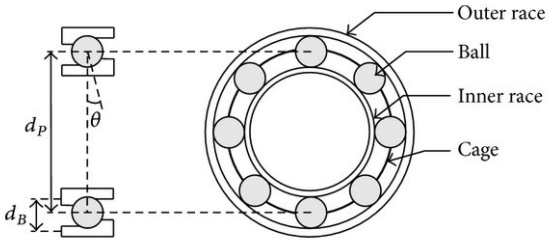


Figure 1: Bearing structure and characteristic dimensions.

Localized faults will produce characteristic vibration frequency components. These bearing fault frequencies are a function of the bearing geometry and the relative speed of the outer and the inner ring. Characteristic

vibration frequencies can be calculated from the bearing's physical dimensions, Fig. 1. In particular, considering the outer ring fixed to the frame:

$$F_{cage} = \frac{1}{2} F_r \left(1 - \frac{D_b \cos\theta}{D_p}\right) \tag{2}$$

$$F_{outer} = \frac{Z}{2} F_r \left(1 - \frac{D_b \cos\theta}{D_p}\right) \tag{3}$$

$$F_{inner} = \frac{Z}{2} F_r \left(1 + \frac{D_b \cos\theta}{D_p}\right) \tag{4}$$

$$F_{ball} = \frac{D_p}{D_b} F_r \left[1 - \left(\frac{D_b \cos\theta}{D_p}\right)^2\right] \tag{5}$$

where D_b stands for the ball diameter, D_p for the pitch diameter, Z for the number of rolling elements, θ for the ball contact angle, Fig.1. Table 1 summarizes the corresponding vibration related components on the machine current for the torque fluctuation model.

	Model based on torque fluctuations
Outer raceway defect	$f \pm k F_{outer}$
Inner raceway defect	$f \pm k F_{inner}$
Ball defect	$f \pm k F_{ball}$
Cage defect	$f \pm k F_{cage}$

Table 1: Vibration related components in the stator current spectrum.

Considering the torque ripple model, vibration effects on machine currents are caused by small speed fluctuations of the rotor. Because of electromechanical filtering effects (due to the rotor inertia and winding inductance) MCSA is in general more sensitive to low frequency phenomena.

To sum it up, it is usually very difficult to retrieve bearing fault signature components by MCSA. Especially because when dealing with realistic incipient faults, the fault signature is buried in noise or it is only a small fraction of the fundamental supply current component, especially when operating at rated load. In [16] the torque ripple associated with a realistic (not drilled) localized fault was experimentally measured: the peak value of the torque ripple was 3-4 orders of magnitude smaller than the nominal torque of the machine.

3 Test setup

The experimental setup, Fig. 2, comprises the electrical motor under test (MUT) that is installed on a test bench in order to vary both the radial and the torque load conditions. The chosen MUT is a three phase induction machine operated directly connected to the 50 Hz mains grid. The test bench also houses a brake/dynamometer consisting of a vector controlled induction machine in order to vary the load torque on the MUT. Table 2 summarizes the nameplate data of the MUT.

Nominal Power	1100	$[W]$
Number of poles	6	
Nominal current	2.8	$[A]$
Power Factor	0.76	
Nominal Torque	11.5	$[Nm]$
Stator Resistance	5.65	$[\Omega]$

Table 2: Nameplate data of the Motor Under Test

Radial load on the MUT shaft is provided by a specially designed test fixture, comprising a pneumatic cylinder coupled to a manifold with pressure regulator and transducer to modulate the radial load. The cylinder is connected to a crosshead carrying an extension shaft that allows to apply a variable radial force at the motor

Table 3: Specifications of the ball bearing and expected fault frequencies

Supply frequency (f)	50 [Hz]	F_r	15.9 [Hz]
Inner diameter	25 [mm]	F_{inner}/F_r	5.41
Outer diameter	52 [mm]	F_{outer}/F_r	3.585
Number of spheres	9	F_{cage}/F_r	0.398
Basic static load rating	7800 [N]	F_{ball}/F_r	4.715

shaft's end. The cylinder is supplied with compressed air at a pressure up to 6 bar, corresponding to a radial force of up to 1180 N exerted on the front bearing.

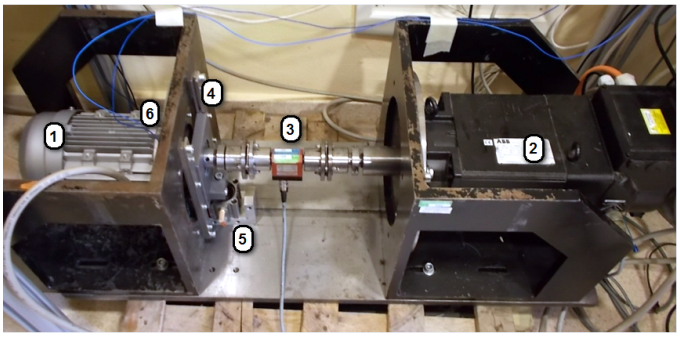


Figure 2: Test setup overview: (1) MUT; (2) brake/dynamometer; (3) torque meter; (4) crosshead; (5) pneumatic cylinder; (6) accelerometer position. The compressed air hose connecting the cylinder to the manifold with pressure regulator and transducer is not shown.

The test bearing is a SKF 6205 deep grove ball bearing. Table 3 summarizes the characteristic fault frequencies with the MUT operated at nominal frequency.

Two different damages were artificially made on the bearings, in order to apply DOE to different data test set: A single defect on the outer raceway, created by chemical etching of the bearing outer race, Fig. 3-left. A simulated brinelling defect, generated applying a mechanical load of 4 tons (40 kN) to the bearing, Fig. 3-right.

Figure 3: Micrography of the chemically etched outer race defect (left). Photo of the hydraulic press employed to impart bearing brinneling damage (right).

The physical quantities monitored are: the radial vibration of the motor, the stator currents fed to the machine, the radial force exerted by the pneumatic cylinder and the load torque at the motor shaft. The vibration of the motor is measured by means of a mono-axial accelerometer placed on the frame of the test rig (sensitivity: 10.28 mV/g). The currents are measured by means of LEM LTSR 6-NP closed loop Hall current transducers (nominal current: 6 Arms; output voltage: 104,16 mV/A, accuracy: $\pm$0,2%) and the pressure by mean of a pressure sensor (output voltage: 0–10 V; measurement range: 0–10 bar, accuracy: $\pm$0,5%). The torque is measured by a torque meter mounted between the motor and the brake shafts (maximum torque: 20 Nm, linearity: $\pm$0,2% of full scale).

4 Methods

4.1 Design of experiment

The Design of Experiment (DOE) procedure, a powerful statistical technique based on the analysis of variance (ANOVA), can be conveniently applied to multivariable problems. The DOE is the branch of science that deals with designing the correct sequence of experiments to minimize measurement errors and maximize the evidence of dependencies between causes and results [12].

The inputs to the system are called “factors” and each of them could have more than one value (usually called “levels”). The number of factors and the number of levels determine the complexity of the experimental plan and the total number of tests to be done. The DOE approach fixes some procedures in order to minimize the influence of parameters other than the selected inputs, and proposes statistical tools to determine the significance of dependencies between the input and output of the system [17].

In this paper, the design of experiment for each test set consists of three independent factors with two levels each: the value of radial load, the load torque and the type of damage on the front bearing. Table 4 summarizes the factors and levels used in the first and second test run.

First test set			
Factors:	Radial load	Load torque	Bearing status
Level 1	3 bar / 590 N	50%	healthy
Level 2	6 bar / 1180 N	100%	Outer race
Second test set			
Factors:	Radial load	Load torque	Bearing status
Level 1	3 bar / 590 N	50%	healthy
Level 2	6 bar / 1180 N	100%	Brinneling

Table 4: Factors and levels of the independent variables used in the Design Of Experiment (DOE).

Two test set were performed in the laboratory trial, one employing a bearing with a localized fault on the outer race, the other employing a bearing with artificial brinneling. The resulting full factorial experimental plans consist of 8 randomized tests each, which are not replicated. The statistical analysis software Minitab was used to lay out the randomized test plan and to perform the analysis of variance (ANOVA) on the results. The levels of the independent variables were normalized: torque was normalized to the rated torque of the machine, radial load was normalized to the maximum value obtainable, while the fault is modeled as a binary variable (healthy = 0, faulty = 1).

The current signal has been filtered by a series of notch filters in order to the remove the 50 Hz fundamental mains supply frequency and its higher harmonics over all the frequency range. The Root Mean Square (RMS) value of the filtered current is taken as response factor for the ANOVA anlysis. The RMS is used to take into account the energy of the residual signal, considering that any damage to the motor requires additional energy, appearing in the spectrum as sideband modulations according to Table 1. Since the torque ripple due to the bearing fault impacts is independent of the torque load on the machine, the residual signal was chosen as a good candidate for a robust scalar fault index.

Figure 4 shows the post-processing flowchart of the current signal, where X is the input current signal, Y is the residual signal after the filtering and Z is the RMS value.

Figure 4: Flowchart of the current signal processing.

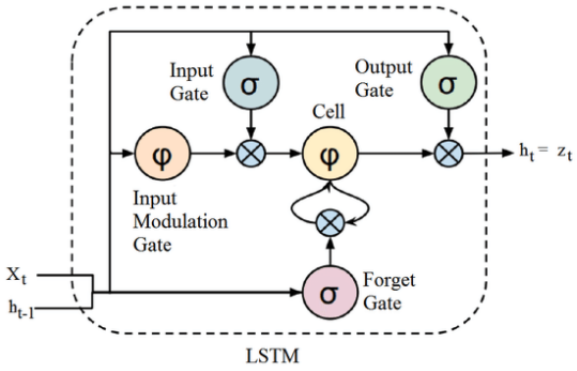


Figure 5: Depiction of LSTM cell. σ (input, output, forget gates) are typically the sigmoid function. Recurrent connections h_t and X_t propagate information through time.

4.2 Long Short-Term Memory Networks

Long Short-Term Memory Networks (LSTMs) are a class of recurrent neural networks. They are designed specifically to process periodic input sequences, such as time-series. They were first introduced in the late nineties [11], and demonstrated high performance in different tasks pertaining pattern recogniti and sequence classification. Only during the last decade and thanks to the largely increased interest in neural networks, LSTMs became the state-of-the-art in different application domains pertaining both to the industry and the research [20]). Recurrent networks are capable of processing time sequences of arbitrary length. The output of a RNN is not simply a function of the actual input, but it is also function of the internal *state* of the network. The output of a generic cell h_t^ℓ in layer ℓ at time t is a function of both the input of that layer (x_t) and of the previous output of the same cell at the time step (h_{t-1}^ℓ). During training, classic RNNs suffer from a problem named "vanishing gradient" [21]. That limitation greatly hinders their applicability in many practical classification tasks. LSTMs were introduced to the particular aim of overcoming this limitation, thanks to the implementation of a *memory cell* with a more complex structure. In simple terms, LSTM cells can maintain their state over time, or forget what they have learned and also to allow new information in. In order to overcome overfitting issues, LSTMs make use of a form of regularization called dropout [22]: during the training phase the inter-layer connections between cells are randomly dropped with probability $(1-p)$ and ,at prediction time, all the weights are multiplied by p.

Figure 5 illustrates the schematic structure of such a cell. The cell's status X_t depends on three non-linear gates that control the information flow: the input gate, which allows new information to enter the state; the forget gate, that controls how much of the value contained in the previous state is kept or forgotten at a given time and the output gate that transfers the information to the upper layers.

The network input data, in the present work, is the time-series of the three phase current signals, pre-processed to obtain the residual signal (Y signal in Fig.4). More specifically, each example fed to the network is a $3 \times W$ matrix, where W is a preset window length of samples of each of the current signals.

5 Results

5.1 DOE results

The test runs were performed according to the DOE and the physical quantities defined in Section 4.1 were acquired using 24 Bit, 51.2 kS/s data acquisition modules: for each test run, a 10 s length file was recorded for post processing and analysis.

Figure 6 shows - on the same picture - a comparison of the spectrum of the current signal, in case of healthy and faulty front bearing (localized defect on the outer race). For uniformity of presentation, both the faulty and the healthy case shown are for a machine operated at rated output torque and 6 bar pressure on the radial load cylinder. The sidebands already present are due to the intrinsic unbalance of the electric machine caused by manufacturing tolerances.

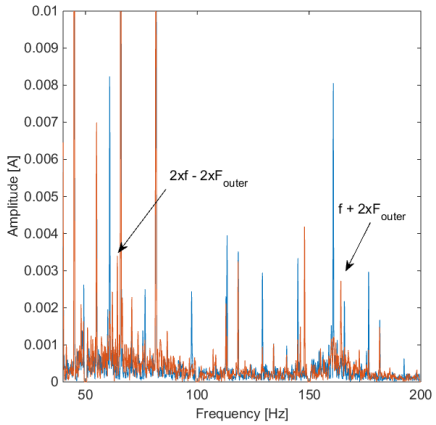


Figure 6: Comparison of FFT spectra of the notch-filtered current signal of the MUT operated at rated output torque and 6 bar pressure on the radial load cylinder: healthy bearing (blue), faulty bearing with localized outer race defect (red).

In a deep groove ball bearing the impact forces due to the bearing fault act along a radial direction mainly. It is reasonable to expect the torque ripple related to the bearing fault to be dependent upon radial load, but to be independent of the torque transmitted by the shaft. From preliminary observations on the radial vibration signals, no significant difference was observed between the test run with the MUT operating at no torsional load and at rated-torque: if the fault is present, fault signature is evident regardless of output torque.

There are a number of approaches to represent the results graphically to demonstrate the effects of the variables on the system outputs. One of the most popular is the normal plot, used to estimate whether a certain set of data follows a Gaussian distribution or not. If the data approximates a straight line the phenomenon is statistically "normal" i.e. follows a stochastic law. The variables affecting the system response will then fall outside the normal distribution line, thus their effect cannot be ascribed to a stochastic process. The greater the deviation of the point from the normal line the larger the confidence interval (i.e. the probability that the variables are significant is higher). The half normal plot, used in this paper is interpreted in the same way as the normal plot but allows absolute values of the effects to be considered. Figure 7 shows the half-normal probability plots from an analysis of the variance (ANOVA) test of current signals fault indicator in case of single point defect (on the left) and brinneling defect (on the right).

The ANOVA highlights the effects that significantly influence a physical phenomenon by comparing these effects of the output variable with a stochastic effect. This is obtained by comparing the results coming from the several levels of the selected experimental variable, driven by the stochastic experimental error. These values represent the probability that the effect of the variable is significant.

Half normal plots show the magnitude of the experiment's effects ordered in increasing magnitude along the x-axis. The effect for a factor is the difference of the average response variable over "high" factor levels minus the average response over the "low" factor levels. As said before, half normal plot show the distribution of the abs(X) with X having a normal distribution with mean zero. The points comprising factors with small and/or insignificant effects on the response will describe (roughly) a straight line on the plot. The points for factors with a 'large' and thus significant effects will visually fall off of the straight line described by the insignificant factors. A line through the insignificant factors helps to graphically delineate the difference between significant and insignificant factors. To visually interpret half normal plots: selecting the factor points which lie reasonably off of the line describing insignificant factors is an easy graphical way to identify important factors and start the process of optimizing the model. Further details and additional statistical information on the half normal plot construction can be found in [18] and [19].

Concerning the first test set on a single point defect on the outer race, the left part of Fig. 7 shows the half normal plot of the current fault indicator: as it can be seen, only the effect of the fault presence is significant, while load torque and radial load and their interactions have negligible effect.

Concerning the second test set on a simulated brinneling fault the right part of Fig. 7 shows the half normal plot of the current fault indicator: as it can be seen, only the effect of the fault presence is significant, while load torque and radial load and their interactions have negligible effect.

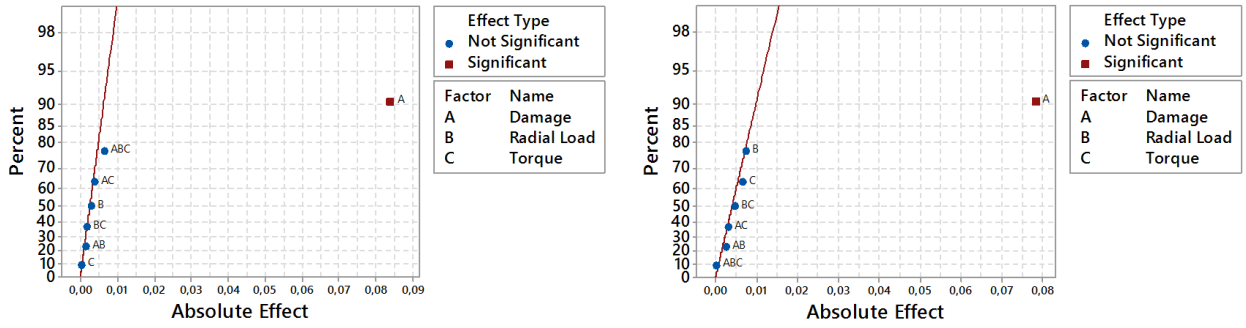


Figure 7: Test set 1 (single point defect on outer race): Half normal plot of the scalar current fault indicator (left). Test set 2 (brinneling defect): Half normal plot of the scalar current fault indicator (right).

It is worth pointing out that with the present test setup, the maximum pressure in the pneumatic cylinder (6 bar) results in a radial loading force that is $1/6$ of the maximum permissible static force provided by the bearing's manufacturer. This represents a normal working condition: much higher values of load could cause an increase of the importance of radial load factor, but would result in reduced lifetime of the bearing and not typical of practical applications in everyday use.

5.2 LSTM results

The deep network used in the fault detection experiments is composed of two LSTM layers, each of them containing 100 cells, followed by other two dense layers containing 100 neurons each (this is a classical feed-forward neural network), followed by one output neuron. A dropout value $p = 0.5$ and a batch size equal to 64 were used for training. The network was trained with the RMSprop algorithm, using an initial learning rate $\eta = 5 \cdot 10^{-5}$. Concerning the data input, a window $W = 300$ was used and, as customary with neural networks, all the features were normalized in the $[0,1]$ interval using the statistics computed on the training set.

The data analysis task began with the simple binary classification of distinguishing faulty cases (class F) from normal healty operation (class H). To obtain the training set, data coming from the experiment with 50% radial load and 50% output torque was used, the data set related to experiments with a single defect on the outer raceway represented the faulty class. To construct the validation set, the same cases were employed, but pertaining to experiments with the radial load level equal to 100%. A relevant point is that no datasets realted to the brinnelling fault experiments were used for training nor for validation sets.

In order to evaluate the proposed approach, we employ standard metrics in machine learning, such as accuracy, precision, recall, and F_1. When dealing with a binary classification task True Positives (TP) and True Negatives (TN) are respectively the number of examples of the positive/negative class that are correctly classified. False Positives (FP) are the negative (healthy) examples that are mistakenly classified as positives, the opposite cases are False Negatives (FN). Accuracy A is defined as the total percentage of correct predictions:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

Precision P and recall R are related to the ratio of false positives and false negatives:

$$P = \frac{TP}{TP + FP} \tag{7}$$

and

$$R = \frac{TP}{TP + FN} \tag{8}$$

Finally F_1 is the harmonic mean between P and R,

$$F_1 = \frac{2PR}{P + R} \tag{9}$$

In our first experiment, the test set is the setting with half radial load, and maximum output torque: in this case we achieve $A = 0.759$, $P = 0.924$, $R = 0.564$, $F_1 = 0.700$, so that the network detects over than half of the faulty examples, with less than 8% of false positives (see confusion matrix in Table 5, left). If we apply a post-processing filter, where we detect a sample as faulty if at least m of the past n samples where predicted as faulty, then performance rapidly improves. The parameters m and n can be chosen so as to balance the number of FP and FN. For example, keeping $n = 1,000$ we achieve $P = 0.840$ and $R = 0.990$ with $m = 100$, whereas we obtain $P = 0.940$ and $R = 0.960$ with $m = 200$. In the former setting, almost all the faults are detected, but the number of FP raises to 16% while in the latter the two types of errors are more balanced.

In our second set of experiments, we considered the network trained in the first setting, and we evaluate its performance with several different test sets. In absence of radial load, we achieve $P = 0.849$ and $R = 0.818$ with full output torque, and $P = 0.931$ and $R = 0.594$ with half output torque (see confusion matrices in Table 5, center and right, respectively). The brinnelling fault is almost perfectly detected as well, with an accuracy equal to 0.984 and 0.998 with half and full output torque, respectively (radial load is kept fixed at 50%).

This result confirms that the datasets obtained with experiments at different radial loads can be used both for training and for validation, to avoid overfitting of the LSTM.

%	H	F
H	47.68	2.33
F	21.79	28.20

%	H	F
H	47.80	2.21
F	20.31	29.68

%	H	F
H	42.74	7.26
F	9.09	40.91

Table 5: Confusion matrices for three different test sets. Left: 50% radial load, 100% output torque; center: no radial load, 50% output torque; right: no radial load, 100% output torque. Rows: true labels, columns: predicted values.

6 Conclusions

This paper details the results of a laboratory trial comprising different test sets on the condition monitoring and fault diagnostic of a six-poles induction motor, using a design of experiment (DOE) approach. The DOE allows some considerations about the influence of external radial loads applied to the output shaft of the motor. A scalar fault indicator based on current signal has been proposed as fault detector. This indicator proved to positively identify the faulted case versus the healthy one. The main results of the DOE are that output torque, radial load and their interactions have negligible effect on scalar fault indicators. Only the effect of the fault presence is significant. This indicates a fair robustness of the chosen scalar fault indicators under different operating conditions and in case of different faults. Finally,the radial load has no relevant effect on the current signals in case of healthy bearing. The occurrence of false positive fault detection due to radial load is avoided. The same scalar fault indicator based on current signal has been taken as input to a Long Short-Term Memory network to test a machine learning approach for the fault detection of the bearings. The main results are that the proposed LSTM fault detection positively identify the fault case with respect to the healthy one. It can be efficiently trained on an easily reproducible single defect case, and then be employed to identify a more complex fault. Finally, the proposed LSTM fault detection showed a fair robustness of detection under different operating conditions and in case of different faults.

Acknowledgements

The authors are grateful for the National University Research Fund (FAR 2018) of the University of Modena and Reggio Emilia - Departmental and Interdisciplinary Projects - DEep LEarning Condition moniToring (DEfLECT).

References

[1] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors-a review," *IEEE Transactions on Energy Conversion*, vol. 20, no. 4, pp. 719–729, Dec 2005.

[2] A. Garcia-Perez, R. d. J. Romero-Troncoso, E. Cabal-Yepez, and R. A. Osornio-Rios, "The application of high-resolution spectral analysis for identifying multiple combined faults in induction motors," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 2002–2010, May 2011.

[3] A. H. Bonnett, "Root cause ac motor failure analysis with a focus on shaft failures," *IEEE Transactions on Industry Applications*, vol. 36, no. 5, pp. 1435–1448, Sep 2000.

[4] C. Bianchini, F. Immovilli, M. Cocconcelli, R. Rubini, and A. Bellini, "Fault detection of linear bearings in brushless ac linear motors by vibration analysis," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 5, pp. 1684–1694, May 2011.

[5] F. Immovilli. G. Curcurú, M. Cocconcelli and R. Rubini, "On the detection of distributed roughness on ball bearings via stator current energy: Experimental results," *Diagnostyka*, vol. 51, no. 3, pp. 17–21, 2009.

[6] R. Rubini and U. Meneghetti, "Application of the envelope and wavelet transform analysis for the diagnosis of incipient faults in ball bearings," *Mechanical Systems and Signal Processing*, vol. 15, no. 2, pp. 287 – 302, 2001.

[7] J. R. Stack, T. G. Habetler, and R. G. Harley, "Fault-signature modeling and detection of inner-race bearing faults," *IEEE Transactions on Industry Applications*, vol. 42, no. 1, pp. 61–68, Jan 2006.

[8] M. Blodt, M. Chabert, J. Regnier, and J. Faucher, "Mechanical load fault detection in induction motors by stator current time-frequency analysis," *IEEE Transactions on Industry Applications*, vol. 42, no. 6, pp. 1454–1463, Nov 2006.

[9] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini, and R. Rubini, "Bearing fault model for induction motor with externally induced vibration," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3408–3418, Aug 2013.

[10] S. Choi, B. Akin, M. M. Rahimian, and H. A. Toliyat, "Performance-oriented electric motors diagnostics in modern energy conversion systems," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 2, pp. 1266–1277, Feb 2012.

[11] S. Hochreiter, and J. Schmidhuber, "Deep learning," *Neural computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[12] R. A. Fisher, *The design of experiments*. Hafner Publishing Company, 1971.

[13] M. Blodt, D. Bonacci, J. Regnier, M. Chabert, and J. Faucher, "On-line monitoring of mechanical faults in variable-speed induction motor drives using the wigner distribution," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 2, pp. 522–533, Feb 2008.

[14] C. M. Riley, B. K. Lin, T. G. Habetler, and R. R. Schoen, "A method for sensorless on-line vibration monitoring of induction machines," *IEEE Transactions on Industry Applications*, vol. 34, no. 6, pp. 1240–1245, Nov 1998.

[15] M. Blodt, P. Granjon, B. Raison, and G. Rostaing, "Models for bearing damage detection in induction motors using stator current monitoring," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 4, pp. 1813–1822, April 2008.

[16] F. Immovilli, A. Bellini, R. Rubini, and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Transactions on Industry Applications*, vol. 46, no. 4, pp. 1350–1359, July 2010.

[17] D. C. Montgomery, *Design and Analysis of Experiments*. Wiley and sons, 2012.

[18] D. Cuthbert, "Use of half-normal plots in interpreting factorial two-level experiments", *Technometrics*, vol. 1, no. 4, pp. 311 – 341, 1959.

[19] M. J. Anderson and P. J. Whitcomb, *DOE Simplified: Practical Tools for Effective Experimentation, Second Edition*. Productivity Press, 2007.

[20] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey", *IEEE Transactions on Neural Networks and Learning Systems*, vol. 99, pp. 1 – 11, 2017.

[21] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult", *Neural Networks, IEEE Transactions on*, vol. 5, no. 2, pp. 157 – 166, 1994.

[22] N.. Srivastava, G.. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting", *Journ. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929 – 1958, 2014.

Uncertainties, stochastic, robustness

A model reduction method to analyze the dynamic behavior of vibrating structures with uncertain parameters

Duc Thinh Kieu, Marie-Laure Gobert, Sébastien Berger, Jean-Mathieu Mencik

INSA CVL, Univ. Orléans, Univ. Tours, LaMé EA 7494, F-41034, 3 Rue de la Chocolaterie, CS 23410, 41034 Blois Cedex, France
sebastien.berger@insa-cvl.fr

Abstract

Assessing the dynamic response of vibrating structures which are described by means of finite element (FE) models with many degrees of freedom (DOFs) is usually computationally cumbersome. The fact that the manufacturing process and the material properties of structures are usually subject to variability means a dispersion of the physical parameters which can be important. The parameters are therefore considered as uncertain DOFs, which makes FE models more complex. The Monte Carlo (MC) method is commonly used to analyze the propagation of uncertainties through FE modeling. However, it requires a large number of simulations which are therefore very cumbersome in terms of CPU times. This work aims at developing a low cost computational strategy to compute the harmonic response of vibrating structures having uncertain parameters. The strategy works by considering the Craig-Bampton method to reduce the physical DOFs of a FE model [1]. Also, a sparse Polynomial Chaos (sPC) expansion is considered to describe the propagation of uncertainties and estimate the Quantities of Interest (QoIs), e.g., the displacement at some measurement points, or an energy quantity. In this work, the sPC expansion is applied through a non-intrusive method, which requires a non-negligible number of simulations of the FE model to be performed to estimate the sPC coefficients, and further the statistics (i.e., mean and variance) of the QoIs. The probability law of the QoIs can be obtained by considering the sPC expansions along with the MC method with 10000 trials. The strategy is here applied to model an academic structure composed of three rectangular Kirchhoff-Love plates made up of various materials and connected together across one of their edges by means of a lineic density of springs with an uncertain stiffness. Comparisons with the results obtained from a reference MC solution involving 10000 simulations of the FE model show good agreement and substantial reduction of the computational effort. The influence of the Craig-Bampton reduction method on the estimation of the QoIs of the FE model is discussed through numerical comparisons.

1 Introduction

Applying the FE method to perform the dynamic analysis of complex industrial structures usually involves models characterized by large numbers of DOFs and leads to large computational costs. It may therefore be necessary to reduce the size of the models to solve. One of the efficient model reduction strategies is the well-established Craig-Bampton method, which is based on the projection of the system internal DOFs onto bases of reduced sizes [1, 2].

In addition, the effect of uncertainties are of growing concern in the analysis and design of engineering structures. These uncertainties in the parameters of the system result from the inevitable variability in the manufacturing process of the structures and the fact that the material properties of the structures can change with time [3, 4]. Taking into account these uncertainties in the dynamic analysis of vibrating structures is therefore a crucial issue. The classic MC approach is usually used to analyze the propagation of uncertainties, and involves a large number of FEM evaluations for various sets of values of the uncertain parameters. For complex industrial systems owning a large number of DOFs and/or uncertain parameters, the computational cost associated with this method becomes prohibitive. Polynomial chaos (PC) expansions have proved efficient to solve this issue, among which sparse PC expansions (sPC) [6, 7, 8] are of particular interest when a large number of uncertain parameters is at stake. For instance, Kieu et al. [5] have recently applied sPC expansions to analyze the stability of a clutch system having uncertain parameters. In that study, comparisons with the generalized polynomial

chaos (gPC) and the multi-element generalized polynomial chaos (ME-gPC) previously applied to the same clutch system [9] shew that the sPC expansions ensured substantial time reduction with respect to the other PC expansions, while providing a high accuracy of the results.

The aim of this paper is to associate an sPC expansion to a CB model reduction in order to reduce the computational costs in the analysis of the dynamic behavior of a vibrating structure having uncertain parameters. The Craig-Bampton method is briefly recalled in section 2, and the building of a sparse PC expansion is detailed in section 3. The method is then applied to a system consisting of several plates connected with spring of uncertain stiffness. The details of the model and the results are finally given in section 4.

2 The Craig-Bampton method

The Craig Bampton method [1] aims at reducing the sizes of FE models involving large numbers of DOFs. Within the FE framework, the equation of motion of a structure is

$$[M]\{\ddot{q}\} + [K]\{q\} = \{F\}, \tag{1}$$

where $[M]$ and $[K]$ denote respectively the mass and stiffness matrices of the structure, $\{q\}$ the vector of displacements and $\{F\}$ the vector of external forces. Considering harmonic forces and distinguishing the internal DOFs q_I from the boundary DOFs q_B, the above equation may be rewritten as:

$$-\omega^2 \begin{bmatrix} M_{BB} & M_{BI} \\ M_{IB} & M_{II} \end{bmatrix} \begin{Bmatrix} q_B \\ q_I \end{Bmatrix} + \begin{bmatrix} K_{BB} & K_{BI} \\ K_{IB} & K_{II} \end{bmatrix} \begin{Bmatrix} q_B \\ q_I \end{Bmatrix} = \begin{Bmatrix} F_B \\ F_I \end{Bmatrix}. \tag{2}$$

In the classical FE procedure, the unknown DOFs are obtained by inverting the above system, which can be cumbersome in terms of computational time if the number of DOFs involved is important. The CB method consists in decomposing the vector of internal DOFs onto a basis of static and fixed interface modes as follows:

$$\begin{Bmatrix} q_B \\ q_I \end{Bmatrix} = \begin{bmatrix} I & 0 \\ X_{st} & X_{el} \end{bmatrix} \begin{Bmatrix} q_B \\ \alpha \end{Bmatrix} \tag{3}$$

where X_{st} is the matrix of static modes, which are computed as $-K_{II}^{-1} K_{IB}$, X_{el} is the matrix of fixed interface modes, i.e. the matrix of the eigenvectors of (K_{II}, M_{II}), and α is the vector of the modal amplitudes.

To reduce the size of the problem, only a limited number of fixed interface modes of amplitudes $\tilde{\alpha}$ is retained. The internal DOFs are then approximated by

$$\{\tilde{q}_I\} \approx [X_{st}]\{q_B\} + [\tilde{X}_{el}]\{\tilde{\alpha}\} \tag{4}$$

where $\tilde{X}_{el}$ is a matrix of reduced size. Inserting Eqs. (3) and (4) into Eq. (2) leads to a system of reduced size easier to invert.

3 Sparse Polynomial Chaos

3.1 Generalized polynomial chaos

The generalized polynomial chaos (gPC) has been proposed by Xiu and Karniadakis [10]. It consists in expanding a random process $X(\xi)$ depending on r independent random variables $(\xi_1, ..., \xi_r) = \xi$ as follows:

$$X(\xi) = \sum_{\alpha \in \mathbb{N}^r} \bar{x}_\alpha \phi_\alpha(\xi), \tag{5}$$

where $\phi_\alpha(\xi)$ are orthogonal polynomials which represent the stochastic components of the process, and $\bar{x}_\alpha$ are the PC coefficients that account for the deterministic components of the process.

The Wiener theory as well as the generalized Cameron-Martin theorem [11] state that the series is convergent in the mean square sense. According to the Askey scheme, if ξ is a uniform random vector, the polynomial functions ϕ_α are most suitably obtained from Legendre polynomials [12, 10, 13].

In practice, the random process $X(\xi)$, which constitutes the quantity of interest (QoI), is approached by a truncated expansion as

$$X(\xi) \approx \sum_{\alpha \in \mathcal{A}^{r,p}} \bar{x}_\alpha \phi_\alpha(\xi), \tag{6}$$

where p is the order of the PC expansion and $\alpha = \{\alpha_1, ..., \alpha_r\} \in \mathbb{N}^r$.
The index set used in the truncated expansion (6) is then defined as

$$\mathcal{A}^{r,p} = \{\alpha \in \mathbb{N}^r : \|\alpha\|_1 \leq p\}, \tag{7}$$

with

$$\|\alpha\|_1 = \sum_{i=1}^{r} \alpha_i. \tag{8}$$

Computing the QoI X comes down to finding the coefficients $\bar{x}_\alpha$ of the truncated gPC expansion Eq. (6). The number of terms N_p is linked to the order p and to the number of uncertain parameters r as [10]

$$N_p = card(\mathcal{A}^{r,p}) = \frac{(p+r)!}{p!r!}. \tag{9}$$

In this study, the QoIs are quantities such as a displacement or an energy quantity that are solutions of a FE model. The PC coefficients are here determined from a non-intrusive regression method that does not require any modification of the FE model: they are built from a finite number $Q = kN_p$ (with k a small integer usually equal to 2, 3 or 4) of values of the QoI X, computed from numerical Q simulations of the FE model. In practice the Q sets of values of the uncertain parameters for which the QoI is computed, which will be referred to as the nested expermimental design (NED) in the following, may be chosen with a Latin Hypercube Samples (LHS) method [14].

Within the regression framework, the evaluation of the coefficients results from the minimization of the following criterion [15]

$$\varepsilon^2_{reg} = \sum_{q=1}^{Q} \left[X\left(\xi^{(q)}\right) - \sum_{\alpha \in \mathcal{A}^{r,p}} \bar{x}_\alpha \phi_\alpha\left(\xi^{(q)}\right) \right]^2, \tag{10}$$

where $\xi^{(q)} = (\xi_1^{(q)}, ..., \xi_r^{(q)})$ (with $q = 1, ..., Q$) denotes the Numerical Experimental Design (NED), that is the set of Q vectors of uncertain parameter values generated from the probabilistic support of the parameters; $X\left(\xi^{(q)}\right)$ denotes the vector of the corresponding FE model evaluations. The PC coefficients are finally calculated as

$$\bar{x} = \left(\phi^T(\xi^{(q)})\phi(\xi^{(q)})\right)^{-1} \phi^T(\xi^{(q)}) X(\xi^{(q)}), \tag{11}$$

with $\phi(\xi^{(q)})$ the matrix defined by

$$\phi(\xi^{(q)}) = \begin{pmatrix} \phi_0(\xi^{(1)}) & \dots & \phi_{N_p-1}(\xi^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi_0(\xi^{(Q)}) & \dots & \phi_{N_p-1}(\xi^{(Q)}) \end{pmatrix}. \tag{12}$$

If the number of uncertain parameters and the order p of the gPC expansion are high, the number of PC coefficients and therefore the necessary number of simulations to build them become quickly prohibitive. Strategies to reduce this number of simulations are consequently necessary.

3.2 Sparse Polynomial Chaos

The sparse Polynomial Chaos (sPC) can reduce the number of PC coefficients. In this paper, sPC with anisotropic hyperbolic index sets will be used.

3.2.1 Anisotropic hyperbolic index sets

The strategy to truncate the PC expansions favors input random variables ξ_i with large total sensitivity indices S_i^T. For this purpose, the truncation is based on the following anisotropic norm

$$\|\alpha\|_{m,w} = \left(\sum_{i=1}^{r} |w_i \alpha_i|^m\right)^{1/m}, \qquad w_i \geq 1. \tag{13}$$

The corresponding anisotropic index set is then chosen as

$$\mathcal{A}_{m,w}^{r,p} = \{\alpha \in \mathbb{N}^r : \|\alpha\|_{m,w} \leq p\}, \tag{14}$$

where w is a set of weights w_i defined by

$$w_i = 1 + \frac{max_{1\leq j\leq r} S_j^T - S_i^T}{\sum_{k=1}^{r} S_k^T}. \quad i = 1, ..., r. \tag{15}$$

In the above equation, S_i^T is the PC-based total sensitivity index [16] of the QoI with respect to the input random variable ξ_i, and is computed as

$$S_i^T = \frac{1}{D_{PC}} \sum_{\alpha \in \mathcal{I}_i^+} \bar{x}_\alpha^2 \mathrm{E}[\phi_\alpha^2(\xi)], \tag{16}$$

where $\mathcal{I}_i^+$ denotes the set of indices having a non-zero i^{th} component

$$\mathcal{I}_i^+ = \{\alpha \in \mathcal{A}_{m,w}^{r,p} : \alpha_i \neq 0\}, \tag{17}$$

and D_{PC} the variance of the QoI

$$D_{PC} = \sum_{\alpha \in \mathcal{A}_{m,w}^{r,p}} \bar{x}_\alpha^2 \mathrm{E}[\phi_\alpha^2(\xi)]. \tag{18}$$

The anisotropic hyperbolic polynomial chaos expansions are finally defined with the index sets $\mathcal{A}_{m,w}^{r,p}$ as

$$X_{\mathcal{A}_{m,w}^{r,p}}(\xi) = \sum_{\alpha \in \mathcal{A}_{m,w}^{r,p}} \bar{x}_\alpha \phi_\alpha(\xi). \tag{19}$$

3.2.2 Error estimates of the polynomial chaos approximations

The building of a sparse PC expansion is based on an iterative search of the significant PC coefficients, and therefore requires the use of error estimates to assess the accuracies of the consecutive PC approximations.

A relevant theoretical error in this context is defined as follows:

$$Err = \mathrm{E}[(X(\xi) - \widehat{X}_{\mathcal{A}}(\xi))^2], \tag{20}$$

which is based on the difference between the deterministic evaluation $X(\xi)$ of the QoI and its PC approximation $\widehat{X}_{\mathcal{A}}(\xi)$ computed from a finite non empty subset $\mathcal{A} \subset \mathbb{N}^r$, that is

$$\widehat{X}_{\mathcal{A}}(\xi) = \sum_{\alpha \in \mathcal{A}} \bar{x}_\alpha \phi_\alpha(\xi). \tag{21}$$

The generalization error is estimated in practice by the following empirical error:

$$Err_{emp} = \frac{1}{Q} \sum_{q=1}^{Q} \left[\left(X(\xi^{(q)}) - \widehat{X}_{\mathcal{A}}(\xi^{(q)})\right)^2\right] \tag{22}$$

in which the differences are computed specifically at the Q observations of a NED $\xi^{(q)} = \left(\xi_1^{(q)}, ..., \xi_r^{(q)}\right)$. The latter will be used in the following to compute a coefficient of determination R^2 defined as

$$R^2 = 1 - \frac{Err_{emp}}{\widehat{\mathrm{V}}[X]}, \tag{23}$$

where $\widehat{\mathrm{V}}[X]$ is the variance of $X\left(\xi^{(q)}\right)$:

$$\widehat{\mathrm{V}}[X] = \frac{1}{Q-1}\sum_{q=1}^{Q}\left(X(\xi^{(q)}) - \bar{X}\right)^2 \qquad \text{with} \qquad \bar{X} = \frac{1}{Q}\sum_{q=1}^{Q}X(\xi^{(q)}).$$

An overfitting phenomenon is likely to occur when using the empirical error, which, as a consequence, underestimates the generalization error. The leave-one-out error [17], which is based on a sum of squared predicted residuals $\Delta^{(i)}$ defined hereafter, may be useful to avoid this drawback. A predicted residual expresses the difference between the deterministic evaluation $X(\xi^{(i)})$ of the QoI at the i^{th} observation of the NED $\xi^{(q)}$, and its prediction $\widehat{X}_{\mathcal{A}}^{(-i)}(\xi^{(i)})$ obtained with a PC expansion $\widehat{X}_{\mathcal{A}}^{(-i)}$ built from a reduced NED $\left(\xi^{(1)},...,\xi^{(Q)}\right)\backslash\xi^{(i)}$ (that is the original NED from which the observation $\xi^{(i)}$ has been discarded) [7]:

$$\Delta^{(i)} = X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}^{(-i)}(\xi^{(i)}). \tag{24}$$

The leave-one-out error is then defined as

$$Err_{LOO} = \frac{1}{Q}\sum_{i=1}^{Q}(\Delta^{(i)})^2. \tag{25}$$

In practice, the predicted residual $\Delta^{(i)}$ may be computed as [17]

$$\Delta^{(i)} = \frac{X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}(\xi^{(i)})}{1-h_i}, \tag{26}$$

where h_i is the i^{th} diagonal term of the matrix $\phi(\xi^{(q)})(\phi^T(\xi^{(q)})\phi(\xi^{(q)}))^{-1}\phi^T(\xi^{(q)})$. The leave-one-out error is in that case given by

$$Err_{LOO} = \frac{1}{Q}\sum_{i=1}^{Q}\left(\frac{X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}(\xi^{(i)})}{1-h_i}\right)^2. \tag{27}$$

A determination coefficient S^2 equivalent to that of the empirical error, R^2, may be defined for the leave-on-out error:

$$S^2 = 1 - \frac{Err_{LOO}}{\widehat{\mathrm{V}}[X]}. \tag{28}$$

The two coefficients R^2 and S^2 defined above will be used in an algorithm whose aim is to build an optimal sparse PC expansion involving the most significant terms from an adapted NED of reduced size. This algorithm is described in the next section.

3.2.3 sPC expansion building algorithm

As explained previously, the efficiency of the method may be increased by retaining only the most significant PC polynomials [7] among those corresponding to the index sets $\mathcal{A}_{m,w}^{r,p}$. In the following, the final index sets of the kept terms are denoted as $\mathcal{A}_{m,w}^{p}$.

The search for those most significant coefficients is performed through an iterative procedure which is summarized below in 5 basic steps.

Step 1

Select a NED $(\xi^{(q)})$, e.g. a random design based on LHS [14], of arbitrary size $Q_k = 4N_p$, where N_p is determined by Eq. (9) with r uncertain parameters and $p = 1$. The FE model evaluations at the NED points are gathered in the vector $X(\xi^{(q)})$. Set arbitrarily the values of the parameters corresponding to the chosen sparse PC method: the maximal PC order p_{max} and the coefficient m used for the m-norm of truncation, as well as the target accuracy S^2_{target} and two thresholds ϵ_1 and ϵ_2.

Step 2

Initialize the algorithm: the PC order is set to $p = 0$, and the truncation index set to the null element of $\mathbb{N}^r$, $\{0\}$; the vector of weights w_i is set to $w = \{1,...,1\}$. The corresponding initial values of the determination coefficients are denoted as R^2_0 and S^2_0.

Step 3: Training step - Enrichment of the PC basis

Increment the order value: $p \to p+1 \in [1, ..., p_{max}]$.
$\Rightarrow$ **Forward step** (Addition step): For each term from the candidate set $\{\alpha \in \mathbb{N}^r : p-1 \leq \|\alpha\|_{m,w} \leq p\}$, add it to the set $\mathcal{A}_{m,w}^{p-1}$ and compute, as above, the PC coefficients and the determination coefficient R^2. Retain only terms which lead to a significant increase in the value of the coefficient R^2, i.e. $\Delta R^2 = R^2 - R_0^2 \geq \epsilon_1$. Let $\mathcal{A}_{m,w}^{p+}$ be the final truncation set at this stage.
$\Rightarrow$ **Backward step** (Elimination step): Remove in turn each term in $\mathcal{A}_{m,w}^{p+}$ of order strictly lower than p, and compute again the PC expansion coefficients and the associated coefficient R^2 in each case. Discard from $\mathcal{A}_{m,w}^{p+}$ the terms that lead to an insignificant decrease in R^2, i.e. $\Delta R^2 = R_0^2 - R^2 < \epsilon_2$. Let $\mathcal{A}_{m,w}^{p}$ be the final truncation set. The total sensitivity indices S_i^T of the current PC approximation are computed and the weights w_i are updated (Eq. (15)).

Step 4: Verification of the conditioning of the regression information matrix

If the conditioning is satisfying, i.e. the size Q_k of the NED ($\xi^{(q)}$) is larger than 2.card($\mathcal{A}_{m,w}^{p}$) go to step 5. If the conditioning is poor, i.e. the size Q_k of the NED ($\xi^{(q)}$) is smaller than 2.card($\mathcal{A}_{m,w}^{p}$), an enrichment of the NED is done using nested Latin Hypercube designs [18, 7] to reach a size Q_{k+1}. In this case, the truncation set is reset to {0} and the enrichment procedure is restarted from step 2.

Step 5: Test step

Stop if either the leave-one-out error S_0^2 is larger than the target value S_{target}^2 or if the order of the PC expansion is equal to p_{max}. Otherwise, go back to step 3.

The detailed algorithms are presented in Figure 1.

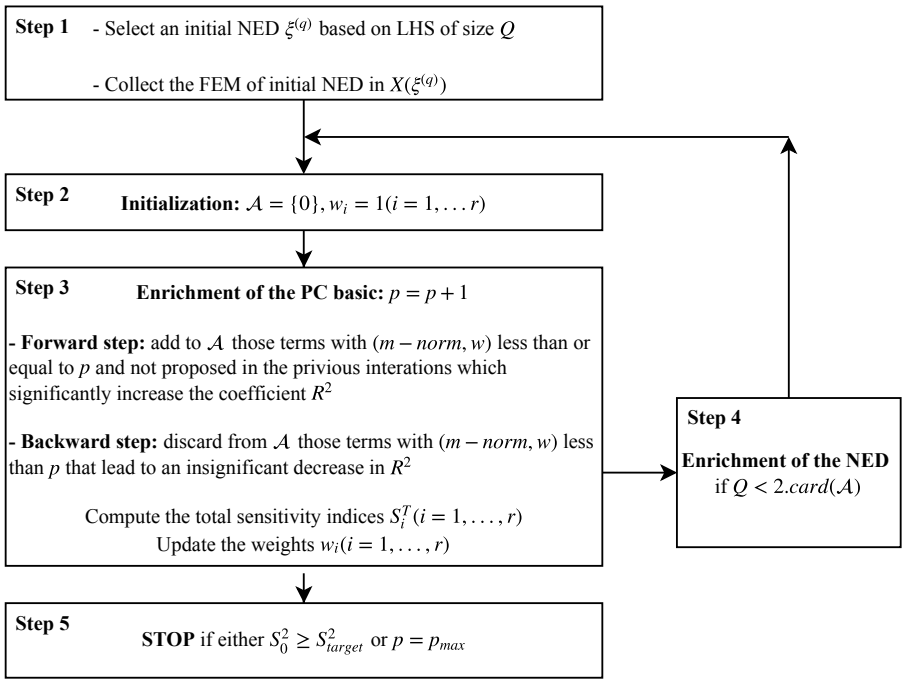


Figure 1: Algorithm applied to build a sparse polynomial chaos expansion with anisotropic hyperbolic index sets

4 Results

4.1 Application model of three plates connected by springs with uncertain stiffness

The method is applied to a system composed of three plates connected together through torsional and translational lineic springs as shown in Figure 2. Plates 1 and 3 are identical 1 m x 1 m square plates made of steel, with the following material properties: density $\rho = 7850 kg/m^3$, Young's modulus $E = 2 \times 10^{11} Pa$ and Poisson's ratio $\nu = 0.3$. The remaining plate 2, of dimensions 0.2 m x 1 m, represents a soft junction made up of rubber; the corresponding material properties are chosen as $\rho = 950 kg/m^3$, $E = 15 \times 10^7 N/m^2$ and $\nu = 0.48$. Within the FE framework, the three plates of same thickness 5 mm are meshed using square plate elements of

length $0.025\,m$ having three DOFs per node, namely the displacement u_z and two rotations θ_x, θ_y. The meshes of plates 1 and 3 therefore involves 1600 elements and 5043 DOFs, while 320 elements and 1107 DOFS are used for plate 2.

The torsional stiffness of the springs along the $x-$ and $y-$ directions is assumed to be uniform and equal to 20 Nm/rad, whereas in the $z-$direction the stiffness k_z is supposed to represent a random variable following a uniform probability law over the range $[100, 200]\,N/m$. The whole structure is clamped at both extremities (i.e., left edge of plate 1 and right edge of plate 3) and subjected to a harmonic point force of amplitude $F = 40N$ in the $z-$direction, located at the node of coordinates (0.25 m, 0.25 m) if the origin is chosen as the lower left corner of plate 1.

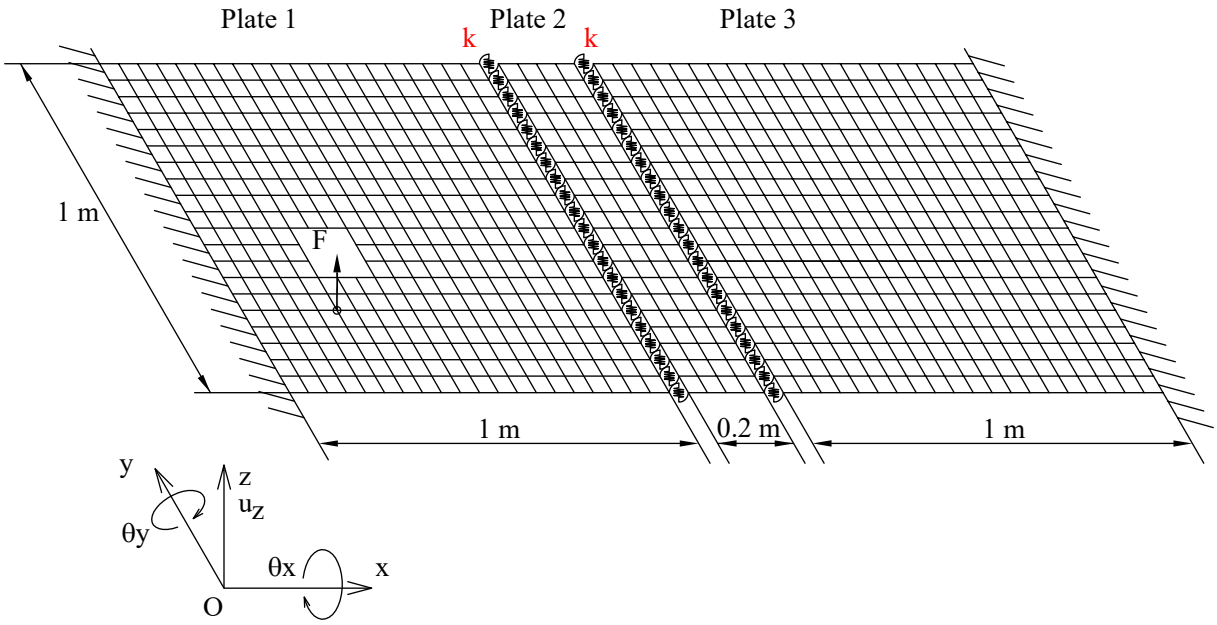


Figure 2: Model of three plates connected by springs

The frequency response function of the structure is studied within the frequency band $[0, 50]Hz$ using a frequency step of $10^{-3}\,Hz$. Examples of FRFs are provided in Figure 3, which represents the frequency evolutions of the deformation energy of plate 1, E_{def1}, for three values of the spring stiffness k_z (namely the two extreme values $100\,N/m$ and $200\,N/m$, and the nominal value $150\,N/m$). As it can be seen in Figure 3, the curves present a similar trend with extrema at the resonance frequencies of the system, but the amplitudes and the frequencies of those peaks depend on the value of the stiffness k_z.

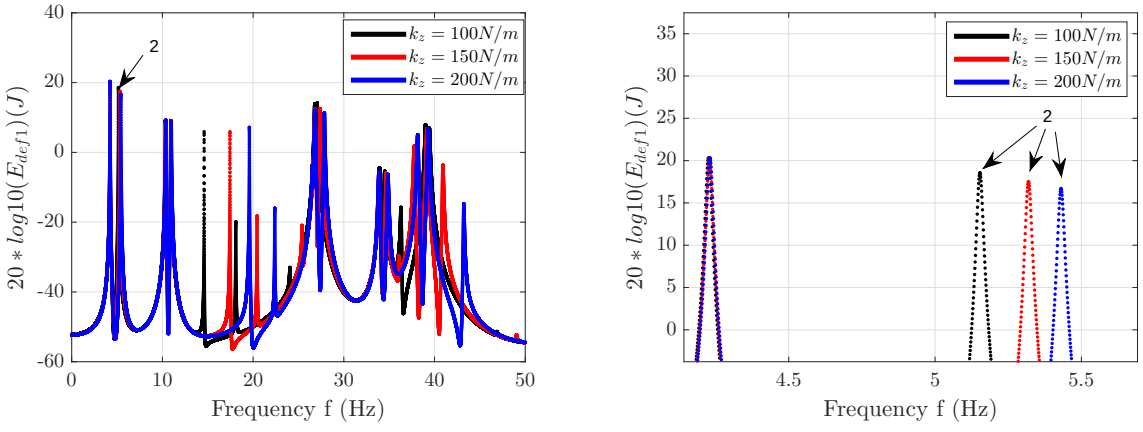


Figure 3: Deformation energy E_{def1} of plate 1: (a) whole frequency range (b) zoom on the first two peaks

4.2 Quantities of interest and statistics

In the following, we will focus on two particular QoIs related to the second peak visible on the E_{def1} curves, which is displayed in a detailed view in Figure 3(b): the frequency and the amplitude of the peak. Both QoIs are determined from a maximum search over a small frequency band around the frequency of the peak obtained at the nominal value of k_z.

The statistics (e.g. mean value and variance) of each QoI are computed within a Monte-Carlo procedure involving a number N of simulations that correspond to N values of k_z chosen uniformly within the range $[100, 200]N/m$. Those N simulations may involve:

- the initial FEM model, which involves a total number of 11193 DOFs; the statistics obtained with this model will be considered as the reference solution in the following;
- a reduced model resulting from a Craig-Bampton decomposition in which a limited number of fixed interface modes is retained; two CB models will be proposed in the following: the first one, denoted as CB50130, involves 50 fixed interface modes for plates 1 and 3 (out of 4563 modes), and 130 modes for plate 2 (out of 819 modes); in the second one, denoted as CB3080, 30 modes are retained for plates 1 and 3 and 80 for plate 2;
- the sPC expansion of the QoI in which the coefficients have been built from a limited number of simulations using the initial FE model, denoted as sPC-FEM;
- an sPC expansion whose coefficients are built from simulations involving one of the two aforementioned reduced CB models, denoted respectively as sPC-CB50130 and sPC-CB3080.

For each QoI, comparisons between the statistics computed from the first three models will provide information on the direct influence of the CB reduction on the accuracy of the results with respect to the full initial FE model. At a second level, comparisons between the results from the FEM and sPC-FEM methods will give insight into the influence of the use of a sparse PC expansion on the accuracy of the statistics. Finally, the influence of the model reduction on the sPC expansions will be studied through the last sPC-CB50130 and sPC-CB3080 methods.

In the following, the results related to the first QoI, namely the resonance frequency of the second peak of the deformation energy E_{def1}, are first gathered in section 4.3. The results related to the amplitude of the peak are then presented in section 4.4.

4.3 E_{def1} peak 2 resonance frequency

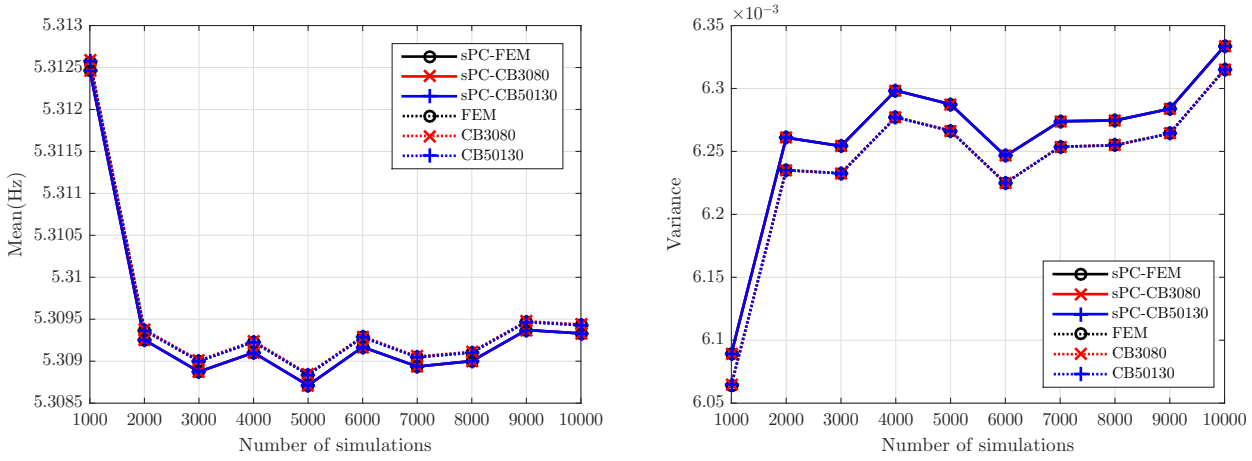
4.3.1 Building of the sPC expansions

The different sPC expansions of the two QoIs are built in accordance with the iterative procedure detailed in section 3.2.3. The same set of parameters is used in all the cases: the maximal PC order is set to $p_{max} = 6$, the target accuracy is chosen as $S^2_{target} = 0.999$, and we use two identical thresholds $\epsilon_1 = \epsilon_2 = 0.001(1 - S^2_{target})$.

As explained previously, the optimal order p of the sPC expansion of the QoI depends on the leave-one-out error S^2_0 that is computed at each step of the iterative procedure and compared to the target accuracy S^2_{target}. For the first QoI (peak 2 resonance frequency), the values of S^2_0 obtained for the three sPC expansions (sPC-FEM, sPC-CB50130 and sPC-3080) are $S^2_0 = 0.9820$ for $p = 1$ and $S^2_0 = 0.9997$ for $p = 2$. The optimal order for the three sPC expansions is therefore $p = 2$.

4.3.2 Accuracy of the QoI statistics obtained from the different methods

The statistics (mean and variance) of the QoI depend on the number of simulations N chosen within the MC procedure. Several tests ranging from $N = 1000$ to $N = 10000$ have been performed for each of the six methods detailed in the previous section. Figure 4 compares the statistics obtained with the six strategies detailed in the previous section for ten values of N ranging from $N = 1000$ to $N = 10000$. The mean value and the variance of the frequency are seen to reach a stabilized value from $N = 2000$. Two groups of curves are visible on each graph, meaning that the use of an sPC expansion (whatever the model chosen to build the PC coefficients)



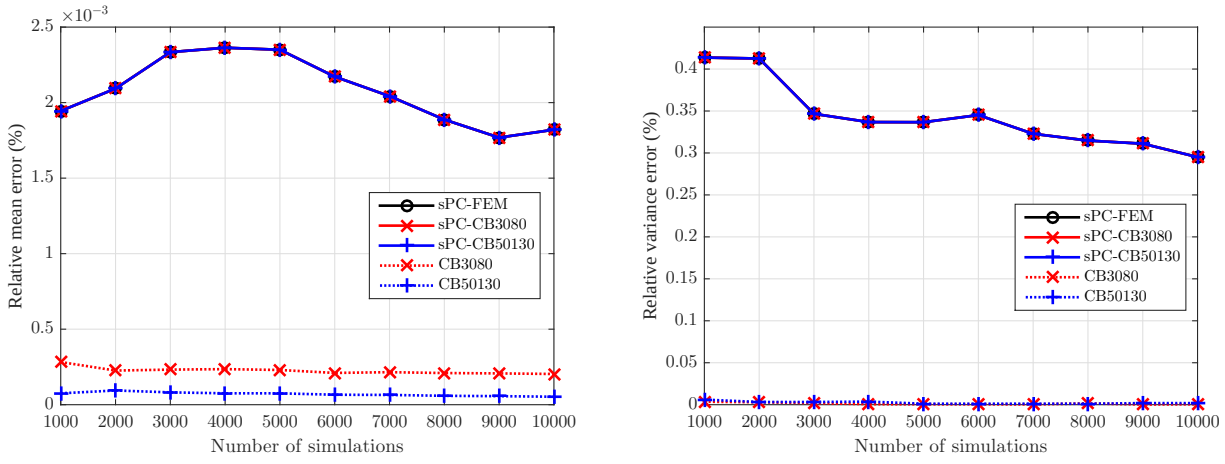
(a) (b)

Figure 4: (a) Mean and (b) variance of the QoI according to the number of MC simulations

induces a slight change in the resulting mean and variance values; the use of a model reduction does not modify the statistics significantly.

The low variance levels obtained in Figure 4(b) also reduce the confidence intervals relative to the mean resonance frequency. For instance, for $N = 10000$, the mean resonance frequency of peak 2 lies in the interval $\bar{f}_2 \pm 1.2\,10^{-4}$ with a confidence level of 95 %.

To further analyze the accuracies of the different method, the relative errors of the mean value of the QoI and its variance with respect to the reference solution (i.e. N simulations from the initial FE model) are displayed in Figure 5. The lowest error levels are logically obtained with the CB50130 method, which involves N direct simulations with the reduced CB model with the largest basis of fixed interface modes. Further reducing the size of the mode basis slightly increases the error levels, but the accuracy remains in both cases excellent, with error levels close to 0. The use of an sPC expansion increases the error levels, which remain however lower than $2.5\,10^{-3}$ % for the mean value and 0.5 % for the variance. Choosing a reduced model instead of the original FE model to compute the coefficients of the sPC does not increase the error levels for this QoI.

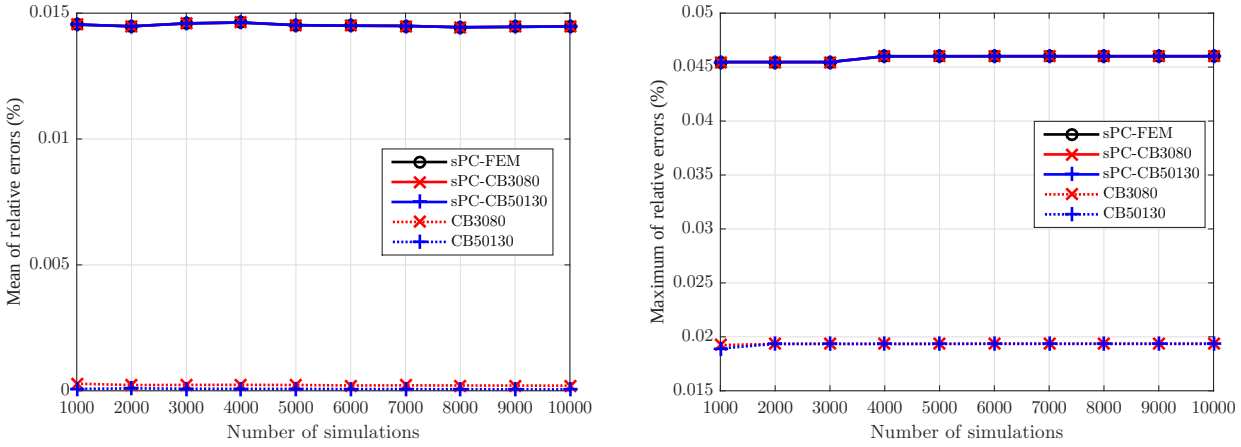


(a) (b)

Figure 5: Relative errors of (a) the mean value of the QoI and (b) the variance of the QoI with respect to the reference solution according to the number of MC simulations

Same conclusions arise when studying the mean (Figure 6(a)) and maximum (Figure 6(b)) values of the relative errors of the QoI for a given N number of simulations (i.e. a relative error is computed for each of the $N k_z$ values between the peak 2 frequency obtained with the considered method and the frequency found using the initial FE model; the mean value is then computed over the k_z range $[100, 200]\,N/m$, along with the maximum value). The errors mainly come from the sPC expansions, the influence of the model reduction being again very limited; the accuracy level is also extremely satisfying with a mean relative error lower than 0.015 %

and maximum error lower than 0.05 % whatever the retained N number of simulations.

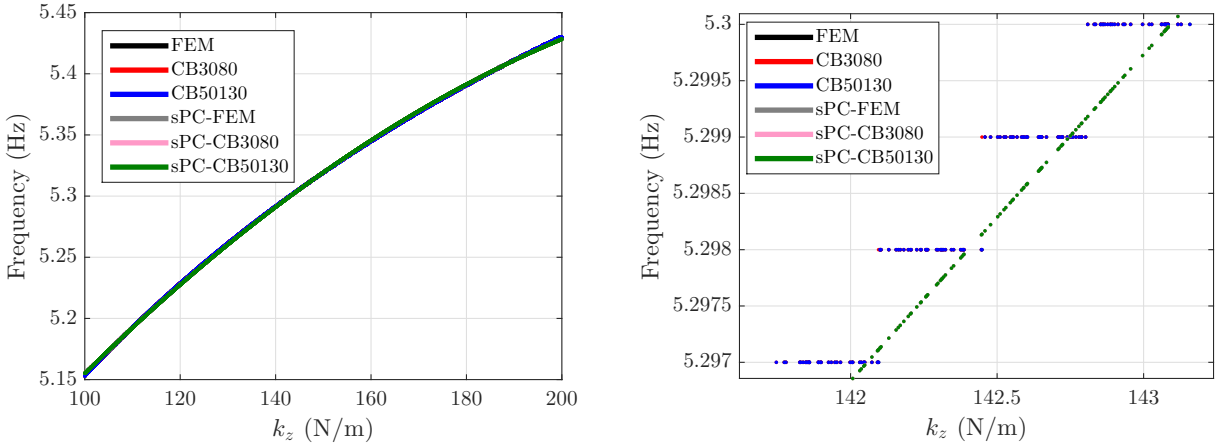


(a) (b)

Figure 6: (a) Mean and (b) maximum values of the relative errors of the QoI with respect to the reference solution according to the number of simulations

4.3.3 Evolutions of the resonance frequency according to k_z

The results in this section are presented for the highest number $N = 10000$ of simulations, to ensure that the QoI is predicted with the highest confidence level. Figure 10 represents the values of the second resonance frequency obtained for each of the 10000 k_z values chosen within the range $[100, 200]\,N/m$ with the different methods: direct simulations with the initial FE model or the reduced ones, or estimations from the three sPC expansions. The QoI exhibits a non linear increasing dependency to the stiffness k_z, with a global variation of about 5 % with respect to its mean value. An overall good agreement is found between the different strategies, as the six curves appear superimposed of the whole k_z range (Figure 10(a)). However, when zooming on a smaller range of k_z values, such as in Figure 10(b), the slight differences behind the error levels presented previously become visible. The curves corresponding to the three methods that involve direct FE simulations (FEM, CB50130 and CB3080) present discontinuities that are linked to the frequency step used in this study, $0.001\,Hz$; the peak frequency is either underestimated or overestimated according to the k_z value as its precision cannot exceed the frequency resolution of the simulations. Predicting the frequency value from an sPC expansion avoids this behavior, as the frequency becomes a polynomial function of k_z. The curves corresponding to the three sPC expansions are therefore continuous plots that can be hardly distinguished from one another.



(a) (b)

Figure 7: Peak 2 resonance frequency according to the spring stiffness k_z for $N = 10000$

4.4 Amplitude of the deformation energy peak 2

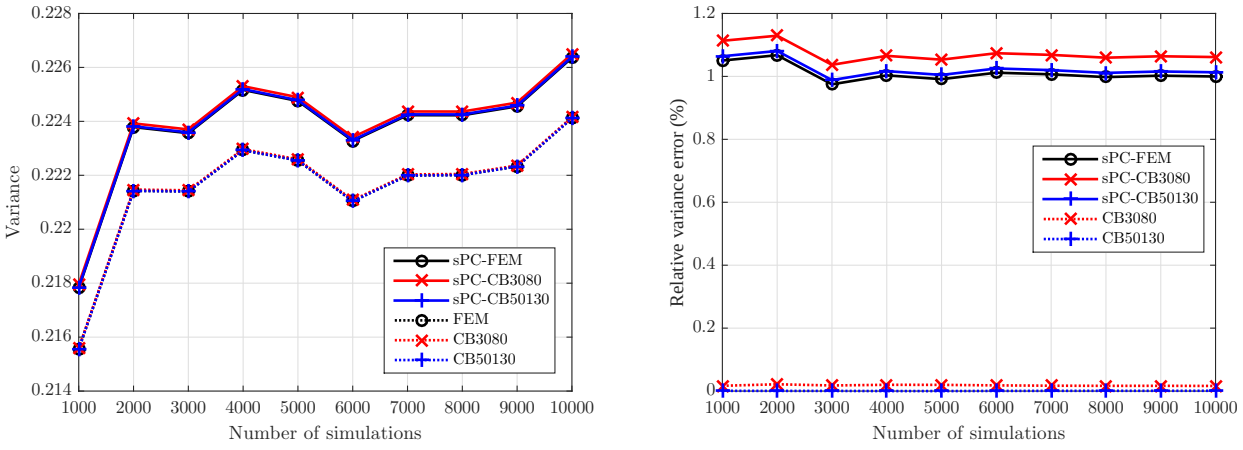
4.4.1 Building of the sPC expansions

The sPC expansions of the second QoI are built using the same parameters and iterative procedure as the resonance frequency. Leave-one-out error S_0^2 is using to choose the optimal order p of the polynomial chaos to calculate QoI that is E_{def1} in this section. The leave-one-out errors S_0^2 for the second QoI are similar for $p = 1$ and $p = 2$ to those obtained for the frequency, the optimal PC order being again $p = 2$.

4.4.2 Accuracy of the QoI statistics

As for the resonance frequency, the statistics related to the amplitude of the deformation energy at the second peak are computed from N simulations corresponding to N values of k_z, from $N = 1000$ to $N = 10000$. The same trends as for the first QoI are retrieved, that will be illustrated hereafter for the variance only.

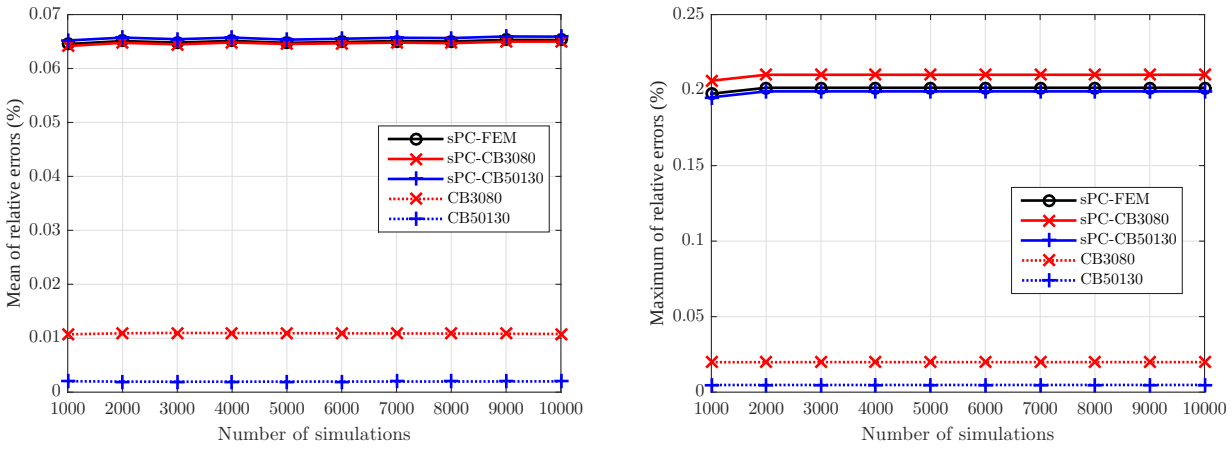
Figure 8(a) displays the variance of the QoI with respect to the number of simulations N, a stabilized value being again reached from $N = 2000$. As for the first QoI (frequency), two groups of curves emerge from the figure, which correspond respectively to the simulations involving the initial or reduced FE models, and to those based on the sPC expansions. The relative errors between the variance values with respect to the reference solution are shown in Figure 8(b), and exhibit overall very satisfying levels although they are higher than those obtained for the the resonance frequency. The use of a reduced basis instead of the initial FE model has again little impact in terms of variance errors, as the highest level error, corresponding to the CB3080 method, is about 0.02 %. The influence of the sPC expansions on the accuracy of the variance estimation is more important, as they induce error levels of about 1 %. The best results are obtained when the sPC coefficients are computed using the initial FE model (sPC-FEM case) while the error levels increase up to 1,1 % with the sPC-CB3080 methods.



(a) (b)

Figure 8: (a) Variance and (b) relative error of the variance of the second QoI according to the number of MC simulations

Figure 9 displays the mean and maximum error levels attained when computing the relative errors, for each k_z value, between the peak 2 amplitudes resulting of a given method (involving a reduced basis or an sPC expansion) and those obtained with the reference solution (involving the initial FE model). As previously, the error levels are constant from $N = 2000$ and result mainly from the use of an sPC expansion, while the CB reduction induces low additional errors. All the methods provide an excellent accuracy with overall very low error levels. For instance, the mean error level is about 0.002 % with the CB50130 method, 0.01 % with the CB3080 method and 0.065 % with any of the sPC expansions. The maximum error values are of the same order of magnitude, from 0.005 % and 0.02 % with the reduced models (CB50130 and CB3080 respectively) to 0.2 % with the sPC expansions, the highest value being obtained when the sPC coefficients are computed using the CB2080 model.



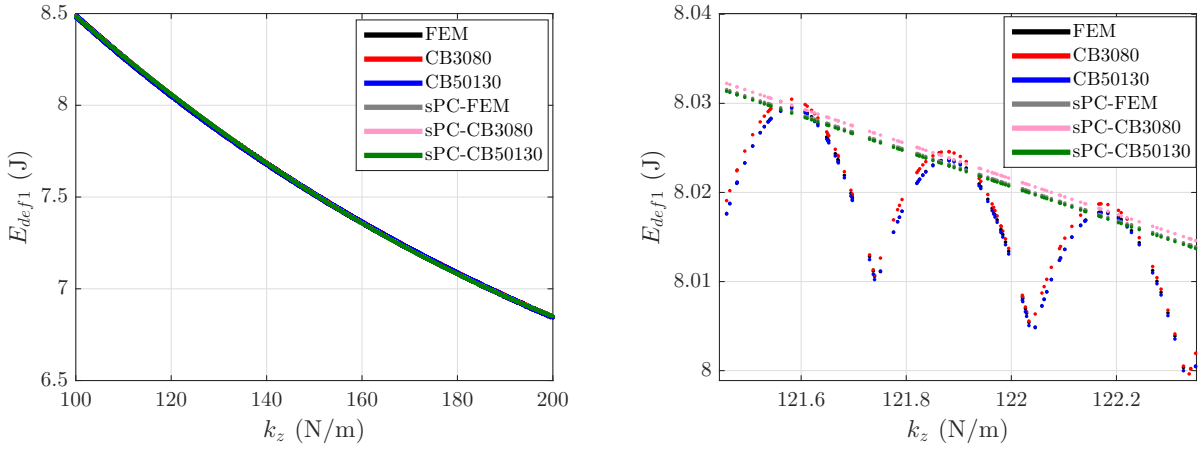
(a) (b)

Figure 9: (a) Mean and (b) maximum values of the relative errors of the QoI with respect to the reference solution according to the number of simulations

4.4.3 Evolutions of the amplitude of peak 2 according to k_z

The evolutions of the amplitude of the second resonance peak of E_{def1}, directly computed with $N = 10000$ simulations of the initial FE model or a reduced CB model, or predicted by one of the three sPC expansions, are finally plotted in Figure 10. The QoI is here a decreasing non linear function of the stiffness k_z, and the six curves appear again superimposed throughout the whole k_z range. The discrepancies between the solutions are revealed in the detailed view (Figure 10(b)), where two groups of curves are again visible. The first group gathers the simulations involving direct FE simulations, for which the frequency of peak 2 could be under- or overestimated due to the limited frequency step of $0.001\,Hz$ (cf. Figure 10(b)). These approximations in the resonance frequency values lead to an underestimation of the peak amplitude and results in the uneven appearances of the curves. The latter reach their maximum values when the peak resonance frequency value coincides with a multiple value of the frequency step. The use of an sPC expansion to predict the peak amplitude appears again efficient to solve this issue, as the resulting curves are smooth and match the maximum values of the first group.

Regarding the accuracy of the predictions, it can be seen that within the first group, the CB50130 is perfectly superimposed with the reference FEM curve, while the CB3080 is slightly shifted (but remains very close to the first two curves). This trend is retrieved in the second group, where the curve corresponding to the sPC-CB3080 does not perfectly superimpose with the two other curves, although the differences are extremely weak.



(a) (b)

Figure 10: Peak 2 amplitude according to the spring stiffness k_z for $N = 10000$

4.5 Computational costs

The previous results have clearly established that an sPC expansion was efficient to produce accurate predictions of the different QoIs evolutions with respect to the uncertain parameter (i.e. the spring stiffness k_z). The main interest of the method is that the computational costs are also extremely decreased due to the limited number of simulations necessary to compute the sPC coefficients. To illustrate this benefit, the computational costs per QoI related to the six aforementioned methods are presented in table 1 for $N = 10000$. The reference solution is obviously the most expensive as it implies $N = 10000$ simulations with the FE model owning the highest number of DOFs. Substantial cost reductions are already achieved when using a FE model of reduced size instead of the original model to perform those 10000 simulations. With the most reduced basis (CB3080), the computational time reduction reaches 73.16 % without loss of accuracy. The advantage of the sPC expansions is that the number of simulations necessary to compute the PC coefficients is very limited. Applying the iterative procedure depicted previously, we obtain the coefficients for $p = 2$ and one uncertain paramters with only 14 simulations. Once the coefficients are known, performing 100000 computations to calculate the QoI is costless. The computational time reduction therefore exceeds 99 % for the three proposed sPC expansions. The highest reduction (99.96 %) is logically obtained when the coefficients are computed using the most reduced CB3080 model.

Method	FEM	CB50130	CB3080	sPC-FEM	sPC-CB50130	sPC-CB3080
Computational time formula	Nb of FEM x Unit time per full FEM	Nb of reduced FEM x Unit time per reduced FEM	Nb of reduced FEM x Unit time per reduced FEM	Nb of FEM to build chaos x Unit time per full FEM	Nb of reduced FEM to build chaos x Unit time per reduced FEM	Nb of reduced FEM to build chaos x Unit time per reduced FEM
Nb of simulations with full or reduced FEM	10000	10000	10000	14	14	14
Computational time	15 days 20h	5 day 6h	6 days 6h	32 min	11 min	9 min
Computational time reduction (%)	-	66.84	73.16	99.86	99.95	99.96

Table 1: Computational costs for one QoI and $N = 10000$ with the different methods

5 Conclusion

In this paper, we have proposed a strategy to analyze the dynamic response of a structure having a large number of DOFs and uncertain parameters in the FE framework. The retained method associates the use of a sparse PC expansion to compute the quantities of interest (e.g. a displacement or a energy quantity such as a deformation energy) and a model reduction based on the Craig-Bampton decomposition to obtain at low cost the PC coefficients.

This strategy has been successfully applied to the case of a structure, composed of several plates connected with springs, presenting one uncertain parameter, namely the spring stiffness k_z. The CB reduction has shown to produce a negligible loss of accuracy while ensuring a substantial reduction of the computational cost. Combining the use of such a reduced model with an sPC expansion, the accuracy of the results remains fully satisfying and the computational time reduction reaches exceptional levels.

In a near future, the proposed strategy will be applied to complex systems characterized by a larger number of uncertain parameters. The benefits of using a sparse PC expansion instead of the classic generalized PC expansion will then become more evident, as the computational cost necessary to compute the PC coefficients can also become prohibitive when a large number of uncertain parameters is involved.

References

[1] R.R. Craig and M.C.C. Bampton. Coupling of Substructures for Dynamic Analyses. *AIAA Journal*, 6(7):1313–1319, 1968.

[2] *Dynamic Model Reduction Methods*, pages 265–280. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[3] B.R. Mace, K. Worden, and G. Manson. Uncertainty in structural dynamics. *Journal of Sound and Vibration*, 288(3):423 – 429, 2005. Uncertainty in structural dynamics.

[4] G.I. Schuëller. On the treatment of uncertainties in structural mechanics and analysis. *Computers & Structures*, 85(5):235 – 243, 2007. Computational Stochastic Mechanics.

[5] Kieu, D.T., Bergeot, B., Gobert, M.-L., and Berger, S. Stability analysis of a clutch system with uncertain parameters using sparse polynomial chaos expansions. *Mechanics & Industry*, 20(1):104, 2019.

[6] G. Blatman and B. Sudret. Anisotropic parcimonious polynomial chaos expansions based on the sparsity-of-effects principle. In *Proc ICOSSAR'09, International Conference in Structural Safety and Relability*, 2009.

[7] G. Blatman and B. Sudret. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25(2):183–197, 2010.

[8] G. Blatman and B. Sudret. Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *Comptes Rendus Mécanique*, 336(6):518 – 523, 2008.

[9] M.H. Trinh, S. Berger, and E. Aubry. Stability analysis of a clutch system with multi-element generalized polynomial chaos. *Mechanics & Industry*, 17(2):205, 2016.

[10] D. Xiu and G.E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, 2002.

[11] R.H. Cameron and W.T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[12] R. Askey and J.A. Wilson. *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, volume 319. American Mathematical Soc., 1985.

[13] D. Xiu and G.E. Karniadakis. Modeling uncertainty in steady state diffusion problems via generalized polynomial chaos. *Computer Methods in Applied Mechanics and Engineering*, 191(43):4927–4948, 2002.

[14] M.D. McKay, R.J. Beckman, and W.J. Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245, 1979.

[15] M. Berveiller, B. Sudret, and M. Lemaire. Stochastic finite element: a non intrusive approach by regression. *European Journal of Computational Mechanics/Revue Européenne de Mécanique Numérique*, 15(1-3):81–92, 2006.

[16] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964 – 979, 2008. Bayesian Networks in Dependability.

[17] G. Blatman. *Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis*. PhD thesis, Université BLAISE PASCAL - Clermont II, 2009.

[18] G.G. Wang. Adaptive response surface method using inherited Latin Hypercube design points. *Journal of Mechanical Design*, 125(2):210–220, 2003.

On the Monitoring of Noisy Data as a Multidimensional Shell

Martin GAGNON [1], François LEONARD [1], James MERLEAU [1], Dominique TAPSOBA [1]

[1] Institut de recherche d'Hydro-Québec (IREQ), 1800 boulevard Lionel-Boulet, Varennes, QC, Canada J3X 1S1
gagnon.martin@ireq.ca

Abstract

Based on the idea that multidimensional data is better summarized as a shell rather than a cloud, we have developed a surveillance approach that can detect with high sensitivity behaviour changes in a monitored process and alert the operator. Our methodology uses the time series of a high number of monitored indicators which we cluster together dynamically as a function of operating conditions. These clusters represent groups of similar realizations used to characterize a multidimensional manifold that can be interpolated to assess each new realization of the process behaviour. We evaluated the methodology on the data from a hydroelectric turbine. The event of interest was the loss of the turbine propeller runner cone. The results are good and the approach is able to detect the abnormal behaviour months before the event happened. We are currently looking at larger scale deployment to benchmark the approach's performance.

1 Introduction

One of the primary objectives of monitoring is the early detection of changes in a monitored system or process. Some of these changes can stem from modifications with time of operating conditions (i.e. system input) or changes in the behaviour (system response or output). Usually, changes in the monitored system inputs are intentional hence already known. Generally, we are interested in detecting changes in the system response. The objective of this paper is to account at the same time for changes in the expected behaviour and associated dispersions for any number of monitored inputs in order to detect significant changes while being able to explain in detail the contribution from each of these inputs.

The basis for the proposed approach has been put forward by Léonard and Gauvin, 2013 [1]. They studied the sphere-hardening phenomenon in multidimensional signal projection problems. In fact, this is not a new concept and was first proposed by Shannon, 1949 [2]. While common in communication theory, it seems relatively unknown in the field of equipment and process monitoring. By looking at the cumulative combined random response and measurement noise of a given process over a high enough number of variables in an experiment $\boldsymbol{R}$ repeated many times ($\boldsymbol{R}_i, i = 1, \ldots, M$), a shell will be formed at a given distance $\mu_{\perp S}$ from the expected value $\boldsymbol{S}$ as shown in Figure 1 for the two-dimensional case. This means that looking at a deviation from the shell ($d_i - \mu_{\perp S}$) rather than the deviation from the expected value (d_i) of the noisy process in multidimensional space is more efficient.

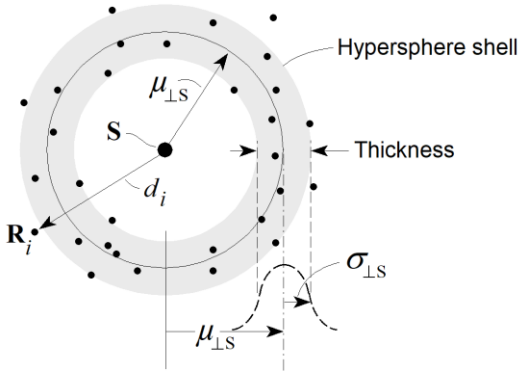


Figure 1: Multidimensional shell from noisy data

However, we cannot do statistics with only one realization of a given event and when monitoring equipment in operation, a difficulty arises since the exact same conditions usually never repeat themselves. The operating conditions of the equipment are always changing hence the need to group similar events together in order to use the sphere hardening concept. Monitored events in similar operating conditions need to be clustered together. Our approach uses a dynamic clustering approach [3] similar to the *k*-mean methodology [4]. Furthermore, since monitoring cannot be restrained to events that are members of a limited number of known clusters, we need to interpolate in-between the clusters and properly account for the uncertainty induced to prevent false alarms that would lead to unwanted downtime and maintenance costs.

The concepts of shell hardening, clustering and interpolation are used to build the monitoring methodology put forward in this paper. To our knowledge, the approach is novel for situations where many channels or indicators are considered simultaneously. Similar to other monitoring approaches, our methodology starts by modelling the equipment response, then estimates the response under the current operating conditions and finally determines the deviation of the current observed response. Dynamic clustering is used to first model the response while also modelling the dispersion. Then, we use kriging to obtain the behaviour across all possible operating conditions. Finally, we assess the deviation in the obtained multidimensional subspace.

Our paper is structured as follows. We start with the concept of a multidimensional shell resulting from noisy data. Next, the full methodology proposed is explained. Then a study case is presented to illustrate the capability of the proposed methodology. Finally, we discuss some of the limitations of the proposed approach.

2 Noisy data and multidimensional shell

At the root of the proposed methodology is the concept that noisy data over a large number of monitored dimensions generate a multidimensional shell with relative thickness that is inversely proportional to the number of dimensions as proposed by Shannon, 1949 [2], see also [5]. If we consider the information in the form of an equipment signature $\boldsymbol{S}$ that we transmit over N dimensions contaminated by noise, the received signal is given by:

$$\boldsymbol{R} = [S_1 + \varepsilon_1,\ S_2 + \varepsilon_2,\ \ldots,\ S_N + \varepsilon_N\ ] = \boldsymbol{S} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{R}$ is the received signal and $\boldsymbol{\varepsilon}$ the random noise vector. However, notice that in the case where the monitored signature is unknown it needs to be estimated using a sample of received signals; relying on the mean as an estimate, one has:

$$\widehat{\boldsymbol{S}} = \frac{1}{M}\sum_{i=1}^{M} \boldsymbol{R}_i \tag{2}$$

Furthermore, since the signature transmitted is constantly changing with the operating conditions of the equipment, the signature $\boldsymbol{S}$ is a manifold rather than a single location as shown in Figure 2. This point is discussed further in the following sections.

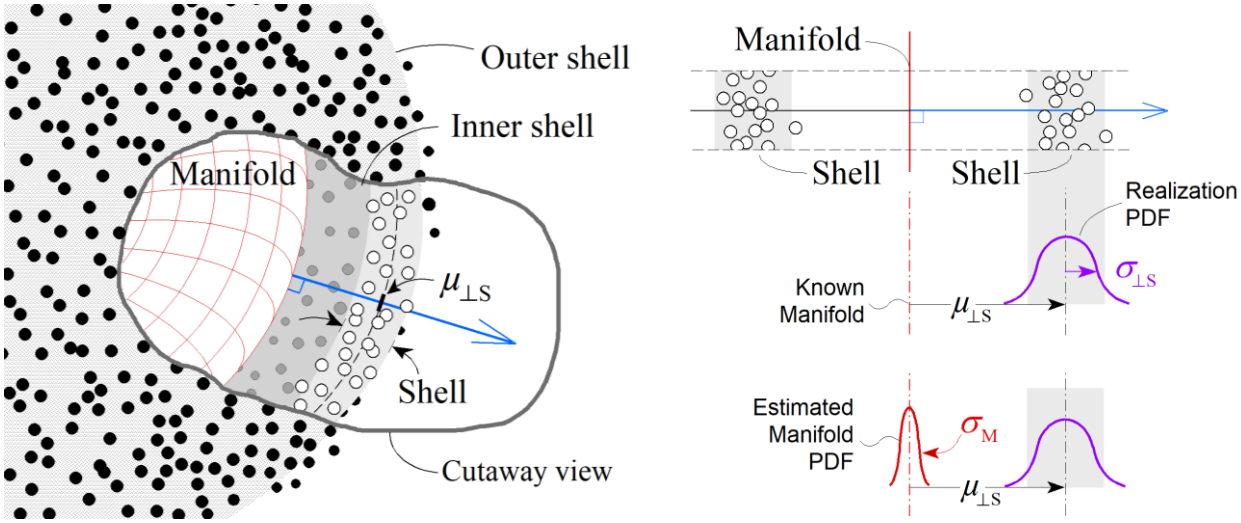


Figure 2: Multidimensional shell and manifold

In this multidimensional space, the distance $\mu_{\perp S}$ between the received signal $\boldsymbol{R}$ and the transmitted signature $\boldsymbol{S}$ at any given location on the manifold can be defined as the average of the realizations at this location as follows:

$$\hat{\mu}_{\perp S} = \frac{1}{M}\sum_{i=1}^{M}\left\|\boldsymbol{R}_i - \widehat{\boldsymbol{S}}\right\| = \frac{1}{M}\sum_{i=1}^{M} d_i \tag{3}$$

where $\widehat{\boldsymbol{S}}$ is the estimate of the transmitted signature of interest $\boldsymbol{S}$, a location on the manifold. In a similar manner, the dispersion can be obtained with:

$$\hat{\sigma}^2_{\perp S} = \frac{1}{M}\sum_{i=1}^{M}(d_i - \hat{\mu}_{\perp S})^2 \tag{4}$$

where $\sigma_{\perp S}$ represents the standard deviation or half-shell thickness. In the present case, where the manifold is also estimated, as shown in Figure 2, the quadratic sum of the manifold dispersion σ^2_M and shell dispersion $\sigma^2_{\perp S}$ can be used to assess the likeliness of a given data point $\boldsymbol{R}_i$. Note that the shell wraps around the manifold when N is greater than the number of operating condition indicators.

3 Methodology

In applications, the use of the multidimensional shell concept is not that simple. As shown in Figure 3, realizations need to first be assembled in clusters. Then, we need to interpolate and extrapolate in the hyperspace between clusters. Finally, a fast estimate of the likeliness of a given new realization needs to be made in order for the information about an alert to be relevant in the context of equipment monitoring.

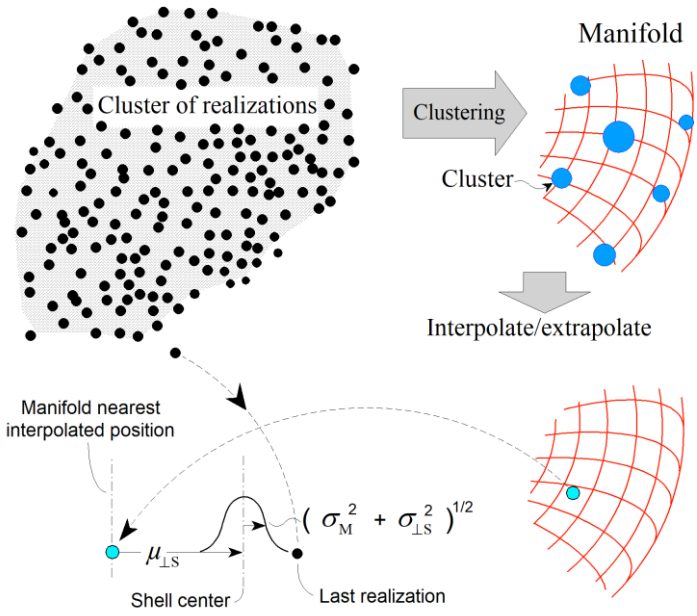


Figure 3: Monitoring methodology

3.1 Data acquisition

The first step is to ensure the data quality before using our algorithm because any error in the data might trigger unwanted alarms. Such preliminary processing is highly dependent on the context of the study and will not be discussed here. For the purpose of this study, let us define the input data as a time series of snapshots $\boldsymbol{X}_i \equiv \boldsymbol{R}_i \cup \boldsymbol{O}_i$ where $\boldsymbol{R}_i$ is a set of response indicators and $\boldsymbol{O}_i$ is a set of operating condition indicators. These $\boldsymbol{X}_i$ cannot be used directly and first need to be formatted and filtered properly to remove unwanted input operating conditions and/or output values. Then, each $\boldsymbol{R}_i$ needs to be normalized in order to ensure that all indicators are represented on similar scales.

3.2 Clustering

Having filtered and normalized the $\boldsymbol{X}_i$ vectors, our goal is to generate clusters of similar $\boldsymbol{O}_i$ to estimate the multidimensional manifold $\boldsymbol{S}$. Initially, for the creation of the clusters, it is important to have a reference dataset of validated history of $\boldsymbol{X}_i$ that cover most operating conditions $\boldsymbol{O}_i$ with corresponding responses $\boldsymbol{R}_i$. Afterwards, with each new $\boldsymbol{X}_i$, it is the dynamic clustering methodology that will determine if a new $\boldsymbol{X}_i$ should be included in the clustering data history. The clustering is dynamic in the sense that the centroid locations are updated every time a new $\boldsymbol{X}_i$ enters the data history. Figure 4 shows the typical process every new $\boldsymbol{X}_i$ is subjected to.

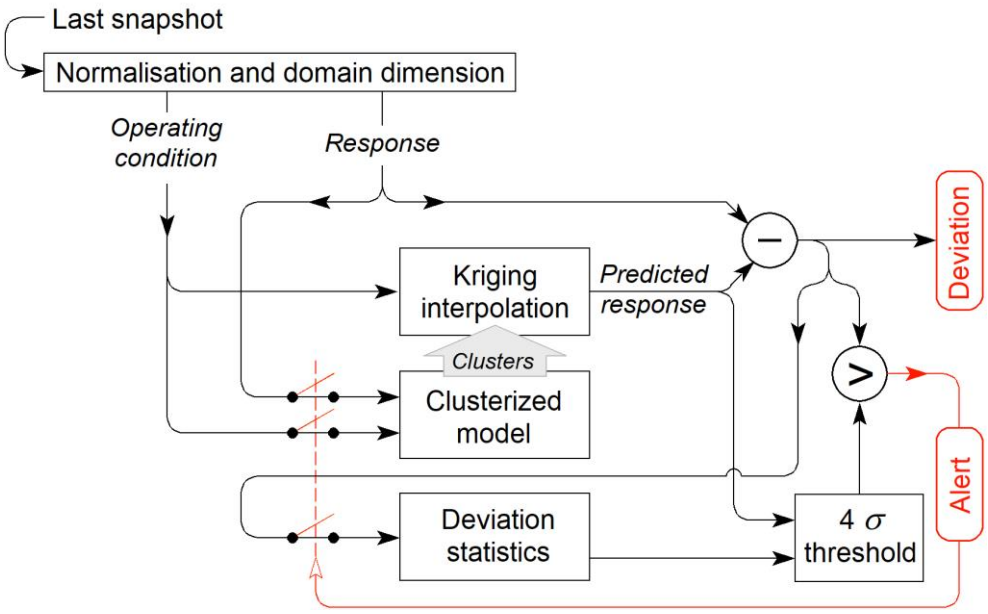


Figure 4: Data snapshots processing flowchart

The clustering process for a new snapshot $\boldsymbol{X}_i$ looks like this:

- If the number of clusters $k < k_{max}$ then the centroid location is $\boldsymbol{Centroid}_k = \boldsymbol{O}_i$ and the population $p_k = 1$ which processes the new snapshot $\boldsymbol{X}_i$
- If the number of cluster $k \geq k_{max}$ then find the smallest distance to an existing cluster c_{min}

$$c_{min} = min_k \|\boldsymbol{Centroid}_k - \boldsymbol{O}_i\| \tag{5}$$

- If $c_{min} > c_{max}$ then merge the two closest clusters together, $\boldsymbol{Centroid}_k = \boldsymbol{O}_i$, $p_k = 1$ and process the new snapshot $\boldsymbol{X}_i$.

$$c_{max} = c\frac{1}{K}\sum_{k=1}^{K}\|\boldsymbol{Centroid}_k - \boldsymbol{O}_i\| \tag{6}$$

- If $c_{min} \leq c_{max}$ then merge snapshot $\boldsymbol{X}_i$ to the nearest cluster, update the population p_k and location $\boldsymbol{Centroid}_k$

Here, in equation 6, c is an arbitrary decision level usually set between 1 and 3. Furthermore, note that any abnormal $\boldsymbol{X}_i$ above the alert thresholds will be rejected and not used for the clustering data history (see section 3.4 for the alert threshold definition). More details about the clustering algorithm used in this study can be found in [5].

3.3 Interpolation

Having clusters of similar data over a large set of different operating conditions $\boldsymbol{O}_k$ with estimated expected response vectors $\boldsymbol{E}(\boldsymbol{O}_k)$ and standard deviation vectors $\boldsymbol{C}(\boldsymbol{O}_k)$ enables us to use a multidimensional interpolator to estimate the response vector for any new operating condition $\boldsymbol{O}_i$. The interpolated $\boldsymbol{E}(\boldsymbol{O}_i)$ with corresponding $\boldsymbol{C}(\boldsymbol{O}_i)$ can then be used to set an alert threshold and assess abnormal behaviour of the monitored system. For simplicity and to limit the computational cost of this interpolation step, the dual kriging formulation was chosen [6]. Kriging is a well-known and extensively used interpolation method. The two traditional formulations which assume a wide sense stationary field, known expected value and variance are the simple kriging formulation and the dual kriging formulation. Implementation of the simple kriging formulation can be either a $O(MN^3) + O(MN)$ or $O(N^3) + O(MN^2)$ process depending on the implementation compared to the dual kriging formulation which is a $O(N^3) + O(MN)$ process for an N

positions over M dimensions problem [6]. Furthermore, because the number of clusters is limited to k_{max} during clustering, we ensure that the numerical cost of the interpolation does not become unmanageable.

3.4 Comparison metric

With the estimated vectors of the expected response $\boldsymbol{E}(\boldsymbol{O}_i) = \widehat{\boldsymbol{S}}_{\boldsymbol{i}}$ (in the notation of section 2) and standard deviation $\boldsymbol{C}(\boldsymbol{O}_i)$ at any new operating condition $\boldsymbol{O}_i$, it is possible to establish an alert threshold above which the behaviour of the monitored process is considered different from past typical responses. Our metric is based on the distance d_i between the snapshot response $\boldsymbol{R}_i$ and the expected response at the operating condition $\boldsymbol{E}(\boldsymbol{O}_i)$:

$$d_i = \|\boldsymbol{R_i} - \boldsymbol{E}(\boldsymbol{O}_i)\| = \sqrt{\sum_{n=1}^{N}\left(r_{i,n} - e(\boldsymbol{O_i})_n\right)^2} \tag{7}$$

More precisely, the alert threshold defines the acceptable relative deviation w_i with regards to the expected value of a given ensemble of similar operating condition as shown in Figure 5. However, to have a faster algorithm, we recommend initially using a single average deviation for all the operating conditions. This can be refined as needed. The average distance $\bar{d}$ and relative deviation w_i are expressed as follow:

$$\hat{\mu}_{\perp S} = \bar{d} = \frac{1}{\sum_{j\in \boldsymbol{A}} 1}\sum_{j\in \boldsymbol{A}} d_j \text{ with } j \in \boldsymbol{A} \text{ if } \boldsymbol{O_j} \approx \boldsymbol{O_k} \tag{8}$$

$$w_i = d_i - \bar{d} \tag{9}$$

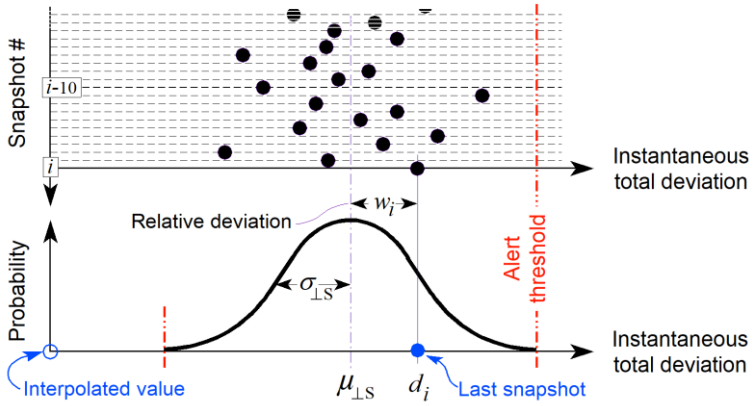


Figure 5: Illustration of the comparison metric

The alert threshold is a function of the global standard deviation of the ensemble of snapshots d_i which accounts for the interpolation standard deviation $\hat{\sigma}_{\mathrm{M}} = \|\boldsymbol{C}(\boldsymbol{O}_i)\|$ and standard deviation of the relative deviation w_i as follows:

$$\sigma' = \sqrt{\hat{\sigma}^2_{\perp S} + \hat{\sigma}^2_{\mathrm{M}}} \tag{10}$$

$$\hat{\sigma}^2_{\perp S} = \frac{1}{\sum_{j\in \boldsymbol{A}} 1}\sum_{j\in \boldsymbol{A}} {w_j}^2 \text{ with } j \in \boldsymbol{A} \text{ if } \boldsymbol{O_j} \approx \boldsymbol{O_k} \tag{11}$$

For the case study given next, we have used an alert threshold of $4\sigma'$.

4 Case study

In this paper, we focus on a case study which is the loss of a hydroelectric turbine propeller runner cone. Figure 6 shows a view of the runner before and after the loss of the cone. The cone structure minimizes hydraulic losses and improves efficiency. Without the cone, we should expect reduced efficiency (around 0.6%) and increased vibration due to the vortex rope which is normally dampened by the cone's presence.

Here, the question is not if we can detect the loss of the cone but rather how early we can alert the operator that something is happening to the runner. The earlier we can detect a problem related to the cone, the more time is available for maintenance outage planning which reduces the unexpected downtime. In this case, we have used approximately two years of snapshots history prior to the event and limited the study to the following indicators:

- Mean spiral case pressure ($\in \boldsymbol{R}$)
- Water temperature
- Peak to peak thrust bearing axial acceleration ($\in \boldsymbol{R}$)
- RMS thrust bearing axial acceleration ($\in \boldsymbol{R}$)
- RMS generator guide bearing radial displacement X ($\in \boldsymbol{R}$)
- RMS generator guide bearing radial displacement Y ($\in \boldsymbol{R}$)
- Mean turbine guide bearing radial displacement X ($\in \boldsymbol{R}$)
- RMS turbine guide bearing radial displacement X ($\in \boldsymbol{R}$)
- Peak to peak turbine guide bearing radial displacement Y b($\in \boldsymbol{R}$)
- Mean turbine guide bearing radial displacement Y ($\in \boldsymbol{R}$)
- RMS turbine guide bearing radial displacement Y ($\in \boldsymbol{R}$)
- Excitation tension
- Wicket gates opening ($\in \boldsymbol{O}$)
- Mean power output ($\in \boldsymbol{R}$)

Figure 6: View of the propeller runner cone, before (left) and after (right) the loss

5 Results

With our methodology, we observe seven different phases in the behaviour of the hydroelectric turbine and two types of transient events (see Figure 7). In phase 1, the snapshots serve as reference data for the algorithm to dynamically define the clusters' centroid and dispersion. We observe that the uncertainty bands gradually stabilize. In phase 2, the method is ready to be used to alert the user of unexpected behavior. Notice that the sudden increase in dispersion after phase 1 is artificial and helps highlight the transition between the learning and monitoring regimes. In phase 3, we systematically observe deviations above the alert level. The deviations increase gradually at each subsequent phase until phase 7 is reached and the cone is lost at the end of the snapshots' time history. One can notice some holes in the time history because some snapshots were unsuitable for the methodology and automatically removed during the data acquisition step. Furthermore, two types of transient events are clearly visible in Figure 7. The first, event 8, is the largest of a family of such events that are due to a cooldown period where the monitored unit was stopped. When the unit is restarted, the generator temperature needs to first stabilize then the surrounding structure temperature also needs to stabilize. This generates a transient state that is not a real alert in the sense that the unit is working as expected; this type of event could easily be filtered out if needed. The second, event 9, is simply due to the initialization of the methodology and one can see that the alert bands rapidly stabilize after a sufficient number of data points have been processed.

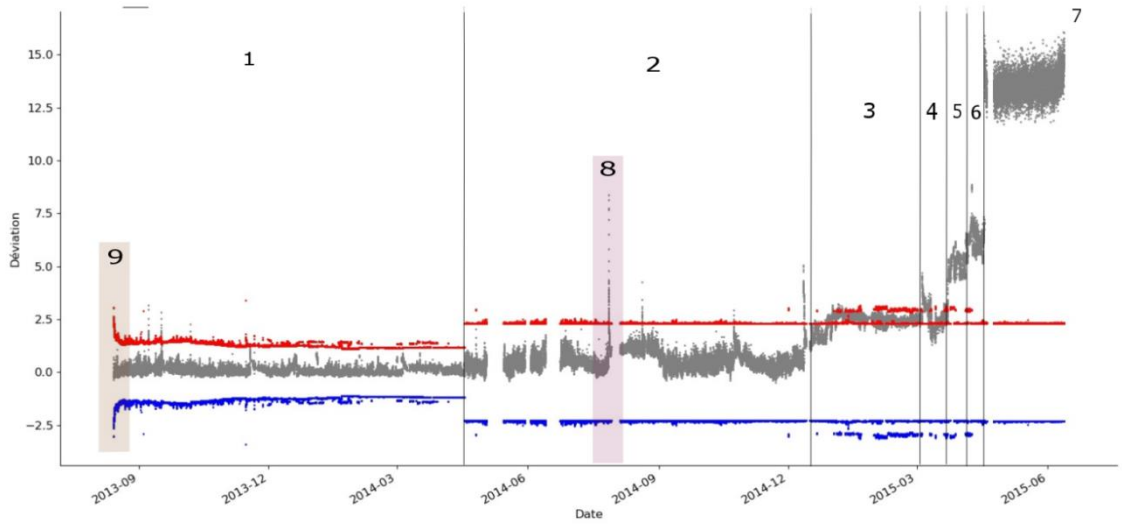


Figure 7: Multidimensional deviation

An advantage of the proposed methodology is that we have access to the contribution of each snapshot response $\boldsymbol{R}_i$ for a given relative deviation w_i. This is of high importance to do a diagnostic of the alert and justify appropriate maintenance outage. In Figure 8, we present an excerpt of the evolution of the individual response contributions for timestamps in each phase from 2 to 7. At first, in phase 3, we observe a highly localised contribution with a slow but gradual increase in contribution from the other response indicators as we move towards phase 5. In subsequent phases, a sudden increase across many of the indicators becomes manifest.

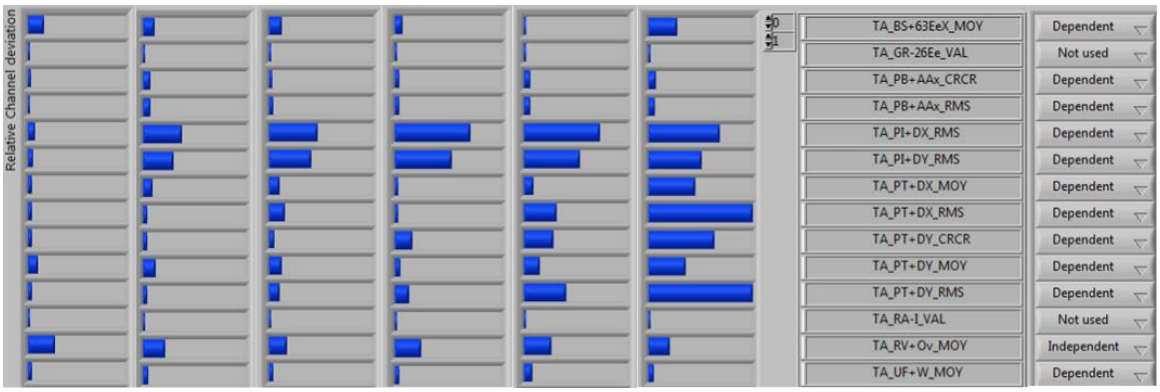


Figure 8: Screenshots for a given timestamp of the relative multidimensional deviation for each channel. From left to right: phase 2, 3, 4, 5, 6 and 7

In comparison, if we look at the time series of certain selected response indicators, by for example intuitively selecting the ones related to the guide bearing which are the closest to the propeller cone, we get the results shown in Figure 9. The problem becomes noticeable only at the end of March 2015 in phase 5, even if the selected indicators are the closest to the propeller cone. By using a larger ensemble of indicators, the approach proposed in this paper is able to alert the operator of an abnormal behaviour more than three months beforehand in phases 3 and 4.

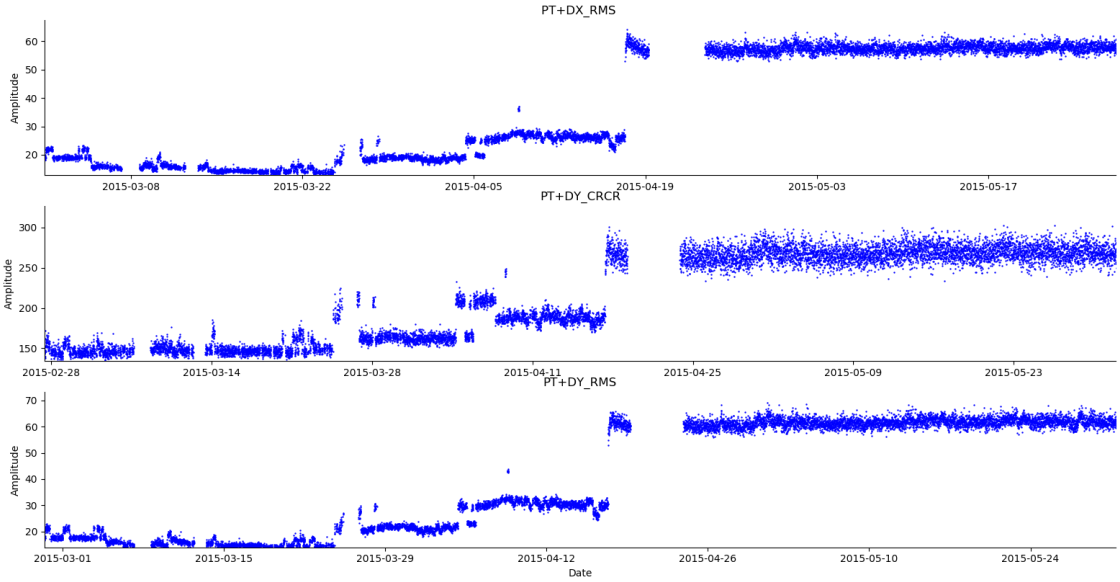


Figure 9: Selected individual responses prior to the propeller runner cone loss

6 Discussions

The case study shows that the sphere hardening principle is applicable to equipment monitoring and performs well in practice. For this application, the method alerts the user months before the actual failure was detected. However, no comparison was made with other mathematical approaches to detect abnormal operational behaviours. Moreover, we simplified the problem by limiting the number of data input to reduce the validation burden. In the proposed approach no difference is made between a change in the behaviour of the sensor and a change in the monitored system. The high sensitivity of the approach to deviations from previous behaviour relies on having data of good quality to avoid false alarms.

Furthermore, no effort was made to optimize the methodology; our initial goal was speed and ease of implementation. In fact, we might be able to optimize the clustering approach to improve kriging performance. Even then, numerical performance might not be the right criteria. The numerical cost of the interpolation might not be a limitation if computational possibilities such as parallelization are considered, given that the necessary infrastructure is becoming more easily available. The same might be true for the alert threshold that could be assessed using conditional numerical simulations.

7 Conclusions

We demonstrated that the data gathered over a group of indicators can be reduced to a single global metric that can be used to monitor equipment behaviour and alert an operator of abnormal equipment behaviour months before an actual failure. The approach is statistically based and is highly sensitive to any deviation from normal past behaviour. An important advantage of the approach is that we can easily track the contribution of each individual indicator and thus explain an alert to establish a diagnostic. We are currently looking at implementing the proposed approach on our hydroelectric turbine fleet in order to benchmark its performance.

References

[1]F. Léonard and M. Gauvin, Trending and pattern recognition using the sphere hardening phenomenon of signal multidimensional projection, Surveillance 7 conference, IUT, Chartres 2013.

[2]C. Shannon, *Communication in the Presence of Noise*, Proceeding of the IRE (1949), Vol. 37, No.1, pp. 10–21.

[3]F. Léonard, *Dynamic clustering of transient signals*, patent US 2014/0100821, priority 2011.

[4]J. B. MacQueen, Some Methods for classification and Analysis of Multivariate Observations, in Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability , no. 1, p. 281–297, 1967.

[5]F. Léonard, *Noisy data clusters are hollow*, Joint Statistical Meetings 2015, Aug 2015, Seattle, United States. JSM 2015 Proceedings, 2015.

[6]F. G. Horowitz, P. Hornby, D. Bone and M. Craig, *Fast Multidimensional Interpolations* Chapter 9 (pp. 53-56) of the 26th Proceedings of the Application of Computers and Operations Research in the Mineral Industry (APCOM 26), R.V. Ramani (ed.), Soc. Mining, Metall., and Explor. (SME), Littleton, Colorado, U.S.A., 538 p., 1996.

Confounding factors analysis and compensation for high-speed bearing diagnostics

Alessandro Paolo Daga[1], Luigi Garibaldi[1], Alessandro Fasana[1], Stefano Marchesiello[1]

[1]Politecnico di Torino, Italy
Department of Mechanical and Aerospace Engineering (DIMEAS),
Dynamic & Identification Research Group (DIRG),

Abstract

In recent years, machine diagnostics through vibration monitoring is gaining a rising interest. Indeed, in the literature many advanced techniques are available to disclose the fault establishment as well as damage type, location and severity. Unfortunately, in general, these high-level algorithms are not robust to operational and environmental variables, restricting the field of applicability of machine diagnostics. Most of industrial machines in fact, work with variable loads, at variable speeds and in uncontrolled environments, so that the finally measured signals are often non-stationary. The very common time-series features based on statistical moments (such as root mean square, skewness, kurtosis, peak value and crest factor) undergo variations related to changes in the machine operational parameters (e.g. speed, load, …) or in the environmental parameters (e.g. temperature, humidity, …), which can be seen as non-measured, and then latent, confounding factors with respect to the health information of interest.

In order to face such issue, statistical techniques like (in a first exploratory stage) the Principal Component Analysis, or the Factor Analysis, are available. The pursuit of features insensitive to these factors, can be also tackled exploiting the cointegration property of non-stationary signals.

In this paper, the most common methods for reducing the influence of latent factors are considered, and applied to investigate the data collected over the rig available at the DIRG laboratory, specifically conceived to test high speed aeronautical bearings monitoring vibrations by means of 2 tri-axial accelerometers while controlling the rotational speed (0 – 30000 RPM), the radial load (0 to 1800 N) and recording the lubricant oil temperature. The compensation scheme is based on two procedures which are established in univariate analyses, but not so well documented in multivariate cases: the removal of deterministic trends by subtraction of a regression, and the removal of stochastic trends in difference stationary series by subtraction of the one-step ahead prediction from an autoregressive model. The extension of these methods to the multivariate case is here analysed to find an effective way of enhancing damage patterns when the masking effect due to the non-stationarities induced by latent factors is strong.

Keywords: trend stationary, difference stationary, regression, autoregressive prediction, residuals, orthogonal regression, PCA whitening, Mahalanobis distance, cointegration, vector autoregression, novelty detection, damage detection, vibration monitoring, bearings.

1 Introduction

Vibration Monitoring (VM) is a particular kind of condition monitoring which exploits vibration as a condition indicator. Indeed, an online non-destructive testing (NDT) based on vibration can be set up to monitor the health condition of the machine while in operation. This turns out to be fundamental in Condition-Based Maintenance (CBM) regimes, in which the maintenance is not programmed but preventive, and must then rely on diagnostics and prognostics. The advantage of VM against other techniques such as Oil debris analysis, Performance analysis, thermography, Acoustic analysis or Acoustic Emissions AE, etc. is related to the high reactivity to sudden changes in a machine, and to the flexibility of the accelerometers (i.e. the vibration sensors), which are not only cost effective and reliable, but also small and light so that almost any machine can be easily instrumented.

A vibration-CBM is basically a Data-to-Decision (D2D) process [1] but the present work will focus mainly on the signal and the pattern processing parts, namely that of selecting and extracting damage-sensitive features and that of building and validating a statistical model based on the data whose scope is the detection of a damaged condition (***data-based modelling***). Damage detection is commonly considered the first fundamental step of diagnostics and consists in producing an indication of the presence of a damage, possibly at a given confidence, so that an alarm can be eventually triggered in case of danger. This can be performed by looking for the symptoms which indicates the presence of a faulty condition (i.e., recognizing patterns in the data). Such damage distinguishing characteristics can be effectively highlighted only by the extraction of proper features, namely quantities which show:

- Damage Consistency (i.e. they vary with damage)
- Damage Sensitivity and Noise-Rejection ability (i.e. they are sensitive also to small, incipient damage),
- Low sensitivity to unmonitored confounding factors.

A perfect feature is then able to reject any influence other than damage, producing stationary sets of indices for which the departure from the normal condition can only be ascribed to a malfunctioning. In this case the detection of novelty corresponds to the detection of damage and is then a relatively easy task.
In reality, researchers will always deal with features affected by operational (e.g. speed, load, …) and environmental (e.g. temperature, humidity, …) variations, which can be seen as latent (i.e., non-measured), confounding factors that can compromise the correct damage detection.
The scope of this paper is to highlight some technique for compensating such confounders with a particular focus on damaged bearings data. In order to cope with the need of a fast and automated real-time damage detection, an analysis is here proposed, based on the common time-series features (i.e., root mean square, skewness, kurtosis, peak value and crest factor (peak/RMS)) which are known to be sensitive to bearing damage but also to the operational conditions of the machine under analysis. The Novelty Detection then, must be preceded by some algorithm compensating for the confounders. In particular, the **regression** and the **cointegration** will be discussed in section 2. The experimental data used in this work refers to the high-speed aeronautical bearings test rig available at the Department of Mechanical and Aerospace Engineering of Politecnico di Torino, shortly introduced in next section.

1.1 The test rig and the dataset

The dataset considered in this analysis comes from a test rig built by the Dynamic & Identification Research Group (DIRG), part of the Department of Mechanical and Aerospace Engineering of Politecnico di Torino, to test high-speed aeronautical bearings. The rig is fully described in [2], but the main information is summarized hereinafter. The rig is made by a single direct-drive rotating shaft supported by two identical high-speed aeronautical roller bearings (B1 and B3 in Figure 1). B3 is known to be healthy while B1 is damaged by purpose with indentations of different size in different parts of the bearing (Rolling Element and Inner Ring) as described in Table 2. The third central bearing B2 is mounted on a sledge meant to load the shaft with an increasing force of 0, 1000, 1400 and 1800 N, while the speed is reducing from 470 to 0 Hz (run-down acquisitions). Table 1 summarizes the operational conditions. Two tri-axial accelerometers located respectively on the B1 bearing support (accelerometer A1, as reported in Figure 1) and on the loading sledge (accelerometer A2). The acquisitions last for a duration of about $T = 50\,s$ at a sampling frequency fs = 102400 Hz. In order to perform a significant analysis, the five selected features *root mean square*, *skewness*, *kurtosis*, *peak value* and *crest factor* are extracted on one hundred independent chunks (about $0{,}5\,s$ each) for each of the 6 channels of the 4 original acquisitions in all the 7 health conditions (from 0A, healthy, to 6A).
Finally, 100 observations in a 30-dimensional space (6 channels, 5 features) per each health conditions are obtained. A part of the dataset is visually summarized in Figure 2.

Table 1. The operational conditions: the different loads while the speed is decreasing from 470 to 0 Hz (run-down acquisitions).

Label	1	2	3	4
F [kN]	0	1	1,4	1,8

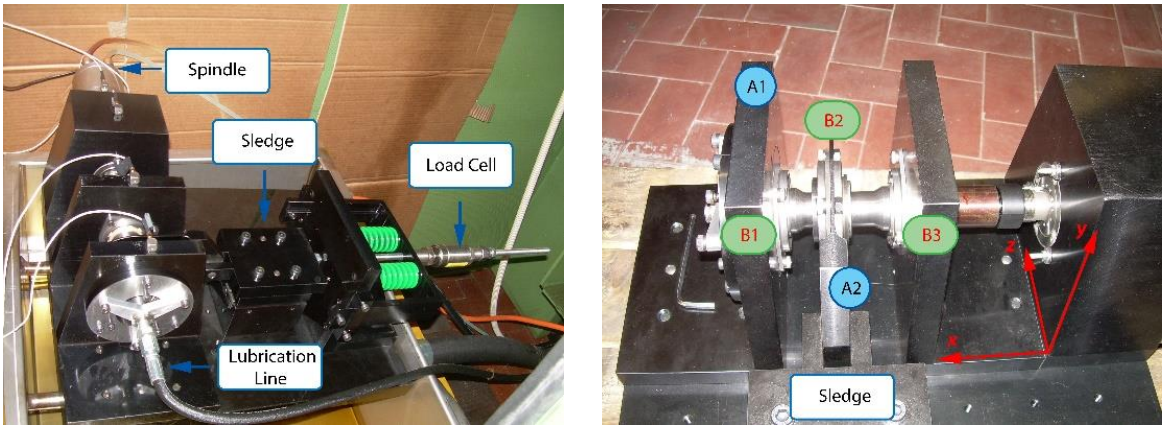


Figure 1. The experimental setup, the triaxial accelerometers location (A1 and A2) and orientation.

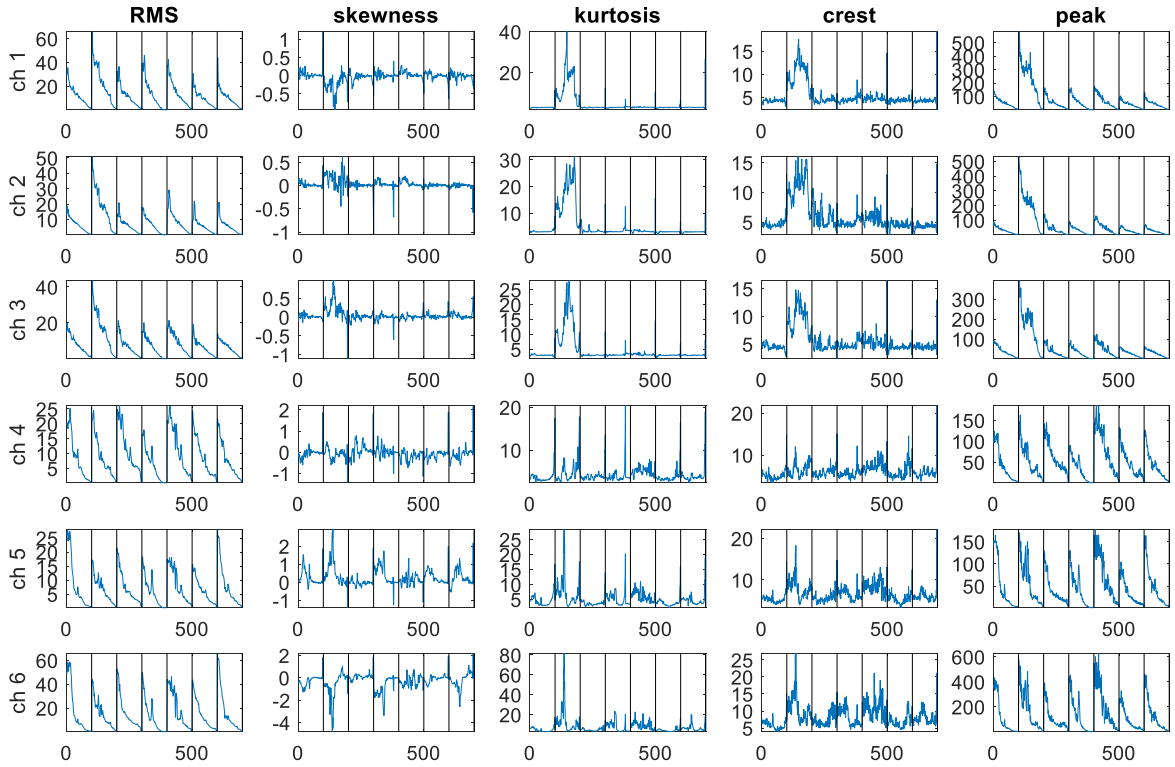


Figure 2. The considered dataset after features extraction for load condition 1 (0 N) while the speed is decreasing until a stop starting from 470 Hz. The black dotted lines divide the different damage conditions (0A to 6A). For each, 100 observations are plotted sequentially.

Table 2. Bearing B1 codification according to damage type (Inner Ring or Rolling Element) and size. The damage is obtained through a Rockwell tool producing a conical indentation of maximum diameter reported as characteristic size.

Code	0A	1A	2A	3A	4A	5A	6A
Damage type	none	Inner Ring	Inner Ring	Inner Ring	Rolling Element	Rolling Element	Rolling Element
Damage size [µm]	-	450	250	150	450	250	150

Figure 3. Stationary stochastic process and the biconditional relationship of novelty and damage.

2 Methodology

This work is devoted to the application of Novelty Detection on time-series features extracted from the raw acceleration data of a test rig. The idea of "novelty" is commonly related to that of "outlier", a discordant measure inconsistent with the others and then believed to be generated by an alternate mechanism.
The judgment on discordancy will depend on a measure of distance from a reference distribution (e.g., healthy), usually called Novelty Index (NI), on which a threshold can be defined [3].
This very simple idea can be exploited for Damage Detection when the healthy vibration signal can be modelled as a stationary stochastic process, meaning that the joint probability distribution function is invariant under time translation, so that damage is left as the only possible cause of discordancy (Figure 3).
Unfortunately, hidden latent (non-measured) factors will always affect the measurements. When their effect is important, non-stationarities will arise, leading to misinterpretations of the novelty (and then damage), so that they are often referred to as confounders (Figure 4).

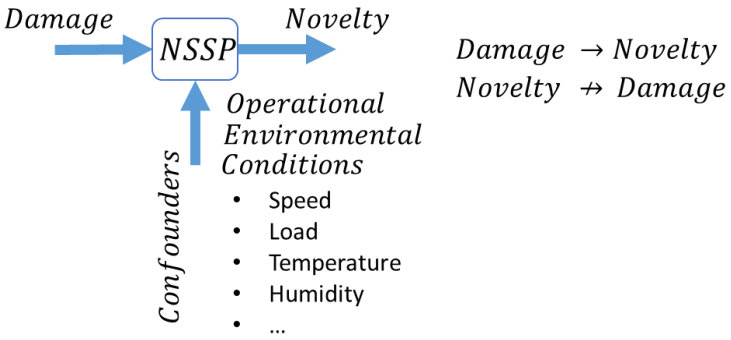


Figure 4. Non-stationary stochastic process and the effect of confounders

In these cases, a wise feature selection able to reduce the effect of the confounders may be important but is usually not enough for solving this issue. Algorithms for compensating such effects become then essential.
In this work, the temperature is controlled, and the load is kept constant so that the only confounder is the variable speed, which strongly affect time features as the peak level and the RMS of the acceleration signal (see Figure 2). The measurement involves an uncontrolled braking of the machine from full speed (470 Hz) to a stop. The features from the first channel of the first accelerometer with a null load are reported in Figure 5.

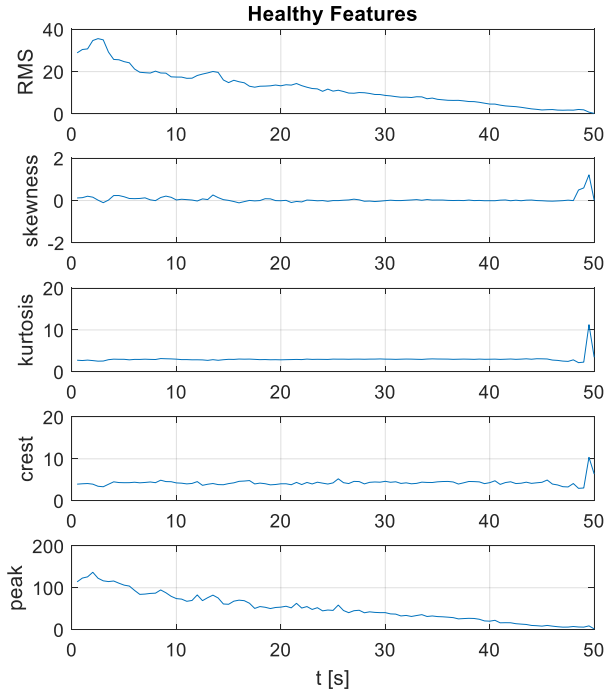


Figure 5. The five selected features from the first channel of the first accelerometer – 0N.

Stationarity is noticeably violated as a trend is clearly visible in Figure 5. In the literature [4] two simple models for such violations can be found.
The first involves a deterministic trend, so that the resulting signal takes the name of **trend stationary**. A polynomial fitting can be used in this case to find and subtract the trend, leading to a stationary residual which is said to be "white" as the resulting frequency spectrum turns out to be flat (i.e. the residual is a white noise) or "decorrelated" as its autocorrelation is null for any lag different from 0.

$$y_t = \beta t + \varepsilon_t \qquad \varepsilon_t \sim N(0, \sigma^2)$$

A second model on the contrary involves a stochastic trend. In the simplest case, this means that the increment in the signal from time to time (innovation) is defined as a stochastic process ε such that

$$y_t - y_{t-1} = \varepsilon_t$$

In this case the signal y is the result of the integration of the considered stochastic process ε and is then called integrated of order 1 or $I(1)$. This process which corresponds to a random walk is **difference stationary** as its first difference is stationary. Again, it is possible to get a stationary signal which can be considered white.

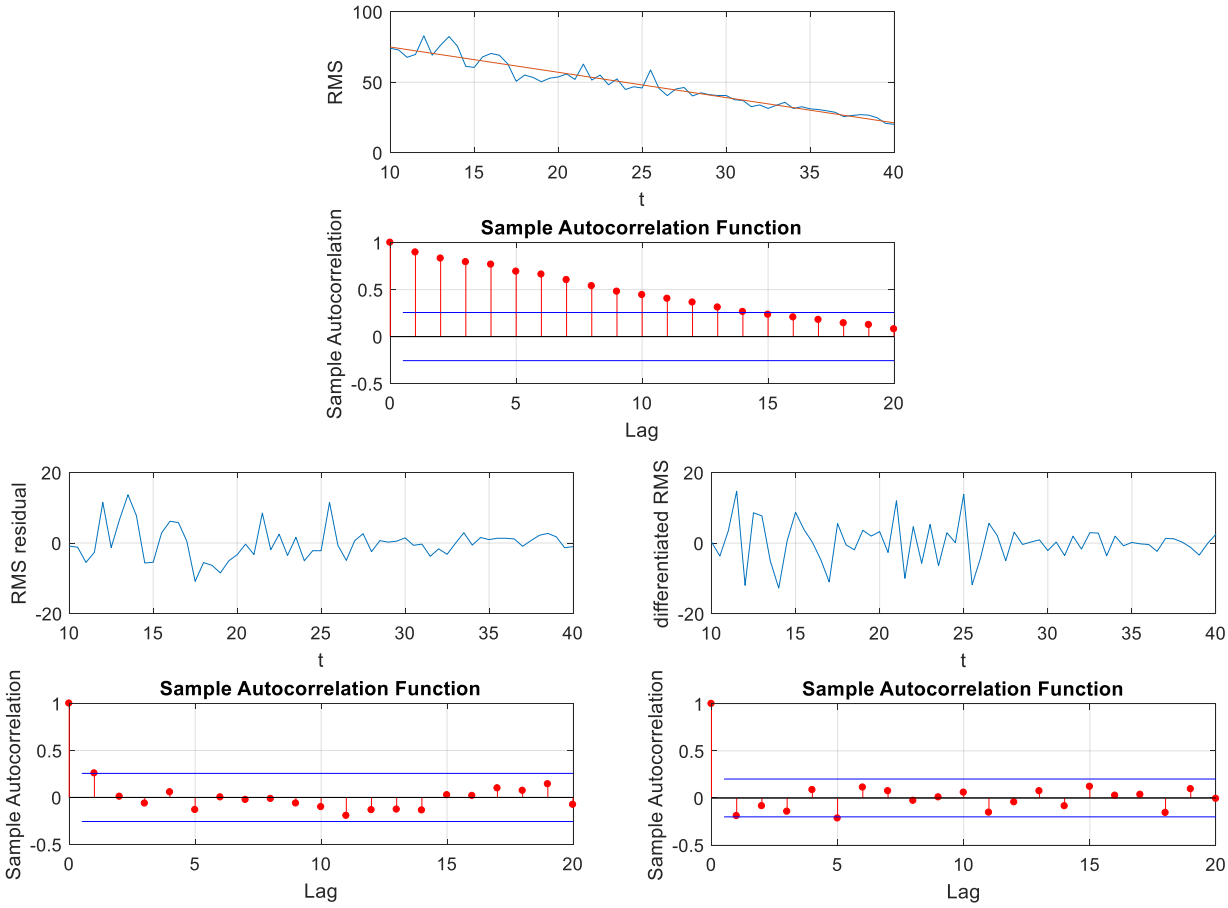


Figure 6. The raw RMS from 10 to 40s and its autocorrelation function (ACF). Below, on the left the residual from removing the linear regression and its ACF, on the right, the differencing (equivalent to the residual after an AR(1) fit) and its ACF.

To generalize, the random walk can be considered as a particular case of an autoregressive AR(1) model with a unitary coefficient. That explains why it is very common in the literature to whiten the data by fitting an AR(1) to the series and focus on the residual, as done in [5] to highlight the damaged bearing signature.
These concepts can be extended to multivariate spaces. When the features are affected by the same confounder in fact, they turn out to be strongly correlated (in simple terms, they vary in sympathy).

Under the first assumption (trend stationarity) then, a multivariate regression can be used. In this case, considering that both the variables are affected by measurement errors and that it is not easy to find a dependent and an independent quantity, the orthogonal regression [6] based on PCA is proposed.

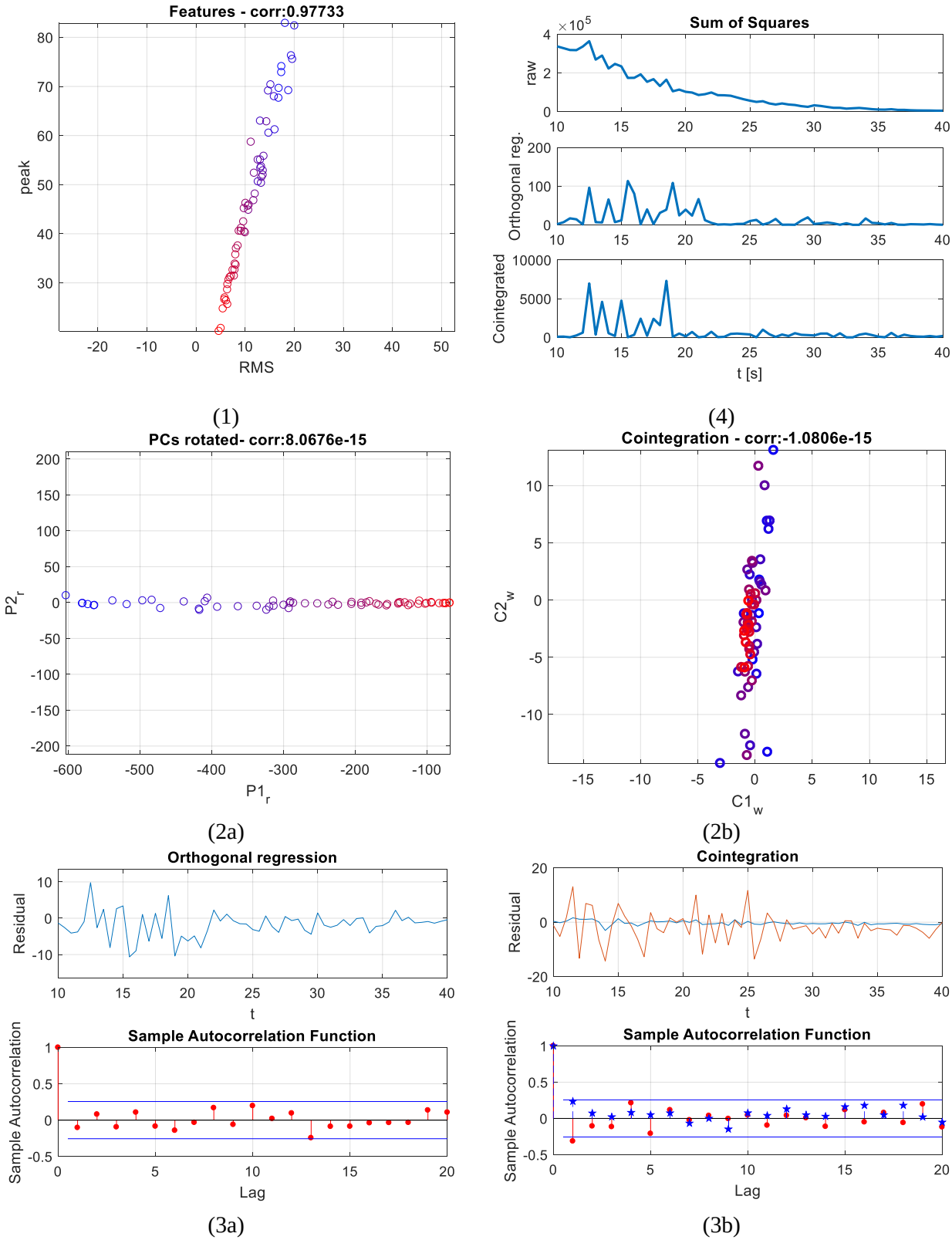


Figure 7. (1) The bivariate scatterplot with time evolution (from blue – 10s to red – 40s)
(2a) PCA rotation (3a) Rotated PC2 corresponding to the residual and its ACF
(2b) coitegration scatterplot (3b) stationary residual from cointegration and its ACF
(4) sum of squares of 3a and 3b components compared to the raw Euclidean squared distance

Orthogonal regression is fundamentally a reconstruction of the dataset in a subspace obtained by neglecting the first principal component. This methodology can be merged to the PCA whitening, to directly obtain a white, unitary covariance residual.
PCA orthogonal regression and whitening are mathematically tackled in section 2.1.
On the contrary, difference stationary multivariate series can be whitened by fitting a vector-auto-regressive VAR(1) model and computing the multivariate residual. In this case the dimensionality is not affected. Nevertheless, the considered variables must be I(1), so that a statistical hypothesis test such as the Augmented Dickey-Fuller is needed [4]. This procedure commonly takes the name of cointegration [7,8] and is described in section 2.2.

The results of a simple bivariate analysis on the RMS and the Peak value of the first channel are reported in Figure 7 to show the ability of the two methods on a real acquisition. In case of real measurements in fact, as confirmed by this simple analysis, it may be difficult to confidently identify the underlying model as both may work in a quite proper way.
Nevertheless, keeping focused on the final scope of detecting novelty (and damage), a relevant consideration can be made about novelty indices (NI). Novelty detection in fact, is commonly based on the Mahalanobis distance [3,9] which is known to be equivalent to a Euclidean distance on a features space rotated to match the principal components and normalized to obtain unitary variance PCs [2].
Hence, the squared Mahalanobis distance equals the sum of the squared whitened principal components. Obviously, it involves also the first PC which pictures the confounding factor, so that it is not stationary, as shown in Figure 8. A good idea then is to use as novelty index the sum of the squared principal components rejecting the first or first few.
The same idea of summarizing the residuals with a single novelty index computed as the sum of squares of the cointegrated series can be extended to cointegration (Figure 7.4 and 8).

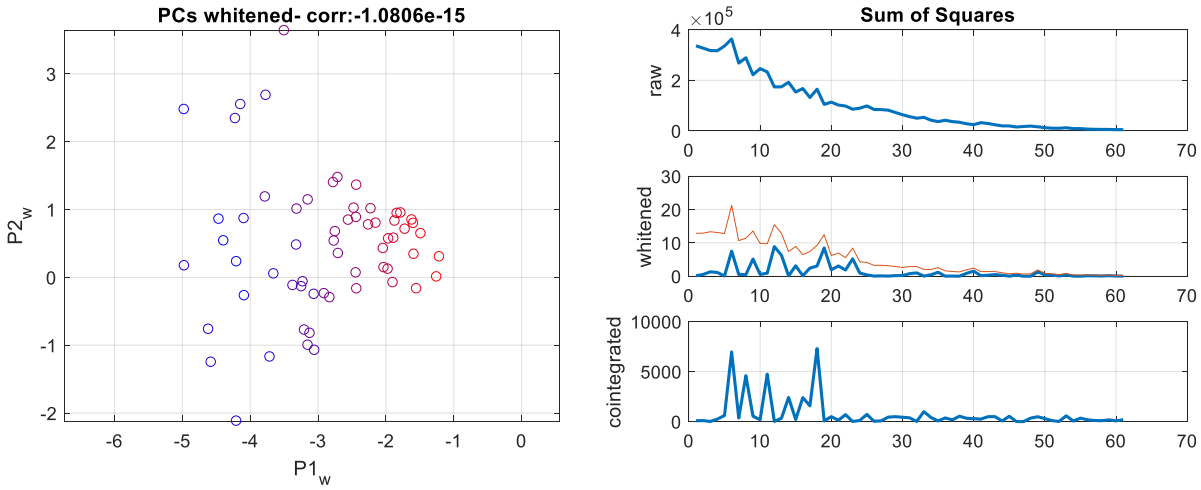


Figure 8. Standardized principal components (whitening) and sum of squares of the two PCs (in red) and of the second alone (blue, mid graph) compared to the squared Eulerian distance (raw) and to the sum of squares of the cointegrated residuals.

2.1 PCA orthogonal regression and whitening

Orthogonal regression is an extension of traditional regression for datasets in which the independent variable is not assumed to be perfectly known but admits errors. Indeed, in statistical literature this is known as "errors-in-variables" model, or also, "total least squares". A simple but effective way to perform it is based on Principal component analysis PCA [10,11].
Mathematically, given a d-dimensional centred dataset of n observations $X \in R^{d \times n}$, an unbiased estimator for the covariance can be used to obtain:

$$S = \frac{1}{n-1} XX'$$

PCA corresponds to the solution of the eigenproblem:

$$S\,V = V\Lambda$$

where V is the orthogonal matrix whose columns are the d eigenvectors v_j while Λ is the diagonal matrix of the d eigenvalues λ_j (usually sorted in descending magnitude) of the matrix S.
The matrix V can be used then to decorrelate the dataset X, that is, to rotate the reference frame to the one identified by the eigenvectors (i.e. the principal components, PCs) of matrix S:

$$Z = V'X$$

If the eigenvectors in V are normalized to have unit length ($v_j'v_j = 1$), the transform is a pure rotation, and it can be proved that $\sigma_j^2 = var(z_j) = \lambda_j$. Namely, the diagonal matrix Λ is the covariance of Z.
Focusing on linear orthogonal regression, the direction given by the first eigenvalue corresponds to the regression line, so that the residuals X_L can be simply found as a projection on the subspace generated by the $L = d - 1$ components other than the first:

$$z_j = v_j'X = v_{j1}x_1 + v_{j2}x_2 + \cdots + v_{jd}x_d = \sum_{k=1}^{d} v_{jk}x_k$$

$$Z_L = V_L'X$$

$$X_L = V_L Z_L = V_L V_L'X$$

The orthogonal regression is visualized in Figure 9.

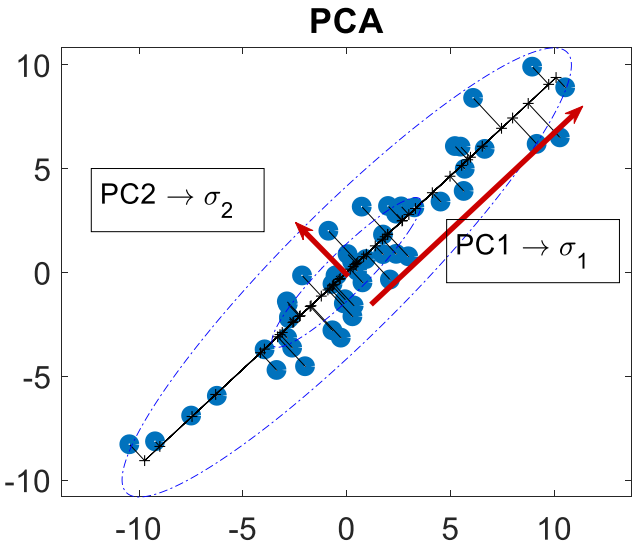


Figure 9. Visualization of the PCA orthogonal regression – the residuals corresponding to PC2 are highlighted.

Different normalizations for the eigenvectors are obviously possible. Another quite common one consists in normalizing for $v_j'v_j = \lambda_j$. In this case $var(z_j) = 1$ so that the covariance matrix of Z is the identity matrix I. In this case, on top of the rotation, a rescaling on the principal component occurs. V is then commonly called a "whitening matrix" W or also sphering matrix as it transforms the data covariance ellipsoid to a spheroid [6].

$$Z_W = W'X = \Lambda^{-1/2}V'X = \Lambda^{-1/2}Z$$

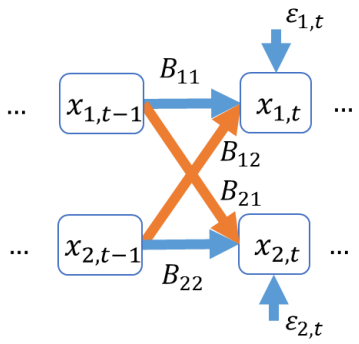
Finally, the squared Mahalanobis distance can be then written as

$$SMD = X'S^{-1}X = Z'V'S^{-1}VZ = Z'\Lambda^{-1}Z = \sum_j \frac{z_j^2}{\lambda_j} = Z_W'Z_W = \sum_j z_{Wj}^2$$

This proves that the squared Mahalanobis distance corresponds to the sum of squares of the whitened features. Hence, removing the first whitened component(s) from the sum corresponds to merging orthogonal regression and PCA-whitening: the so found distance is therefore robust to confounders. This makes it a good candidate to substitute the Mahalanobis distance as NI in the presence of non-stationary operational or environmental conditions.

2.2 Cointegration

Cointegration is a property of multiple nonstationary time series which can be linearly combined through a cointegrating matrix B to produce stationary series as the residual from a one-step-ahead prediction from a Vector Auto Regressive (VAR) model:



$$\hat{X}(t) = \mathrm{B}\, X(t-1)$$

$$Z(t) = X(t) - \hat{X}(t)$$

Figure 10. The VAR(1) model and residual.

The resulting $Z(t)$ can be proved to be stationary if the considered series in X share the same order of integration equal to 1. This can be verified by an Augmented Dickey Fuller univariate test [4,7] run on all the series.
In simple terms, ADF is based on the estimation of an AR(1) model

$$y_t = \rho\, y_{t-1} + \varepsilon_t$$

Obviously,
Obviously, different cases can be found:

- If $\rho = 0$ the signal is a pure white noise,
- If $|\rho| < 1$ the signal is stationary, or I(0),
- If $\rho = 1$ the signal is a pure random walk, or I(1),

Therefore, ADF tests the null hypothesis $H_0: \hat{\rho} = 1$ against the alternative $H_a: \hat{\rho} \neq 1$.
Hence, a confidence interval can be built around the estimated

$$\hat{\rho} = \frac{\sum_t y_t y_{t-1}}{\sum_t y_{t-1}^2} = \rho + \frac{\sum_t \varepsilon_t y_{t-1}}{\sum_t y_{t-1}^2}$$

If $\hat{\rho} - 1$ falls within the interval centred around 0, then the signal is proved to be I(1) at the selected confidence.

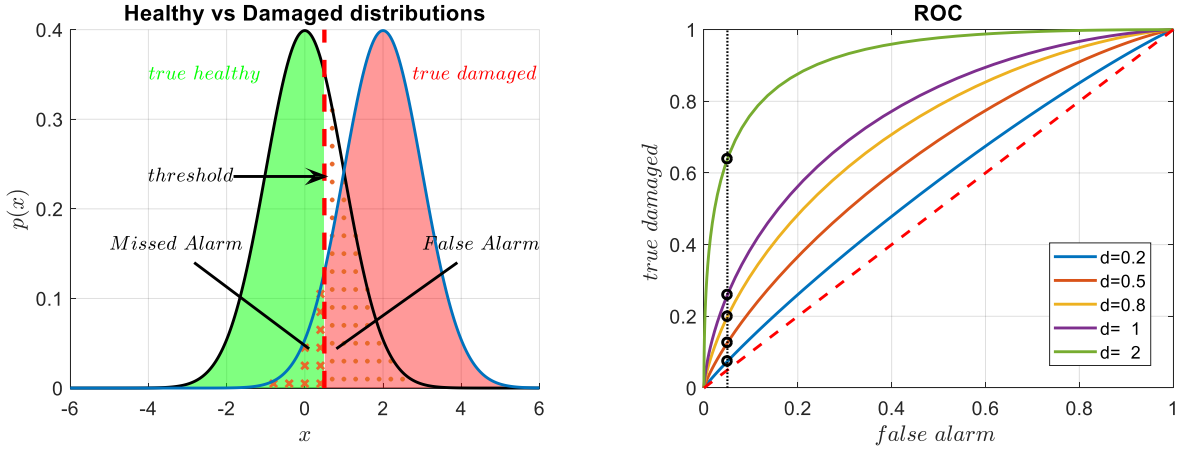
Once the I(1) series are highlighted, the VAR model can be fitted, and the residual vector $Z(t)$ computed.
By analogy to Mahalanobis distance and PCA orthogonal regression & whitening, the sum of squares of the Z residuals can be used as a NI, proved that these residuals are stationary and uncorrelated.
In practical cases, PCA orthogonal regression & whitening performs decorrelation, but not ensures necessarily the stationarity of the series. On the contrary cointegration enforces stationarity but does not guarantee uncorrelatedness. That is why it may be a good idea to merge the to procedures into a single approach.
By exploiting cointegration on pre-whitened series in fact, both uncorrelatedness and stationarity can be obtained.

2.3 Novelty detection and classification: performance assessment

Diagnostics, and in particular damage detection, can be considered a classification problem. The simplest binary classification can be tackled via novelty detection: when novelty is found (i.e. when the NI exceeds a threshold) the measure is assigned to the "damaged" class and an alarm is triggered, otherwise the acquisition is believed to be "healthy".
This implies the possibility of two kind of errors:

- type I error, which corresponds to triggering a False Alarms (FA or False Positive)
- type II error, which is a missed indication of damage although present (Missed Alarm, MA or 1-True Positive).

These error rates are usually collected in tables such as Table 8, which are very common when binary classification is considered. If the classification involves more than two groups, larger tables can be found with the name of confusion matrices.
On the contrary, in the field of Operational Research (OR), a discipline that deals with the application of analytical methods for making better decisions, the Receiver Operating Characteristic (ROC) is usually preferred for assessing the diagnostic ability of a binary classifier while its discrimination threshold is varied.
Figure 7(b) summarizes the true damaged rate as a function of the false alarm rate for some relevant effect sizes $d = {\mu_1 - \mu_2}/{\sigma}$ (the variance-normalized distance of the healthy and the damaged distributions) while the threshold takes all the possible values. The threshold corresponding to the very common 5% false alarm rate is highlighted. In general, anyway, the farthest is the ROC curve from the 1st – 3rd quadrant bisector, the better the classification, which obviously improves as the effect size is increasing.



(a) **(b)**

Figure 11. Receiver Operating Characteristic (ROC) as a function of the threshold (Gaussian distributions). **(a)** graphical summary of the table of type I and type II errors in yellow (Table 3). **(b)** ROC for binary classification with different effect sizes d and the position of the 5% false alarm rate. For $d = 0{,}2$ the performance is very poor as the ROC is near to the 1st-3rd quadrant bisector (nearly a random classifier).

Table 3. Type I and II errors in damage detection.

		True Health Condition:	
		Healthy	Damaged
Classification:	Healthy	No Alarm true healthy	Missed Alarm type II error
	Damaged	False Alarm type I error	Alarm true damaged

3 Results and Discussion

The methodology introduced in section 2 is tested on the experimental acquisition regarding high speed aeronautical bearings described in section 1.1. The dataset collecting 5 simple time features per each of the 6 channels (i.e., 30 features) is analysed separately for the 4 different load conditions. The confounding effect of the reducing speed (from 470 Hz to 0 Hz) will be compensated during the healthy training with 5 different approaches:

- Plain Euclidean distance (raw)
- Cointegration of the standardized features,
- Mahalanobis novelty,
- PCA orthogonal detection and whitening,
- Cointegration of the PCA-pre-whitened features.

The standardization which precedes the cointegration is necessary as the considered features have different order of magnitude, and this could lead to wrong estimates of the cointegrating matrix (poorly conditioned problem).
The novelty indices for the healthy reference and for the damaged conditions are reported in Figure 12 for condition 4 (load 1800N, decreasing speed). This graph highlights relevant considerations previously introduced in section 2.

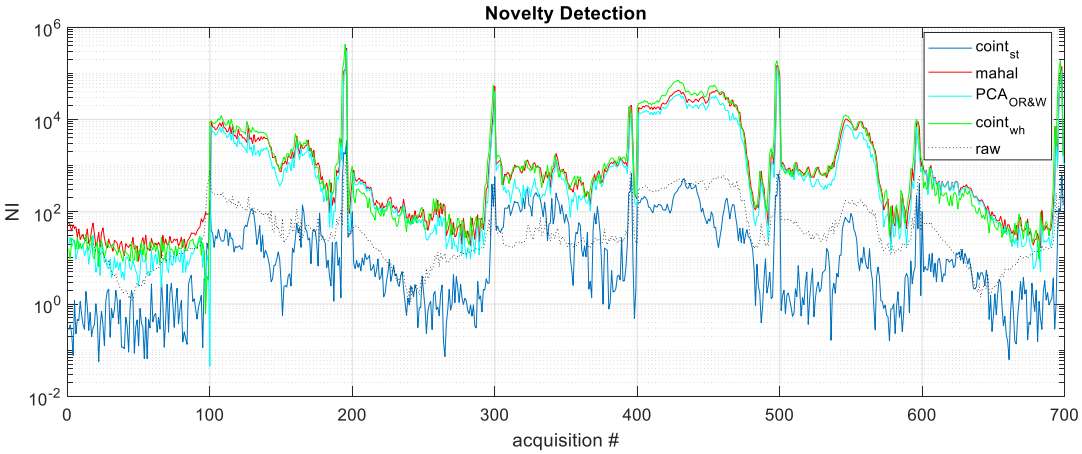


Figure 12. The Novelty Detection with the different NIs. 0-100 samples are the healthy reference, 100-200 corresponds to 1A damage, and so on until 600-700 coming from 6A damage.

In particular, cointegration based healthy NIs are stationary, while the Mahalanobis NIs are not. A trend is clearly visible, as this algorithm is targeted on decorrelation. An improvement in stationarity is obtained by neglecting the first 20 whitened principal components and focusing on the subspace individuated by the last 10. An additional note should be added to explain that the behaviour of the NIs at the end of all the run-down is ascribable to the fact that the record is not stopped exactly when the machine stops, so that the last points are practically acquiring just noise as the machine is already at a stop.

Despite the NIs curve already gives a qualitative impression of good detectability of almost all the different damages (from 1A to 7A), a quantitative comparison of the performances of the different methods is necessary. At this scope, Figure 13 reports the ROC curves for all the 4 different load conditions.

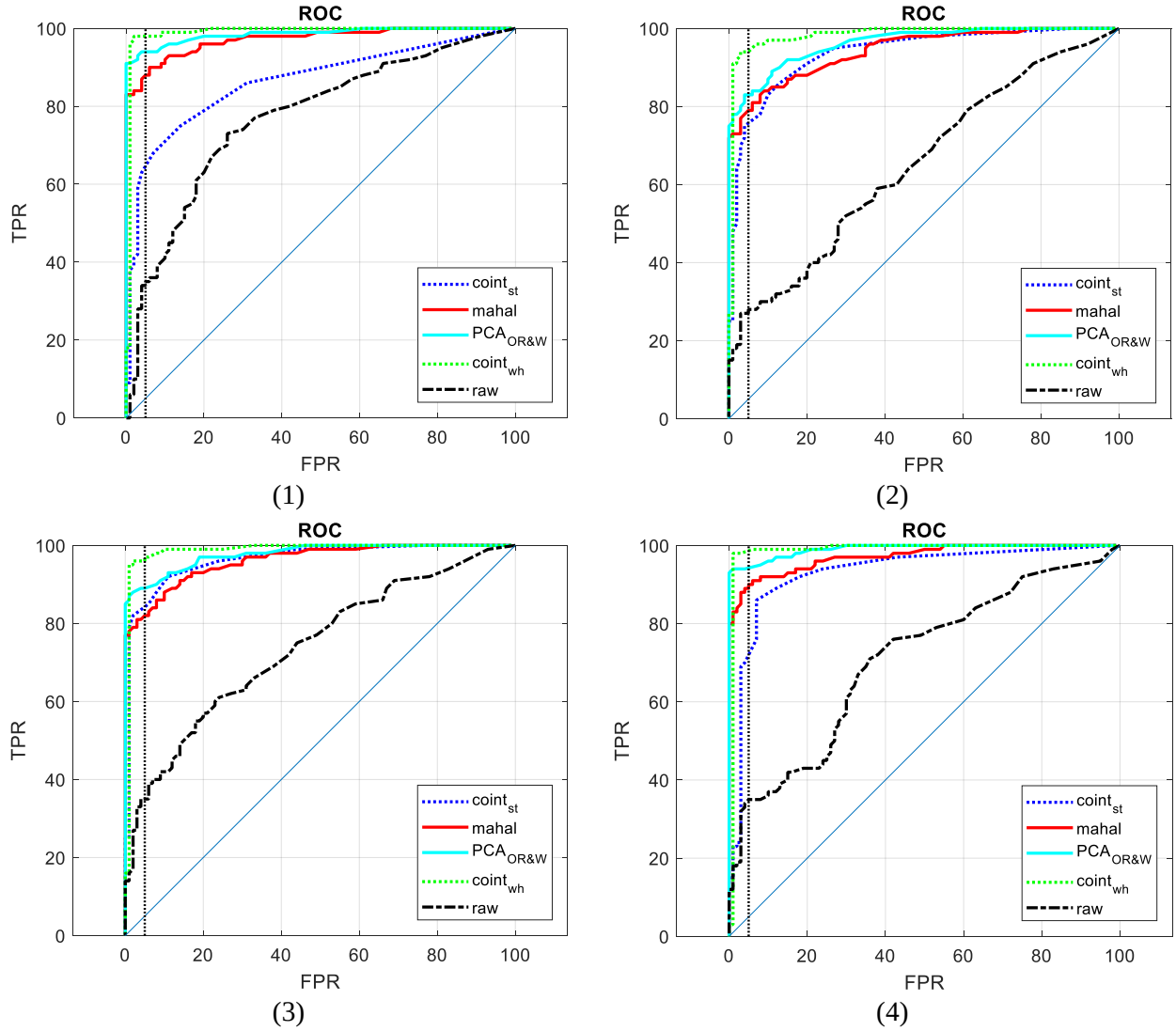


Figure 13. The ROC for Novelty Detection in the 4 loading conditions, respectively 1800, 1400, 1000, 0N

All the 4 graphs lead to a similar result: in this particular application, cointegration of the standardized features can enhance the damage detection in case of non-stationary rotational speed, but its performances are always comparable or inferior to Mahalanobis novelty detection. Removing the first 20 principal whitened components from the NIs computation (PCA-Ortogonal Regression & whitening), is able to further improve the results. The best performances anyway are given by cointegration of the PCA-pre-whitened features, as, accepting a 5% of false alarms, the missed alarms are always lower than 5% (100-95 % in the graphs).
The only issue with cointegration in this application is that decreasing the acceptable false alarm rate to 1% a knee of the curve is reached, so that the missed alarm rate increases dramatically. This phenomenon is not evidenced my Mahalanobis or PCA-Orthogonal Regression & whitening NIs.

4 Conclusions

This work focused on the compensation of confounders through two procedures which are established in univariate analyses, but not so well documented in multivariate cases:

- removal of linear deterministic trends in series by subtraction of a linear regression,
- removal of stochastic trends of difference stationary series by subtraction of the one-step ahead prediction from an autoregressive model.

In order to extend these approaches to the multivariate case, PCA orthogonal regression was used in the first case, while a vector autoregressive VAR(1) model was estimated in the second case.
In order to obtain consistent results from the VAR estimation, the analysed features were first standardized.
This is unnecessary for PCA, as a standardization can be more wisely performed on the principal space at a paltry expense. Indeed, just by normalizing the eigenvectors of the data covariance matrix so as to have the modulus equal to the corresponding eigenvalue, a sphering transform can be obtained. This is called PCA whitening and is "hidden" inside the Mahalanobis distance.
On the contrary, focusing only on the 1-D reduced dimensionality space individuated by the first principal component, orthogonal regression was performed. Therefore, a residual was found by removing the first principal component and focusing on the reduced dimensionality space.
Doing this on the whitened principal space, a Novelty index analogous to the Mahalanobis distance was found by summing the squares of the residuals (PCA-Orthogonal Regression & whitening).
Finally, cointegration was performed also on the PCA-pre-whitened dataset, to get the best of the two detrending strategies.
All the introduced methods enhanced the damage detection with respect to the raw Euclidean-distance-novelty-detection. The best damage detection in terms of reducing the missed alarms at a fixed maximum false alarm rate of 5% is without doubts the PCA-pre-whitened cointegration, which ensures missed alarm rates lower than 5% in all the loading conditions. Nevertheless, if the acceptable false alarm rate is decreased to 1%, the PCA-Orthogonal Regression & whitening proved to outperform all the other methods.
In general, the here proposed methodology, gives a quick, human independent and simple but effective way of performing damage detection also in case of non-stationary operational conditions. The nice compensation of the confounders in fact allows to enhance the damage, which can be easily highlighted by novelty detection.

References

[1] Daga A. P., Fasana A., Marchesiello S., Garibaldi L., *The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data*, Mechanical Systems and Signal Processing, 2019. DOI: 10.1016/j.ymssp.2018.10.010.

[2] Farrar C.R., Doebling S.W., *Damage Detection and Evaluation II*. In: Silva J.M.M., Maia N.M.M. (eds) Modal Analysis and Testing. NATO Science Series (Series E: Applied Sciences). Springer, Dordrecht, 1999. ISBN 978-0-7923-5894-7.

[3] Worden K., Manson G., Fieller N. R. J., *Damage detection using outlier analysis*, Journal of Sound and Vibration (2000).

[4] Stadnitski T., *Deterministic or Stochastic Trend: Decision on the Basis of the Augmented Dickey-Fuller Test*. Methodology European Journal of Research Methods for the Behavioral and Social Sciences. 6. 83-92. 2010. DOI: 10.1027/1614-2241/a000009.

[5] Sawalhi N., Randall R.B., *Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size*, Mechanical Systems and Signal Processing, 2011, DOI: 10.1016/j.ymssp.2010.09.009.

[6] Kessy A., Lewin A., Strimmer K., *Optimal Whitening and Decorrelation,* The American Statistician. 2015.DOI: 10.1080/00031305.2016.1277159.

[7] Cross E. J., Worden K., Chen Q., *Cointegration: a novel approach for the removal of environmental trends in structural health monitoring data,* Proc. R. Soc. A, 2011. DOI: 10.1098/rspa.2011.0023

[8] K. Worden, T. Baldacchino, J. Rowson and E. J. Cross, *Some Recent Developments in SHM Based on Nonstationary Time Series Analysis,* Proceedings of the IEEE, vol. 104, no. 8, pp. 1589-1603, Aug. 2016. DOI: 10.1109/JPROC.2016.2573596

[9] A. Deraemaeker, K. Worden, *A comparison of linear approaches to filter out environmental effects in structural health monitoring*, Mechanical Systems and Signal Processing, 2018, DOI: 10.1016/j.ymssp.2017.11.045.

[10] Jolliffe I.T., *Principal Component Analysis*, Springer, 2002. DOI: 10.2307/1270093

[11] Van Huffel S., Vandewalle J., *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, 1991, ISBN: 978-0-898712-75-9

Smart Structures

EMBEDDED SENSING MICRO-COMPONENTS FOR FIBRE REINFORCED COMPOSITE MATERIAL SYNTHESIS AND MONITORING

M. F. Zhang[1], X. P. Cui[1], M. Salvia[1*] and O. Bareille[1]
[1] Ecole Centrale de Lyon – LTDS, University of Lyon, 36 avenue Guy de Collongue 69134 Ecully, France
* Corresponding author (olivier.bareille@ec-lyon.fr)

Keywords: *bio-epoxy matrix composites, self-healing, dual-curing process*

1.Introduction

The failure of materials involved in large infrastructures, such as bridges or roads, may cause a huge loss of life, economy or a loss of services. An attractive solution to upgrade existing infrastructures and increase their life-span is to bond fibre reinforced thermosetting composite patches to strengthen the damaged structure and to prevent failure growth. And in doing so, understanding the properties as well as failure mechanism of the bonded thermosetting composites is essential. Initiations and growths of defects occur during almost all life of the composite up to final failure which occurs by coalescence of previous damages (cumulative damage). Hence, how to detect the latent cracks in materials and how to fix the damaged materials with more cost-effectively ways are research emphasizes.

As a non-destructive inspection method, in-situ Structural Health Monitoring (SHM) implies the use of sensors, such as piezoelectric ceramics (PZT), which are embedded within a structural material and provide real time performance feedback based on the measurement, such as electrical impedance. These sensors can be used to monitor the health state of thermosetting composites, from their curing process to the propagation of microcracks then to the end of their life cycle, which is useful for analyzing the fabrication or failure procedure of materials. By monitoring the change in the impedance spectrum which is linked to the changes of matrix viscoelastic properties as curing progressed, the different steps of the epoxy curing regarding molecular motion, viscosity, crosslinking density and their consequences on the mechanical properties of the material can be understood.

2. Experiment and method

2.1. Monitoring method with PZT

The monitoring circuit is shown as figure.1, an oscilloscope is applied as a voltage signal source. In output option, the parameters are set as followingg: sinusoidal, frequency 4kHz (the resonance frequency of PZT), 2.5V offset. An electric resistance with fixed value (6.8MΩ or other values in the order of mega ohms, which is in similar to the order of magnitude of impedance of PZT, in order to minimize the measuring error) and a PZT ceramic disk are connected in series with the source. A data acquisition system (DAQ) is connected to the circuit, detecting the electric tension of the source and the electric resistance mainly by a differential operational amplifier, and then the data is sent to a laptop. According to the tension relation in the series circuit, the following equation can be given:

$$\frac{U_0 - U_R}{U_R} = \frac{Z_P}{R}$$

where U_0 represents the total electric tension applied to the circuit, U_R represents the electric tension applied on the piezoelectric ceramic, R is the electric resistance value, Z_P is the impedance of PZT concerned in this monitoring. The expression of impedance Z_P can be written as following:

$$Z_P = A \cdot R$$

where

$$\mathrm{A} = \frac{U_0 - U_R}{U_R}$$

A series of programs based on MATLAB are running with the monitoring process. Variables such as total monitoring time and sampling frequency should be set before experiments, and should be adapted to monitoring requirement. For long-time monitoring such as epoxy resin curing monitoring or epoxy/flax composite monitoring, the system calculates and records the impedance every minute. But if this system is applied for monitoring the behaviour of material during tensile test, there would be greater sampling frequency.

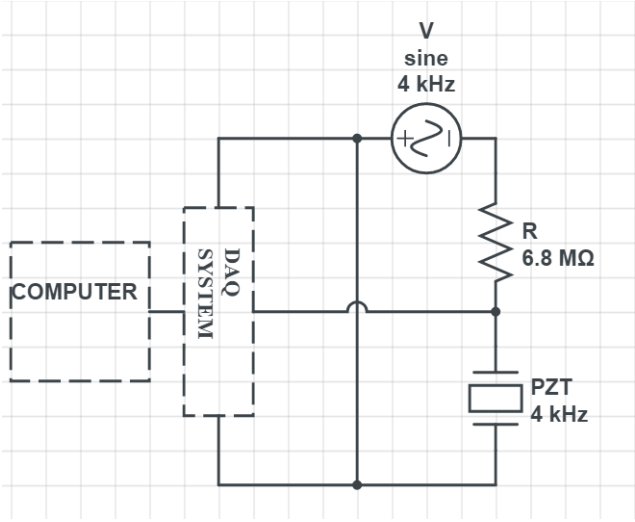


Fig. 1. Circuit diagram of monitoring system

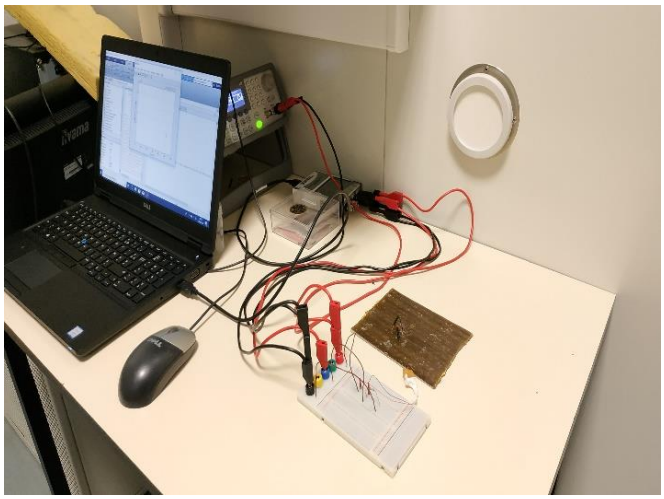


Fig. 2. Monitoring electric circuit and equipment

2.2. Material preparation

In this work, the resin for fabricating the patch composite is designed for ambient temperature curing. Its prepolymer is a low molecular weight "green" epoxy resin that made from bio-based epichlorohydrin. The hardener used is Cardolite NX5619, a solvent-free, low viscosity phenalkamine curing agent made through the Mannich reaction of cardanol from cashewnuts, formaldehyde, and amines. The natural fiber used is flax, and a quasi-unidirectional fabric made of untwisted rovings was used. The weft and warp ratio is 9/91 and the areal density is 200 g/m^2. The manufacturing of the natural fibers reinforced epoxy composite was accomplished by the use of wet hand lay-up process (2 plies) at room temperature.

2.3. Epoxy resin curing monitoring

Epoxy and its hardener are fully mixed in room temperature with mass ratio of 1.8:1, and stored in a cylindrical container of 50ml, which is show as figure.3. A PZT sensor (0.2 mm thick disk with a diameter of 7 mm) is placed inside the resin right after the mixture. Then the epoxy resin and PZT are transferred to the inside of a constant temperature chamber which is preheated to 40℃. The monitoring is carried on within this temperature during 16 hours. For the epoxy resin curing monitoring, a fixed value electric resistance of 9.68MΩ is used.

Fig. 3. Specimen of epoxy resin curing monitoring

2.4. Epoxy/flax composite curing monitoring

As a consequence of hand lay-up molding process, the proportion of natural fibers in the final composite can not be precisely controlled. In this experiment, composites containing about 14%wt flax fiber are used. A PZT sensor is embedded between the two plies of flax fabric during the manufacturing process at room temperature. Before the experiment, the composite plate has cured for 4 days under ambient temperature, and preheated 1 hour inside the constant temperature chamber. Different from the epoxy resin curing monitoring experiment, a fixed value electric resistance of 1.0MΩ is used, which indicates more intense electric current in the monitoring circuit. The composite plate with embedded sensor is similarly keep inside of the constant temperature chamber under 40℃

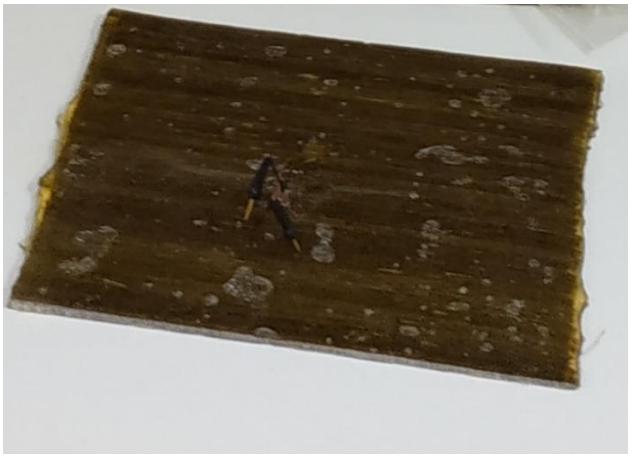


Fig. 4. Specimen of epoxy/flax composite monitoring

2.5. Composite monitoring in water environment

In order to measure the influence of water environment to the composite, another composite plate (manufactured with the same process) with PZT embedded is dipped in deionized water during a week. The change of impedance is recorded by the monitoring system. This experiment is implemented under ambient temperature, which is variable in a day.

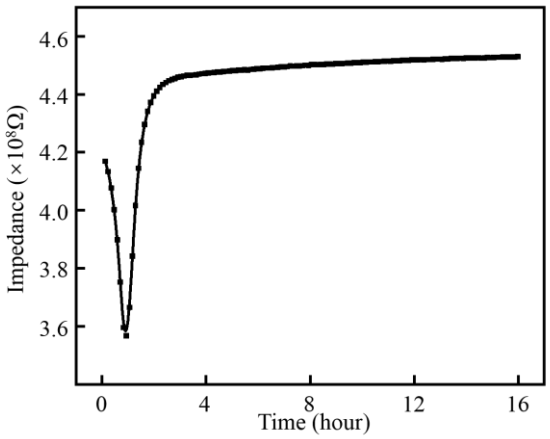


Fig. 5. Impedance of PZT sensor embedded in the specimen of epoxy resin

3.Results

3.1. Epoxy resin curing monitoring

The results of epoxy resin curing are shown in Fig. 5. According to the figure, the impedance of PZT has a rapid decrease in the first hour from the beginning of experiment. Then the impedance increases similarly fast as the decrease period. After about 2 hours, the impedance arrives at a platform, where only slight but continuous augmentations can be observed.

The decreasing section is considered caused by the change of temperature, which indicates that, comparing to room temperature, the augmentation of temperature surly reduce the viscosity of epoxy resin. The fast increase section indicates an abrupt change of viscosity of epoxy resin, which matches the gelation of polymer. The gel point of epoxy resin should be found before the lower pole of impedance curve, synthesizing the heat-related decrease.

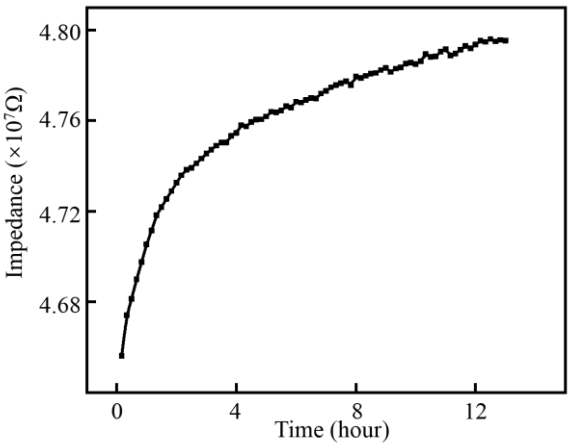


Fig. 6. Impedance of PZT sensor embedded in the specimen of epoxy/flax composite

3.2. Epoxy/flax composite curing monitoring

The behavior of pre-cured epoxy/flax composite is principally different from that of pure epoxy resin. After the four-day-long curing, the epoxy resin has already passed its gel point for rather long time, the cross-linking reaction has well developed. The epoxy resin has mainly lost its fluidity. The characteristic of composite should be described as viscoelasticity, and storage modulus is used as a reference of curing process.

The increasing rate of impedance during the whole composite curing monitoring is about 0.2% per hour, which is very close to the rate found in epoxy resin monitoring, at the platform section. Although, a relatively faster augmentation is observed at the first hours of monitoring, as a result of higher temperature in monitoring experiment than curing temperature. The curing process is accelerated by heat, as shown in Fig. 6.

Aiming at qualitative analysis of the relation between storage modulus and curing of epoxy, a

series of DMA (Dynamic mechanical analysis) tests were carried on. The composite was tested after 1, 2, 3, 4, 6 and 12 days of curing. The storage modulus under 40℃ are shown in Fig. 7. An evident increasing tendency of storage modulus is found, which is associated with polymerization of epoxy.

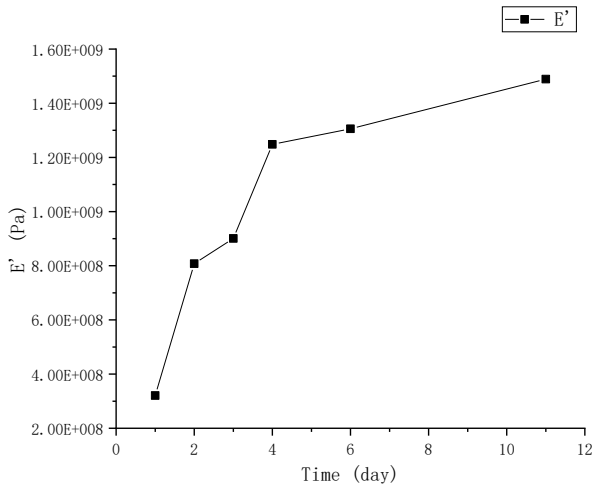


Fig. 7. Storage modulus of epoxy/flax composite measured by DMA tests in a series of days

3.3. Composite monitoring in water environment

After a week's monitoring, the impedance variation is shown as Fig. 8. A curve of tendency (blue line in Fig. 8) is added by Gaussian fitting which is only utilized to give a clearer view.

According to the figure, the curve is a kind of combination of two parts. One is the decreasing tendency which is driven by water absorption of composite. The hydrophilic groups in cellulose or other molecules allow flax fiber to have relatively high water absorption rate, comparing to synthetic fibers, which causes degradation of mechanical properties of composite. The epoxy resin was still curing, thus a balance appears with saturation of water absorption, and then the curing is more dominant, which explains the final slight augmentation of impedance. The other part is a sinusoidal within an envelope. This part is caused by the variation of room temperature during a day, which can be proved by its period about 8000 seconds

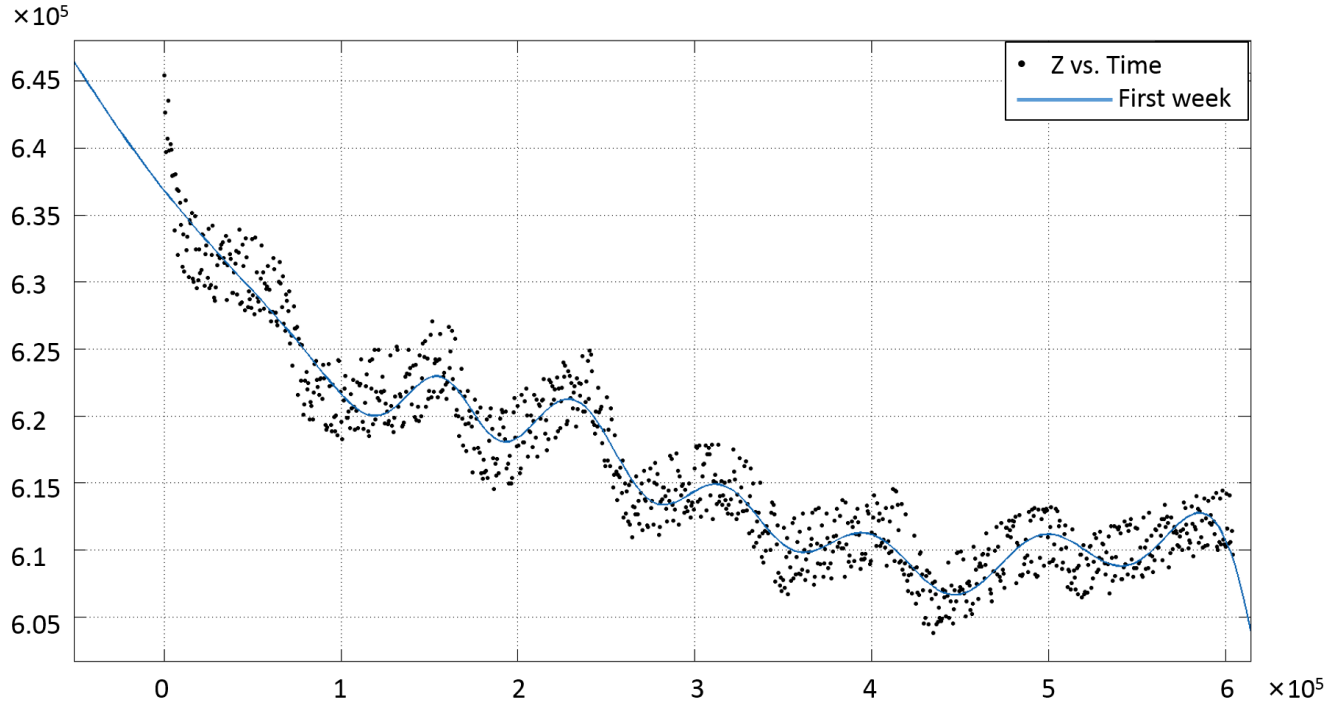


Fig. 8 Impedance of PZT embedded in epoxy/flax composite steeped in water

4.Conclusion

The change in the impedance spectrum which is linked to the changes of matrix viscoelastic properties as curing progressed was used to follow the different steps of the epoxy matrix curing.

The results showed that the piezoelectric transducers are well suited to in-situ monitor the reaction progress during isothermal curing of a flax reinforced epoxy materials. Other factors being able to influence the mechanical properties of composite can be correlated with the change of impedance and be monitored by PZT sensor. After curing, the sensor was used as damage detector. In order to assess the efficiency of such a system for health monitoring, tests will be performed based on tensile measurements using digital image correlation (DIC), classical acoustic emission and scanning electron microscope (SEM).

References

[1] F. Regina, R. M. William, B. Derrick, A. Purvis, A. Tiwari, and G. D. M. Serugendo "Self-healing and self-repairing technologies". *The International Journal of Advanced Manufacturing Technology*, 69(5-8),1033-1061, 2013.

[2] G. Park, H. Sohn, C. R. Farrar and D. J. Inman "Overview of piezoelectric impedance-based health monitoring and path forward". *Shock and Vibration Digest*, 35(6), 451-463, 2006.

[3] D. G. Bekas, K. Tsirka, D. Baltzis and A.S. Paipetis "Self-healing materials: a review of advances in materials, evaluation, characterization and monitoring techniques". *Composites Part B: Engineering*, 87, 92-119, 2016.

[4] Dae-Hyun Han, Lae-Hyong Kang "Piezoelectric characteristics of PNN-PZT/Epoxy paint sensor accordingto the poling conditions". *Sensors and Actuators*, A269 (2018), 419–426.

[5] P. Negi, T. Chakraborty, Naveet Kaur, Suresh Bhalla "Investigations on effectiveness of embedded PZT patches at varyingorientations for monitoring concrete hydration using EMI technique". *Construction and Building Materials*, 169 (2018), 489–498.

[6] Hidemi Mutsuda, Yoshikazu Tanaka, Rupesh Patel, Yasuaki Doi, Yasuo Moriyama, Yuji Uminod "A painting type of flexible piezoelectric device for ocean energyharvesting". *Applied Ocean Research*, 68 (2017), 182–193.

[7] T. Visalakshi, S. Bhalla, A. Gupta "Monitoring early hydration of reinforced concrete structures using structural parameters identified by piezo sensors via electromechanical impedance technique". *Mechanical Systems and Signal Processing*,99 (2018), 129–141.

[8] Gyuhae Park, Hoon Sohn, Charles R. Farrar and Daniel J. Inman "Overview of Piezoelectric Impedance-Based Health Monitoring and Path Forward" *The Shock and Vibration Digest November 2003*, 451-463.

[9] Ismael Payoa, J.M. Hale "Sensitivity analysis of piezoelectric paint sensors made up of PZT ceramic powder and water-based acrylic polymer" *Sensors and Actuators A*, 168 (2011), 77–89.

Vibration Control of Cable-Driven Parallel Robot for 3D Printing

Florian LACAZE[1], Simon CHESNE[1], Didier REMOND[1]

[1]Univ Lyon, INSA-Lyon, CNRS UMR5259, LaMCoS, F-69621, France
florian.lacaze@insa-lyon.fr

Abstract

These last years, the additive manufacturing and 3D printing technologies have known some major breakthroughs. The motion of a printer head can be made with cable transmission. The deployment of the cable-driven parallel robots (CDPR) in the industry is studied in very various application fields for their low cost and large workspace. Furthermore, the use of cables for the transmission induces a reduction of the mobile parts' masses, compared to a rigid transmission, which enables to reach higher accelerations. Moreover, the structure of a CDPR is modular and reconfigurable thanks to the repositioning of the actuators' anchor points. However, the lack of rigidity of a CDPR raises issues of accuracy and the rise of vibrations, which can be generated by the trajectory of the mobile parts, the actuators, the friction between pulleys and cables or disturbances.
Several dynamic models of cables have been studied to understand the vibrating behaviour of a CDPR: a simple elastic model of springs with positive tensions, a lumped mass-spring model and a finite elements model based on a continuous one for the cables dynamics. The numerical simulation of the dynamic behaviour of the CDPR with these models enables the analysis of an appropriate control system and the design of a controller. It aims at ensuring an accurate positioning and a decrease of vibrations.
In this contribution, we will firstly present the dynamic behaviour's model and the issue of the actuation's redundancy, systematically present on these robots to guarantee stiffness with the tension in the cables. A comparison will be done between the effects of the models on the conception and the performance of the controller. Thus, we explain that significant decreases in the vibration levels may be observed with the use of PID controllers. The generalisation of the command, the use of active control technologies and an experimental validation will be the next steps of this study.

1 Introduction

Cable-Driven Parallel Robots (CDPR) are a type of parallel kinematic machines in which cables link a mobile platform to a fixed base. Reels allow the control of cables length and cables tension. Several applications of CDPR have been studied, such as high speed manipulation [1], heavy materials handling [2], haptic perception [3]. Their large workspace enables to visualise [4, 5] or print [6] large 3D objects.

Modelling CDPR requires to take into account the cables dynamics, which are complex and non-linear. The main models of cables dynamics are the following :

- Elastic models (valid and efficient for low-density and thin cables);
- Analytic models with non-linear equations (Irvine model [7]);
- Semi-analytic methods for cable with small sag [8, 9];
- Lumped mass methods [10];
- Finite element models, using cables with time-dependent length [11, 12]

.

A CDPR constituted of a punctual mass linked by two cables driven by two actuators is considered in the following paper. It is an over-actuated CDPR, since the only controlled Degree Of Freedom (DOF) of the effector is the horizontal position.

The contribution of this paper is to compare three dynamic models of a CDPR in Section 2: an elastic spring model, a lumped mass-spring model and a variable length finite elements model. Section 3 is dedicated to the strategies of command. Results are discussed in Section 4.

2 Models of dynamic behaviour

2.1 Elastic model

An effector of mass M is linked with two cables, modelled by two elastic springs. The cables tension is controlled by means of actuators, allowing the effector motion control. The distance between the two cable reels A_1 and A_2 is $d = 1$ m. This configuration is shown in Figure 1.

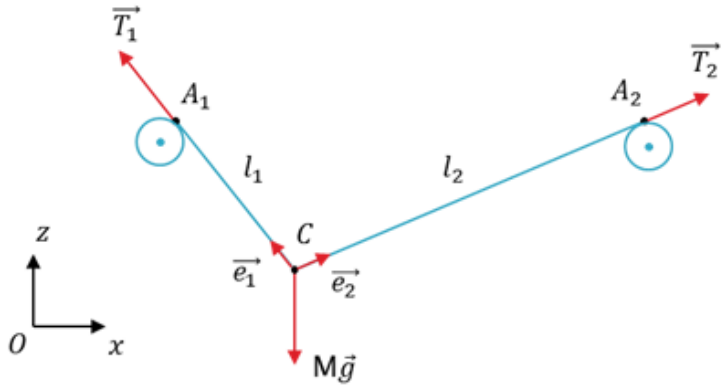


Figure 1: CDPR with 2 cables

Given the Young's modulus $E = 102$ GPa of the cable and its cross-sectional area $A = 1.76e^{-6}$ m^2, the cable tension can be described as :

$$T = EA\frac{l - l_0}{l_0} \tag{1}$$

with l the strained cable length and l_0 the unstrained cable length.

The effector of mass M is constrained under the two cable tensions T_1 and T_2, and its dynamic equation is given by :

$$\begin{pmatrix} \ddot{x} \\ \ddot{z} \end{pmatrix} = \frac{1}{M}(T_1\vec{e_1} + T_2\vec{e_2}) + \vec{g} \tag{2}$$

Unit vectors state the direction of efforts transmitted through the cables. For the i^{th} cable, we have :

$$\vec{e_i} = \frac{\overrightarrow{OA_i} - \overrightarrow{OC}}{||\overrightarrow{OA_i} - \overrightarrow{OC}||} \tag{3}$$

2.2 Lumped mass-spring model

Each cable is now lumped into $N = 40$ mass-spring elements. Each of them is formed by a spring, of unstrained length $l_0 = \frac{L_0}{N}$ and stiffness $k_0 = \frac{EA}{l_0}$, and by a mass $m = \rho A l_0$.

As shown in Figure 2, the length of modelled cable is larger than the length between A_1 and the end-effector. The purpose is to have always the same amount of elements when the effector is moving. The cable located to the left of A_1 and to the right of A_2 corresponds to the cable rolled in the reels.

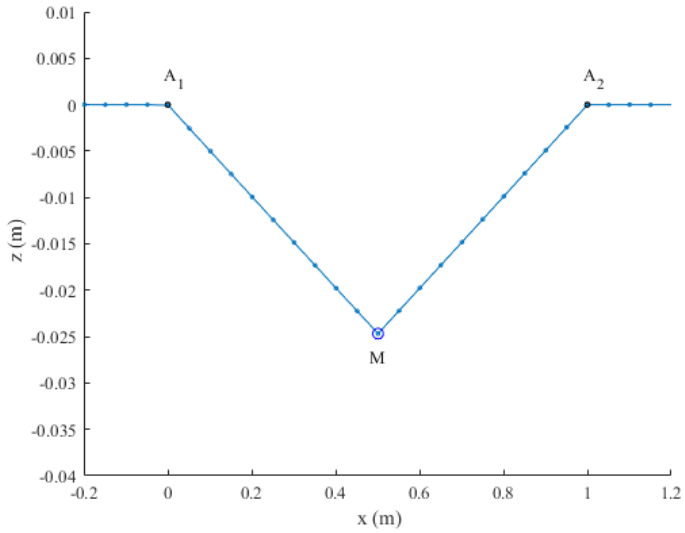


Figure 2: Lumped mass-spring model

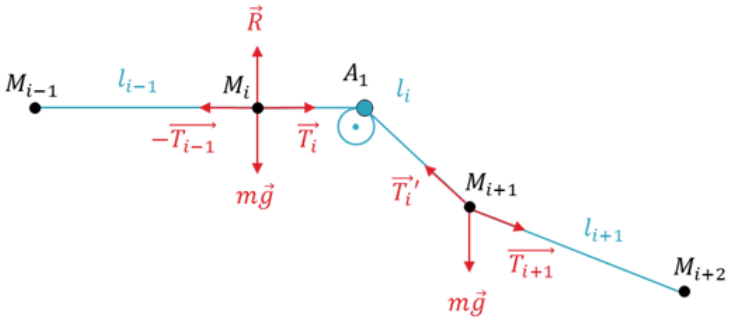


Figure 3: 3 elements of the lumped mass-spring model

To understand the difference of dynamics between the rolled cable and the free one, the Figure 3 gives a closer look at the lumped mass-spring model near A_1. We consider the points M_j with $j = \{i-1, i, i+1, i+2\}$. The corresponding lengths are :

$$\begin{cases} l_{i-1} = ||\overrightarrow{M_{i-1}M_i}|| \\ l_i = ||\overrightarrow{M_iA_1}|| + ||\overrightarrow{A_1M_{i+1}}|| \\ l_{i+1} = ||\overrightarrow{M_{i+1}M_{i+2}}|| \end{cases} \tag{4}$$

The tensions, computed using these lengths, are the following :

$$\begin{cases} T_{i-1} = k(l_{i-1} - l_0) \\ T_i = k(l_i - l_0) \\ T_{i+1} = k(l_{i+1} - l_0) \end{cases} \tag{5}$$

We can also determine the accelerations $\ddot{X}_i = \begin{pmatrix} \ddot{x}_i \\ \ddot{z}_i \end{pmatrix}$ and $\ddot{X}_{i+1}$ before and after A_1. In M_i the mass element is only constrained by horizontal tensions and in M_{i+1} unit vectors give the directions of the efforts :

$$m\ddot{X}_i = \begin{pmatrix} -T_{i-1} + T_i \\ 0 \end{pmatrix} \tag{6}$$

$$m\ddot{X}_{i+1} = \begin{pmatrix} T_i \frac{x_{A_1} - x_{i+1}}{||\overrightarrow{A_1M_{i+1}}||} + T_{i+1}\frac{x_{i+2} - x_{i+1}}{l_{i+1}} \\ T_i \frac{z_{A_1} - z_{i+1}}{||\overrightarrow{A_1M_{i+1}}||} + T_{i+1}\frac{z_{i+2} - z_{i+1}}{l_{i+1}} \end{pmatrix} - m\vec{g} \tag{7}$$

2.3 Variable length finite elements model

We consider that each cable is divided in $N = 10$ elements with the same unstrained length $l = \frac{L}{N}$. The length l is time-depending during the movement of the CDPR and its variations are taken into account in the dynamics. The details of the equations can be found in the article of J. Du and al. [13].

For a cable of length l, the kinetic energy is :

$$T = \int_0^l \frac{1}{2}\mu \dot{\vec{r}}^T . \dot{\vec{r}} \mathrm{d}s \tag{8}$$

where μ is the mass per unit length and $\vec{r}$ is the position vector of the element.

The elastic potential and gravitational potential energy of the same cable is :

$$U = \int_0^l (\frac{1}{2}EA\varepsilon^2 - \mu g \vec{r}^T . \vec{z}) \mathrm{d}s \tag{9}$$

where ε is the element strain.

The changing mass of the system can be written as :

$$\delta H = \delta \vec{r}_j{}^T f_j + \mu \dot{\vec{r}}_1{}^T . \delta \vec{r}_1 v_1 + \mu \dot{\vec{r}}_2{}^T . \delta \vec{r}_2 v_2 \tag{10}$$

with v_1 the winding speed in A_1 and v_2 the winding speed in A_2.

The application of the Hamilton's principle on the cables gives the following equation :

$$\int_0^t T \mathrm{d}t - \delta \int_0^t U \mathrm{d}t + \int_0^t \delta H \mathrm{d}t = 0 \tag{11}$$

Eventually, this leads to the dynamic equation, which describes the position $\vec{r}_j$ of the j index point :

$$\mathbf{m_j} \ddot{\vec{r}}_j + \mathbf{c_j} \dot{\vec{r}}_j + \mathbf{k_j} \vec{r}_j = \vec{f}_j + \vec{f}_j^g \tag{12}$$

In Eq. 12, $\mathbf{m_j}$ is the conventional mass matrix, $\mathbf{c_j}$ describes, with convective terms, an energy transfer due to length variations. The stiffness matrix $\mathbf{k_j}$ is composed of the axial deformation of the element and of a second term due to the first and second derivatives of the element length variation with respect to time. $\vec{f}_j$ is the nodal force acting on the cable. $\vec{f}_g$ is the equivalent nodal force of the cable element self-weight.

3 Strategies of command

3.1 Trajectory

The chosen trajectory is a *step5*-function (Equation 13), which enables to avoid chocs and discontinuities and to lower vibrations. The principle is to go smoothly from x_1 to x_2 between t_1 and t_2 ; it means that the velocities at t_1 and t_2 are both equal to zero.

$$\mathrm{step5}(t) = \begin{cases} x_1 & \text{if } t < t_1 \\ x_1 + a\Delta^3(t)(10 - 15\Delta(t) + 6\Delta^2(t)) & \text{if } t_1 \leq t < t_2 \\ x_2 & \text{if } t \geq t_2 \end{cases} \tag{13}$$

with $a = x_2 - x_1$ and $\Delta(t) = \frac{t - t_1}{t_2 - t_1}$.

Here we consider $t_1 = 0$ s, $t_2 = 0.2$ s, $x_1 = 0.5$ m and $x_2 = 0.6$ m. The maximum speed in the case of the parameters mentioned below is 0.9375 m.s^{-1} at $t = 0.1$ s. The maximum acceleration is ± 14.434 m.s^{-2} at $t = (0.0423; 0.1577)$ s.

3.2 Control

There are two ways of controlling the end-effector of a CDPR : by controlling the cables length or their tension. The choice, here, has been made to control the cables tension, mainly for observed stability reasons. Moreover, the choice is led by the fact that the mechanism is redundantly constrained, which means that the number of cables driving the end-effector is one greater than the number of the robot's DOF.

Because of the actuation's redundancy, we face an issue of tension distribution in the cables. The next step is to chose a set of additional equality or inequality constraints before designing an algorithm of optimisation that finds an optimal tension distribution [14]. The most commonly used constraints in order to get the optimal tension distribution are :

- positive cables tension (to avoid an unstressed cable) [15];
- minimal sum of cables tension (to minimise the actuators' energy);
- tensions in an interval $[t_{min}; t_{max}]$ [16];
- continuous cables tension.

For the tensions distribution, we chose to force 100 N of pretension on each cable, which means that $\sum T = T_1 + T_2 = 200$ N. PID controllers are robust enough to be used in CDPR control, even if CDPR behaviour is non-linear [17]. Our PID controller provides the effort needed by the effector so that it follows the desired trajectory. Thus, the tensions distribution in the controllers will be written : $T_1 = \frac{\sum T - F}{2}$ and $T_2 = \frac{\sum T + F}{2}$.

The controllers' gains have been settled on the elastic model, which enables a faster and easier setting. We do not present the optimisation of the parameters for the controller, but a comparison between the controlled models. The PID controllers' gains for each model are :

$$\begin{cases} P = 1860 \\ I = 8700 \\ D = 100 \\ \text{filter coefficient } N = 540 \end{cases}$$

The blocks containing the calculation of the trajectory, the controller and the dynamic model are shown in the Figure 4.

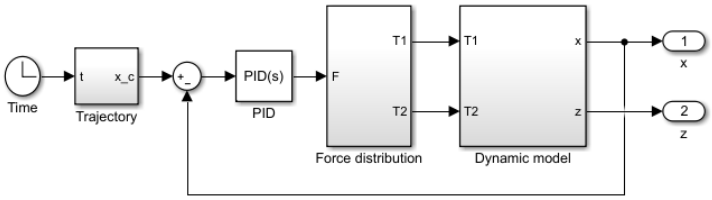


Figure 4: Simulink diagram

4 Results and discussion

Computation of the dynamic models is done using MATLAB and SIMULINK. Parameter settings of the numerical simulation are as follows : MATLAB solver used is *ode45*, the relative tolerance is $1e^{-5}$, the minimum time step is $1e^{-5}$ s and the sample time is $1e^{-}5$ s.

The initial position of the elastic model (Section 2.1) is computed by solving the static equilibrium of the Equation 14, which gives the starting position for the end-effector.

$$T_1\vec{e}_1 + T_2\vec{e}_2 + M\vec{g} = \vec{0} \tag{14}$$

To get the initial positions of the models presented in 2.2 and 2.3, we add a damping in the dynamic parameters of the model and we compute a free simulation in order to determine the equilibrium of each element of the model. The Figure 5 outlines the simulation results of the three dynamic models introduced in Section 2 for a computing time of $t_{\text{final}} = 5$ s. Very small oscillations, in the range of $1e^{-5}$ m, are observed for x. They are similar in all the models. The error, computed by $error = x_{\text{trajectory}} - x_{\text{effector}}$ and presented in Figure 6, is lower than $1e^{-2}$ in the transition phase, and lower than $1e^{-5}$ after 1 s of simulation.

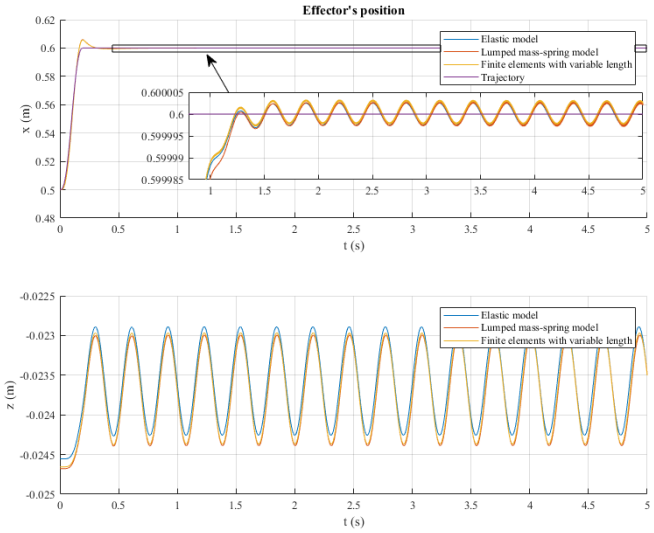


Figure 5: Evolution of x and z effector's position

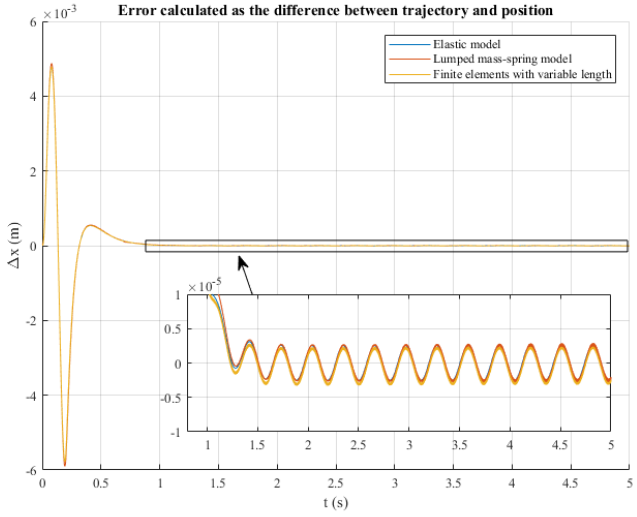


Figure 6: Position error of the CDPR

For z, oscillations are observed at the same frequency, with an amplitude of 1 mm. The mean of z oscillations in the elastic model is higher than in the two other models, because this one does not take into account the mass of the cables. The Figure 7 shows a spectrum analysis of the error signals between $t = 1$ s and $t = 5$ s. A peak at 20.11 rad.s^{-1} is observed for the three models. It corresponds to the end-effector's vertical oscillations,

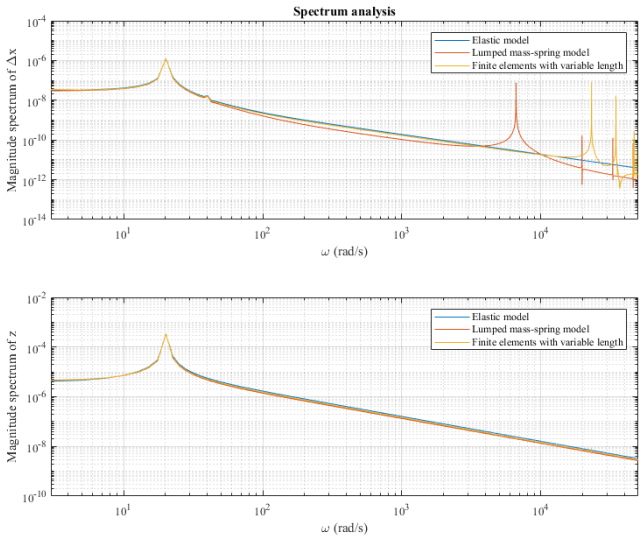


Figure 7: Spectrum analysis of Δx and z

as shown in Figure 5. The vertical resonance frequency of a system constituted of two strings and one mass is given by the following equation :

$$\omega = \sqrt{\frac{T}{M}\left(\frac{1}{x} + \frac{1}{d-x}\right)} \tag{15}$$

With the parameters values used in the three models, this resonance frequency is equal to 20.41 rad.s^{-1}, which is closed to the frequency of the peak on the spectrum. The high frequency peaks, observed for the lumped mass-spring model and the variable length finite elements model, corresponds to the discretization of the cables.

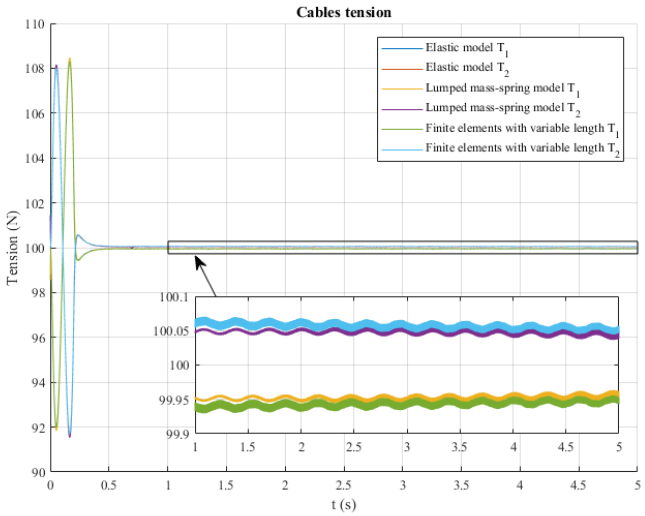


Figure 8: Cables tension T_1 and T_2

Figure 8 represents the evolution of the cables tension. In order to follow the trajectory acceleration, the tension T_1 decreases down to 92 N, then increases up to 108 N, and finally stabilises around 99.95 N. The tension T_2 does an inverse evolution and finally stabilises around 100.05 N. In the configuration where the effector's

position is $x = 0.6$ m, the two tensions are not exactly the same because the cables are not symmetrical. This enables the end-effector to get the desired position .

5 Conclusion

Three cable models for a CDPR have been presented in this paper : an elastic model, a lumped mass-spring model and a finite elements model. Eventually, these models' results are consistent with each others. For the three models, the PID controller enables to achieve good performances. Indeed, the static error is lower than $1e^{-5}$ and the overshoot is lower than 1 cm. It can be noticed that the PID controller is robust enough to control the effector's position and keeps constant gains, regardless of the dynamic model used in the simulation. The gains' adjustment was not problematic to control numerical models, thus we can hope that adjusting the gains on an experimental model will not be problematic either.

In further studies, we shall test other strategies of control, such as inverse dynamics control or model-free control, add the behaviour of the actuators to the dynamic model, and build a test bench to bring an experimental validation to the results that are introduced in this paper.

References

[1] Sadao Kawamura, Hitoshi Kino, and Choe Won. High-speed manipulation by using parallel wire-driven robots. *Robotica*, 18(1):13–21, 2000.

[2] Roger Bostelman, James Albus, Nicholas Dagalakis, Adam Jacoff, and John Gross. Applications of the nist robocrane. In *Proceedings of the 5th International Symposium on Robotics and Manufacturing*, pages 14–18, 1994.

[3] Martin Otis. *Analyse, commande et intégration d'un mécanisme parallèle entraîné par des câbles pour la réalisation d'une interface haptique comme métaphore de navigation dans un environnement virtuel*. PhD thesis, Faculté des études supérieures de l'Université Laval, 2009.

[4] Clément Gosselin and Samuel Bouchard. A Gravity-Powered Mechanism for Extending the Workspace of a Cable-Driven Parallel Mechanism: Application to the Appearance Modelling of Objects. *International Journal of Automation Technology*, 4(4):372–379, July 2010.

[5] Samuel Bouchard. *Géométrie des Robots Parallèles Entraînés par des Câbles*. PhD thesis, Faculté des études supérieures de l'Université Laval, 2008.

[6] Eric Barnett and Clément Gosselin. Large-scale 3d printing with a cable-suspended robot. *Additive Manufacturing*, 7:27–44, July 2015.

[7] H. Max Irvine. *Cable structures*, volume 17. MIT press Cambridge, MA, 1981.

[8] Younes Achkire. *Active Tendon Control of Cable-Stayed Bridges*. PhD thesis, Université Libre de Bruxelles, 1997.

[9] Frederic Bossens. *Amortissement actif des structures câblées : de la théorie à l'implémentation*. PhD thesis, Université Libre de Bruxelles, 2001.

[10] Donya Mohammadshahi. *Dynamics and Control of Cables in Cable-Actuated Systems*. PhD thesis, McGill University, Montreal, 2013.

[11] Haluk Ozdemir. A finite element approach for cable problems. *International Journal of Solids and Structures*, 15(5):427–437, 1979.

[12] P.-H. Wang, R.-F. Fung, and M.-J. Lee. Finite Element Analysis of a Three-Dimensional Underwater Cable with Time-Dependent Length. *Journal of Sound and Vibration*, 209(2):223–249, January 1998.

[13] Jingli Du, Chuanzhen Cui, Hong Bao, and Yuanying Qiu. Dynamic Analysis of Cable-Driven Parallel Manipulators Using a Variable Length Finite Element. *Journal of Computational and Nonlinear Dynamics*, 10(1):011013, September 2014.

[14] G. Meunier, B. Boulet, and M. Nahon. Control of an Overactuated Cable-Driven Parallel Mechanism for a Radio Telescope Application. *IEEE Transactions on Control Systems Technology*, 17(5):1043–1054, September 2009.

[15] So-Ryeok Oh and S.K. Agrawal. Cable-suspended planar parallel robots with redundant cables: controllers with positive cable tensions. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*, volume 3, pages 3023–3028, Taipei, Taiwan, 2003. IEEE.

[16] Johann Lamaury. *Contribution à la commande des robots parallèles à câbles à redondance d'actionnement*. PhD thesis, Université Montpellier II - Sciences et Techniques du Languedoc, 2014.

[17] A. Mohammad Khosravi and Hamid D. Taghirad. Robust PID control of fully-constrained cable driven parallel robots.

Tuesday

Angular approaches

NUMERICAL AND EXPERIMENTAL LOADS ANALYSIS ON A HORIZONTAL-AXIS WIND TURBINE IN YAW

Davide Astolfi[1], Francesco Castellani[1], Francesco Natili[1], Matteo Becchetti[1]

[1]University of Perugia, Department of Engineering, Via G. Duranti 93 - 06125 Perugia (Italy)
davide.astolfi@unipg.it, francesco.castellani@unipg.it,
francesco.natili@collaboratori.unipg.it, matteo.becchetti@unipg.it

Abstract

The characterization of wind turbines in yawed conditions is one of the most important topics as regards the latest advances in optimizing power production and mechanical behavior in wind farms. The classical wind turbine control strategy consists in keeping the rotor constantly aligned with wind direction: whereas this approach maximizes the power coefficient of each single turbine, it might not be the best solution when, in a wind farm, upwind turbines generate wakes on downwind ones. Considering this, yawing the rotors gives a steer to wakes, improving the flow on downwind turbines. This new kind of control strategy has been attracting the scientific interest not only by an energetic point of view, but also as regards the mechanical behavior of turbines operating not aligned with wind, in particular for what concerns generation of forces and vibrations. On these grounds, the aim of this paper is to study in deep how a wind turbine works on yawed configurations. In order to do this, wind tunnel tests have been performed with yaw angles that range over from -45° to 45° on a 2 m. diameter small scale wind turbine. Experimental measurements of forcespower and tower vibrations are then compared with the results of simulations from two different codes. The first, called BEM, is internally developed following the principles of Blade Element Momentum theory and it is used to estimate forces and torque acting on the rotor. The second implemented model is developed using the FAST (Fatigue, Aerodynamics, Structures and Turbulence) software, developed at the National Renewable Energy Laboratory (NREL). FAST simulations provide in output forces, torque and vibrations of tower and blades. Simulations are set up with similar conditions as the wind tunnel tests, with many yaw angles and steady wind speed. One of the main results of this study is that there is a remarkable agreement between simulations and measurements as regards the estimate of the power coefficient C_P in yawed and non-yawed configurations. In spite of this, thrust coefficient C_T is not faithfully estimated when the yaw angles is vanishing. This matter of fact is then explained by the fact that low-fidelity numerical models are not capable in reproducing reliably the effect of the tower blockage, slowing down the air stream in its proximity. As a consequence, when a blade passes close this area of reduced flow speed, the generation of aerodynamic forces decreases. In yawed configurations, this phenomenon is less relevant because of the increased distance between blades and tower on air flow direction.

1 Introduction

The optimization of power production and mechanical behavior of wind turbine through advanced control strategies [1, 2] has been recently becoming one of the main topics in wind energy research. The blade pitch [3, 4] and the yaw management are two very fertile fields of investigation for the research in wind turbine control optimization [5, 6].

The classical approach to control wind turbines nacelle orientation consists in continuously following the wind direction, in order to maintain the rotor axis constantly parallel to air flow. This method guarantees that the single turbine has always the maximum tip-speed ratio and so the maximum energetic production. In spite of this, in wind farm configurations an aspect that has to be considered is that upstream wind turbines generate wakes affecting the downstream ones and this affects the power production [7] and the mechanical loads [8]. By this point of view, new methods for active wind farm control are oriented: the general idea is that a slight decrease of the power produced by the single upwind turbine can optimize the total production of the entire farm. Haces-Fernandes et al. [9] states that a selective turbine deactivation allows an enhancement on wind

farm production and found that improvement is more prominent as the size of the turbines, and so rotor diameter, increases. Another method is the derating [10], that consists on running some turbines with a non-optimal rotational speed in order to catch less energy from air stream but generating less wake effect on downwind turbines: in this case the increase of energy production can range from 1.86% to 6.24%. A novel approach, instead, called wake steering [11, 12, 13], consists in keeping the upwind turbines not aligned with wind direction with the purpose to deviate the wakes and let the downwind turbines to be invested by an air stream with a more energetic content. For a single wind turbine, the power is related to ϕ, the yaw angle between the wind flow and the rotor, with a cosine cube law [14]: as a consequence, the energy production decreases as the yaw angle increases. To understand the net increase of generated power at wind farm level, Archer et al. [12] found that yawing the first row of a turbine array of 20°, the power of the following rows increases profitably (more than the losses of the first row).
To have a wider outlook on wake steering wind farm control, it is necessary to consider not only the effect on energy production, but also the possible side effects on wind turbine structural integrity. For example, Bakshi [15] estimated the reliability of blades in yawed asset, performing a stress analysis in different yaw configurations.

On these grounds, the present study aims at providing a contribution to the experimental analysis and numerical characterization of horizontal-axis wind turbines in yawed conditions. Wind tunnel measurements on a 2 m. diameter turbine are performed, with yaw angles ranging from 0 to $\pm 45°$: forces generated by the rotor, nacelle accelerations, rotational speed and generator power are monitored. Experimental data are then compared to numerical results of simulations performed with two different algorithms. The first is called BEM and has been internally developed according to Blade Element Momentum theory. This code allows an estimation of forces generated by the rotor. The second code, FAST, is developed by NREL, National Renewable Energy Laboratory, Colorado, and is one of the most used software for aeroelastic wind turbine modeling. In FAST, it is possible to obtain in output information concerning power, forces, moments, torque, accelerations and deformations. This software is frequently used to simulate large size wind turbines: by this point of view, one of the purposes of this study is to investigate the reliability of the FAST environment for small wind turbine simulation too. Actually, the critical point is that small wind turbines are strongly affected by fatigue, as a result of their size and the variability of loads , induced by the unsteady wind conditions (especially in urban environment [16]), and modulated by a very high rotational speed [17]. It is therefore interesting to understand the capability of simplified numerical models in reproducing reliably the dynamical behaviour of this kind of devices, especially in yawed conditions.

This paper is organized in the following sections: Section 2 presents the methods and facilities and a discussion on the equipment used. In section 3 the results are presented and examined. Finally Section 4 is devoted to conclusions and future developments.

2 Experimental Set Up and Numerical Models

In this study, experimental tests in wind tunnel and numerical tools are used to characterize the behaviour of a small wind turbine in yawed configuration.

2.1 Wind Turbine and Wind Tunnel for Experimental Tests

The HAWT prototype selected for this work has these main features:

- 40 kg nacelle mass;
- rotor diameter: 2 meters;
- hub height: 1.2 meters;
- hub radius is 0.13 meters;
- minimum chord of the profile: 5 cm. Maximum: 15 cm;
- angle of attack variable between 1.7° and 32°;

- the prototype is equipped with three polymer reinforced with fiberglass blades;
- fixed pitch angle;
- operative rotational speed between 200 and 700 RPM;
- 3 kW of maximum power;
- electric control based on experimental optimal power curve.

In Fig. 1 the test case wind turbine placed inside the wind tunnel is represented; the configuration in the Figure is at 0° of yaw angle but many tests have been performed with yaw angles of up to ±45°.
The wind tunnel used for this research is located at the Department of Engineering at the University of Perugia, Italy (www.windtunnel.unipg.it). The facility consists on a closed loop, open test chamber wind tunnel with a squared cross section of the ducts of 2.2 m. per side. The recovery section is about 2.7 m. x 2.7 m. A 375 kW electric motor puts in rotation a fan that is able to produce variable wind speed in the test section up to 45 m/s. A peculiar characteristic of this tunnel is the extremely low turbulence of the air flow that can be quantified in 0.4%. The wind speed is measured by a Pitot tube and a cup anemometer placed at the inlet section. In Figure 2 a scheme of the wind tunnel is reported.

Figure 1: The small HAWT in the wind tunnel open test section.

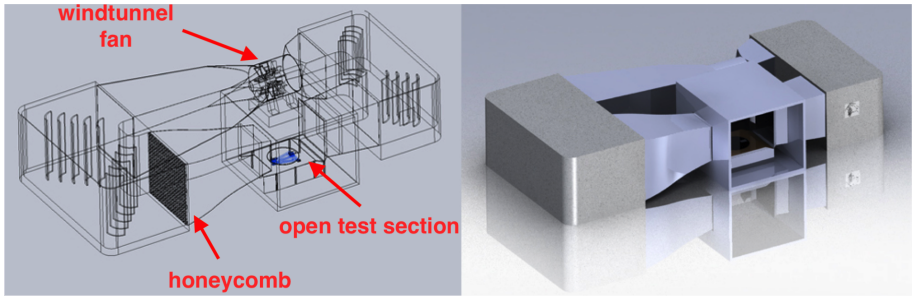


Figure 2: A sketch of the wind tunnel.

As the test section of the wind tunnel is not a free field, it has to be considered a confined environment where the airflow gets modified by the presence of the turbine itself: this phenomenon is called blockage. To consider the blockage, in this discussion wind velocities and thrust or power coefficients will be scaled by a corrective factor, BF (Blockage Factor), estimated following Kinsey and Dumas [18] as in eq.1:

$$BF = \frac{U}{U'}, \tag{1}$$

where U is the free stream wind speed in the wind tunnel with the rotor and U' without the presence of the rotor. Using the blockage factor, it is possible to correct both power and thrust coefficients as expressed by eq.2 and eq. 3:

$$C'_P = C_P \cdot \left(\frac{U}{U'}\right)^3 = C_P \cdot BF^3 \tag{2}$$

$$C'_T = C_T \cdot \left(\frac{U}{U'}\right)^2 = C_T \cdot BF^2, \tag{3}$$

where C'_P and C'_T are the corrected power and thrust coefficient. Previous experimental and numerical studies, in particular Eltayesh et al. [19], have been devoted to the analysis of the blockage factor of the wind tunnel of University of Perugia and the results have been employed for the purpose of this study to correctly estimate the reference free wind speed. According to this, to reliably compare numerical and experimental tests, it should be intended that the C_P and C_T factors obtained from simulations are the corrected ones.
The HAWT has been subjected to steady wind time series having duration of 60 s. During each time series, the yaw angle has a fixed value. The tested yaw angles are:

- 0°;
- $\pm 22.5^\circ$
- $\pm 45^\circ$.

The selected wind intensity is 10 m/s and some tests have been performed at 8 m/s too.

2.2 The FAST Software

FAST (Fatigue, Aerodynamics, Structures,and Turbulence) is an open-source aeroelastic software developed by NREL (National Renewable Energy Laboratory) and it is used to perform simulations of energetic and mechanical behaviour of horizontal axis wind turbines.
This software offers many alternatives to customize the modeling of turbine components. Electric generator, yaw controller, pitch controller and shaft brake can be modeled in many ways; the most used includes the use of subroutines, look up tables and the interface with external software environments. The number of input files depends on the characteristics of the simulation. In this test the employed input files are:

- Primary: is the main file where simulation parameters can be setted and contains the link to the other files.
- InflowWind: this file describes the wind characteristics. Data about wind speed magnitude,vertical and horizontal components has to be implemented in this file. In addition, it contains the spatial discretization resolution.
- AeroDyn: it includes environment air condition, links to the table of blade airfoils polars, and tower aerodynamic properties.
- ElastoDyn: in this file, the wind turbine mechanical design (pre-cone, tilt angle, masses and inertia) is described. Links to blades and tower shape modes are also included.
- ServoDyn: it manages the behavior of the controllers.Through this file it is possible to implement generator, pitch, yaw and braking models.

Figure 3 is a block chart representing how the files interact during the simulation running.

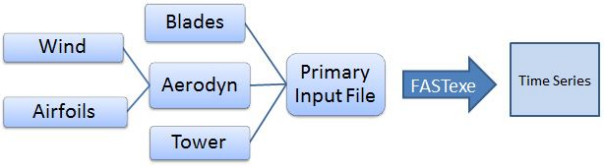


Figure 3: A flowchart of the FAST simulation.

As usual for small wind turbines, the model studied in this paper does not have an active pitch or yaw control, as discussed for example in Scappaticci [20]. To meet market requests and in consideration of the lack of adequate spaces to house actuators, small wind turbines are typically not equipped with advanced control systems. For this reason, in ServoDyn, only the electric generator is modeled. FAST offers many solutions to set up the simulation of generator, in this case, a look up table is considered the best solution. The of external software (like Simulink) is a better choice when PID (Proportional, Integrative, Derivative) controllers have to be implemented, especially for unsteady simulations. In the present paper, instead, simulations are always performed in steady conditions. Neither the default generator model, present in FAST, can be profitably used because it is arranged for large wind turbines. According to this, the choice of look-up table is the most suited among all the possibilities.
Look-up table creates a relationship between the instant rotational speed of the shaft and the resistant torque that has to be applied. Electric generator response is tested experimentally in wind tunnel steady runs with variable wind speed. In this way, once the system reaches the equilibrium, shaft speed and the corresponding torque is collected and than used to create the look up table. It has to be noticed that turbine power controller works according to MPPT (Maximum Power Point Tracking) in order to always find the best performance in terms of power production. FAST allows to impose and keep fixed the the yaw angle in ElastoDyn file. The tested yaw angles are the same as the experimental ones: 0° , $\pm 22.5^\circ$ and $\pm 45^\circ$.

2.3 The BEM Algorithm

The second numerical framework used to estimate mechanical loads on wind turbine is internally developed on the grounds of the BEM (Blade Element Momentum) theory. Many handbooks on wind turbine aerodynamics explain this mathematical approach; in the following, we refer to a summary by Burton[21]. Drag and lift coefficients are defined in eq.4 and eq.5:

$$C_{lift} = 2\frac{L}{\rho A_{ref} U_\infty^2} \tag{4}$$

$$C_{drag} = 2\frac{D}{\rho A_{ref} U_\infty^2}, \tag{5}$$

where L and D are the aerodynamic lift and drag forces; ρ the air density; A_{ref} the area of wind turbine rotor and U_∞ the free stream wind speed. Moreover, labeled as U_d the wind speed at the disk, it is possible to introduce the axial induction factor a', eq.6:

$$a = \frac{U_\infty - U_d}{U_\infty} \tag{6}$$

To keep in account the rotation effect that the disk imparts to the downstream flow, the a' coefficient is introduced, eq.7

$$a' = \frac{1}{2\Omega}\omega, \tag{7}$$

labeled as ω the angular velocity of the wake imparted by the rotor, whose velocity is Ω. Using the a factor, it is possible to rewrite the axial and tangential speeds as eq.:

$$V_x = U_\infty(1-a) \qquad\qquad V_y = \Omega R(1+a). \tag{8}$$

After the calculation of speed components, the angle of attack(ϕ) on each section of the blades is obtained using polar charts available from aerodynamic simulation software (i.e. Xfoil). Knowing ϕ, the C_x and C_y coefficients (eq. 9) can be computed:

$$C_x = C_l \cos(\phi) + C_d \sin(\phi) \qquad C_y = C_l \sin(\phi) + C_d \cos(\phi). \tag{9}$$

As stated by Ning[22], it is possible to obtain tip and loss coefficients using eq.10 and eq.11:

$$f_{tip} = \frac{B}{2}\frac{(R-r)}{|sin\phi|} \qquad F_{tip} = \frac{2}{\pi} acos(e^{-f_{tip}}) \tag{10}$$

$$f_{hub} = \frac{B}{2}\frac{(r-R_{hub})}{R_{hub}|sin\phi|} \qquad F_{hub} = \frac{2}{\pi} acos(e^{-f_{hub}}), \tag{11}$$

where:

- F_{tip}: tip loss correction;
- B: blade number;
- R: rotor radius;
- r: distance from center of the rotor to root blade section;
- F_{hub}: hub loss correction.

Introducing the solidity σ as eq.12:

$$\sigma = \frac{Bc}{2\pi r}, \tag{12}$$

with c representing the chord length, then one can write eq13:

$$k = \frac{\sigma C_x}{4\sin\phi\sin\phi F} \qquad k' = \frac{\sigma C_x}{4\sin\phi\cos\phi F}, \tag{13}$$

considering $F = f_{tip}^2$. Many formulations of a are available according to the values of ϕ and k; in particular for $\phi > 0$ and $k > 2/3$ equations 14 and 15 can be used:

$$\gamma_1 = 2Fk - (\frac{10}{9} - F) \qquad \gamma_2 = 2Fk - F(\frac{4}{3} - F) \qquad \gamma_2 = 2Fk - (\frac{25}{9} - 2F) \tag{14}$$

$$a = \frac{\gamma_1 - \sqrt{\gamma_2}}{\gamma_3}. \tag{15}$$

If $\phi < 0$ and $k > 1$, the axial induction factor has the following formulation:

$$a = \frac{k}{k-1}. \tag{16}$$

Instead, if $\phi > 0$ and $k < 2/3$, one obtains

$$a = \frac{k}{k+1} \tag{17}$$

For a', a unique formulation is obtained:

$$a' = \frac{k'}{k'+1} \tag{18}$$

From the aforementioned equations, induction flow factors can be estimated in each solution region. When a turbine is yawed, otherwise, it is necessary to consider additional corrective factors: in this case the induction factor with yaw correction is eq.19:

$$a_{yaw} = a(1 + K\frac{r}{R}\sin(\psi)), \tag{19}$$

where ψ is the azimuth angle and K, from Shen[23], is given in eq20:

$$K = \frac{15}{32}\pi \tan\frac{\chi}{2}. \tag{20}$$

χ is known as skew angle and it is obtained from eq.21:

$$\chi = (0.6a + 1)\gamma \tag{21}$$

In literature, different formulations for correcting the induction factor considering yaw are available [24]. For example Coleman[25] proposed eq.22:

$$K = \tan(\frac{\chi}{2}). \tag{22}$$

For White and Blake [26] one has eq.23:

$$K = \sqrt{2}\tan(\chi). \tag{23}$$

Moreover Shen[23] proposed eq.24:

$$a_{yaw} = a\left[1 + \frac{15\pi}{32}\sqrt{\frac{1-cos\gamma}{1+cos\gamma}}\frac{r}{R}Ksin\psi\right]. \tag{24}$$

Different, additional formulations have been proposed by Ackermann [27] and Bianchi [28]. All these formulations have been compared but noticeable differences that may cause substantial changes to the algorithm have not been found.

3 Results

3.1 Analysis of Power and Thrust Coefficients

In this section, results from simulation codes and experimental measurements are shown. Figure 4 compares the measured and simulated C_p values in different yaw configurations. C_p is computed using eq. 25:

$$C_P = \frac{P}{\frac{1}{2}\rho A_{ref} U_\infty^3}, \tag{25}$$

where P is the generator power. Physically C_p measures the ratio between the power that is produced and the kinetic energy of the flow. The maximum theoretical limit of this coefficient is indicated by the Betz law [29].

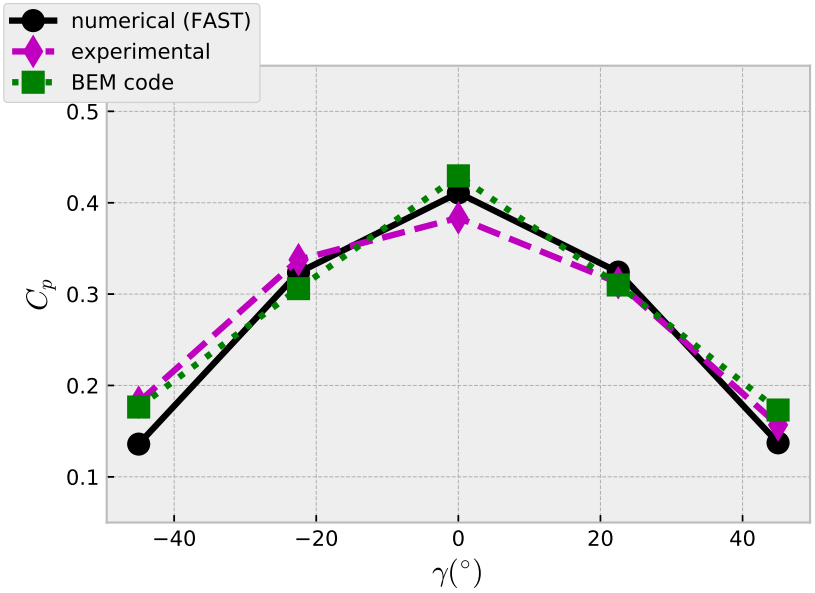


Figure 4: Power coefficients at 10 m/s: experimental vs numerical results.

Similarly to power coefficient, thrust coefficient is used to characterize turbine behavior. Its definition is given in eq.26:

$$C_T = \frac{F}{\frac{1}{2}\rho A U_\infty^2} = 4a(1-a), \tag{26}$$

where F is the thrust force acting on the rotor in the flow direction. In figure 5, the measured and simulated behavior of C_T is shown.

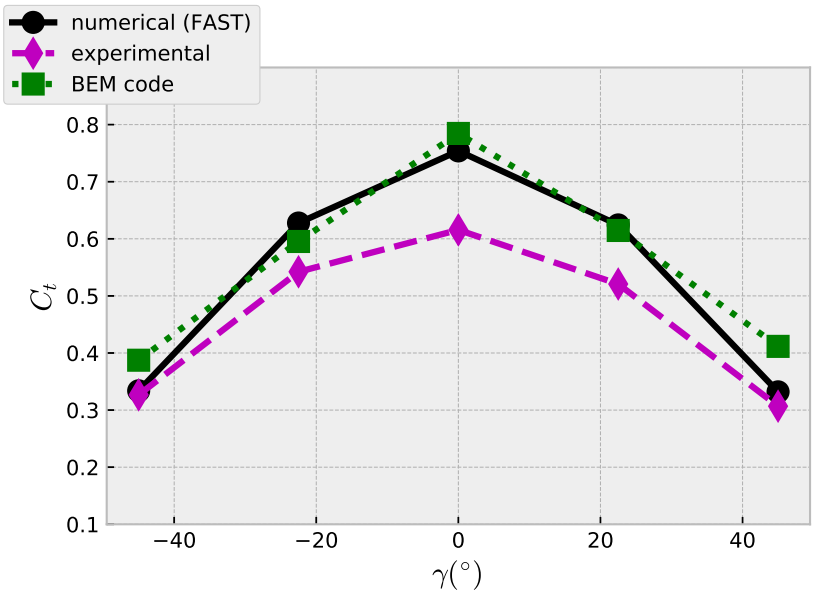


Figure 5: Thrust coefficients at 10 m/s: experimental vs numerical results.

It can be seen that both numerical models reproduce the power coefficient fairly but they largely overestimate the thrust one. The maximum percentage error for C_P is about 5% for the BEM code and 8% for FAST. In spite of this, for C_T errors are up to 25% for BEM and 20% for FAST. In numerical simulations a 95% of generator efficiency has been considered and the small errors on C_P for vanishing yaw angles shows that it can be considered a reliable estimation.

The mismatch between measured and simulated C_T coefficient can be imputable to multiple causes: the most important can be supposed to be the fact that the numerical models do not take into account blades deformations. In section 3.2, it will be discussed how the combined effects of blade deformation and tower blockage are linked to yaw configuration. In fact, the aerodynamic thrust generation depends on the distance between blade and tower in stream direction. When the turbine is yawed, this distance tends to be increased and the blades are affected by a lower tower blockage effect producing more thrust: in this case, the mismatch as regards the C_T, where blockage is not implemented, is negligible.

The slight asymmetry, visible in experimental tests, can be related to wind tunnel layout. In the open test chamber the lateral walls are placed at different distances respect the air stream and the turbine rotor. Because of this, the flow that impacts the rotor at negative or positive yaw angles is slightly different. Anyway, the discrepancies are estimated to be 3% on C_P and 8% on C_T.

3.2 Study of Thrust Cyclic Variation

Because of the critical issues revealed by the previous analysis, it has been considered useful to set up a study devoted to the cyclic variations of the aerodynamic forces during a complete rotation of a blade. The reference angle is denoted as azimuth. From Figure 6, it arises that there is periodic component in correspondence of the first blade passing frequency 3P. This phenomenon is well known and can be interpreted as due to the interaction between tower and airflow, causing cyclic decrease of aerodynamic forces. As previously explained, the intensity of the fluctuations is lower for increasing yaw angles because the blade passes farther from the low velocity air situated close to the tower.

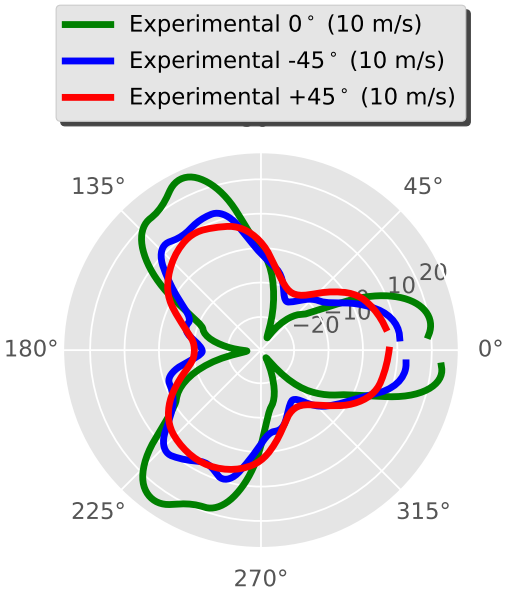


Figure 6: The rotor thrust variation(N), with respect to the average value, as a function of the azimuth angle.

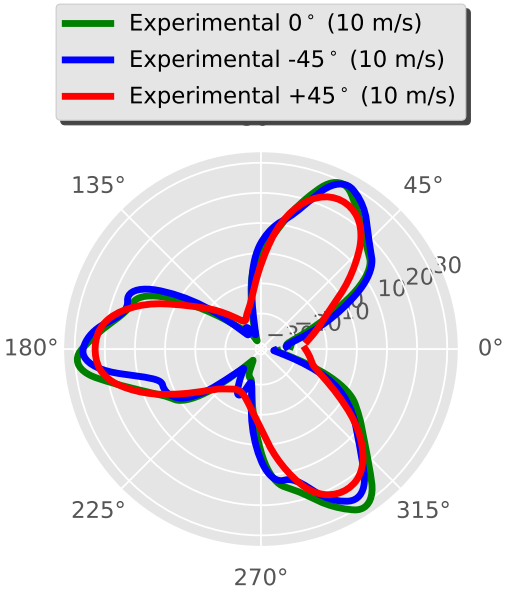


Figure 7: The rotor thrust relative variation(N), with respect to the average value, as a function of the azimuth angle.

The curves of thrust as a function of the azimuth angle have been scaled with respect to the corresponding mean value in fig. 7. It can be seen that there is a consistent overlap: this means that the amplitude of oscillations is not dependant on the yaw angle.

Additional experimental tests with a wind speed of $8m/s$ has been performed to discover the dependence of thrust oscillations with flow characteristics at 0° yaw value. The results are reported in fig. 8.

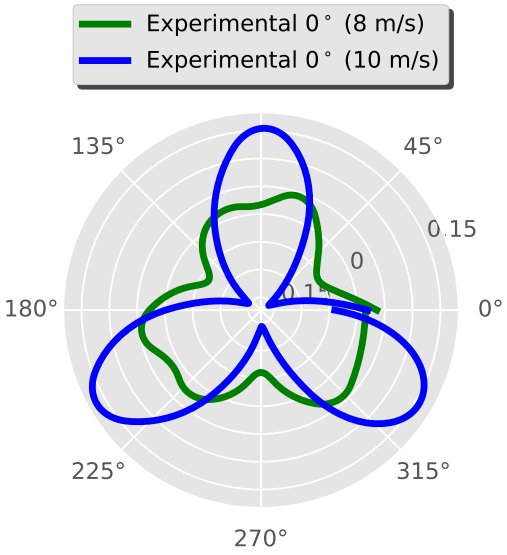


Figure 8: The rotor thrust coefficient cyclic variation in different wind tunnel tests with a wind speed of 8 m/s and 10 m/s.

From the comparison between the $10m/s$ and the $8m/s$ tests, it results that the C_T fluctuations at 10 m/s are more than doubled with respect to 8 m/s. According to this, it can be stated that the tower interference has a less relevant effect as the wind speed tends to be lower. Blade deflections in facts are strictly dependant on the aerodynamics loads and, since they decrease when the wind speed decreases, the space between the deflected blade and the tower increases and therefore the thrust is less affected by blockage. To quantify the blade deflection, it has to be considered that in previous measurements campaigns the blade deflection at the tip has been measured to be around 7% with a wind speed of $32m/s$, so it is expected that at 10 m/s it is of the order of 1% of the rotor radius: it is remarkable that this small value can induce such large tower interaction effects.

3.3 Analysis of Tower Interference on Accelerations

The above results indicate that the tower inference is a non-negligible aspect of the dynamical behavior of the small scale wind turbine considered in this study. Wind tunnel tests have been useful to deeply understand how tower interference effect gets modified by the yawing the turbine: this has been possible thanks to the triaxial accelerometer located on the nacelle, recording the aerodynamic induced vibrations. Fast Fourier Transform (FFT) theory has been used to analyze the spectrum of vibrations and making a comparison between $0°$ yaw and $+45°$, fig.9.

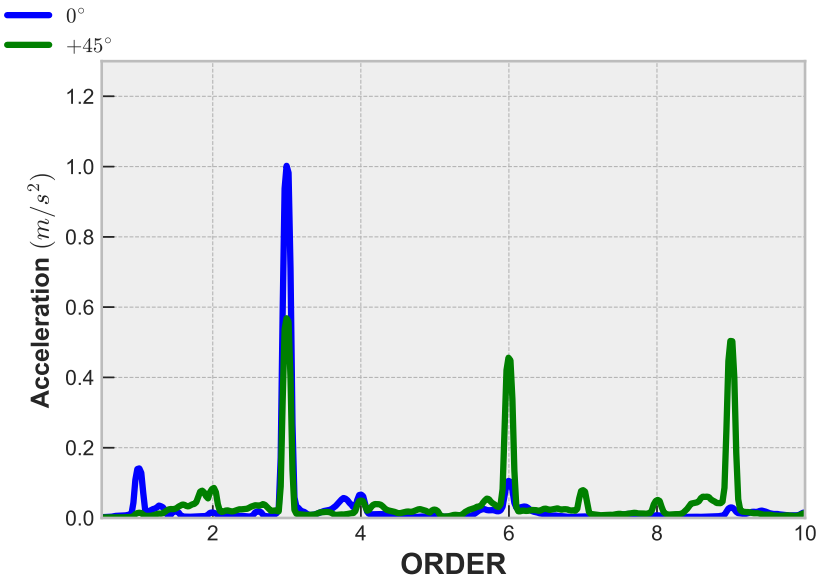


Figure 9: Experimental normalized order spectrum of the acceleration (fore-aft component normalized on the amplitude of the 3P component with zero yaw).

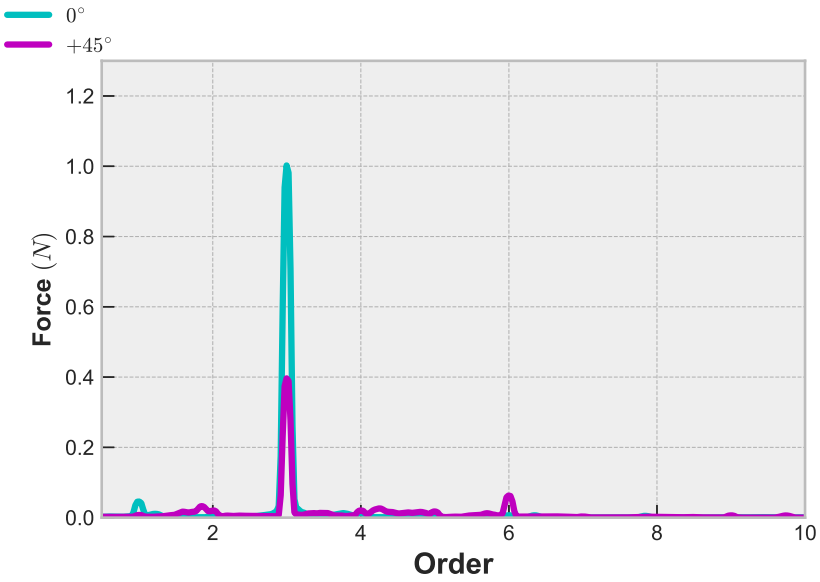


Figure 10: Experimental normalized order spectrum of the forces (fore-aft component normalized on the amplitude of the 3P component with zero yaw).

Figure 10 is the FFT of the thrust force measured by load cell placed between tower top and nacelle. Spectral analysis shows that order 3P, related to tower blockage, undergoes a substantial decrease passing from 0° yaw to 45°. This is an additional proof that gives consistency to the thesis that tower inference is more prominent for vanishing yaw angle and that justifies the thrust overestimation obtained with numerical codes, where it is not possible to account for tower blockage.

4 Conclusion

The objective of this study was the characterization of the mechanical behavior of horizontal axis wind turbines in yaw configuration. This field of study is attracting the scientific interest because yawing turbines allows several types of wind farm cooperative control, as for example the wake steering, that is useful to

optimize the energy production.

A small scale horizontal axis wind turbine, with 2 m. of rotor diameter has been tested at the wind tunnel facility of the University of Perugia. The prototype has been equipped with accelerometers, load cell, electric power meter and tachometer in order to collect information about its operative conditions when undergoing different yaw angles. Experimental tests have been performed with a wind speed of 10 m/s and $\pm 45°$,$\pm 22.5°$ and $0°$ of yaw angle; additional tests have been carried out with 8 m/s of wind speed at $0°$ yaw. In addition, two numerical models have been adopted with two different software: an internally developed BEM algorithm and the open source FAST code. The numerical models have been set up in order to reproduce the conditions of the experimental test: the results are then compared in therms of power coefficient C_P and thrust coefficient C_T.

The main result is that the numerical models fairly reproduce the C_P coefficient. There are more critical points as regards the thrust coefficient: whereas for the cases of yawed configurations, the simulations fairly agreed with experimental test, for vanishing yaw angle the discrepancy is remarkable. This fact has motivated further analysis of the experimental data and the interpretation is that the mismatch between simulation and measurements is given by the fact that the tower blockage is a relevant phenomenon that the numerical models employed in this work do not take into account. The tower blockage is related to the streamwise distance between blades and tower, and so it is related to blade deflection too.

The cyclic variation of the thrust as a function of the azimuth angle has been analyzed for different yaw configurations, confirming the presence of a $3P$ periodicity which testifies the presence of a tower induced blockage effect variable respect to yaw angle. Under this circumstance an experimental test, with a wind speed of 8 m/s has been useful to confirm that with lower aerodynamic loads also the blade deflections decrease and so they do the lower thrust oscillations.

In addition, FFT analysis has been used to compare order spectra of nacelle accelerations and thrust for the $0°$ and $45°$ yaw configurations. The accelerations and thrust at 3P order appear diminished when the rotor is yawed. As this order is characteristic of tower interference effect, the order analysis brings an additional argument is support of the fact that blockage phenomenon is strictly correlated to the yawing behavior and cannot be neglected when yaw angle tends to vanish.

The results of this study can be useful to increase the knowledge of the behaviour of small wind turbines in yawed configuration and to evaluate the ability of low-fidelity numerical models predicting loads on yawed rotors. Future improvements of this study regard the possibility of better characterizing the interactions between blade tip and turbine tower, possibly using CFD codes. This study can also be useful for the implementation of wake steering wind farm control where it is expected that turbines will runs in yawed configuration for long periods and so an accurate estimation of loads is a crucial step to guarantee best performances and to assess the fatigue loading of the machine.

References

[1] Torben Knudsen, Thomas Bak, and Mikael Svenstrup. Survey of wind farm controlâpower and fatigue optimization. *Wind Energy*, 18(8):1333–1351, 2015.

[2] Jackson G Njiri and Dirk Soeffker. State-of-the-art in wind turbine control: Trends and challenges. *Renewable and Sustainable Energy Reviews*, 60:377–393, 2016.

[3] Richie Gao and Zhiwei Gao. Pitch control for wind turbine systems using optimization, estimation and compensation. *Renewable Energy*, 91:501–515, 2016.

[4] Bing Han, Lawu Zhou, Fan Yang, and Zeng Xiang. Individual pitch controller based on fuzzy logic control for wind turbine load mitigation. *IET Renewable Power Generation*, 10(5):687–693, 2016.

[5] Dongran Song, Xinyu Fan, Jian Yang, Anfeng Liu, Sifan Chen, and Young Hoon Joo. Power extraction efficiency optimization of horizontal-axis wind turbines through optimizing control parameters of yaw control systems using an intelligent method. *Applied energy*, 224:267–279, 2018.

[6] Aitor Saenz-Aguirre, Ekaitz Zulueta, Unai Fernandez-Gamiz, Javier Lozano, and Jose Manuel Lopez-Guede. Artificial neural network based reinforcement learning for wind turbine yaw control. *Energies*, 12(3):436, 2019.

[7] Rebecca Jane Barthelmie and LE Jensen. Evaluation of wind farm efficiency and wind turbine wakes at the nysted offshore wind farm. *Wind Energy*, 13(6):573–586, 2010.

[8] Soo-Hyun Kim, Hyung-Ki Shin, Young-Chul Joo, and Keon-Hoon Kim. A study of the wake effects on the wind characteristics and fatigue loads for the turbines in a wind farm. *Renewable Energy*, 74:536–543, 2015.

[9] Francisco Haces-Fernandez, Hua Li, and David Ramirez. Improving wind farm power output through deactivating selected wind turbines. *Energy Conversion and Management*, 187:407 – 422, 2019.

[10] Mikel De-Prada-Gil, César Guillén AlÃas, Oriol Gomis-Bellmunt, and Andreas Sumper. Maximum wind power plant generation by reducing the wake effect. *Energy Conversion and Management*, 101:73 – 84, 2015.

[11] Paul A. Fleming, Pieter M.O. Gebraad, Sang Lee, Jan-Willem van Wingerden, Kathryn Johnson, Matt Churchfield, John Michalakes, Philippe Spalart, and Patrick Moriarty. Evaluating techniques for redirecting turbine wakes using sowfa. *Renewable Energy*, 70:211 – 218, 2014. Special issue on aerodynamics of offshore wind energy systems and wakes.

[12] Cristina L. Archer and Ahmad Vasel-Be-Hagh. Wake steering via yaw control in multi-turbine wind farms: Recommendations based on large-eddy simulation. *Sustainable Energy Technologies and Assessments*, 33:34 – 43, 2019.

[13] Paul Fleming, Jennifer R King, Katherine L Dykes, Eric J Simley, Jason M Roadman, Andrew K Scholbrock, Patrick Murphy, Julie K Lundquist, Patrick J Moriarty, Katherine A Fleming, et al. Initial results from a field campaign of wake steering applied at a commercial wind farm: Part 1. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[14] Shuting Wan, Lifeng Cheng, and Xiaoling Sheng. Effects of yaw error on wind turbine running characteristics based on the equivalent wind speed model. *Energies*, 8(7):6286–6301, 2015.

[15] Roozbeh Bakhshi and Peter Sandborn. The effect of yaw error on the reliability of wind turbine blades. In *ASME 2016 10th International Conference on Energy Sustainability collocated with the ASME 2016 Power Conference and the ASME 2016 14th International Conference on Fuel Cell Science, Engineering and Technology*, pages V001T14A001–V001T14A001. American Society of Mechanical Engineers, 2016.

[16] L Battisti, E Benini, A Brighenti, S DellâAnna, and M Raciti Castelli. Small wind turbine effectiveness in the urban environment. *Renewable energy*, 129:102–113, 2018.

[17] Francesco Castellani, Davide Astolfi, Matteo Becchetti, Francesco Berno, Filippo Cianetti, and Alessandro Cetrini. Experimental and numerical vibrational analysis of a horizontal-axis micro-wind turbine. *Energies*, 11(2):456, 2018.

[18] Thomas Kinsey and Guy Dumas. Impact of channel blockage on the performance of axial and cross-flow hydrokinetic turbines. *Renewable energy*, 103:239–254, 2017.

[19] Abdelgalil Eltayesh, Massimiliano Burlando, Francesco Castellani, and Matteo Becchetti. Experimental and numerical study of the wind tunnel blockage effects on the behaviour of a horizontal axis wind turbine. In *Lecture Notes in Civil Engineering*. Springer, 2019.

[20] Lorenzo Scappatici, Nicola Bartolini, Francesco Castellani, Davide Astolfi, Alberto Garinei, and Marcello Pennicchi. Optimizing the design of horizontal-axis small wind turbines: From the laboratory to market. *Journal of Wind Engineering and Industrial Aerodynamics*, 154:58–68, 2016.

[21] Tony Burton, Nick Jenkins, David Sharpe, and Ervin Bossanyi. *Wind energy handbook*. John Wiley & Sons, 2011.

[22] Andrew Ning, Gregory Hayman, Rick Damiani, and Jason M Jonkman. Development and validation of a new blade element momentum skewed-wake model within aerodyn. In *33rd Wind Energy Symposium*, page 0215, 2015.

[23] Wen Zhong Shen, Wei Jun Zhu, and Jens Nørkær Sørensen. Actuator line/navier–stokes computations for the mexico rotor: comparison with detailed measurements. *Wind Energy*, 15(5):811–825, 2012.

[24] Daniel Micallef and Tonio Sant. A review of wind turbine yaw aerodynamics. In *Wind Turbines-Design, Control and Applications*. InTech, 2016.

[25] Robert P Coleman, Arnold M Feingold, and Carl W Stempin. Evaluation of the induced-velocity field of an idealized helicoptor rotor. Technical report, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION HAMPTON VA LANGLEY RESEARCH CENTER, 1945.

[26] Fred White and Bruce B Blake. Improved method predicting helicopter control response and gust sensitivity. American Helicopter Society, 1979.

[27] MC Ackerman. Yaw modelling of small wind turbines. *Journal of wind engineering and industrial aerodynamics*, 39(1-3):1–9, 1992.

[28] Serena Bianchi, Alessandro Bianchini, Giovanni Ferrara, and Lorenzo Ferrari. Small wind turbines in the built environment: influence of flow inclination on the potential energy yield. *Journal of Turbomachinery*, 136(4):041013, 2014.

[29] A. Betz. *Introduction to the theory of flow machines,*. Oxford Pergamon Press, 1966.

Gears and Bearings faults Detection: from Instrumentation to Classification.

Renaud Bertoni[1], Sylvain Barcet[1], Alexandre Carbonelli[1], Marion Cortes[1]

[1]VIBRATEC
28 Chemin du petit bois, 69130 ECULLY, France
renaud.bertoni@vibratec.fr

Gears and bearings, used in many industrial areas are subject to failure that may lead to costly shutdowns. The current trend is to detect failures (cracks, spall, pitting …) and to identify and control their evolution. Such monitoring leads to a huge amount of data. With a double skill in test and simulation, Vibratec proposes an approach based on measurements coupled with Machine Learning (ML) processing.

This presentation defines the fault detection global approach used by Vibratec, from signal acquisition to the classification of indicators. The methodology is firstly applied on a specific HMS test bench. Then, the machine learning strategy is deployed on a database. The numerical simulations are in good agreement with the measurement results obtained on the test bench, and the machine learning indicators provides encouraging results. In the upcoming months, this complete methodology will be applied on a collaborative project aiming to improve the maintenance of aircraft engines.

Measurement and use of transmission error for diagnostics of gears

Robert RANDALL[1], Zhan Yie CHIN[1], Wade SMITH[1], Pietro BORGHESANI[1]

[1]UNSW Sydney, School of Mech. and Manuf. Engineering, Sydney 2052, Australia
b.randall@unsw.edu.au

Abstract

Transmission error (TE) has long been thought to be a major contributor to gear vibration and noise, but insufficient consideration has been given to the different types of TE and how they generate vibrations. TE is defined as the difference in torsional vibration of two meshing gears, scaled so as to represent linear motion along the line of action. There are three distinct types of TE; 1) Geometric TE (GTE) given by deviations of the (combined) tooth profiles from ideal involute; 2) Static TE (STE) including elastic deformation of the teeth and therefore being load dependent; 3) Dynamic TE (DTE) including inertial as well as stiffness effects, and thus being speed as well as load dependent. It has long been recognized that TE can be measured very accurately by phase demodulation of the signals of shaft encoders rigidly attached to each of the gears in mesh, but only recently realized that all three types can be measured; GTE at low speed and low load, STE at low speed and higher load, and DTE at higher speed and higher load. This paper demonstrates that TE has several advantages over vibration acceleration (or even the raw torsional vibrations) as a diagnostic parameter, being close to the source (the gearmesh) and with "common mode rejection" from the two gears, thus being much less sensitive to operating conditions and rig parameters, including the much greater number of transfer paths, modulations, and resonances in the casing vibration measurements. The measurements in this paper were made on a single stage gearbox, over an input gear speed range from 2 – 20 Hz, and input shaft torque range from 0 – 20 Nm. Earlier measurements on the same gearbox were for soft gears which developed distributed pitting over an operating period of many hours. Unfortunately, the encoders used at that time (actually included in slip rings) had a low torsional resonance frequency, which precluded obtaining TE at higher than 2 Hz shaft speed, so only GTE and STE could be estimated. New results are presented here for ground, hardened gears with a simulated tooth root crack on one tooth. Not only does this illustrate the differences with a local fault, but new encoders were mounted, valid up to a shaft speed of 20 Hz, so that DTE could also be measured.

1 Introduction

Gear transmission error (TE) is defined as the difference in torsional vibration of two gears in mesh, scaled so as to represent linear motion along the line of action, this being common to the two gears. Already in 1996 [1], it was shown that TE could be measured simply and accurately by phase demodulation of the pulse signals from high quality shaft encoders on the free ends of the shafts on which the gears are mounted. The measured torsional vibrations, in terms of angular displacement, are scaled by the respective base circle radii, and subtracted to give relative motion along the line of action. The accuracy of the encoders themselves corresponds to fractions of a micron of TE, and virtually no further error is introduced by the phase demodulation processing by Hilbert transform techniques (as compared with the earlier use of analogue phase meters, or polynomial interpolation between pulses). It is often possible to mount the encoders on the free ends of the gear shafts (the section not transmitting torque) so that they follow the gear motions up to a very high frequency. The proposed application in [1] was to the measurement of TE in design, development and manufacture, to add to information gained from measurements using gear metrology machines, but it has also been proposed as a tool for gear diagnostics in [2]. However, at the time, that was limited by the necessity to mount encoders on the machines.

It is now becoming more common for encoders to be built into machines, to provide valuable information for both control and monitoring of, for example, variable speed machines such as wind turbines, and this will presumably increase with the adoption of the Internet of Things, so it is likely that measurement of TE will become more available as an indicator of gear faults.

Transmission error (TE) has long been thought to be a major contributor to gear vibration and noise, but the relationship between them has not been fully understood. For a start there are three distinct types of TE: 1) Geometric TE (GTE) given by deviations of the (combined) tooth profiles from ideal involute; 2) Static TE (STE) including elastic deformation of the teeth, and therefore being load dependent; 3) Dynamic TE (DTE) including inertial as well as stiffness effects, and thus being speed as well as load dependent.

Measurement and application of these three types of TE as a diagnostic tool were discussed in [3], but it was found that the encoders used there (actually included in slip rings) had a low resonance frequency, which precluded measurements at high enough speed to give DTE. The same test rig has now been equipped with high quality encoders, and the current paper uses new measurements with that system. Another difference is that the old measurements were made with soft gears, run for an extended period so that (uniformly distributed) pitting developed, but no distinct local faults. The current paper uses measurements made with hardened ground gears, but with a simulated tooth root crack seeded in one tooth on the pinion, to give information on local faults, and tooth root cracks in particular, this being one of the most critical faults, and most important to distinguish from less critical faults such as local spalls.

2 Test rig and measurements

The overall layout of the spur gear test rig is shown in Figure 1.

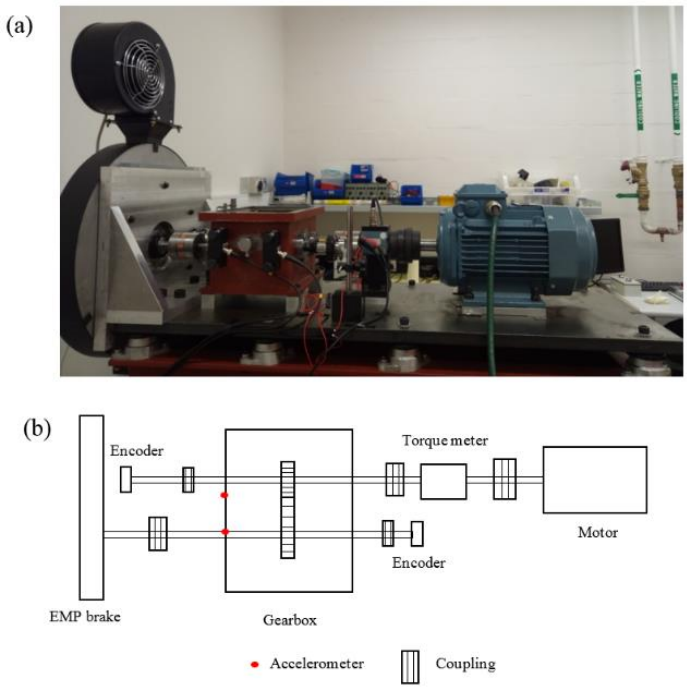


Figure 1: The spur gear test rig at UNSW. (a) photo; (b) schematic diagram

For the original measurements in [3] the reduction ratio was 19:52, and the gears were of mild steel. The original encoders were also slip rings, and had a low frequency resonance so that the highest valid input speed was 2 Hz.

For the new measurements, the reduction ratio was changed to 27:44 (same centre distance) and the gears were of hardened steel to avoid surface distress. The encoders were replaced by Heidenhain type ROD426, with 1000 pulses per rev, as well as a one per rev tacho pulse as a phase marker, and they gave valid results up to at least 20 Hz shaft speed. An EDM-generated half-tooth root crack (a 45° slot across the entire facewidth, 2.86 mm deep, extending to the tooth centreline, and 0.35mm wide) was seeded on one pinion tooth (input gear). Measurements were made at speeds 2, 5, 10, 15, 20 Hz, and loads 0, 5, 10, 20 Nm (all referenced to the input pinion). In addition to the encoder and tacho recordings, accelerometer measurements were made in the vertical direction on the casing above the input shaft at the motor end, and above both shafts at the brake end.

3 Results and discussion

3.1 Earlier results from the spur gearbox

A short summary is given here of the results published in [3], because they contain some findings which are different from those of the more recent measurements, published for the first time here. As mentioned, the gears were of mild steel and were run for a long period (nearly 50 hours) during which time they developed surface pitting fairly uniformly distributed around the gears. This was much more pronounced on the 19 tooth pinion than on the 52 tooth gear, because each tooth had a much greater number of contacts in inverse ratio to the tooth numbers, so only the pinion is discussed here.

Wear was monitored by trending the amplitude of the TE gearmesh harmonics (and the corresponding component of the synchronously averaged TE signal) in two conditions: low speed-low load (GTE) and low speed-high load (STE).

The effect of wear on GTE and STE showed an unexpected trend. The growth of the gearmesh harmonics was more pronounced on GTE during the first 6 hours of operation (mild pitting), and on STE later (severe pitting). The greater sensitivity of GTE in the initial phase was interpreted as being due to the fact that the unloaded GTE would have been dominated by (a few) local high spots at the edges of the pits, which would be easily deformed under relatively light load to give a reduced STE. On the other hand, with severe pitting more continuously distributed along the contact line, high spots would reduce the visibility of wear in GTE, and increased load would tend to give an increase in TE. Figure 2 shows a schematic representation of this interpretation, together with snapshots of the surfaces after about 2.5 and 42.5 hours of operation. For a detailed description of this test campaign the reader is referred to [4].

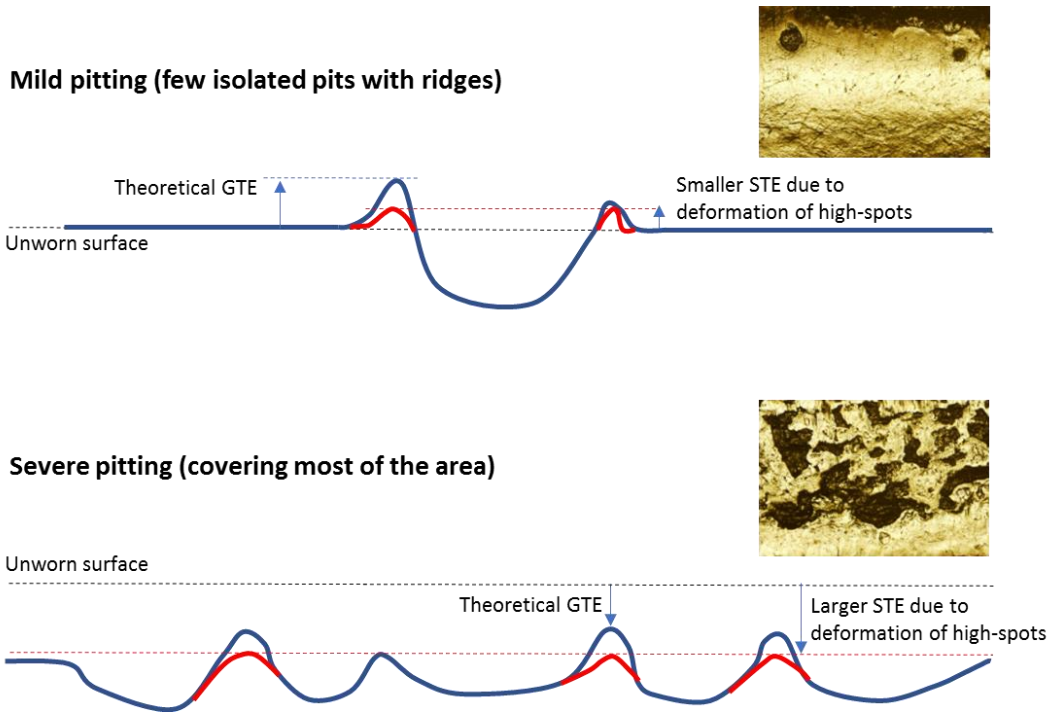


Figure 2: Schematic example of the interpretation of the effect of mild (top) and severe (bottom) pitting on GTE and STE, with corresponding example images of the gear surface.

In simulation models it is quite common to have GTE as a fixed value in series with the toothmesh stiffness. The latter is not always constant, but any nonlinearity is usually taken to correspond just to the extra compliance of the Hertzian component at low load, which still does not give a large difference in the overall stiffness, since the Hertzian component typically only represents about 25% of the total compliance, with the dominant bending stiffness component being almost linear. The above experience with “high points” does seem to indicate that, to obtain a reasonable match between such a simplified model and

experiment, it would be better to use a value of GTE measured at a low, but non-zero, load sufficient to negate the effect of the high spots, and giving a more sudden transition to the Hertzian affected section of the stiffness curve.

Another interesting finding from the same study showed that, differently from TE, vibration was almost entirely insensitive to wear in both unloaded and loaded cases, at low speed. This was attributed to the fact that the proportion of the STE due to tooth deflection is still relatively small, but in fact it is only the dynamic tooth load, giving this deflection, which gives rise to vibration. At low speed there is no inertial resistance to rotation, so the driven gear can simply absorb the GTE by relative torsional motion, with almost no change in the GM spring force, even for the loaded case where the static load is almost constant. It could be expected that for DTE the much greater angular accelerations involved might prevent the driven gear from simply "moving out of the way" and thus force tooth deflection and increased vibration. This was actually found in [3] for the higher harmonics of gearmesh. Unfortunately, the encoders mounted at the time of this first test had a low resonance preventing reliable measurements of TE at speeds higher than 2 Hz (i.e. DTE) and their comparison with the vibration.

3.2 New results from the spur gearbox

As mentioned above, the new measurements were for a different gear ratio, and the gears were hardened and ground, to mitigate against surface distress. Moreover, they were reduced in face-width from 20 mm to 5 mm to reduce the gearmesh stiffness proportionately. The tests are to check the effects of the simulated half tooth-root crack described in section 2. It should be noted that the gearbox test rig is non-ideal (and non-typical) because the shafts are relatively long and slender (to give access inside the casing), but this means that the TE tends to be dominated by shaft deflections rather than tooth defections, making it difficult to detect changes in tooth stiffness, such as result from a crack. The tooth stiffness is at least an order of magnitude greater than the shaft stiffness. Both TE and vibration acceleration were measured over a range of speeds and loads, but speeds of 2 Hz and 20 Hz, and loads from zero (nominal) to 20 Nm are presented here. There was a small friction load corresponding to nominal zero, which was sufficient to keep the gears in contact, and allow measurement of the GTE at low speed.

Figure 3 shows the measured TE, synchronously averaged with respect to the pinion, for loads of 0, 5, 10, 15 and 20 Nm, for four different conditions:

1) Original TSA at 2 Hz
2) Original TSA at 20 Hz
3) Filtered TSA at 2 Hz
4) Filtered TSA at 20 Hz

Two (identical) rotational periods are shown. Bandpass filtering was performed to remove the masking effect of the gearmesh (GM) components and the first two harmonics of the input shaft speed, and so shaft harmonics from the 3rd to the 13th were retained in the TE signals. It was checked that the main effect of the crack was additive rather than multiplicative (modulation of the GM harmonics) so the signals were lowpass filtered just under half the GM frequency to enhance additive impulses from the crack, having components above the first two rotational harmonics, but removing modulation sidebands along with the GM harmonics.

Considering first the unfiltered results at low speed in Fig. 3(a), the increasing load gives a corresponding increase in the gearmesh component, but no change in a shaft speed component, which is likely due to a small eccentricity of the pinion. The TE for zero load could be taken as the GTE for this gear. The increasing GM component with load corresponds to the static deflection component of the STE.

For the equivalent results at 20 Hz, in Fig. 3(b), it is seen that the DTE is substantially different from the STE, at least with respect to the GM component. This can be explained by the fact that the GM frequency (540 Hz) is very close to a resonance of the system. This interpretation is also consistent with the fact that the increased GM component is dominated by the first harmonic, whereas that in Fig. 3(a) has many GM harmonics.

The filtered low speed results in Fig. 3(c) reveal the effect of the crack, at about 50 degrees along the scale, although the effect becomes less evident with increasing load. With this knowledge, it will be seen that the crack can also be detected in the unfiltered signal in 3(a), though only at the lowest load.

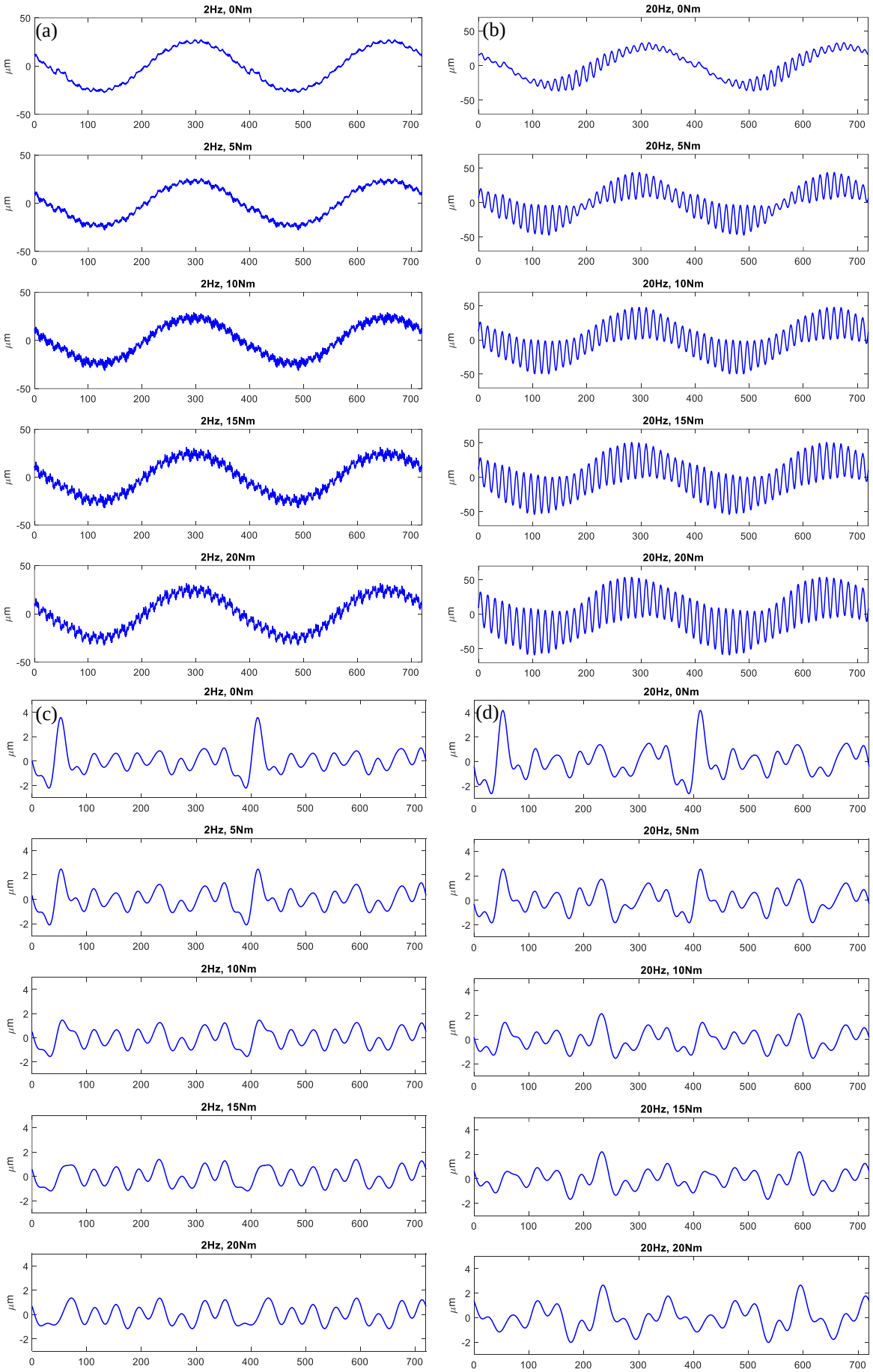


Figure 3: Comparison of original and filtered TE measurements
(a) Original TE, 2 Hz (b) Original TE, 20 Hz (c) Filtered TE, 2 Hz (d) Filtered TE, 20 Hz

The situation is very similar for the high-speed results in Fig. 3(d) (and 3(b)), and it is quite remarkable that once the effect of the resonance on the GM component is removed, the STE of Fig. 3(c) and DTE of 3(d) are very similar, at least for the lowest two loads. This illustrates one of the advantages of TE rather than vibration (including torsional vibration) as a diagnostic parameter, since the effects of operating conditions are greatly reduced.

The unexpected reduction in TE with increase in load gave rise to speculation as to the cause, and it was realised that it must be due to the fact that the "crack" has actually started slightly closed with respect to the undamaged gear, and the effect of increasing load is to counteract this with increasing tooth deflection under load. This is the opposite to what is expected to happen in the case of a genuine natural crack, where it has been demonstrated [5] that there is a tendency for the crack to be permanently open, in the unloaded condition, because of the plastic deformation at the crack tip which is an intrinsic part of crack development. The reason for the "crack" closure in this case is undoubtedly because of relief of residual stresses from heat treatment when the slot was machined, but this should never occur with real crack development, where STE due to loading would be in the same direction as the original GTE.

The change in TE as a result of tooth deflection is not easy to see, even from the filtered results in Fig. 3(c) and (d), but Figure 4(a) and (b) show a zoom of the differential TE in the vicinity of the crack. This represents the difference with respect to the curve at the highest load (20 Nm), but with reversed sign so as to show the increase of deflection with load. This is seen to be monotonic and close to linear. The corresponding linearised compliance can be derived from the deflection vs load curves in Fig. 4(c, d). These differ by only 33%, and indicate that it may be possible to estimate gearmesh stiffness from DTE as well as STE, even where measurements cannot be made at low speed.

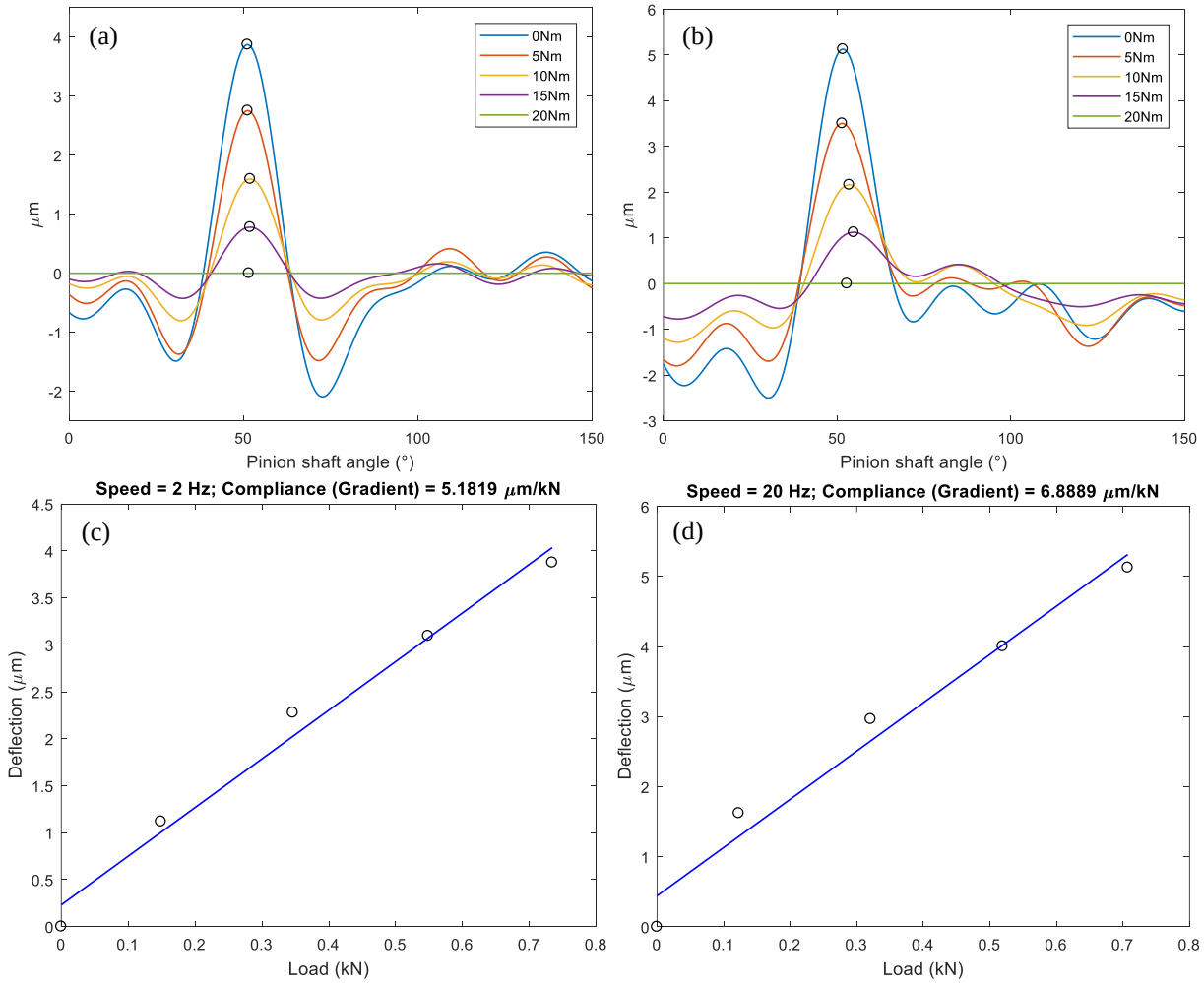


Figure 4: (a, b) Zoom on differential TE in vicinity of crack (c, d) corresponding compliance curves
(a, c) 2 Hz shaft speed (b, d) 20 Hz shaft speed

It is interesting to compare the (differential) compliance values in Fig. 4 with the typical value given for total stiffness by Smith in [6] as "A generally accepted figure for the mesh stiffness of normal teeth is 1.4 $\times 10^{10}$ N/m/m", which works out in this case to be 7×10^7 N/m, or 14 μm/kN in terms of compliance. This constant value (per unit facewidth) is based only on the bending stiffness component, and is independent of scale for a given shape of tooth since the stiffness varies directly with the cube of the depth, and inversely with the cube of the length. The values in Fig. 4 represent the differential compliance (additional deflection for the same load), which would be 5.18 and 6.89 μm/kN, respectively. In Ref. [7], an estimate is made of the change in stiffness of the toothmesh due to cracks of various sizes, using FEM and an improved simplified method, which agree. For their largest crack, which extends to 48.4% of the tooth thickness, and which has a sharp tip, the increase in compliance is 33% in the single tooth pair zone and 25% in the double tooth pair zone. Considering that the "crack" in the current results has a depth of 50%, and is actually a slot, it is likely the increase in compliance would be greater than those from [7], giving good agreement with the results from Fig. 4.

It is interesting to compare these TE results with those from response accelerations. Figure 5 shows synchronously averaged signals (over two rotation periods) at zero and 20 Nm load, and 2 and 20 Hz input shaft speed. Only the response at highest speed and highest load shows the tooth root crack. Although not shown here, even the responses at 20 Hz and 15 Nm did not show the crack. From Fig. 5(d) it appears that the effect of the crack is mainly multiplicative (local amplitude modulation) so it could be that the resonance near the GM frequency has also amplified the effect of the crack.

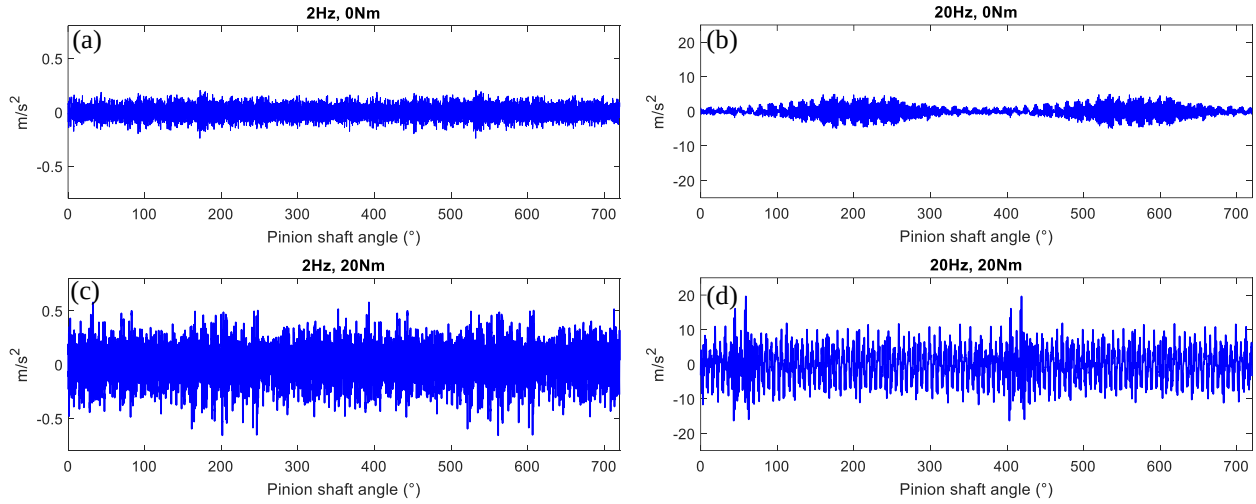


Figure 5: Synchronously averaged acceleration signals for two speeds and two loads
(a, b) 0 Nm (c, d) 20 Nm (a, c) 2 Hz (b, d) 20 Hz

It is quite possible that further signal processing could extract evidence of the crack from more of the response signals, but the main point with respect to this paper is that the TE and vibration responses give quite different information about a tooth root crack, with perhaps the main point being that it only excites a vibration response when teeth are deflected, and therefore not under zero load. The GTE, on the other hand, does show the crack at zero load, in this case because the "slot" had actually closed because of relief of residual stresses. However, in the case of normally developing cracks, they would be partially open because of plastic deformation at the crack tip, and would open further under load, this being detectable by measurement of STE and DTE, the latter at higher speeds, where it would not be possible to measure the GTE.

The fact that information was obtainable, from the measured TE, of toothmesh stiffness, even at higher speed where the GM frequency excited a resonance, emphasises the fact that the TE is measured right at the source, whereas vibration response measurements at different measurement points would all be different, and correspond to different (possibly time-varying) transmission paths.

4 Conclusion

This paper gives a number of examples of how measured gear TE can be useful in gear diagnostics, as an alternative, or supplement, to vibration measurements. It explains how GTE, STE and DTE can be measured if it is possible to run the machine at low speed and low load (GTE), low speed and high load (STE) and high

speed and high load (DTE). An earlier paper demonstrated some of the characteristics for generalised distributed wear and pitting of the teeth, giving changes on tooth profiles, whereas the current paper shows a number of advantages, compared with vibration measurement, for the critical case of a tooth root crack. Of particular interest was that it was possible to obtain estimates of the change in toothmesh stiffness (actually compliance) due to the crack, and indirectly of the toothmesh stiffness itself. The latter would probably require comparison with simulations of the cracked tooth, for example with an FE model.

Potential advantages of using TE for gear diagnostics are:

1) The measurement is closer to the source, and less disturbed by transfer function effects than vibration responses, which not only vary considerably between different positions, but can also be time-varying.

2) It is easier to get a good correspondence with simulations, because the torsional parts of simulated systems are simpler, and affected by fewer resonances than lateral vibrations, so model updating should be simpler.

3) The measurement of GTE at different times during the life of a gearbox, as well as giving a more direct measurement of wear, will make possible the inclusion of more accurate versions of this parameter in simulation models, including those giving lateral vibrations as outputs.

The technique does require the mounting of accurate encoders on at least the input and output shafts of the gear transmission, but does not necessarily require them to be mounted on all shafts [3], which can be difficult for internal components. However, the inclusion of such encoders is already implemented in some machines, for operational purposes, and this is likely to increase with the wider implementation of the Internet of Things.

Acknowledgments

The authors would like to thank Mr Haichuan Chang for acquiring the impressions and providing illustrations of the gear tooth surfaces, for the wear tests, at different stages of the wear process.

References

[1] Sweeney, P.J. and Randall, R.B., *Gear Transmission Error Measurement using Phase Demodulation*. Proc. I.Mech.E., Part C, J.Mech.Eng.Sci. 210(C3), 1996, pp. 201-213.

[2] Randall, R.B., Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications. Wiley, Chichester, UK, 2011.

[3] Randall, R.B., Peng, D., Smith, W.A., *Using measured transmission error for diagnostics of gears*, SIRM conference, Copenhagen, February 2019.

[4] Chang, H., Borghesani, P., Smith, W.A., Peng, Z. *Application of surface replication combined with image analysis to investigate wear evolution on gear teeth - A case study*, Wear, 430-431, 2019, pp. 355-368.

[5] Mark, W.D., Reagor, C.P., Static-transmission-error vibratory-excitation contributions from plastically deformed gear teeth caused by tooth bending-fatigue damage, Mechanical Systems and Signal Processing, 21, 2007, pp. 885-905.

[6] Smith, J.D., Gear Noise and Vibration, 2nd Ed., Marcel Dekker Inc., NY, 2003.

[7] Mohammed, O.D., Rantatalo, M., Aidanpää, J.-O., *Improving mesh stiffness calculation of cracked gears for the purpose of vibration-based fault analysis*, Engineering Failure Analysis, 34, 2013. pp. 235–251.

Squared envelope analysis based on the H_∞ filter order tracking: Application for bearing diagnosis

A. Assoumane [1*], D. Abboud [1], E. Sekko [2], M. Elbadaoui [1,3]

[1]Safran Tech, Rue des Jeunes Bois â Châteaufort, 78772 Magny-les-Hameaux, France
[*]amadou.assoumane@safrangroup.com
[2]Univ Orléans, PRISME EA422921, 21 Rue Loigny-La-Bataille, 28000 Chartres, France
[3]Univ Lyon, UJM-St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France

Abstract

The analysis of the squared envelope spectrum (SES) is one among the most used tools for bearing diagnosis. It can easily reveals the characteristic frequencies related to the bearing fault [1, 2]. Actually, the envelope is estimated through a demodulation process in a selected frequency band. The proper choice of the latter is really challenging in a complex environment [3]. In addition to that, the frequency of the bearing fault is likely to be masked by deterministic components. This can jeopardize the efficiency of classical techniques [3, 4, 5]. In this paper, a new approach for bearing diagnostic is proposed. It is based on a recently proposed order tracking technique using the H_∞ filter [7]. In details, the method starts by computing the squared envelope (SE) of the raw signal over the full demodulation band without prior processing. Next, the SE is modeled in a state space using a trigonometric series expansion. Last, an H_∞ estimator is designed to extract the amplitude of each harmonic related to the bearing fault signature. This estimator is well convenient to track the order of bearing faults, particularly in the presence of deterministic components (i.e. the noise). Since this noise is neither white nor Gaussian, the traditional Kalman filter order tracking is compromised [8, 9, 10]. Contrary to the Kalman filter, the H_∞ filter is based on the minimax optimization. The minimax approach leads to the minimization of the estimation error for the worst possible amplification of the noise signal. More interestingly, no prior knowledge about the statistical properties of the noise signals is required [11, 12]. The efficiency of the proposed approach is demonstrated on simulated and real-world vibration signals in nonstationary regimes.

Keywords: H_∞ filter, state space modelling, order tracking, squared envelope, bearing diagnosis, vibration signal, variable speed condition.

1 Introduction

Rolling element bearings are among the most widely used elements in rotating machines. Because of their common role to carry high loads, bearings are likely to be exposed to sudden failures causing system outage. Thus, there has been an increasing interest in developing appropriate techniques for signal denoising and incipient fault detection. Due to their non-invasive nature and their high reactivity to incipient faults, the development of vibration-based techniques has spiked the interest of the scientific community [1]. In this context, envelope analysis has long been recognized as a powerful bearing diagnosis technique. Typically, it consists of a bandpass filtering step in a frequency band wherein the impulsive response is amplified, followed by a demodulation that extracts the signal envelope. The spectrum of the envelope reveals the desired diagnostic information, including the repetition frequency of the fault as well as possible modulations. It has been shown in [2] that it is preferable to use the squared envelope instead of the envelope as the latter is likely to introduce additional interfering components in the envelope spectrum. Since that time, the envelope spectrum was replaced by the squared envelope spectrum (SES) which has become the benchmark technique for bearing diagnostics. A powerful solution to this issue was proposed through the spectral kurtosis [16] (and some derived tools such as the kurtogram [17], the fast kurtogram [3], etc.) which provides an entirely blind way of identifying the best demodulation band according to the impulsivity criterion. Despite its remarkable relevance in machine

signal analysis, the efficiency of the spectral kurtosis is compromised in many situations; for instance, in the presence of energetic deterministic part or the presence of multiple impulsive sources or strongly in nonstationary conditions. This paper comes in this context aiming at providing a new way to address bearing diagnostic based on tracking bearing characteristic orders (i.e. fault frequencies referenced to the shaft frequency) in the squared envelope, without the need of eliminating the deterministic component neither to filter the signal. The method uses the fact that the SE signal comprises a cyclic patterns related to bearing fault. From this observation, the SE signal is described in the state space model using a trigonometric series expansion. Then, an H_∞ filter is designed to track bearing fault order components. This approach is different from the classical Kalman filter based order tracking. The latter is a widespread method used to track sinusoidal components [8, 9, 10], assuming that the exegeneous noises that affect the state model are white and Gaussian with known statistics. In current situations, those assumptions are not valid. Indeed, the meshing components that mask the bearing ones are neither white nor Gaussian. To address this issue, an H_∞ filter is proposed. The latter minimizes the estimation error for the worst possible amplification of the noises. This leads to a minimax optimization where no prior knowledge about the statistical properties of the noises is required [11, 12].
The proposed approach is presented in this paper as follows. In Section 2, the SE of a discrete vibration signal is described in the state space model using a trigonometric series expansion. In Section 3, the methodology to design an H_∞ filter is exposed. In Section 4, the proposed approach is first applied to a simulated vibration signal. Then, it is applied to analyze real-life vibration signals acquired from a wind turbine under nonstationary conditions. Conclusions of this paper are given in Section 5.

2 State space modeling of the squared envelope signal

Condider the discrete measured vibration signal as follows:

$$y[k] = y_r[k] + y_g[k] + b[k] \tag{1}$$

where $y_r[k]$ is the signal related to the bearing vibration, $y_g[k]$ is the meshing signal and $b[k]$ is the signal composed of all the exegeneous vibrations such as the background noise for all $k = 1, \cdots, N$. N is the number of signal samples. The meshing signal, in the case of a tooth crack, exhibits amplitude and phase modulations [6]. The corresponding signal in nonstationary regimes can be written as:

$$y_g[k] = \kappa(\omega[k]) \sum_m a_m[k] e^{j\phi_m[k]} e^{j\theta_m[k]} \tag{2}$$

in which $\kappa(\omega[k])$ is a modulation function depending on the machine regime, $\omega[k] = 2\pi f_r[k]$ is the shaft angular speed and f_r is the machine rotating frequency, a_m and ϕ_m are respectively the amplitude and phase modulations, θ_m is the instantaneous meshing angular displacement and j is the complex number such as $j^2 = -1$. Concerning the bearing vibration signal, it exhibits a series of impulses which can be modelled as [1]:

$$y_r[k] = \kappa(\omega[k]) M[k] \sum_i^d A_i I[k - \lceil T_i f_s \rceil] \tag{3}$$

in which:

- $M[k]$ is the load distribution function for an inner-race under radial load. In stationary conditions, this function is periodic at the shaft rotating period [18];
- A_i is the amplitude of the ith impact so that $A_i = A + \delta A_i$. A is the mean value of the distribution and δA_i is a zero-mean random part with σ_A its standard deviation;
- T_i is the instant of apparition of the ith impact;
- I is the damping response that depends on the damping factor and the resonance frequency of the bearing structure;
- d is the number of impacts resulting from the bearing fault;

- f_s is the sampling frequency;
- $\lceil \cdot \rceil$ stands for the integer part of a decimal number.

Since the bearing's rolling elements are subject to slipping phenomena, the time of occurence from one impact to another is not constant. This time exhibits a random part and, as mentionned in [1], can be modelled in stationary conditions as:

$$T_i = iT + \delta T_i \tag{4}$$

where T is the time isntant between two consecutive impacts and δT_i is a random variable with a Gaussian distribution. This modelling is no longer valid in nonstationary conditions. In this context, Borghesani et *al.* [19] and Abboud et *al.* [20] have written the instant of impact occurrence as follows:

$$T_i = t(i\theta_d + \delta\theta_i) \tag{5}$$

in which θ_d is the angular period of the bearing fault and $\delta\theta_i$ is a zero-mean Gaussian distribution.
The squared envelope of the measured vibration signal, which is of interest in this work and denoted SE, is given as follows:

$$\begin{aligned} \mathrm{SE}[k] &= \mathbb{E}\{y[k]\bar{y}[k]\} & (6)\\ &= \mathbb{E}\{(y_r[k]+y_g[k]+b[k])(\bar{y}_r[k]+\bar{y}_g[k]+\bar{b}[k])\} & (7) \end{aligned}$$

where $\bar{a}$ is the conjugate of the complex number a and $\mathbb{E}\{\cdot\}$ stands for the expectation symbol.
In this paper, it is assumed that the bearing, the meshing and the noise signals are mutually not correlated. Hence, the squared envelope becomes:

$$\begin{aligned} \mathrm{SE}[k] &= \mathbb{E}\{y_r[k]\bar{y}_r[k]\} + \mathbb{E}\{y_g[k]\bar{y}_g[k]\} + \mathbb{E}\{b[k]\bar{b}[k]\} & (8)\\ &= \mathbb{E}\{y_r[k]\bar{y}_r[k]\} + n[k] & (9) \end{aligned}$$

where $n[k] = \mathbb{E}\{y_g[k]\bar{y}_g[k]\} + \mathbb{E}\{b[k]\bar{b}[k]\}$ is considered as a noise signal. Otherwise, the SE can be expressed using the autocorrelation function (ACF) denoted by $\mathscr{R}$. The latter, applied to the bearing signal y_r in equation (3), can be written as [20]:

$$\mathscr{R}[k,j] \approx (A^2+\sigma_A^2)\kappa^2(\omega[k])M^2[k]\sum_{i=1}^{d}\mathbb{E}\{g[k-\lceil T_i f_s\rceil, j]\} \tag{10}$$

where $g[k,j] = I[k]I[k-j]$. By writing this function in the angular domain, one gets:

$$\mathscr{R}[k_a,j] \approx (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-\lceil \theta_i N_a\rceil, j]\} \tag{11}$$

with $\tilde{x}$ the angular transformation of the time variable x, k_a the sample index in the angular domain, Na the angular sampling frequency and θ_i the angle instant of the ith impact occurrence. Refering to equation (5), the latter is modelled as $\theta_i = i\theta_d + \delta\theta_i$ [22]. Thus, the ACF becomes:

$$\mathscr{R}[k_a,j] \approx (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-k_{a,i}-\delta k_{a,i}, j]\} \tag{12}$$

where $k_{a,i} \approx \lceil i\theta_d N_a\rceil$ is the angular sample of the ith impact occurrence and $\delta k_{a,i} \approx \lceil \delta\theta_i N_a\rceil$ is a random integer. The above equation of ACF has been proven by Abboud et *al.* in [20]. By taking advantage of this equation, the SE, using the ACF, can be expressed as:

$$\begin{aligned} \mathrm{SE}[k_a] &= \mathscr{R}[k_a, j=0] + \tilde{n}[k_a] & (13)\\ &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-k_{a,i}-\delta k_{a,i}, j=0]\} + \tilde{n}[k_a] & (14)\\ &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{h}^2[k_a-k_{a,i}-\delta k_{a,i}]\} + \tilde{n}[k_a] & (15) \end{aligned}$$

Assume that the random variable $k_a - k_{a,i} - \delta k_{a,i}$ has a probability density function $f[\delta k_{a,i}]$ centered at $k_a - k_{a,i}$ with a constant standard deviation. According to the *law of the unconscious statistician* [15], the above equation is written as:

$$\begin{aligned} \mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\sum_{i=1}^{d}\tilde{h}^2[k_a-k_{a,i}-\delta k_{a,i}]f[\delta k_{a,i}]+\tilde{n}[k_a] && (16)\\ &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}(\tilde{h}^2 \circledast f)[k_a-k_{a,i}]+\tilde{n}[k_a] && (17)\\ \mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}s[k_a-k_{a,i}]+\tilde{n}[k_a] && (18) \end{aligned}$$

in which $\circledast$ stands for the convolution symbol and $s[k_a] = (\tilde{h}^2 \circledast f)[k_a]$ is the convolution between the function $\tilde{h}^2$ and f. The function $\tilde{M}^2[k_a]$ is deterministic and can be approximated by a Fourier series such as $\tilde{M}^2[k_a] = \sum_s \lambda_s[k_a]e^{j\psi_s[k_a]}e^{js\theta_r[k_a]}$ where λ_s and ψ_s are respectively the sth variable amplitude and phase of the Fourier series and θ_r is the angular period of the shaft. In same way, the sum in the SE formula can also be expressed by $\sum_{i=1}^{d} s[k_a-k_{a,i}] = \sum_z \rho_z[k_a]e^{j\varphi_z[k_a]}e^{jz\theta_d[k_a]}$ with ρ_z and φ_z respectively the zth variable amplitude and phase of the Fourier series and θ_d the angular period of bearing fault. This leads to:

$$\mathrm{SE}[k_a] = (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\Re\{\sum_{s,z}\lambda_s[k_a]\rho_z[k_a]e^{j(\psi_s[k_a]+\varphi_z[k_a])}e^{j(z\theta_d[k_a]\pm s\theta_r[k_a])}\}+\tilde{n}[k_a] \quad (19)$$

where $\Re\{.\}$ defines the real part of a complex number. In this paper, all the components related to the shaft angular period are not of interest. Therefore, the SE is written as:

$$\begin{aligned} \mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\Re\{\sum_{s=0,z}\lambda_0[k_a]\rho_z[k_a]e^{j(\psi_0[k_a]+\varphi_z[k_a])}e^{j(z\theta_d[k_a])}\}+v[k_a] && (20)\\ &\approx \sum_{z=1}^{l}\alpha_z[k_a]\cos(z\theta_d[k_a]+\phi_z[k_a])+v[k_a] && (21) \end{aligned}$$

in which $\alpha_z[k_a] = (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\lambda_0[k_a]\rho_z[k_a]$, $\phi_z[k_a] = \psi_0[k_a]+\varphi_z[k_a]$ are respectively the zth amplitude and phase of the Fourier series, $v[k_a]$ is the noise comprising the initial noise $\tilde{n}[k_a]$ and all the components related to the shaft angular period $\theta_r[k_a]$ and l is the higher order of the series. The latter defines the number of the bearing component of interest in the estimation procedure. In the state modelling approach that is proposed in this paper, $v[k_a]$ is the so-called *measurement error* or *measurement noise*.

At this stage, the detection of the bearing fault is reduced to the estimation of the amplitude α_z and the phase ϕ_z of the zth order component. This estimation can be done using a linear or a non-linear filtering approach. However, the non-linear approach are subject to a divergence issue. To obtain a linear model, the SE signal is presented in the following form:

$$\mathrm{SE}[k_a] = \sum_{z=1}^{l}\boldsymbol{h}_z^T[k_a]\boldsymbol{x}_z[k_a]+v[k_a] \quad (22)$$

where:

- $\boldsymbol{h}_z[k_a] = \begin{pmatrix} \cos(z\theta_d[k_a]) & \sin(z\theta_d[k_a]) \end{pmatrix}^T \in \mathbb{R}^{2\times 1}$ is the zth measurement vector. $\mathbb{R}$ stands for the ensemble of the real number and $(\cdot)^T$ for the transpose symbol. In the rest of this paper, the lowercase symbols in bold stand for vectors and the uppercase ones in bold stand for matrices;
- $\boldsymbol{x}_z[k_a] = \begin{pmatrix} \alpha_z[k_a]\cos(\phi_z[k_a]) & \alpha_z[k_a]\sin(\phi_z[k_a]) \end{pmatrix}^T \in \mathbb{R}^{2\times 1}$ is the zth state variable.

Using equation (22), the estimation of the amplitude and the phase of the Fourier series reduces to the estimation of the state variable $\boldsymbol{x}_z$. In this context, the simultaneous estimation of $\boldsymbol{x}_z$ leads to write the SE as:

$$\mathrm{SE}[k_a] = \boldsymbol{h}^T[k_a]\boldsymbol{x}[k_a]+v[k_a] \quad (23)$$

in which $\boldsymbol{h}[k_a] = \begin{pmatrix} \boldsymbol{h}_1^T[k_a] & \cdots & \boldsymbol{h}_l^T[k_a] \end{pmatrix}^T \in \mathbb{R}^{2l\times 1}$ and $\boldsymbol{x}[k_a] = \begin{pmatrix} \boldsymbol{x}_1[k_a] & \cdots & \boldsymbol{x}_l[k_a] \end{pmatrix}^T \in \mathbb{R}^{2l\times 1}$ are respectively the *measurement vector* and the *state variable*. Assuming that the angular period of the bearing fault is known, the detection of the latter is reduced to the estimation of the state variable $\boldsymbol{x}[k_a]$. In this context, it is proposed to estimate $\boldsymbol{x}[k_a]$ in a recursive manner using a state space modelling approach.
For this reason, the SE signal is described in a state space. That is to say the dynamic of the state variable has to be defined. In this paper, all parameters include in the state variable $\boldsymbol{x}[k_a]$ are supposed to follow, roughly speaking, a *random walk* so that:

$$\boldsymbol{x}[k_a+1] = \boldsymbol{x}[k_a] + \boldsymbol{w}[k_a] \tag{24}$$

where $\boldsymbol{w}[k_a]$ is a random or deterministic signal with bounded energy. Equations (23) and (24) form the state space model of the *SE* signal. From the latter, an H_∞ filter is designed in the next section for the state variable estimation.

3 H_∞ filter order tracking

Considering equations (23) and (24), the H_∞ filter will be designed to estimate some arbitrary linear combination of the state, say:

$$\hat{s}[k_a] = \boldsymbol{h}^T[k_a]\hat{\boldsymbol{x}}[k_a] \tag{25}$$

where $\hat{x}[k_a]$ satisfies the following recursion:

$$\hat{\boldsymbol{x}}[k_a] = \hat{\boldsymbol{x}}[k_a-1] + \boldsymbol{g}[k_a]\left(\mathrm{SE}[k_a-1] - \boldsymbol{h}^T[k_a-1]\hat{\boldsymbol{x}}[k_a-1]\right) \tag{26}$$

where $\boldsymbol{g}[k_a]$ is the H_∞ gain and $\hat{\boldsymbol{x}}[k_a]$ is the estimate of $\boldsymbol{x}[k_a]$. The state variable is estimated for any $v[k_a]$ and $\boldsymbol{w}[k_a]$ of bounded energy.
Let $e[k_a] = s[k_a] - \hat{s}[k_a]$ be the estimation error, then the H_∞ gain is found by minimizing the following cost function given by [12]:

$$J = \frac{\sum_{k_a=1}^{N} \|e[k_a]\|^2}{\|e[1]\|^2_{\boldsymbol{P}[1]^{-1}} + \sum_{k_a=1}^{N}\left(\|\boldsymbol{w}[k_a]\|^2_{\boldsymbol{Q}^{-1}} + \|v[k_a]\|^2_{R^{-1}}\right)} \tag{27}$$

where $(e[1], \boldsymbol{w}[k_a], v[k_a]) \neq (0,0,0)$, $e[1]$ represents the initial error, $\boldsymbol{P}[1] > 0$, $\boldsymbol{Q} > 0$ and $R > 0$ are positive definite weighting matrices, N is the number of samples and $\|e[k_a]\|_{\boldsymbol{S}} = e[k_a]^T S e[k_a]$. This can be interpreted as the energy gain from the unknown disturbances $\boldsymbol{P}^{-1/2}[1]e[1]$ and $\{\boldsymbol{Q}^{-1/2}\boldsymbol{w}[k_a], R^{-1/2}v[k_a]\}_{k_a=1}^N$ to the estimation error $\{e[k_a]\}_{k_a=1}^N$. It is quite clear that if the ratio in (27) is small then the estimation is better, and vice versa. However, this ratio depends on the quantities $e[1]$, $\boldsymbol{w}[k_a]$ and $v[k_a]$ which are unknown. In this context, the worst case is considered below:

$$\underset{e[1],\boldsymbol{w}[k_a],v[k_a]}{\mathrm{Sup}} \quad J \leq 1/\gamma \tag{28}$$

where "Sup" stands for the supremum and γ is the performance bound. Otherwise, the goal of the H_∞ problem is to find an estimation $\{\hat{s}[k_a]\}_{k_a=1}^N$ that minimizes the worst-case energy. This is equivalent to minimize the following scalar quadratic form:

$$J_f = e^T[1]\boldsymbol{P}^{-1}[1]e[1] + \sum_{k_a=1}^{N} \boldsymbol{w}[k_a]^T\boldsymbol{Q}^{-1}\boldsymbol{w}[k_a] + \sum_{k_a=1}^{N} v[k_a]^T R^{-1} v[k_a] - \gamma \sum_{k_a=1}^{N} e[k_a]^T e[k_a] \tag{29}$$

so that $J_f > 0$ for all vectors $e[1]$, for all nonzero signals $\boldsymbol{w}[k_a]$ and $v[k_a]$ of bounded energy.
Giving the cost function J_f, the worst case minimization is reduces to minimize J_f in respect to $\hat{s}[k_a]$ and to maximize J_f in respect to $e[1]$, $\boldsymbol{w}[k_a]$ and $v[k_a]$ for all $k_a = 1, \cdots, N$. This leads to a minmax optimization formulated in such a way that:

$$\{\hat{s}[k_a]\}_{k_a=1}^N = \arg\left(\min_{\hat{s}} \max_{e[1],\boldsymbol{w},v} (J_f)\right) \tag{30}$$

This optimization problem can be solved by the well known Lagrange mutliplier approach. The solution to the above optimization is given by the theorem quoted below [11, 12, 14].

Theorem 1 *Let $\gamma > 0$ be the user-specified performance bound. Then, there exists an H_∞ estimation for $s[k_a]$ if and only if there exists a symmetric positive definite matrix $\boldsymbol{P}[k_a] \in \mathbb{R}^{2l\times 2l}$ that satisfies the following discrete-time Riccati equation:*

$$\boldsymbol{P}[k_a] = \boldsymbol{P}[k_a-1]\boldsymbol{\Gamma}[k_a-1] + \boldsymbol{Q} \tag{31}$$

when

$$\boldsymbol{\Gamma}[k_a] = \left(\boldsymbol{I}_{2l} - \gamma \boldsymbol{h}^T[k_a]\boldsymbol{h}[k_a]\boldsymbol{P}[k_a] + \boldsymbol{h}[k_a]R^{-1}\boldsymbol{h}^T[k_a]\boldsymbol{P}[k_a]\right)^{-1} \tag{32}$$

and $\boldsymbol{I}_{2l} \in \mathbb{R}^{2l\times 2l}$ is the identity matrix.
Then, the H_∞ gain $\boldsymbol{g}[k_a] \in \mathbb{R}^{2l\times 1}$ is given by:

$$\boldsymbol{g}[k_a] = \boldsymbol{P}[k_a]\boldsymbol{\Gamma}[k_a-1]\boldsymbol{h}[k_a]R^{-1} \tag{33}$$

It should be noted that for some weighting matrices $\boldsymbol{P}[1]$, $\boldsymbol{Q}$ and R the performance criterion in (28) is achieved if and only if the performance bound γ satisfies the following inequality:

$$\gamma < R^{-1} \tag{34}$$

Since γ defines the noise level attenuation or the performance bound, it should be as high as possible. And, it has been shown in a previous paper [7] that when γ is greater than its optimal value, the matrix $\boldsymbol{P}$ is not symmetric positive definite. Otherwise, when γ tends to zero, the H_∞ filter is not constrainted. Then, it is equivalent to the standard Kalman filter for which R and $\boldsymbol{Q}$ are defined respectively as the covariance matrix of the measurement noise and the state noise.

4 Application

4.1 Synthetic signal analysis

Here, a synthetic signal is presented to evaluate the performance of the proposed approach in estimating bearing order components. The signal is composed of the bearing, the meshing and the noise signal. They represent respectively 20%, 30% and 50% of the synthetic signal energy. The non-stationary condition is simulated using a non-linear rotating frequency varying between 5 Hz and 30 Hz such as $f_r[k] = 5 + 25\sin\left(\frac{\pi(k-1)}{2(N-1)}\right)$ for $1 \leq k \leq N$ where N is the number of samples. The time duration of the signal is 5 s.
Concerning the meshing signal, sampled at the frequency $f_s = 10$ kHz, it is computed using equation (2) and contains five meshing components. The latter is composed of the 45th and 49th shaft order component with amplitude and phase modulations. The generation of these modulations is presented in details in Appendix A. About the bearing signal, it is generated using equation (3) in which the resonance frequency and the damping factor are respectively equal to 4 kHz and 2000.

The rolling bearing considered is subject to a local defect occurring on the outer-race. Its characteristic order, ball-pass-order on the outer-race denoted BPOO, is equal 8.7 times the rotating frequency with a slight random variation. Last, the additive noise is generated using a white Gaussian noise modulated by the rotating frequency.
The different contributions of the signal are displayed in Figure 1. It can be seen in the SES of the raw signal (Figure 1 (g)) that the bearing fault order is masked by deterministic components. Thus, the proposed approach is applied to track the order component of the bearing fault. The number of order harmonic of interest l is equal to 15. Also, the parameters of the filter take the following values: $R = 1$, $\boldsymbol{Q} = 0 \times \boldsymbol{I}_{2l}$, $\gamma = 0.9R^{-1}$ and the filter initial values are $\boldsymbol{P}[1] = \boldsymbol{I}_{2l}$ and $\boldsymbol{x}[1] = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{2l\times 1}$. The choice of $\boldsymbol{Q}$ is motivated by the fact the coefficients of the trigonometric series expansion are assumed to be constant. This means that the state noise $\boldsymbol{w}[k]$ is null for all $k = 1, \cdots, N$.
The result of the estimation provided by the H_∞ estimator for the choosen parameters is presented on Figure 2. There, the number of peaks estimated by the proposed approach correponds to the expected ones, i.e 15 peaks. Moreover, the deterministic components present in the SES of the raw signal have been greatly attenuated. Besides of that, some peaks, with lower energy level, appear around the bearing order components. This is due to the fact that the H_∞ filter, like all type of filter, don't atteanuate uniformely all the frequencies outside the

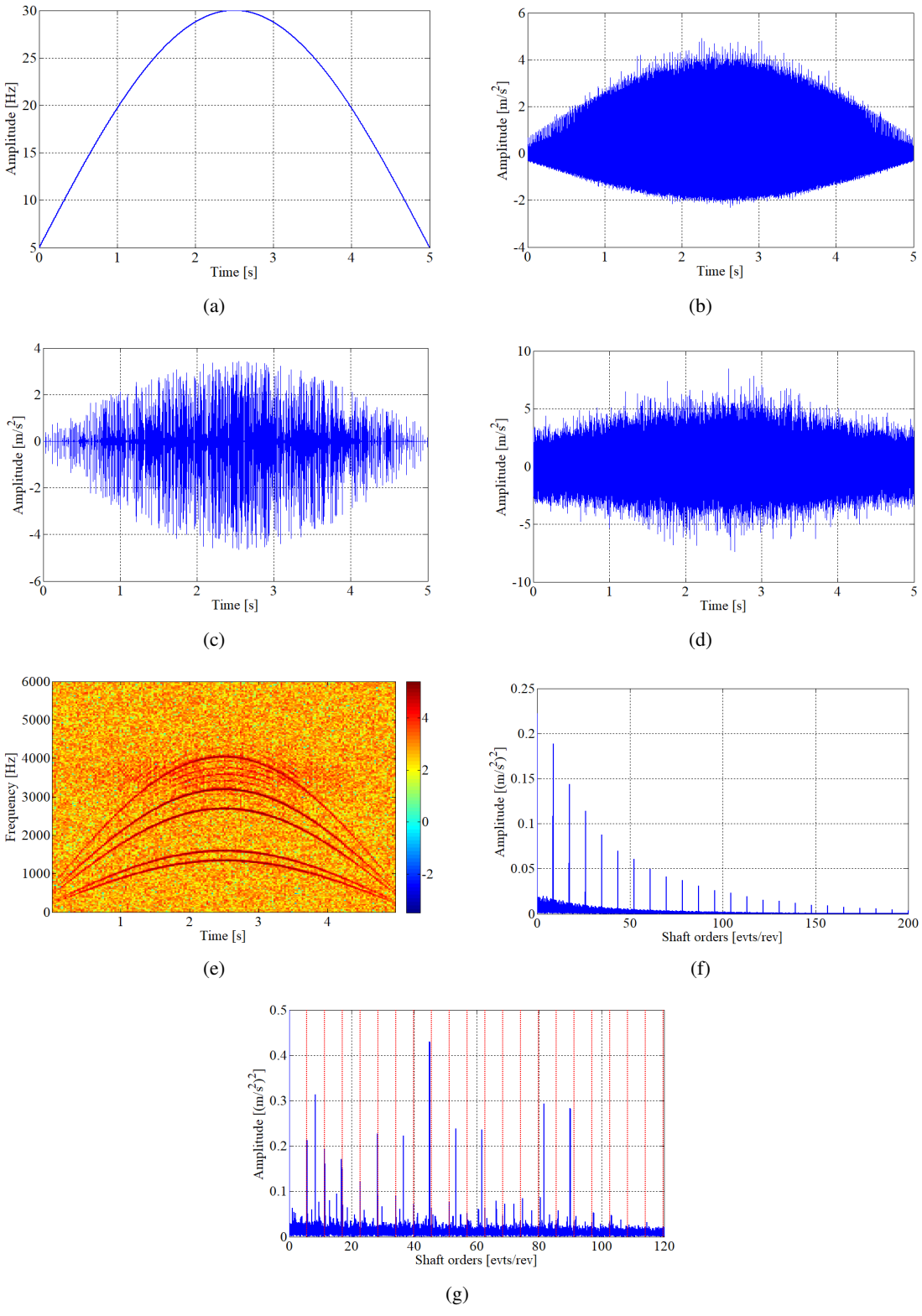


Figure 1 – Synthetic vibration signal: (a) Rotating frequency, (b) Meshing signal, (c) Bearing signal, (d) Complete vibration signal, (e) Time-frequency representation, (f) SES of the bearing signal and (g) SES of the complete signal.

band of interest.
To evaluate the performance of the proposed approach, the signal-to-noise ratio is calculated by the following formula $\text{snr} = 10 \times \log_{10}\left(\frac{\sum_{k=1}^{N} s^2[k]}{\sum_{k=1}^{N} (\hat{s}[k]-s[k])^2}\right)$ where s is the SE related only to the bearing signal and $\hat{s}$ is its estimate. When the bearing signal is totally influenced by the meshing and the noise signal, that is the worst case estimation, the estimation error is $\hat{s}[k] - s[k] = b[k]$. In this case, the snr is equal to -6 dB and defines the lower limit of the performance bound. On this basis, it can be stated that all estimations provided by the proposed approach should have a snr greater than -6 dB. It follows that a higher snr leads to a better estimation. For this first simulation, after a 200 Monte-Carlo simulations, the estimation error leads to a snr equal to 1.48 dB. This value is greater than the performance bound and corroborates the quality of the estimation displayed on Figure 2.

Influence of the fault frequency incertitude

In the results presented above the frequency (or order) of the bearing was exactly known. In real situations, this frequency is known with uncertainty. This can be due (i) to the fluctuation of the shaft rotating frequency, (ii) to the imperfection of the speed sensor or (iii) to the sliping phenomena of the bearing rolling's elements. In this section, the influence of the uncertainty of the bearing fault order on the estimation quality is investigated. Thus, the snr is evaluated for different values of uncertainty on the bearing fault order. This uncertainty varies from 0 % to 5 % of the real value of the order of the bearing fault. The snr obtained in the range of the uncertainty is displayed on Figure 3. There, the snr remains constant when the uncertainty varies from 0% to 2.5%. This is interesting for the bearing health monitoring since the uncertainty on a potential fault frequency can reach 2% in real situations as mentionned by Randall and Antoni [1]. Beyond this value, the snr decreases; thus the quality of the estimation is degraded and the proposed approach is no longer robust to track the bearing order components.

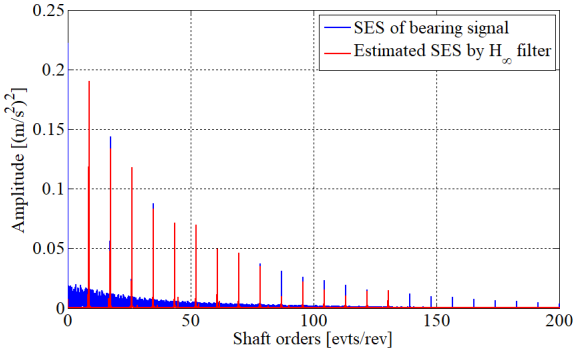


Figure 2 – Estimation of the SES of the bearing signal provided by the H_∞ filter. Here, the uncertainty on the bearing fault order is equal to zero.

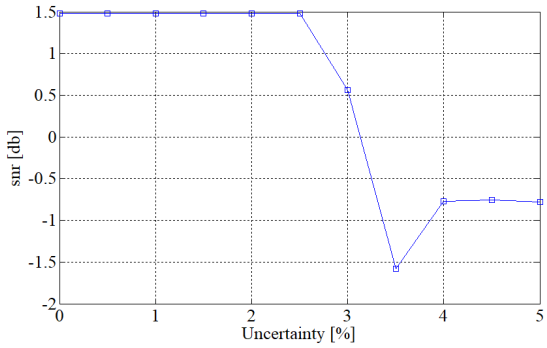


Figure 3 – Signal-to-noise ratio (snr) evolution against the uncertainty on the bearing fault order.

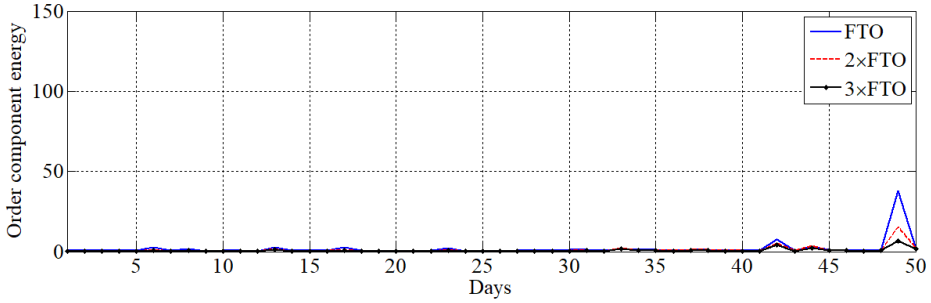
4.2 Experimental signal analysis

This subsection deals with signals acquired from a 2 megawatts wind turbine high speed shaft on which a condition monitoring system is installed. The bearing has an inner race fault which is increasing in severity across the 50-day period. At the end of the test, the bearing was inspected and a crack has been identified in the inner race. Acceleration signals were recorded on a daily basis (one signal per day) together with tachometer signals, over a 6 s duration each with a sampling frequency equal to 97656 Hz. The nominal speed of the bearing shaft is 1800 rpm (30 Hz). Note that the speed variability has reached 15 % of the nominal speed in some records; the reason why the regime is considered nonstationary. The theoretical fault frequencies referenced to the shaft frequency are as follows:

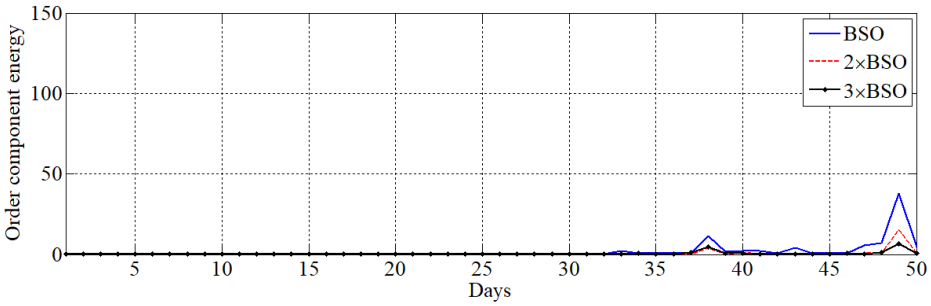
- Ball pass order on outer-race: BPOO $= 6.72$;
- Ball pass order on inner-race: BPOI $= 9.47$;

- Ball spin order: BSO = 1.435;
- Fundamental train order: FTO = 0.42.

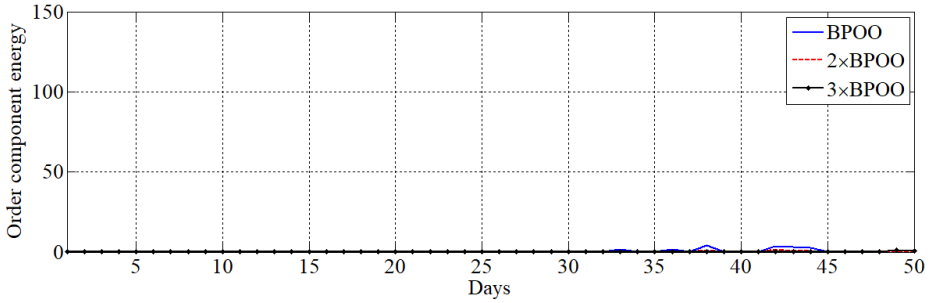
More information can be found in [21]. In this section, the proposed approch is applied to track the order components locolated at $i \times$ BPOO, $i \times$ BPOI, $i \times$ BSO and $i \times$ FTO where $i = 1,2,3$. The parameters of the H_∞ filter take the following values: $R = 1$, $\boldsymbol{Q} = 0 \times \boldsymbol{I}_3$ and $\gamma = 0.95$ and the filter initial values are $\boldsymbol{P}[1] = \boldsymbol{I}_3$ and $\boldsymbol{x}[1] = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{6\times 1}$. To monitor the health state of each component of the bearing during the 50-day periods, the energy of each order component is evaluated. Figure 4 shows the evolution of this energy during



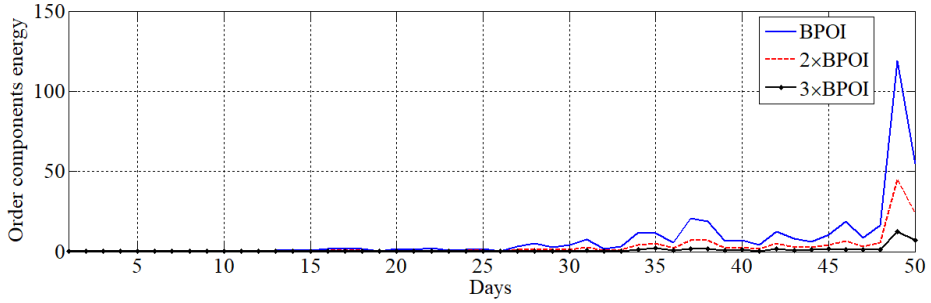
(a) FTO



(b) BSO



(c) BPOO



(d) BPOI

Figure 4 – Evolution of the order component energy for the different type of fault that can occur on the bearing over the 50 days of measurement: (a) cage fault, (b) Rolling elements fault, (c) Outer-race fault and (d) Inner-race fault.

the 50 days. Concerning the order components located at $i \times$ FTO and $i \times$ BSO with $i = 1,2,3$, their energy remains constant till the 49th day where an amplification is observed. This may be related to a degradation of the bearing train and rolling elements at the end of the test. About the BPOO components, their energy is almost constant and negligible compare to that of the FTO and BSO. Therefore, the bearing outer-race is healthy. Otherwise, the energy of the order components located at $i \times$ BPOI increases along the days. This amplification is related to the evolution of the inner-race fault severity. A significant jump in the energy can be seen from the 30th day. According to the inspection done after the 50th day, a crack in the inner-race has been noticed at the 50th day. Thus, the proposed order tracking approach is efficient to detect earlier a bearing fault. Based on the bearing order component estimation provided by the H_∞ filter, different indicators can be designed to monitor bearing health state.

5 Conclusion

In this paper, an order tracking technique was proposed to diagnose a bearing fault under a nonstationary condition. The proposed method consists of estimating a certain number of bearing order components whitout removing the deterministic components. The method described the squared envelope signal in the state space model using a trigonometric series expansion. Then, an H_∞ filter is designed to track bearing fault order components. Firstly, the theoretical foundation of the proposed approach has been described in details. Secondly, a synthetic signal has been generated to evaluate the performance of the proposed approach. It has been shown that the approach was able to track the bearing order components without removing the deterministic components. Moreover, the performance of the proposed approach remains stable for an uncertainty error on bearing orders less than 2.5%. Finally, the efficiency of the proposed approach has been demonstrated with wind turbine vibration signals under a nonstationary condition. The order components related to the bearing fault has been estimated by the proposed approach throughout 50-days of measurement. A fault on the bearing inner-race has been successfully detected earlier at the 30th day. In terms of perspective for this research, the authors will work on the design of a robust H_∞ filter to deal with a large uncertainty on the order of the bearing fault.

Acknowledgements

Acknowledgement is made for the measurements used in this work provided through data-acoustics.com Database.

A Synthetic meshing signal generation

The meshing signal is composed of five components and presented as follows:

$$y_g[k] = \kappa(\omega[k]) \sum_{m=1}^{5} y_{g,m}[k] \tag{35}$$

where $y_{g,m}$ is the mth meshing component. The latter is defined by the expression below:

$$y_{g,m}[k] = A_m[k](1 + a_m[k])e^{o_m \theta[k]} \tag{36}$$

in which:

- a_m is the amplitude modulation of the mth meshing component;
- A_m is the amplitude of the mth meshing component so that it is a random value comprises between 0 and 1 and $\kappa(\omega[k]) = \omega^2[k]/\max(\omega[:])$ is the modulation correponding to the variation of the regime in which ω is the angular speed of the machine;
- $\theta[k] = 2\pi t_s \sum_{c=1}^{k} f_r[c]$ is the instantaneous angular displacement of the shaft rotating at frequency f_r. t_s is the sampling period;

- o_m corresponds to the mth order of the meshing signal. It takes respectively the following values:45, 2×45, 3×45, 1.089×49 and $2 \times 1.089 \times 49$.

It is well known that when a fault appears on a gear tooth, the resonance frequency of the gear structure is excited by an impulse [6]. The generated impulse signal modify the shape of the amplitude modulation so that the latter can be described by a series of impulsive. Each impulse is modeled by a narrow bandwidth gausian function. This function is defined by as below:

$$ext[k] = \sum_i e^{-\frac{1}{2}\left(\frac{(k-1)t_s-\mu_i}{\sigma^2}\right)^2} \tag{37}$$

where μ_i is the center of the Gaussian function determined by the time instant for which the impulse occurs on the gear tooth and σ defines the width of the Gaussian function. In this simulation $\sigma = 10^{-3}$. Since the gear attenuates the impulse generated by the fault, a transfert function is included in the model so that the amplitude modulation becomes:

$$a_m[k] = \rho_m \times s[k] \circledast ext[k] \tag{38}$$

in which $s[k] = e^{-\xi^{(g)}(k-1)t_s} \sin\left(2\pi f_{res}^{(g)}(k-1)t_s\right)$ is the transfert function of the gear structure and ρ_m is a random value comprises between 0 and 1. $\xi^{(g)}$ and $f_{res}^{(g)}$ are respectively the damping factor and the resonance frequency of the gear structure. They are respectively equal to 5000 and 4000 Hz.

References

[1] Randall, R. B., and Antoni, J. Rolling element bearing diagnostics-a tutorial. *Mech. Syst. Signal Process.*, 25(2), 485-520, 2011.

[2] Ho, D., & Randall, R. B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mech. Syst. Signal Process.*, 14(5), 763-788.

[3] Antoni, J. (2007). Fast computation of the kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.*, 21(1), 108-124.

[4] Wang, D., Peter, W. T., & Tsui, K. L. (2013). An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.*, 35(1-2), 176-199.

[5] Sawalhi, N., Randall, R. B., & Endo, H. (2007). The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mech. Syst. Signal Process.*, 21(6), 2616-2633.

[6] McFadden, P. D. (1986). Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *Journal of vibration, acoustics, stress, and reliability in design*, 108(2), 165-170.

[7] Assoumane, A., Sekko, E., Capdessus, C., & Ravier, P. (2017). Order tracking using H_∞ estimator and polynomial approximation. *Mechanics & Industry*, 18(8), 808.

[8] Pan, M. C., & Lin, Y. F. (2006). Further exploration of Vold-Kalman-filtering order tracking with shaft-speed information-I: Theoretical part, numerical implementation and parameter investigations. *Mech. Syst. Signal Process.*, 20(5), 1134-1154.

[9] Pan, M. C., and Wu, C. X. Adaptive Vold-Kalman filtering order tracking. *Mech. Syst. Signal Process.*, 21(8), 2957-2969, 2007.

[10] Pan, M. C., Chu, W. C., & Le, D. D. (2016). Adaptive angular-velocity Vold-Kalman filter order tracking-Theoretical basis, numerical implementation and parameter investigation. *Mech. Syst. Signal Process.*, 81, 148-161.

[11] Shen, X. (1995, May). Discrete H_∞ filter design with application to speech enhancement. In *Acoustics, Speech, and Signal Processing*, 1995. *ICASSP*-95., 1995 *International Conference on* (Vol. 2, pp. 1504-1507). IEEE.

[12] Shen, X. M., & Deng, L. (1997). Game theory approach to discrete H/sub/spl infin//filter design. *IEEE Trans. Signal Process*, 45(4), 1092-1095.

[13] Pinkus, A. (2000). Weierstrass and approximation theory. *Journal of Approximation Theory*, 107(1), 1-66.

[14] Banavar, R. N., & Speyer, J. L. (1991, June). A linear-quadratic game approach to estimation and smoothing. In 1991 *American control conference* (pp. 2818-2822). IEEE.

[15] Ross, S. M. (2013). Introduction to stoachastic process. In *Applied probability models with optimization applications* (p. 2). Courier Corporation.

[16] Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mech. Syst. Signal Process.*, 20(2), 308-331.

[17] Antoni, J. (2006). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mech. Syst. Signal Process.*, 20(2), 282-307.

[18] McFadden, P. D., & Smith, J. D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of sound and vibration*, 96(1), 69-82.

[19] Borghesani, P., Ricci, R., Chatterton, S., & Pennacchi, P. (2013). A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mech. Syst. Signal Process.*, 38(1), 23-35.

[20] Abboud, D., Antoni, J., Eltabach, M., & Sieg-Zieba, S. (2015). Angle \ time cyclostationarity for the analysis of rolling element bearing vibrations. *Measurement*, 75, 29-39.

[21] Bechhoefer, E., Van Hecke, B., & He, D. (2013, October). Processing for improved spectral analysis. In *Annual Conference of the Prognostics and Health Management Society, New Orleans, LA, Oct* (pp. 14-17).

[22] Antoni, J., Bonnardot, F., Raad, A., El Badaoui, M. (2004). Cyclostationary modelling of rotating machine vibration signals. Mechanical systems and signal processing, 18(6), 1285-1314.

Dynamic Characterization of Hydroelectric Turbine with Transient Data Records Using OBMA and Phase-Shift Analysis

Quentin DOLLON[1], Antoine TAHAN[2], Jérôme ANTONI[3], Martin GAGNON[4], Christine MONETTE [5]

[1]ÉTS, Montréal, INSA Lyon
quentin.dollon.1@ens.etsmtl.ca
[2] ÉTS, Montréal, Dpt. Mécanique, H3C 1K3, Montreal, Qc., Canada
antoine.tahan@etsmtl.ca
[3]Univ Lyon, INSA-Lyon, Lab. Vibrations Acoustique, F69621 Villeurbanne, France
jerome.antoni@insa-lyon.fr
[4]Institut de Recherche de HydroQuébec (IREQ), J3X 1S1, Varennes, Qc., Canada
gagnon.martin11@ireq.ca
[5]Andritz Hydro Ltd., H9R 1B9, Pointe-Claire, Qc., Canada
christine.monette@andritz.com

Abstract

The purpose of this paper is to investigate the possibility of estimating Francis hydroelectric turbine modal parameters in transient conditions by focusing on resonance regions generated by the interaction of a structural mode with a frequency-variant harmonic pressure pulsation. Especially when numerous modes are in the same bandwidth, this method separates them by exciting only matching mode shapes. To extract a specified harmonic from the signal, the resonance retrieval is done using Order Tracking method. A classical ambient modal identification algorithm is then used to feature the isolated mode. Furthermore, using the phase-shift between measured locations, modes can be localized and shape determined.

List of Symbols

i	Complex unit
$diag[\boldsymbol{A}]$	Diagonal matrix of vector $\boldsymbol{A}$
$<\boldsymbol{A}\|\boldsymbol{B}>$	Complex inner product of $\boldsymbol{A},\boldsymbol{B}$
$\boldsymbol{A}^T$	Transpose of $\boldsymbol{A}$
$\boldsymbol{A}^*$	Conjugate of $\boldsymbol{A}$
$\|\boldsymbol{A}\|$	Determinant of $\boldsymbol{A}$
δ_r	Kronecker Symbol
N_S	Number of sensors
N_f	Number of frequency samples
$\boldsymbol{\Phi}$	$(N_S \times n)$ Global modal matrix for a system of n modes
MAC	Modal Assurance Criterion
MAC_{thr}	Threshold MAC in E-FDD
$\boldsymbol{\theta}_r$	$(4+N_S \times 1)$ Parameter vector of mode r
$\boldsymbol{\theta}$	$(4+N_S \times 1)$ Modal parameter vector as a variable
$\mathscr{L}(\boldsymbol{\theta})$	Negative Log-Likelyhood Function
ω_k	Discrete pulsation $k \in [\![1,N_f]\!]$
$\widehat{\boldsymbol{X}}_k$	$(N_S \times 1)$ Estimated frequency response in ω_k $(\mathbb{C})$
$\boldsymbol{X}_k$	$(N_S \times 1)$ Theoretical frequency response in ω_k $(\mathbb{C})$
$h_{r,k}$	Modal transfer function of mode r in ω_k $(\mathbb{C})$
p_k	Scaled FFT of modal Excitation in ω_k $(\mathbb{C})$
$\boldsymbol{\varepsilon}_k$	$(N_S \times 1)$ Scaled FFT of channel noise in ω_k $(\mathbb{C})$
$\boldsymbol{E}_k$	$(N_S \times N_S)$ Theoretical density matrix in ω_k $(\mathbb{C})$
$s_{1,k}$	First singular value of E_k
$\boldsymbol{\varphi}_r$	$(N_S \times 1)$ Theoretical mode shape of mode r ($\mathbb{C}$ in FDD)
$\widehat{\boldsymbol{\varphi}_r}$	$(N_S \times 1)$ Estimated mode shape of mode r
β_r	Characteristic real scalar value of mode r
ω_r	Natural pulsation of mode r
ξ_r	Damping Ratio of mode r
S_r	Modal force of mode r
S_{er}	PSD Error for mode r

1 Introduction

Design and exploitation of hydroelectric turbines relies on the knowledge of their dynamic behavior. This enables one to generate and validate models to either get a good assessment of life duration or plan predictive-based maintenance. Two sources of information are useful to properly characterize the mechanical behavior of a structure: numerical simulations and experimental data processing. Giving high computing power, the first

source could give a whole and detailed analysis of the behavior in any expected regime through Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) but needs to be validated by the second to be reliable. It is a straightforward consequence of the strong assumptions made to reduce computational burden and model the highly turbulent characteristics of the flow. The second approach relies on in-situ measurements to extract dynamic features [1].

The increase of computational power over years allows getting more accurate simulations for startup regimes [2, 3], no-load or part-load configurations [4, 5, 6] and even hydrodynamic damping estimations [7]. However, the results still show discrepancies in structural parameters due to deviations from real operating conditions: rotating machinery [8], fluid-structure interaction added mass, damping and stiffness [9, 10], cavitation influence [11, 12, 13] or boundary condition sensitivity [14]. On the other hand, experimental characterization is highly fragmented, but in general closer to reality for a given measured operating condition. The features obtained from experimental data relie on statistical models [15, 16], indirect measurements [17], time-frequency analysis [18], but can also be obtained by modal parameter identification using Operational Modal Analysis (OMA) [19] or Experimental Modal Analysis (EMA) [20]. In addition, the experimental hydraulic instability study can be used to compare different computational turbulence models [21, 22, 23]. Typically, the two sources of information (simulations and experiments) are crossed to obtain a hybrid representation of the dynamic behavior, which is used to obtain accurate load levels and allow a better prediction of fatigue [24]. Those predictions are used to assess the runner life duration and reliability of the capacity [25, 26, 27].

One of the problems with experimental analysis is the cost of data acquisition. To reduce financial burden of measurements, the idea is to extract a maximum of information from transient records instead of several stationary records, which would make the measurement less time-consuming. Furthermore, the processing of transient records allows obtaining real structural parameters of highly damaging regimes [23, 28] (what numerical analysis still struggles to perform, as aforementioned). Our goal is to determine whether a signal processing methodology is able to extract precise and suitable features from these transient measurements. For this, a combined methodology using Order Tracking, OMA and Phase-Shift Analysis is implemented and performed on a case study. The case study data come from a medium-head Francis Turbine in Quebec (Canada). The paper first presents the theoretical background, including literature and the different OMA tools to be used. Afterwards, the model is tested on the case study of an operational runner prototype.

2 Resonance Detection Using Phase-Shift Diagrams

Resonances are usually found with the study of experimental correlograms where the amplified regions are treated as Operating Deflecting Shape (ODS). But there is another alternative to detect resonances with more confidence: Phase-Shift Analysis (PSA) [29, 30]. Resonance amplitudes are time-dependent and phases are relative to a reference in experimental data, but the modal phase-shift from one sensor to another is a theoretical time invariant absolute quantity that is specific to each mode. Especially, when a harmonic (time-variant pulsation) and a mode (almost constant pulsation) intersect with the same phase-shift, the observed ODS is very likely the resonance of the only excited mode. This resonance can be extracted and processed with OBMA through a Single Degree of Freedom (SDoF) formulation (Section 3 & 4). Once the mode has been detected, it is possible to feature its shape: the mode shape is assumed to be the nodal diameter that fits the best the modal phase-shift (e.g. [18] in which a self-excited vibration of a hydroelectric runner is studied during load rejection). It is also possible to determine the shape by identifying the pattern of the exciting harmonic, particularly if this last comes from a well-known phenomenon (vortex rope [27, 31, 32, 33, 34], Rotor-Stator Interactions (RSI) [26, 27, 31, 35, 36]).

3 Order Tracking Procedure

Once PSA and resonance mapping is achieved, it is still required to extract accurate damping ratios and frequencies, and eventually other modal properties (modal force *etc.*). In order to do this, identification algorithms are implemented to process multi-channel resonance signals. The first pre-processing step is to extract the resonance component and isolate it from the rest of the signal. This is the purpose of Order Tracking. This class of method gathers all the tools able to extract one harmonic from the signal by shifting the time

domain to a harmonic one, called order domain. Orders, measured in times per revolution, are analog to frequencies. Order Tracking is a classical and very used diagnosis tool for rotating machineries. There are four main techniques that are commonly used: direct method using Fourier Transform of a time series (FS), Angular Resampled-based Order Tracking (AD), Time-Variant Discrete Fourier Transforms (TVDFT) and Vold-Kalman filters (VK).

FS extracts the $n-th$ harmonic from a signal by tracking the $\widehat{\boldsymbol{X}}[n\omega_{0,k}]$ response with a short-time Fourier transform at each time step, where $\omega_{0,k}$ is the runner angular velocity at time step k. This procedure is highly biased due to tapering, leakage effects and bandwidth control. In the classical approach with constant time intervals, low rotating frequencies are less accurate than higher ones. If time intervals are non-constant, some power spectral density rescaling issues rise and must be taken into account.

Another technique relies on an adaptative Fourier transform with settable kernel [37]: the kernel of the analytic exponential function tracks the frequency of interest. Consequently, a precise targeted order is extracted. In early works, the kernel orthogonality was lost and a compensation matrix had to be introduced to partly fix the problem. This issue is now easily fixed by introducing a change of variable in the integration domain, and gives a Velocity Synchronous Fourier Transform [38]. Vold-Kalman (VK) Bank Filters can extract orders from a signal with an instantaneous analysis instead of an averaging procedure [39]. Consequently, VK filtering is the most accurate technique in terms of resolution but entails a heavy computational burden, that is irreconcilable with industrial applications.

AD is a resampled-based method that avoids any leakage effect and phase issues [40]. The asynchronous time series are turned into synchronous time series (constant $\Delta\alpha$ instead of Δt) by the means of interpolation and tachometer record (Computed Order Tracking [41]). Then, a short-time Fourier transform is performed on the resampled signal, with intervals corresponding to one runner revolution (so that the spectrum resolution coincides with orders). Intervals are neither overlapped nor windowed. An order spectrum is obtained for each studied revolution. Each of those revolutions is converted into frequency by averaging the rotational speed over the lap. It can be noticed that the lower the studied dynamic, the weaker the quasi-static assumption over a revolution, the higher the response estimation quality. The bias of AD comes from both interpolation and synchronous interval split. Interpolation bias is due to interpolating method (*e.g.* linear, quadratic, splines) and shaft torsion that induces tachometer signal fluctuations. The issue with synchronous interval split is that each interval must represent exactly one revolution, that is not necessarily the case. Most of those biases can be reduced if data are recorded with a extensively high sampling frequency compared to the structure natural frequencies. For the purpose of this paper, the classical COT-based AD will be used, because the data sampling frequency is far higher than the natural frequencies of studied modes.

4 Operational Modal Analysis

Few has been done in the field of OMA for hydroelectric runner dynamic featuring. Gagnon et *al.* used this technique to characterize guide-vane behavior for different operating conditions [19]. The same point is made for EMA for which the study achieved on a runner obtained results that were in well agreement with simulations, but for experimental setting not representative of actual operating conditions [20]. Moreover, in many cases EMA cannot be implemented and when it is possible, suffers from major drawbacks like experimental set-up cost or structural size and complexity. The point of OMA is to extract modal parameters from output-only measurements containing both unknown excitation and response of the system. When those signals are extracted with Order Tracking, the procedure is called OBMA (Order Based Modal Analysis) [40, 42, 43, 44]. OMA is of interest for several reasons: it is fast in terms of computing effort and measurement (mere sensors replace excitation set-up), ambient excitation is appropriate to linearize the dynamic behavior and so on.

OMA techniques are divided into different classes: they can process in the time domain (TD) or frequency domain (FD), and can be parametric or non-parametric [45]. TD approaches are straightforward, and are generally parametric. They mainly study the auto-regression degree with (AR)MA-(X) models [46] and Subspace Identification technics [47] or the output correlations between channels (Polyreference, LSCE, Subspace Identification, ERA) [48]. Frequency approaches can be parametric (Polymax or Polyreference) [49] or non-parametric (Pick- Peaking, (E)-FDD) [50, 51]. Non-parametric methods often rely on Single Degree of Freedom (SDoF) theory, so that a pre-processing step is mandatory to separate modal contributions.

In OBMA, Polymax model has typically been implemented as identification support [40, 42]. However,

Polymax does not seem to be the best candidate for such an identification, because parametric models always generate spurious modes (due to noise and numerical bias). Furthermore, the model order is always difficult to define (methods are based on stabilization diagrams or parsimony principle through the minimization of criteria, *e.g.* Akaike and Bayesian Information Criterion, AIC or BIC). In order-tracked signals, it is easy to know in advance the number of excited modes, which allows using non-parametric methods. Thus, the authors propose to perform the following procedure: different modes are decomposed into SDoF responses and bandlimited using a partial E-FDD procedure. Then, each mode is identified using a classical ambient SDoF transfer function with a maximum likelihood estimator.

4.1 SDoF Separation

SDoF separation is performed using a modal coherent criterion applied on the singular vectors of the discrete spectral density matrix. The classical input-output relation of density matrices under the condition of white-noise input, low damping and uncoupled modes, can be developed according to the Heaviside partial-fraction expansion theorem in the vicinity of its modal pulsations [52, 53].

$$\boldsymbol{E}_k|_{\omega_k \approx \omega_r} \approx \boldsymbol{\Phi}^* diag\left[\frac{\beta_r}{(\xi_r \omega_r)^2}\delta_r\right]\boldsymbol{\Phi}^T \tag{1}$$

Eq. 1 shows that the excitation density matrix is diagonal, and thus the output density matrix E is equivalent to a diagonal one. The change of basis is done with the modal matrix $\boldsymbol{\Phi} = (\boldsymbol{\varphi}_1, ...\boldsymbol{\varphi}_n)$. The diagonal matrix contains only one non-zero term expressed as the contribution of the investigated mode through $(\beta_r, \xi_r, \omega_r)$, respectively a characteristic scalar value, the damping ratio and the natural pulsation of mode r. The Kronecker symbol δ_r is 1 for the $r-th$ position of the diagonal matrix, else 0. In the vicinity of a natural pulsation, the associated mode is the only contributor to the global dynamic of the system. The associated mode shape is $\boldsymbol{\varphi}_r$, $r-th$ column of $\boldsymbol{\Phi}$. In other terms, it is shown that the Singular Value Decomposition of the experimental density matrix in the vicinity of a mode returns only one dominant singular value. The associated singular vector in $\omega_k = \omega_r$ (within the limit of frequency resolution) is the mode shape estimator. The set of the first singular values $\{s_{1,k}\}$ is called Complex Modal Identification Function (CMIF) and is a unilateral representation of the previous spectral density functions. The resonance function of each SDoF is identified from the CMIF thanks to a discriminating criterion, the Modal Assurance Criterion (MAC) [54]. MAC varies from 0 to 1 as the modal coherence increases. It compares the degree of agreement of two vectors:

$$MAC(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) = \frac{<\boldsymbol{\varphi}_i|\boldsymbol{\varphi}_j>^2}{<\boldsymbol{\varphi}_i|\boldsymbol{\varphi}_i><\boldsymbol{\varphi}_j|\boldsymbol{\varphi}_j>} \tag{2}$$

This criterion is able to separate two uncoupled close modes and discriminate spike noises. Brinker et *al.* set the threshold to $MAC_{thr} = n/\sqrt{N_S}$, where n is an integer so that MAC_{thr} is close but lower than 1, and N_S the number of studied sensors [55]. If $MAC(\widehat{\boldsymbol{\varphi}_r}, \boldsymbol{\varphi}_j) > MAC_{thr}$, where $\widehat{\boldsymbol{\varphi}_r}$ is the shape estimator and $\boldsymbol{\varphi}_j$ a singular vector of the CMIF, then $s_1[\omega_j]$ belongs to the resonance function of the SDoF. This ensures to select a bandwidth with high modal coherence. Figure 1 shows an example of the use of MAC. The E-FDD theory shifts back in time domain to make the identification. But this procedure is a bad damping estimator, especially in the case of short signals [56]. For this reason, the identification support is different and presented in the next subsection.

A last point can be raised about the E-FDD limits: this procedure is only proper to separate uncoupled signals. In the case of coupled modes, it is unable to differentiate modal contributions. In future works, a Frequency-Domain Blind Source Separation developed by Castiglione et *al.* should be used instead [57]. FD-BSS is able to separate coupled modes with an impressive accuracy, and relies on a more rigorous mathematical approach.

4.2 Identification Using Maximum Likelyhood

After being extracted using AD method and bandlimited with MAC, the N_S experimental frequency responses are concatenated into a vector $\widehat{\boldsymbol{X}}_k$ and are modelled with the classical SDoF response described in eq. (3, 4), where $\boldsymbol{X}_k$ is the theoretical response vector, $h_{r,k}$ is the modal transfer function depending on modal parameters (ω_r, ξ_r) and $p_k, \boldsymbol{\varepsilon}_k$ are the normalized Fourier transforms of excitation and noise respectively.

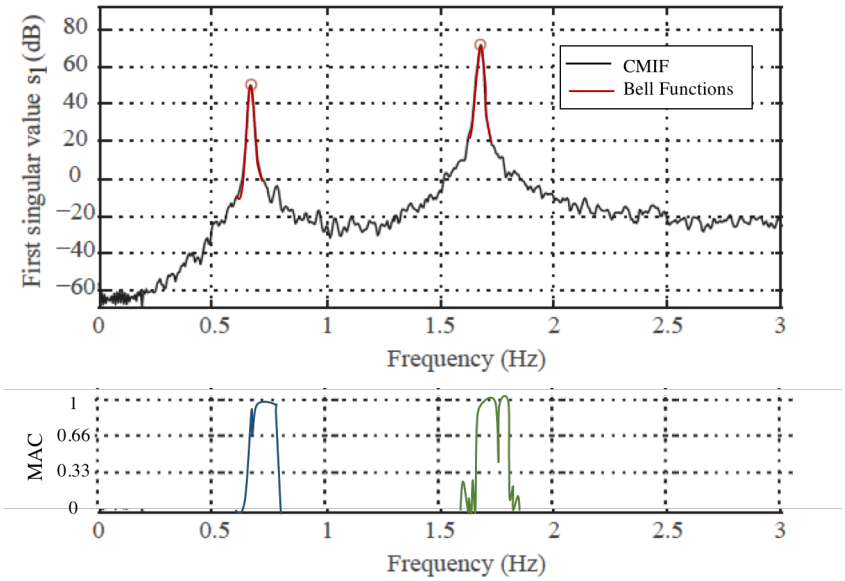


Figure 1: Modal contributions are framed using MAC, which compares the agreement level between two mode shapes.

$$\boldsymbol{X}_k = \boldsymbol{\varphi}_r h_{r,k} p_k + \boldsymbol{\varepsilon}_k \tag{3}$$

$$h_{r,k} = -\frac{1}{\omega_k^2 - \omega_r^2 - 2i\xi_r\omega_r\omega_k} \tag{4}$$

The associated Negative Log-Likelihood Function (NLLF) is given in eq. (5), where N_f is the number of point per channel, $\boldsymbol{E}_k[\boldsymbol{\theta}]$ the theoretical SDoF density matrix arising from eq. (3) and $|\boldsymbol{E}_k[\boldsymbol{\theta}]|$ the determinant of the density matrix ; the analytical determination of both determinant and inverse matrix of $\boldsymbol{E}_k[\boldsymbol{\theta}]$ is far from being trivial, and described in [58]. $\boldsymbol{\theta}_r = (\omega_r, \xi_r, S_r, S_{er}, \boldsymbol{\varphi}_r)$ is the parameter vector, including natural pulsation, damping ratio, modal force, PSD error and mode shape. $\boldsymbol{\theta}$ is the parameter variable, used to estimate $\boldsymbol{\theta}_r$. The NLLF is minimised using a Nelder-Mead algorithm [59]. A such identification method was chosen because it shows the best asymptotic properties.

$$\mathscr{L}(\boldsymbol{\theta}) = N_S N_f ln(\pi) + \sum_{k=1}^{Nf} ln(|\boldsymbol{E}_k[\boldsymbol{\theta})]|) + \sum_{k=1}^{N_f} \widehat{\boldsymbol{X}}_k^* \boldsymbol{E}_k^{-1}[\boldsymbol{\theta}] \widehat{\boldsymbol{X}}_k \tag{5}$$

5 Case Study

The studied measurements come from a vertical medium head Francis hydroelectric runner exploited in Quebec, Canada. This facility was chosen because the turbine was designed and is operated by two partners of the current project. The measurement data were recorded during a slow transient from no-load overspeed to stop. Two blades separated with an angle of 111° were instrumented with strain gauges. Intrados were instrumented with three strain gauge rosettes, located in the band junction to blade leading edge and trailing edge, and in the middle crown-blade weld, as shown in Figure 2. Extrados were instrumented with two uniaxial gauges, one close to the crown, the other close to the band. The locations are the same from one blade to another to ensure a redundant signal. Accelerometers and pressure sensors are located in different points (blade, structure and penstock) and sensors are also installed on the shaft to record torque, flexion and thrust. The rosette and uniaxial gauges are oriented in agreement with the expected strain flow direction, *.e.* in the direction of the principal stresses.

An analysis of experimental correlograms (amplitude of short-time Fourier transforms) and absolute phase-shift spectra between redundant sensors was made first. Some examples are shown in Figures 3 and 4, related to

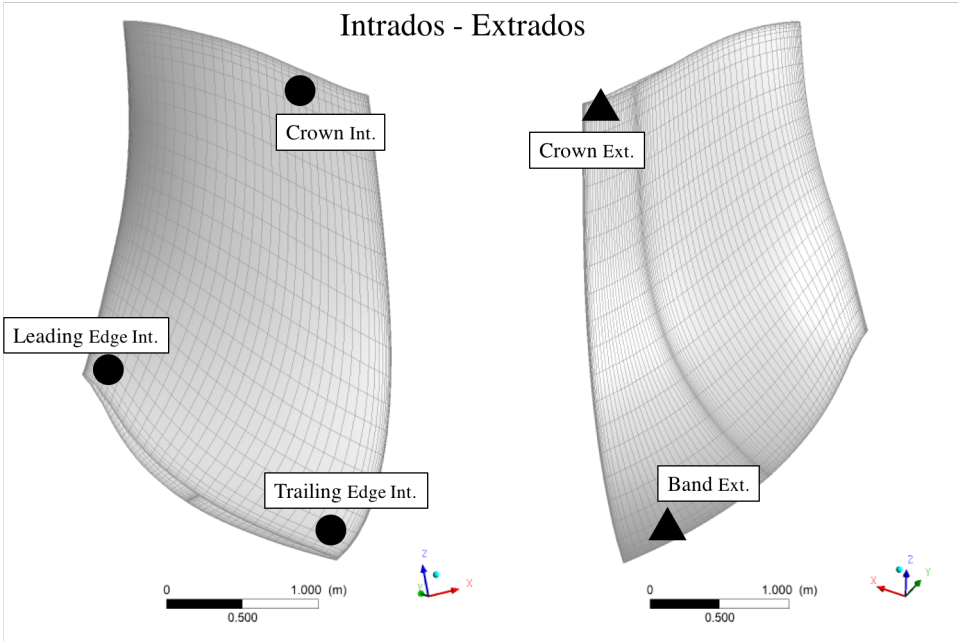


Figure 2: Strain gauge location on the instrumented blades. Circles represent rosette gauges, triangles represent uni-axial gauges.

both sides of the crown (time and frequency axes are empty for the purpose of clarity). When the windowing is long enough, correlograms show five ODS. Several modes can be contained inside. A clear resonance of mode 1 is detected on the intrados in the lower part of the diagram (below blade passing frequency signature, abusively denoted "RSI") ; the other resonances are in the upper part. Phase-Shift spectra are amplitude-filtered, and show only phases associated with a high enough correlogram. They show four single-mode bands and a multi-mode band, due to multiple phase-shift detection. Resonance of mode 2 is detected on both sides, and mode 3 is vaguely detected on the extrados. Only one mode of the multi-shifted band is excited by a harmonic, thus leading to a SDoF resonance. Table 1 summarizes all the detected modes in the range $[0, 100]Hz$, and reports the related resonant harmonic index. Phase-shifts are averaged over the region where the mode is found.

In the studied data, all the time series are recorded with an extensively large sampling frequency, and the use of AD technique to extract harmonics should be straightforward. The next subsections present typical case studies based on previous identified modes.

Mode Reference	Frequency [Hz]	Phase-Shift $\times\pi$[rad]	Harm. Order
1	18.0	0	13
2	28.0	$\pm$3/4	42
3	50.0	$\pm$1/7	63
4	91.0	$\pm$1/9	$\emptyset$
5	58.0	$\pm$6/7	61

Table 1: Experimentally Detected Modes

5.1 Identification Example: Shaft Torsion Mode with $f_0 = 18Hz$

The first mode to be studied, mode 1, is excited by the $13-th$ harmonic of the rotating speed. This corresponds to the blade passing frequency. Such an excitation can come for instance from the spiral case intake or the drat tube elbow that can create a stationary disturbance that is seen by the rotating runner each time a blade passes in front of the intake or the draft tube direction. The investigation of torsion measurements shows that the studied mode is a natural torsion mode of the shaft line. What is observed on blades is only the propagation of shaft natural vibrations. Thus, all the runner is excited with the same phase, and the nodal diameter is 0, that

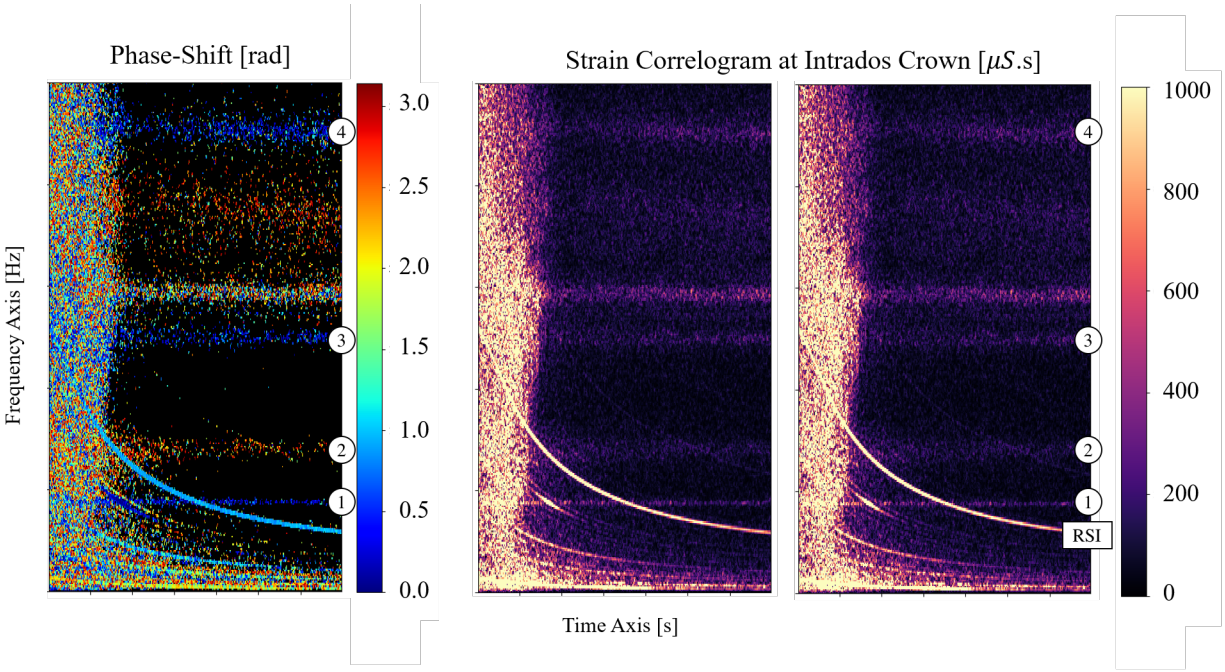


Figure 3: ODS Analysis of principal direction of intrados crown Rosette gauge. On the left, phase-shift spectrum of the redundant gauges. On the right, redundant amplitude spectra.

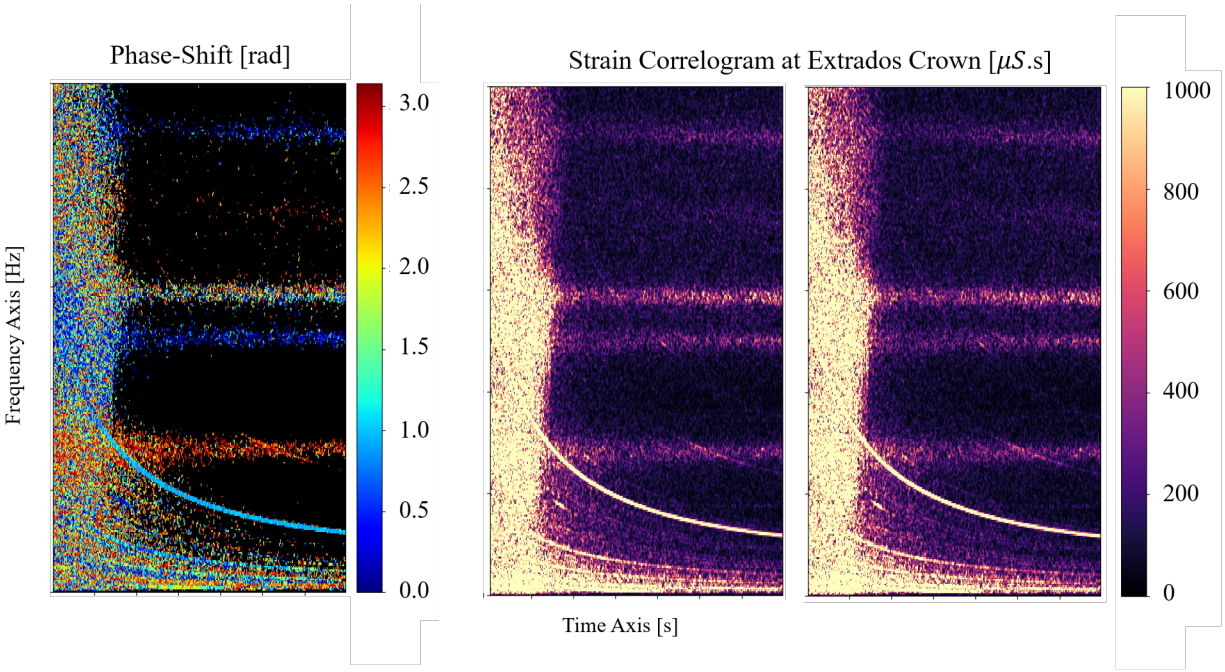


Figure 4: ODS Analysis of extrados crown Rosette gauge.

is confirmed by the absence of phase-shift between blades. An axial thrust pulsation is measured on the shaft, and indicates that the inflow to the runner is not symmetric to the guide vane orientation [31]. Examples of extracted resonances are shown in Figure 5. The *MAC* narrows the bandwidth with a threshold $MAC_{thr} = 0.875$, as depicted in Figure 6. The maximum likelihood estimator raises optimal parameters, shown in Table 2 and Figure 7.a.

Natural Frequency	Damping Ratio	Modal Force	PSD Error
f_0 [Hz]	ξ [%]	S [ms^2/Hz]	S_e [μ S/Hz]
17.43	1.26	2.22E6	1.73E-2

Table 2: Torque Mode Featuring

The shape relative amplitudes are the same on the two blades, as testifies Figure 7.a.. The mode shape is in phase opposition from leading edge to crown, and is not spotted neither on trailing edge intrados or on band extrados signals. That attests a $ND-0$ "in umbrella", as depicted in Figure 7.b. The modal force is very difficult to extract and is likely very biased. The bias on damping ratio mainly depends on experimental data. The leakage and tapering bias due to the windowing is avoided thanks to COT-based AD. However, the global uncertainty level remains likely high because of the unknown excitation.

5.2 Results

Table 3 shows the result of the identification process performed on all detected resonance harmonics. The information presented is: the exciting harmonic (indexed on the rotating frequency), the most likely nodal diameter, the Signal-to-Noise Ratio (SNR), the bandwidth and the associated method (*MAC* or SENS for sensitivity analysis) and the modal parameters. Except for the torsion mode, resonance signals have a low SNR that renders impossible the use of *MAC*, because the singular value spectrum is still buried in noise. Instead, a sensitivity analysis was made on modal parameters as a function of the bandwidth. The selected band corresponds to the parameter convergence. This method gives wider bands (around ten times the width of a *MAC* selected bandwidth), where noise has a significant influence. Rather, *MAC* criterion selects a narrower frequency band with very few noise. The two methods return quite equivalent results. The experimental data reveals five isolated modes numbered from 1 to 4 in Figure 4, and a multi-mode band. Amongst the 4 well-separated modes, only the first 3 are excited by a harmonic and then identifiable. Into the multi-mode band, one mode is excited by a harmonic. It is thus possible to feature it, but the SNR is particularly low. Notice that the first mode of table 3 is the torsion mode featured in table 2.

6 Conclusion

This paper shows that Francis runner structural modes can be identified from ambient vibration data during transient conditions. These modes have been successfully extracted and identified through an enhanced OBMA technique (E-OBMA). E- OBMA combines three existing techniques and takes benefit from the best of each: Order Tracking separation quality, MAC bandlimiting rigor and maximum likelihood accuracy. This work shows that experimental transient data contains accurate frequency information that can be used to assess numerical model validity. The presented results are the first effort in creating OMA strategy tailored for Francis runners. The E-OBMA still has to be validated on an analytical case, which is now being developed. The Order Tracking quality should be evaluated in conjunction with EMA sine-sweep excitation theory. Also, further improvements will make possible the uncertainty quantification which is a major stake in signal processing.

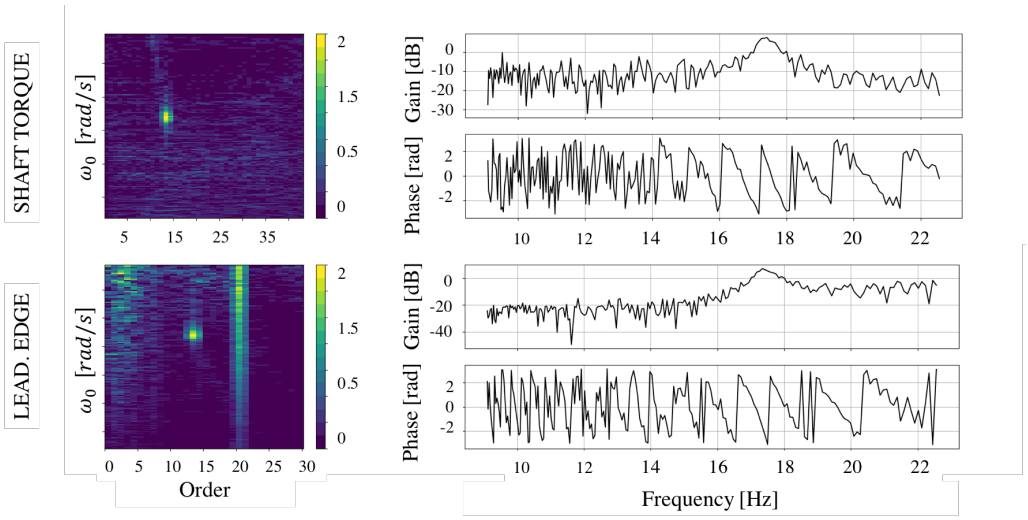


Figure 5: On the left, Order Spectra. All the harmonic contents are on the same line. On the right, the related Bode Diagrams of the harmonic 13.

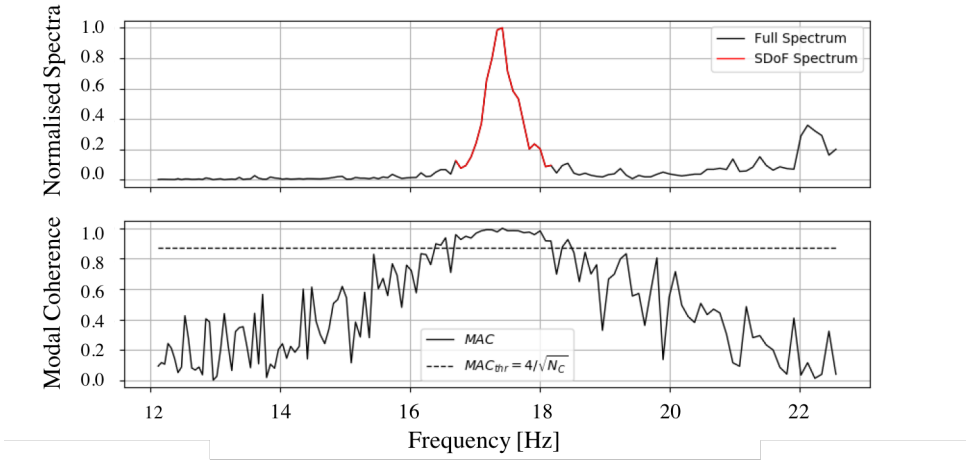


Figure 6: Resonance function extraction from the first singular values spectra (CMIF).

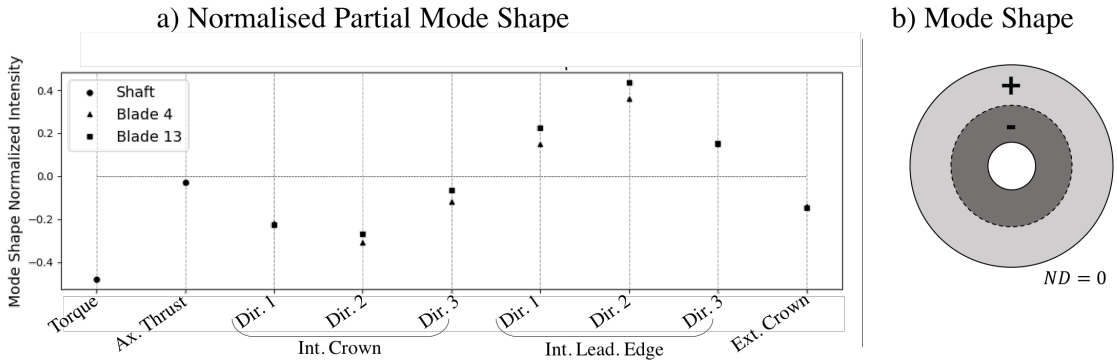


Figure 7: a) Partial mode shape extracted with maximum likelihood. b) Schematics of the observed mode shape.

		Identification BandWidth	Modal Parameters	
Mode	Nodal Diameters	Bandwidth (Hz)	Frequency (Hz)	Modal Force (ms^2/Hz)
Harm. Index	SNR (dB)	Method	Damping ratio (%)	PSD Error (μ s/Hz)
1	0	1.5	17.43	2.22E6
13	7.4	MAC	1.26	1.73E-2
2	1	8	28.71	3.31E5
42	1.8	SENS	3.26	1.18E-3
3	3/5	9	49.84	1.40E6
63	1.9	SENS	1.67	8.00E-4
Multi.	6	12	59.35	4.31E6
61	0.96	SENS	2.30	2.00E-3

Table 3: OBMA Identification Results

References

[1] A. Presas, Y. Luo, Z. Wang, and B. Guo. Fatigue life estimation of francis turbines based on experimental strain measurements: Review of the actual data and future trends. *Renewable and Sustainable Energy Reviews vol. 102*, page 96 to 110, March 2019.

[2] J. Nicolle, J.F. Morissette, and A.M. Giroux. Transient cfd simulation of a francis turbine startup. *IOP Conf. Ser.: Earth Environ. Sci. vol. 15 062014*, October 2012.

[3] J. Nicolle, A.M. Giroux, and J.F. Morissette. Cfd configurations for hydraulic turbine startup. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032021*, September 2014.

[4] T. Krappel, A. Ruprecht, S. Riedelbauch, R. Jester-Zuerker, and A. Jung. Investigation of francis turbine part load instabilities using flow simulations with a hybrid rans-les turbulence model. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032001*, March 2014.

[5] B. Nennemann, J.F. Morissette, J. Chamberland-Lauzon, and C. Monette. Challenges in dynamic pressure and stress predictions at no-load operation in hydraulic turbines. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032055*, 2014.

[6] J.F. Morissette, J. Chamberland-Lauzon, B. Nennemann, C. Monette, A.M. Giroux, A. Coutu, and J. Nicolle. Stress predictions in a francis turbine at no-load operating regime. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072016*, July 2016.

[7] B. Nennemann, C. Monette, and J. Chamberland-lauzon. Hydrodynamic damping and stiffness prediction in francis turbine runners using cfd. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072006*, page 854 to 865, July 2017.

[8] A. Presas, D. Valentin, E. Egusquiza, C. Valero, and U. Seidel. Influence of the rotation on the natural frequencies of a submerged-confined disk in water. *Journal of Sounds and Vibrations vol. 337*, page 161 to 180, February 2015.

[9] C. Trivedi and M.J. Cervantes. Fluid-structure interactions in francis turbines: A perspective review. *Renewable and Sustainable Energy Reviews vol. 68*, page 87 to 101, February 2017.

[10] A. Presas, D. Valentin, E. Egusquiza, C. Valero, M. Egusquiza, and M. Bossio. Accurate determination of the frequency response function of submerged and confined structures by using pzt-patches. *Sensors*, March 2017.

[11] A. Ducoin, J. Andre-Astolfi, and J.-F. Sigrist. An experimental analysis of fluid structure interaction on a flexible hydrofoil in various flow regimes including cavitating flow. *European Journal of Mechanics - B/Fluids vol. 36*, page 63 to 74, December 2012.

[12] A. Lelong, P. Guiffant, and J.A. Astolfi. An experimental analysis of the structural response of flexible lightweight hydrofoils in cavitating flow. *Journal of Fluids Engineering vol. 140*, November 2017.

[13] X. Liu, Y. Luo, A. Presas, Z. Wang, and L. Zhou. Cavitation effects on the structural resonance of hydraulic turbines: failure analysis in a real francis turbine runner. *Energies vol. 11*, September 2018.

[14] D. Valentin, D. Ramos, M. Bossio, A. Presas, E. Egusquiza, and C Valero. Influence of the boundary conditions on the natural frequencies of a francis turbine. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072004.*, November 2016.

[15] M. Gagnon, A. Tahan, P. Bocher, and D. Thibault. On the stochastic simulation of hydroelectric turbine blades transient response. *Mechanical Systems and Signal Processing vol. 32*, page 178 to 187, October 2012.

[16] M. Poirier, M. Gagnon, A. Tahan, A. Coutu, and J. Chamberland-Lauzon. Extrapolation of dynamic load behaviour on hydroelectric turbine blades with cyclostationary modelling. *Mechanical Systems and Signal Processing vol. 82*, page 193 to 205, January 2017.

[17] I. Diagne, M. Gagnon, and A. Tahan. Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072014*, July 2016.

[18] E. Moisan, D.-B. Giacobbi, M. Gagnon, and F. Leonard. Self-excitation in francis runner during load rejection. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032025*, September 2014.

[19] M. Gagnon, A. Tahan, A. Coutu, and M. Thomas. Oma en presence d'excitations harmoniques : etude de cas sur des composantes de turbine hydroelectrique. *24th Seminar on Machinery Vibration*, October 2006.

[20] D. Valentin, A. Presas, M. Bossio, and M. Egusquiza Montagut. Feasibility to detect natural frequencies of hydraulic turbines under operation using strain gauges. *Sensors*, January 2018.

[21] A. Presas, D. Valentin, E. Egusquiza, and C. Valero. Detection and analysis of part load and full load instabilities in a real francis turbine prototype. *J. Phys.: Conf. Ser. 813 012038*, April 2017.

[22] D. Liu, X. Lui, and Y. Zhao. Experimental investigation of inter-blade vortices in a model francis turbine. *Chinese Journal of Mechanical Engineering vol. 30*, February 2016.

[23] R. Goyal and B.K. Gandhi. Review of hydrodynamics instabilities in francis turbine during off-design and transient operations. *Renewable Energy vol. 116*, page 697 to 709, October 2017.

[24] X. Liu, Y. Luo, and Z. Wang. A review on fatigue damage mechanism in hydro turbines. *Renewable and Sustainable Energy Reviews vol. 54*, February 2016.

[25] A. Gagnon, A. Tahan, P. Bocher, and D. Thibault. Influence of load spectrum assumptions on the expected reliability of hydroelectric turbines: A case study. *Structural Safety, vol. 50*, March 2014.

[26] X. Huang, J. Chamberland-Lauzon, C. Oram, A. Klopfer, and N. Ruchonnet. Fatigue analyses of the prototype francis runners based on site measurements and simulations. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 012014*, 2014.

[27] U. Seidel, C. Mende, B. Hubner, W. Weber, and D. Otto. Dynamic loads in francis runners and their impact on fatigue life. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032054*, 2014.

[28] C. Trivedi, B. Gandhi, and M.J. Cervantes. Effect of transients on francis turbine runner life: a review. *Journal of Hydraulic Research vol. 51-2*, page 121 to 132, 2013.

[29] F. Leonard. Spectrogramme de phase et spectrogramme de frequence. *Traitement de signal vol. 17 nb 4*, page 269 to 286, 2000.

[30] F. Leonard. Phase spectrogram and frequency spectrogram as new diagnostic tools. *Mechanical Systems and Signal Processing vol. 21*, page 125 to 137, January 2007.

[31] P. Dorfler, M. Sick, and A. Coutu. Flow-induced pulsation and vibration in hydroelectric machinery. *Engineer's Guidebook for Planning, Design and Troubleshooting*, April 2012.

[32] G. Blommaert. Etude du comportement dynamique des turbines francis : controle actif de leur stabilite de fonctionnement. *These de l'Université Polytechnique Federale de Lausanne*, 2000.

[33] A. Favrel, C. Landry, A. Muller, and F. Avellan. Experimental identification and study of hydraulic resonance test rig with francis turbine operating at partial load. *IOP Conf. Ser.: Earth Environ. Sci. vol. 15 062064*, August 2012.

[34] A. Favrel, C. Landry, A. Muller, K. Yamamoto, and F. Avellan. Hydro-acoustic resonance behavior in presence of a precessing vortex rope: observation of a lock-in phenomenon at part load francis turbine operation. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032035*, September 2014.

[35] C. Nicolet, N. Ruchonnet, S. Alligne, J. Koutnik, and F. Avellan. Hydroacoustic simulation of rotor-stator interaction in resonance conditions in francis pump-turbine. *IOP Conf. Ser.: Earth Environ. Sci. vol. 12 012005*, 2010.

[36] H. Tanaka. Vibration behavior and dynamic stress of runners of very high head reversible pump-turbines. *International Journal of Fluid Machinery and Systems vol.4*, page 289 to 306, June 2011.

[37] J. Blough, D. Brown, and H. Vold. The time variant discrete fourier transform as an order tracking method. *SAE Noise and Vibration Conf. and Exp.*, May 1997.

[38] P. Borghesani, P. Pennachi, S. Chatterton, and R. Ricci. The velocity synchronous discrete fourier transform for order tracking in the field of rotating machinery. *Mechanical Systems and Signal Processing vol. 44*, page 118 to 133, February 2014.

[39] D.B. Stephens and H. Vold. Order tracking signal processing for open rotor acoustics. *Journal of Sounds and Vibration vol. 333*, page 3818 to 3830, August 2014.

[40] E. Di Lorenzo. Operational modal analysis for rotating machines, challenges and solutions. *PhD - KL Arenberg Doct. School and Frederico II Naple Univ.*, April 2017.

[41] K.R. Fyfe and D.S. Munck. Analysis of computed order tracking. *Mechanical Systems an Signal Processing vol. 11*, page 187 to 205, March 1997.

[42] K. Janssens, Z. Kollar, B. Peeters, S. Pauwels, and H. Van der Auweraer. Order-based resonance identification using operational polymax. *ResearchGate Article*, January 2006.

[43] E. Di Lorenzo, S. Manzato, A. Dabizzi, B. Peeters, F. Marulo, and W. Desmet. Industrial applications of advanced modal identification on operational rotating machineries. *Proceeding of ISMA*, page 2833 to 2848, September 2016.

[44] E. Di Lorenzo, A. Palermo, S. Manzato, A. Dabizzi, B. Peeters, W. Desmet, and F. Marulo. Gear dynamics characterization by using order-based modal analysis. *Proceedings of the International Modal Analysis Conference (IMAC)*, page 387 to 404, 2016.

[45] H. Van der Auweraer, P. Mas, B. Peeters, K. Jannsens, and A. Vecchio. Modal and path contribution models from in-operation data: Review and new approaches. *Shock and Vibration vol. 15*, page 403 to 411, 2008.

[46] A.G. Poulimenos and S.D. Fassois. Parametric time-domain methods for non-stationary random vibration modelling and analysis : A critical survey and comparison. *Mechanical Systems and Signal Processing vol. 20*, page 764 to 816, January 2006.

[47] L. Ljung. System identification. *University of Linkoping, Sweden*, 1998.

[48] R Brinker and C.E. Ventura. Introduction to operational modal analysis. *Wiley Editions*, 2015.

[49] B. Peeters, H. Van der Auweraer, P. Guillaume, and J. Leuridan. The polymax frequency-domain method: a new standard for modal parameter estimation? *Shock and Vibration vol. 11*, page 395 to 409, August 2004.

[50] R. Brinker, L. Zhang, and P. Andersen. Modal identification from ambiant responses using frequency domain decomposition. *Proceedings of the International Modal Analysis Conference (IMAC)*, page 625 to 630, February 2000.

[51] R. Brinker, L. Zhang, and P. Anderser. Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures vol. 10*, page 441 to 445, January 2001.

[52] L. Zhang, T. Wang, and Y. Tamura. A frequency-spatial domain decomposition (fsdd) method for operational modal analysis. *Mechanical Systems and Signal Processing vol. 24*, page 1227 to 1239, July 2010.

[53] C.-X. Qu, T.-H. Yi, H.-N. Li, and B. Chen. Closely spaced modes identification through modified frequency domain decomposition. *Measurement vol. 128*, page 388 to 392, November 2018.

[54] M. Pastor, M. Binda, and T. Harcarik. Modal assurance criterion. *Procedia Engineering vol. 48*, page 543 to 548, 2012.

[55] R. Brinker, P. Andersen, and N.-J. Jacobsen. Automated frequency domain decomposition for operational modal analysis. *Proceedings of the International Modal Analysis Conference (IMAC)*, 2007.

[56] M. Danial, A. Hasan, Z.A.B. Ahmad, M. Salman Leong, and L.M. Hee. Enhanced frequency domain decomposition algorithm: a review of a recent development for unbiased damping ratio estimates. *Journal of Vibro-Engineering vol. 20*, page 1919 to 1936, August 2018.

[57] R. Castiglione, J. Antoni, and L. Garibaldi. Separation and identification of structural modes in largely underdetermined scenarios using frequency banding. *Journal of Sound and Vibration, vol. 414*, page 192 to 217, February 2018.

[58] S.-K. Au. Operational modal analysis, modeling, bayesian inference, uncertainty laws. *Springer*, 2017.

[59] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal vol. 7*, page 308 to 313, 1965.

A new method for identifying diagnostic rich frequency bands under varying operating conditions

Stephan Schmidt[1] Alex R. Mauricio[2] P. Stephan Heyns[3] Konstantinos C. Gryllias[4]

[1]Centre for Asset Integrity Management, University of Pretoria, Pretoria, South Africa
stephan.schmidt@up.ac.za

[2]Department of Mechanical Engineering, PMA Division, KU Leuven, Leuven, Belgium
Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium
alex.ricardomauricio@kuleuven.be

[3]Centre for Asset Integrity Management, University of Pretoria, Pretoria, South Africa
stephan.heyns@up.ac.za

[4]Department of Mechanical Engineering, PMA Division, KU Leuven, Leuven, Belgium
Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium
konstantinos.gryllias@kuleuven.be

Abstract

Performing condition monitoring under time-varying operating conditions is challenging. The varying operating conditions impede the ability of conventional fault diagnosis methods to detect damage on rotating machine components such as bearings and gears. This paper investigates a new method for identifying diagnostic rich frequency bands under time-varying operating conditions. This method uses the order-frequency spectral coherence and a feature, which is dependent on the cyclic order of interest and the frequency resolution of the spectral coherence, to decompose the signal into a feature plane. Thereafter, the spectral frequency and the spectral frequency resolution that maximise the feature plane are used to design a bandpass filter. The bandpass filter extracts a diagnostic rich signal, which can be analysed by using the squared envelope spectrum or the synchronous average. The proposed method is compared to the fast kurtogram on a numerical gearbox dataset as well as on an experimental gearbox dataset, with very promising results obtained.

1 Introduction

Effective fault diagnosis techniques are important for expensive assets such as wind turbines, because this can result in early detection of faults, their characteristics can easily be understood (e.g., which component is damaged) and subtle changes in the damage (i.e. deterioration) can be monitored. Many rotating machines inherently operate under time-varying operating conditions, which impede effective fault diagnosis. Hence, it is important to use condition monitoring techniques that are able to diagnose damaged machine components under time-varying operating conditions.

Damaged rotating machine components such as bearings result in periodical excitations of the structure at a rate dependent on the kinematic characteristics of the component (e.g. ball pass order of the outer race, shaft rotation). This angle-dependent periodical excitation of the time-invariant structure generates signals that can be approximated as angle-time cyclostationary [1]. Abboud et al. [1, 2, 3] extended the suite of conventional time and angle cyclostationary techniques to time-varying speed conditions with tools such as the Order-Frequency Spectral Coherence (OFSCoh) being one of the most powerful fault diagnosis techniques for bearings under varying speed conditions.

However, in condition monitoring it is usually desired to utilise simple metrics or representations for making decisions (e.g. a spectrum is preferred instead of a time-frequency spectrum). Hence, the enhanced envelope spectrum and the even more powerful Improved Envelope Spectrum (IES), both calculated from the spectral coherence or the spectral corelation, can be used to diagnose the machine. For the IES, it is very important to select carefully the integration band to ensure that the IES has an optimal signal-to-noise ratio. This means that it is important to be able to identify frequency bands that are rich with diagnostic information. Identifying diagnostic rich frequency bands is also important for calculating the synchronous average and the squared envelope spectrum [4].

The spectral kurtosis and the related kurtogram are effective for identifying frequency bands with much impulsive information [5, 6]. This is very appropriate for diagnostics, because bearing damage [4, 5] and gear damage [7] result in vibration signals containing bandlimited impulses. However, the kurtogram is sensitive to transients not related to the condition of the machine and it is not possible to investigate the optimal frequency band to detect damage associated with a specific cyclic order. Recently, new methods such as the infogram [8] and the IESFOgram [9] have been proposed for identifying frequency bands that are rich with diagnostic information by improving the shortcomings of the kurtogram.

A new method is investigated in this paper that is able to identify a frequency band that contains diagnostic information related to a specific machine component under time-varying operating conditions. This has a significant advantage over conventional methods, because incipient damage components that are normally masked by other dominant signal components and distorted by time-varying operating conditions, can be extracted from the signal and used to diagnose the machine. The performance of this method is compared to the Fast Kurtogram on numerical gearbox data as well as on experimental gearbox data, both acquired under time-varying operating conditions.

The outline of this paper is as follows: In Section 2, the proposed method is presented, whereafter it is investigated on phenomenological gearbox data in Section 3 and experimental gearbox data in Section 4. In the last section, Section 5, some conclusions are extracted and some recommendations are made for future investigations.

2 Methodology

2.1 Overview of the methodology

An overview of the methodology is presented in Figure 1. The measured vibration signal and the corresponding rotational speed (or phase) is given as inputs, whereafter an Order-Frequency Spectral Coherence (OFSCoh) is calculated for a specific window length. A feature is extracted from each frequency band of the calculated OFSCoh. This process is repeated for the set of window lengths under consideration, whereafter a feature plane is constructed. The feature plane contains the value of the feature for different combinations of centre frequencies and window lengths (or frequency resolutions). Thereafter, the feature plane is maximized to obtain the parameters of a bandpass filter. This bandpass filter is used to extract a signal that is rich with diagnostic information from the original signal, whereafter the filtered signal can be analysed to infer the condition of the machine component.

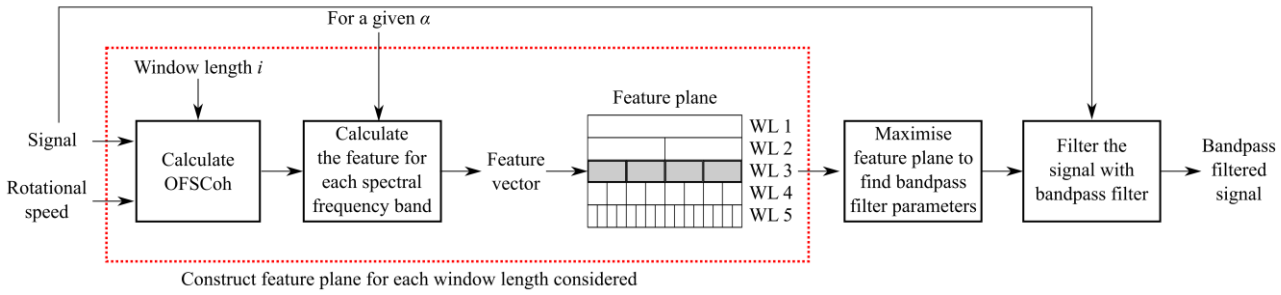


Figure 1: The proposed method for identifying frequency bands that are rich with diagnostic information.

The subsequent sections give detailed information on each step in the proposed method.

2.2 Order-Frequency Spectral Coherence (OFSCoh)

The impulses generated by components such as bearings are periodic in the angle domain, while they manifest in the time-invariant frequency bands. This means that the OFSCoh can be used to identify the resonance bands that are excited at specific cyclic orders. The OFSCoh [2]

$$\gamma_{xx}(\alpha, f) = \frac{S_{xx}(\alpha, f)}{\left(S_{xx}(0, f) S_{x_\alpha x_\alpha}(0, f)\right)^{1/2}} \tag{1}$$

provides a two-dimensional view of the modulating frequencies (i.e. cyclic orders) and their carriers (i.e. spectral frequencies) in the signal *x(t)*. The Order-Frequency Spectral Correlation (OFSC) [2]

$$S_{xx}(\alpha, f) = \lim_{W \to \infty} \frac{1}{\Phi(W)} E\left\{F_W\left(x(t)\right)^* F_W\left(x(t) e^{-j\alpha\theta(t)} \dot{\theta}(t)\right)\right\} \tag{2}$$

is used to calculate the OFSCoh in Equation (1). The expectation operator is denoted *E*, the Fourier transform is denoted F_w and $\Phi(W)$ denotes the phase of the shaft during the measurement time period *W*. The instantaneous phase of the shaft is denoted θ. It is easier to detect non-dominant components by using the OFSCoh as opposed to the OFSC.

Estimators need to be used to calculate the OFSCoh for the measured data, with the Welch estimator as proposed in Ref. [2], used in this work. The Welch estimate of the OFSCoh is denoted $\gamma_{xx}(\alpha, f; \Delta f)$, where Δf is the frequency resolution that is used to obtain the estimate.

2.3 Frequency Band Identification (FBI)

It is possible to use a one-dimensional metric such as the kurtosis to identify the frequency band of interest. However, one-dimensional metrics do not allow different signal components to be distinguished from one another, which may result in a frequency band to be identified that is not necessarily of interest. Hence, a more advanced metric is required.

2.3.1 Feature extraction

Ref. [4] uses a metric to quantify the quality of the Squared Envelope Spectrum (SES). If their metric is large, it means that the diagnostic information is dominant with respect to the noise level in the SES, while a small metric indicates that it could be difficult to detect the cyclic components in the SES. The authors estimated the noise level with the median because the median is robust to outliers generated by the cyclic components in the SES.

We used this metric as inspiration for designing the feature to identify the frequency band of interest, with the following feature obtained for the cyclic order set $\{\alpha_f\}$:

$$\Psi_{xx}(f, \Delta f; \{\alpha_f\}) = \sum_{\{\alpha_f\}} \frac{|\gamma_{xx}(\alpha_f, f; \Delta f)|^2}{median\left(|\gamma_{xx}(\alpha_f, f; \Delta f)|^2\right)} \tag{3}$$

The numerator contains the squared magnitude of the spectral coherence for a specific window length Δf. The denominator contains the median function, which is calculated for the squared magnitude of the spectral coherence and is used to estimate the noise level in the OFSCoh. The following points are important considerations when calculating the feature for practical signals:

1. The analytical cyclic orders may be different from the actual cyclic orders due to slip and therefore the maximum of a range of $[0.9\alpha_f, 1.1\ \alpha_f]$ is calculated to estimate the numerator.
2. The median of the squared magnitude OFSCoh cannot be calculated at $\alpha=\alpha_f$ and therefore it needs to be estimated from the discrete OFSCoh data. Hence, the median of the squared magnitude of the OFSCoh in the range of $[\alpha_f - 1, \alpha_f + 1]$ is used to estimate the denominator.

This feature also has similarities to the feature used by the IESFOgram [9]. In the latter method the ratio of the signal components in the IES are calculated with respect to the mean of the IES in the predefined bandwidth.

2.3.2 Feature plane construction and maximisation

The feature is calculated for each frequency band in the OFSCoh. The Welch estimator of the OFSCoh depends on a number of parameters, namely, the window length, the window overlap as well as the number of points used to calculate the FFT. It is best to use an overlap longer than 75% of the window length, however, the window length needs to be determined prior to the analysis. It is also necessary to estimate the frequency bandwidth and not only the centre frequency for designing the bandpass filter parameters. Hence, the following procedure is used to simultaneously optimise the centre frequency and frequency bandwidth of the frequency band of interest: Firstly, the OFSCoh is calculated for a specific window length, whereafter the feature is calculated for each spectral frequency band in the OFSCoh. This process is repeated for each window length under consideration, whereafter the feature plane is obtained. The frequency band parameters are identified by finding the centre frequency and frequency bandwidth that maximise the feature plane. This is a very similar procedure to the kurtogram and the infogram, but instead of using the short-time Fourier transform, the OFSCoh is used, and instead of maximising a scalar value (e.g. spectral kurtosis), the maximisation is done for a set of cyclic orders. This allows the optimal frequency band to be determined to detect a set of cyclic orders.

The identified frequency band parameters can be used to calculate the IES or to extract a bandlimited signal. In this work, we used the frequency band parameters to design a bandpass filter, whereafter the bandpass filtered signal is interrogated. The bandpass filtered signal can subsequently be analysed with techniques such as the Synchronous Average (SA) [10] and the Squared Envelope Spectrum (SES) [3].

2.4 Computational aspects

Even though real-time condition monitoring is rarely required in practice, it is still necessary to provide answers in a reasonable time. The Welch-based estimator of the OFSCoh has very good bias and variance properties, but is very expensive to calculate for large datasets, especially for high rotational speed applications. If the cyclic orders of interest are known a priori, it is possible to only estimate the OFSCoh for specific cyclic orders; however, even this may be impractical for complex gearboxes found in wind turbines and helicopters, which may have many cyclic orders of interest. Fortunately, there has been very exciting developments in this field, where fast (and faster) estimators of the spectral correlation are proposed, which could make this method significantly faster to be calculated [11, 12].

3 Numerical gearbox data

In this section, we investigate the method and compare it to the kurtogram on data generated from a phenomenological gearbox model. In the next section, an overview is given of the model and the generated data, whereafter the Fast Kurtogram (FK) is used on the dataset in Section 3.2. The results of the proposed method are presented and discussed in Section 3.3.

3.1 Phenomenological Gearbox Model (PGM)

The Phenomenological Gearbox Model (PGM) proposed in Ref. [3] is used to generate a casing vibration signal. The casing vibration signal

$$x(t) = x_b(t) + x_{rg}(t) + x_n(t) \tag{4}$$

contains a bearing component $x_b(t)$, a random gear component $x_{rg}(t)$ and a broadband noise component $x_n(t)$. The generalised synchronous average can be used to attenuate the deterministic gear components attributed to the meshing of gears as described by Abboud et al. [3] and therefore they are not included in this model. The bearing component is generated by bearing damage on the outer race

$$x_b(t) = M(\omega(t)) \cdot h_b(t) \otimes \sum_{k=1}^{K_b} A_k \cdot \delta(t - T_k) \tag{5}$$

where T_k denotes the time-of-arrival of the kth bearing impulse, which incorporates the varying speed conditions and the slip. The amplitude of the kth impulse, denoted A_k, is sampled from a uniform distribution. The raw bearing impulses are filtered through the structure, which is assumed to have an impulse response

function of a single degree-of-freedom system h_b. The modulating function $M(\omega(t)) = \omega^2$ is used to simulate the varying amplitude induced by time-varying operating conditions and is assumed to be the same for all signal components for the sake of simplicity.

The random gear component

$$x_{rg}(t) = M(\omega(t)) \cdot h_{rg}(t) \otimes \left(\varepsilon(t) \cdot \sum_{k=1}^{Krg} B_k \cdot \sin\left(k \cdot \int_0^t \omega(t)dt + \varphi_k \right) \right) \tag{6}$$

is attributed to gear damage and contains the random variable $\varepsilon(t)$ which is sampled from a zero mean, unit variance normal distribution, and B_k and φ_k are, respectively, the amplitude and the phase of the kth harmonic of the component. There are K_{rg} harmonics in the vibration signal. The noise component

$$x_n(t) = M(\omega(t)) \cdot \varepsilon(t) \tag{7}$$

is generated by a zero mean Gaussian distribution with its amplitude dependent on the rotational speed of the system. The natural frequency of the impulse response function of the bearing and the gear components are 7 kHz and 1.3 kHz respectively. The fundamental cyclic order of the distributed gear damage is 1.0 shaft order, while the fundamental cyclic order of the outer race bearing damage component is 4.12 shaft orders.

A single dataset is investigated in this paper with the time-varying speed profile $\omega(t)$ and the different signal components shown in Figure 2. This system operates under constant load conditions.

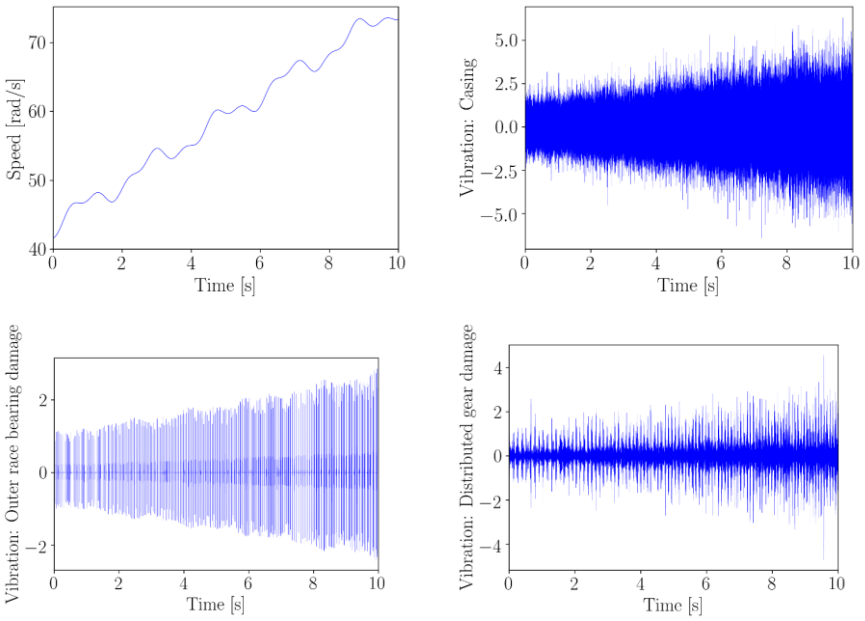


Figure 2: The speed profile, the casing vibration signal and the bearing and random gear components of the phenomenological gearbox model are presented.

The varying speed conditions result in the amplitude and the instantaneous frequency of the signal components to be dependent of time. The relative magnitudes of the components were chosen so that the dominant distributed gear damage component impedes the ability to detect the bearing component. Hence, the focus of the subsequent investigations is to highlight how the proposed method can be used to detect weak components in the presence of dominant components and to show that it is possible to distinguish between the two. In the next section the kurtogram is investigated on the generated dataset.

3.2 Application of the Fast Kurtogram (FK)

The Fast Kurtogram (FK), developed in Ref. [6], is a faster estimator of the kurtogram than the conventional short-time Fourier transform-based estimator and is used in this work. The kurtogram is based on the spectral kurtosis [5], a very useful technique to identify frequency bands that contain transient information (as typically seen by bearing and gear damage). The FK is applied to the casing vibration signal (see Equation (4)) of the PGM with the result shown in Figure 3.

The FK is maximum at a frequency band with a centre frequency of 1328.12 Hz. This is the frequency band associated with the distributed gear damage component. The frequency band of the bearing damage at 7.0 kHz can also be seen in Figure 3; however, its magnitude is significantly smaller than the magnitude of the gear component.

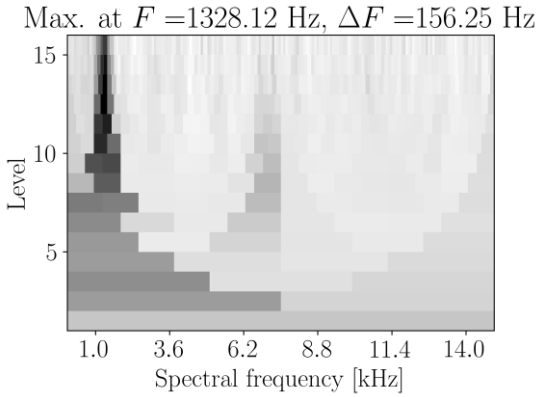


Figure 3: The kurtogram of the PGM's vibration signal.

The implication of this is that without careful consideration, only the dominant impulsive frequency band will be detected by the FK, with a non-dominant frequency band easily missed in the condition interrogation process.

This is corroborated by the results of the Squared Envelope Spectrum (SES) seen in Figure 4. The SES of the raw signal (i.e. without bandpass filtering the signal) and the SES of the filtered signal contains the same information. The fundamental component of the distributed gear damage at one shaft order and its harmonics are clearly seen in both spectra, while the bearing component is not seen.

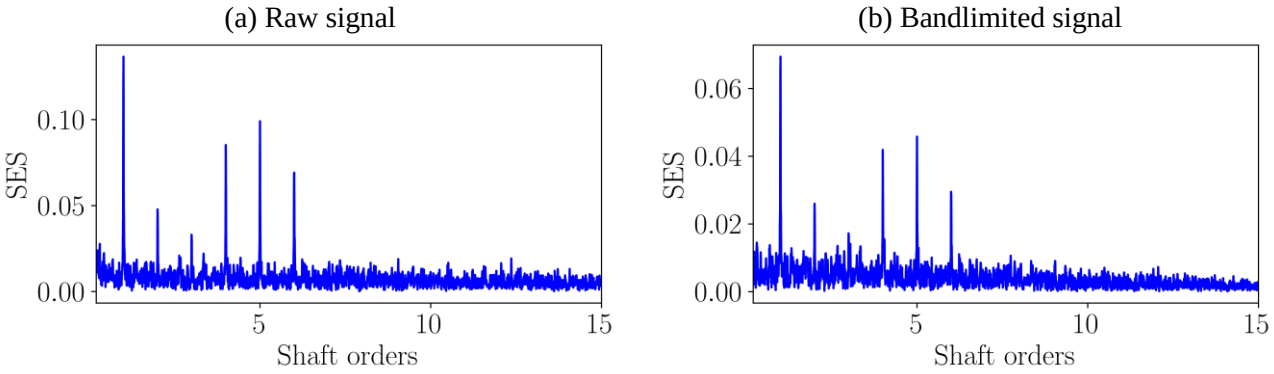


Figure 4: The Squared Envelope Spectrum (SES) of the raw vibration signal and of the bandlimited signal obtained with the Fast Kurtogram (FK) for the PGM.

It is important to emphasise that due to the statistical characteristics of the distributed gear damage component, it is not possible to remove it using cepstrum pre-whitening or the generalised synchronous average [3]. The proposed method is investigated in the next section.

3.3 Application of the proposed method

The proposed method is applied with the procedure discussed in Section 2, with the bearing and gear being monitored for damage. Therefore, the feature, calculated with Equation (3), is calculated for the gear with $\{\alpha_f\}=\{1.0, 2.0, 3.0\}$ (denoted α=1.0 in the figures) and for the bearing with $\{\alpha_f\}=\{4.12, 8.24, 12.36\}$ (denoted α=4.12 in the figures), which result in two feature planes that are maximised independently. The feature plane of the gear and the bearing are shown in Figure 5(a) and Figure 5(b) respectively.

It is evident that the feature plane is clearly very dependent on the cyclic order that is used. Large values are obtained in Figure 5(a) in the region of 1.3 kHz, while large values are obtained in Figure 5(b) in the

region of 7 kHz. The optimal value for the gear in Figure 5(b) differs slightly from the analytical value, because the gear component is very dominant, which results in the different blocks to have features with very similar values, i.e. any of the blocks, could be used for detecting the gear.

The SES of the raw and bandlimited signals of the two signal components are shown in Figure 6. The SES of the bandlimited gear signal, presented in Figure 6(b), does not improve the SES of the raw signal, presented in Figure 6(a), because the gear component is already very dominant in the SES.

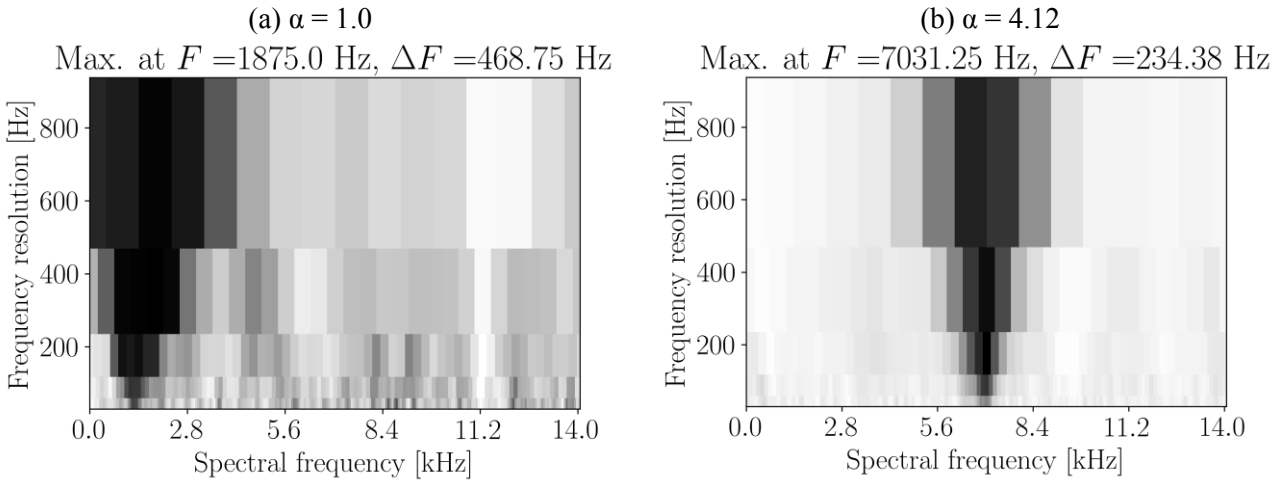


Figure 5: The feature plane obtained with the proposed method for the gear component (a) and the outer race bearing component (b) of the PGM. The colour scales are not the same in the two plots.

A significant improvement can be seen for the SES of the bearing component. The bearing component cannot be detected in Figure 6(c), but after identifying the appropriate frequency band with the proposed method, it is possible to obtain a SES that clearly highlights the damaged bearing component as seen in Figure 6(d).

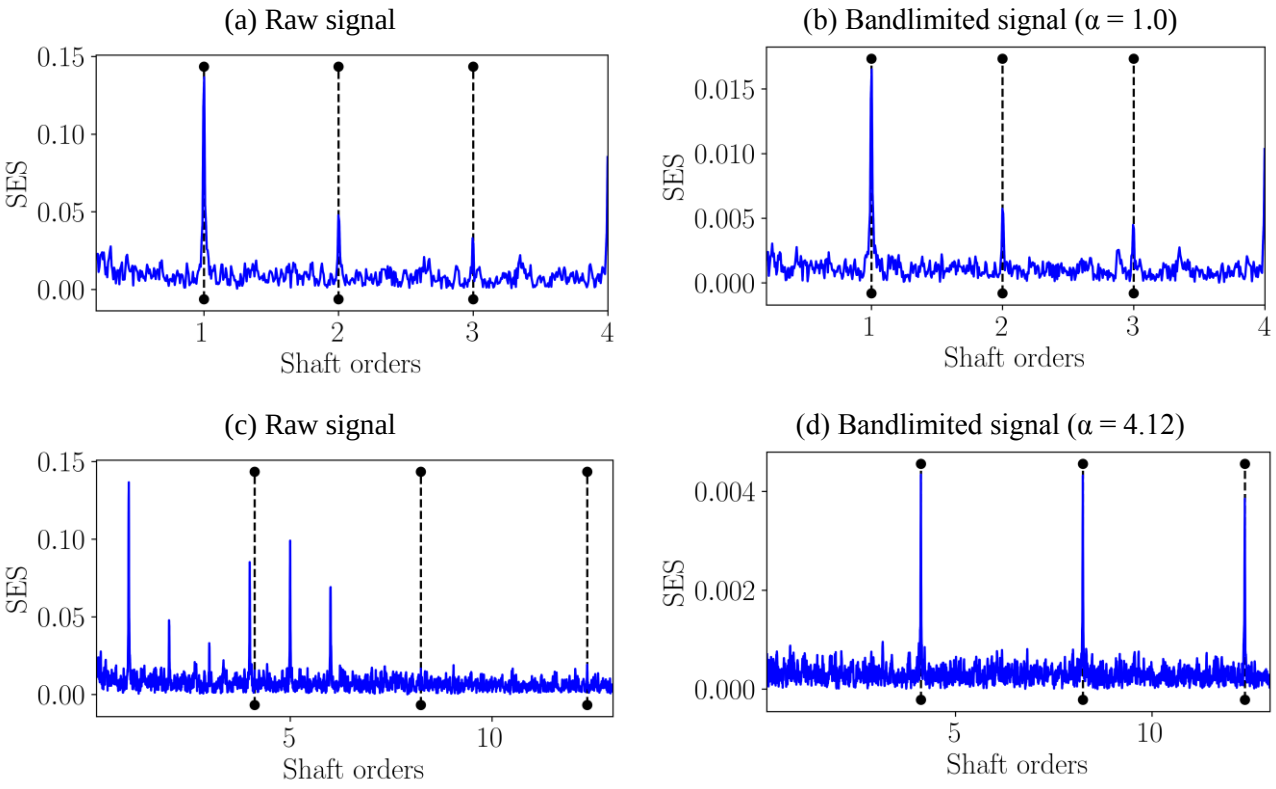


Figure 6: The Squared Envelope Spectra (SES) of the raw and bandlimited signals are shown for the gear component in (a) and (b) and for the outer race bearing component in (c) and (d) for the PGM.

This highlights the benefit of using the proposed method; if the signal component is dominant in the spectrum then the kurtogram can lead to similar results (as seen when comparing the results in Figure 4(b) and Figure 6(b)). However, the proposed method has sufficient flexibility to identify frequency bands for signals with low signal-to-noise ratios as well.

4 Experimental investigation

In this section, the proposed method is investigated on an experimental dataset. A brief overview of the experimental data is given in Section 4.1, whereafter the FK is applied to the dataset in Section 4.2 and the proposed method is investigated in Section 4.3.

4.1 Overview of the experimental dataset

The method is applied and verified in this section on an experimental gearbox dataset that has been acquired in the Centre for Asset Integrity Management (C-AIM) laboratory at the University of Pretoria. The experimental setup contains three helical gearboxes, an alternator and an electrical motor. The alternator and the electrical motor were used to induce the time-varying speed and load conditions shown in Figure 7 on the monitored gearbox. One of the helical gearboxes was damaged with the damaged gear shown in Figure 8(a) and operated for approximately 20 days whereafter the tooth failed as shown in Figure 8(b). A vibration and a tachometer measurement, taken after approximately five days of testing, are used in this paper. The gear rotates at 1.0 shaft order, while the pinion rotates at 1.85 shaft order. More information on the experimental setup can be found in Ref. [13].

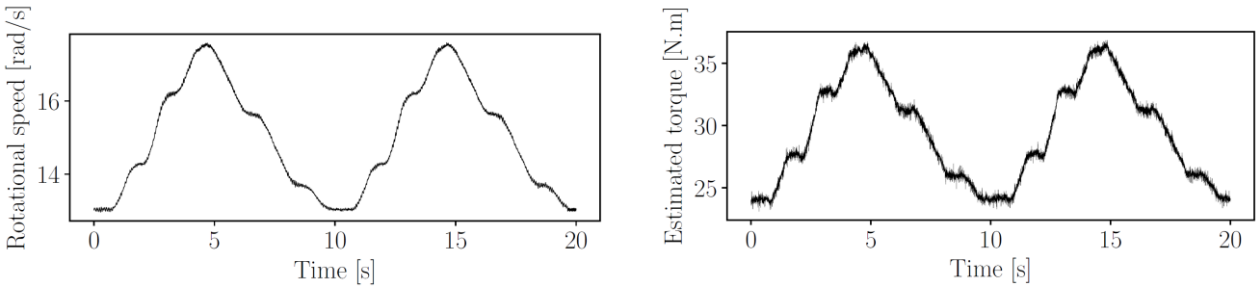


Figure 7: The operating conditions during the measurement period.

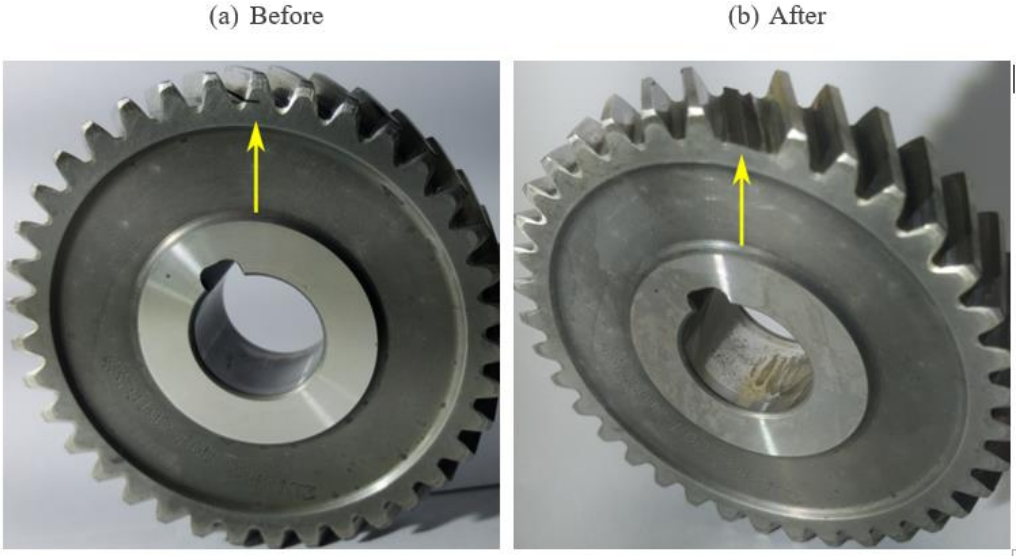


Figure 8: The gear of the helical gearbox with the seeded fault before the fatigue experiment (a) and after the fatigue experiment was completed (b).

4.2 Application of the Fast Kurtogram (FK)

The FK is applied on the dataset with the decomposition shown in Figure 9. Very large values are seen in the higher frequency bands. This is attributed to the presence of bandlimited transients that manifest at the frequency band 8-12 kHz at a cyclic order of approximately 5.5 shaft orders.

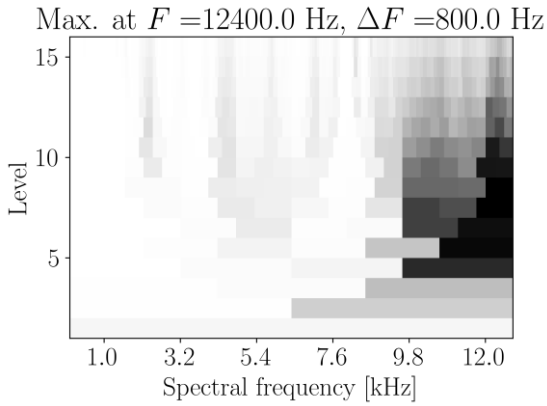


Figure 9: The kurtogram of the experimental gearbox dataset.

The SA is used to interrogate the presence of damage on the gear in Figure 10. The SA of the raw and the bandlimited signals are shown in Figure 10(a) and (b). It is not clear from the raw signal in Figure 10(a) what the condition of the gear is, but the transients that are retained by the bandpass filtering process dominate the synchronous average and make it especially difficult to infer the condition of the machine from the result in Figure 10(b).

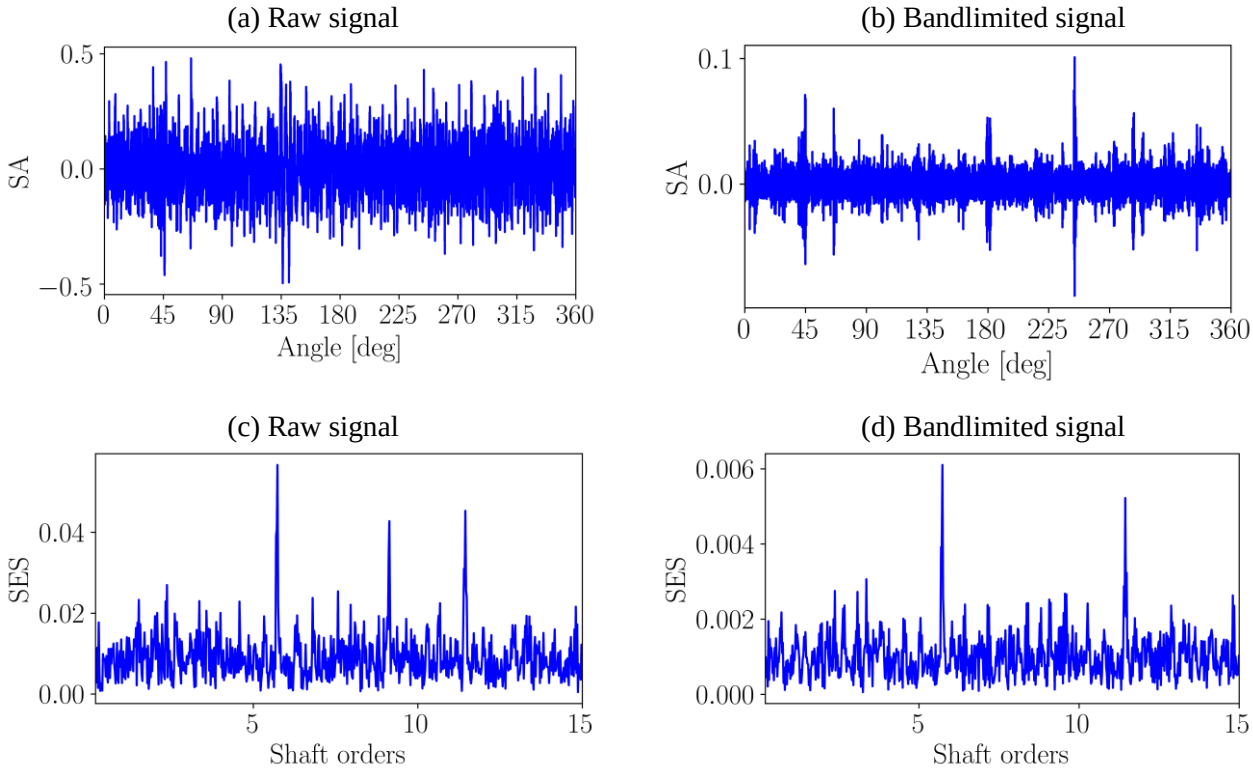


Figure 10: The Synchronous Average (SA) and the Squared Envelope Spectrum (SES of the raw and the bandlimited signals are shown as obtained with the Fast Kurtogram (FK). The damaged gear tooth is located at approximately 135 degrees in the SA plots.

The SES of the raw and the bandlimited signals are also investigated in Figure 10. Three peaks are observed in the SES of the raw signal; the components at 5.72 and 11.44 shaft orders are attributed to the

transients in the signal and the component at 9.12 shaft orders is attributed to the alternator's shaft being slightly unbalanced which resulted in periodical excitations. After, the filtering process, only the transient at 5.72 shaft orders and its harmonics are retained. Hence, it is evident from the results that the kurtogram fails to recognise the important frequency band for diagnosing the gear.

4.3 Application of the proposed method

The proposed method is applied on the same signal as investigated in the previous section. The gear and the pinion are monitored and therefore the decomposition is performed for $\{\alpha_f\}=\{1.0, 2.0, 3.0\}$ shaft orders (denoted $\alpha=1$) and $\{\alpha_f\} = \{1.85, 3.7, 5.55\}$ shaft orders (denoted $\alpha=1.85$), respectively. The feature plane is shown in Figure 11 for the two monitored components, where it can be seen that the feature planes are dependent on the cyclic order of interest, however, the identified frequency bands may not necessarily be completely separated. It is completely reasonable that the same cyclic order band is optimal for different mechanical components and therefore care should be taken to interpret the statistics (e.g. kurtosis) of the bandlimited signals.

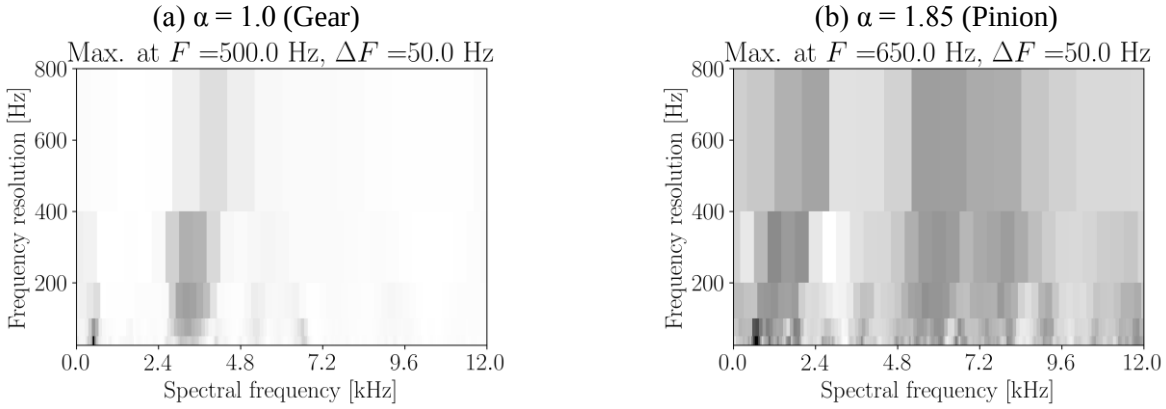


Figure 11: The feature plane obtained with the proposed method. The feature plane of the gear is shown in (a) and the feature plane of the pinion is shown in (b).

The SA in Figure 12 do not clearly reveal damage on either the gear or pinion with only small peaks seen at 135deg for the gear. This is attributed to the fact that the damage is still small and that helical gears are used with large contact ratios. Hence, the synchronous average is ineffective for detecting the incipient gear damage.

The SES of the raw and bandlimited signals in Figure 13 perform significantly better than the SA for the gear and the pinion. It is possible to see that there is a clear 1.0 shaft order component, which is attributed to the damaged gear. In contrast, the SES of the healthy pinion does not contain any dominant components at 1.85 shaft orders, which is indicative that the pinion is healthy. Hence, it is possible to use the proposed method and the SES to detect the incipient gear damage in the presence of dominant frequency components and time-varying operating conditions.

5 Conclusions

In this paper, a new method is investigated for identifying frequency bands that are rich with diagnostic information. The method uses the spectral coherence and a very carefully designed feature to allow specific frequency bands to be detected which can be analysed using the squared envelope spectrum and the synchronous average.

The method is evaluated on two datasets; the first one is a numerical gearbox dataset that simulates bearing damage and gear damage under time-varying speed conditions. The results indicate that it is possible to identify the appropriate frequency band to identify the cyclic components of interest, while the fast kurtogram only identifies the frequency band with the most impulsiveness. Similar results are obtained on the experimental dataset where incipient damage was present on the gear of a helical gearbox. The fast

kurtogram maximised on frequency bands with strong impulsive content, with the incipient gear damage only detected by using the proposed method. It was also found that the synchronous average is not very effective for incipient gear damage detection and the squared envelope spectrum performs significantly better.

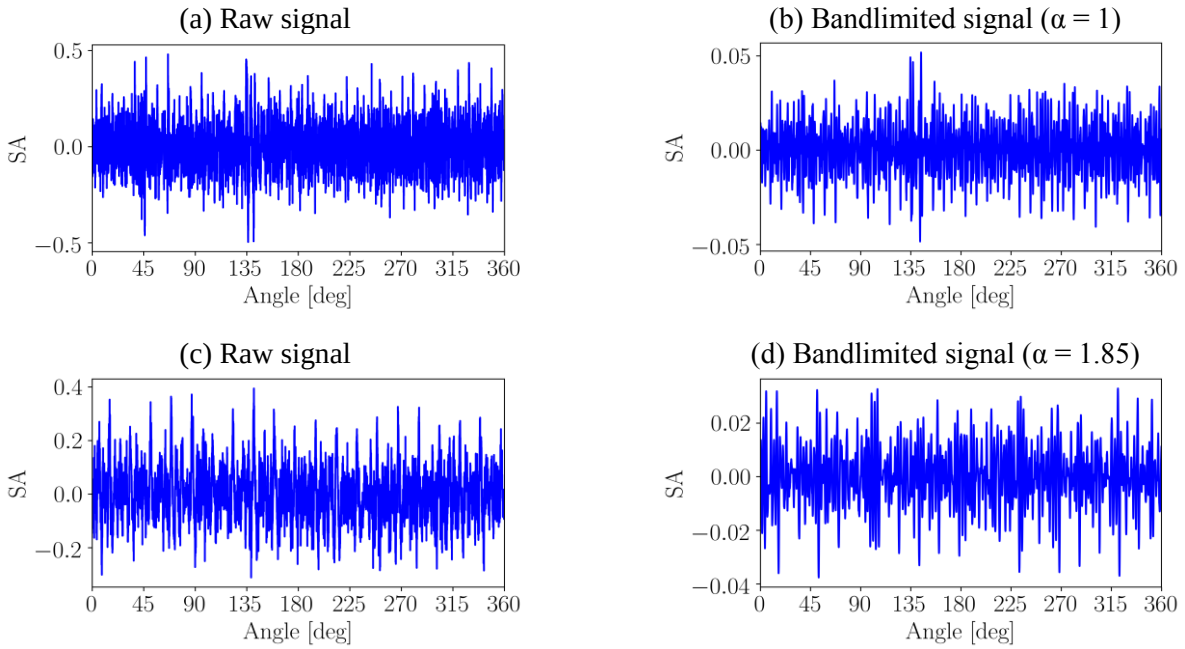


Figure 12: The Synchronous Averages (SA) of the raw and the bandlimited signals, obtained with the proposed method, are shown. The result for the gear is shown in (a) and (b), while the result of the pinion is shown in (c) and (d).

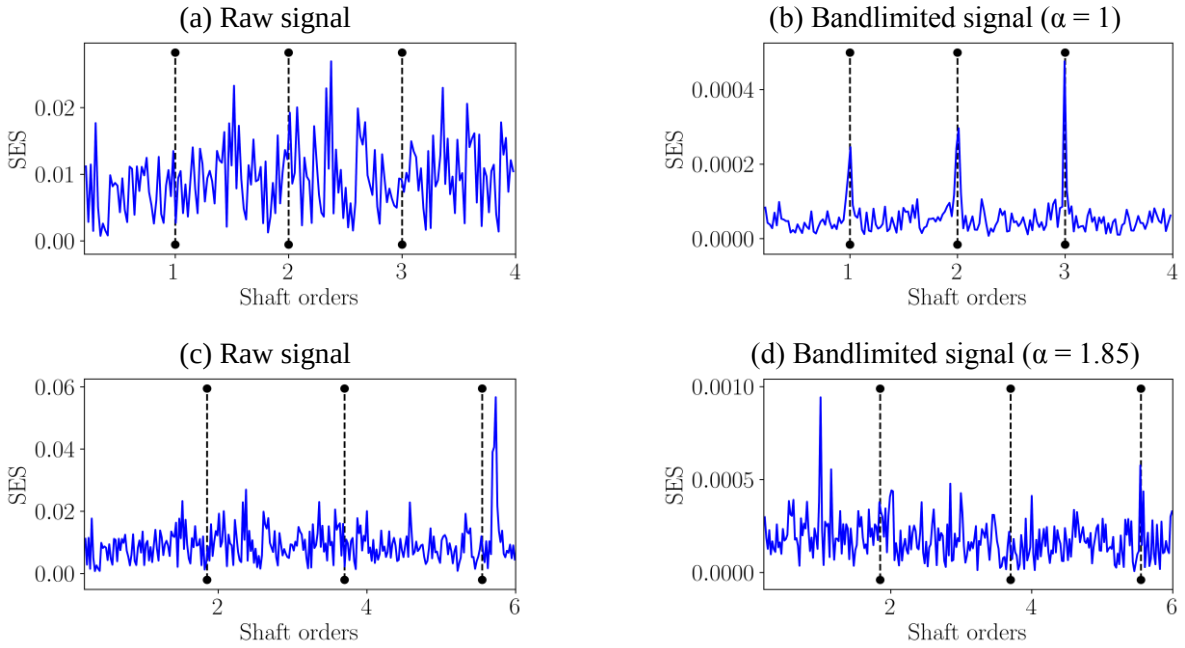


Figure 13: The Squared Envelope Spectra (SES) of the raw and the bandlimited signals, obtained with the proposed method are shown. In (a) and (b) the results for the gear are shown, while the results for the pinion are shown in (c) and (d).

In future investigations, the method will be compared to the more recent developments in the informative frequency band identification field (e.g. infogram) and the suitability of this method for fault diagnosis under time-varying operating conditions will be investigated on more datasets. It is also suggested that the spectral

coherence needs to be estimated with the fast or faster spectral correlation instead of the Welch estimator used in this work. This would improve the computational efficiency of the proposed method.

Acknowledgments

The authors gratefully acknowledge the support that was received from the Eskom Power Plant Engineering Institute (EPPEI) in the execution of the research.

References

[1] Abboud, D. & Antoni, J., 2017. Order-frequency analysis of machine signals. Mechanical Systems and Signal Processing, 87, pp.229–258.

[2] Abboud, D., Baudin, S., Antoni, J., Rémond, D., Eltabach, M. and Sauvage, O., 2016. The spectral analysis of cyclo-non-stationary signals. Mechanical Systems and Signal Processing, 75, pp.280-300.

[3] Abboud, D., Antoni, J., Sieg-Zieba, S. and Eltabach, M., 2017. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. Mechanical Systems and Signal Processing, 84, pp.200-226.

[4] Schmidt, S., Heyns, P.S. & Gryllias, K.C., 2019. A pre-processing methodology to enhance novel information for rotating machine diagnostics. Mechanical Systems and Signal Processing, 124, pp.541-561.

[5] Antoni, J. & Randall, R.B., 2006. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. Mechanical Systems and Signal Processing, 20(2), pp.308–331.

[6] Antoni, J., 2007. Fast computation of the kurtogram for the detection of transient faults. Mechanical Systems and Signal Processing, 21(1), pp.108–124.

[7] Barszcz, T. & Randall, R.B., 2009. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mechanical Systems and Signal Processing, 23(4), pp.1352–1365.

[8] Antoni, J., 2016. The infogram: Entropic evidence of the signature of repetitive transients. Mechanical Systems and Signal Processing, 74, pp.73–94.

[9] Mauricio A, Qi J, & Gryllias K. 2018. Vibration-based condition monitoring of wind turbine gearboxes based on cyclostationary analysis. ASME. Journal of Engineering Gas Turbines Power.

[10] Stander, C.J., Heyns, P.S. & Schoombie, W., 2002. Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions. Mechanical Systems and Signal Processing, 16(6), pp.1005–1024.

[11] Antoni, J., Xin, G. & Hamzaoui, N., 2017. Fast computation of the spectral correlation. Mechanical Systems and Signal Processing, 92, pp.248–277.

[12] Borghesani, P. & Antoni, J., 2018. A faster algorithm for the calculation of the fast spectral correlation. Mechanical Systems and Signal Processing, 111, pp.113–118.

[13] Schmidt, S., Heyns, P.S. & De Villiers, J.P., 2018. A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process. Mechanical Systems and Signal Processing, 100, pp.630-646.

Diagnostics and Dynamic models

Challenging the traditional model of gear vibration signals

E. Hubert[1], P. Borghesani[2], R.B. Randall[2], M. El-Badaoui[1]
[1]SAFRAN Tech, Rue des Jeunes Bois – Châteaufort, 78772 Magny les Hameaux, France
[2]School of Mechanical and Manufacturing Engineering, UNSW Sydney, NSW 2031, Australia

Abstract

Despite rarely being made explicit, signal models (and assumptions) for gear vibration have been fundamental in the development of both condition monitoring and operational modal analysis techniques. The analysis (for condition monitoring) or the removal (for OMA) of the dominant gear-meshing component is in fact dependent on the assumption on the number, location and patterns of the corresponding spectral harmonics. This paper discusses in detail the common modelling choices and their consequences for condition monitoring and OMA. The traditional gearmesh-carrier/shaft-modulated model is analysed and two main limitations of current models are highlighted: the additive assumption on the two gear modulating functions and the regularity of their effect on different gearmesh harmonics. The paper uses experimental gear signals to prove the validity of the newly introduced assumptions and to assess their practical significance.

1 Introduction

Gearbox condition monitoring has often been based on simple signal models. Empirical signal models are used as a first approximation of the vibration signal to justify and guide the development of diagnostic signal processing techniques. For this purpose, they are preferred to more detailed physical models (e.g. FEM or lumped parameters) due to their generality and ease of implementation. Empirical signal models aim at reproducing the overall properties of vibration signals, retaining the main time and frequency features that are observed in real signals and can be used for condition monitoring. Despite not requiring the fine-tuning of structural and geometric parameters typical of detailed physical models, correct assumptions on the phenomena generating the vibration are fundamental in developing appropriate empirical signal models.
In the case of spur gears, the main source of vibration comes from the time-varying meshing force generated at the contact point between the pinion and driven gear teeth. The most widely accepted gear-signal model is therefore represented as [1]

$$y(t) = h(t) \otimes g(t) \tag{1}$$

where the gear-related component $y(t)$ of a measured vibration signal results from the convolution (symbol $\otimes$) of the system impulse response $h(t)$ with the gear-meshing forcing function $g(t)$. As usual in rotating machines [2], the true nature of this signal has a hybrid time-angle definition. However, for nominally constant speed, the approximation of linearity between time and angular domain is often assumed. In this case, the easiest way to represent the signal in the frequency domain is probably to define (with approximation) the system transfer function $H(f) = \mathcal{F}\{h(t)\}$ in an equivalent shaft-order domain, adopting the approximate relationship $f \approx \Omega f_1$ where f_1 is the average (and almost constant) shaft speed of a reference gear (in this study we will always use the pinion/input shaft) and Ω is the order coordinate of the same shaft. In this case we can rewrite eq. (1) in the order domain as:

$$Y(\Omega) = H(\Omega f_1) \cdot G(\Omega). \tag{2}$$

In the case of perfect and healthy teeth, the contact force $g(t)$ is theoretically expected to show a gearmesh fundamental frequency (i.e. for a pinion with Z_1 teeth $G(\Omega) \neq 0$ only for $\Omega = kZ_1$), but in practice even imperfections in the manufacturing stage result in tooth-to-tooth variations. These variations are expected to be even more accentuated in the case of localised gear faults (e.g. tooth crack) and are modelled as gear-synchronous modulations of the gearmesh harmonics.

The excitation g is usually modelled in time or in the angular domain θ of the reference shaft (in this case shaft 1) as an amplitude/frequency modulated signal, where the carrier is represented by the dominant gearmesh harmonics and the two modulating functions are synchronous with the two shafts. Actual explicit mathematical expressions of AM/FM signal models are rare, and many studies focus on the simpler AM case only. In this paper we will mainly focus on the implications that arise for condition monitoring and OMA in considering an AM/FM model. A full analytical discussion will be provided for AM-only models, including limitations of common assumptions and further issues encountered when dealing with actual signals. However, considerations on FM and AM/FM models will be provided, without the explicit formulation of full AM/FM signal models (due to their cumbersome expression), but keeping in mind all the major and minor spectral components arising from the combination of all modulating functions.

2 Secondary sidebands

The simplest model of the gear force (often implicitly considered in many condition monitoring studies) includes two purely amplitude modulation components:

$$g(\theta) = c(\theta) \cdot \{a(\theta) + b(\theta)\} \tag{3}$$

where:

- $c(\theta)$ is the gearmesh-periodic dominant effect of the tooth-meshing

$$c(\theta) = \sum_h C_h e^{jhZ_1\theta} \tag{4}$$

  with Z_1 representing the number of teeth of the gear on shaft 1,

- $a(\theta)$ and $b(\theta)$ are a shaft-periodic amplitude modulation functions due to irregularities among the teeth of shaft 1 and 2 respectively (and/or geometric/misalignment issues on the same shaft)

$$a(\theta) = \sum_k A_k e^{jk\theta} \quad \text{and} \quad b(\theta) = \sum_k B_k e^{jk\tau\theta} \tag{5}$$

where $\tau = Z_1/Z_2$ is the gear ratio.
Such AM signal is usually represented as:

$$g(\theta) = \sum_h \sum_k C_h A_k e^{j(hZ_1+k)\theta} + \sum_h \sum_k C_h B_k e^{j(hZ_2+k)\tau\theta}. \tag{6}$$

This formulation shows the main feature of AM gear models: the presence of sidebands around the gearmesh harmonics at orders $hZ_1 + k$ (effect of shaft 1) and $hZ_1 + \tau k$ (effect of shaft 2).

Such a simplified model already poses risks for OMA. In fact, the removal of these harmonics is often considered straightforward by means of established synchronous averaging techniques, using encoders installed on both shafts (or at least on a reference shaft). In most cases, shaft-1 and shaft-2 sidebands are thus removed separately by synchronous averaging over the respective periods. Even in a number of highly rigorous approaches, only one or a few combined periods of the two gears are used for these synchronous averaging operations. This "grand-period" is defined as the interval between the meshing of the same tooth pair, and equivalent to Z_2 periods of shaft 1 or Z_1 periods of shaft 2.

However, considering the physical nature of the AM functions $a(\theta)$ and $b(\theta)$, a multiplicative model is much more justified, i.e. there is no reason why carrier and modulations should be treated differently and a three-term multiplication is more appropriate. This results in the modification of eq. (3) into the following:

$$g(\theta) = c(\theta) \cdot a(\theta) \cdot b(\theta) \tag{7}$$

with a consequent proliferation of sidebands:

$$g(\theta) = \sum_{h} \sum_{k} \sum_{\ell} C_h A_k B_\ell e^{j(hZ_1 + k + \tau\ell)\theta}. \tag{8}$$

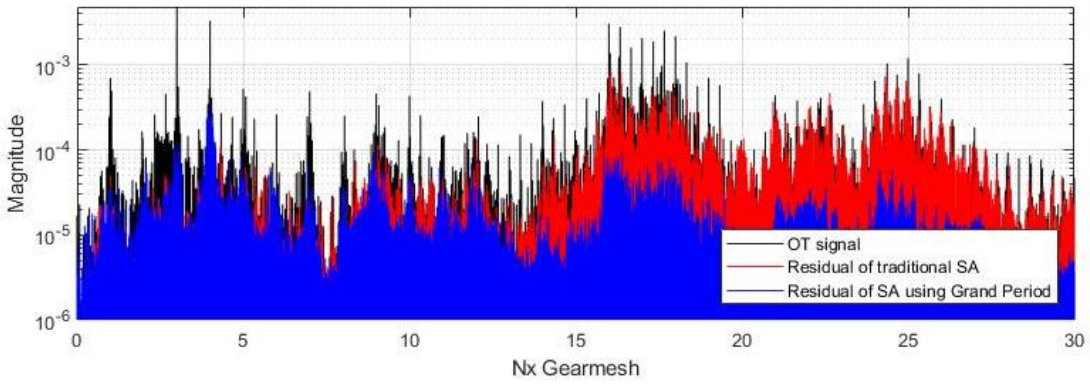
Under this modelling assumption – and actually also for model (3) –, the fundamental period of the signal is the "grand-period". However, differently from model (3) the signal shows a vast number of secondary sidebands $hZ_1 + k + \tau\ell$, with $k, \ell \neq 0$ (in addition to the $\ell = 0$ and $k = 0$ primary sidebands present also in the previous model). These secondary sidebands are expected to have (in the spectrum of the excitation) a lower amplitude (as the zero-frequency component of the modulation signals must be dominant to ensure positive-only modulating functions), but they could still significantly compromise OMA attempts based on the assumption that noise dominates the spectrum of the vibration signal, once the primary sidebands are removed.

An experimental test to verify the extent of this effect has been carried out on the UNSW spur-gear test-rig. The test-rig is composed of a speed-reducing spur gear pair ($Z_1 = 27$ and $Z_2 = 44$) powered by an electric drive and connected to a magnetic particle brake. The set of gears (module 2 with 5 mm face width) are built in mild steel, and surface hardened. A gear crack was simulated by means of an artificial slot on the pinion starting at the base of the tooth and reaching the centreline with an angle of 45°. The test rig was operated with constant speed and load (10 Hz /10 Nm on the input shaft) and a vibration signal was measured by means of a B&K4396 accelerometer, installed on the top of the casing in proximity of the DE input-shaft bearing. The signal was sampled at a rate of 100 kSamples/s for a duration of 101 s, sufficient to ensure the observation of 22 "grand-periods". In addition, a phase-reference signal was obtained synchronously to the vibration signal, thanks to an encoder with 1000 pulses/rev installed on the NDE of the pinion shaft.

The vibration signal was order-tracked ensuring that an integer number of samples N_{GM} was taken within a gearmesh period, thus also ensuring integer numbers of samples $Z_1 N_{GM}$ in a revolution of shaft 1 and $Z_2 N_{GM}$ in a revolution of shaft 2. Residuals were obtained following two synchronous averaging (SA) procedures: the first was the traditional approach which removed all harmonics of shaft 1 and shaft 2 (primary sidebands of the gearmesh harmonics), whereas the second used the entire grand-period as reference for the synchronous averaging, thus removing all primary and secondary sidebands.

The results are shown in Figure 1. The raw order-tracked spectrum clearly shows a significant amount of gearmesh harmonics and sidebands, which are only partly removed by the traditional SA approach. Whereas the low-frequency range (up to 10-15 gearmesh harmonics) seems mostly unaffected by the presence of secondary sidebands, in the range from 15 to 30 gearmesh harmonics the removal of the primary harmonics still leaves a large quantity of discrete components, which are identified as secondary sidebands.

(a)



(b)

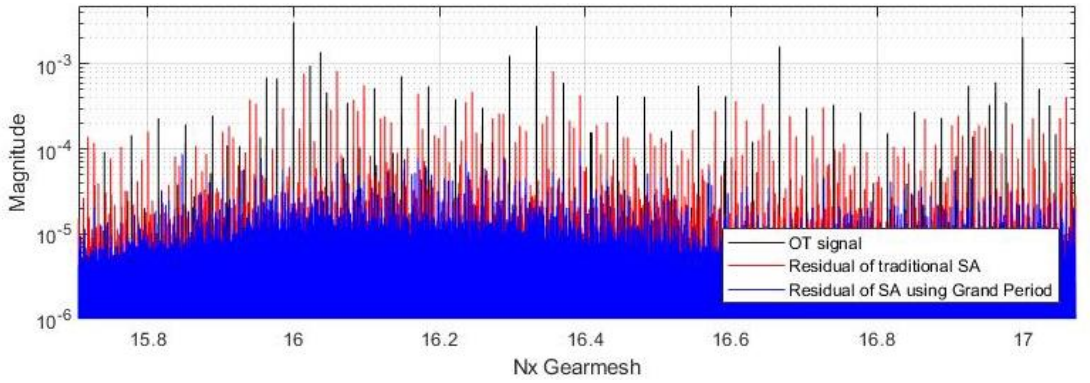


Figure 1. Result of the removal of primary-only (red) and primary + secondary (blue) sidebands from a gear vibration signal (black): (a) frequency range 0-30 gearmesh harmonics, (b) zoom of the most affected area.

Adding frequency modulation to the model, the considerations made so far become even more relevant. Even taking pure frequency modulation, the signal consists of

$$g(\theta) = c\big(\theta + \phi(\theta) + \psi(\theta)\big) \tag{9}$$

where $\phi(\theta)$ and $\psi(\theta)$ are the phase modulations introduced by shaft 1 and 2, respectively:

$$\phi(\theta) = \sum_k \Phi_k e^{jk\theta} \quad \text{and} \quad \psi(\theta) = \sum_k \Psi_k e^{jk\frac{Z_1}{Z_2}\theta} \tag{10}$$

This signal can be expressed as a Fourier Series as:

$$g(\theta) = \sum_h C_h e^{jhZ_1\left(\theta+\phi(\theta)+\psi(\theta)\right)} = \sum_h C_h e^{jhZ_1\left(\theta+\sum_k \Phi_k e^{k\theta}+\sum_k \Psi_k e^{jk\frac{Z_1}{Z_2}\theta}\right)} \tag{11}$$

The Bessel expansion of such signal (too cumbersome to report in this paper and whose details are of little significance) is composed of a large series of harmonics, at all the multiples of the fundamental frequency obtained from the "grand period". Moreover, whereas the bandwidth of AM sidebands is expected to keep constant for each carrier harmonic, the bandwidth of FM sidebands grows proportionally to the harmonic order of the carrier [3], potentially amplifying this phenomenon at high orders.

A detailed analysis of the motivation (AM, FM or mixed) of the high-frequency location observed for strong secondary side-bands is outside the scope of this paper, but a preliminary sensitivity analysis (varying shaft speed) indicates that the phenomenon has a stable location in the frequency domain (independent of

speed), rather than in the order domain. This suggests that the relevance of the secondary harmonics is linked to a dynamic amplification due to the system transfer function, rather than an FM bandwidth problem.

3 Irregularity of the sideband patterns

According to the AM model of eq. (3), each sideband-pattern should repeat identically at each carrier harmonic. This is simply explained by the convolutive nature of the spectrum of an amplitude modulated signal. For instance, simply dividing each set of sidebands $G(hZ_1 + k) = C_h A_k$ by the corresponding carrier harmonic C_h, the following equivalence should be obtained:

$$\frac{G(hZ_1 + k)}{C_h} = \frac{G(h'Z_1 + k)}{C_{h'}} \quad \forall h, h' \in \mathbb{Z} \tag{12}$$

This ideal property is explicitly at the basis of Ref. [4], which proposed a multi-carrier demodulation method, but partially and implicitly adopted by most studies which arbitrarily use the first or second harmonic for demodulation.

This concept is challenged in this section using a vibration signal measured on the same gearbox as discussed in the previous section, albeit with healthy 20 mm face width gears and a transmission ratio of 19/52. The test was operated at 20 Hz with a load of 20 Nm (all measured on the pinion/input shaft). The sampling frequency was set at 100 kHz for a total duration of the acquisition of 10 s.

The order-tracked signal was split into frequency bands corresponding to the neighbourhood of the first 4 gearmesh harmonics and the different sidebands patterns were shifted to overlap with each other. All sideband amplitude coefficients were divided by the corresponding gearmesh harmonics amplitude coefficient, in order to compute an amplitude ratio. The amplitude ratio of those patterns is reported in Figure 2.

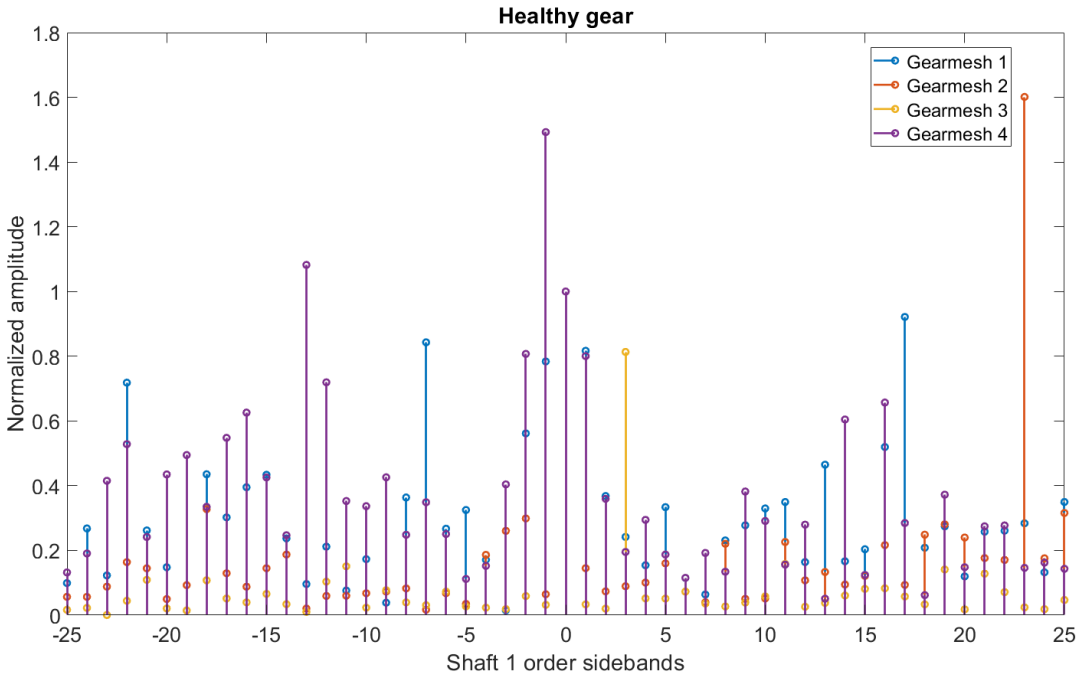
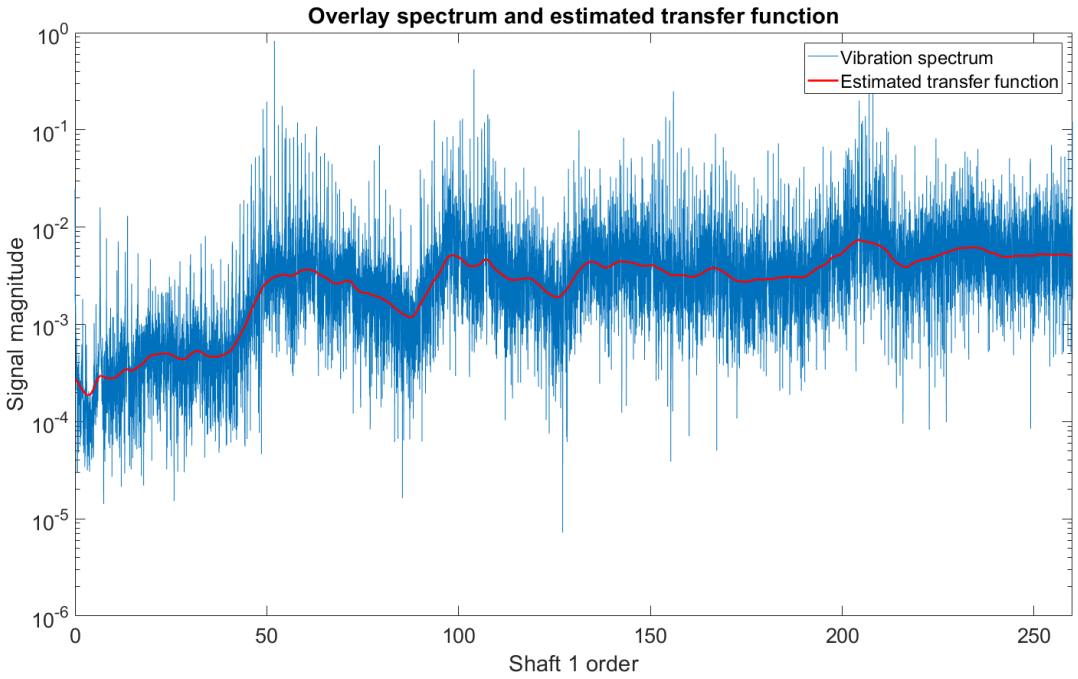


Figure 2. Sideband amplitude patterns of the first 4 gearmesh harmonics. 20 Nm - 20 Hz test.

This result shows how the patterns are massively different even disregarding their phase, which should also coincide after normalisation by the carrier harmonics. Two possible explanations for such behaviour were suggested: amplifications due to the system transfer function, or dominant frequency-modulation effects.

The first option was further investigated trying to remove the transfer function effect by means of cepstral liftering [5], [6]. An exponential lifter was applied, with a cut-off angular quefrency of 0,7 radians,to the original spectrum and the result was used to remove the short-quefrency transfer function effects. The result of the liftering operation and the "normalised" spectrum are shown in Figure 3 (a-b) respectively.

(a)



(b)

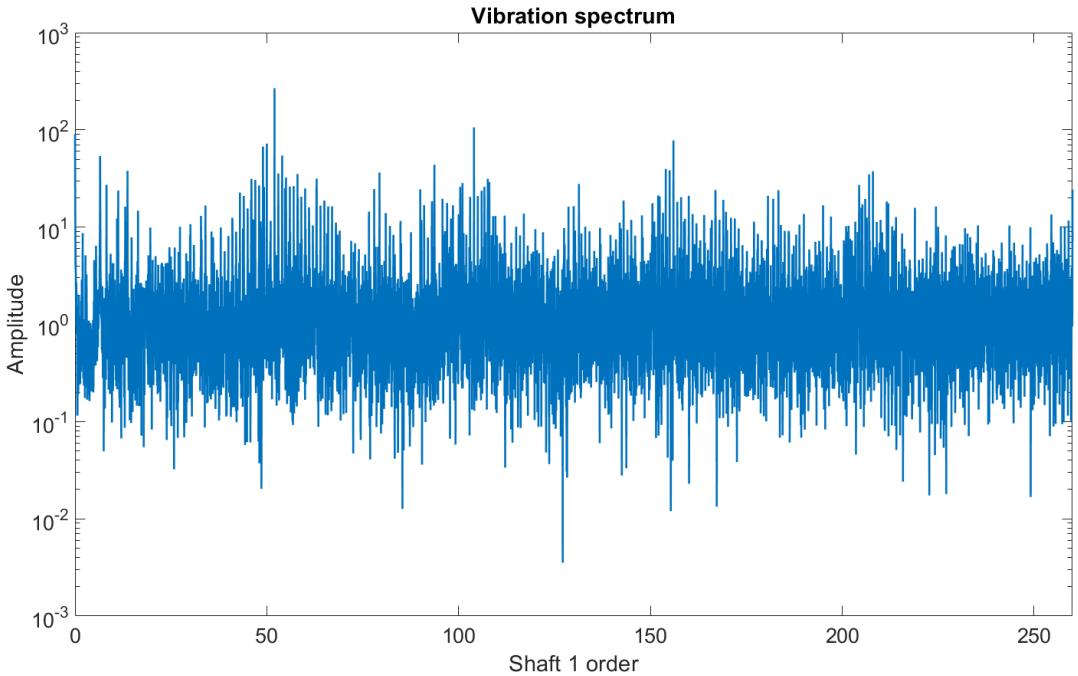


Figure 3. Removal of the TF effect: (a) short-pass liftering operation, (b) normalised spectrum.

Despite the good result in terms of spectral liftering, the problems observed in Figure 2 continue to be as severe in the normalised spectrum harmonics reported in Figure 4.

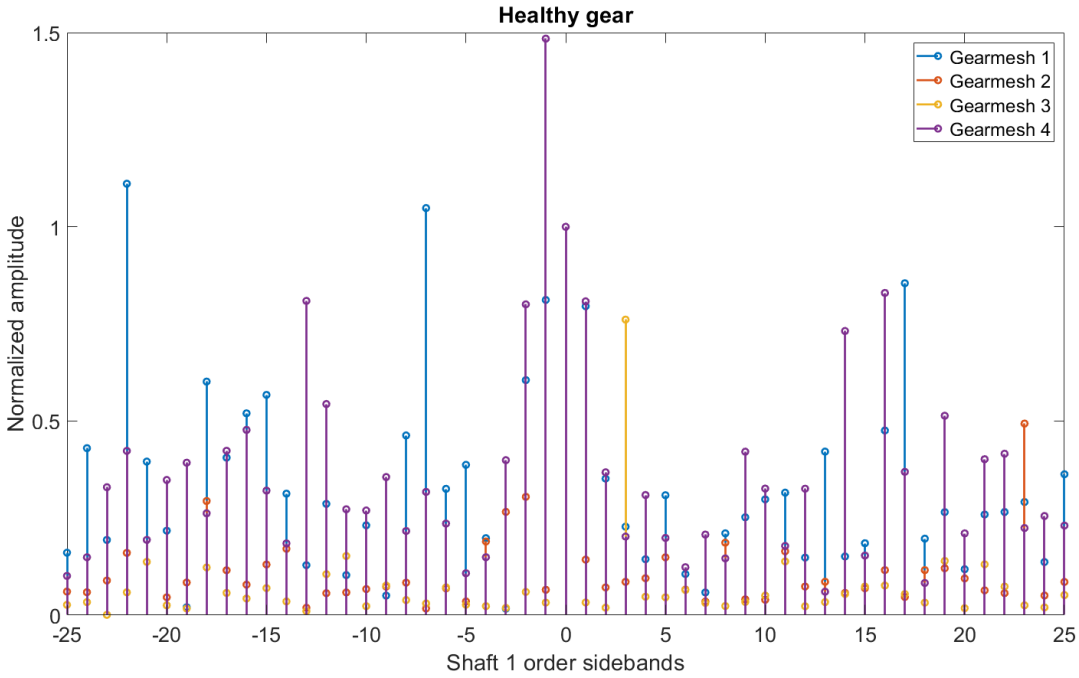


Figure 4. Sideband amplitude patterns of the first 4 gearmesh harmonics after cepstral liftering. ~1.5 Nm - 20 Hz test.

As the shape of the FM patterns of a multi-carrier modulated signal do not seem to match with the ones observed in Figure 2 and Figure 4, the authors are of the opinion that, despite possibly contributing to the difference among the sideband patterns, other phenomena (unmodelled in the current approach) must be influencing the vibration signal. FM multi-carrier modulation in fact usually results in sideband patterns showing a similar "shape", yet with a bandwidth proportional to the carrier harmonic order.

A possible explanation of the differences in the sideband patterns could be found in the different roles played by two different root-cause mechanisms resulting in gear vibration: geometric and static transmission error (each potentially resulting in a separate AM/FM modulated signal, with different carrier and modulation). The first is due to profile irregularities, whereas the second is due to the angular dependence of the gear-meshing compliance under load. An additional test was therefore executed at very low load (~1.5 Nm, just enough to maintain contact between the gear teeth) and speed (2 Hz), where geometric transmission errors were expected to dominate. Since under these operating conditions significant electrical noise was present in the lowest frequency range, in this case harmonics 2-5 were analysed.

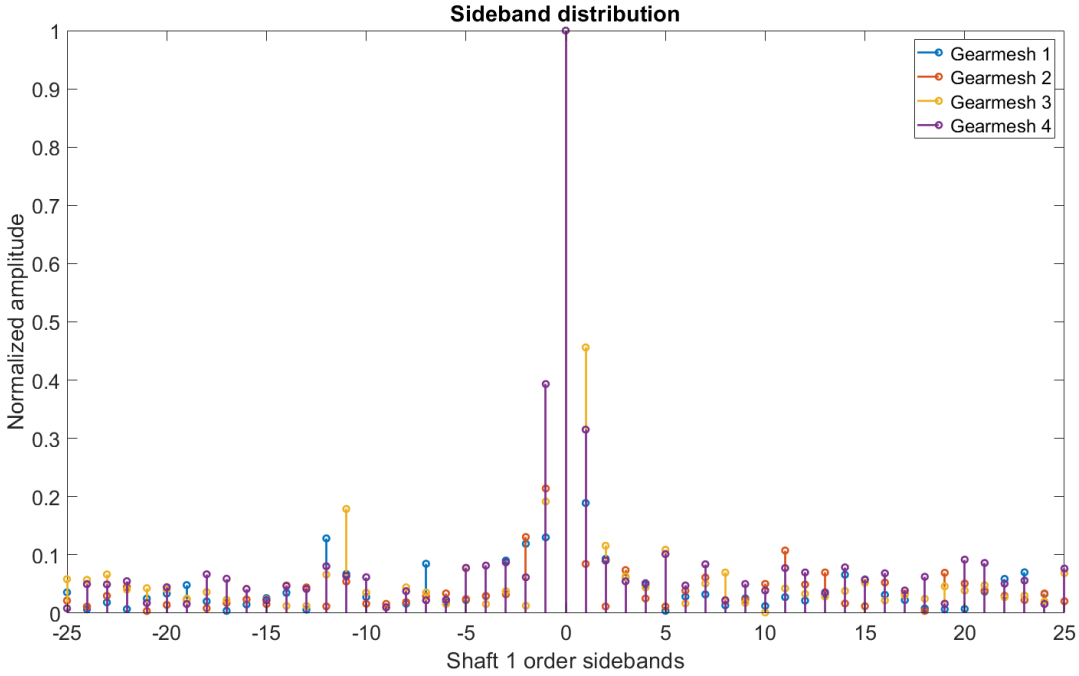


Figure 5. Sideband amplitude patterns of the first 4 gearmesh harmonics. Low load – low speed test.

The patterns shown in Figure 5 are much more consistent, even if discrepancies are still present, thus supporting the idea of a potential two-mechanism root-cause of the observed pattern inconsistency.

In order to investigate more deeply the origin of the pattern inconsistencies observed in the vibration signal, the sideband distribution of the transmission error signal is also studied, in a low-speed and low-load test. As illustrated in Figure 6, the pattern distributions have been plotted for the two cases of a heathy and a faulty gear. The transmission error is computed as the relative difference between the rotation of the input and output shafts. The pattern shown for the sidebands of the healthy gear is almost as consistent as that obtained with the vibration signal, although differences persist in the amplitudes of the sidebands. However, in the case of the faulty gear, the distribution of the sidebands is quite similar for all the sidebands of every gearmesh harmonic.

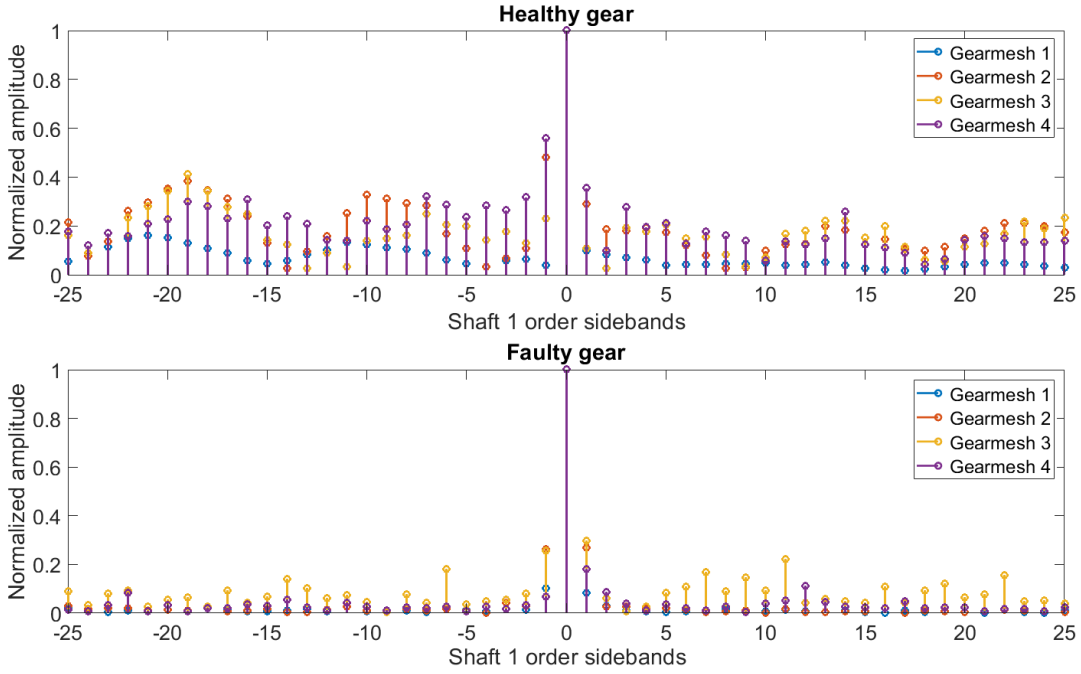


Figure 6: Sideband amplitude patterns of the first 4 gearmesh harmonics of the TE signal. Low load – low speed test.

4 Conclusions

This study has highlighted two major limitations in the current modelling (and assumptions) of gear vibration signals. Neglecting secondary sidebands has been shown to be not always reasonable, and strong discrete components were still observed in the spectrum of a gear signal after removing the primary sidebands by means of traditional synchronous averaging procedures. This problem, which could bias the identification of system transfer functions with OMA approaches, is easily solved if it is possible to observe a sufficient number of "grand-periods". In practice this could be possible for a series of machines operating at reasonably constant speed, but might be impractical for complex transmissions with more than one stage or planetary arrangements (very long grand-period).

Experimental evidence also casts doubts on the validity of AM and even AM/FM models of gear vibration signals, and suggests the possibility of multiple forcing functions (with different spectral distribution) acting simultaneously to create complex modulations. In particular, the geometric vs static transmission error components seem to be potential candidates for future investigations.

Significant investigative efforts are required to clarify these issues and give rise to more reliable models, in turn enabling new and more effective condition monitoring and OMA approaches.

[1] W. D. Mark, “Analysis of the vibratory excitation of gear systems: Basic theory,” *The Journal of the Acoustical Society of America*, 2005.
[2] D. Abboud and J. Antoni, “Order-frequency analysis of machine signals,” *Mechanical Systems and Signal Processing*, 2017.
[3] P. Borghesani, P. Pennacchi, R. B. Randall, and R. Ricci, “Order tracking for discrete-random separation in variable speed conditions,” *Mechanical Systems and Signal Processing*, vol. 30, 2012.
[4] E. Hubert, A. Barrau, and M. E. Badaoui, “New Multi-Carrier Demodulation Method Applied to Gearbox Vibration Analysis,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2141–2145.
[5] Y. Gao and R. B. Randall, “Determination of frequency response functions from response measurements - I. Extraction of poles and zeros from response cepstra,” *Mechanical Systems and Signal Processing*, 1996.
[6] R. B. Randall, B. Peeters, J. Antoni, and S. Manzato, “New cepstral methods of signal pre-processing for operational modal analysis,” in *{P}roceedings of the {I}nternational {C}onference on {N}oise and {V}ibration {E}ngineering {ISMA 2012}*, 2012.

Detection sensitivity study of local faults in spur gears based on realistic simulations

Lior Bachar, Ido Dadon, Renata Klein, Jacob Bortman

Keywords: Gear Model, Vibration Signature, Condition Monitoring, Local Tooth Fault, Spur Gear, Gear Diagnostics

Abstract:

The dynamic response of gear transmissions holds essential information for the recognition of an incipient fault and its propagation. A realistic and validated dynamic model is used to predict the vibration regime of gear transmissions [1]-[2]. This model was validated experimentally for both healthy and damaged conditions [1]. A great virtue of a model is the ability to examine each phenomenon separately and to isolate its contribution to the dynamic response. The model considers the nonlinear behavior of the gear mesh stiffness, integrating the geometric profile errors of the gears. The scattering in the data, which is generated by the random factor of the simulated surface roughness, simulates the reality better than data of an ideal profile. The ability to determine what is possible to monitor for each surface roughness is not trivial and cannot be achieved experimentally, due to the immense span of cases to consider.

This work presents an analysis of spur gears transmissions that can be separated into two integral but still different studies. The first study examines the effect of the operating conditions, including speed load and surface roughness, on the vibration signature of a healthy gearbox. The two main evidences from this study are related to the levels of the gear mesh frequencies (GMF) and to the sidebands (SB's) in the spectra, which are caused by the frequency modulation (FM) of the rotational speed. It was found that there is a strong dependency of the energy at the gear mesh frequency on the applied load. Figure 1.a presents the total GMF's energy for different rotational speeds (R1 is the lowest speed, R3 is the highest speed) under different loads (L1 is the lowest load, L4 is the heaviest load). It is noticeable that under the same rotational speed, the total GMF's energy sharply rises as load increases. On the other hand, it was found that the total spectral energy of the FM sidebands sharply rises as speed increases, but is not affected by load. Figure 1.b shows the spectral energy of the FM SB's for different speeds under different loads around the first six harmonics of the GMF. As for the influence of the surface roughness [3], it was found that a coarser surface roughness tends to obscure the effects of load and speed on the signature due to the gears profile error. Furthermore, the energy level of the FM sidebands sharply rose as the surface roughness got coarser, while the energy level of the GMF's barely changed due to the surface roughness.

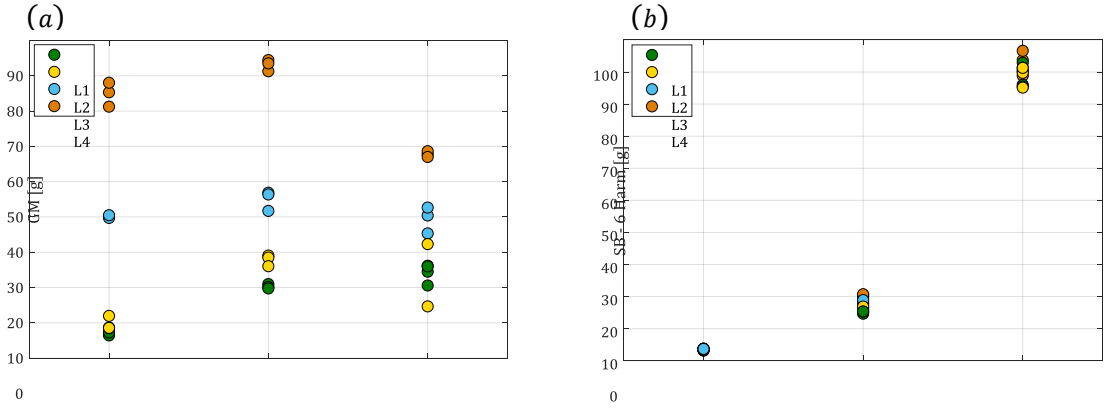


Figure 1: total energy against rotational speed and load of the (a) GMF's (b) FM sidebands around the first six GMF's

The second study examines the expression of local tooth faults in the vibrations signature. Load effects and other AM phenomena including eccentricity and misalignment may obscure the expression of local tooth faults. The comprehensive study of the effects of the operating conditions on the signature was necessary in order to fit a robust and sensitive monitoring process for the local faults detection capability. The optimal process should reflect the expression of the fault in the signature, while extinguishing the effects of the operating conditions, which are not related to the fault itself. The difference signal removes from the synchronized vibrations signal the GMF's components that are strongly affected by load, and components which are related to AM phenomena. Hence, the difference signal let us focus on the fault expression while diminishing the effects of the operating conditions. Figure 2.a shows the RMS level against kurtosis, both of the difference signal, for five different fault severities (where "Fault 1" is the least severe fault and "Fault 5" is the most severe fault). It can be seen that a separation of the faulty conditions from the healthy condition can be achieved for most faults severity levels, due to the significant differences in their locations on the graph. Besides the analysis of the difference signal, we can also utilize the total spectral energy of the FM sidebands. For each random signature, the total energy of the FM sidebands can be calculated and be compared to the healthy condition by statistical distances. The statistical distance may determine whether the examined signature can be attributed to the healthy population or not, meaning that a separation of the faulty condition from the healthy condition may be achieved. It was found that the detection capability is clearer when examining the spectral energy of the FM sidebands around the gear mesh frequencies overlapping the natural frequencies of the gearbox. Figure 2.b presents the Mahalanobis statistical distances (D) of the total FM sidebands energy around the GMF harmony which was found to be the most affected by the natural frequencies of the gearbox, against the level of the fault severity level. A Mahalanobis distance of $D = 10$ was determined to be the threshold for excluding a signature from the healthy population [4]. It can be seen that a separation of most of the local faults was achieved, as well as ranking of the three most severe faults.

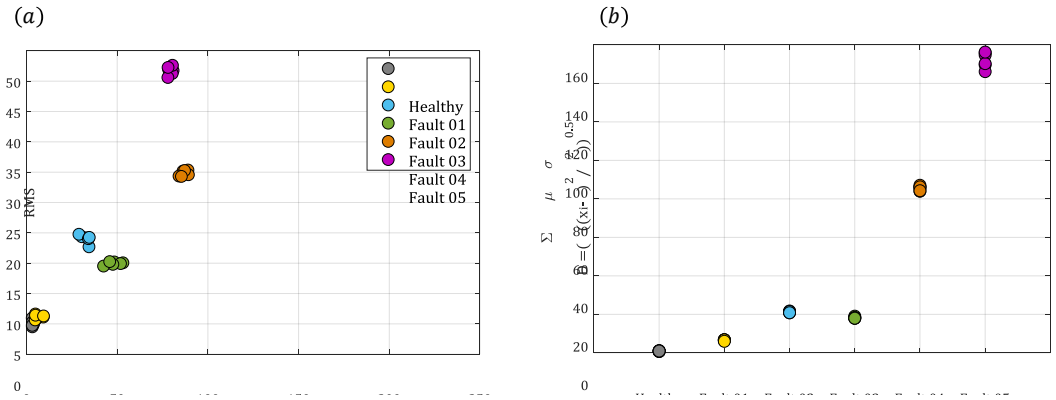


Figure 2: (a) RMS Vs. Kurtosis of the difference signal (b) Statistical distance of FM sidebands around the harmony of the GMF which was found to be the most affected by natural frequencies

References

[1] Dadon, I., Koren, N., Klein, R., & Bortman, J. (2018). A realistic dynamic model for gear fault diagnosis. Engineering Failure Analysis, 84, 77-100.

[2] Koren, N., Dadon, I., Bortman, J., & Klein, R. (2017). Steps towards fault prognostics of gears. International Journal of Condition Monitoring, 7(1), 10-15.

[3] Dadon, I., Koren, N., Klein, R., & Bortman, J. "The Effect of Gear Tooth Surface Quality of Diagnostic Capability ." Surveillance 9, 2017, surveillance9.sciencesconf.org.

[4] Klein, R. (2012, September). Condition indicators for gears. In Annual conference of prognostic and health management society (pp. 1-8).

Towards a better understanding of helical gears vibrations – dynamic model validated experimentally

Silverman, N; Dadon, I; Bortman, J; Klein, R

In order to simulate the vibration signature of gears, an accurate calculation of the gear mesh stiffness (GMS) is required. The time varying GMS, which is the main excitation that determines the dynamic response of transmissions' vibrations, is well understood for spur gears, but that of helical gears was less investigated. Although there is work dedicated to helical gears vibrations, a comprehensive analysis of their GMS compared to spur gears and their time and spectral domains have yet to be made. This paper deals with the dissimilarities and provides a better understanding of helical gears behavior, as they are a key component in many complicated and costly machines. With this new knowledge a more educated approach to diagnostics might be achieved.

The main difference between spur and helical gears is in the contact line pattern. In spur gears the contact line is parallel to the tooth's base and so calculating the GMS in any given moment is rather easy. Helical gears on the other hand have a diagonal line of contact which makes the moment applied by the meshing gear in respect to the tooth's base change along the tooth's width. To overcome this challenge a 'multi slice' method is utilized [1-4], in which the helical tooth is divided into many infinitesimally narrow slices which are treated each as a spur tooth. The total helical tooth stiffness is the sum of all those spur slices.

For the purpose of simulating the vibrations of helical gears a fourteen degree of freedom spur-teeth dynamic model [5] was upgraded to include helical gears a well. The dynamic equations and stiffness calculation were not changed and thus are discussed only briefly. The focus is dedicated to the modeling of the contact line using the multi slice method and other adjustments made to the model.

The challenge with the slice method is determining how many slices are in mesh at every given time, along with determining their mesh "height" (distance from the tooth's base). The solutions found in the literature are rather complicated and require knowing niche data about the gears, such as the transverse operating pressure angle, which are often not provided by the manufacturer. In contrast, the method suggested in this work is based on only a few common parameters such as the gears module, number of teeth and the involute profile.

The model was validated by an experiment conducted on a helical gearbox and recorded with a tri-axial accelerometer. The signals were compared in terms of their load and RPM dependency and exhibited a similar behavior, as can be seen in Figure 1. After obtaining a healthy baseline, a broken tooth case with three severity levels was studied. The fault severities were removal of 25% of the tooth's width in a diagonal line, removal of 50% and a missing tooth (Figure 1). This kind of diagonal material removal was chosen because when helical teeth breaks it happens in a pattern parallel to the contact line. The light and medium fault were challenging, but the missing tooth

was seen clearly, mainly in the Kurtosis and Crest Factor of the difference signal. The statistical distances of the SA spectrum around the first and fifth GM harmony proved to provide better sensitivity, and showed clear detection of various fault severities, mainly in the tangential direction. A calculation of the Z-score index around the first GM even showed capability of ranking by fault severity (Figure 2).

Figure 1: Three level of fault severity. Left to right: removal of 25% of the tooth's width, removal of 50%, and a missing tooth.

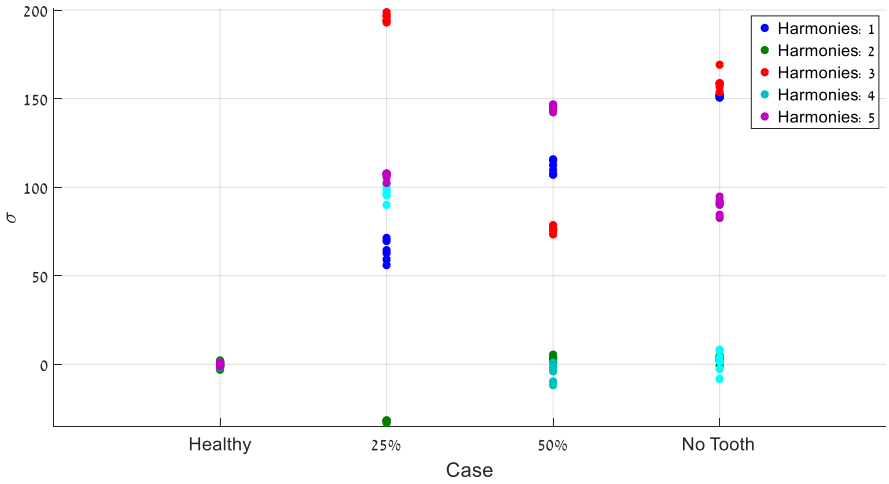


Figure 2: The Z-Score index for each GM harmony. Notice the first harmony which shows fault ranking and the fifth, which shows detection even at the smallest fault.

Keywords: Helical gears, Dynamic model, Multi-slice method, Broken tooth.

Bibliography

1. Smith, J. D. "Estimation of the static load distribution factor for helical gears." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science209.3 (1995): 193-199.

2. Chen, Kangkang, et al. "Comparison of meshing characteristics of helical gears with spalling fault using analytical and finite-element methods." Mechanical Systems and Signal Processing121 (2019): 279-298.
3. Wang, Qibin, et al. "An improved time-varying mesh stiffness model for helical gear pairs considering axial mesh force component." Mechanical Systems and Signal Processing 106 (2018): 413-429.
4. Wang, Qibin, and Yimin Zhang. "A model for analyzing stiffness and stress in a helical gear pair with tooth profile errors." Journal of vibration and control 23.2 (2017): 272-289.
5. Dadon, I., et al. "A realistic dynamic model for gear fault diagnosis." Engineering Failure Analysis 84 (2018): 77-100.
6. Dadon, I., et al. "A Step Toward Fault Type and Severity Characterization in Spur Gears." Journal of Mechanical Design141.8 (2019): 083301.

Modeling and identification of mechanical systems

Localization and quantification of damage by frequency-based Method: Numerical application on bending vibration beam

Anurag DUBEY[1], Vivien DENIS[1], Roger SERRA[1]

[1]INSA Centre Val de Loire, Laboratoire de Mécanique Gabriel Lamé - EA 7494, Blois, France
anurag.dubey@insa-cvl.fr

Abstract

The sudden growth of damages can cause catastrophic failure of structures or mechanisms that lead to unplanned shutdowns of machines and production lines. If a damage remains undetected and reaches a critical size, sudden collapses and failures can happen. To overcome these problems, it is essential to detect these damages before they reach their critical state. The presence of damages can alter the structure which reduces the bending stiffness and modify the modal parameters and the natural frequencies. One of the most suitable monitoring methods to define the presence of damage and assess the structure is vibration based structure health monitoring (VBSHM). The objective of the work is to localize and quantify the damages with the consideration of eigenfrequencies of healthy and tested structures. Hence, a methodology for damage identification in structure using frequency shift coefficient (FSC) is presented. Numerical finite element models (2D and 3D) are performed and correlated to obtain a damage library for the cantilever beam structure. Based on the cost function, Young's modulus of 2D and 3D models are iteratively updated to closely match the frequencies of the reference beam. The approach also quantified geometry damage with vibration measurements on cantilever beams, which is related to an equivalent bending stiffness reduction by the use of FSC. The effect of severity of the damage is considered. Finally, the result is validated numerically through the identification of geometry damage.

1 Introduction

Damages or cracks are inevitable in aerospace, aeronautical, mechanical and civil structures during their service life. Any changes in the structures such as material, physical or geometrical properties which affects their performance are considered as damages. The study of damages is an important perspective in order to ensure safety or to avoid any serious losses. Sudden occurrence of damages in the structure can cause catastrophic failure and reduction in load carrying capacity. However, it is necessary to improve the durability and reliability of structure as expressed in the design and maintenance specifications. The presence of the damage makes local stiffness vary in the structure and it also affects the mechanical behavior and performance of the structure. However, preventing the formation of damages is almost impossible as they propagate along the structure due to fluctuating stress or fatigue conditions. If these cracks remain undetected and reach a critical size sudden collapse can happen. Indeed, damage identification has significant life safety implications.

Structure Health Monitoring (SHM) is an efficient way for the diagnosis of the constituent's materials or structures. SHM involves the integration of sensors, data transmission, computational techniques, and processing ability to respond the behavior of a structure. Consequently, it aims to provide maintenance services throughout the life of the structure. Nowadays, structural damages are identified by Non-Destructive Testing (radiographic, ultrasonic testing, X-ray, eddy-current etc.) [1]. Vibration based structural health monitoring (VBSHM) is one of these categories based on the fact that a loss of stiffness caused by damages affects the dynamic response of the structure. VBSHM consists of five levels (existence, location, type, extent and prognosis) [2] which are efficient and widely accepted because of their ability to monitor and detect damage from global testing of the structure.

Many researchers from the last few decades, natural frequencies of a damaged structure are found as an identification parameter for both damage location and size. The first study developed by Cawley and Adams [3] depends on the shift of more than one frequency that could yield the location of the damage. In a review

of the literature, Salawu [4] found that the natural frequencies are a sensitive indicator to detect the damage in the structure. The important technique is analyzing the changes (shifts) in natural frequencies in a structure with and without damage. Hilmy et al. [5] have presented frequency shifting as a function of damage evolution for a plate structure. The method proves shifting of the natural frequency is greater at higher frequency values and determines the location of the void damage. Messina et al. has proposed Damage Location Assurance Criterion (DLAC) [6] and after extended to Multiple Damage Location Assurance Criterion (MDLAC) [7] to measure the frequency variation due to damage between experimental and numerical values correlation. More recently, a method proposed by Serra et al. [8] demonstrates a correlation of 2D and 3D FE models to identify the typical damages (like hole, crack, notch) based on numerical and experimental study. Masoumi and Ashory [9] presented numerical and experimental studies to localize cracks.

In this paper, an approach for damage identification by using the frequency-shift coefficient is proposed. This method was first introduced by Silva and Gomes [10] for solving the damage detection problem. The method requires numerical models as a function of damage position and size for the frequency shift. First, vibration based strategy is used with detection, localization and classification (Size/Severity/Geometry) of damages. The study is followed by simulating a beam in commercial software (COMSOL, MATLAB) as a numerical case and 2D and 3D FE models are correlated to obtain geometry damage properties (size, location and severity...). Finally, numerical example is validated in order to localize and quantify geometry damage.

2 Cantilever beam bending vibration background

The eigenvalue problems and the analytic formulas concerning the modal parameters of a cantilever beam were described by the partial differential equation of the linear model with viscous damping as:

$$M(x)\frac{\partial^2 v(x,t)}{\partial t^2}+C(x)\frac{\partial v(x,t)}{\partial t}+\frac{\partial^2}{\partial x^2}\left(EI(x)\frac{\partial^2 v(x,t)}{\partial x^2}\right)=F(x,t) \tag{1}$$

where $v(x,t)$ is the transverse deflection, $M(x)$ is the mass per unit length, $C(x)$ is the damping coefficient, $EI(x)$ is the bending stiffness and $F(x,t)$ is the external force per unit length of the beam. The equation of the motion for dynamic systems are easily obtained from Newton's second law. This gives an equation for each degree of freedom within the system. When discretized, the equation of the motion may take the following matrix form:

$$[M]\{\ddot{X}\}+[C]\{\dot{X}\}+[K]\{X\}=\{F\} \tag{2}$$

where $[M]$ is the mass matrix for the system, $[C]$ is the damping matrix and $[K]$ is stiffness matrix, $\{\ddot{X}\}$, $\{\dot{X}\}$, $\{X\}$ are vectors containing acceleration, velocity and displacement in all degree of freedom of the model; and $\{F\}$ contains external forces actuating in the system. If we assume free motion and negligible damping, one possible solution for the equation is:

$$\{x\}_i=\{y\}_i\sin(\omega_i t-\theta_i) \tag{3}$$

y_i are the amplitudes for each mode shape, ω_i are the natural pulsations (in rad.s^{-1}) of vibration for each mode shape and θ_i are phase angles. The natural frequencies (in Hz) are given by $f_i=\frac{\omega_i}{2\pi}$. The following equation is obtained for the healthy case:

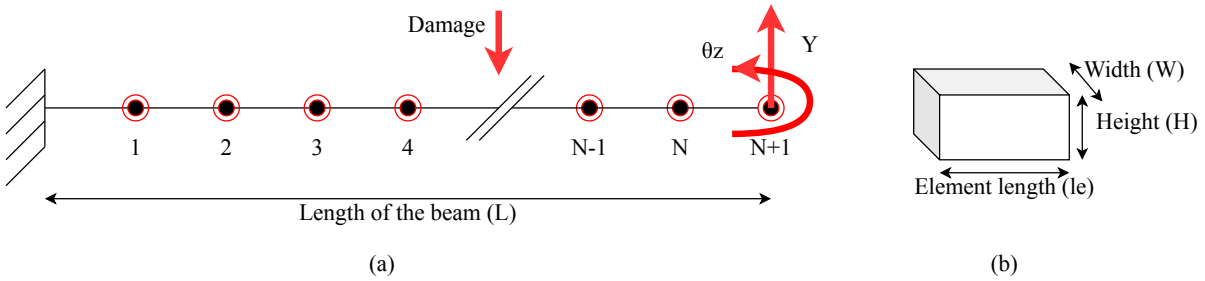
$$([K]-\omega_i^2[M])y_i=0 \tag{4}$$

Damage to the structure changes its dynamic response. Therefore, natural frequencies and natural modes are changed. The equation of a damaged case can be expressed as:

$$([K]-(\omega_i^2)^*[M])y_i^*=0 \tag{5}$$

2.1 2D Finite element model

The studied model is a cantilever beam, which has two degrees of freedom, a vertical translation y and a rotation θ_z. As can be seen in (Figure 1) this beam is divided into equal size of N elements and $N+1$ nodes.



(a) (b)

Figure 1 – Cantilever beam model. (a) 2D Finite element model; (b) element properties

In 2D FE Model, the damage is represented by an elemental stiffness reduction coefficient α_i which is the ratio of the stiffness reduction to the initial stiffness. The stiffness matrix of damaged beam is defined as a sum of elemental matrices multiplied by reduction coefficient by the following equation:

$$[K_d] = \sum_{i=1}^{N} (1 - \alpha_i)[K_i] \tag{6}$$

where K_d is global stiffness matrix for damaged beam, K_i is elemental stiffness matrix, N is number of elements, and α_i is a reduction coefficient , which varies from 0 to 1 for the damaged structure. The value of α_i = 0 indicates a healthy structure.

2.2 3D Finite element model

Simulation of damaged beam structure is performed using COMSOL multiphysics software. The damage model is built and the mesh is 3D tetrahedron element. The number of mesh is controlled by the software and depends on the shape of the structure, thus it changes with the size of the crack. A high meshing density is applied near the damaged area mainly to have the behavior correctly modeled.

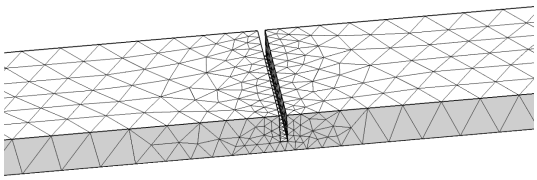


Figure 2 – 3D finite element mesh of the beam with damage

Geometry case (rectangular) is studied in order to quantify the severity of the damage. Figure 2 shows the meshed beam zoomed near the damaged area and width of the crack is set to 0.5 mm while the height is a parameter. The sensitivity of the 3D model is determined by mesh size. As the mesh is finer the model is more sensitive but computing cost is higher.

3 Frequency shift coefficient based strategy

The first type of modal method for damage detection relied on changes in dynamic properties of the structure and particularly natural frequencies. Any changes in the properties of the structure, such as reduction in stiffness will cause changes in the natural frequencies. One of the important advantages of natural frequency is that it can be quickly and easily conducted when measurements required. Classical measurements procedure can be used for the determination of experimental resonant frequencies. In this context, the frequency shift criterion is first presented by Silva and Gomes [10] for damage identification problems. The technique requires experimental measurements or numerical solution for the frequency shifts as a function of size and position of damage The

frequency shift coefficient (FSC) is defined as:

$$FSC = \text{argmin}\left(\sqrt{\frac{1}{m}|\sum_{i=1}^{m}\left(\frac{(R_i)_X - (R_i)_A}{(R_i)_X}\right)|}\right) \quad \text{and} \quad R_i = \frac{f_i^u}{f_i^h}, \tag{7}$$

where m is the total number of modes, X refers to the tested case, A refers to the reference case, f_i^u is the unknown beam frequencies, f_i^h is healthy beam frequencies and i is denotes modes indices.

It is well known that the presence of damages modifies dynamic parameters and behavior of the structure. The location, classification and size of damages in the structure are identified by changes in the vibration parameters. At first, a set of reference state frequencies are identified. Numerical correlation of 2D and 3D FE models is performed to fit the frequencies with the references. Then, in the 3D FE model, the damage was materialized as a geometrical discontinuity of rectangular considering the position, type, size, geometry of the damage. At the same time, the damage was materialized as a local reduction of bending stiffness in an element for the 2D FE model. Finally, numerical correlation result will specify the position, size, depth and geometry of the damage from the damage library.

4 Numerical rectangular geometry damage identification

The numerical simulation test is performed to verify the efficiency of the proposed VBSHM strategy. A cantilever steel beam was taken into consideration for the numerical test and beam properties are given below in Table 1. A beam 2D FE model was divided into equal size of 100 elements and each element size is 10 mm.

Beam Properties	**Value**
Length (L)	1000 mm
Width (W)	24.9 mm
Height(H)	5.3 mm
Young's modulus (E)	210 GPa
Mass density (ρ)	7850 kg/m^3
Poisson's ratio (v)	0.33

Table 1 – Beam dimensions and properties

However, 2D-3D model correlation is done for the damage geometry (rectangular) by using software COMSOL with MATLAB. The modal responses of the structure were generated using FE models before and after the damaged case. The first seven modes are retained. The criterion is employed as a tool for identification of the damage by measuring frequencies. The final goal of correlation is to localize and quantify the severity of geometry damage that can link to the percentage reduction in stiffness of a beam structure.

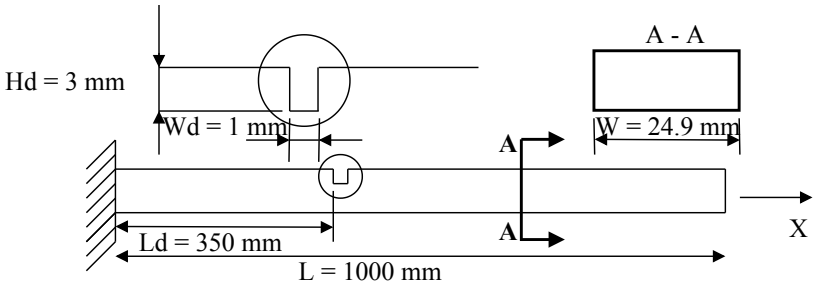


Figure 3 – 2-D clamped free beam plan with rectangular damage

A detailed 2D beam view as shown in Figure 3. The damage case is tested for position 350 mm with width (Wd) and the height (Hd) of damage are 1 mm and 3 mm respectively. Meanwhile, the FSC is computed for every position and severity, in order to illustrate its variations.

In Figure 4, the FSC is shown as a function of tested position and severity where color levels represent the FSC values. The minimum value (coordinates and value of the minimum) allows the identification of given

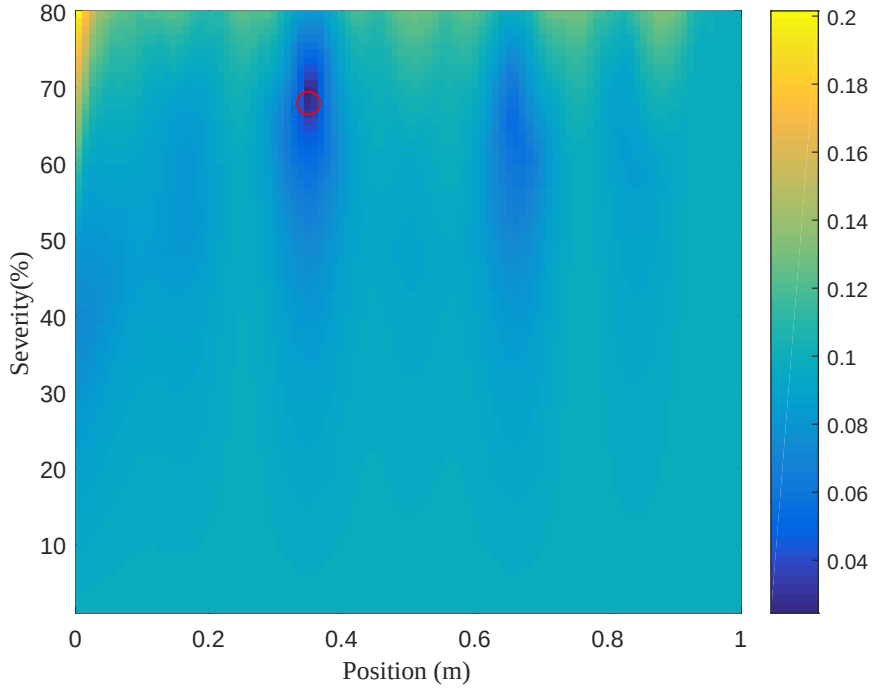


Figure 4 – FSC as a function of tested position and severity. The red circle indicates the severity and position identified by the algorithm.

damage in 2D clamped free beam. In this case, a defect of 68% severity localized at 350 mm is found : the position is thus well identified by the FSC and the identified severity corresponds to the parameters chosen for the rectangular damage. These values relate to the other damages properties and information about the size and type. In addition, Table 2 shows the identified damage properties for this particular case of rectangular damage. This is one item of the damaga library the presented strategy is intended to build.

Type	**Rectangular**
Position	350 mm
Width	24.9 mm
Length	1 mm
Height	3 mm
Severity	68 %

Table 2 – Geometry damage properties

5 Conclusion

This paper presents a method to identify damage in structure by using natural frequencies. The formulation of the method based on stiffness reduction has been validated with the localization and quantification of the rectangular geometry damage in beam like structure. The simulation correlation with COMSOL and MATLAB are presented and the robustness of the present method is examined. A numerical example with 3D geometry damage case is identified. Based on natural frequency, the damage localization and quantification is accurate because of the sensitivity of the frequency shifts to the damage states. Both 2D and 3D models of the beam were used to link the size of damage to the reduction in stiffness. Geometrical damage properties were successfully accomplished by linking FE models. In the future, more experiments and simulations should be investigated in order to validate the methodology in real cases.

Acknowledgements

The work was funded by the Région Centre Val de Loire, France through a doctoral contract. The authors are thankful for the financial support.

References

[1] B. P. Chandra Rao, *Non-destructive Testing and Damage Detection*, In: Aerospace Materials and Material Technologies, Springer, Singapore, 2017. p. 209-228.

[2] A. Rytter, *Vibrational Based Inspection of Civil Engineering Structures*, Department of Building Technology and Structural Engineering, Aalborg University, 1993.

[3] P. Cawley, R.D. Adams, *The location of defects in structures from measurements of natural frequencies*, The Journal of Strain Analysis for Engineering Design, 1979, vol. 14, no 2, pp. 49-57.

[4] O. S. Salawu, *Detection of structural damage through changes in frequency: a review* Engineering Structures, 1997, vol. 19, no 9, pp. 718-723.

[5] I. Hilmy, M. A. Adesta, E. Y. T., and T. Firdaus, *Damage Detection based on the Natural Frequency shifting of a clamped rectangular plate model*, In : Journal of Physics: Conference Series. IOP Publishing, 2015, pp. 012034.

[6] A. Messina, I. Jones, and E. Williams, *Damage detection and localization using natural frequency changes*, In Proceedings of the 1st Conference on Identification, Cambridge, England, UK, 1992, p. 67-76.

[7] A. Messina, E. J. Williams, and T. Contursi, *Structural damage detection by a sensitivity and statistical-based method*, Journal of sound and vibration, 1998, vol 216, no 5, pp. 791-808.

[8] R. Serra, L. Lopez, and G. Gautier, . *Tentative of damage estimation for different damage scenarios on cantilever beam using numerical library*, In : Proceedings of the 23th International Congress on Sound and Vibration, Athens, Greece, 2016, pp. 10-14.

[9] M. Masoumi, and M. R. Ashory, *Damage identification from uniform load surface using continuous and stationary wavelet transforms*, Latin American Journal of solids and structures, 2014, vol. 11, no 5, pp. 738-754.

[10] J. Silva, and A. Gomes, *Crack identification on simple structural elements through the use of natural frequency variations: The inverse problem*, In Proceedings of the 12th International Modal Analysis, 1994, p. 1728.

ARX model for experimental vibration analysis of flexible manipulator during grinding

Quoc-Cuong Nguyen[1], Viet-Hung Vu[1] and Marc Thomas[1*]
[1]Mechanical Engineering Department, École de Technologie Supérieure (ÉTS),
Montréal, Québec, H3C 1K3, Canada
[*]Corresponding author: marc.thomas@etsmtl.ca

Abstract

Using a flexible manipulator for grinding process in situ has become a cost effective engineering service in the recent years, especially for repair and refurbish of mechanical systems and components. In comparison with traditional rigid robot manipulators, the flexible manipulator has proved its efficiency in terms of accuracy and facility. However, because of its compact and flexible structure, concerns arise regarding its dynamic behavior during a grinding process. This paper proposes a method using an ARX (autoregressive with exogenous excitation) model for experimentally analyzing the vibrations of a flexible robot during a grinding operation in different cases: Single Input–Single Output (SISO) and Multi Input–Multi Output (MIMO). Simultaneously, a dynamometer allows for triaxial input force measurement while three accelerometers mounted at the end effector record the vibration outputs. Due to the Operational Modal Analysis (OMA), the dynamical properties of the robot can be identified directly during operation. The results have shown that the ARX model is efficient for analyzing the operational vibration in complex systems with multi degrees of freedom and multi directions. The determination of modal parameters and identified Frequency Response Functions (FRFs) enable to predict the dynamical behavior of the system and to simulate the vibration in real working conditions. Further studies on inverse problem are promising for estimating the excitation forces while these later are not available and not practically measured in industrial applications.

Keywords: Operational modal analysis, flexible manipulator, grinding process, ARX model, transfer functions, force identification.

1. Introduction

Nowadays robots sufficiently conduct manifold manipulation works with a high degree of autonomy and rigorousness. Portable manipulator systems are regarded as an effective and profitable solution for the automation maintenance tasks on large hydroelectric equipment. The SCOMPI (Super COMPact robot Ireq) was developed at IREQ (Hydro Quebec's research institute) and is particularly designed with flexible links and flexible joints for working in the hard-to-reach areas or confined spaces of hydraulic turbines in a hostile environment [1]. Because of its flexible structure, vibration problems of Scompi become crucial since producing chatter and bad surface finish. A numerical simulation [2] has been constructed in MSC/Adams in different configurations included impact force, sinusoidal and operational forces. There is a great number of researches that focus on identifying the modal parameters of the system in order to understand the dynamical behavior of robot [3-7], and estimate the operational forces from the actual accelerations measured on the robot [8]. Knowing a system's frequency response function is a key to many system analysis and control synthesis

methods [9]. The main problems are due to the fact that these modal parameters are changing with the robot motion and position and thus a time-varying method is proposed for studying this kind of non-stationary structure [10, 11]. Researchers are particularly interesting to identification of continuous-time system by using discrete data [12].

This paper presents a technique to identify the modal parameters as well as the transfer function of Scompi robot by applying the Autoregressive with eXogenous input (ARX) model [12-14]. This method reveals a convenient and advantageous for Operational Modal Analysis of structures (OMA), which allows for determining operational modal model excited by ambient noise and vibration. The modal parameters are estimated and identified by applying straightforward method such as Ordinary Least Squares (OLS) [12], [15]. The results are validated by another approach based on updated Auto Regressive (AR) model in [4] and shown a great accuracy of identified modal parameters. This study enables us to predict the dynamical behavior of Scompi for identifying excitation forces during operations of grinding and consequently improve the quality of the surface finish.

2. Auto Regressive Exogenous Excitation model (ARX)

The ARX model [12-15] is a primary choice because of its simplicity. It has been applied to numerous practical applications especially in control systems. However, critical motivation for choosing the ARX model, is its correlation to the state space model [16-19] which can be implemented for inverse problem with the aim of reconstructing the excitation forces acting on vibrating structures [8], which is impossible to obtain from direct measurement in the real systems. The ARX model is a convenient model to obtain the general relation between input and output signals for different cases, such as Single Input – Single Output (SISO) or Multiple Input – Multiple Output (MIMO), which can reliably represent the dynamic properties of the system. Figure 1 illustrates the block diagram of ARX model.

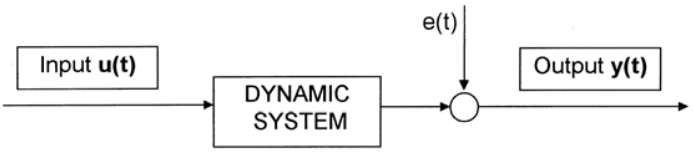


Figure 1. Block diagram of ARX model

This model has a simple structure and strong robustness. It is very efficient when the noise is low. However, when the noise is large, the order of the model must increase to compensate the impact to system identification precision from noise [18].

Examine a *c* dimensional vector input **u**(t) and a *d* dimensional vector output **y**(t) of a Multiple Input and Multiple Output (MIMO) system.

The ARX model can be described as a linear difference equation:

$$\mathbf{y}(t) + \mathbf{A}_1\mathbf{y}(t-1) + \ldots + \mathbf{A}_{n_a}\mathbf{y}(t-n_a) = \mathbf{B}_0\mathbf{u}(t) + \mathbf{B}_1\mathbf{u}(t-1) + \ldots + \mathbf{B}_{n_b}\mathbf{u}(t-n_b) + \mathbf{e}(t) \quad (1)$$

where:

$\mathbf{A}_i$ – are $d \times d$ matrices and

$\mathbf{B}_i$ – are $d \times c$ matrices.

The general ARX model can be rewritten in the polynomial form:

$$\mathbf{A}(q)\mathbf{y}(t)=\mathbf{B}(q)\mathbf{u}(t)\ +\mathbf{e}(t) \tag{2}$$

where:

$$\mathbf{A}(q)=\mathbf{I}\ +\mathbf{A}_1q^{-1}+\mathbf{A}_2q^{-2}+\ ...+\mathbf{A}_{n_a}q^{-n_a} \tag{3}$$

$$\mathbf{B}(q)=\mathbf{B}_0+\ \mathbf{B}_1q^{-1}+\mathbf{B}_2q^{-2}+...+\mathbf{B}_{n_b}q^{-n_b} \tag{4}$$

The model (2) is an ARX model where AR refer to the Autoregressive part **A**(q)**y**(t) and X refer to the extra input **B**(q)**u**(t) called the exogenous input. **y**(t) is considered as the output of the model while **u**(t) is the input to the model and **e**(t) is innovation term at the time *t*. **A**(q) and **B**(q) are polynomials in the delay operator q^{-1} and n_a, n_b are the model order of **A**(q) and **B**(q) respectively. **A**(q) is a matrix whose elements are polynomials in q^{-1}.

This results in Matrix Fraction Description (MFD).

Defining the parameter matrix:

$$\theta=[\mathbf{A}_1\ \ \mathbf{A}_2\ \ ...\ \ \mathbf{A}_{n_a}\ \ \mathbf{B}_0\ \ \mathbf{B}_1\ \ ...\ \ \mathbf{B}_{n_b}\ ]^T \tag{5}$$

We may rewrite (2) as a linear regression:

$$y(t)=\theta^T\phi(t)+e(t) \tag{6}$$

If we consider N consecutive values of the responses from *y(k)* to *y(k+N-1)*, the model parameters can be obviously estimated by least square method [15] by minimizing the norm of *e(t):*

$$\Phi=\arg\min\left(\frac{1}{N}\sum_{t=k}^{k+N-1}\left\|e(t)\right\|^2\right)=\arg\min\left(\frac{1}{N}\sum_{t=k}^{k+N-1}\left\|\mathrm{y}(t)-\theta^T\phi(t)\right\|^2\right) \tag{7}$$

After obtaining the measured force and acceleration signals on all channels, the model ARX can be used to fit the data. The ARX model creates a regressive connection between the input vector **u**(t) and the output vector **y**(t) through a residual vector **e**(t). By applying the least square method, the modal parameters matrices **A** and **B** can be estimated. In vibration measurement application, it can be seen that force (input) and acceleration (output) are normally synchronized, thus the two parts may be modeled with the same order $n_a = n_b$.

Once the model parameters of the system are identified, the state matrix can be determined as in the form of autoregressive parameters:

$$A\big|_{(naxnb)}=\begin{bmatrix}-A_1 & -A_2 & -A_3 & ... & -A_p\\ I & 0 & 0 & ... & 0\\ 0 & I & 0 & ... & 0\\ ... & ... & ... & ... & ...\\ 0 & 0 & 0 & I & 0\end{bmatrix} \tag{8}$$

There is a remarkable coincidence that the poles of model are also the roots of characteristic polynomial of the state matrix. Consequently, the continuous eigenvalues, system natural frequencies and damping rates of the structure can be calculated for each pole by using the subsequent standard equations:

Eigenvalues: $$[V,\lambda]=eig\,(A) \tag{9}$$

Frequencies: $$f_i = \frac{\sqrt{\mathrm{Re}^2(\lambda_i) + \mathrm{Im}^2(\lambda_i)}}{2\pi} \quad (10)$$

Damping rates: $$\xi_i = -\frac{\mathrm{Re}(\lambda_i)}{2\pi f_i} \quad (11)$$

When the modal parameters are estimated, we can construct the transfer function which is regarded as the frequency response function of the system. All the system can be described by linear constant coefficients and represented by transfer functions that are "rational polynomial in q".

$$\mathbf{G}(q) = \frac{\mathbf{B}(q)}{\mathbf{A}(q)} = q^{-n_k} \frac{\mathbf{B}_1 q^{-1} + \mathbf{B}_2 q^{-2} + \ldots + \mathbf{B}_{n_b} q^{-n_b}}{\mathbf{I} + \mathbf{A}_1 q^{-1} + \mathbf{A}_2 q^{-2} + \ \ldots \ + \mathbf{A}_{n_a} q^{-n_a}} \quad (12)$$

with n_k is the transport delay.

3. Application to a flexible manipulator during grinding process

3.1 Brief introduction of the SCOMPI robot

The proposed approach is now implemented to the portable robot Scompi. Figure 2 presents the structure of Scompi, which is used for repair tasks in Hydro Quebec power plants, particularly for grinding or welding jobs [1]. Because of its compact and flexible structure, the question is raised up from its dynamical behavior under operating conditions. Hence, the flexibility of the joints and links needs to be taken into consideration, which might affect the stabilization of robot at the end effector during operational process [4]. The aim of Scompi is to achieve both a high Material Removal Rate (MMR) and a polished surface finish with great precision. However, because of the portable and lightweight design, undesired chatter vibrations can appear during machining process which produces an undesirable waviness surface. Therefore, the monitoring of its modal parameters as well as the transfer functions of the structure in the grinding operation are necessary for minimizing vibration at the end effector while controlling chatter phenomenon and improve the quality of grinding surface.

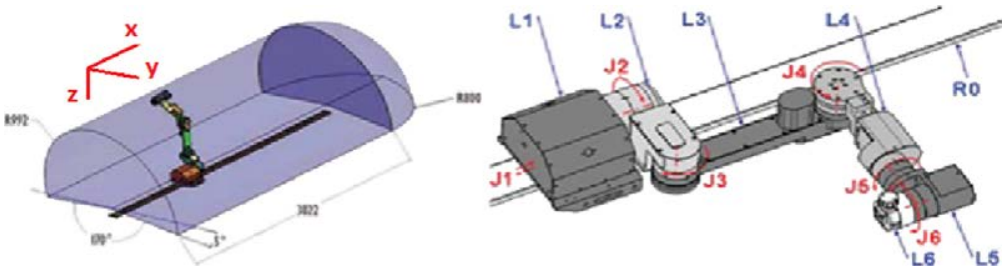


Figure 2. Scompi robot

3.2. Presentation of the Experimental setup

As can be seen from figure 3, a Scompi robot is tested under real grinding operation. Due to the interest in typical dynamic behavior of the robot at the end effector, the Scompi is set to its home configuration. Three accelerometers are mounted at the end effector in triaxial directions X, Y and Z. Meanwhile, a Kistler table

dynamometer CH8408 is placed under the work-piece for measuring the forces. The power is set up at 1500 W and grinding motor is rotated at a constant speed of 3225 (rpm) for conducting each single grinding pass within 12 seconds. A multi-component dynamometer is used for measuring the grinding forces in three directions at the tool piece contact point. After obtaining the measured signals from dynamometer and accelerometers, we acquired them to the frequency rate of 512 (Hz) (Figure 4, 5).

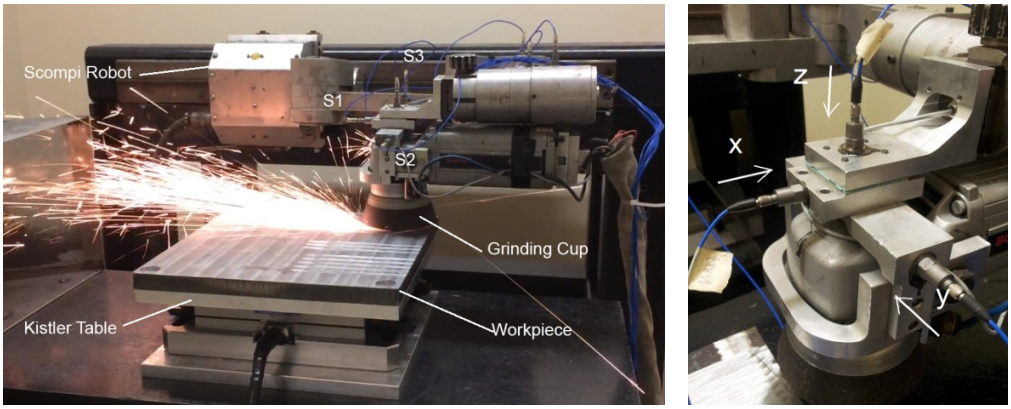


Figure 3. Overall configuration of the experimental setup

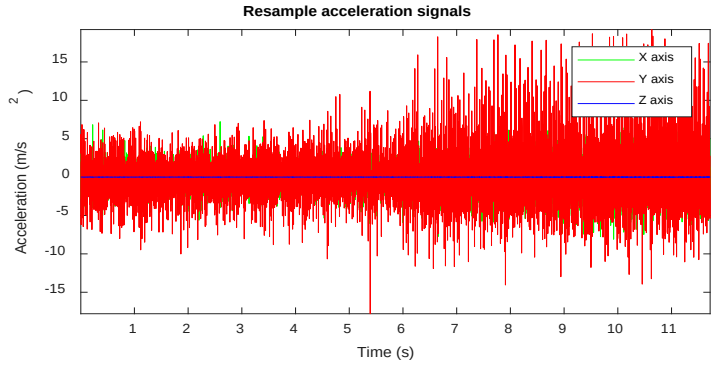


Figure 4. Measured acceleration signals during grinding process.

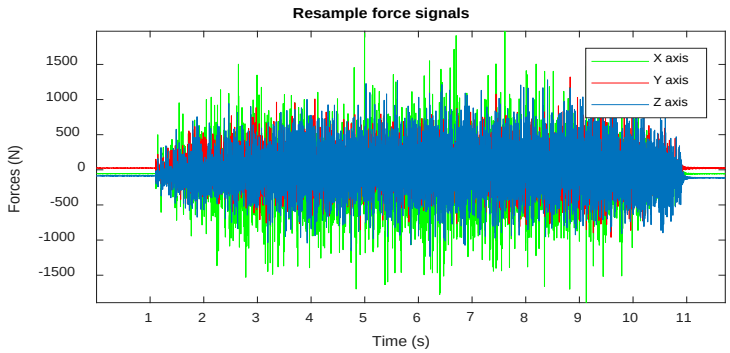


Figure 5. Measured force signals during grinding process.

Taking three measured acceleration signals in X, Y and Z directions, by applied Fast Fourier Transform (FFT) analysis, we can easily see the measured signals in both time and frequency domain as shown in Figure 6. As

indicated, there are some significant frequencies in frequency domain such as 53.9 (Hz) - the first harmonic; 93.7 (Hz); 106.2 (Hz) - the second harmonic; 146.8 (Hz) and 200.8 (Hz).

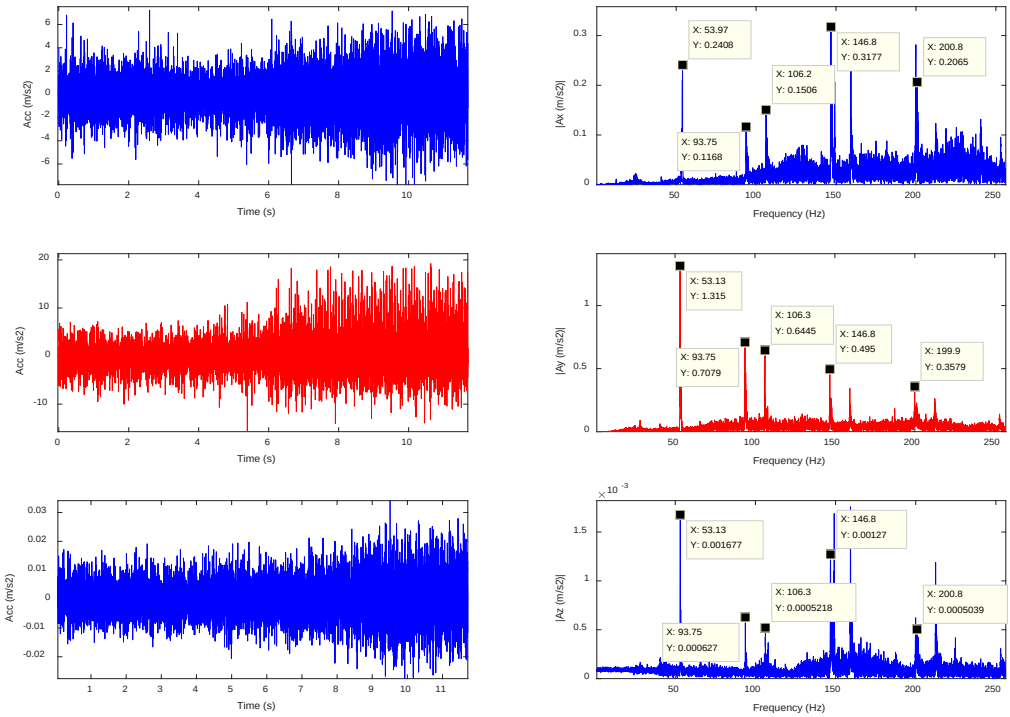


Figure 6. Time domain and frequency domain of the acceleration signals in three directions.

4. Results and discussion

Operating in a tridimensional space, the ARX model is applied on Scompi structure to fit the measured signals on each direction ($S_1 – F_x$); ($S_2 – F_y$); ($S_3 – F_z$) for constructing frequency stabilization in different cases: Single Input – Single Output (SISO) and Multi Input – Multi Output (MIMO). The figures 7–10 demonstrated the frequency stabilization diagrams up to 250 (Hz) with a model order up to 100 where all the interesting frequencies may be observed. The model order is chosen at 100 for computation of the modal parameters with low uncertainties. In addition, another stabilization given in figure 11 is computed by MODALAR based on updated AR model [4] with an aim of validation between two approaches. The 53.75 (Hz) electric frequency of grinding and its harmonics are clearly revealed in the stabilization diagrams.

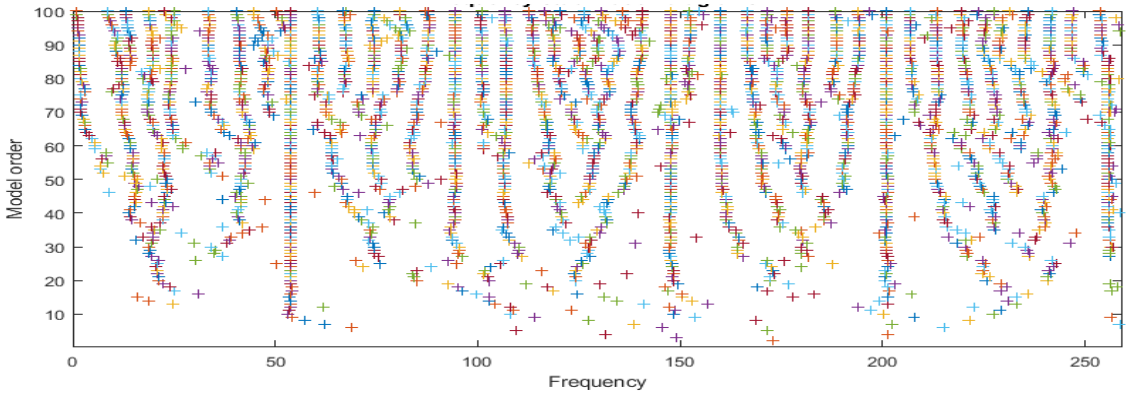


Figure 7. Frequency stabilization diagram on X direction

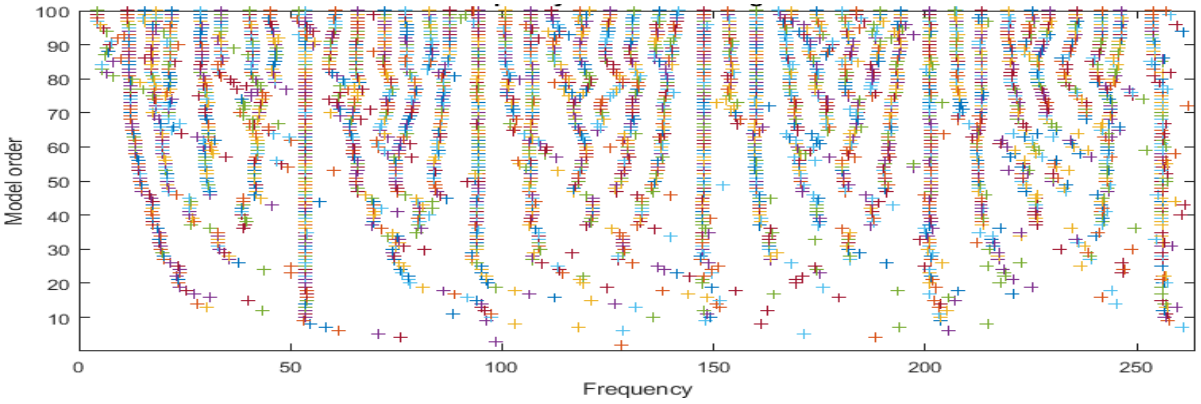


Figure 8. Frequency stabilization diagram on Y direction

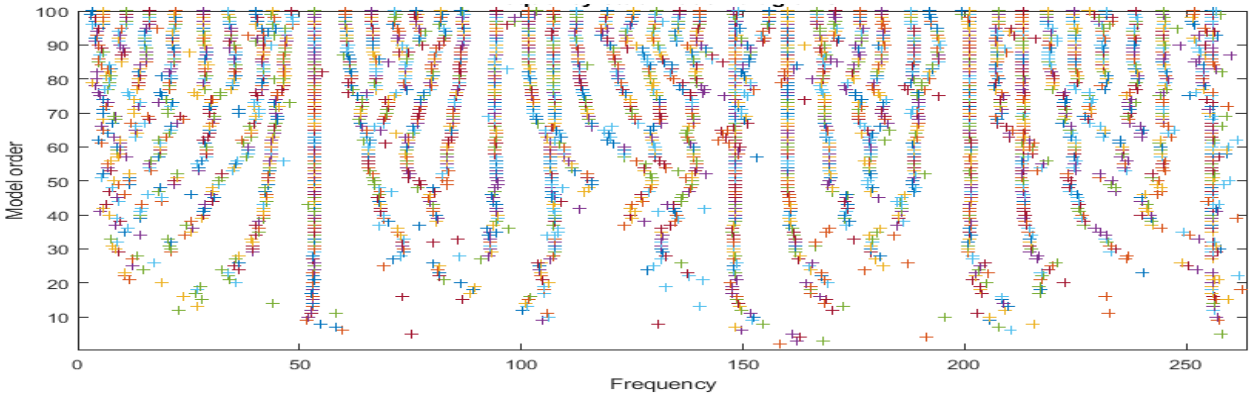


Figure 9. Frequency stabilization diagram on Z direction

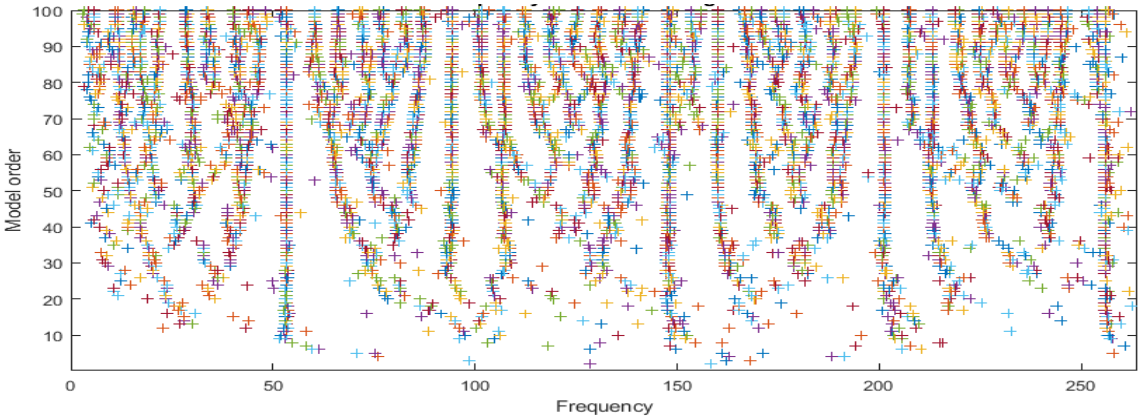


Figure 10. Frequency stabilization diagram in MIMO case (three inputs and three outputs)

Synthetically, the natural frequencies and damping ratios are estimated directly from the frequency stabilization diagram of MIMO case, where all the excited frequencies can observed clearly in multi directions. Figure 12 illustrated the stabilization diagrams of damping ratio with 95% uncertainties. The natural and harmonic frequencies identified by two methods with their damping ratio are given in the table 1. The harmonic frequencies are identified with their damping rates close to zero.

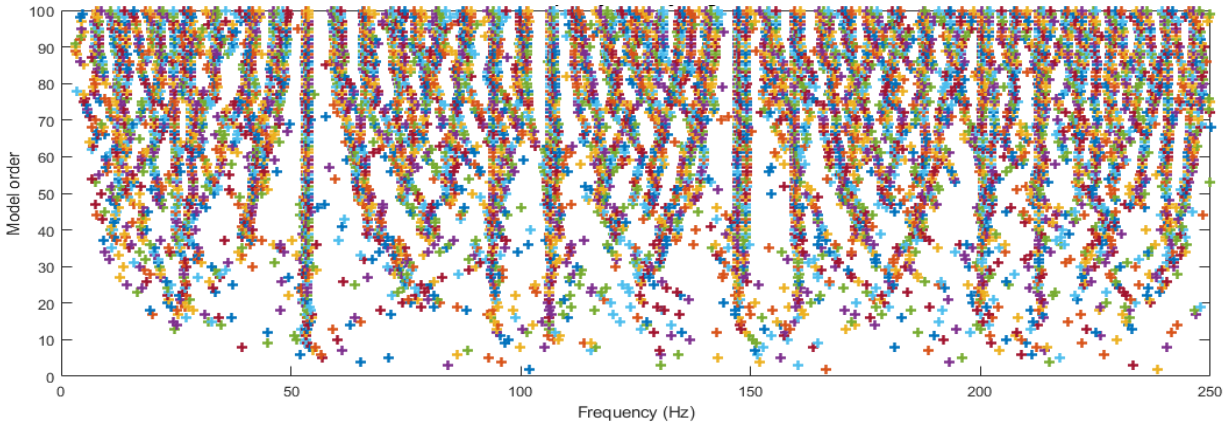


Figure 11. Frequency stabilization diagram by MODALAR

Figures 13-21 present the transfer functions identified by ARX model at the order 100. The identified transfer function from the working condition is crucial for the assessment of the robot dynamics and for further simulations under different loadings.

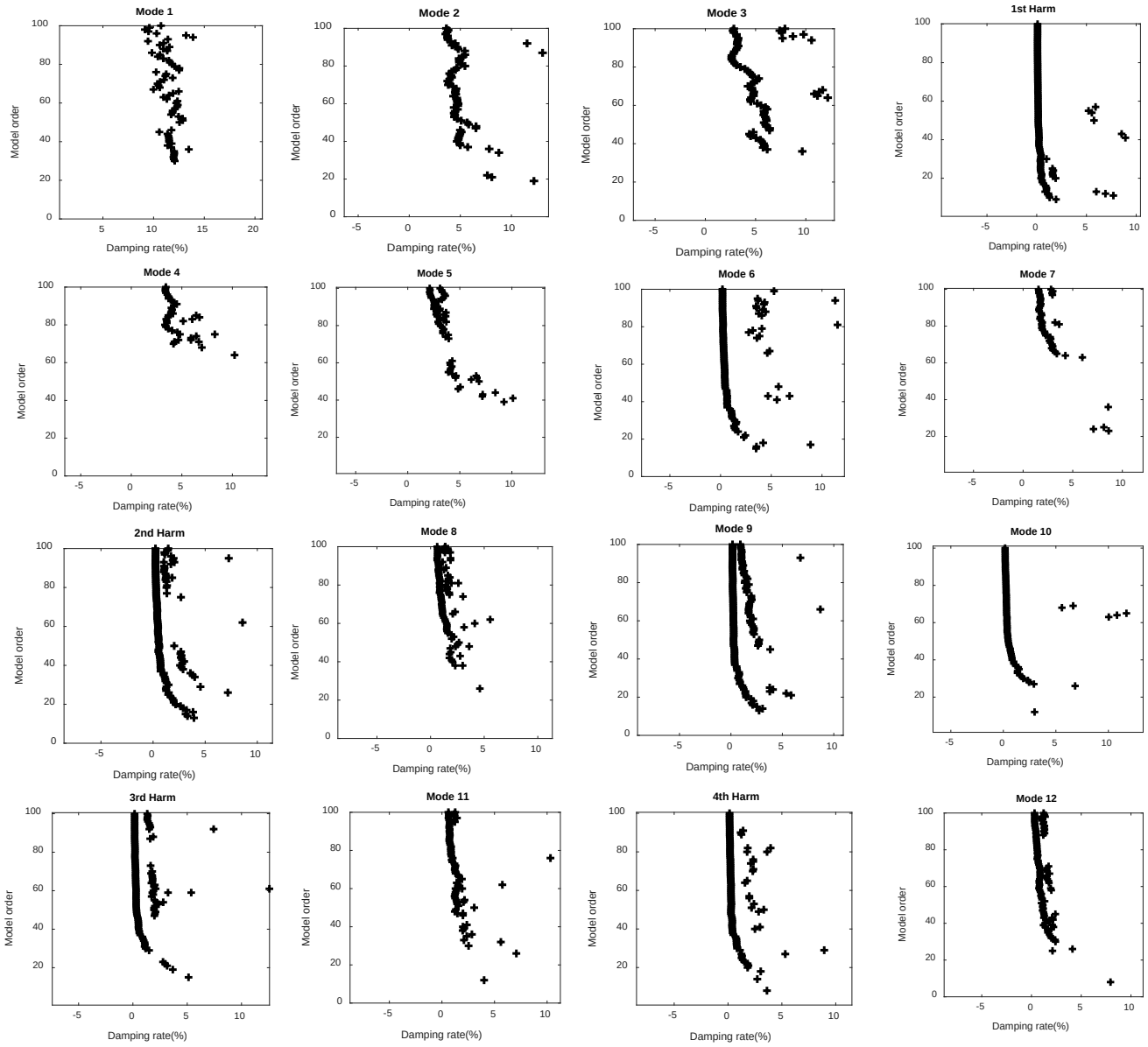


Figure 12. Damping ratio stabilization diagrams

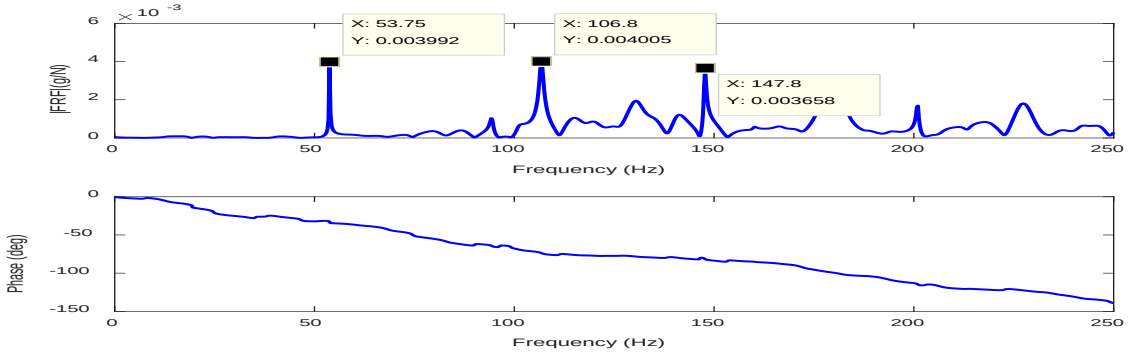


Figure 13. Identified Transfer Function FRFxx

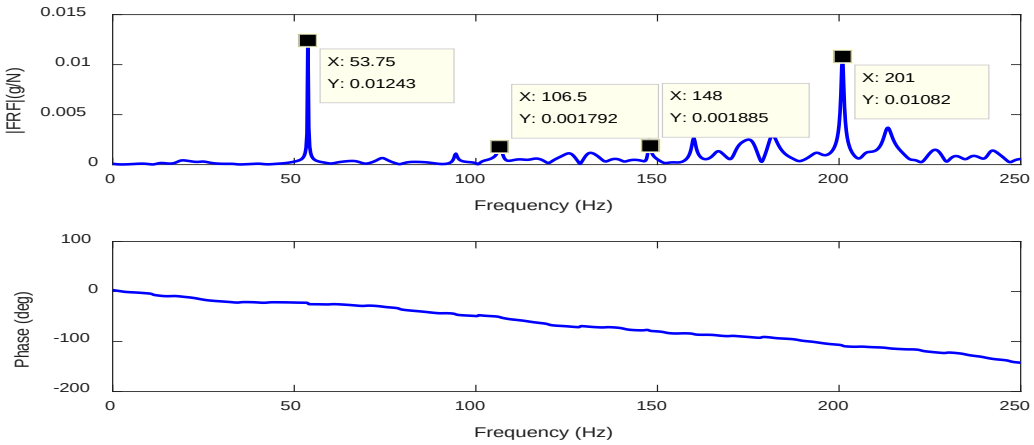


Figure 14. Identified Transfer Function FRFxy

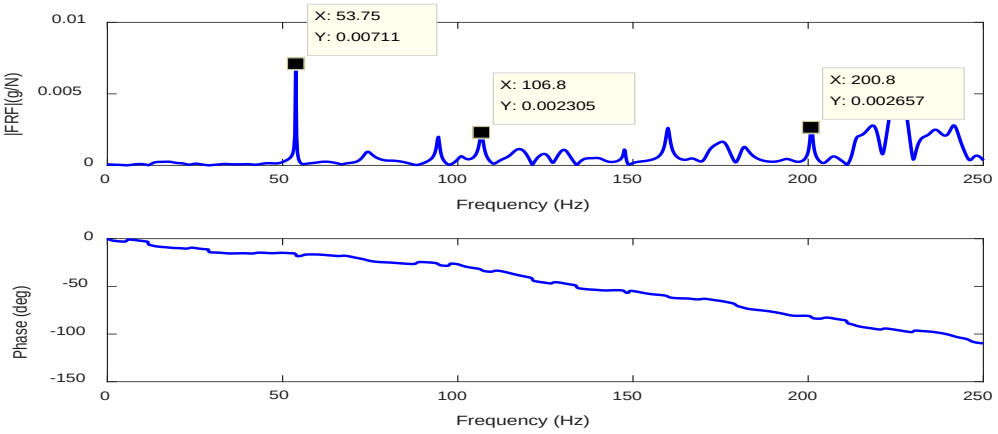


Figure 15. Identified Transfer Function FRFxz

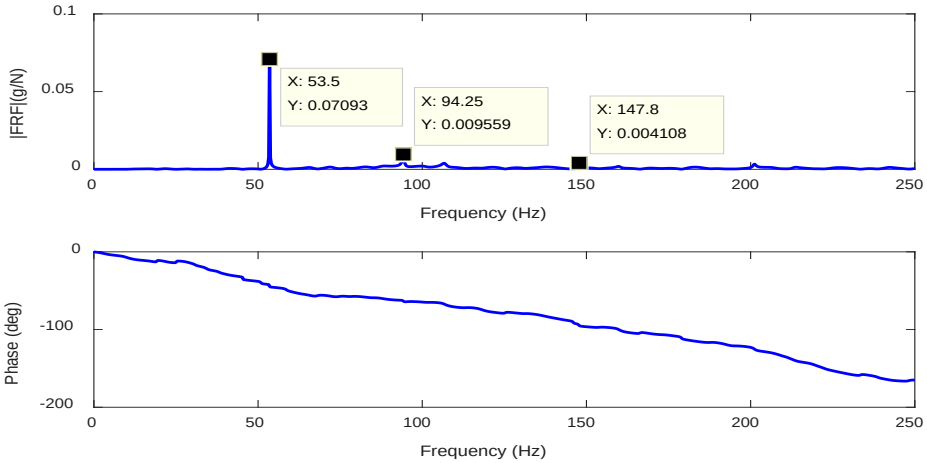


Figure 16. Identified Transfer Function FRFyx

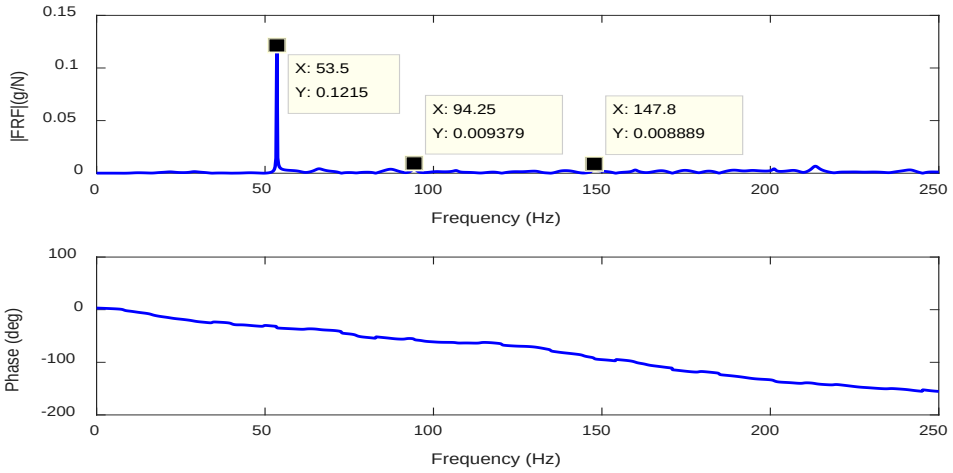


Figure 17. Identified Transfer Function FRFyy

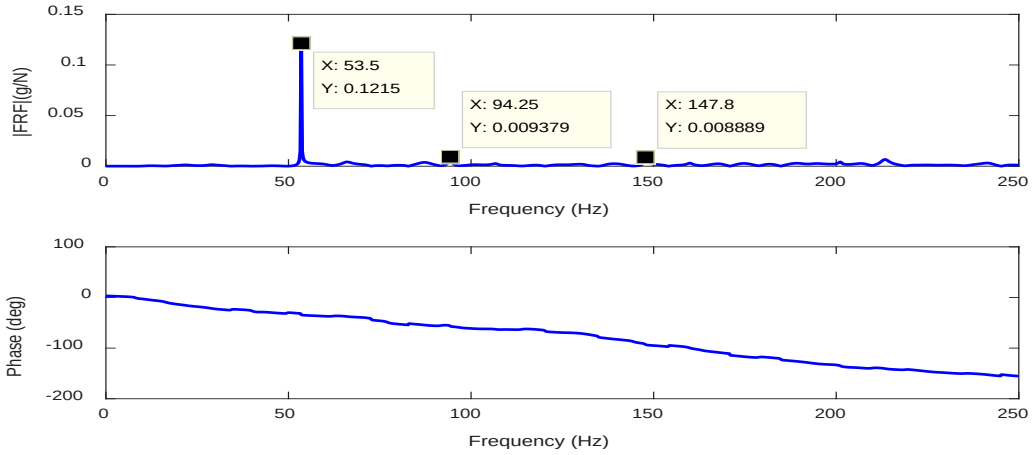


Figure 18. Identified Transfer Function FRFyz

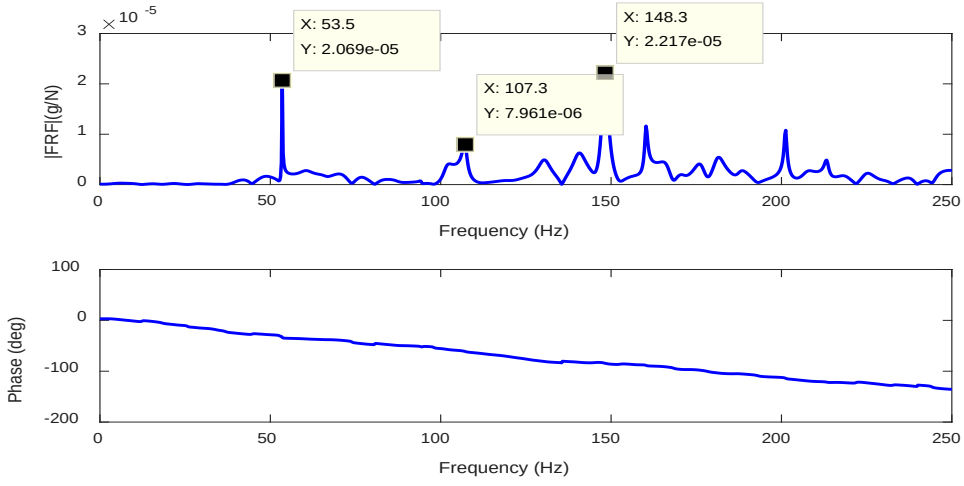


Figure 19. Identified Transfer Function FRFzx

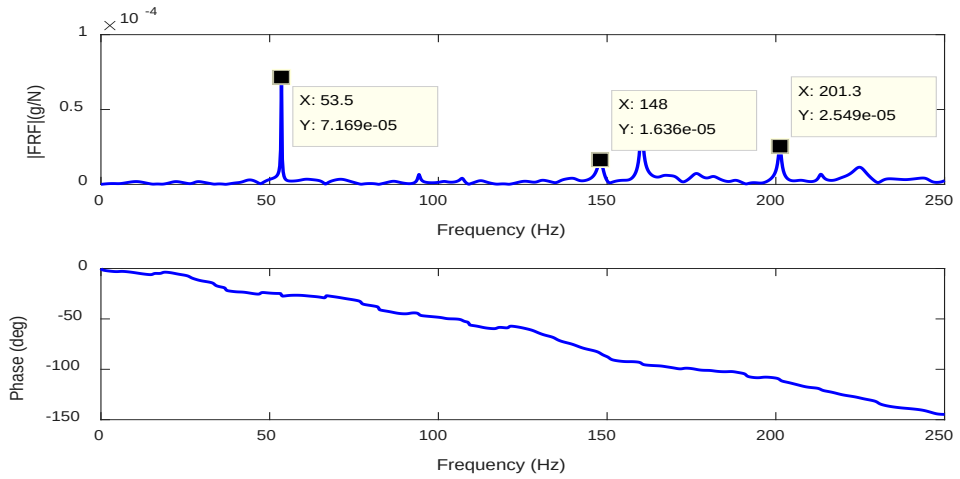


Figure 20. Identified Transfer Function FRFzy

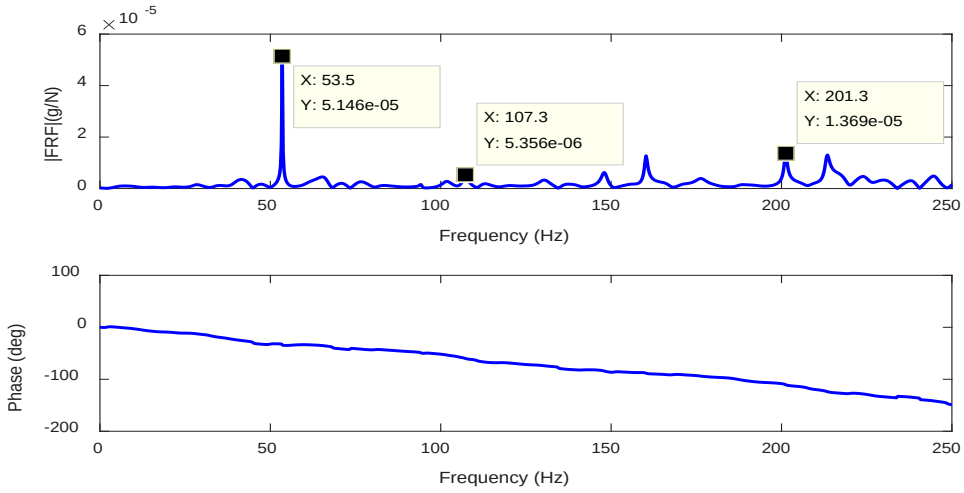


Figure 21. Identified Transfer Function FRFzz

Mode	**M 1**	**M 2**	**M 3**	**1st harm**	**M 4**	**M 5**	**M 6**	**M 7**
Frequencies (Hz) - ARX	14.2	24.4	28.2	53.3	61.2	66.2	94	101.5
Frequencies (Hz) - AR	12.7	24.6	28.2	54.2	60.1	66.2	93.9	101.3
Damping (%) - ARX	11.5	3.7	3.3	0	3.4	2.7	0.2	1.5
Mode	**2nd harm**	**M 8**	**M 9**	**M 10**	**3rd harm**	**M 11**	**4th harm**	**M 12**
Frequencies (Hz) - ARX	106.3	111.7	147.5	149.2	159.5	200.4	212.6	224.6
Frequencies (Hz) - AR	106.3	113.5	147.3	149.1	159.5	201.3	213.8	225.1
Damping (%) - ARX	0	0.6	0.2	0.2	0	0.2	0	0.7

Table 1. Identified frequencies and damping ratios

By comparison to the identified frequencies by MODALAR based on updated AR model [4] shown in table 1, the approach reveals high accuracy identified natural and harmonic frequencies with their damping ratios.

Moreover, frequency response function is directly identified from grinding operation based on ARX model. The results are better observed on the X and Y directions, this can be explained by the configuration of Scompi when working in horizontal surface to perform the grinding task.

5. Conclusion

This work is a part of an ongoing research program on investigating vibration problems of flexible manipulator. The frequencies, damping ratios and operational FRFs can be constructed and most excited modes are revealed during the grinding process. In this paper, operational FRFs of a structure are identified directly from measured signals via an ARX model. The results illustrated the sensibility of the acceleration in the X and Y directions while the contrary is proved in the Z direction with low magnitudes of the FRFs. Furthermore, as damping of the grinding process and equivalent stiffness are in command of cutting stability, so their identification is crucial to predict and avoid detrimental chatter occurrence. In the ongoing research, the inverse of ARX model will be applied in order to estimate the excitation force in the working conditions, with the integration of phase and coupling between directions. The interest lies in the reconstruction of excitation forces that gave rise to measured response signals based on ARX model. This approach is expected to serve for monitoring and vibration control design of the robot during machining operation.

Acknowledgments

The authors would like to thank IREQ (Institut de recherche d'Hydro Québec) for technical support.

References

[1] Hazel B, J. Côté, Y. Laroche and P. Mongenot, A portable, multiprocess, track-based robot for in situ work on hydropower equipment, *Journal of Field Robotics*, vol. 29, no. 1, pp. 69-101, 2012.

[2] Santos T, Z. Liu and B. Hazel, Study and validation testing of the dynamics of a robotic system with flexible links and joints, *IEEE International Conferrence on Mechatronics and Automation, Beijing, China, 2-6 Aug.* 2015, pp. 1891-1897.

[3] Vu V.H, Z. Liu, M. Thomas and B. Hazel, Modal analysis of a light-weight robot with a rotating tool installed at the end effector, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 9, pp. 1664-1676, 2015.

[4] Vu V.H, Z. Liu, M. Thomas, W. Li, and B. Hazel, Output-only identification of modal shape coupling in a flexible robot by vector autoregressive modeling, *Mechanism and Machine Theory*, vol. 97, pp. 141-154, 2016.

[5] Li. W, Vu V.H, Z. Liu, M. Thomas and B. Hazel, Extraction of modal parameters for identification of time-varying systems using data-driven stochastic subspace identification, *Journal of Vibration and Control*, vol. 24, nº 20. p. 4781-4796, 2018.

[6] Mokdad. F, Vu V.H, M. Thomas, F. Rafieian, Z. Liu and B. Hazel, Online modal analysis of a flexible robot during grinding, *Proceedings of the International operational modal analysis conference* (IOMAC 2011). Istanbul, Turkey, May-2011, paper 409, 10p.

[7] Rafieian F., Z. Liu, & B. Hazel, Dynamic model and modal testing for vibration analysis of robotic grinding process with a 6DOF flexible-joint manipulator. *In Proceedings of the IEEE International Conference on Mechatronics and Automation* (IEEE ICMA), Changchun, China, 2009.

[8] Vu V.H, Z. Liu, M. Thomas and A.M Tahvilian, Dynamic forces identification and modeling of a robotic grinding process, *Surveillance 8*, Oct. 2015, Roanne, Fr, 15p.

[9] Vu V.H, Z. Liu, M. Thomas and H. Bruce, Identification of frequency response functions of a flexible robot as tool-holder for robotic grinding process, *IEEE, IECON2016, Florence, Italy.*, 6p, Oct. 2016.

[10] Li, W., Vu V.H, Z. Liu, M. Thomas and B. Hazel, Application of adaptable functional series vector time-dependent autoregressive model for extraction of real modal parameters for identification of time-varying systems. *Measurement: Journal of the International Measurement*, vol. 103. p. 143-156, 2017.

[11] Li, W., Vu V.H, Z. Liu, M. Thomas and B. Hazel, Automatic structural mode extraction for time-varying structures via vector time-dependent autoregressive model. In *Surveillance 8*, Roanne, France, Oct. 20-21, 2015.

[12] Soderstrom T, H. Fan, B. Carlsson and S. Bigi, Least squares parameter estimation of continuous time ARX models from discrete-time data, *IEEE Transactions on Automatic Control*, vol. Vol 42, No .5, pp. 659-673, 1997.

[13] Alessandro C., The role of vector autoregressive modeling in predictor-based subspace identification, *Automatica*, vol. 43, no. 6, pp. 1034-1048, 2006.

[14] Da Silva S, M. Dias J. and V. Lopes J, Damage detection in a benchmark structure using AR-ARX models and statistical pattern recognition. *J. Braz. Soc. Mech. Sci. & Eng*. vol.29 no.2, Rio de Janeiro Apr./June 2007.

[15] Ljung L., System Identification: Theory for the User, *Prentice-Hall, 2nd Ed,* 1999.

[16] Juang J-N, State space system realization with input and output data correlation, *NASA Technical Paper 3622*, 1997.

[17] Raath A. D. and C. C. Waveren, A time domain approach to load reconstruction for durability testing, *Engineering Failure Analysis*, vol. Vol.5, No. 2., pp. 113-119, 1998.

[18] Wan Z., T. Wang, and W. Zheng, An improved state space method for force identification based on function interpolation in the presence of large noise, *Journal of Vibroengineering*, vol. 19, no. 2, pp. 751-768, 2017.

[19] Strejc V, State Space Theory of Discrete Linear Control. *IEEE Transactions on Systems*, 12 (5), 1982.

Use of virtual sensors for the analysis of forces exerted by the load inside a tumbling mill

José Venegas Pulgar[1], M. Aníbal Valenzuela[2] and Cristián Molina Vicuña[1]

[1]Laboratorio de Vibraciones Mecánicas, Departamento de Ingeniería Mecánica,
Universidad de Concepción, Chile
{jvenegasp, crimolin}@udec.cl
[2]Laboratorio de Accionamientos Eléctricos, Departamento de Ingeniería Eléctrica,
Universidad de Concepción, Chile
anivalenz@udec.cl

Abstract

In this paper, a methodology is presented to obtain average representative forces exerted by the load inside a tumbling mill on the different faces of the lifters and liners, which are directly related to its power consumption. The methodology is based on the use of virtual sensors included in DEM simulations combined with signal processing and allows obtaining the magnitude of the forces based on the angular position of the lifters as the mill rotates. The methodology is validated by comparing numerical and experimental results obtained from a test bench mill. The variables considered are the power, movement of the load inside the mill, and average forces. The latter are experimentally measured using instrumented lifters specially designed for this task. The results obtained show differences in the magnitude of the average forces in specific angular positions, depending on the operating conditions of the mill. These differences explain the behavior of power consumption with respect to operating conditions reported in the literature.

1 Introduction

Tumbling mills are critical machines of the mining industry. They are used to reduce the size of mineral particles, and their operation has associated high economic costs. The grinding of minerals requires high energy consumption and represents the most expensive stage in the production of metals. The economies of scale and the decrease in the grade of the ores, has led to the development of large grinding mills that currently reach throughput of 80,000 - 100,000 ton/day with powers of up to 28 [MW] [1][2].

The main component of the mill is the horizontal hollow cylinder called drum, inside which the ore is grinded (Figure 1a). The drum rotates on its axis and is delimited at its sides by two ends through which occurs the entry and exit of the material, respectively. Fixed to the inner wall of the drum are the coatings, composed by liners and lifters. The liners protect the drum from wear due to contact with the particles, while the lifters transmit the energy associated with the rotation of the mill to the load, producing its movement inside the mill. The comminution of the mineral occurs due to the contact forces that are generated between different particles and between particles and the internal surfaces of the drum.

The liners and lifters wear out over time as a result of the continuous contact with the particles inside the mill. The replacement of these elements constitute the primary maintenance operations and generate high costs associated not only with the purchase of replacement parts and labor but also with production losses during

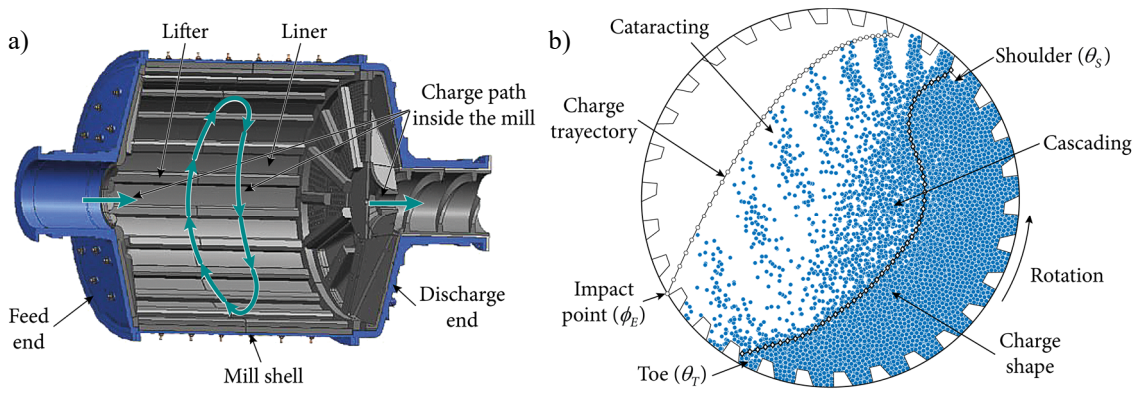


Figure 1: Tumbling mill. (a) General view of the drum. (b) View of the load inside the drum.

maintenance works. In addition, the continuous wear of the coatings produces changes in the power consumption and throughput of the mills.

While rotating, the drum lifts the grinding load along one side of the mill until reaching the point called shoulder of the load, as shown in Figure 1b. In this position, the particles located near the lifters begin to move independently of the movement of the drum and then fall describing free fall movements called cascade type or cataract type. The particles that describe cascade movements from the shoulder characterize by falling continuously, while those that describe cataract movements do so in the form of waves driven by the lifters. In the movement that describes the load inside the mill, a set of relevant angular positions, measured with respect to its axis of rotation, are identified: the position of the shoulder (θ_S), the position of the point of impact (ϕ_E) and the position of the toe (θ_T). The point of impact corresponds to the highest position at which the particles fall from the shoulder on the opposite end of the mill ring. The position of the toe is where the lifters begin to lift the load after it falls from the shoulder.

Figure 1b presents an example of the disposition of the particles inside the mills, showing the shape and trajectory of the load, the position of the toe, of the shoulder, and the point of impact. The shape of the load corresponds to the shape adopted by the set of particles that do not fall in free fall (cascading or cataracting) while the mill rotates, which commonly compares to the shape of a kidney bounded at its ends by the toe and the shoulder. Most of the particles fall from the areas near the shoulder on the internal surface of the shape of the kidney, describing a cascade-type movement, as can be seen in Figure 1. The lifters lift a portion of the particles in the shoulder to higher positions and then fall on the toe or the coatings of the drum near the toe describing a cataract movement. The lifters that leave the position of the shoulder drag small portions of particles that become independent from it gradually, forming waves of particles that are thrown into the free space inside the drum. The formation of these waves then depends on the passage of the lifters out of the shape of the load and, therefore, the impacts of the particles that describe a cataract movement on the area near the toe are not continuous, but linked to the movement of lifters. The trajectory of the load corresponds to the free fall movement that describes the particle that reaches the point of impact.

The Discrete Elements Method (DEM) is a numerical methodology that describes the behavior of granular materials. It allows simulating the movement of each of the particles forming the grinding load inside the mill by modeling the interactions between the different particles and between particles and surfaces, and solving the equations of motion of each particle. DEM has been used by multiple researchers to study tumbling mills focused on, for example, the analysis of the load movement [2][3][4][5][6][7][8][9][10][11][12][13][14][15][16][21][23][26][27][28][32][33][34][45], the study of wear of coatings [2][14][15][30][43][44] and the modeling of the comminution process [16][5][8][17][18][19][20][22][31][35][37][38][39][40][41][42]. This paper focuses on analyzing the power requirements of the mills, a topic that has also been discussed in the literature [13][14][45][49][21][24][25][26][29][33]. Different from other researches, this is done by determining the average forces exerted by the particles on the lifters and liners as a function of the angular position in which they are located. Some studies relate globally the behavior of the load with the contact forces by using different methods [47][48][49][50][51][52][54][55][56]. The correlation between the average forces and the power obtained in this work allows identifying the physical phenomena that explain the observed power variations as a function of the operating conditions.

2 Test bench: SetupD100

The analyses are carried out based on a laboratory scale mill, called SetupD100, shown in Figure 2a. This mill consists of three main components: the ring, a back cover, and a front cover. The ring is a hollow cylinder representing the drum of the mill with lifters mounted on its inside. The ring and lifters are made of technyl. The ring has an internal diameter of 945 [mm], an internal length of 60 [mm] and is delimited at its ends by the back and front cover, respectively. Both covers are made of acrylic. The back cover is gray, while the front is transparent, which allows observing the movement of the load while the mill runs. The mill is connected to the electric drive by a drive shaft in a cantilever arrangement. The drive includes a frequency converter that allows controlling the speed of rotation.

2.1 Lifter geometry

The internal geometry of the mills is one of the main aspects to analyze in order to understand the behavior of the load inside the mill. It is defined mainly by the number of lifters (N_{lft}) and their geometry. Figure 2a shows the dimensions of the lifters installed in the test bench mill.

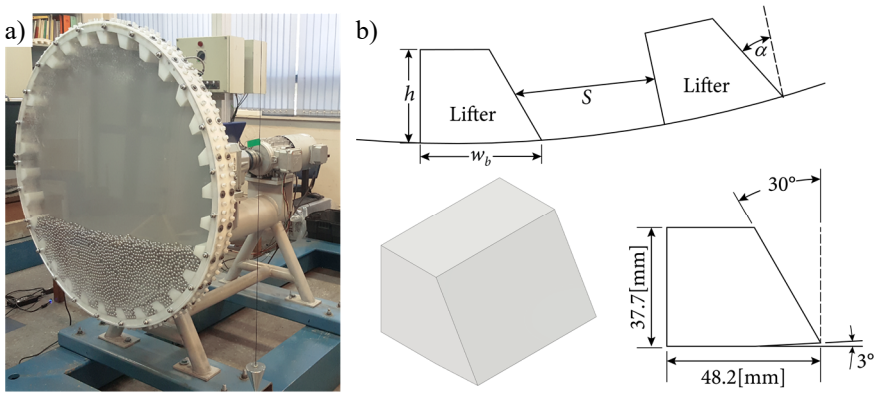


Figure 2: (a) View of the Setup D100. (b) Dimensions of the lifters.

2.2 Operating conditions

The rotating speed and fill level of the mill define its operating condition. The numerical and experimental analyses presented in this paper consider 81 different combinations of speed and fill levels presented in Table 1.

The rotation speed of the mill is defined as a fraction of its critical speed (N). The critical speed of a mill (N_c) corresponds to the speed of rotation from which the load begins to centrifuge, adhering to the internal surfaces of the drum. It is calculated as [47]:

$$N_c = \frac{42.3}{\sqrt{D_M}} \tag{1}$$

The critical speed of the SetupD100 is 4.556 [rad/s], and the 9 speeds analyzed vary from 55% to 95% of it.

The fill level (J_c) is the fraction of the internal volume of the mill that is occupied by the grinding load. The SetupD100 has an internal volume of 0.0396 [m^3] and is filled with 11 [mm] diameter steel balls. The 9 fill levels analyzed range from 25% to 45% of the mill's internal volume, with 2.5% jumps

3 Numerical modeling of the SetupD100

The numerical model of the test bench consists of two main components: a geometric model and a contact model. The geometric model represents the surfaces of the mill with which the particles come into contact, while the contact model describes the interactions between the particles located inside the mill and the components of the geometric model, and between different particles.

3.1 Geometric model of the SetupD100

The geometric model used in this investigation is composed of 3 elements: the ring and the two covers. In the geometric model, the ring and the lifters are considered as a single element. Figure 3a shows the three components of the geometric model, and Figure 3b shows a view of the load inside the mill together with the cartesian system used as a reference for the analyses.

Fraction of the critical speed	Rotation speed, [rad/s]
55%	2.506
60%	2.734
65%	2.961
70%	3.190
75%	3.418
80%	3.645
85%	3.874
90%	4.101
95%	4.329

Fill level	Mass, [kg]
25%	46.25
27.5%	50.87
30%	55.49
32.5%	60.12
35%	64.74
37.5%	69.37
40%	73.99
42.5%	78.61
45%	83.24

Table 1: Operating conditions.

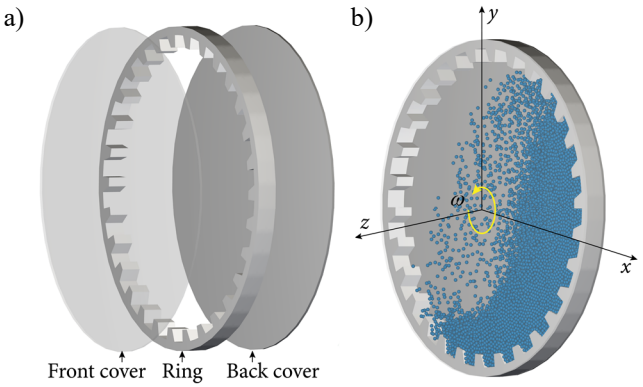


Figure 3: (a) Components of the geometric model. (b) Cartesian reference system.

3.2 Contact model

The particles inside the mill can contact other particles, the covers, or the ring. This means the existence of three different contact types: steel-steel contact between the steel balls, steel-acrylic contact between the steel balls and the acrylic covers, and steel-technyl contact between the steel balls and the ring. The contact models used in DEM allow calculating the forces associated with the contacts to which all the particles are subject, but for this, it is necessary to define the physical parameters that characterize all possible contacts. These parameters are the static friction coefficient, the dynamic friction coefficient, the coefficient of restitution, and the coefficient of rolling resistance. Table 2 shows the values of the contact parameters used in the DEM simulations.

3.3 Power due to the movement of the load inside the mill

Considering the SetupD100 operating as shown in Figure 3b, it can be noted that in any instant of time, not all particles are in contact with the internal surfaces of the mill. It is also noted that a particle can be in contact with the internal surfaces of the mill in more than one point (maximum three) and that there is a given number of contacts (n) between particles and internal surfaces of the mill.

Now, consider a particle that is in contact with one of the internal surfaces of the mill, and that this contact i occurs in a position $\vec{r}_i$ with respect to the axis of rotation of the mill. Let $\vec{F}_i$ be the force exerted by the particle on the point of contact i, as shown in Figure 4a:

$$\vec{r}_i = r_{x,i}\hat{i} + r_{y,i}\hat{j} \tag{2}$$

$$\vec{F}_i = F_{x,i}\hat{i} + F_{y,i}\hat{j} \tag{3}$$

The torque exerted by the particle on the surface at contact i is, thus, given by:

$$\vec{T}_i = \vec{r}_i \times \vec{F}_i \tag{4}$$

Finally, taking into account the n existing contacts between particles and internal surfaces of the mill during a time instant, the torque associated with the movement of the particles inside the mill is given by:

$$\vec{T}_{Mill} = \sum_{i=1}^{n} \vec{T}_i \tag{5}$$

The torque allows to calculate the power due to the operation of the mill at speed ω:

Contact	Coefficient of rolling resistance (μ_r)	Static friction coefficient (μ_s)	Dynamic friction coefficient (μ_k)	Coefficient of restitution (c_R)
	0.01	-	-	-
Steel–Steel	-	0.43	0.36	0.91
Steel–Acrylic	-	0.42	0.38	0.44
Steel – Technyl	-	0.23	0.22	0.69

Table 2: Contact parameters used in the DEM simulations.

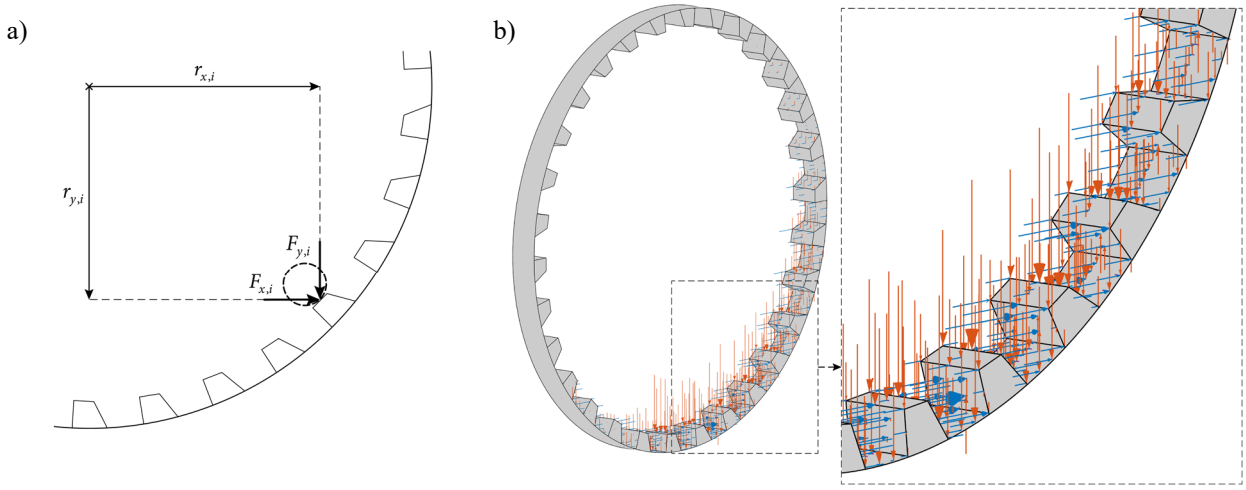


Figure 4: (a) Contact forces of a single particle in contact with the ring. (b) All contact forces acting on the ring with a close view of the contact forces.

$$P_{Mill} = T_{Mill}\omega \tag{6}$$

Thus, the power associated with the movement of the mill originates from the contact between the particles and their internal surfaces. The contacts between the particles contribute to the power indirectly by action/reaction forces, and the Discrete Element Method (DEM) is based on calculating the forces associated with the contacts between particles and between particles and surfaces.

3.3.1 Forces acting on the ring and covers

Figure 4b shows the components of the nodal forces acting on the ring at a given time instant, with a close view of a section of the ring. The blue arrows represent the components of the forces on the *x*-axis; the red arrows represent the components of the forces on the *y*-axis. There are also small purple arrows representing the force components on the *z*-axis. The size of the arrows is scaled based on the magnitude of the forces they represent. As observed, the components of the forces on the *z*-axis are much smaller than the other two components, being so small that they are not recognizable in Figure 4b. The latter, added to the fact that these forces do not contribute to the power, allows treating the analysis of the forces as bidimensional.

Consider the disposition of particles during a given time instant shown in Figure 5a and the forces excerted by these particles on the ring, shown in Figure 5b. It can be seen that the magnitude of the forces exerted by the particles impacting the ring before the toe has a low magnitude and that the amplitude of the forces grows as they come into contact with the ring in the zone between the toe and the shoulder. Also, it is possible to notice that the vertical components of the nodal forces act mainly in the negative direction while the forces in the horizontal direction are positive or negative depending on the face of the lifter in which they act. The sense of the horizontal forces determines the sense of the torque they generate.

Figure 6 shows the forces of Figure 5 decomposed in the ring tangential direction. These components are those that contribute to the torque associated with the movement of the particles. It can be seen that the forces acting on the back face of the lifters (purple arrows) generate torque in the same direction as the rotation of the

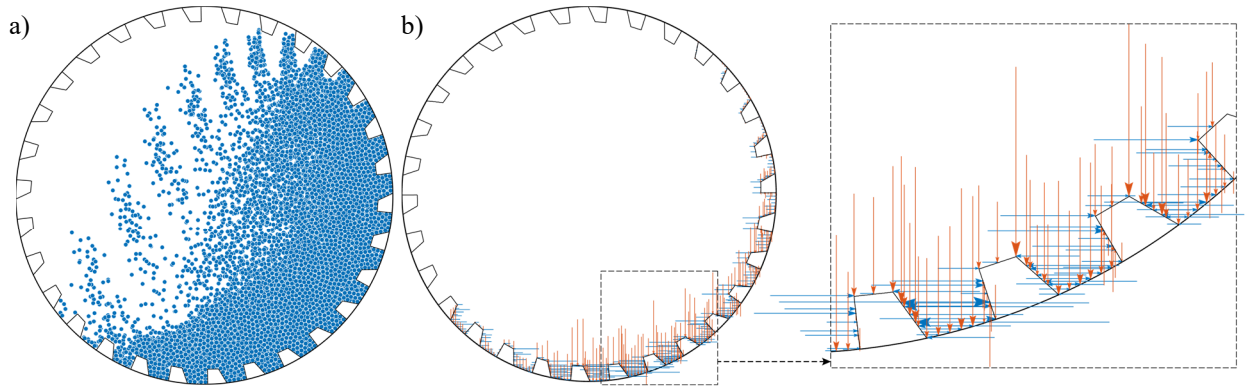


Figure 5: (a) Disposition of particles at a given time instant. (b) Forces excerted by the particles on the lifters and liners at a given instant of time.

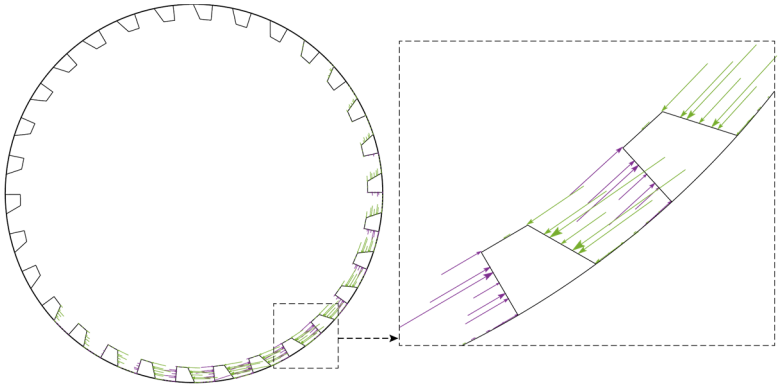


Figure 6: Decomposition of forces in tangential direction.

mill, while the forces acting on the front face of the lifters (green arrows) generate torque in the opposite direction. It is also possible to notice that the components of the forces that contribute to the torque and that act on the liners or the top faces of the lifters are of very low amplitude when compared with the components that act on the front and back faces. Finally, it can be seen that the magnitude of the tangential forces varies significantly with the angular position. In this work, the power associated with the forces exerted by the particles on the ring and obtained using DEM is called $P_{DEM,Ring}$.

The main consequence of this analysis is that, depending on the face of the lifter in which the forces act, they can contribute or oppose the total torque exerted by the load, which determines the required driving torque. The evolution of these forces, depending on the internal geometry of the mill and the operating conditions, have a significant influence on the power consumption associated with its operation.

Figure 7a shows the components of the forces in x-axis and y-axis acting due to friction on one of the acrylic covers of the mill. Figure 7b shows the tangential decomposition of these forces, where the size of the arrows is according to their magnitude. As observed, the forces on the covers that generate torques are of similar magnitude to those acting on the ring. The magnitude of the power associated with the back cover and front cover is similar and, therefore, both covers are treated indistinctly in this work.

The power associated with the movement of both covers is defined as $P_{DEM,Covers}$ and is calculated by adding the powers associated with each cover.

3.3.2 Total power due to the movement of the load inside the mill

The total power associated with the operation of the mill obtained by modeling the SetupD100 using DEM corresponds to the sum of the powers associated with each of its components:

$$P_{DEM,Mill} = P_{DEM,Ring} + P_{DEM,Cover1} + P_{DEM,Cover2} \tag{7}$$

Most of the power associated with the movement of particles inside industrial scale mills comes from the contact forces acting on the liners and lifters. In the case of the power associated with the operation of the SetupD100, both the power associated with the ring and the covers are of similar magnitude because the length of the ring is too short for its diameter.

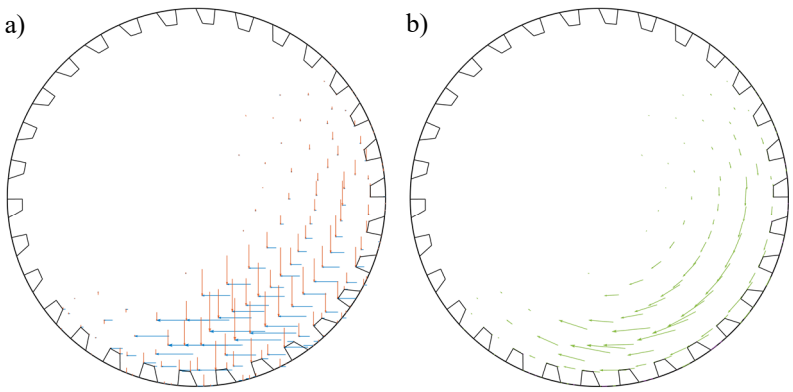


Figure 7: Forces acting on one cover (a) in x and y direction. (b) Tangential decomposition of the forces.

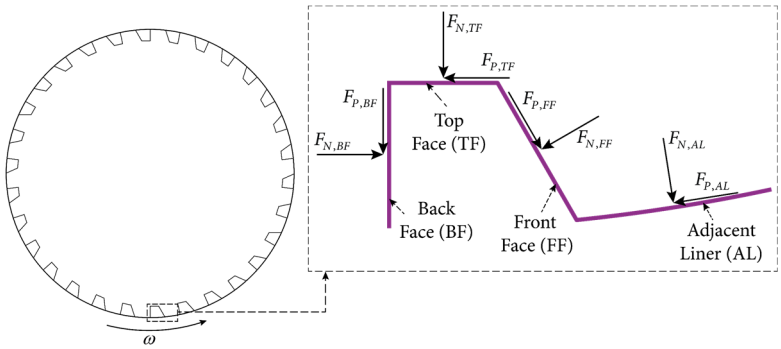


Figure 8: Normal and parallel forces acting on the 3 faces of the lifter and adyacent liner.

4 Model of forces acting on the lifters

As explained in section 3.3.1, the magnitude of the forces exerted by the particles on the ring when the mill rotates depends on the angular position of the liners and lifters, and on the lifter face on which they act. Motivated by this, a force model is developed, based on DEM simulation results. The model allows obtaining the magnitude of the average forces acting on the different faces of the lifters and liners as a function of its angular position θ while the mill operates. This model groups the contact forces exerted by the particles on the liners and each face of the lifters of the mill. The internal geometry of the mill divides into four faces: the front face (FF), the back face (BF), the top face (TF) and the adjacent liner (AL). The forces acting on the ring are divided into 8 forces that act normal and tangential to each face, as shown in Figure 8.

4.1 Modification of the geometric model: implementation of virtual sensors

To obtain the magnitude of the average forces exerted by the particles on the 3 faces of the lifters and adjacent liners as a function of their angular position, a set of three virtual sensors is incorporated in the geometrical model of the DEM model of the SetupD100, as shown in Figure 9a. The virtual sensors (VS) consist of 4 independent sheets (one for each face of the lifter plus the adjacent liner) fixed to the lifters they cover. The fact that the sheets are independent geometries allows obtaining the resultant force exerted by the particles on each sheet as a function of time separately.

Figure 9b shows an example of the forces acting on one of the sheets covering the front face (FF) of the lifter of one of the virtual sensors, as directly obtained from the DEM software, that is, in terms of the *xy*-coordinates. Since there are 3 virtual sensors with their respective 4 sheets, a total of 12 *x* and *y* force signals as a function of time is obtained from each simulation.

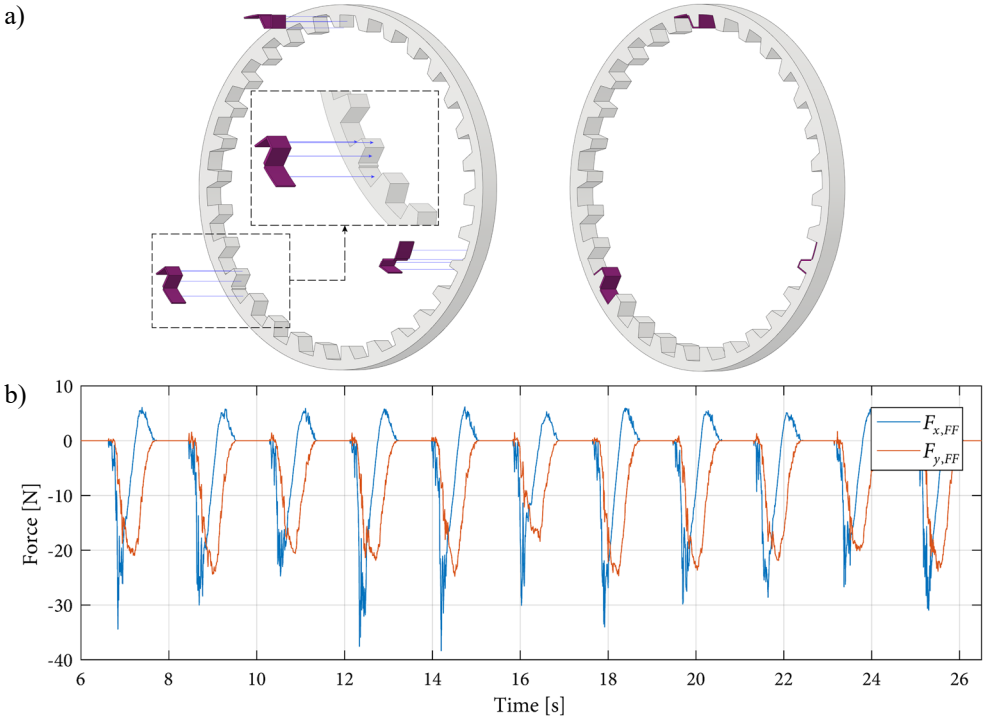


Figure 9: (a) Set of virtual sensors included in the geometric model. (b) Example of time history of force signal in *xy* coordinate for the front face of one lifter.

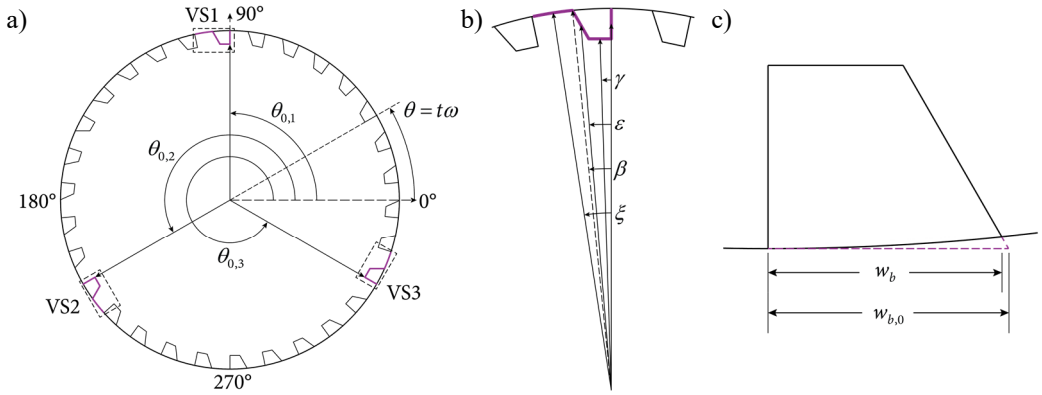


Figure 10: a) Starting position of the virtual lifters. b) Angles used to define the center position of the lifter faces and adjacent liner. c) Projected width of lifter.

4.2 Transformation of forces into local coordinates

The first step in the processing of the forces is to transform the forces captured by the virtual sensors, based on the global *xy* coordinates (Figure 9b), in terms of the local coordinates defined for each face according to Figure 8. The global coordinates are fixed in space so that the mill rotates with respect to the *xy* plane. On the other hand, the virtual sensors move with the lifters on which they are mounted and, consequently, rotate with the mill. Because of this, all local coordinates, defined for each of the faces of the VS, also rotate with the mill. Hence, to transform the forces from global to local coordinates, it is necessary to know the location of each of the faces of the VS as a function of time, based on the rotation of the mill.

4.2.1 Angular displacement equations of virtual sensor faces

In order to know the location of the different faces of the virtual sensors with respect to the axis of the mill, the equations describing the angular displacement of the centers of the 3 lifter faces and adjacent liner are considered. Figure 10a shows the starting position of the three virtual sensors (VS), where the back face of VS1 is located at 90° with respect to the horizontal, that is, $\theta_{0,1} = \pi/2$. The back face of the VS2 and VS3 are located at 120° and 240° from the back face of the SV1, respectively. The angular position of the centers of the other faces is defined relative to the back face of the respective virtual sensor, as shown in Figure 10b. The angles shown in the figure are defined according to the geometry of the lifters and the ring. Due to how the lifters are installed in the ring (Figure 10c) it is convenient to define a projected width of the lifter $w_{b,0}$, which represents the width of the base of the lifter if the curvature of the ring did not exist. Doing so allows defining the angles between the centers of the faces with respect to the back face of the lifter as:

$$\gamma = \tan^{-1}\left(\frac{w_{b,0} - h\tan(\alpha)}{D_M - 2h}\right) \tag{8}$$

$$\varepsilon = \tan^{-1}\left(\frac{2w_{b,0} - h\tan(\alpha)}{D_M - h}\right) \tag{9}$$

$$\beta = 2(\varepsilon - \gamma) \tag{10}$$

$$\xi = \frac{\pi}{N_{lft}} + \frac{\beta}{2} \tag{11}$$

For the geometry and number of lifters of the SetupD100, the angles are $\gamma = 1.836°$, $\varepsilon = 4.762°$, $\beta = 5.853°$ y $\xi = 8.927°$. After locating the different faces of the VS with respect to their corresponding back faces, it is possible to describe the movement of each of the faces of the 3 VS as a function of time, based on the rotation speed of the mill as follows:

$$\theta_{BF,i}(t) = \theta_{0,i} + \omega t \tag{12}$$

$$\theta_{TF,i}(t) = \theta_{BF,i} + \gamma \tag{13}$$

$$\theta_{FF,i}(t) = \theta_{BF,i} + \varepsilon \tag{14}$$

$$\theta_{AL,i}(t) = \theta_{BF,i} + \xi \tag{15}$$

Where ω is the rotation speed of the mill, t corresponds to time, the index $i=1,2,3$ refers to the virtual sensor and $\theta_{0,i}$ represents the initial position of the back face of the virtual sensor i.

4.2.2 Transformation of forces into local coordinates

Once the location of the 4 faces of the 3 virtual sensors is known as a function of time, it is possible to transform the forces obtained from the virtual sensors from global xy coordinates into local coordinates (normal and parallel) by using the following equations:

For forces acting on the front face:

$$F_{N,FF,i}(t) = F_{x,FF,i}(t)\sin\left(\alpha + \theta_{FF,i}(t)\right) - F_{y,FF,i}(t)\cos\left(\alpha + \theta_{FF,i}(t)\right) \quad (16)$$

$$F_{P,FF,i}(t) = F_{x,FF,i}(t)\cos\left(\alpha + \theta_{FF,i}(t)\right) - F_{y,FF,i}(t)\sin\left(\alpha + \theta_{FF,i}(t)\right) \quad (17)$$

For forces acting on the adjacent liner:

$$F_{N,AL,i}(t) = F_{x,AL,i}(t)\cos\left(\theta_{AL,i}(t)\right) + F_{y,AL,i}(t)\sin\left(\theta_{AL,i}(t)\right) \quad (18)$$

$$F_{P,AL,i}(t) = F_{x,AL,i}(t)\sin\left(\theta_{AL,i}(t)\right) + F_{y,AL,i}(t)\cos\left(\theta_{AL,i}(t)\right) \quad (19)$$

For forces acting on the back face:

$$F_{N,BF,i}(t) = -F_{x,BF,i}(t)\sin\left(\theta_{BF,i}(t)\right) + F_{y,BF,i}(t)\cos\left(\theta_{BF,i}(t)\right) \quad (20)$$

$$F_{P,BF,i}(t) = F_{x,BF,i}(t)\cos\left(\theta_{BF,i}(t)\right) + F_{y,BF,i}(t)\sin\left(\theta_{BF,i}(t)\right) \quad (21)$$

For forces acting on the top face:

$$F_{N,TF,i}(t) = F_{x,TF,i}(t)\cos\left(\theta_{TF,i}(t)\right) + F_{y,TF,i}(t)\sin\left(\theta_{TF,i}(t)\right) \quad (22)$$

$$F_{P,TF,i}(t) = F_{x,TF,i}(t)\sin\left(\theta_{TF,i}(t)\right) - F_{y,TF,i}(t)\cos\left(\theta_{TF,i}(t)\right) \quad (23)$$

Figure 11a shows the forces of Figure 9b, acting on the front face of one of the virtual sensors, after the change of coordinates. After the change of coordinates, it is possible to notice that the magnitude of the normal forces acting on the faces of the lifters is significantly larger than that of the parallel forces.

4.1 Separation in force pulses

As observed in Figure 11a, the magnitude of the forces changes from zero to non-zero in consecutive time intervals, a situation that repeats with every rotation of the mill. This is because the VS is not in contact with the particles when it is located between the shoulder and the toe or the point of impact (whatever occurs first) and, therefore, is not subjected to forces during this time interval. Contrary, when the VS is in contact with particles, the force is non-zero. This process of loading and unloading the VSs is approximately periodic so that lifters and liners are subject to similar forces as they pass through the same angular position.

Based on the above, the forces are subjected to a process of pulse separation. A pulse corresponds to a set of the 8 forces (two for each face, normal and parallel) exerted by the particles on one of the virtual sensors

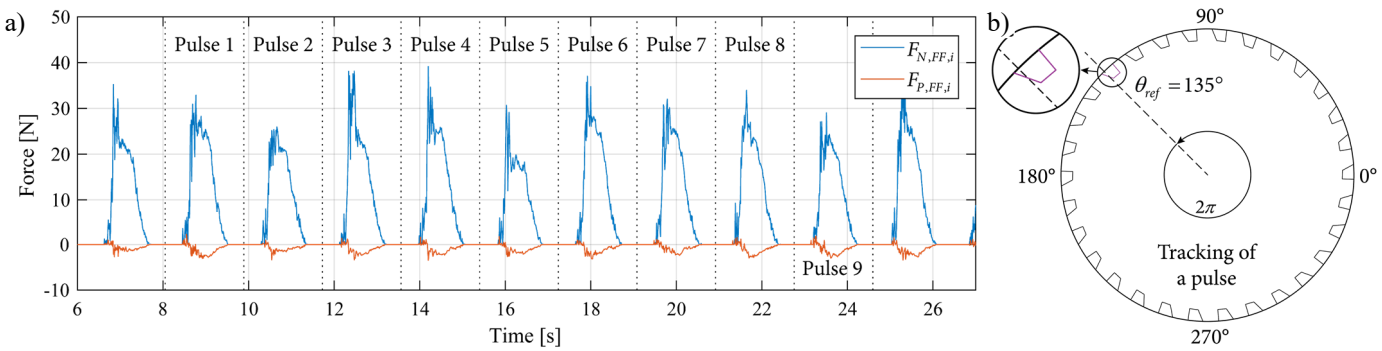


Figure 11: (a) Forces acting on the front face of a SV. Dotted lines mark the beginning and end of the force pulses. (b) Path followed by a SV for pulse separation with reference position of 135°.

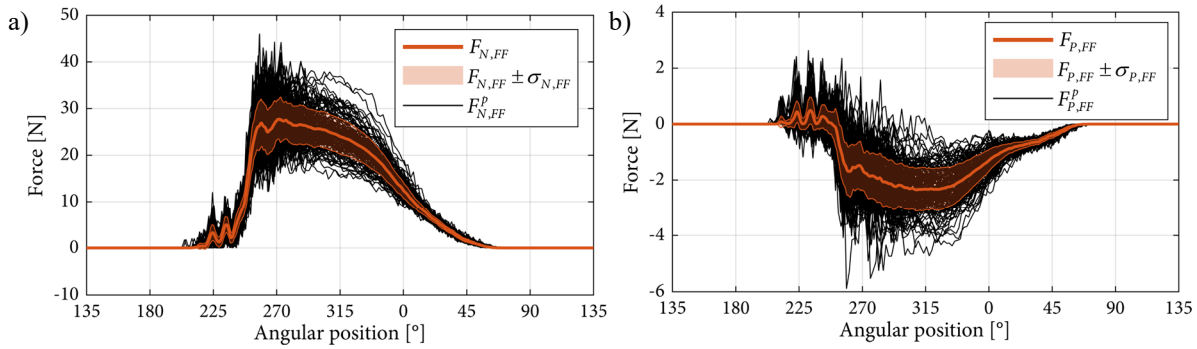


Figure 12: Averaged (red) and individual (black) pulse forces acting on the front face of the lifter. (a) Normal force. (b) Parallel force.

for a complete revolution of the mill around its axis of rotation. The pulse separation allows relating the magnitude of the forces exerted by the particles on the virtual sensors as a function of the angular position instead of time.

The pulse separation is carried out using an algorithm that defines a reference position (θ_{ref}) and identifies the instants of time where the front face of the VS pass through said position. Figure 11b shows the path followed by a VS during which one pulse is obtained, with respect to the reference position (135° in this case). Thus, the pulses correspond to the set of forces exerted by the particles during one rotational period of the mill based on a reference position. In Figure 11a, the vertical dotted lines represent the initial and final time instants of the different pulses. The first pulse is not considered representative because the movement of the load develops only after the first seconds of the simulation.

4.2 Averaging process of pulse forces

Once separated into pulses, the forces are subjected to a process of averaging for deterministic/random separation. The process consists of averaging the magnitude of the pulse forces of the respective faces of the SV when they are in corresponding angular positions. The result of the process is the set of 8 average pulse forces that act on each of the faces of the lifters as a function of the angular position: $F_{N,FF}$, $F_{P,FF}$, $F_{N,AL}$, $F_{P,AL}$, $F_{N,BF}$, $F_{P,BF}$, $F_{N,TF}$ y $F_{P,TF}$. Figure 12a and Figure 12b show, respectively, the average pulse forces $F_{N,FF}$ and $F_{T,FF}$ superimposed on the individual pulses from which they are calculated. Figure 13 shows the forces of Figure 11 (in blue), along with the average force pulses (in red) and the random part of the forces (in yellow). In general, the magnitude of the random part of the forces is low compared to the magnitude of the average forces, especially for normal forces. The average forces represent the stationary forces acting on any of the lifters as

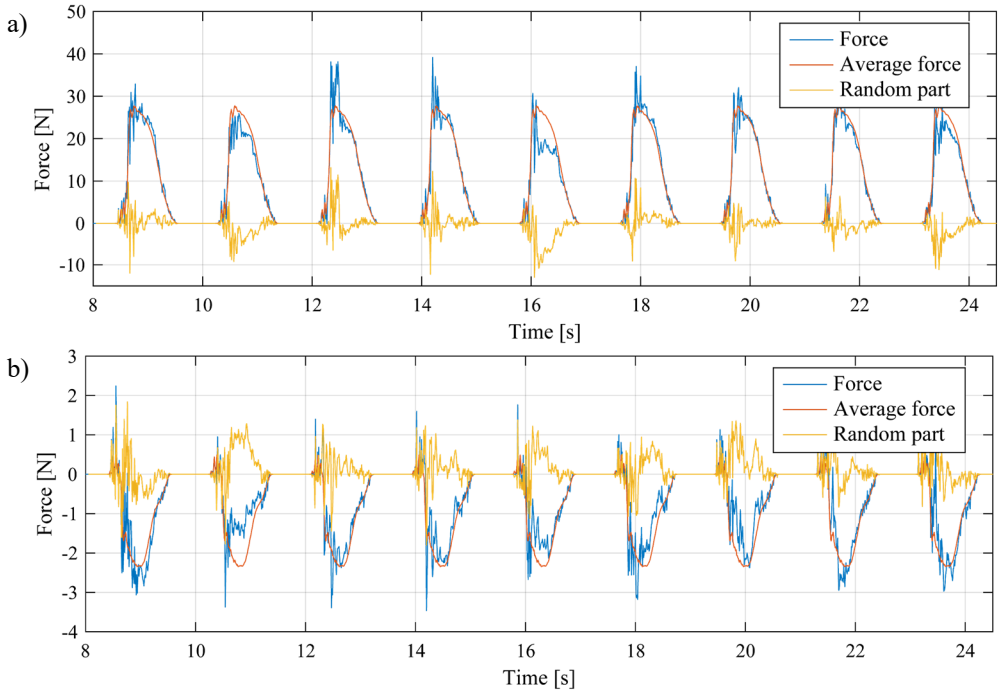


Figure 13: Average/random decomposition of the forces acting on the front face of the lifter. a) Normal force. b) Parallel force.

a function of the angular position, being the random part responsible for deviating the magnitude of the average forces until reaching the magnitude of the individual pulses.

4.3 Calculation of torques and power from the forces acting on the virtual sensors

4.3.1 Time delay of virtual sensor forces

The ring of the mill is composed of a set of liners and lifters, of identical geometry, located in different angular positions. As the VS collect the forces that act on the 3 faces of a lifter and its adjacent liner, it is possible to calculate the power associated with the movement of the ring from the forces they register, assuming that the average forces acting on a lifter i is equal to the forces acting on the reference lifter, when its located in the same angular position.

It is considered that the stationary force acting on each face of a lifter i is equal to the force of the respective face of the reference lifter, including the time delay according to the difference between their angular positions. That is:

$$F_{i,j}(t) = F_{1,j}\left(t - (i-1)T_{lft}\right) \tag{24}$$

Where:

$$T_{lft} = \frac{2\pi}{\omega N_{lft}} \tag{25}$$

N_{lft} is the number of lifters, j represents the force (i.e., face, normal or parallel), i represents the lifter for which the force is being expressed based on the reference lifter on which the stationary force $F_{1,j}$ acts, t is time and T_{lft} is the lifter pass period. Figure 14 illustrates the time delay process, where the forces acting on the rest of the lifters are determined from the forces acting on the lifter 1. Figure 14a shows the result of the phase shift process for 3 lifters. Figure 14b shows the forces acting on all lifters at a given instant, calculated from the average forces acting on the reference lifter.

4.3.2 Calculation of the driving torque and power

Of the eight forces shown in Figure 8, five of them have tangential components and, therefore, exert torques in the direction of rotation of the mill. This set of forces is determinant in the driving torque and, thus, in the

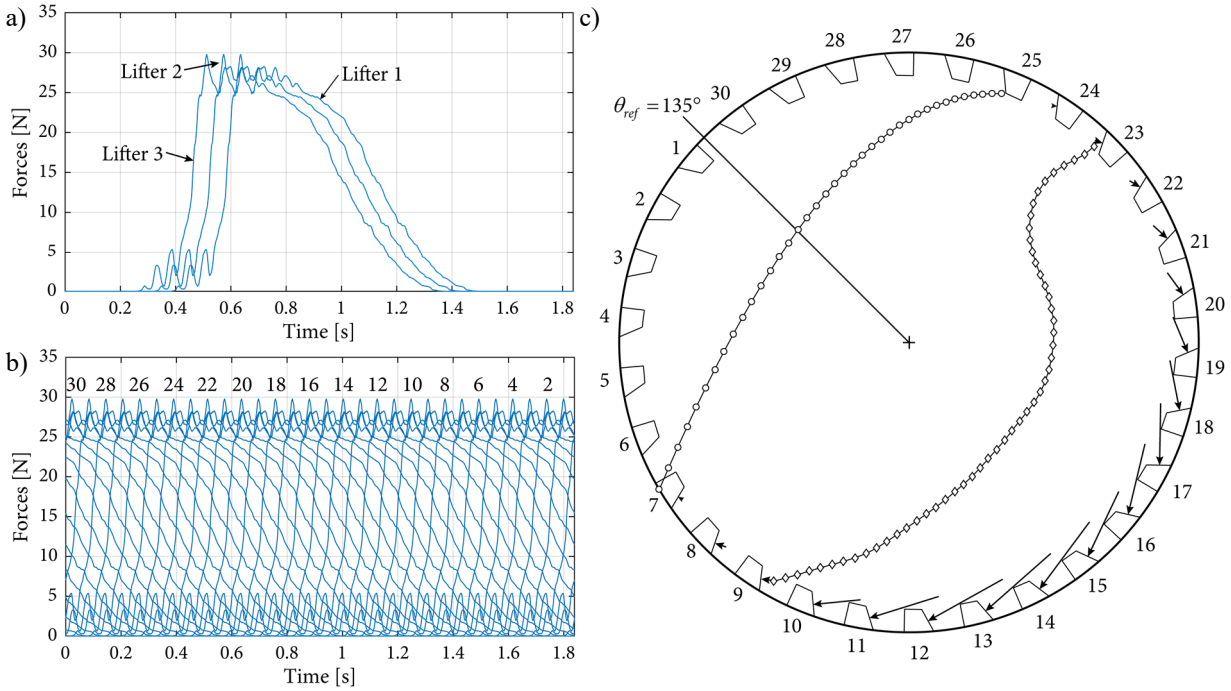


Figure 14: Illustration of the time delay process for determining the forces acting on all lifters and liners. (a) Time delay for three lifters. (b) Time delay for all lifters. (c) Global representation of the forces.

power consumption. Of these five forces, four exert resistant torques, as shown in Figure 6. The force $F_{N,BF}$, acting on the back face of the lifter, exerts torque in the rotation direction, thus favoring the mill's movement and diminishing the required driving torque. As the forces are treated separately for each face, it is possible to obtain the torque associated with the front face, top face, back face, and adjacent liner independently. These torques consider the total number of lifters, due to the time delay process explained in the previous section.

The torques exerted by the different forces acting on any face of a single lifter are given by:

$$\vec{T}_{i,j} = \vec{r}_j \times \vec{F}_{i,j} \tag{26}$$

Where $T_{i,j}$ is the torque exerted by force $F_{i,j}$ acting on the face j of the lifter i, and r_j is the arm of this force. Hence, the torques associated with the different forces acting on each of the faces of a lifter i are given by:

For normal and tangential forces acting on the front face of the lifter (FF):

$$T_{FF,i} = \left(F_{N,FF,i}\cos(\alpha) + F_{P,FF,i}\sin(\alpha)\right)\left(0.25D_M\left(1+\cos(\beta)\right) - h\right) \tag{27}$$

For the tangential force acting on the adjacent lifter (AL):

$$T_{AL,i} = 0.5F_{P,AL,i}D_M \tag{28}$$

For the normal force acting on the back face of the lifter (BF):

$$T_{BF,i} = -0.5F_{N,BF,i}\left(D_M - h\right) \tag{29}$$

For the tangential force acting on the top face of the lifter (TF):

$$T_{TF,i} = F_{P,TF,i}\left(0.5D_M - h\right) \tag{30}$$

And the total torque contribution per face is obtained by considering the respective torques of all N_{lft} lifters:

$$T_{FF} = \sum_{i=1}^{N_{lft}} T_{FF,i};\ T_{AL} = \sum_{i=1}^{N_{lft}} T_{AL,i};\ T_{BF} = \sum_{i=1}^{N_{lft}} T_{BF,i};\ T_{TF} = \sum_{i=1}^{N_{lft}} T_{TF,i} \tag{31}$$

The total torque per face of the lifter and adjacent liner includes the effect of the magnitude change of the forces and its varying angular position as the mill rotates. The total torques associated with each face, obtained from the set of forces acting on the ring, are shown in Figure 15 as a function of time. The figure also indicates the average values.

Finally, the total power due to the movement of the load inside the mill is given by the sum of the average torques associated to each face:

$$P_{Forces} = \left(\overline{T}_{FF} + \overline{T}_{AL} + \overline{T}_{BF} + \overline{T}_{TF}\right)\cdot\omega \tag{32}$$

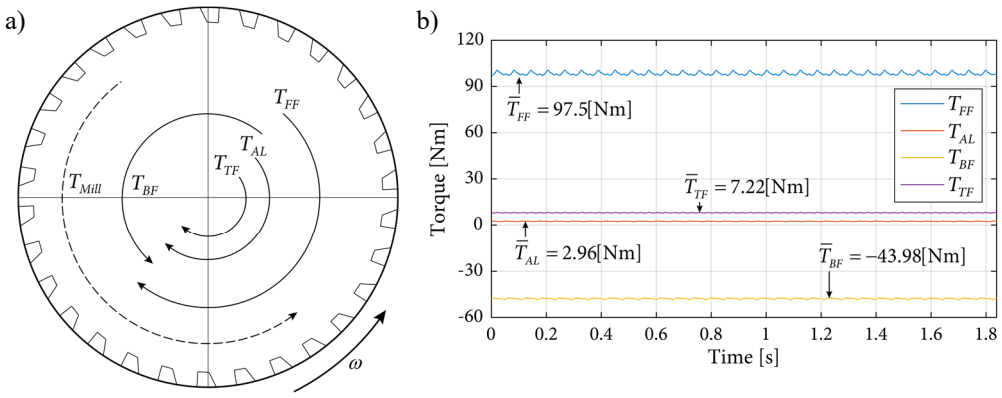


Figure 15: (a) Illustration of the components of the total torque per face of the lifter and adjacent liner. (b) Results of total torque per face as a function of time (average values indicated).

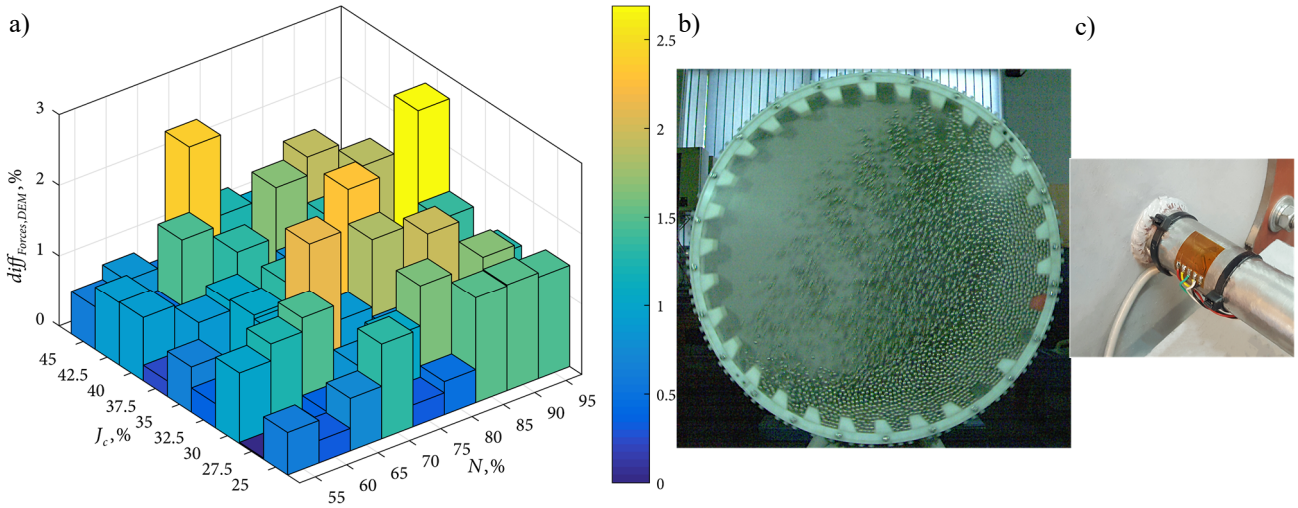


Figure 16: (a) Power differences for different operating conditions. (b) Front view of the mill while operating. (c) Strain gage for experimental measurement of driving torque.

4.3.3 Comparison between the power calculated from the forces and the power obtained from DEM

In order to validate the methodology, the power P_{Forces} is compared to the power $P_{DEM,Ring}$, which is obtained directly from the DEM model of the SetupD100.

Figure 16a shows the difference between the two powers calculated as:

$$diff_{Forces,DEM} = \frac{\left| P_{Forces} - P_{DEM,Ring} \right|}{P_{DEM,Ring}} \tag{33}$$

As observed, the differences between the power obtained directly from the DEM software and the power calculated from the forces are small (maximum 2.69%), thus validating the method to obtain the set of 8 forces as a function of the angular position using the virtual sensors.

5 Instrumentation used in SetupD100

The instrumentation used in SetupD100 focuses on analyzing the movement of the load inside the mill, analyzing the interaction between the lifters and the particles, and the experimental measurement of the power associated with the operation of the mill.

5.1 High-speed camera

One of the main features of the SetupD100 is that it is possible to observe the movement of the particles inside while operating, thanks to its transparent front cover. A high-speed camera (200 fps) is used to record the movement of the load inside the mill. Figure 16b shows an example of the images captured by the camera.

5.2 Driving torque

The measurement of the driving torque is made by using strain gages in full-bridge configuration installed on the mill shaft, between the mill and the bearing, Figure 16c.

The strain measurement in this position allows determining the torque due to the movement of the particles only, eliminating the need to estimate losses in other elements of the powertrain such as couplings and gear transmission. By multiplying the motor torque by the angular speed of the mill, the experimental power associated with the movement of the mill P_{Exp} is obtained, which is equivalent to the power $P_{DEM,Mill}$.

5.3 Instrumented lifter and tachometer

In order to validate the average forces obtained numerically, instrumented lifters resembling the virtual sensors are built. The instrumented lifters are capable of sensing the interaction between the particles and two of the lifter faces: front face (FF) and back face (BF).

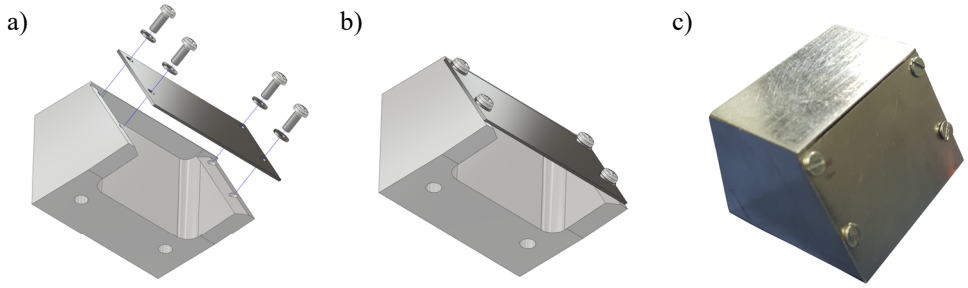


Figure 17: Instrumented lifter for measurement of front face interactions (LI FF). (a) and (b) Assembly. (c) Actual lifter.

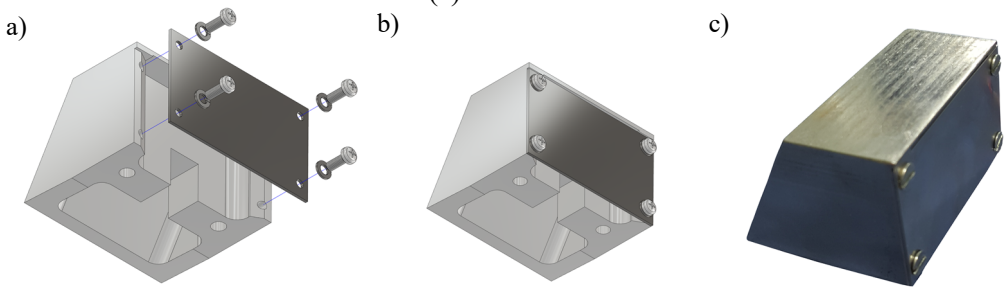


Figure 18: Instrumented lifter for measurement of back face interactions (LI FF). (a) and (b) Assembly. (c) Actual lifter.

The instrumented lifters are composed of two parts: a base block and a thin sensing plate. The sensing plate is made of steel and is fixed to the base block using 4 bolts at its side ends, as shown in Figure 17a. When the instrumented lifter contacts the particles, the sensing plates deflect. The deflection is measured by a pair of bi-axial strain gauges in half bridge configuration installed in the back of the plate. Figure 17 shows the assembly of the instrumented lifter for the measurement of front face interactions (LI-FF); whereas Figure 18 shows the same for measurement of back face interactions (LI-BF).

Figure 19 shows the results of the measurements with the two instrumented lifters for a rotation speed of 75% of the critical speed and a fill level of 30%. As in the case of the forces measured with the VS, the strain measured by the instrumented lifters are separated in pulses. The angular position of the instrumented lifters is determined from the reference signal provided by a photo-tachometer. Subsequently, the experimental pulses (EP) are averaged to obtain an estimator of the stationary component of the pulse forces, Figure 20.

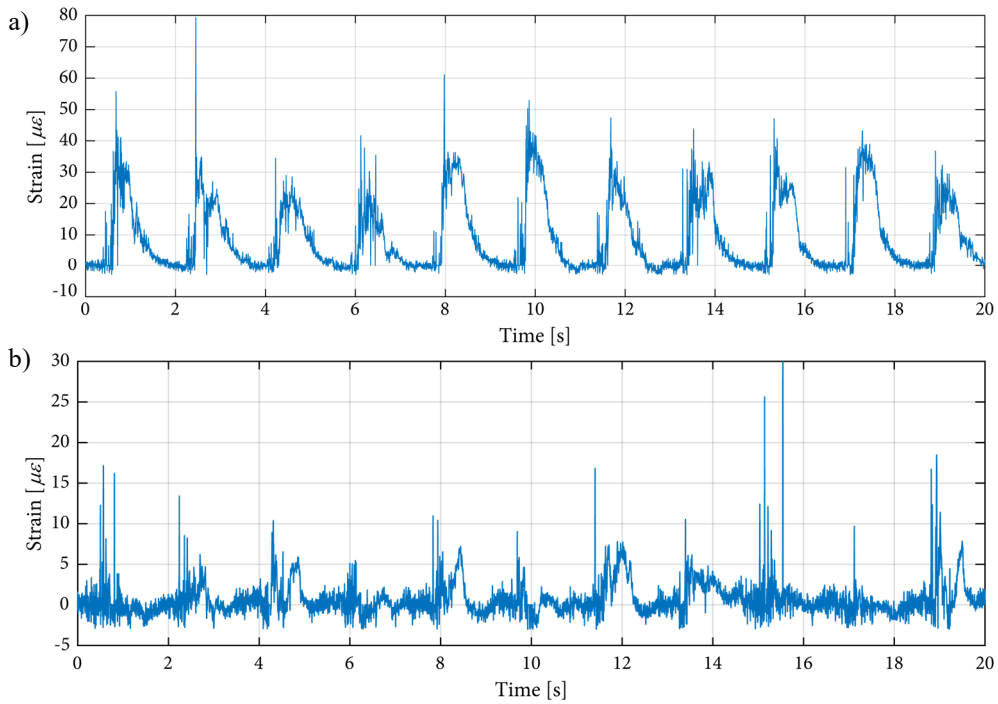


Figure 19: Strain waveform measured by the instrumented lifters. (a) Front face (LI-FF). (b) Back face (LI-BF).

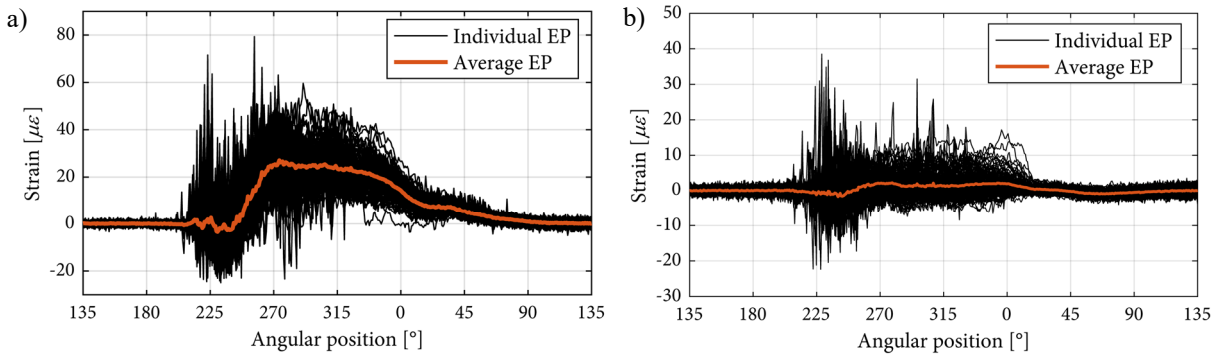


Figure 20: Experimental average pulses obtained with the instrumented lifters. (a) Front face (LI-FF). (b) Back face (LI-BF).

6 Results and discussion

The presented methodology allows a detailed analysis of the behavior of the load inside the mill, which in turn defines the overall behavior of the machine. Of particular interest is the behavior of the power consumption depending on the operating conditions. This behavior has been described in the literature based on the observation of the phenomenon, but a physical explanation has not yet been presented.

This section presents experimental results that validate the proposed methodology. Afterwards, the methodology is used to provide the physical explanation mentioned.

The average forces associated to each face as a function of the angular position, and the corresponding torques they exert, allows analyzing the behavior of the power associated with the movement of the ring from the viewpoint of the interaction of the particles with its internal surfaces.

6.1 Average forces

Figure 21 shows the average forces obtained after processing the forces from the virtual sensors included in the DEM model of the SetupD100, for the mill operating at 75% of its critical speed and a fill level of 30%. As can be observed, the magnitude of the parallel forces is much smaller than the magnitude of the normal forces. Five of the forces acting on the lifter faces are tangential to the ring. From these, the normal forces acting on the front face (FF) and back face (BF) are the most significant in the total torque associated with the rotation of the ring. The 3 remaining forces are parallel forces of magnitude considerably lower than that of normal forces. Based on this, the torque components T_{FF} and T_{BF} are the most relevant in the total torque and thus in the power consumption.

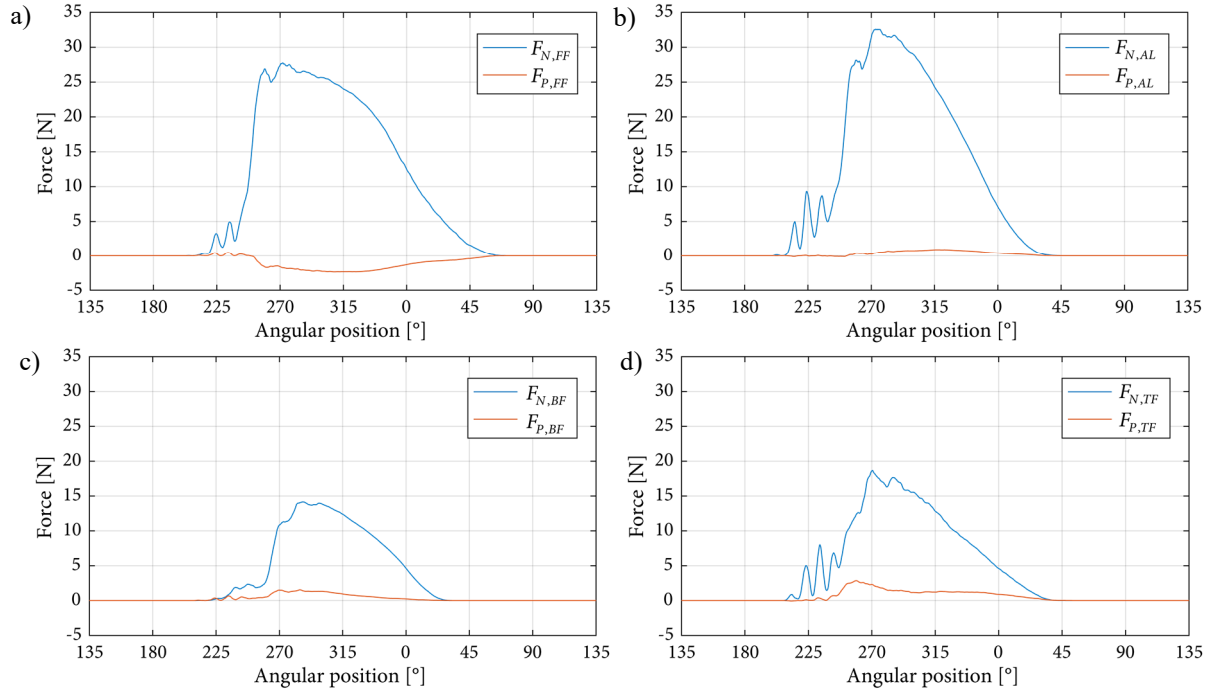


Figure 21: Normal and parallel average forces from the virtual sensors. (a) Front face. (b) Adjacent lifter. (c) Back face. (d) Top face.

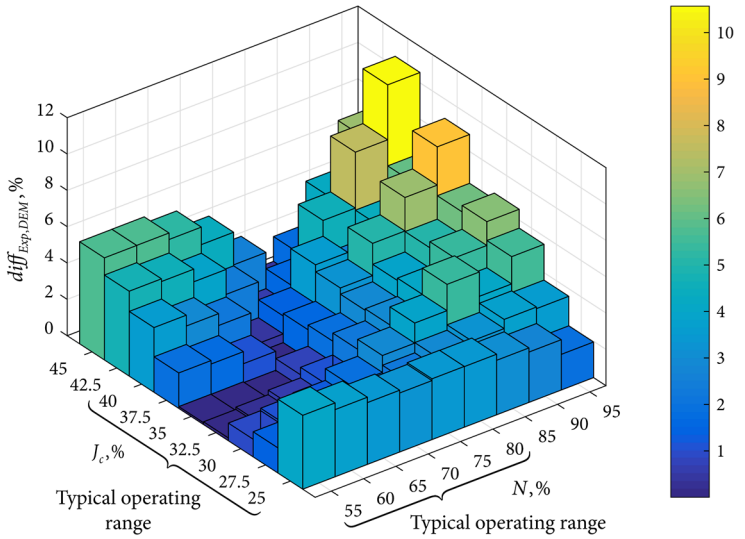


Figure 22: Difference between P_{Exp} and $P_{DEM,Mill}$ for each of the conditions analyzed.

6.2 Experimental validation

The numerical model of the SetupD100 is validated by comparing the numerical results obtained by processing the data extracted from the DEM simulations against their experimental equivalents obtained using the instrumentation of the test bench. The comparison is made first in terms of the power associated with the movement of the mill obtained by DEM ($P_{DEM,Mill}$) and the power obtained experimentally by the strain gauge installed on the drive shaft of the mill (P_{Exp}). As a second approach, the average forces obtained from the virtual sensors and the average experimental pulses obtained with the instrumented lifters are considered. Finally, the disposition of the particles inside the mill obtained by DEM and by the high-speed camera are contrasted.

Figure 22 shows the difference between P_{Exp} and $P_{DEM,Mill}$, calculated as:

$$diff_{Exp,DEM} = \frac{\left| P_{DEm,Mill} - P_{Exp} \right|}{P_{Exp}} \tag{34}$$

The maximum difference is 10.5% for the mill operating at 95% of its critical speed and 42.5% fill level. In mining industries, grinding mills typically operate with speeds ranging between 55% to 80% of the critical speed and fill levels between 25% to 40% of the internal volume. Within this range of operating conditions, the difference between the numerical and experimental power is less than 5%.

Figure 23 shows the comparison between the forces $F_{N,FF}$ and $F_{N,BF}$, and the experimental average pulses obtained with the LI-FF and the LI-BF, for different operating conditions. The differences observed are due to the fact that the instrumented lifters do not measure the contact forces directly, but the strain caused by these forces. Even though the strain is a consequence of these forces, there is no direct relation to the resultant force magnitude, because the force distribution has an influence too. It can also be seen that the magnitude of the measurements made with the LI-BF is much lower than those obtained with the LI-FF. Despite this, there is a clear agreement between the angular intervals in which the forces and the average experimental pulses are non-zero. Also, the fact that the average experimental pulse obtained with LI-BF is mostly positive confirms that the forces exerted by the particles on the BF compress it, generating torques in the sense of rotation of the mill.

Figure 24 compares the disposition of particles inside the mill obtained numerically and experimentally for 3 different operating conditions, showing good correlation.

Based on the similarities observed between the numerical and experimental results, it is reasonable to assume that the numerical model of the SetupD100 and the proposed methodology are valid.

6.3 Effect of rotating speed on the power, torque, and forces

Figure 25a shows the behavior of the power $P_{DEM,Ring}$ as a function of the speed, for the mill operating with different fill levels. Figure 25b shows the same for the torque $T_{DEM,Ring} = P_{DEM,Ring} / \omega$. This torque is equivalent to the sum of the torques $\overline{T}_{FF}$, $\overline{T}_{AL}$, $\overline{T}_{BF}$ and $\overline{T}_{TF}$. It can be seen that the power increases steadily, for all fill levels,

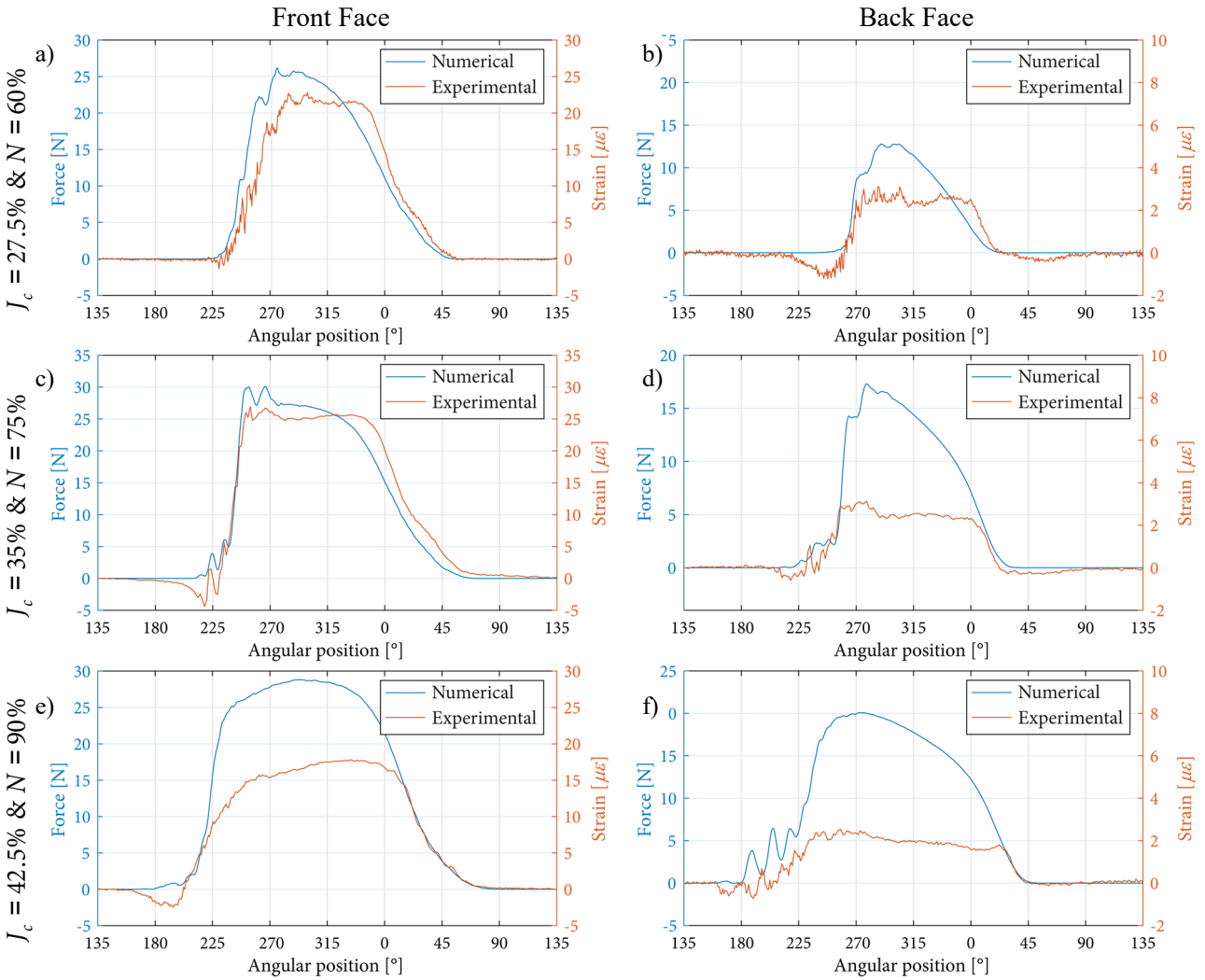


Figure 23: Comparison of numerical and experimental average pulses for the front face and back face, and different operating conditions.

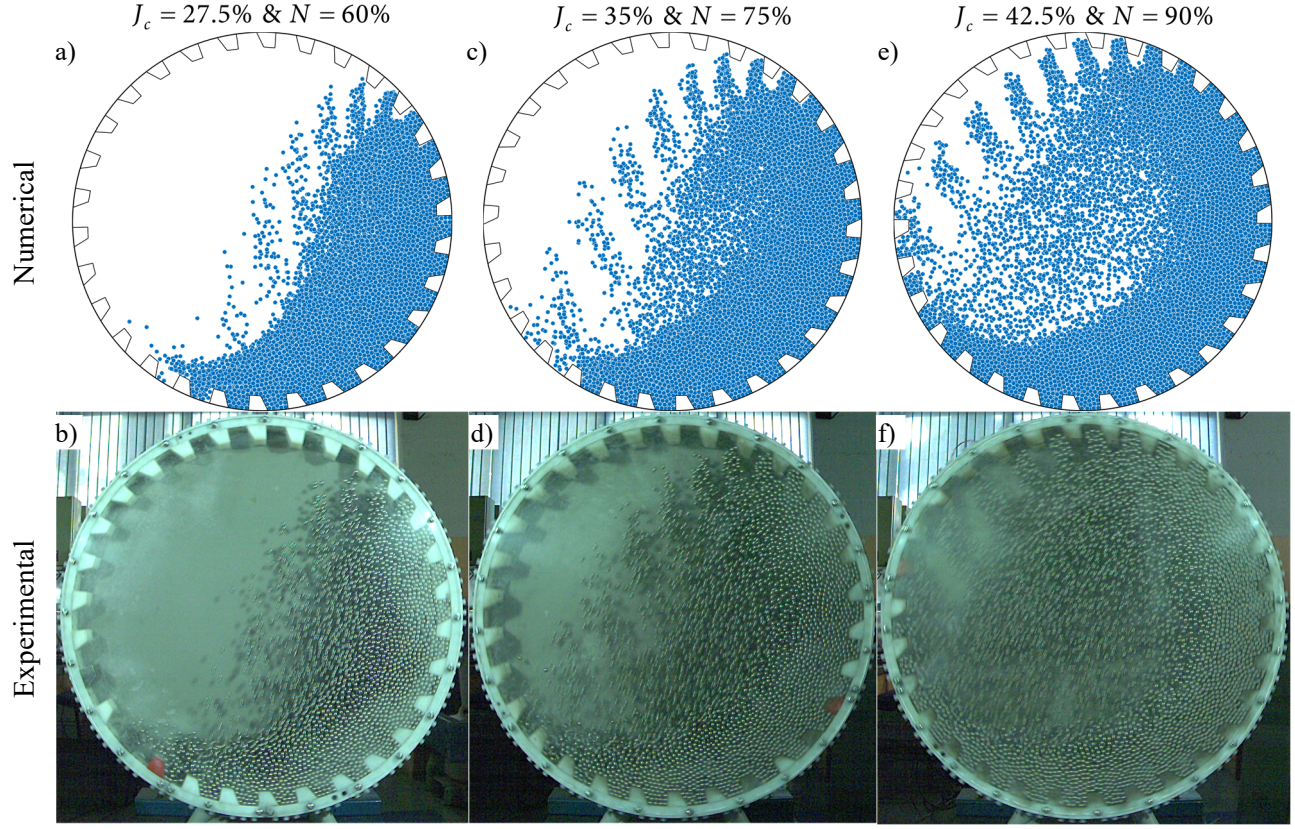


Figure 24: Disposition of particles inside the mill (a),(c),(e) Numerical results. (b),(d),(f) Expermiental results.

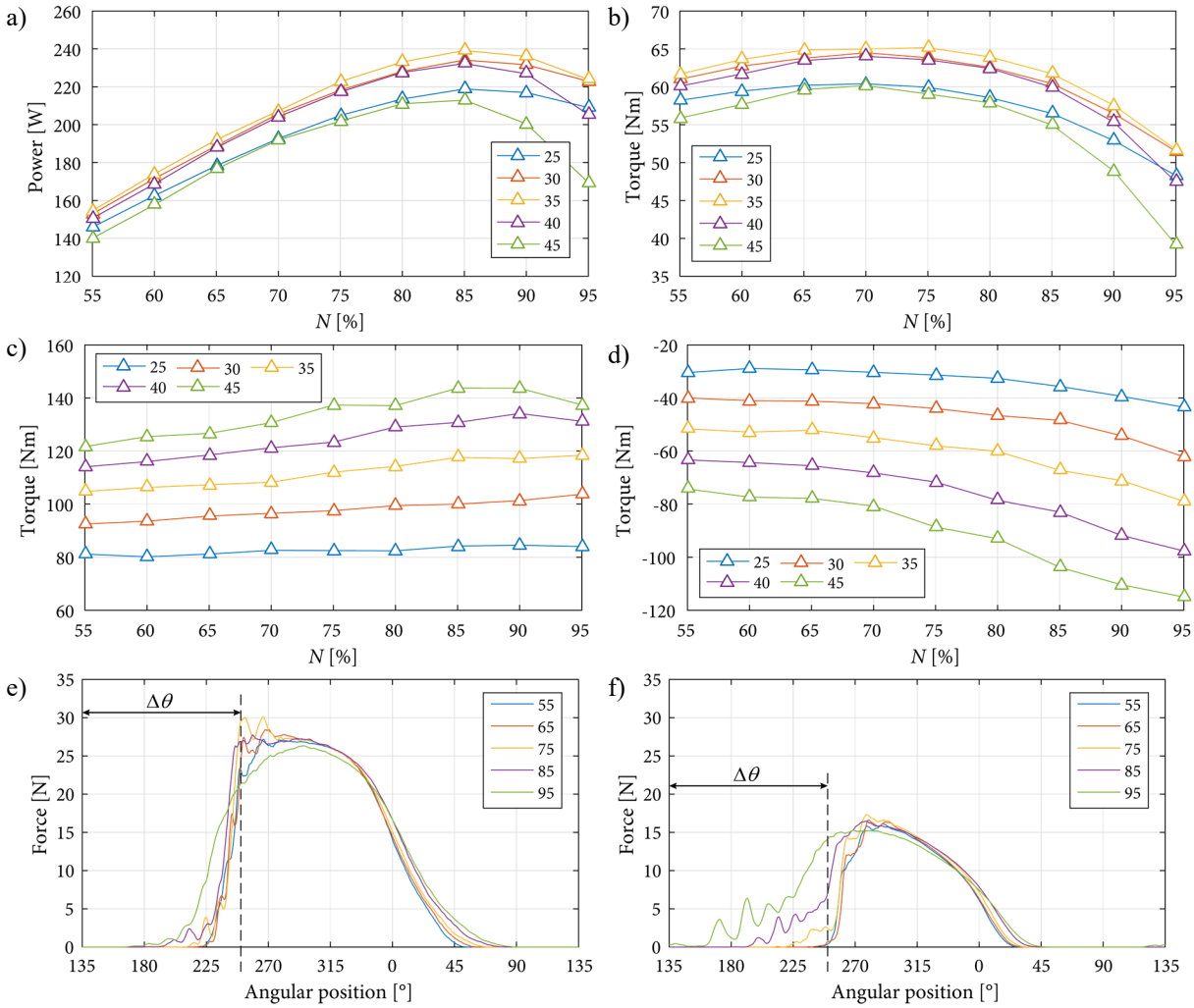


Figure 25: (a) Power $P_{DEM,Ring}$ as a function of speed. (b) Torque $T_{DEM,Ring}$ as a function of speed. (c) Torque T_{FF} as a function of speed. (d) Torque T_{BF} as a function of speed. (e) Force $F_{N,FF}$ as a function of the angular position. (f) Force $F_{N,BF}$ as a function of the angular position.

from 55% to 80% of the critical speed. For higher speeds, the power decreases at a rate that becomes more pronounced as the fill level increases. The torque $T_{DEM,Ring}$ increases up to 70% of the critical speed. From this point on, it begins to decrease. This behavior has been reported in the literature [13][14][21][57], and in order to explain it, the torques and the average forces associated with the different faces of the lifters are further analyzed.

Figure 25c and Figure 25d show the behavior of the torques T_{FF} and T_{BF} as a function of the speed. It can be seen that in the range between 55% and 70% of the critical speed, the increase rate in the magnitude of T_{FF} is slightly higher than the increase rate in the magnitude of T_{BF}, which explains the increase in the magnitude of $T_{DEM,Ring}$ within this speed range. For speeds above 70% of the critical speed, it can be noted that the magnitude of T_{BF} increases at a higher rate than the magnitude of T_{FF}, and that this is more pronounced for higher fill levels. This indicates that the drop in the torque $T_{DEM,Ring}$ for speeds higher than 70% of the critical speed observed in Figure 25b originates because the magnitude of the force $F_{N,BF}$ increases more significantly than the magnitude of the force $F_{N,FF}$ as the rotating speed increases.

Figure 25e and Figure 25f show, respectively, the magnitude of the forces $F_{N,FF}$ and $F_{N,BF}$ as a function of the angular position for the mill operating at 35% of fill level and speed from 55% to 95% of the critical speed with 10% jumps. It can be noted that the magnitude of $F_{N,FF}$ and the angular interval where this force is non-zero increase slightly as a function of the speed of rotation, which produces the slight and sustained increase in the magnitude of T_{FF} observed in Figure 25c. On the other hand, the force $F_{N,BF}$ remains approximately constant between 55% and 65% of the critical speed but begins to increase in the angular range $\Delta\theta$ between 135° and 250° for fill levels above 75%. This behavior originates the increase suffered by the magnitude of

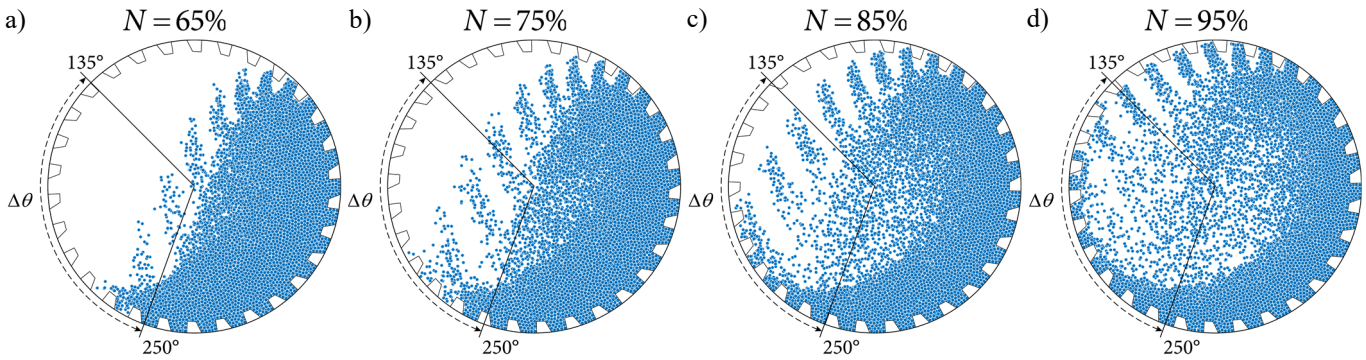


Figure 26: Disposition of particles inside the mill operating with 35% fill level and rotating speed of (a) 65%, (b) 75%, (c) 85% and (d) 95% of the critical speed.

T_{BF} between 70% and 95% of the critical speed, and that causes the decrease of $T_{DEM,Ring}$ in this speed range. The physical phenomenon that gives rise to the forces exerted by the particles in the angular interval between 135° and 250°, for speeds higher than 70% of the critical speed—and that is accentuated with speed--, corresponds to the impact of the particles describing cataract movements and that fall over the toe position, as shown in Figure 26. The particles impacting the internal surface of the mill over the position of the toe exert forces on the back face of the lifters and not on the front face, due to their orientation relative to the path of the particles. These forces exert torques in the sense of rotation of the mill, thus diminishing the total torque requirement. This phenomenon becomes more significant as the rotation speed increases, because both the number and speed of the particles falling on the back face of the lifters above the toe position, increase.

6.4 Conclusions

A methodology is proposed to obtain the magnitude of the average forces exerted by the particles inside the mill on the different faces of the lifters and liners of the ring as a function of their angular position, and the torques associated with each of these faces for the complete set of lifters in the ring. The results obtained from the numerical modeling are in agreement with the corresponding experimental measurements obtained in the test bench, thus validating the methodology.

The forces and torques provide a useful tool to understand the behavior of the power consumed by the mill depending on its geometry and operating conditions. These variables provide the link between the behavior of the load inside the drum and the resulting behavior of the power consumption of the mill. This is especially relevant considering that the energy costs associated with the operation of tumbling mills and the evolution of their internal geometry due to wear, represent a significant fraction of the costs associated with the refinement of minerals.

The results of the analyses carried out in this work show that the decrease in power consumption observed when the speed of the mill increases is due to the impacts of the particles falling on the zone of the ring above the toe of the load. These impacts exert forces on the back face of the lifters, thus increasing the torque component associated with this face, which acts in the sense of rotation. At the same time, as the speed increases, the torque associated with the front face of the lifters also increase, but its magnitude grows at a lower rate. This difference in the growth rates of both torques provides the physical explanation of the power consumption drop at higher speeds.

The methodology seems to be a promising tool in view of gaining further insight about the behavior of tumbling mills and could be useful, for example, to develop strategies to reduce O&M costs and increase efficiency.

References

[1] 'The world's largest sag mill', Outotec, South East Asia Pacific Output, Dec. 2012, page 6.

[2] Guerrero, V. & Pontt, J. 2011. 'Oscillatory Torque Caused by Dead Time in the Current Control of High Power Gearless Mills', *IECON 2011* - 37th Annual Conference on IEEE Industrial Electronics Society, 2011, pp. 1966 – 1970.

[3] Martins, S., Li, W., Radziszewski, P., Faucher, A. & Makni, S. 2013. 'Experimental and simulated instrumented ball in a tumbling mill - A comparison'. *Minerals Engineering*, vol. 43-44, pp.79-84.

[4] Hosseini, P., Martinsm S, Martin, T. Radziszewski, P. & Boyer, F. 2011. 'Acoustic emissions simulation of tumbling mills using charge dynamics'. *Minerals Engineering*, vol. 24, pp. 1440-1447.

[5] Cleary, P., Morrison, R. & Morrell, S. 2003. 'Comparison of DEM and experiment for a scale model SAG mil'. *International Journal of Mineral Processing*, vol. 68, pp. 129-165.

[6] Cleary, P. & Morrison, R. 2016. 'Comminution mechanisms, particle shape evolution and collision energy partitioning in tumbling mills'. *Minerals Engineering*, vol. 86, pp. 75-95.

[7] Cleary, P. 2015. 'Prediction of coupled particle and fluid flows using DEM and SPH'. *Minerals Engineering*, vo. 73, pp. 85-99.

[8] Owen, P. & Clearly, P. 2015. 'The relationship between charge shape characteristics and fill level and lifter height for a SAG mill'. *Minerals Engineering*, vol. 83, pp. 19-32.

[9] Cleary, P. & Morrison, R. 2011. 'Understanding fine ore breakage in a laboratory scale ball mill using DEM'. *Minerals Engineering*, vol. 24, pp. 352-366.

[10] Kalala, J., Breetzke, M. & Moys, H. 2008. 'Study of the influence of liner wear on the load behaviour of an industrial dry tumbling mill using the Discrete Element Method (DEM)'. *International Journal of Mineral Processing*, vol. 86, pp. 33-39.

[11] Powell, M. & McBride, A. 2004. 'A three-dimensional analysis of media motion and grinding regions in mills'. *Minerals Engineering*, vol. 17, pp. 1099-1109.

[12] Mishra, B. 2003. 'A review of computer simulation of tumbling mills by the discrete element method: Part I-contact mechanics'. *International Journal of Mineral Processing*, vol. 71, pp. 73-93.

[13] Cleary, P. 2001. 'Charge behaviour and power consumption in ball mills: Sensitivity to mill operating conditions, liner geometry and charge composition'. *International Journal of Mineral Processing*, vol. 63, pp. 79-114.

[14] Cleary, P. 1998. 'Predicting charge motion, power draw, segregation and wear in ball mills using discrete element methods'. *Minerals Engineering*, vol. 11, pp. 1061-1080.

[15] Mishra, B. & Rajamani, R. 1992. 'The discrete element method for the simulation of ball mills'. *Applied Mathematical Modelling*, vol. 16, pp. 598-604.

[16] Kalala, J & Moys, M. 2004. 'Discrete element method modelling of liner wear in dry ball milling'. *Journal of The South African Institute of Mining and Metallurgy*, vol. 104, pp. 597-602.

[17] Morrison, R. & Cleary, P. 2004. 'Using DEM to model ore breakage within a pilot scale SAG mil'. *Minerals Engineering*. Vol. 17, pp. 1117-1124.

[18] Jiménez-Herrera, N., Barrios, G. & Tavares, L. 2017. 'Comparison of breakage models in DEM in simulating impact on particle beds'. *Advanced Powder Technology*, vol. 29, pp. 692-706.

[19] Cleary, P., Sinnott, M. & Pereira, G. 2015. 'Computational prediction of performance for a full scale Isamill: Part 1 - Media motion and energy utilisation in a dry mill'. *Minerals Engineering*, vol. 79, pp. 220-238.

[20] Tuzcu, E. & Rajamani, R. 2011. 'Modeling breakage rates in mills with impact energy spectra and ultra fast load cell data'. *Minerals Engineering*, vol. 24, pp. 252-260.

[21] Cleary, P. 2001. 'Recent advances in DEM modelling of tumbling mills'. *Minerals Engineering*, vol. 14, pp. 1295-1319.

[22] Cleary, P. 2001b. 'Modelling comminution devices using DEM'. *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 25, pp. 83–105.

[23] Cleary, P.W., Sinnott, M. & Morrison, R., 2008. 'DEM prediction of particle flows in grinding processes'. *International Journal for Numerical Methods in Fluids*, vol. 58, pp. 319–353.

[24] Cleary, P. 2015. 'A multiscale method for including fine particle effects in DEM models of grinding mills'. *Minerals Engineering,* vol. 84, pp. 88–99.

[25] Djordjevic, N. 2003. 'Discrete element modelling of the influence of lifters on power draw of tumbling mills'. *Minerals Engineering,* vol. 16, pp. 331–336.

[26] Djordjevic, N., 2005. 'Influence of charge size distribution on net-power draw of tumbling mill based on DEM modelling'. *Minerals Engineering*, vol. 18, pp. 375–378.

[27] Geng, F., Gang, L., Wang, Y., Li, Y. & Yuan, Z., 2015. 'Numerical investigation on particle mixing in a ball mill'. *Powder Technology*, vol. 292, pp. 64–73.

[28] Govender, I., McBride, A. & Powell, M. 2004. 'Improved experimental tracking techniques for validating discrete element method simulations of tumbling mills'. *Experimental Mechanics*, vol. 44, pp. 593–607.

[29] Govender, I., Cleary, P. & Mainza, A., 2013. 'Comparisons of PEPT derived charge features in wet milling environments with a friction-adjusted DEM model'. *Chemical Engineering Science,* vol. 97, pp. 162–175.

[30] Hlungwani, O., Rikhotso, J., Dong, H., Moys, M.H., 2003. 'Further validation of DEM modeling of milling: effects of liner profile and mill speed'. *Minerals Engineering,* vol. 16, pp. 993– 998.

[31] Kalala, J., Bwalya, M & Moys, M. 2005. 'Discrete element method (DEM) modelling of evolving mill liner profiles due to wear Part II. Industrial case study'. *Minerals Engineering,* vol. 18, pp. 1386–1391.

[32] Makokha, A & Moys, M. 2006. 'Towards optimising ball-milling capacity: effect of lifter design'. *Minerals Engineering* ,vol. 19, pp. 1439–1445.

[33] Mayank, K., Malahe, M., Govender, I. & Mangadoddy, N. 2015. 'Coupled dem-cfd model to predict the tumbling mill dynamics'. *Procedia IUTAM*, vol. 15, pp. 139–149.

[34] Mishra, B. & Rajamani, R. 1994a. 'Simulation of charge motion in ball mills. Part 1: experimental verifications'. *International Journal of Mineral Processing,* vol. 40, pp. 171–186.

[35] Mishra, B. & Rajamani, R. 1994b. 'Simulation of charge motion in ball mills. Part 2: numerical simulations'. *International Journal of Mineral Processing*, vol. 40, 187–197.

[36] Morrison, R. & Cleary, P. 2008. 'Towards a virtual comminution machine'. *Minerals Engineering*, vol. 21, pp. 770–781.

[37] Mulenga, F. & Moys, M. 2014. 'Effects of slurry filling and mill speed on the net power draw of a tumbling ball mill'. *Minerals Engineering*, vol. 56, pp. 45–56.

[38] Powell, M. & McBride, A. 2006. 'What is required from DEM simulation to model breakage in mills?' *Minerals Engineering*, vol. 19, pp. 1013–1021.

[39] Powell, M., Weerasekara, N., Cole, S., LaRoche, R. & Favier, J. 2011. 'DEM modelling of liner evolution and its influence on grinding rate in ball mills'. *Minerals Engineering,* vol. 24, pp. 341–351.

[40] Rezaeizadeh, M., Fooladi, M., Powell, M., Mansouri, S. & Weerasekara, N. 2010. 'A new predictive model of lifter bar wear in mills'. *Minerals Engineering*, vol. 23, pp. 1174–1181.

[41] Toor, P., Franke, J., Powell, M., Bird, M. & Waters, T. 2013. 'Designing liners for performance not life'. *Minerals Engineering*, vol. 43, pp. 22–28.

[42] Wang, M., Yang, R. & Yu, A. 2012. 'DEM investigation of energy distribution and particle breakage in tumbling ball mills'. *Powder Technology*, vol. 223, pp. 83–91.

[43] Weerasekara, N., Liu, L. & Powell, M. 2016. 'Estimating energy in grinding using dem modelling'. *Minerals Engineering*, vol. 85, pp. 23-33.

[44] Xu, L., Luo, K. & Zhao, Y. 2018. 'Numerical prediction of wear in SAG mills based on DEM simulations', Powder Technology, vol. 329, pp. 353-363.

[45] Qui, Xx, Potapov, A., Song, M., & Nordell, L. 2001. 'Prediction of Wear of Mill Lifters Using Discrete Element Method', In SAG 2001, University of British Columbia, Vancouver, B.B, Canada, pp. 2–7.

[46] Boemer, D & Ponthot, J. 2017. 'DEM modeling of ball mills with experimental validation: influence of contact parameters on charge motion and power draw', *Computational Particle Mechanics*, vol. 4, pp. 53-67.

[47] Wills, B. & Napier-Munn, T. 2006. 'Wills' Mineral Processing Technology', 7th edn, Elsevier Science & Technology Books.

[48] Tano, K., Pålsson, B. & Sellgren, A. 2005. 'On-line lifter deflection measurements showing flow resistance effects in grinding mills', *Minerals Engineering*, vol. 18, pp. 1077–1085.

[49] Owen, P. & Cleary, P. W. 2015. 'The relationship between charge shape characteristics and fill level and lifter height for a SAG mill', *Minerals Engineering*. vol. 83, pp. 19–32.

[50] Cleary, P. W. & Owen, P. 2018. 'Development of models relating charge shape and power draw to SAG mill operating parameters and their use in devising mill operating strategies to account for liner wear', *Minerals Engineering*. vol. 117, pp. 42–62.

[51] Kalala, J., Bwalya, M. and Moys, M. 2005. 'Discrete element method (DEM) modelling of evolving mill liner profiles due to wear. Part I: DEM validation', *Minerals Engineering*, vol. 18, pp. 1386–1391.

[52] Kalala, J., Bwalya, M. & Moys, M. 2005. 'Discrete element method (DEM) modelling of evolving mill liner profiles due to wear. Part I: DEM validation', *Minerals Engineering*, vol. 18, pp. 1386–1391.

[53] Jonsén, P., Pålsson, B. & Häggblad, H. 2012. 'A novel method for full-body modelling of grinding charges in tumbling mills', *Minerals Engineering*, vol. 33, pp. 2–12.

[54] Moys, M., van Nierop, M., van Tondero, J. & Glover, G. 2000. 'Validation of the discrete element method (DEM) by comparing predicted load behaviour of a grinding mill with measured data', *Developments in Mineral Processing,* vol. 13, pp. C3-39-C3-44.

[55] Tano, K. 2005. 'Continuous Monitoring of Mineral Processes with Special Focus on Tumbling Mills – A Multivariate Approach', Doctoral Thesis, Luleå University of Technology.

[56] Tano, K., Pålsson, B. & Sellgren, A. 2005. 'On-line lifter deflection measurements showing flow resistance effects in grinding mills', *Minerals Engineering*, vol. 18, pp. 1077–1085.

[57] Bian, X., Wang, G., Wang, H., Wang, S. & Lv, W. 2017. 'Effect of lifters and mill speed on particle behaviour, torque, and power consumption of a tumbling ball mill: Experimental study and DEM simulation'. *Minerals Engineering*, vol. 105, pp. 22-35.

Signal processing

Comparison and Improvement

of Techniques for Transmission-Path Restoring

Authors: Omri Matania, Renata Klein, Jacob Bortman

Keywords: vibration signal, transmission-path, pre-whitening, adaptive clutter separation (ACS), Cepstrum-liftering, AR model, restoration process.

Abstract:

Condition based monitoring by vibration sensors is a widely spread technique for monitoring the status and condition ("health") of rotating machines. In most cases, the monitoring is based on the ability to isolate specific elements of the vibration signals, generated by the different rotating components.

The generated signals are propagating through various transmissions paths of the machine, that distort the original signals, hence affect the assessment of the machine's condition. While these effects are usually ignored by most vibration analysis techniques, first steps towards mitigating this problem have been taken place during the last years. These techniques used pre-whitening methods, which usually served to separate the signal from its background, as well as to reduce the transmission path effects.

In this study, we aim to go a step further in transmission path restoring through deepening our understanding of their effects on the vibration signals. We start by reviewing three main pre-whitening methods: liftering low quefrencies at the Cepstrum (Cepstrum-liftering), adaptive clutter separation (ACS), and pre-whitening using auto-regressive (AR) models.

We first show that signal's pre-whitening by the AR model has large errors and therefore is less adequate for transmission path restoration processes. Through several simulations, we show that the AR model succeeds in extracting the background spectrum only where the noise is significantly larger than any other component in the PSD. For all other cases, there are large inaccuracies at the restoring background process. We propose a theoretical explanation for this phenomenon, and then strengthen our argument that AR in its current use is less adequate for this purpose.

We then propose a theoretical approach to adjust parameters for ACS and Cepstrum-liftering techniques and examine them and their sensitivity through quantitative methods. By tying the theoretical adequate windows parameters of Cepstrum-liftering and ACS, we show that it is possible to predict the sizes of these two adequate windows. Furthermore, we suggest an adaptive algorithm that is based on the theoretical calculation to reach more accurate values for the parameters of the window in real cases.

After adjustment of the parameters, we compare ACS and Cepstrum-liftering through a variety of simulations and receive slightly different quantities results of the transmission path estimation using ACS and Cepstrum liftering than the former paper [1].

A new technique to restore the transmission path background and its phase is proposed (Figure 1). The restoration of the phase is an important feature due to its deep implications on the ability to correctly restore the original signal in the time domain. The new technique is based on AR model and artificial noise colorization to restore both the background spectrum and its phase. Furthermore, it also improves the magnitude of the restored background, compared to ACS and Cepstrum-liftering techniques.

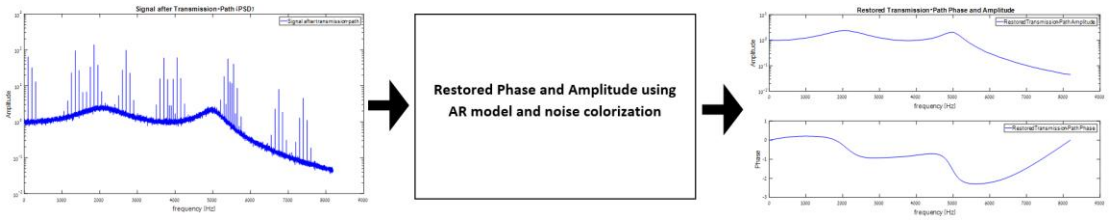


Figure 1 – restored Transmission-Path phase and amplitude

We also suggest exploiting the advantage of the new technique to restore the background in the frequency domain by converting the signal to the order domain and restore it there. The technique is highly beneficial in cases where the signal is smeared in the frequency domain but is sharp in the order domain, cases which the system rotating speed varies during the time.

References:

[1] R. Klein, *Comparison of methods for separating excitation sources in rotating machinery*, The International Conference SURVISHNO 7, 2013.

[2] R.B Randall, B. Peeters, J. Antoni, S. Manzato, *New cepstral methods of signal pre-processing for operational modal analysis*, ISMA2012-USD2012, Leuven, Belgium, pp. 775-764, 2012.

[3] R.B. Randall, W.A Smith, *Cepstrum-based operational modal analysis revisited: A discussion on pole-zero models and the regeneration of frequency response functions*, Article in Mechanical Systems and Signal Processing (MSSP), 2016.

[4] R.B. Randall, W.A Smith, *New cepstral techniques for operational modal analysis*, Conference Paper, 2017.

[5] N. Sawalhi and R.B. Randall, *Spectral Kurtosis Enhancement using Autoregressive Models*, ACAM2005 conference, Melbourne, 2005.

[6] R. Castiglione, J. Antoni, L. Garibaldi, *Separation and identification of structural modes in largely underdetermined scenarios using frequency banding*, Journal of Sound and Vibration 414 pp. 192-217, 2018.

Influence of Gaussian Signal Distribution Error on Random Vibration Fatigue Calculation.

Yuzhu WANG[1], Roger SERRA[1] et Pierre ARGOUL[2]
[1]INSA Centre Val de Loire, Laboratoire de Mécanique G. Lamé, Campus de Blois, Equipe DivS, 3 rue de la chocalaterie, 41000 Blois, France
[2]IFSTTAR, Laboratoire Mast – EMGCU, 14-20 Boulevard Newton, Cité Descartes, 77447 Marne-la-Vallée Cedex 2

Abstract

In the study of random vibration problems, Gaussian vibration and non-Gaussian vibrations are usually classified according to the excitation signal. The skewness and kurtosis are usually used to distinguish. Here we discuss a non-strict Gaussian signal, which is the error that exists in skewness and kurtosis and usually unavoidable in actual experiments or signals analysis. Through experiments and simulation calculations, the influence of this error on the traditional fatigue calculation method is discussed. The PSD approach will be discussed primarily, and time domain signals based on the rain-flow counting method will be recorded and verified. Total nine calculation model studied in this process. Finally, through a threshold, the range of skewness and kurtosis is indicated, that within this range, Gaussian signal-based calculations can be continued. By comparing the performance of different methods, a better method for signal adaptability can be obtained.

Keywords: Random vibration fatigue, Damage cumulative calculation and Gaussian random vibration.

1. Introduction

For many mechanical components, the working load is in the form of random vibrations. In the fatigue design project, the load cycle of the structure is usually obtained by the rain flow counting method according to the conventional time domain signal, with the material property, the damage of the structure could be obtained by using the Miner‘s Law, and then prediction the life of the structure. But the acquisition of time domain signals relies on a large number of experimental records, which obviously sounds expensive. Later, according to the stochastic theoretical method, the power spectral density was used to characterize the random vibration characteristics, and the method of inferring the rain flow counting result of response stress was proposed.[1][2] This theory generally assumes that the load is subject to a Gaussian distribution. The more successful method is the narrow-band approximation method proposed by Bendat in 1968.[3] Later, due to the efforts of more scholars, the broadband approximation method was

also proposed. At the same time, various improvement schemes were proposed to improve the accuracy of the approximate results. [4][5][6]

However, since most of the actual loads are non-Gaussian distributions, it is obvious. Therefore, when the load is a non-Gaussian signal, the original Gaussian-based frequency domain damage analysis method may be used, which may cause poor deviation. Therefore, it is necessary to further discuss the influence of signal non-Gaussian on stress distribution. The research method based on the frequency domain signal discusses the overall distribution of signals in the frequency domain, and the non-Gaussian signal that satisfies this condition is not unique. This leads to the use of frequency domain method to study the distribution of non-Gaussian signals. Large deviations, when calculating damage using this distribution result, pose a significant risk to product design and life estimation.

In this paper, a reference based on kurtosis judgment will be proposed to select the method of fatigue damage calculation. The rain-flow count analysis is performed on the response stress of non-Gaussian load, and the difference between the Gaussian signal and the non-Gaussian signal at the same level is obtained, and an allowable value is obtained, that is, in this range, even if it is not a Gaussian signal, the PSD method also could be used. The fatigue damage results obtained by the method are still within error tolerance. Beyond this range, the non-Gaussian signals must be considered with the special method.

2. Non-Gaussian signal and Kurtosis control

Generally, a signal whose probability density distribution obeys a Gaussian distribution is called a Gaussian signal and is mainly judged by the skewness and kurtosis of the signal. This indicator indicates the distribution of data within the data range. The skewness refers to the zero offsets of the centre, which is represented by S. The kurtosis can be understood as the specific gravity in the central region, denoted by K. The greater the kurtosis, the greater the accumulation of data in the centre. Usually, the Gaussian distribution has a kurtosis of 3 and skewness of 0.

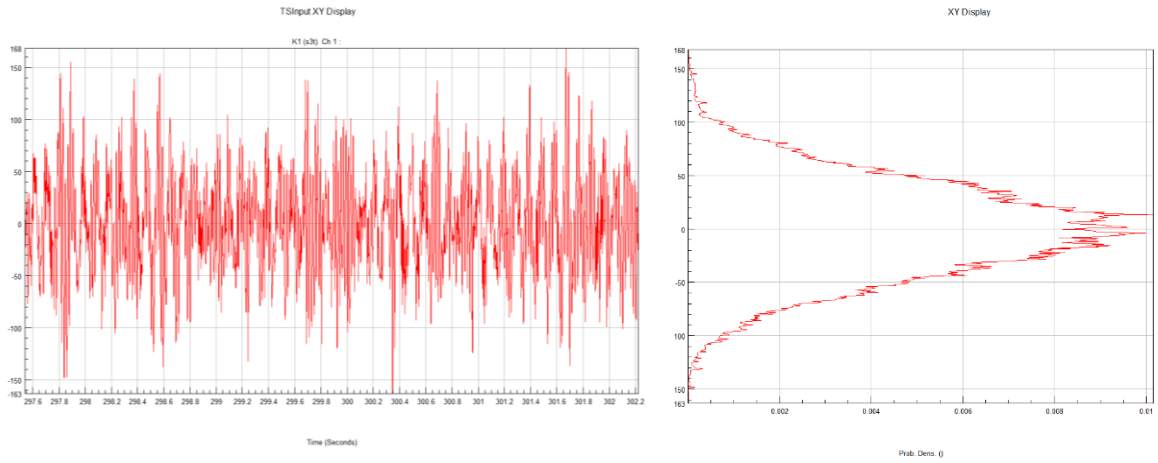


Figure 1 Gaussian signal and probability density distribution

Thus σ is the standard deviation, μ_n is the nth central moment

$$S = \frac{\mu_3}{\sigma^3} \tag{1}$$

$$K = \frac{\mu_4}{\sigma^4} \tag{2}$$

When the kurtosis or skewness has a condition that does not satisfy the Gaussian distribution, the signal is called a non-Gaussian distribution. At the same time, according to the index, when K>3, it is called leptokurtic, and when k<3, it is called platykurtic. According to the probability density distribution of the non-Gaussian signal, it can be found that the kurtosis reflects the distribution of acceleration in the middle region.[7]

In order to obtain an acceleration signal of a non-Gaussian distribution, it is usually obtained by Gaussian signal transformation. There are many methods used, Hermit polynomial, Gaussian mixture model, Phase selection method, Power-law model, Exponential method, Non-parametric method. In this paper, Steinwolf's phase selection method is used to modulate non-Gaussian signals with specific skewness and kurtosis.[8]

3. Phase selection method

By fast Fourier transform, an acceleration time domain signal can be described as a superposition of harmonics in the frequency range.

$$x(t) = \sum_{n=1}^{N} A_n \cos(2\pi n \Delta f t + \varphi_n) \tag{3}$$

The amplitudes of the harmonics are obtained

$$A_n = \sqrt{2\Delta f S(n\Delta f)} \tag{4}$$

Then the nth centre moment could be obtained

$$M_z = \lim_{T\to\infty} \frac{1}{T} \int_0^T \{x(t)\}^z dt = \frac{1}{T} \int_0^T \{x(t)\}^z dt, \ z > 2 \tag{5}$$

After the kurtosis formula could be written as

$$K = \frac{M_4}{M_2^2} = 3 - \frac{3}{2}\left(\frac{\sum_{n=1}^{N} A_n^4}{\left(\sum_{n=1}^{N} A_n^2\right)^2}\right) + \left(\frac{1}{2}\sum_{n=1}^{N} A_n^2\right)^{-2} \{\frac{3}{2}\sum_{\substack{i+2j=k \\ i\neq j}} A_i {A_j}^2 A_k \cos(\varphi_i + 2\varphi_j - \varphi_k) +$$

$$\frac{3}{2}\sum_{i<j} {A_i A_k}^2 A_j \cos(\varphi_i + \varphi_j - 2\varphi_k) + 3\sum_{i<j,k<m,i<k} A_i A_j A_k A_m \cos(\varphi_i + \varphi_j - \varphi_k -$$

$$\varphi_m) + 3\sum_{i<j<k<m} A_i A_j A_k A_m \cos(\varphi_i + \varphi_j + \varphi_k - \varphi_m) + \frac{1}{2}\sum_{i=3j} A_i {A_j}^3 \cos(\varphi_i - 3\varphi_j)\} \tag{6}$$

In the case of ensuring that the mean and RMS of the non-Gaussian signal are not changed, only the distribution is changed. So to make the K fitted for the experiment by modulating the specific φ. The resulting non-Gaussian distribution acceleration time series with specific K could be found. It should be noted that as the bandwidth increases， the amount of calculation becomes very large and also affecting the amount is the sampling frequency and data length. In general, the required non-Gaussian data is obtained in a combination of several methods.

4. Rain-flow count and damage calculation

The rain-flow counting method was proposed in the 1950s by two British engineers, M. Matsuishi and T. Endo.[9] The main function of this counting method is to simplify the measured load history into several load cycles for fatigue life estimation and fatigue test load spectrum. It is based on the two-parameter method and considers two variables of dynamic strength (magnitude) and static strength (mean). The rain flow counting method is mainly used in the engineering field, and is widely used in the calculation of fatigue life.

Through the rain flow counting method, the response stress is cyclically counted, and the stress amplitude-cycle number curve is obtained, which is the ***p(s)*** curve required for the fatigue damage calculation. The ultimate goal of the frequency domain based PSD method is also to approximate the fitting through various frequency models through the frequency domain information, and finally obtain the probability density curve ***p(s)*** of the stress response.

The S-N curve of the material expresses the number of life cycles of the material under different stresses. Usually expressed in the Basquin model.[10]

$$S^m N = C \tag{7}$$

And then according Miner's Law[11], calculation formula of damage in unit time could be written

$$D = v_p C^{-1} t_0 \int_0^{\infty} S^m p(s) ds \tag{8}$$

Then the prediction life could be obtained by D

$$T = \frac{1}{D} \times t_0 \tag{9}$$

5. Case study

In order to study the influence of the probability distribution of the response stress on the structural damage calculation, using the same frequency range, the 0 mean, the RMS is the same, the skewness is 0, and the different time series with the kurtosis between 2.9-10 are applied to the same double notched specimen, the effect of kurtosis on the damage calculation is obtained by comparing the damage conditions.

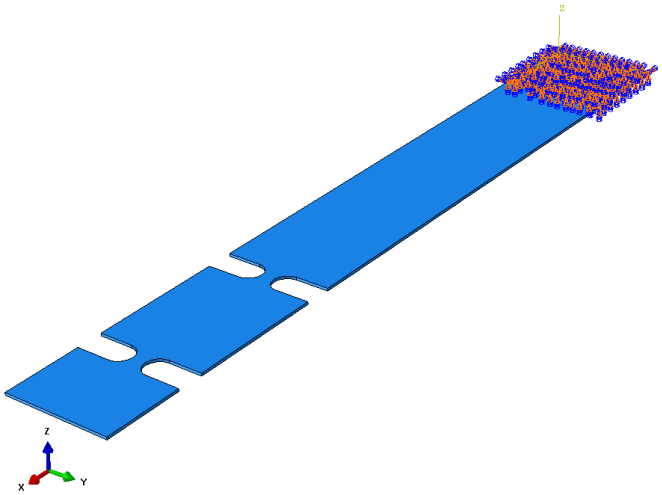


Figure 2 Two notch specimen & excitation

To meet the requirement, 10s time series of acceleration which allow non-Gaussian distribution were generated with different kurtosis. The bandwidth selection is based on the modal analysis result of the specimen, the second-order bending mode is 51.2 Hz, and 15-95 Hz is selected as the frequency range of the excitation signal. The sampling ratio is 1024Hz according to the band width.

Frequency	PSD (g^2/Hz)	(m^2/s^4)/Hz
15	0.25	12.03
95	0.25	12.03
RMS	4.472136g	31.02m/s^2

Table 1 Frequency domain information (g=9.81 m/s^2)

The simulation part is finished by Abaqus. Before taken simulation, the modal analysis was used to check FE model to do verify and validation according mode frequency and stress distribution. Though the modal dynamic, the response stress of different times series could be obtained.

No.	1	2	3	4	5	6	7	8
RMS(m/s^2)	31.44	31.32	31.27	31.31	31.10	30.84	31.54	31.44
Kurtosis	3.01	3.50	4.01	4.53	5.02	8.04	10.24	2.85
No.	9	10	11	12	13	14	15	16
RMS(m/s^2)	31.44	31.44	31.44	31.44	31.44	31.44	31.40	31.52
Kurtosis	2.92	2.80	3.06	2.96	3.12	2.90	3.15	3.21

Table 2 The RMS and Kurtosis of acceleration

After, the response stress result was show that the specimen is not linear structure, the time series of response stress was as below. The direction opposite to the gravitational acceleration is the Z-axis forward direction, and the nodal stress at the left side of the upper surface of the model is selected as the research object.

The overall research strategy is shown below. The focus is on the difference in stress cycling and damage values after rain flow counting. In general, the infinite fatigue stress of the material

is selected as the threshold value of the rain flow count, that is, in this stress cycle and below, the material has an infinite life. However, in order to compare the results of the data, this threshold is set to zero.

6. Results

According the result of response stress, it could be found that the kurtosis were influence by structure. Moreover, as the acceleration kurtosis increases, the kurtosis of the response stress decays more significantly.

No.	1	2	3	4	5	6	7	8
RMS(MPa)	28.89	29.00	28.88	29.74	29.39	30.21	31.77	29.07
Kurtosis	3.07	3.30	3.52	3.92	3.70	6.11	6.49	2.85
No.	9	10	11	12	13	14	15	16
RMS(MPa)	28.89	29.09	29.17	28.34	28.89	28.45	29.00	28.93
Kurtosis	3.02	2.78	2.90	3.07	3.88	3.14	3.04	3.18

Table 3 The RMS and Kurtosis of response stress

By comparing the results of time-domain and frequency-domain methods with different kurtosis, it can be found that as the kurtosis increases, the stress distribution is more concentrated in the low-stress region, but the maximum value of the stress amplitude is significantly increased. The key to damage deviation. Moreover, there is a possibility that the maximum value of the stress amplitude is greater than the allowable stress directly causing structural failure.

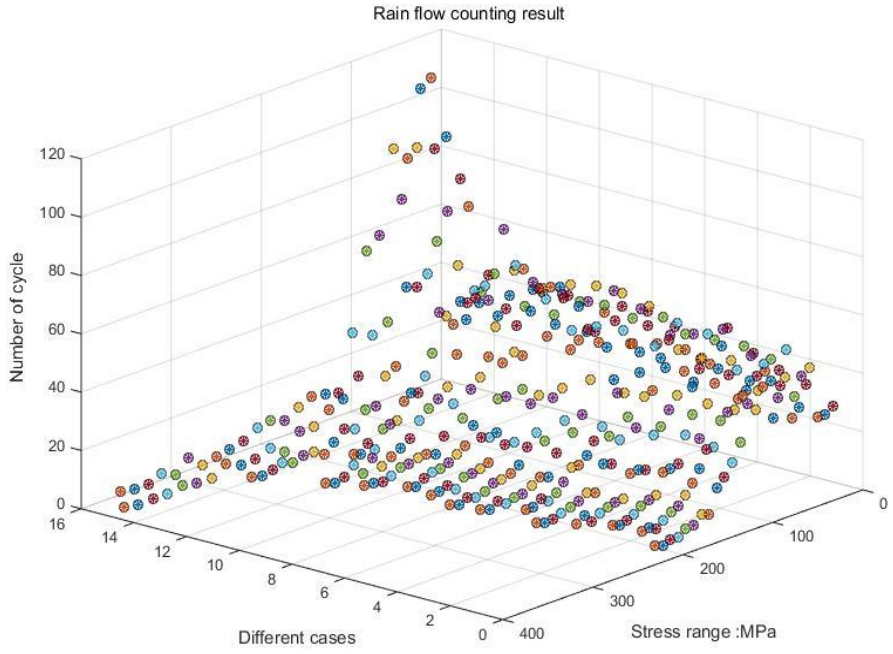


Figure 3 Rain flow counting result

The DIRLIK method is selected to obtain the stress expectation based on the Gaussian distribution of the frequency domain method, which is used as a reference object and compared with the results under the time domain signal.

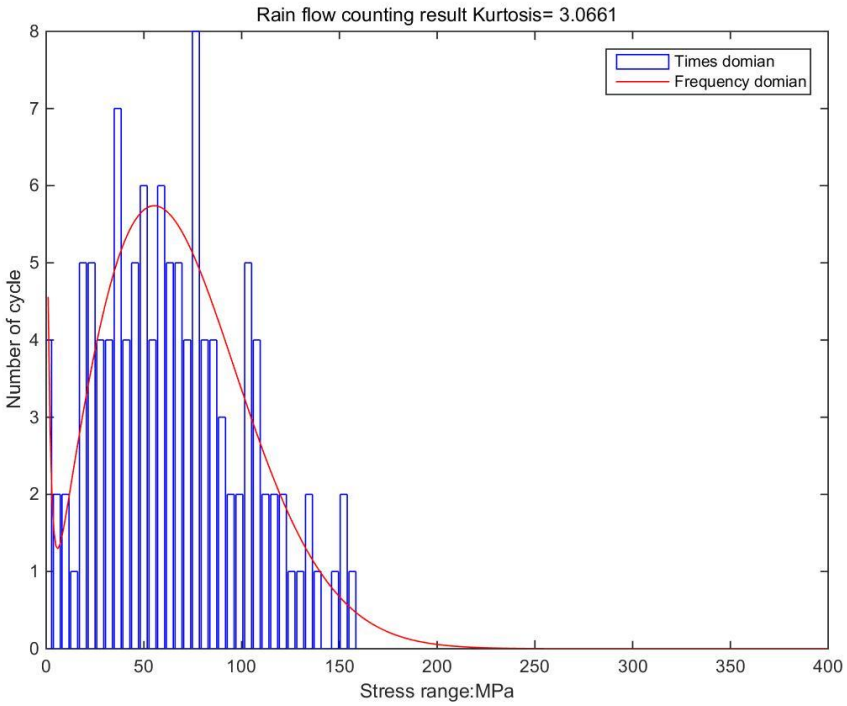


Figure 4 Rain flow counting of K=3.07

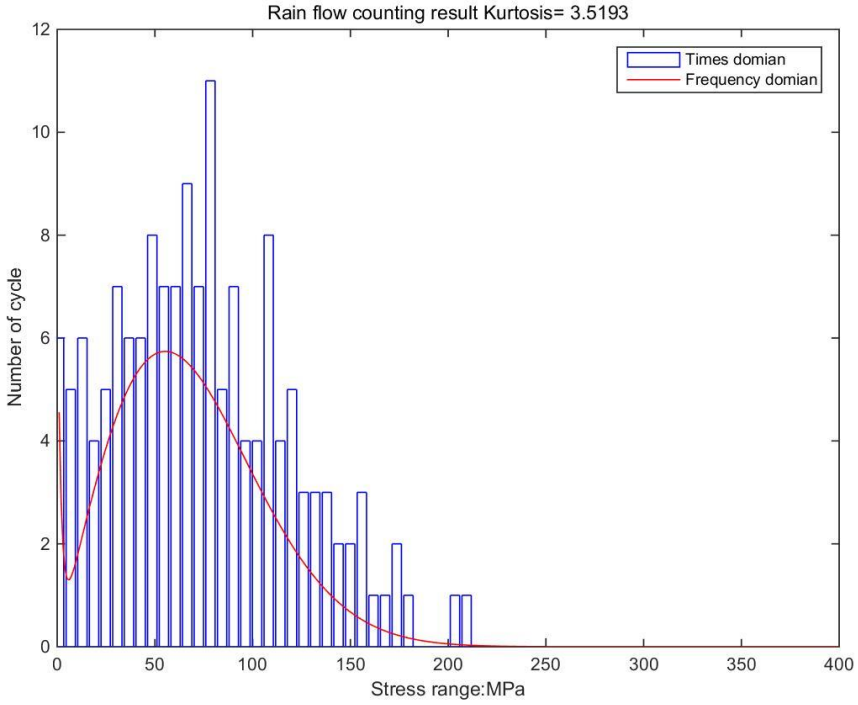


Figure 5 Rain flow counting of K=3.52

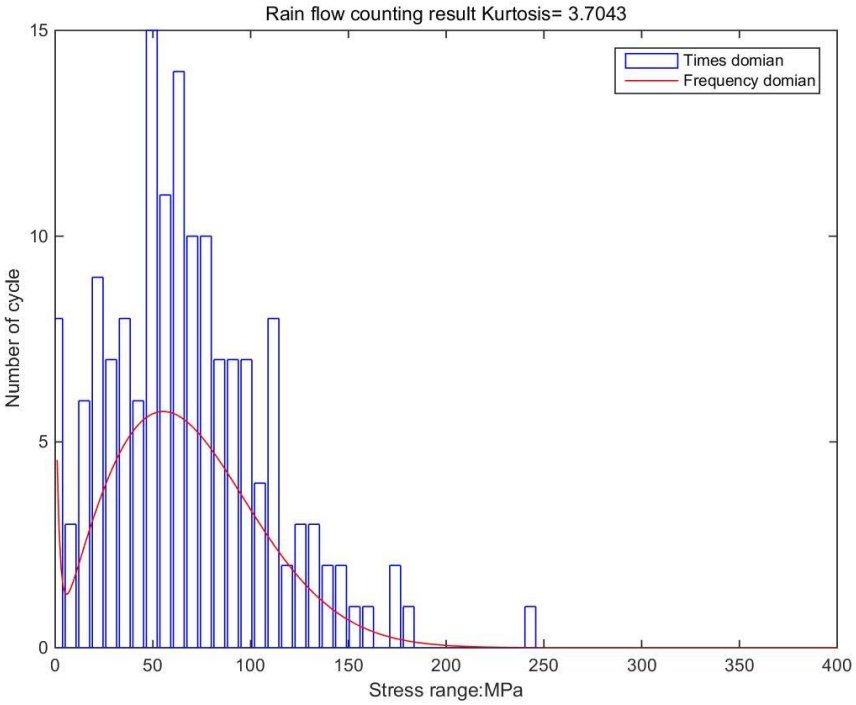


Figure 6 Rain flow counting of K=3.70

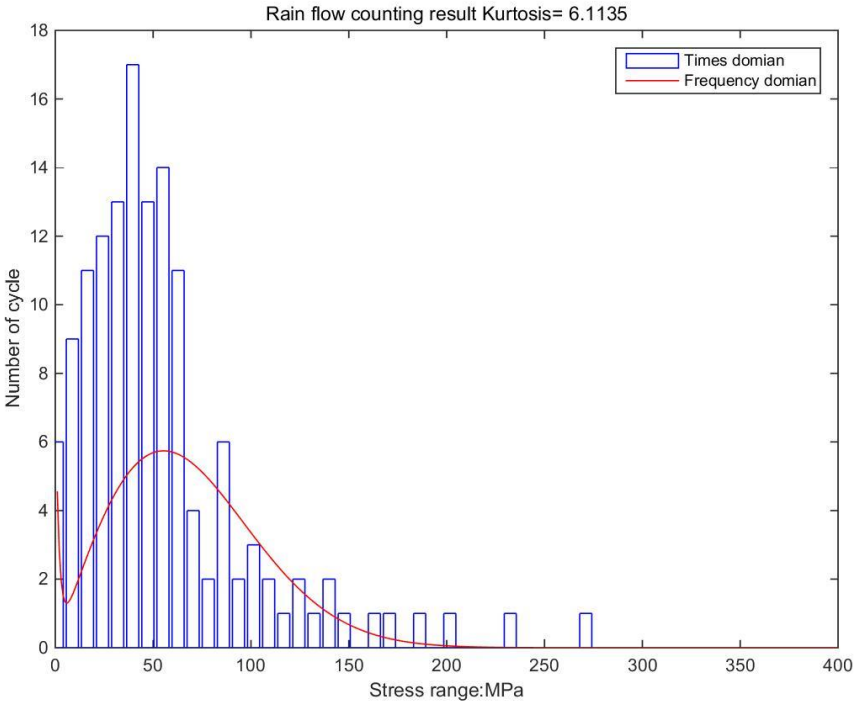


Figure 7 Rain flow counting of K=6.11

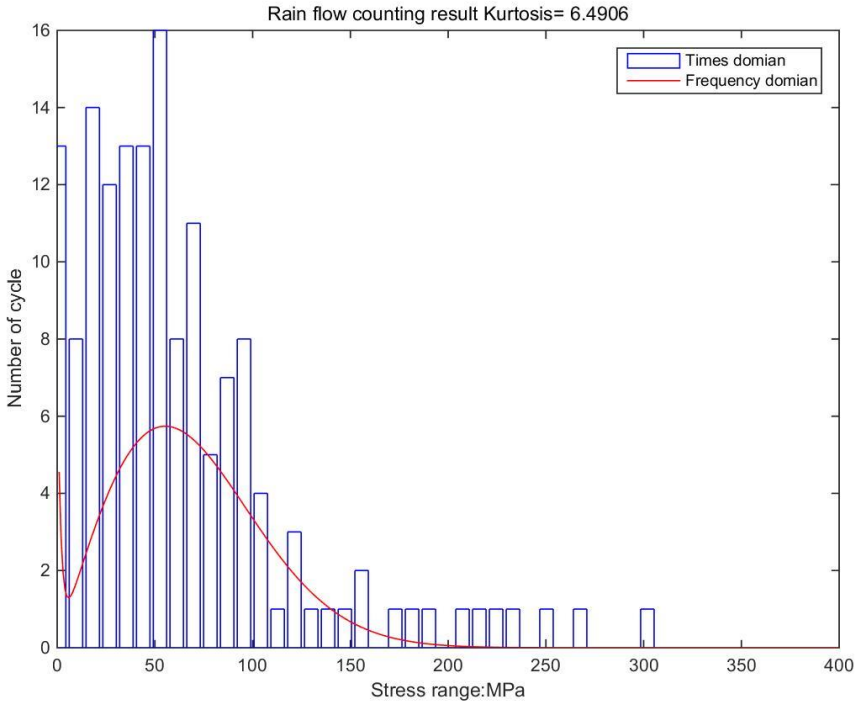


Figure 8 Rain flow counting of K=6.49

Though Miner's Law, the linear cumulative damager could be calculation,

$$D = \frac{n}{N} \tag{10}$$

For the different kurtosis time series data, the error control γ could be written ,

$$\gamma = 1 - \frac{D_K}{D_{PSD}} = \frac{\sum_0^{Smax} \cdot N_k(s) \cdot s^m}{v_p \int_0^{\infty} S^m p(s) ds} \tag{11}$$

When $\gamma \leq 30\%$, this can accept the results of using a PSD-based Gaussian signal for non-strict Gaussian vibration signals. Beyond this range, fatigue calculation methods based on non-Gaussian vibration signals must be used. Generally, when the kurtosis is less than 3.5, the result is within an acceptable range. Of course the accuracy of this result is limited by the length of the data. The longer the data, the more obvious the distribution. Smaller samples are relatively more affected by randomness.

7. Conclusion

Based on the analysis results, it can be found that the data difference is affected by the material property ***m***. At the same time, due to the structural relationship, the actual excitation kurtosis is greater than the threshold, but the response stress after attenuation can be calculated according to the Gaussian distribution.

Because of the high kurtosis, the response stress exceeds the limit of material strength, and should be avoided in industrial design. For response stresses with high kurtosis, there is a possibility that the maximum value exceeds the strength limit. This should be taken seriously.

8. Acknowledgement

This contribution has been elaborated under the **China Scholarship Council**

References:

[1] Y. WANG, R. SERRA *Life assessment of structures using random vibration testing*, ENVIRORISK Bourges 3ème edition June 2018

[2] R. SERRA, L. KHALIJ et al. *Effets de l'environnement vibratoire sur la durée de vie d'une éprouvette en aluminum*. SF2M, 35èmes Journées de Printemps, Mai 2016.

[3] L. Crow, Edwin & S. Bendat, J & G. Piersol, A. (1968). Measurement and Analysis of Random Data. Technometrics. 10. 869. 10.2307/1267469.

[4] H. Wirsching, P & C. Light, Mark. (1980). Fatigue under wide band random stress. ASCE J Struct Div. 106.

[5] D.Benasciutti R.Tovo & A.Cristofori, (2013), Analogies between spectral methods and multi axial criteria in fatigue damage. Probabilistic Engineering Mechanics.

[6] Dirlik, Turan. (1985). Application of computer in fatigue analysis.

[7] Benasciutti, Denis & Tovo, Roberto. (2006). Fatigue life assessment in non-Gaussian random loadings. International Journal of Fatigue. 28. 733-746. 10.1016/j.ijfatigue.2005.09.006.

[8] Steinwolf, Alexander. (2015). Vibration testing of vehicle components by random excitations with increased kurtosis. International Journal of Vehicle Noise and Vibration. 11. 10.1504/IJVNV.2015.067983.

[9] Matsuichi, M & Endo, T. (1968). Fatigue of Metals Subjected to Varying Stress. Proceedings of the Kyushu Branch of Japan Society of Mechanics Engineering. 102.

[10] Basquin, O.H.. (1910). The exponential law of endurance testing. 19. 625-630.

[11] Miner, M.A.. (1945). Cumulative damage in fatigue. J. Appl. Mech.. 12. A159-A164.

Fault diagnosis and prognosis for rolling bearings

Synchronous fitting for deterministic signal extraction in non-stationary regimes: Application to helicopter vibrations

Dany ABBOUD[1], Yosra MARNISSI[1], Amadou ASSOUMANE[1], Mohammed ELBADAOUI[1,2]

[1] Signal & Data Analytics Lab, SafranTech, Rue des Jeunes Bois – Châteaufort 78772 Magny –les-Hameaux, France
[2]Univ Lyon, UJM St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France

Dany.abboud@safrangroup.com

Abstract

Deterministic-random separation is crucial in machine signal processing. The synchronous average is a widely used tool that separates the deterministic contribution from the random one. This tool consists on averaging the cycles of the vibration signal. In fact, it uses the fact that, for a given location in the cycle, the associated samples have a constant mean. This makes possible to estimate the signal mean through synchronous averaging, i.e. by averaging the samples associated with each position in the cycle. However, in many practical applications, the cycle-to-cycle statistics can change according to many factors such as the speed, torque, load, etc. The resulting signal is widely referred in the literature as cyclo-non-stationary. This means that the mean signal is not periodic anymore, thus jeopardizing the synchronous average technique. This paper addresses this issue by proposing a new generalization of the synchronous average. The proposed method takes advantage of the smoothness of the statistics (in particular the mean) variation with respect to cycles. Instead of computing the average of the samples located at a given angular location, the time-varying mean is computed by optimally fitting the data with an appropriate curve. This defines the synchronous fitting idea, being a mean estimator of cyclo-non-stationary signals. Two solutions are proposed to solve the fitting problem. Whereas the first seeks for a global solution, the second adopts a local solution inspired from Savitzky–Golay filter. These two approaches are tested and compared on numerical and real signals captured from a helicopter engine operating under a runup regime. Overall, the results have asserted the superiority of the local approach over the global one.

1. Introduction

The theory of cyclostationary processes has proven to be effective in describing and processing rotating machine signals [Antoni 2009]. The vibration components generated by mechanical sources can be mainly classified into first and second (or higher) order cyclostationary classes. First order cyclostationary components are those deterministic, principally consisting of a set of sinusoids corrupted with stationary noises. Those are described by the (quasi-) periodicity of their mean. Examples of first-order phenomena can include gear meshing vibrations, shaft unbalance and misalignment, fan rotations and others. Second-order cyclostationary components are random in nature, meaning that their mean equals zero or, equivalently, their spectrum does not exhibit clean harmonics. The periodicity of these components is hidden and can be revealed through the instantaneous power or, more generally, the auto-covariance function. In practice, those components are generated by different kind of mechanical phenomena subjected to some randomness. A typical example is the vibrations generated by a local fault in a rolling element bearing wherein the randomness is due to the presence of a slippage in the motion of the rolling elements [Ho 2000]. For this reason, the cyclostationary analysis offers an efficient way to detect and characterize the presence of clear or hidden periodicities in the signals through a thorough differential diagnosis. Obviously, the separation of first and second-order components is crucial for an accurate analysis of the signal. Nowadays, the corresponding state of the art comprises a set of supervised and

unsupervised signal processing tools that deal with this issue. Among these methods, one of the most widely used is the synchronous average (SA) [Braun 1975]. The latter simply consists of cutting the signal into slices of the same length, being equal to the fundamental period of the extracted component and averaging them together. As it will be shown later, this paper deals with its generalization.
The cyclostationary modelling assumes the (hidden-) periodicity to be stable in time, which in turn requires a constant speed. Such a condition is however hard to obtain as the speed often undergoes some fluctuations. This jeopardizes the effectiveness of the SA even if the magnitude of the speed fluctuations is low. Since repetitive patterns in rotating machines are intrinsically locked to specific angular positions, it totally makes sense to rather process the signal in the angular domain. In this case, the cyclostationary property holds in the angle domain and, consequently, the SA is applied on the angular signal, either obtained by angular sampling or resampling [Antoni 2004].
In the case of large speed fluctuations, signals are subjected to significant distortions that jeopardize the effectiveness of the SA. These distortions are basically introduced by (i) variations of the machine power intake and (ii) the effect of linear time-invariant (LTI) transfers. Whereas the former essentially results in amplitude modulation, the latter also induces phase modulation. Non-periodic modulations obviously invalidate the (angle-) CS assumption and call for a more general description of nonstationary signals. Accordingly, the principle of cyclo-non-stationarity was proposed to formalize this specific type of signals. The consideration of cyclo-non-stationary signals requires the extension of the cyclostationary signals. This paper is particularly concerned in extending the synchronous average. Many previous works have addressed this issue. Reference [Coats 2009] proposed the improved synchronous average, being based on resampling the signal with a virtual tachometer signal synthesized via the demodulated phase. Another attempt to generalize the SA was proposed in Ref. [Daher 2010] through a parametric approach. In details, the authors used the Hilbert space representation of the deterministic component in which they decomposed the deterministic components onto a set of periodic functions multiplied by speed-dependent functions apt to capture long-term evolution over consecutive cycles. In ref. [Abboud 2016], the authors proposed a non-parametric approach based on averaging the signal cycles that belong to a given regime, defined by its central speed and a pre-defined width.
This paper proposes a different approach to generalize the SA based on a synchronous curve fitting of the data. The theoretical backgrounds of the proposed technique is exposed in section 2, while its performances are evaluated through numerical simulations in section 3. In section 4, the efficiency of the technique is tested on real vibration signals recorded under a varying speed condition.

2. Description of the synchronous fitting technique

In this section, the fundamentals of the proposed method are provided. First, a mathematical model for CNS signals is reviewed. Then, a global solution for the first order CNS estimation, which corresponds to the one proposed in [Daher 2009], is presented. Finally, the newly proposed technique based on a local solution is introduced.

2.1. General

Let $x[n]$ be a first-order CNS signal with a characteristic period N (i.e. cycle and $1/N$ the normalized frequency) and a length L. One can model such a signal as follows:

$$\forall n \in \{1, \dots, L\} \quad x[n] = d[n] + w[n] = \sum_k d_k[n] e^{\frac{j2\pi kn}{N}} + w[n] \tag{1}$$

where k is an integer, $d_k[n] \in \mathbb{C}$ are deterministic smooth functions (whose real and imaginary part are continuous and differentiable) and whose bandwidths, noted B_k, are much smaller than the half the fundamental frequency i.e. : $\forall k, B_k \ll 1/2N$ and $w[n]$ is a random noise. The discrete-time Fourier transform (DTFT) of (1) reads:

$$\forall f \in]-1/2; 1/2] \quad X[f] = \sum_k D_k[f] * \delta(f - k/N) + W[f] \tag{2}$$

where $D_k[f]$ and $W[f]$ are respectively the DTFTs of $d_k[n]$ and $w[n]$. Since $d_k[n]$ are deterministic smooth functions, and according to the Weirstrass theorem, they can be approximated through a P-order polynomial function, i.e.:

$$\forall k \quad d_k[n] \approx \sum_{p=0}^{P} d_k^p n^p \tag{3}$$

where $d_k^p \in \mathbb{C}$. By inserting Eq. (3) into the expression of $d[n]$, one obtains:

$$\forall n \in \{1, \dots, L\} \quad d[n] = \sum_{p=0}^{P} c_p[n] n^p \tag{4}$$

where $c_p(n) = \sum_k d_k^p e^{j2\pi kn/N}$ is a periodic function of period N. Equation (4) indicates that the deterministic component can be approximated by a sum of periodic functions multiplied with the polynomial basis: it is actually a polynomial with periodic coefficients.

Let's first define $\bar{n} = \lfloor (n-1)/N \rfloor + 1$ as the sample location within the period N ($\lfloor a/b \rfloor$ denotes the remainder of the division of a by b) . Since $c_p(n)$ is periodic with period N, we have $c_p[\bar{n}] = c_p[\bar{n} + (q-1)N]$ for all integer $q = 1, \dots, Q$ (Q is the number of cycles). Thus, Eq. (4) can be equivalently written as follows:

$$\forall q \in \{1, \dots, Q\} \ \forall \bar{n} \in \{1, \dots, N\} \quad d[\bar{n} + (q-1)N] = \sum_{p=0}^{P} c_p[\bar{n}] (\bar{n} + (q-1)N)^p \tag{5}$$

Using the binomial theorem ($(\bar{n} + (q-1).N)^p = \sum_{i=0}^{p} C_i^p N^i \, (\bar{n} - N)^{p-i} q^i$ where C_i^p is the binomial coefficient), one can deduce from Eq. (5) that the samples associated with the same location $\bar{n}$ in the period, $s_q[\bar{n}] = d[\bar{n} + (q-1).N]$ for all integer $q \in \{1, \dots, Q\}$, defines a polynomial of order P with constant coefficient, i.e.:

$$\forall q \in \{1, \dots, Q\} \, \forall \bar{n} \in \{1, \dots, N\} \quad s_q[\bar{n}] = \sum_{p=0}^{P} b_p[\bar{n}] q^p \tag{6}$$

where $b_p[\bar{n}] = N^p \sum_{j=p}^{P} C_p^j \, (\bar{n} - N)^{j-p} c_j[\bar{n}]$. Note that $b_p[\bar{n}]$ is parametrized by $\bar{n}$.

2.2. A global LMS solution

In the case of a noisy signal $x[n]$, a good estimate of the deterministic component $d[n]$ is then to find the best fit of the curve $\boldsymbol{s}[\bar{n}] = \left[s_1[\bar{n}], \dots, s_Q[\bar{n}]\right]^T$ for each $\bar{n} \in \{1, \dots, N\}$ which reduces to find an estimate of $\boldsymbol{b}[\bar{n}] = \left[b_1[\bar{n}], \dots, b_P[\bar{n}]\right]^T$ for each $\bar{n} \in \{1, \dots, N\}$ for a given polynomial of order P. A common way to do this is to find the curve which minimizes the least mean square error, i.e.:

$$\begin{aligned} \forall \bar{n} \in \{1, \dots, N\} \quad \widehat{\boldsymbol{b}}[\bar{n}] &= \operatorname{argmin}\left(\sum_{q=1}^{Q} w[\bar{n} + (q-1)N]^2\right) \\ &= \operatorname{argmin}\left(\sum_{q=1}^{Q} \left(d[\bar{n} + (q-1)N] - x_q[\bar{n}]\right)^2\right) \\ &= \operatorname{argmin}\left(\sum_{q=1}^{Q} \left(\sum_{p=0}^{P} b_p[\bar{n}] q^p - x_q[\bar{n}]\right)^2\right) \end{aligned} \tag{7}$$

where $x_q[\bar{n}] = x[\bar{n} + (q-1)N]$. Let's define the $Q \times (P+1)$ matrix $\boldsymbol{\Phi}$ such that $\boldsymbol{\Phi}_{q,p} = q^{p-1}$ (with $q \in \{1, \dots, Q\}$ and $p \in \{1, \dots, P+1\}$, and $\boldsymbol{x}[\bar{n}] = \left[x_1[\bar{n}], \dots, x_Q[\bar{n}]\right]^T$. One can rewrite the Eq. (7) as:

$$\forall \bar{n} \in \{1, \dots, N\} \ \widehat{\boldsymbol{b}}[\bar{n}] = \operatorname{argmin} \|\boldsymbol{\Phi b}[\bar{n}] - \boldsymbol{x}[\bar{n}]\|^2 \tag{8}$$

whose solution expresses as follows:

$$\forall \bar{n} \in \{1, \dots, N\} \ \widehat{\boldsymbol{b}}[\bar{n}] = \left(\boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{x}[\bar{n}] \tag{9}$$

Once the coefficients are calculated, one can find the estimated deterministic component located at $\bar{n}$ in the form:

$$\forall \bar{n} \in \{1, \dots, N\} \ \hat{\boldsymbol{s}}[\bar{n}] = \boldsymbol{\Phi}\hat{\boldsymbol{b}}[\bar{n}] = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathbf{T}}\,\boldsymbol{x}[\bar{n}] \tag{10}$$

The deterministic signal can then be deduced as follows:

$$\forall n \in \{1, \dots, L\} \ \ d[n] = s_q[\bar{n}] \ \text{ where } \ \bar{n} = \lfloor (n-1)/N \rfloor + 1 \ \text{ and } \ q = 1 + \ (n - \bar{n})/N \tag{11}$$

2.3. A local LMS solution

This subsection describes the proposed method. The basic idea is excerpted from the "Savitzky-Golay filter" which is a widely known method to smooth or fit the data based on the least mean square solution of local polynomial fitting [Savitzky 1964]. Precisely, for every $q \in \{1, \dots, Q\}$, let's consider the data set $x_q[\bar{n}]$ being a function of q and parametrized by $\bar{n}$; we try to find the best LMS polynomial fit, with a fixed order P at the point $\bar{n}$, from the $2M+1$ subset centered at q, i.e. $\{x_{q-M}[\bar{n}], \dots, x_{q+M}[\bar{n}]\}$. That being said, this problem can be stated in a similar way as the previous subsection, i.e.:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\} \quad \widehat{\boldsymbol{b}}^{(\boldsymbol{q})}[\bar{n}] = \operatorname{argmin} \left\| \mathbf{J}\, \boldsymbol{b}^{(\boldsymbol{q})}[\bar{n}] - \boldsymbol{x}^{(q)}[\bar{n}] \right\|^2 \tag{12}$$

where $\boldsymbol{x}^{(q)}[\bar{n}] = \left[x_{q-M}[\bar{n}], \dots, x_{q-M}[\bar{n}]\right]^{\mathsf{T}}$ represents the q^{th} subset, $\boldsymbol{b}^{(\boldsymbol{q})}[\bar{n}] = \left[b_0^{(\boldsymbol{q})}[\bar{n}], \dots b_P^{(\boldsymbol{q})}[\bar{n}]\right]^{\mathsf{T}}$ are the $P+1$ polynomial coefficients associated with the q^{th} subset, and $\mathbf{J}$ the $(2M+1) \times (p+1)$ matrix such that $\forall m \in \{1, \dots, 2M+1\}\, \forall p \in \{1, \dots, P+1\}$ $\mathbf{J}_{m,p} = (m - M + 1)^{p-1}$. The $(2M+1)$-length curve that best fits the q^{th} subset writes:

$$s_m^{(q)}[\bar{n}] = \sum_{p=0}^{P} b_p^{(q)}[\bar{n}]\,.\,(m - M + 1)^p \tag{13}$$

The Savitzky-Golay method suggests to estimate the deterministic component at the q^{th} data point by retaining the value of the polynomial at the central point i.e. at $m = M + 1$:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\} \quad \hat{d}[\bar{n} + (q-1).N] = \hat{s}_{M+1}^{(q)}[\bar{n}] = b_{M+1}^{(q)}[\bar{n}] \tag{14}$$

Following the same lines as for Eq (9), one can show that the coefficients of the polynomial write:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\} \quad \widehat{\mathbf{b}}^{(\boldsymbol{q})}[\bar{n}] = \ \boldsymbol{H}\, \boldsymbol{x}^{(q)}[\bar{n}] \tag{15}$$

with $\boldsymbol{H} = \left(\mathbf{J}^{\mathbf{T}}\mathbf{J}\right)^{-1}\mathbf{J}^{\mathbf{T}}$ a matrix of size $(P+1) \times (2M+1)$ whose elements are independent of $\bar{n}$ and q . The $(M+1)^{th}$ element of the above vector namely $\hat{b}_{M+1}^{(q)}[\bar{n}]$ is actually a linear combination of $\boldsymbol{x}^{(q)}[\bar{n}]$ with the $2M+1$ elements of the $(M+1)^{th}$ row, $\boldsymbol{h}^{\boldsymbol{T}} = [h_{-M}, \dots h_M]$, of $\boldsymbol{H}$ being independent of q and $\bar{n}$:

$$\hat{b}_{M+1}^{(q)}[\bar{n}] = \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{x}^{(q)}[\bar{n}] \tag{16}$$

Considering equations (14) and (15), one can write the estimate of the deterministic component

$$\begin{aligned} \hat{d}[\bar{n} + (q-1).N] &= \sum_{m=-M}^{M} x_{q-m}[\bar{n}]\, h_m \\ &= \sum_{m=-M}^{M} x[\bar{n} + (q-1).N - m.N]\, h_m \\ &= \sum_{i=-MN}^{MN} x[\bar{n} + (q-1).N - i]\, \tilde{h}_i \end{aligned} \tag{17}$$

where $\tilde{\boldsymbol{h}}^{\boldsymbol{T}} = [\tilde{h}_{-MN}, \ldots, \tilde{h}_{MN}\ ]$ is obtained by zero-padding $\boldsymbol{h}$ as follows:

$$\begin{cases} \tilde{h}_i = h_m & if\ \ i = mN,\ \forall -M \le m \le M \\ \tilde{h}_i = 0 & elsewhere \end{cases} \tag{18}$$

It becomes obvious that the estimated deterministic component turns to a LTI filtering of the original signal $x[n]$ with the $(2MN+1)$-length filter $\tilde{\boldsymbol{h}}_i$:

$$\hat{d}[n] = \sum_{i=-MN}^{MN} x[n-i]\,\tilde{h}_i \tag{19}$$

3. Numerical evaluation

In this section, the performance of the synchronous fitting techniques are tested and compared on a synthetic signal. The deterministic signal is modelled as a sum of four speed-varying sinusoids whose envelopes and phases are functions of the cyclo-non-stationary $\lambda[n]$ (which can be in practice the torque, load, speed, etc.):

$$d[n] = \sum_{k=1}^{4} A_k[n] \sin(2\pi k n/N + \Phi_k[n]) \tag{20}$$

Where:

- $A_k[n]$ and $\Phi_k[n]$ are functions of $\lambda[n]$ (see Fig. 1);
- $N = 100$ is the fundamental period;
- $L = 15000$ is the signal length.

The deterministic signal is exposed in Fig 2 together with its noisy version constituted by adding a white Gaussian noise such that the initial signal to noise ratio is equal to -3 dB.

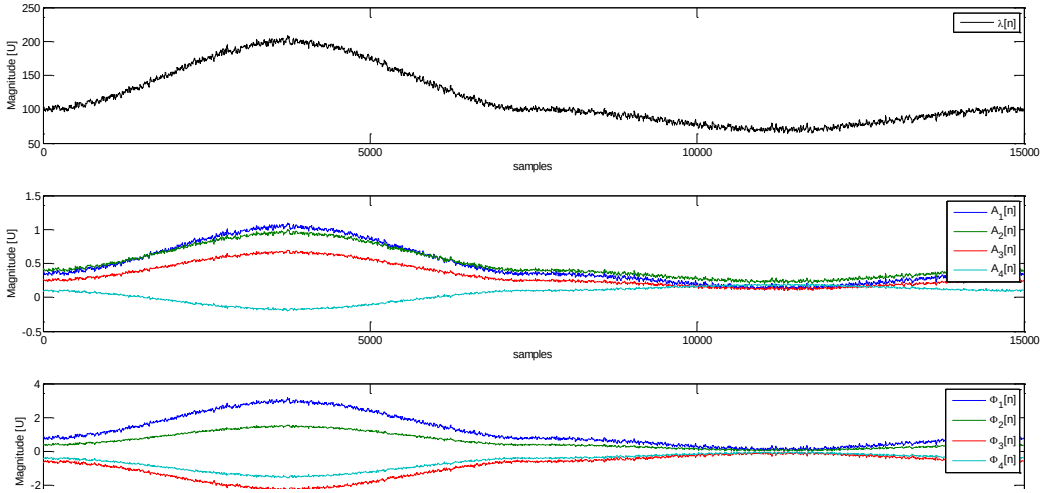


Figure 1: The plot of the cyclo-non-stationary agent $\lambda[n]$ (top), the 4 amplitudes $A_k[n]$ (middle) and the 4 phase modulations $\Phi_k[n]$ (bottom) associated with the sinusoids of the synthetic signal.

In the following, the global and local approaches are applied to the noisy signal with respect to the cycle N with the aim of recovering the deterministic component, being here the signal of interest. For the global approach, the degree of polynomial was set to 30, knowing that the results were stable for higher polynomial degrees. For the local approach, the window length was set to 49 (i.e. $M = 24$) and the

polynomial order to 3. The obtained results are exposed in Fig. 3 together with the error signal obtained by subtracting the estimated signal from the actual one. Both approaches tend to estimate with good accuracy the deterministic signal, with a clear superiority of the newly proposed local approach over the global one: the estimation error of the latter is almost twice larger than the former.
Eventually, the performance of these methods are compared for different signal-to-noise ratios (SNR). For this purpose, the relative error defined as the energy of the error normalized by the signal energy:

$$\epsilon_n[n] = 10\log_{10}\left(\sum_{n=1}^{N}\left(d[n]-\hat{d}[n]\right)^2 / \sum_{n=1}^{N} d[n]^2\right). \quad (21)$$

The obtained results are exposed in Fig. 4. The SA returns poor and consistent results as this latter only estimates the average periodic component existing in the signal. The reason is that the number of average is big so it was slightly affected by the SNR: the average periodic part was almost the same for all SNR. When it comes to the synchronous fitting techniques, the local approach evidences better estimation performances. In fact, the local approach returns an estimation error less than -6 dB when the SNR greater than -6 dB, while the global approach needs a SNR greater than 10 dB to get this accuracy. The results highlight the effectiveness of the local approach as compared with the global one. The reason is that the local approach assumes that the mean is locally smooth which is a more accurate assumption.

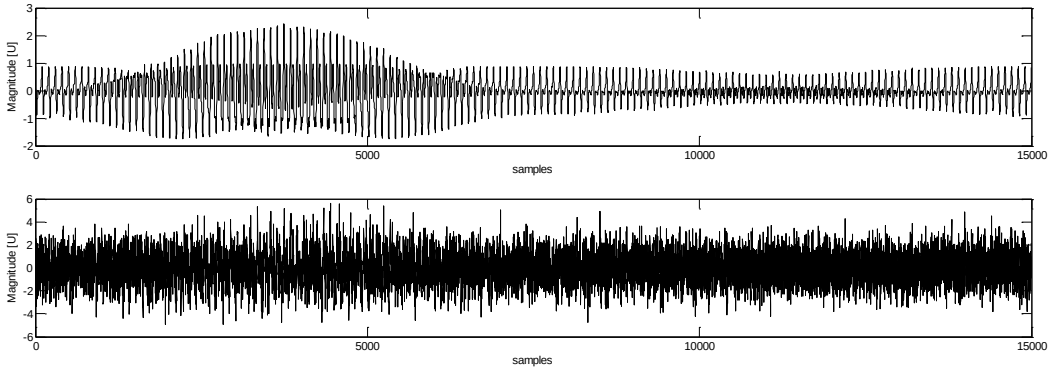


Figure 2: The deterministic component (top plot) and the noisy signal (bottom plot) constituted by adding a white Gaussian noise whose standard deviation equals twice of the former.

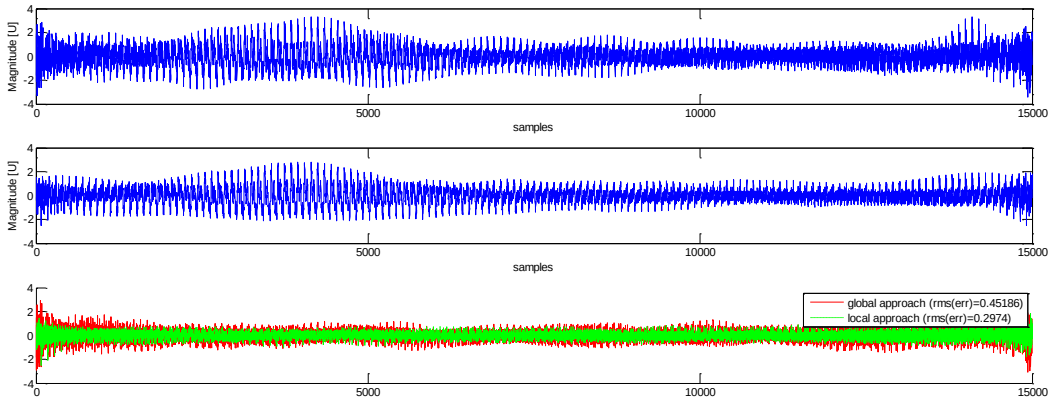


Figure 3: The deterministic component (top plot) and the noisy signal (bottom plot) constituted by adding a white Gaussian noise whose standard deviation equals twice of the former.

4. Application: a helicopter engine

In this section, the proposed approach is applied on real vibration signals captured from the gas generator of a helicopter engine. The aim is to extract the component related to the centrifugal compressor while the engine speed operates under a runup regime. An encoder is also present to measure the shaft location and to provide an accurate estimation of the engine speed. The encoder signal is used to resample the signal in the angular domain and the synchronous fitting techniques are both applied with respect to the blade pass period of the centrifugal compressor. It is worth noting that the blade pass period equals the shaft period divided by the compressor blade number. The blade number as well as the signals magnitude are not given for confidentiality reason. Figure 5 exposes the raw acceleration signal, the synchronous fitting estimations via the global and local approach. It is obvious that the signal associated with the local approach is much more accurate presenting a clear resonance starting at 10s. The related spectrograms are exposed in Figure 6 wherein the speed-varying harmonics of the centrifugal compressor are clearly shown. Though the extraction seems good in both techniques, it was hard to compare the performance of the extraction techniques. The order spectrum of the signals are computed by the Welch estimator applied to the angular resampled signals. The obtained spectra are exposed in Fig. 6 for comparison. Whereas the latter show better noise rejection in the global solution case especially for the noise floor, the close-ups clearly evidence the superiority of the local approach in accurately estimating the peaks. The global approach tends to loose accuracy as the frequency (order) gets high, this is because the global interpolation tends to confuse high frequency sinusoids with high frequency noises. Overall, the newly proposed local approach evidences better performances than the global one.

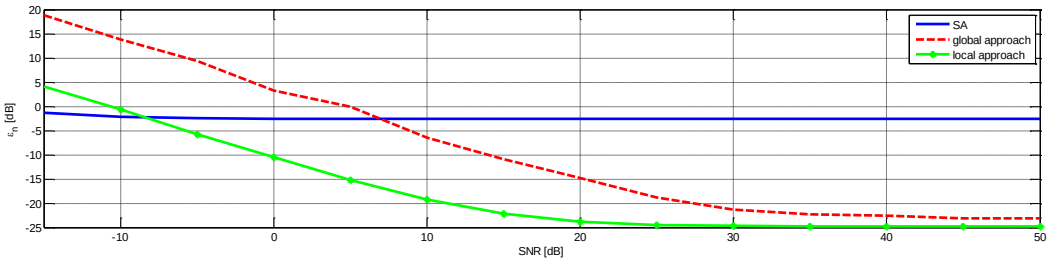


Figure 4: Performance of the SA, the synchronous fitting with the global and the local approach.

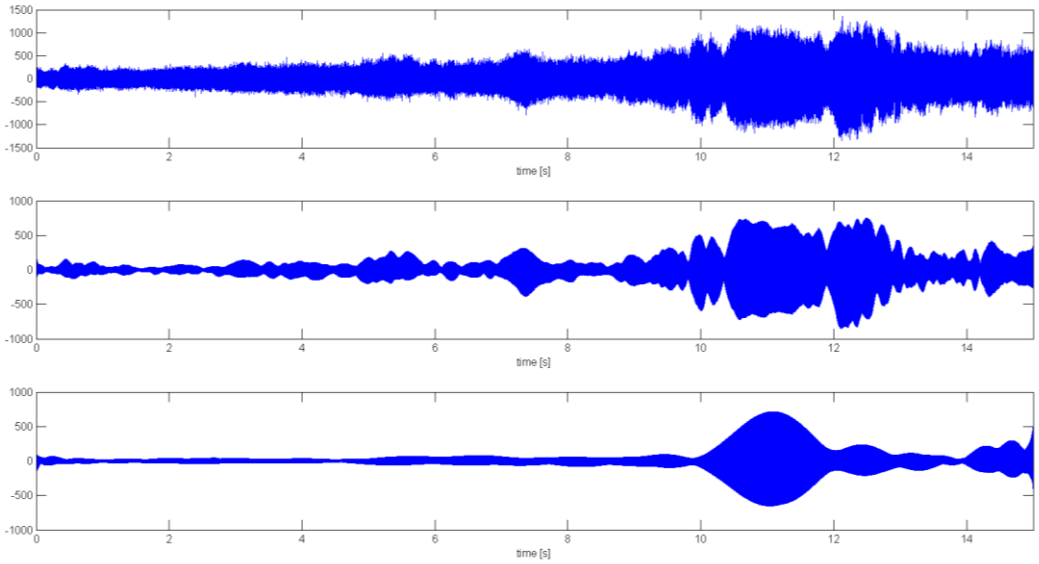


Figure 5: The raw acceleration signal (top), the synchronous fitting with the global approach (middle) and local approach (bottom).

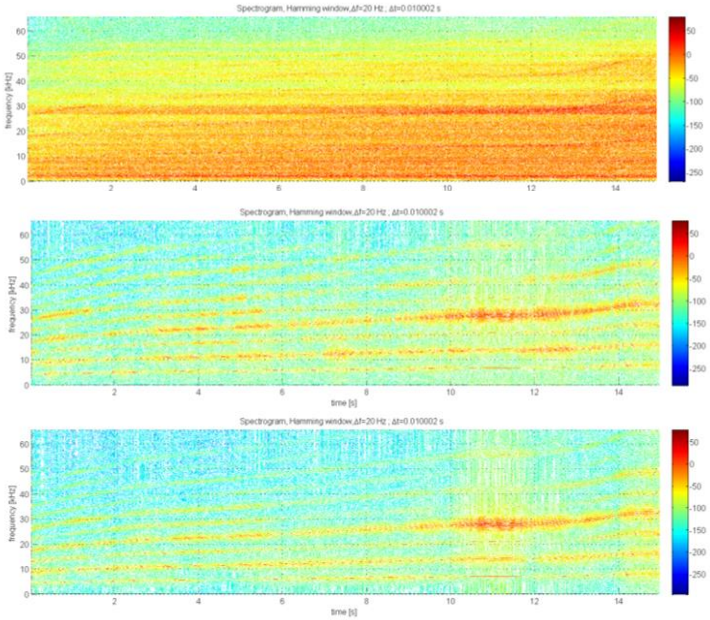


Figure 6: Spectrograms of the raw acceleration signal (top), the synchronous fitting with the global approach (middle) and local approach (bottom).

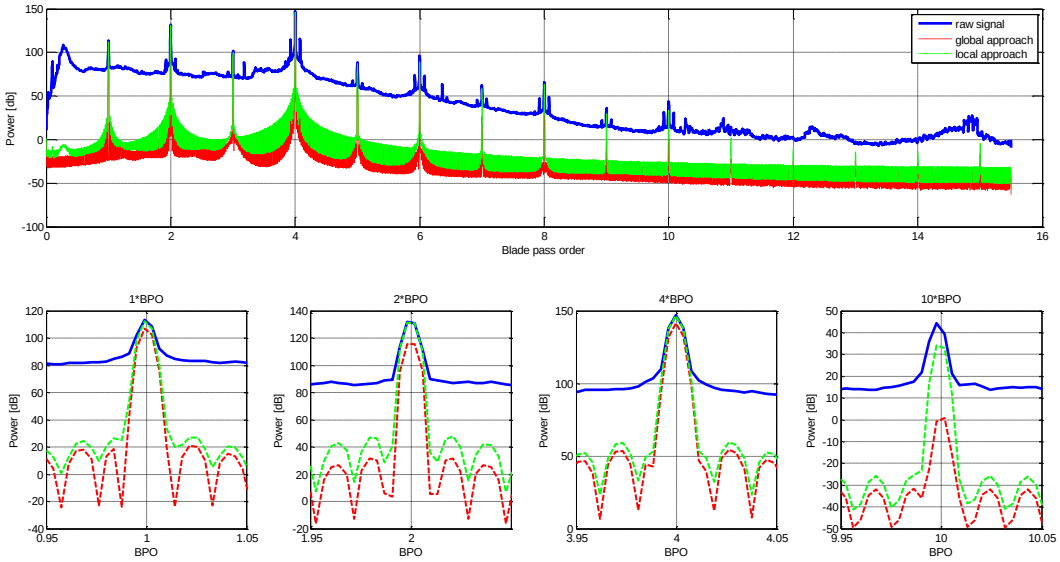


Figure 7: Order spectra of the raw acceleration signal (blue continuous line), the synchronous fitting with the global approach (red dotted line) and local approach (green dotted line).

5. Conclusion

This paper proposes a new technique for the extraction of a deterministic component in variable regimes. It can be seen as an extension of the classical synchronous averaging. In fact, instead of computing the cyclic mean via the synchronous average, the latter is computed via a synchronous fitting. This leads to two ways to tackle the issue. The first way seeks a global solution and compute the mean, for the data associated with a given position in the cycle, by finding the best polynomial that minimizes the least mean square error. It turns out that the solution of this problem was previously proposed in a previous publication. However, the second way is original and addresses differently the same problem by seeking a local solution based on the Savitzky-Golay filter. Numerical simulations are conducted showing a clear superiority of the newly proposed local approach over the global one. In fact, the local approach returns an estimation error less than -6 dB when the SNR greater than -6 dB, while the global approach needs an SNR greater than 10 dB to get this accuracy. The results highlights the effectiveness of the local approach as compared with the global one. An additional advantage of the local approach over the global one is the computational cost. As the first turns to a linear-time-invariant convolution, its implementation is much easier than solving a global least mean square problem which requires a matrix inversion. Eventually, both techniques are successfully tested on a real vibration signal measured on a helicopter engine under a runup condition with the aim of extracting the vibratory component emitted by the centrifugal compressor of the gas generator. Both techniques were able to extract the component of interest, yet the local approach evidences much better extraction than the global one.

References

[Antoni 2009] J. Antoni, Cyclostationarity by examples, Mech. Syst. and Sign. Proc., 23 (2009) 987–1036

[Ho 2000] D. Ho, Randall R. B., Optimization of bearing diagnostic techniques using simulated and actual bearing fault signals. Mech. Syst. and Sign. Proc. 14, 763-788 (2000)

[Braun 1975] S. Braun, The extraction of periodic waveforms by time domain averaging, Acta Acustica United with Acustica, vol. 32, no. 2, pp. 69-77, 1975.

[Antoni 2004] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, *Cyclostationary modelling of rotating machine vibration signals*. Mech. Syst. And Sig. Proc. 18, 1285–1314 (2004)

[Coats 2009] M.D. Coats, N. Sawalhi, R.B. Randall, Extraction of tacho information from a vibration signal for improved synchronous averaging, Proc. Acoust. (2009).

[Daher 2010] Z. Daher, E. Sekko, J. Antoni, C. Capdessus and L. Allam, Estimation of the synchronous average under varying rotating speed condition for vibration monitoring, Proceedings of ISMA 2010.

[Abboud 2016] D. Abboud, J. Antoni, S. Sieg-Zieba, M. Eltabach, *Deterministic-Random separation in non-stationary conditions*, Journal of Sound and Vibration, Volume 362, 3 February 2016, Pages 305-326

[Savitzky 1964] Abraham Savitzky et Marcel J. E. Golay, « Smoothing and Differentiation of Data by Simplified Least Squares Procedures », Analytical Chemistry, vol. 8, no 36, 1964, p. 1627–1639 (DOI 10.1021/ac60214a047)

High Frequency Demodulation Technique for Instantaneous Angular Speed Estimation

Frédéric Bonnardot[1], Khalid Lizoul[1], Saad Errafik [1], Hugo André[1], François Guillet[1]

[1]Univ Lyon, UJM Saint Etienne, LASPI, EA 3059
F-42334, IUT de Roanne, France.
{frederic.bonnardot}@univ-st-etienne.fr

Abstract

This paper adapts the super-heterodyne method from telecommunication field to speed estimation based on optical encoder signal.
This method uses an analog frequency shifting before sampling of the signal. Therefore, the required sampling frequency is reduced and linked with the speed fluctuation frequency. Spectrum of optical encoder signal is analysed and used to explain how to set up the frequency shift and limitations. Finally, a comparison is made with elapsed time.

1 Introduction

Many vibration monitoring methods use an accelerometer sensor in order to diagnose the systems. Instantaneous Angular Speed (IAS) has appeared to provide a new source of information on the rotating machine and it is commonly used to diagnose faults such as bearing faults[1], gearboxes [2], ...

Classically, IAS use an optical encoder as a speed sensor. This sensor is available in many rotating machines (especially for speed and position control). It provides a square / sine wave with a frequency proportional to the rotation speed. Thus, this signal is frequency modulated by the system speed variations. There are two major methods for recording IAS signal: timer / counter technic [3], and ADC- based methods [4]. The latter is limited by the capability of the ADC-board to collect the data (sampling frequency) which could restrict the use of higher encoder resolution and introduces spatial aliasing.

In this study, we are interested in IAS estimation using a super-heterodyne like demodulation technique. The idea is to be able to acquire the speed signal with relatively higher resolution without using a higher sampling frequency. The IAS signal is first analogously shifted in frequency domain in order to be acquired at a lower sampling rate. Then, the sampled signal is further treated to get IAS.

A first part presents the measurement principle, a second part discuss about limitations, finally a last part show a comparison with elapsed time technique.

2 Measurement principle

In frequency demodulation technique, the signal is recorded at a high sample rate in order to exploit only the spectral content between $[f_{cod} - B/2; f_{cod} + B/2]$. The idea of super-heterodyne technique illustrated by figure 1 that come from telecommunication is to shift analogically the optical encoder signal from the band $[f_{cod} - B/2; f_{cod} + B/2]$ to the band $[\varepsilon; \varepsilon + B]$ by using an analog multiplier (ε is a small margin). Therefore, the recorded signal could be recorded with a sampling frequency greater than $2 \times B$. Since the band B (a few kiloHertz) is smaller than optical encoder frequency, it becomes possible to use high resolution (4096 lines per revolution) with a classical data acquisition device. The demodulation effect is simply compensated after computing instantaneous speed by adding a constant frequency.

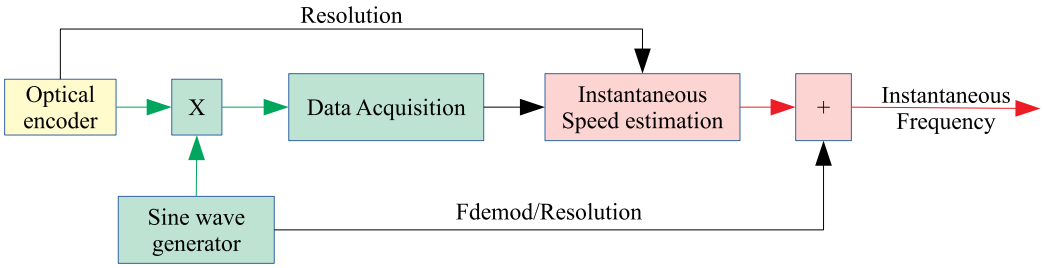


Figure 1: Principle of High frequency Demodulation Technique.

In order to illustrates the analog frequency shifting part, the sine wave generator, optical encoder signal and product signal has been acquired. Due to the limitation of data acquisition device, an optical encoder resolution of 256 lines has been used (to visualise original optical encoder signal). The sine wave generator has been set to $F_{mod} = 4\,kHz$. The spectrum of each signal is presented in figure 2.

The first spectrum is the spectrum of the sine wave generator. On top of the peak at $4\,kHz$, a third harmonic $68\,dB$ lower (i.e. amplitude 2400 lower) indicate the quality of the generator. It will also ask the question latter of the influence of a third harmonic on low-cost generators. The low noise in the spectrum is produced by the engine inverter as well as the generator.

The second spectrum corresponds to the optical encoder and was already studied in the first part. The effect of multiplication of encoder signal by sine wave results in a displacement of each component of the encoder spectrum by F_{mod} on the left and on the right as indicated by the arrow. Each component is identified by a colour to help analysis after demodulation.

The last spectrum shows the demodulated signals by using the analog multiplier. Each optical encoder component has been shifted in both left and right direction. The amplitude ratio of $20\,dB$ between 1^{st} and 3^{rd} optical encoder harmonic was kept after demodulation (on the bottom spectrum).

Unfortunately, components at f_{cod}, $3.f_{cod}$ (marked in gray) and F_{mod} are present after demodulation. It means that the product $p(t)$ signal is:

$$p(t) = [encoder(t) + \varepsilon_e].[sin(2\pi F_{mod} t) + \varepsilon_s] \quad (1)$$

$$= encoder(t).sin(2\pi F_{mod} t) + \varepsilon_e.\sin(2\pi F_{mod} t) + \varepsilon_s.encoder(t) \quad (2)$$

Where :

- $encoder(t)$ is a zero mean version of encoder signal obtained by a high pass filter,
- ε_e is an DC component (introduced by electronic),
- ε_s is an DC component (introduced by electronic despite the fine tune of the sine wave generator).

The attenuation between the gray signal at f_{cod} and the original red signal is $41\,dB$, it means that $\varepsilon_s = 0.009$. By looking at the sine wave the value $\varepsilon_e = 0.001$ could be identified.

The values ε_e and ε_s are in fact small, and, reduce them necessitate higher precision offset adjustment.

In practice the last spectrum corresponds to the only acquired signal. The only interesting part of the signal is the left red pattern corresponding to the demodulated f_{cod} part. This part could be acquired by an acquisition with a sampling frequency of $8\,kHz$.

After acquisition, a classical instantaneous frequency scheme is applied (band pass filtering, analytic signal extraction, differentiation). A simple addition of F_{mod}/N_{ppt} is required to compensate the frequency shift of F_{mod}.

It should be noted that in real application the optical encoder frequency F_{cod} should be higher (for example $102.4\,kHz$ at $1500\,rpm$ and a resolution of 4096 lines).

This example helps to view some limitation of the method.

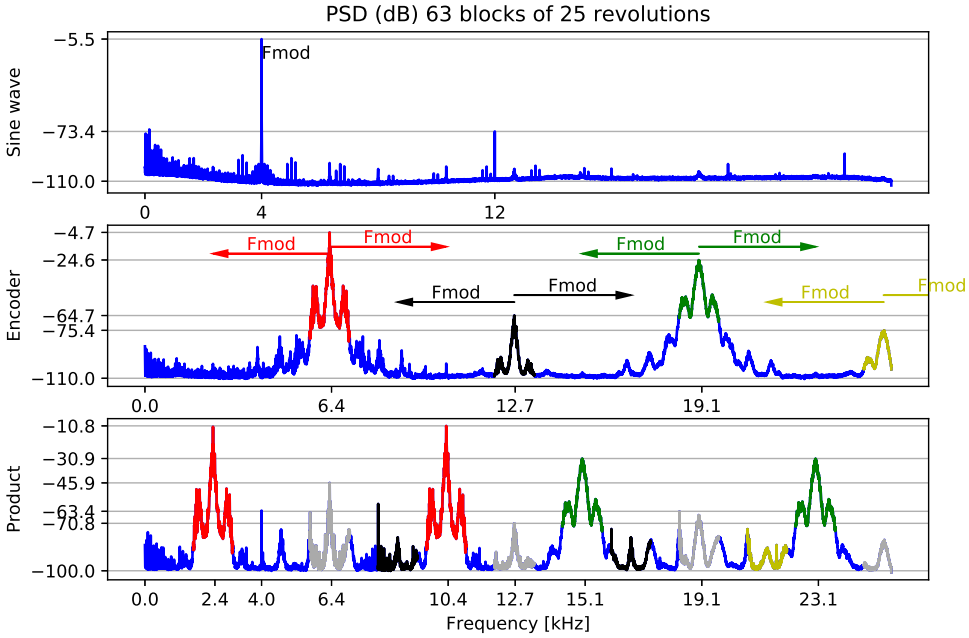


Figure 2: Illustration of High frequency Demodulation Technique.

3 Limit : Frequency overlapping

Let be β the highest fluctuation frequency in Hz (like in the previous section). By looking at the bottom spectrum of figure 2, it is possible to identify the following content :

- $[f_{cod} - f_{mod} - \beta, f_{cod} - f_{mod} + \beta]$: left shifted encoder frequency used for instantaneous frequency estimation,
- $[k.f_{cod} \pm f_{mod} - \beta, k.f_{cod} \pm f_{mod} + \beta]$: all left and right shifted encoder frequencies,
- f_{mod} : modulation frequency (should not be here if $\varepsilon_e = 0$),
- $[k.f_{cod} - \beta, k.f_{cod} + \beta]$: encoder frequencies (should not be here if $\varepsilon_s = 0$).

In order to have a good estimation, the left shifted encoder frequency should not overlap with the other area. Otherwise, the analytic signal will no longer be a sine wave and its phase will have a different meaning.

The frequency shifted red pattern should be kept in positive frequency. It leads to :

$$f_{cod} - f_{mod} - \beta > 0 \tag{3}$$
$$f_{mod} < f_{cod} - \beta \tag{4}$$

In order to access the first harmonic of the encoder, the demodulation frequency f_{mod} is lower than f_{cod}.

If ε_e is not 0, no overlapping between the left frequency shifted red pattern and f_{mod} implies that :

$$f_{cod} - f_{mod} + \beta < f_{mod} \tag{5}$$
$$f_{mod} > \frac{f_{cod} + \beta}{2}. \tag{6}$$

If ε_s is not 0, no overlapping between the left red frequency and the original one implies that :

$$f_{cod} - f_{mod} + \beta < f_{cod} - \beta \tag{7}$$
$$f_{mod} > 2\beta \tag{8}$$

If $\varepsilon_s = 0$ and $\varepsilon_e = 0$, no overlapping between frequency left and right shifted version implies that :

$$f_{cod} - f_{mod} + \beta < f_{cod} + f_{mod} - \beta \quad (9)$$

$$f_{mod} > \beta \quad (10)$$

To conclude, the modulation frequency f_{mod} should be lower than $f_{cod} - \beta$ and depending on whether ε_e, ε_s, or no ε are take into account, f_{mod} should be respectively greater than $\frac{f_{cod}+\beta}{2}$, 2β or β.

Another way of thinking should be to find the demodulation frequency f_{modop} that maximises the highest observable fluctuation frequency β :

$$f_{modop} = \frac{2}{3}.f_{cod} \quad , \quad \beta_{max} = \frac{1}{3}.f_{cod} \quad \text{considering } \varepsilon_e \text{ and } \varepsilon_s, \quad (11)$$

$$f_{modop} = \frac{f_{cod}}{2} \quad , \quad \beta_{max} = \frac{f_{cod}}{2} \quad \text{if } \varepsilon_e = \varepsilon_s = 0. \quad (12)$$

If offsets are not compensated, they limit β_{max} to order $N_{ppt}/3$ instead of $N_{ppt}/2$. The table 1 shows examples of demodulator setup.

Rotation frequency (rpm)	1500				3000	6000
Optical encoder resolution	256	512	1024	4096	4096	1024
Optical encoder mean frequency (kHz)	6.4	12.8	25.6	102.4	204.8	102.4
Optimal f_{modop} frequency (kHz)	4.3	8.5	17.0	68.3	136.5	68.3
Theoretical maximum order	85	170	341	1365	1365	341

Table 1: Example of demodulator setup

Since the sampling frequency of the acquisition device should be twice the f_{modop} frequency in this particular case, this table is not compatible with a $51,2\,kHz$ maximum sampling frequency. To adapt it the constraint $f_{mod} > f_{cod} + \beta - f_s/2$ should be added. This constraint leads to :

$$f_{modop} = \max(\frac{2}{3}.f_{cod}, f_{cod} - \frac{f_s}{4}) \quad , \quad \beta_{max} = min(\frac{f_{cod}}{3}, \frac{f_s}{4}) \quad (13)$$

or by dividing β by f_{rot} the maximum order o_{max} is :

$$o_{max} = \min(\frac{N_{ppt}}{3}, \frac{f_s}{4.f_{rot}}). \quad (14)$$

It means that the demodulation frequency f_{mod} should be increased in order to shift more on the left the red pattern in figure 2. In this case, the needed sampling frequency become f_s.

Rotation frequency (rpm)	1500				3000	6000
Optical encoder resolution	256	512	1024	4096	4096	1024
Optical encoder mean frequency (kHz)	6.4	12.8	25.6	102.4	204.8	102.4
Optimal f_{modop} frequency (kHz)	4.3	8.5	17.0	89.6	192	89.6
Theoretical maximum order	85	170	341	512	256	128

Table 2: Example of demodulator setup to use not more than $f_s = 51.2\,kHz$

These adjustments shows that the sampling frequency will limit the bandwidth in order. It should be kept in mind for comparison that theoretical order correspond to half of the equivalent optical encoder resolution.

4 Comparison with elapsed time

In order to compare both methods, the full resolution optical encoder was recorded using elapsed time technique and high frequency demodulation technique. The frequency f_{mod} is fixed to $90\,kHz$ according to table 2.

The figure 3 compares two results for the two approaches. The instantaneous speed obtained with high frequency demodulation was resampled in angular domain and manually shifted in order to be compared to elapsed time speed. The band β is chosen to be $3.f_{mesh}$ (a lower band gives bad results). The comparison is valid only locally for the first graph. The blue curve corresponds to elapsed time, and the red curve to high frequency demodulation.

The two instantaneous speed curves are superposed. The high frequency method is smoother than elapsed time one due to quantification. The variation in the difference of $\pm 0.06\,Hz$ is not so big be compared to the quantification step of $0.03\,Hz$ for elapsed time method.

The spectrum is also close till order 30. After it seems that an additive noise is present in the high frequency demodulated signal : high peaks are still present but not smallest one. The filtering effect due to the band of $3.f_{mesh}$ (order 69) becomes easily visible near order 100.

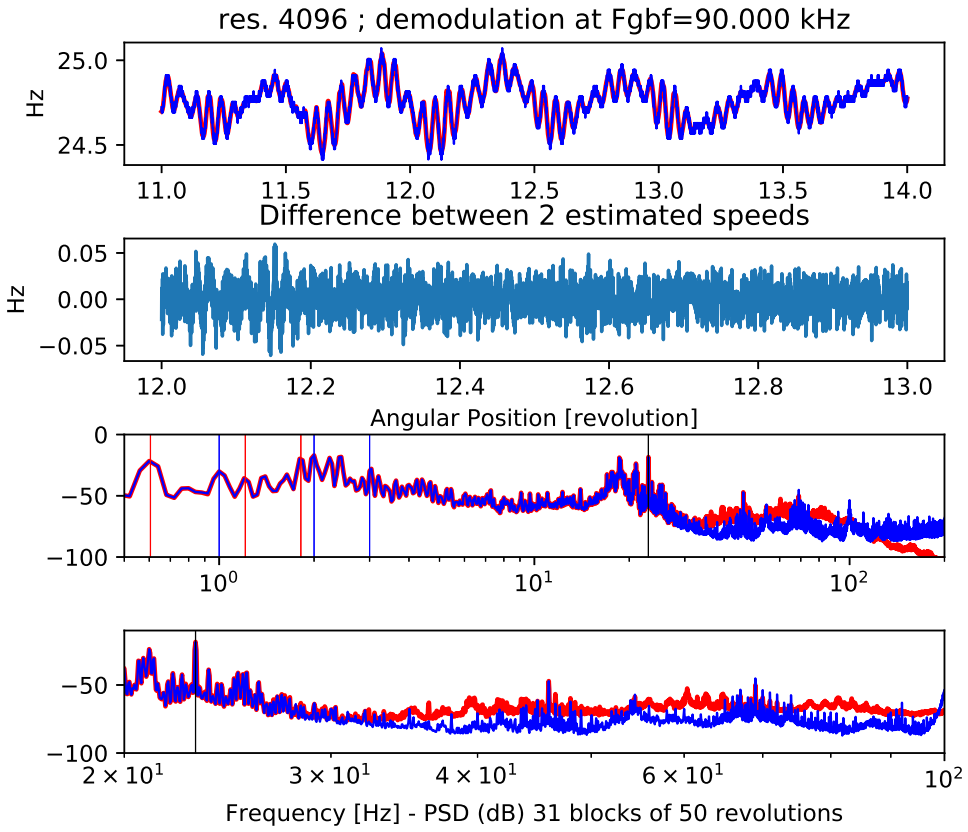


Figure 3: Comparison of Elapsed time and high frequency demodulation Technique.

5 Conclusion

We have proposed a high frequency demodulation technique that performs an analog demodulation prior to acquisition and enable to use the full optical encoder resolution. This technique known in telecommunication as super-heterodyne enables to inspect a restricted bandwidth in high frequency with a low sampling rate. The study of this new technique explains how to choose intermediary frequency for demodulation and some limitations due to the imperfection of electronic components.

A test on a bench to compare this new method with elapsed time shows that the results are similar to the beginning of the spectrum and that an additive noise is present in high frequency. Therefore, this method seems to be interesting when no elapsed time system is present, or it is required to do simultaneous temporal sampling of other signal (acceleration, ...), or, if it is necessary to record very high frequency optical encoder signal.

Acknowledgement

The authors would like to thank Lucien Périsse that design this test bench for us.

References

[1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, and D. Rémond. Natural roller bearing fault detection by angular measurement of true instantaneous angular speed. *Mechanical Systems and Signal Processing*, 24(7):1998 – 2011, 2010. Special Issue: ISMA 2010.

[2] Ming Zhao, Xiaodong Jia, Jing Lin, Yaguo Lei, and Jay Lee. Instantaneous speed jitter detection via encoder signal and its application for the diagnosis of planetary gearbox. *Mechanical Systems and Signal Processing*, 98:16 – 31, 2018.

[3] H. André, F. Girardin, A. Bourdon, J. Antoni, and D. Rémond. Precision of the ias monitoring system based on the elapsed time method in the spectral domain. *Mechanical Systems and Signal Processing*, 44(1):14 – 30, 2014. Special Issue on Instantaneous Angular Speed (IAS) Processing and Angular Applications.

[4] K. Lizoul, H. André, and F. Guillet. Improvement of instantaneous angular speed estimation based on frequency demodulation method under non stationary conditions. In *Proceedings of ISMA2018 and USD2018*, Leuven, 17-19 September 2018.

[5] F. Bonnardot and J. Al Fallah. Non-uniform (re)sampling for optical encoder. In *First World Congress on Condition Monitoring (WCCM)*, London, 13-16 June 2017.

[6] M. El Badaoui and F. Bonnardot. Impact of the non-uniform angular sampling on mechanical signals. *Mechanical Systems and Signal Processing*, 44(1-2):199–210, February 2014.

[7] Heidenhain. *Rotary Encoders brochure*, 2018.

[8] F. Bonnardot, M. El Badaoui, R.B. Randall, J. Danière, and F. Guillet. Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). *Mechanical Systems and Signal Processing*, 19(4):766 – 785, 2005.

Development of a vibration monitoring strategy based on cyclostationary analysis for the predictive maintenance of helicopter gearbox bearings

V. Camerini L. Macchi F. Champavier
G. Naccarato

HUMS department, Airbus Helicopters, 13725 Marignane Cedex, France

Abstract

The scope of this paper is the development of a fault detection and diagnosis method aimed to helicopter gearbox bearings vibration monitoring in an operational context. Bearings are critical components in the gearbox, and their monitoring allows for failure anticipation capabilities, leading to increased safety and improved maintenance planning. Deploying a monitoring strategy for helicopter gearboxes necessitates the development of a methodology which can provide reliable information under varying operating conditions, dealing with a noisy vibration environment and simultaneously considering acquisition system constraints, such as limited acquisition duration and sampling frequency, and operational needs, such as low rate of false alarms and minimal workload for the analyst. The approach proposed in this paper is based on the cyclostationary signals theory and relies on a two-steps procedure of detection and diagnosis. First, bearing fault detection indicators are devised on a statistical basis, leveraging on the theoretical properties of the envelope method. Then, a diagnosis based on the computation of the averaged cyclic periodogram is performed to assess the damage in the eventuality of an alarm. The developed methodology is validated on real helicopter data collected over about twenty thousand flight hours, including four bearings from different machines for which in-service spalling initiation occurred. The fault detection performance is evaluated on the basis of the achieved false alarm rates and the improvement in fault anticipation with respect to chip detectors, whereas the capability of isolating the fault-related signals using cyclostationary signal separation methods is shown for the diagnosis stage.

1 Introduction

Aircraft operations always pose the problem of guaranteeing at any time a compliant level of safety of the machine, achieved through an adequate maintenance plan, without significantly compromising its availability. In helicopters, the drive train sub-system is responsible for transferring power from the engines to the rotors, and represents a critical sub-system for the machine due to non-redundant load paths and the high variability of the dynamic loads acting on the components [1]. As to ensure aircraft airworthiness, the system needs to be maintained following a prescribed preventive maintenance program, resulting in a burden to operating costs and aircraft availability. Searching for an optimal trade-off between keeping the machine operational and reducing safety risk within acceptable hazard levels calls for making as informed as possible decisions. In the last decades, the helicopter industry worked on implementing technical solutions for increasing safety and reducing maintenance costs by enabling Condition Based Maintenance (CBM) [2, 3]. The effort resulted in the widespread adoption of Health and Usage Monitoring Systems (HUMS). Owing to the mechanical degradation of drive train components often resulting in specific vibration symptoms, and considering the widespread availability of vibration measurement systems, helicopter HUMS mostly rely on vibration analysis as a monitoring mean.

A structured breakdown of the failure mechanisms that may affect a helicopter transmission is given in [4], mostly based on [5, 6]. Failure modes may be divided into gear failure modes, bearing failure modes and shaft failure modes. From an operational point of view, the potential of vibration monitoring in driving maintenance operations toward condition based can probably at most be realized by improving the gear and bearing monitoring procedures. Gearbox inspections are expensive and require long-term grounding of the

machine. Therefore, timely detecting an impending mechanical degradation is a great advantage which results in improved maintenance planning and increased machine availability.

In this paper, a two-steps procedure for rolling element bearing fault detection and diagnosis is proposed, with the aim of obtaining a reliable operational procedure able to cope with the monitoring of a fleet of helicopters. First, cyclostationary analysis is recalled as a tool to describe and characterize the characteristic signature of a faulty bearing in section 2. Then, a procedure based on automated statistical bearing fault detection and successive fault diagnosis, and tailored on the specific features of helicopter gearbox vibration environment, is proposed in section 3. The devised strategy is validated in section 4 using data collected from a fleet of operating commercial helicopters. Finally, conclusions are drawn in section 5.

2 Theoretical background

This section is a summary of existing literature on the subject of cyclostationary methods for bearing fault detection, and does not contain original material, except for section 2.4.3.

2.1 Cyclic spectral analysis

Rotating machinery vibration signals have been in the last decades successfully modelled as cyclostationary processes [7–11]. Cyclostationarity is a property characterizing stochastic processes whose statistics vary periodically with respect to some variable (for rotating machinery, typically time or shaft's angular position) [12]. Due to this generality, it is particularly fit to describe rotating machinery signals [13]. The impact forces generated by rolling elements interacting with a local defect on the race are not repeating perfectly periodically due to slippage of the elements in normal operating conditions. Also, the transfer path to the accelerometer varies depending on the relative position of the sensor and the source of the impact. This is the case, for example, for a defect localized on any rotating element in the bearing, where the impact location varies periodically with respect to the transducer position. Such phenomena can be described by their periodic statistics, and therefore the class of cyclostationary signals is suitable to represent the associated excitation. In this paper, the second order cyclostationary descriptors are used to characterize the bearing fault signature. The main quantities of interest when dealing with cyclostationary processes in rotating machinery vibration monitoring are the cyclic spectral correlation (SC) and its normalized version, the cyclic spectral coherence (SCoh). Those quantities are bi-spectral representations, containing information related to the correlation between spectral frequency bands spaced apart by a so-called cyclic frequency α. By considering a signal $y(t)$ recorded in the time T and its Fourier transform $Y_T(f)$, its cyclic spectral correlation can be expressed as:

$$S_{yy}(f,\alpha) = \lim_{T\to\infty} E\left\{ Y_T\left(f+\frac{\alpha}{2}\right) Y_T\left(f-\frac{\alpha}{2}\right)^* \right\} \tag{1}$$

whereas the (squared magnitude) cyclic spectral coherence reads:

$$|\gamma_x(f\alpha)|^2 = \frac{|s_{yy}(f,\alpha)|^2}{s_{yy}(f+(\alpha/2))s_{yy}(f-(\alpha/2))}, \tag{2}$$

A practical estimator of the cyclic spectral correlation can be obtained using the averaged cyclic periodogram method [14]. Based on [14], the averaged cyclic periodogram for the N-length discrete sequence $x[n]$ sampled with sampling frequency F_s, computed using K (possibly overlapping) windows $w[n]$ of length N_w can be computed as:

$$\tilde{S}_{xx}^{(N)}(f;\alpha) = \frac{1}{KF_s\|w\|^2} \sum_{k=0}^{K-1} X_{N_w}^{(k)}\left(f+\frac{\alpha}{2}\right) X_{N_w}^{(k)}\left(f-\frac{\alpha}{2}\right)^*, \tag{3}$$

where:

$$X_{N_w}^{(k)}\left(f\pm\frac{\alpha}{2}\right) = \sum_{n=kR}^{kR+N_w-1} w_k[n]x[n]\mathrm{e}^{\pm j\pi wn/F_s}\mathrm{e}^{-\mathrm{j}2\pi fn/F_s} \tag{4}$$

is the DFT of the kth windowed sequence $w_k[n]x[n]e^{\pm j\pi\alpha n/F_s}$. Practically, the selection of the window length, the window function and the cyclic frequency resolution can be optimized as to minimize the computational time, minimize the cyclic leakage and find the proper trade-off between frequency resolution on the spectral axis f and variance reduction of the estimator [15]. Despite being computationally heavy, the estimator of equation (3) provides reliable results, thanks to its statistical properties, well characterized in [15]. Albeit faster algorithms were developed to estimate the cyclic spectral correlation, e.g. [16], the average cyclic periodogram method still remains a benchmark in terms of estimation accuracy and estimation variance properties.

2.2 Envelope analysis

An important relationship that can be exploited for characterizing a second order cyclostationary process is that connecting the cyclic spectral correlation of the process with its envelope spectrum. It holds from [10] that marginalizing the spectral correlation on the cyclic frequency axis, by integrating out the spectral frequency yields the squared envelope spectrum of the signal. The envelope spectrum has indeed been used in rotating machinery long before the cyclostationary framework was introduced [17]. However, the work in [10] allows to explain the efficiency of the envelope spectrum as an analysis tool for second order cyclostationary processes, framing the technique in the solid theoretical framework of cyclostationary analysis. Other than simplifying the analysis (albeit at the price of losing information on the spectral frequency distribution of the investigated process), the envelope spectrum can be easily estimated from a digitalized realization of the stochastic process by making use of the discrete Hilbert transform and the Fast Fourier Transform, and it is therefore a computationally very convenient quantity. Being equivalent to the integration of the cyclic coherence along the frequency axis, the squared envelope spectrum (SES) as a function of the cyclic frequency α, can be obtained from the N-samples discrete sequence $x[n]$ sampled with sampling frequency F_s as [14]:

$$\begin{aligned} \mathrm{ICC}_x^{(N)}(\alpha) &\propto \left|\frac{1}{N}\sum_{n=0}^{N-1}|x[n]*g[n]|^2 \mathrm{e}^{-\mathrm{j}2\pi n\alpha/F_s}\right|^2 \\ &= \left|\mathrm{DFT}\left\{|x[n]*g[n]|^2\right\}\right|^2 \\ &= \mathrm{SES}_x^{(N)}(\alpha) \end{aligned} \quad , \tag{5}$$

where the convolution with $g[n]$ accounts for whitening of the signal (necessary to have the power normalization leading to cyclic coherence, in place of cyclic correlation); analytic signal transformation; and band-pass filtering in a band comprised between the frequencies F_1, F_2, normally to be chosen as to filter the signal in a band in which the fault symptoms are prominent with respect to the background vibration and the interfering sources. More recently, the logarithm of the envelope spectrum (LES) for a discrete sequence $x[n]$ with $n = \{1,...,N\}$ was introduced in [18] as:

$$LES_x[\alpha] = \left|\frac{\sum_{n=0}^{N-1}\log\left(x[n]^2\right)\mathrm{e}^{-2\pi jn\alpha/F_s}}{N}\right|^2 , \tag{6}$$

The LES is an interesting quantity to be considered in an automated detection framework, thanks to its advantageous statistical properties demonstrated in [18].

2.3 Statistical tests for cyclostationarity

The problem of detecting the second order cyclostationarity is formulated as the decision between the two alternative hypotheses:

$$\begin{aligned} &\mathrm{H}_0 : \text{"The signal does not contain a CS2 component at the cyclic frequency } \alpha\text{"} \\ &\mathrm{H}_1 : \text{"The signal contains a CS2 component at the cyclic frequency } \alpha\text{."} \end{aligned} \tag{7}$$

2.3.1 Testing the SES for cyclostationarity

A rigorous statistical test for the presence of a cyclostationary component at frequency α was given in [14] and is based on the cyclic coherence. By exploiting the link between the cyclic coherence and the SES, a practical statistical test on the SES can be obtained, with the advantage of allowing to work on a simpler, faster

to compute quantity. Namely, for a discrete signal $x[n]$ of length N, the following result is obtained in [14] by extending the statistical test on the cyclic coherence, and in [19] following a direct analysis of the discrete SES:

$$\text{"Reject } H_0 \text{ if:"} \text{SES}_x^{(N)}(\alpha) \geqslant \frac{\sigma_{x*g}^4}{2N} \frac{F_s}{F_2 - F_1} f(\alpha) \cdot \chi_{1-p,2}^2\text{"}, \tag{8}$$

being p the significance level of the test, σ_{x*g} the standard deviation of the filtered signal $x[n] * g[n]$ and:

$$f(\alpha) = \begin{cases} 1 - |\alpha| / (F_2 - F_1), & |\alpha| < F_2 - F_1 \\ 0 & \text{elsewhere} \end{cases} \tag{9}$$

It is important to underline that the optimality of the test is obtained under the assumption of white noise signal for a healthy component. An analysis of the effects of CS1, CS2 components and colored noise on the SES of the signal is exhaustively performed in [19]. The relevant points are summarized below:

- The effects of a set of M additive multi-harmonic CS1 components of frequencies λ_m, $m = \{1, ..., M\}$ are that of biasing the SES at the difference frequencies $\{\Delta\lambda\} = \{\lambda_m - \lambda_n\}$, $m, n = \{1, ..., M\}$; and that of amplifying the variance in large frequency bands.
- The effect of an additive CS2 component in the signal is that of introducing a bias in the estimator of the SES, which is stronger when the average power of the CS2 carrier is dominating over the background noise.
- The generalization to colored noise implies estimating the variance of the signal at each frequency bin, resulting in a statistical threshold which is no longer a linear function of the frequency.

2.3.2 Testing the LES for cyclostationarity

For a white noise, discrete signal $x[n]$ of length N, the distribution of the LES at a cyclic frequency α is given in [18] as:

$$\frac{\text{LES}_x[\alpha]}{\pi^2/4N} \sim \chi_2^2 \tag{10}$$

Therefore, the LES test for cyclostationarity at significance level p reads:

$$\text{"Reject } H_0 \text{ if:"} \text{LES}_x^{(N)}(\alpha) \geqslant \frac{\pi^2}{4N} \cdot \chi_{1-p,2}^2\text{"}, \tag{11}$$

The LES allows estimating the CS2 components in the signal with the following advantages with respect to the SES:

- The estimator is unbiased by the presence of cyclic components of a frequency different than the considered one.
- Under the white noise assumption, the variance of the estimator is independent from the variance of the noise in the signal.

The second point is more a matter of mathematical rigor for long, whitened noise signals, whereas the first point constitutes an important advantage of the LES when it comes to defining automated tests for the presence of cyclostationary components at a frequency of interest. Additionally, the LES was shown to yield better statistical performance in presence of impulsive noise [20]. This last characteristic is not surprising in light of its being unbiased from CS2 components, considering the existing relation between the CS2 components in the signal and its Kurtosis [21].

2.4 Signal pre-processing

In order to leverage on the optimality of the statistical tests discussed in section 2.3, it is necessary to bring the analyzed signal's statistics as close as possible to the white noise conditions. First, it is a good practice to remove CS1 components from the signal, as they have a biasing effect as explained in section 2.3. The removal of CS1 components can be performed, e.g., through estimation and subtraction. Such an estimation can be performed in different ways, depending on whether the fundamental cycle of interest is known or not. In the case it is not, it can be based on blind estimators, as the linear adaptive enhancer (ALE), or the self-adaptive noise canceller (SANC) and its more efficient frequency domain formulation [22–25]. Generally, blind estimators performance is negatively affected by signal to noise ratio. Moreover, blind filters require a proper parameter tuning which may not be trivial in every case. When the cycle of the signal is known, a popular estimator of the periodic mean is the synchronous average (SA) operator, which is also known as Time Synchronous Average (TSA) due to its original formulation in time domain [26]. In order to obtain the periodic mean in the case of a quasi-cyclostationary signal, the SA must be applied for each of the fundamental cycles which are present in the signal, and then the extracted periodic components need to be summed together [13]. Once the deterministic part of the signal is removed, it is necessary to obtain a flat frequency spectrum for the signal, resembling white noise statistics. In order to do so, there are mainly two strategy: one is selecting a narrow-band frequency region and filter it out; the other is to apply any method to "flatten" the spectrum, such as cepstrum pre-whitening (CPW) [27–29]. On the other hand, the estimation bias resulting from the exogenous CS2 components discussed in section 2.3 cannot be simply corrected for, due to its statistical nature. As a summary, two main steps shall be performed before analyzing the signal, i.e. removal of CS1 components and pre-whitening of the residual. In this work, two techniques were found particularly useful for the scope: the angular domain synchronous average and the cepstrum pre-whitening. The first technique is preferred as the cycles of the main additive deterministic components are known for a given gearbox, whereas the second one is preferred over filtering, as it allows to consider the full-band signal in the analysis, avoiding a further optimization step to select a narrow-band filter which is able of isolating the fault signature (e.g., Spectral Kurtosis [30] is a popular tool that can be used for the scope). Also, if compared to other pre-whitening techniques, the CPW excels for the simplicity of use and the lack of configuration parameters to be properly selected.

2.4.1 Synchronous average removal

Under the assumption of cycloergodicity [13], SA is indeed a practical estimation of the periodic mean of a CS signal, which is its first order cyclostationary part of cycle equal to the fundamental period used for averaging. The equation for the SA of a signal $x(\theta)$ of fundamental cycle Θ reads in angle domain [31]:

$$SA[x(\theta)]_\Theta = \frac{1}{N}\sum_{i=0}^{N-1} x(\theta + i\Theta) \tag{12}$$

Synchronous averaging is thus equivalent to applying a comb filter to the signal [13], which extracts the multiples of the reference harmonic. The number of averages controls the bandwidth of the lobes, the amount of noise rejection and the position of the notches of the filter. In order to remove multiple cycles linked to different harmonic families, equation (12) can be applied multiple times to extract the harmonic family of interest and then subtract it from the original signal. It is worth mentioning that an "order tracking" or "angular resampling" step has to be performed to correct for small speed fluctuations, expressing therefore the measured vibration signal in the angle-domain form of equation (12) [27]. This angular resampling step is typically performed using a synchronization signal acquired from an external measurement system, as e.g. a magnetic pick-up sensor mounted on a reference shaft [31].

2.4.2 Cepstrum pre-whitening

The pre-whitening operation consists in setting a zero value for the whole real cepstrum (except possibly at zero quefrency), then, once transformed back to the frequency domain, the obtained signal is recombined with the phase of the original signal and inverse transformed to time domain [29]. Considering a signal x, its

pre-whitened version can be simply computed as [29]:

$$x_{cpw} = IFT\left\{\frac{FT(x)}{|FT(x)|}\right\} \tag{13}$$

2.4.3 Remarks on signal pre-processing

The two steps of signal pre-processing carry with them some hidden difficulties which is worth pointing out. First, it is necessary to observe that the synchronous average removal requires angular resampling of the signal, and when it is performed using the computed order tracking (COT), it involves interpolating the signal in order to obtain its angle-domain values. In [32], a discussion of interpolation methods is given. Interpolating acts as a low-pass filter in the frequency domain. Therefore, it is important to keep into account that any order tracking step has the effect of distorting the signal's spectrum by attenuating the high-frequency components. As a consequence, spectral flattening shall always be performed after order tracking, when envisaging the use of the statistical tests of section 2.3. Furthermore, the cepstrum pre-whitening enhances the sensitivity of the squared envelope results to phase correlation. This last issue is formalized below by using a simple, unit amplitude complex harmonic signal as an example. Consider the following:

$$x(t) = e^{j2\pi\tilde{f}t} \tag{14}$$

then its Fourier transform reads:

$$X(f) = \delta(f-\tilde{f}) \tag{15}$$

being $x(t)$ complex analytic, and noting $\overline{x}(t)$ its complex conjugate, then the envelope spectrum can be computed as:

$$\mathrm{ENV}_x(f) = FT(x(t)\cdot\overline{x}(t)) \tag{16}$$

if a finite-length, discrete signal $y[n] = x[n]\cdot w[n]$, $n = \{1,...,N\}$ is considered (neglecting the sampling step from the notation for simplicity), being $w[n]$ a rectangular, causal observation window of length N, and the Discrete Fourier Transform is used to compute the Fourier transform, it holds:

$$\mathrm{ENV}_y[k] = \frac{1}{N}\sum_{n=0}^{N-1} y[n]\cdot\overline{y}[n]\cdot\exp\left(-j2\pi\frac{nk}{N}\right) \tag{17}$$

then by the convolution theorem, equation (17) can be rewritten as:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} Y[r]\cdot\overline{Y}[r-k] \tag{18}$$

again, for the convolution theorem, the sifting property of the Dirac's delta distribution $\delta(\cdot)$, and considering the Fourier transform of the rectangular window $w[n]$, equation (18) can be expressed as:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} \frac{\sin(\pi(r-\tilde{k}))}{\sin(\pi(r-\tilde{k})/N)} e^{-j\pi(r-\tilde{k})\frac{N-1}{N}} \cdot \frac{\sin(\pi(r-k-\tilde{k}))}{\sin(\pi(r-k-\tilde{k})/N)} e^{j\pi(r-k-\tilde{k})\frac{N-1}{N}} \tag{19}$$

where $\tilde{k}$ denotes the discrete frequency index corresponding to $\tilde{f}$. After the cepstrum pre-whitening operation of equation (13), only the phase terms are left from the DFT of the original signal and equation (19) becomes:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} e^{-j\pi(r-\tilde{k})\frac{N-1}{N}}\cdot e^{j\pi(r-k-\tilde{k})\frac{N-1}{N}} = e^{-j\pi k\frac{N-1}{N}}\left(\frac{N}{2}-k\right) \tag{20}$$

From equation (20), it can be observed that the phase correlation from residual periodic components has a harmful effect on the classical envelope spectrum of the pre-whitened signal, which is amplified when the amplitude spectrum of the signal is equalized to the unit value.

2.5 Bearing fault signature

Healthy bearings vibration does not typically bring a significant contribution to the vibration generated by a helicopter gearbox. On the other hand, a defective bearing generates a characteristic vibration signature, characterized by repeated impacts occurring each time that a bearing element contacts the defective surfaces [5, 33–35]. Typically, four characteristic frequencies can be identified: ball pass frequencies on the outer and inner races (respectively BPFO and BPFI), typically linked to localized defects on one of the races; fundamental train frequency (FTF), generally linked to cage defects; and ball spin frequency (BSF), normally related to localized defects on the rolling elements surface. By indicating with f_i and f_e respectively the inner and outer race rotation frequency, with N_b the number of rolling elements in the bearing, by α_0 the initial contact angle, by d the rolling element diameter and by D the bearing pitch diameter, these characteristic frequencies read [36]:

$$\begin{aligned} BPFI &= \frac{N_b\,|f_e - f_i|}{2}\left(1+\frac{d}{D}\cos(\alpha_0)\right) \\ BPFO &= \frac{N_b\,|f_e - f_i|}{2}\left(1-\frac{d}{D}\cos(\alpha_0)\right) \\ BSF &= \frac{D\,|f_e - f_i|}{2d}\left(1-\left(\frac{d}{D}\cos(\alpha_0)\right)^2\right) \\ FTF &= \frac{1}{2}\left(f_e\left(1+\frac{d}{D}\cos(\alpha_0)\right)+f_i\left(1-\frac{d}{D}\cos(\alpha_0)\right)\right) \end{aligned} \tag{21}$$

Letting $h_j(t)$ be the impulse response to a single impact measured by the sensor located at position j, $q(t)$ the periodic modulation owing to load distribution (or periodic changes in the loading conditions, or sensor orientation/position with respect to the impact point) [37, 38], and letting T be the fundamental impact periodicity (which can be computed by inverting the frequency of interest from equation (21)); then the measured response $x_j(t)$ related to the defective bearing was given in [14, 39–41] as:

$$x_j(t) = \sum_{i=-\infty}^{+\infty} h_j(t - iT - \tau_i)\, q(iT) A_i + n_j(t), \tag{22}$$

where $n_j(t)$ includes the additive background noise and all eventual interference sources, the subscript i indicates the ith impact, τ_i represents the mentioned uncertainty on pulse arrival time and A_i the random amplitude of the impact. Both the variables are modeled in [14] as mutually independent, white, stationary random sequences with respectively zero and unity mean. Those idealized assumptions allow, according to the literature, to gain sufficient insight into the described phenomenon. The fault signature appears as a pseudo-periodic excitation consisting of pulses which are separated by a period close to that of the fault frequency, but affected by a small, random variation typically of the order of one percent of the fundamental period. Such slight fluctuations results practically in destroying the discrete, harmonic structure that would arise if the random fluctuations were neglected as in [34], giving raise to an essentially random vibration signal in the frequency range of interest [14]. The main difference between the model found in [34] and that of equation (22) is that in the latter, the harmonic structure produced by the fault-related impacts rapidly turns into a random signal. As a consequence, the bearing fault signature in the spectrum is likely to be localized in the low-frequency region, and therefore subject to masking from the background noise and from other possibly existing interference sources, as e.g. gear mesh harmonics. Hence, the model can explain the reason for which classical spectral analysis may fail in detecting rolling-element bearing faults, making this signal representation closer to the reality of the phenomenon. A practical solution to this issue resides in making use of the second-order cyclostationary tools presented in section 2.1 in order to isolate the bearing vibration signature from the rest of the measured signal. The cyclostationary approach was actually shown very successful for bearing diagnostic problems in several works, as in [10, 14, 19, 21, 33, 41]. From equation (1), adopting the proper normalization to obtain the cyclic power spectrum from the spectral correlation, it can be shown that for the signal model of

equation (22), it holds [10]:

$$S_{x_j}(f,\alpha) \simeq \frac{1}{T} H_j\left(f+\frac{\alpha}{2}\right) H_j\left(f-\frac{\alpha}{2}\right)^* \left(\Phi(\alpha)\left(1+\sigma_A^2\right) - \Phi\left(f+\frac{\alpha}{2}\right)\Phi\left(f-\frac{\alpha}{2}\right)^*\right) \times \sum_{k,l=-\infty}^{+\infty} Q_l \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right] + \delta[\alpha] S_n(f) \quad , \tag{23}$$

in which $\Phi(f)$ stands for the Fourier transform of the probability density function of the random variable τ associated to the impacts jitter, P for the load variation characteristic period, $H_j(f)$ for the transfer function from the impact point to the measurement location j, obtained as the Fourier transform of the impulse response function $h_j(f)$, σ_A the standard deviation of the random variable A representing the random impact amplitude, and the Q_l are coefficients of the Fourier transform of the modulating function $q(t)$. The weak harmonic contribution was neglected, being it highly attenuated in the high frequency region as an effect of the random jitter of the impact times. From equation (23), the discrete structure of the bearing signature is finally evident in the cyclic power spectrum plane, with continuously distributed values along the spectral frequency lines appearing at multiples of the fundamental impact frequency (along axis α). Also, the values are higher for those spectral frequency bands where the fault signature is dominating. It follows for the cyclic coherence [14]:

$$\left|\gamma_{x_j}(f,\alpha)\right|^2 \simeq \left|\frac{\mathrm{SNR}(f)}{1+\mathrm{SNR}(f)}\right|^2 |\Phi(\alpha)|^2 \sum_{k,l=-\infty}^{+\infty} \left|\frac{Q_l}{Q_0}\right|^2 \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right], \tag{24}$$

where $SNR(f)$ represents the signal-to-noise ratio of the fault. Consequently, as a function of the spectral frequency f, equation (24) shows an increased coherence for increasing amplitude modulation randomness, impact frequency and load modulation intensity, whereas as a function of the cyclic frequency α, it shows a discrete structure consisting of harmonics of the fault signature separated by the characteristic impact frequency, with decreasing amplitude depending on the low-pass filtering function $\Phi(\alpha)$. Therefore, the highest the multiple of the fundamental impact frequency, the lower the intensity of the observed bearing signature in the cyclic coherence. Equation (3), along with the proper normalization, can be used to estimate the quantity appearing at the LHS of equation (24), yielding an efficient diagnostic representation able of highlighting the bearing fault signature according to the structure of equation (24).

3 Proposed monitoring procedure

Operationally, it is desirable to control the risk of false alarms from the health monitoring system, in addition to providing the earliest possible warning. In order to achieve those targets, a two-steps procedure is proposed in this work. First, statistical indicators leveraging on the cyclostationary theory are designed in order to attain a specified false alarm rate. Secondly, a diagnostic step based on the analysis of the cyclic coherence is taken each time that an alarm is raised, in order to confirm that the threshold exceedance is actually due to a mechanical defect. The importance of the first step of the procedure is that of providing an easy-to-read scalar indicator, with known statistical behavior, which can be employed to guarantee a given false alarm rate as low as not to overload the analysts, maintaining contextually an acceptable detection performance. The second step implies confirming the alarms raised in the detection phase before performing any maintenance, as a measure to avoid unnecessary grounding of the concerned helicopters.

3.1 Fault detection stage

With the aim of deriving statistically reliable monitoring indicators, the cyclostationary signal theory is adopted in this work within the frame of a procedure similar to that proposed in [42]. The envelope spectrum is here computed through the methods described in section 2.2. Both SES and LES can be considered. The necessity of an automated bearing monitoring gives raise to the following challenges:

1. The false alarm rate shall be kept under control in order to avoid unnecessary grounding of the machine;
2. The actual fault frequency cannot be accurately predicted using the simplified kinematics relations;
3. Interfering, exogenous components may mask the bearing fault signature.

As to cope with the first point, the statistical tests presented in section 2.3 for LES and SES can be exploited. First, a signal pre-whitening step consisting of the removal of the periodic (CS1) components through synchronous averaging followed by cepstrum pre-whitening for spectral flattening is performed. This allows to get rid of interferring CS1 sources and to bring the signal's statistics closer to those of white noise. A statistical threshold can then be derived from the white-noise envelope spectrum statistics as in [18, 43]. For a number of bearing fault harmonics to be configured, a narrow-band frequency range around the fault frequencies of an extent to be configured, can be defined. This is done in order to allow some margin in considering eventual shifts of the fault frequency from its predicted nominal value, addressing the second problem in automatizing the algorithm. At the same time, if the considered cyclic frequency range is too wide, there is the risk of exogenous CS2 components leaking into the analysis band, leading to incorrect diagnosis. This issue has to be carefully addressed when tuning the algorithm's parameters. Any number of fault harmonics can be considered in the algorithm. However, according to equation (24), best results are obtained for the low bearing harmonics. For each defined range, the values of the SES or LES are compared to the statistical thresholds of equations (8) and (11). If any statistically significant value is present, a fault detection alarm is raised. The advantages of the adopted procedure are two-fold: on one hand, the envelope spectrum is computed in a computationally efficient way; on the other hand, the computation of the theoretical threshold after pre-whitening according to [18, 43] provides solid grounds for statistical testing. For a given signal to be processed, the algorithm can be summarized in the following steps:

1. Define the desired false alarm rate $P\tilde{F}A$ for the indicator according to the operational needs;
2. Calculate the fault frequency of interest F_F (according to equation (21)) in units of the sampling frequency F_s;
3. Set the number of fault harmonics N_h to be monitored;
4. Set a tolerance band ψ as a percentage of the fault frequency of interest, in order to account for the uncertainty on the actual fault frequency;
5. Remove known CS1 components using synchronous average removal;
6. Apply spectral flattening using cepstrum pre-whitening;
7. Compute the full-band SES/LES of the pre-whitened signal;
8. Compute the statistical threshold p according to the defined desired false alarm rate, based on the white noise assumption;
9. For each considered fault harmonic: find the maximum value of SES/LES in the defined tolerance band;
10. Compute the indicator value as the mean of the statistically significant values with respect to the defined threshold (if no significant value is found for any of the considered harmonics of the fault, the indicator value is set to zero).

The expected false alarm rate $P\tilde{F}A$ corresponds to the probability of one value of the (squared or logarithmic) envelope spectrum within the considered range being higher than its statistical threshold computed through equations (8) and (11). Therefore, to attain a desired level for $P\tilde{F}A$, it is necessary to calculate the significance level p as to satisfy:

$$P\tilde{F}A = 1 - p^{\sum_{h=1}^{N_h} r_h} \tag{25}$$

with N_h being the number of the considered fault harmonics and:

$$r_h = \lceil h \frac{\psi}{100} \frac{F_F}{F_s} N \rceil \tag{26}$$

where N is the discrete signal length and F_s its sampling frequency expressed in inverse units of the sampling step. The value of p obtained from equation (25) allows for calculating the statistical thresholds of equations (8) and (11) so that the probability of having one exceedance of the envelope spectrum in the computational range is equal to $P\tilde{F}A$. It is important to point out that where the false alarm rate can be kept under control, the detection performance cannot be predicted a priori. Typically, the higher the tolerance band ψF_f, the lower the detection performance for a fixed $P\tilde{F}A$; the higher the admissible alarm rate $P\tilde{F}A$, the higher the detection performance that can be expected.

3.2 Fault diagnosis stage

Each time an alarm is triggered, a fault diagnosis step is performed by the analyst. It consists of computing the cyclic spectral coherence around the fault frequencies for which the detection algorithm raised an alert and visually assessing the existence of CS2 components compatible with the expected signature from a bearing fault. The spectral coherence is estimated according to equation (3), where the number of averages and the window length are set according to the guidelines discussed in [15]. Within this assessment step, the analyst can additionally assess the presence of diagnostic side-bands carrying supplementary information on the nature of the fault.

4 Results

In this section, the proposed monitoring strategy is applied on in-service helicopter HUMS data in order to assess its performance in terms of reliability and detection. LES and SES indicators, along with different pre-processing treatments are compared, stressing the importance of properly pre-whitening the signal before carrying on with the analysis.

4.1 Data description

In order to validate the proposed procedure, a comprehensive data-set consisting of vibration data recorded from fourteen machines over about twenty-thousand flight hours (FH) is considered. The data-set includes four bearing in-service degradation cases that were detected by the HUMS: two of them concern roller bearings, and the other two concern ball bearings. The acquisitions were performed in various operating conditions, involving different regimes for the rotational speed of the rotor and for the transmitted torque from the engines. Main gearbox (MGB) and accessory gearbox (AGB) acquisitions comprise signals from seven accelerometers and two keyphasor signals. The two keyphasor signals provide respectively one pulse per revolution of the main rotor and of the tail rotor shaft. Accelerometers are typically mounted on the gearbox casing, close to the monitored components. Generally, acquisitions are divided in groups. Each acquisition group is launched when specific flight conditions are matched and consists of a synchronized acquisition from a set of sensors, performed with a configured sampling frequency for a configured duration. The available keyphasor signals are always sampled with the same sampling frequency of the accelerometer signals. Signals were sampled synchronously from all the MGB accelerometers with a sampling frequency of 50 kHz, for one second duration. The four documented bearing fault cases occurred on different machines during the monitoring period. The faults were anticipated by the HUMS in the spalling initiation phase, allowing for timely maintenance. Table 1 summarizes the four selected fault cases, whereas figure 1 show the inspected bearings after component removal. All the cases involve outer race spalling which occurred whether on a roller or on a ball bearings. The HUMS, through the deployed monitoring strategy, triggered regular inspections of the chip detectors and allowed in each case to anticipate the chip warning coming from the oil metal chips detectors. Additionally, for the fault case 4, HUMS had gained enough confidence to trigger the removal without the need of waiting for the metal particles in the chip detector to be out of criteria. The geometrical parameters of the concerned bearings are not reported for proprietary reasons.

Table 1 – Fleet selected bearing fault cases summary

Fault case ID	Machine ID	Damaged bearing
1	1	Roller bearing (bearing 1)
2	2	Roller bearing (bearing 1)
3	3	Ball bearing (bearing 2)
4	3	Ball bearing (bearing 2)

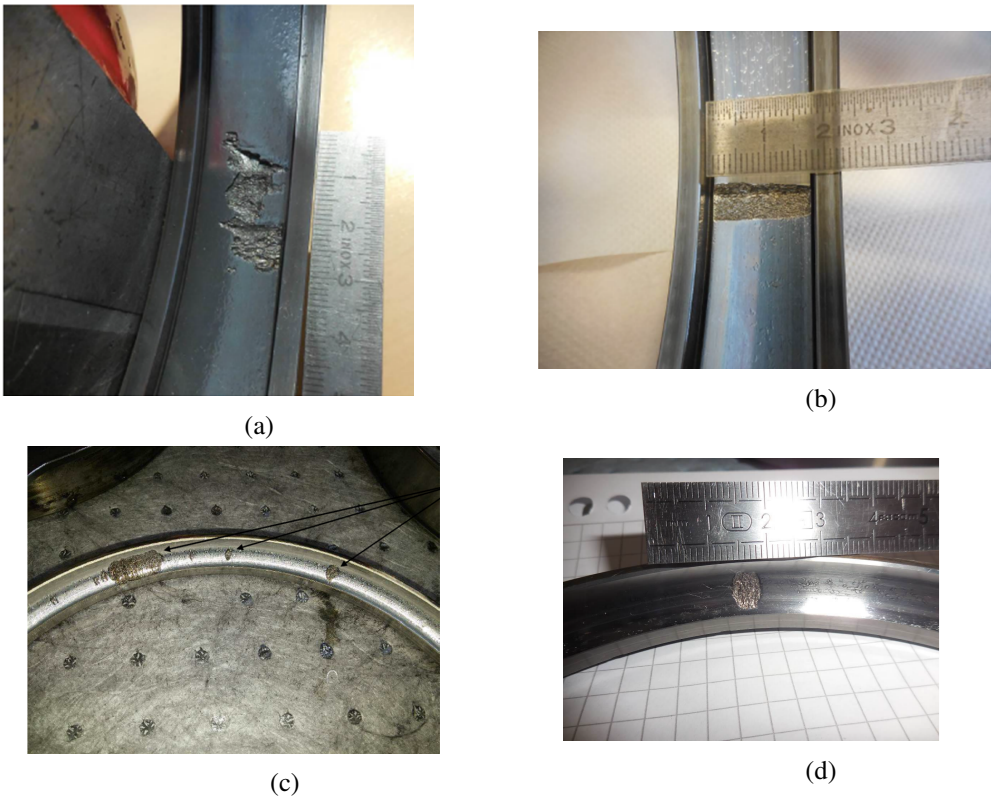


Figure 1 – Damaged bearings after inspection. Outer race presenting spalled area at the time of removal – a) Bearing fault case 1 ; b) Bearing fault case 2; c) Bearing fault case 3 ; d) Bearing fault case 4.

4.2 Fault detection performance

The bearing monitoring strategy proposed in section 3 is applied on the operational data of this section. Bearing anomaly indicators are computed routinely on all the fleet data, and then the detected exceedances are analyzed using cyclostationary analysis in order to complete the diagnosis. Although bearing anomaly indicators are computed for each monitored bearing in the gearbox, the results here presented are restricted to those for which the faults were observed, allowing to validate both detection and the reliability performance. From the bearing geometrical properties, the theoretical characteristic frequencies could be computed according to equation (21). Table 2 reports the computed, nominal bearing defect frequencies for the considered bearings 1 and 2, expressed in orders of the rotational speed of the shaft to which they are attached. According to the nominal design parameters, the fault frequencies are very close (for the outer race defect frequency, the difference between bearing 1 and 2 is less than two percent), posing a challenge in discriminating which one is the faulty bearing in the event of a detection. Both SES-based and LES-based indicators are evaluated, computing the envelope spectra on the pre-whitened signal after a synchronous average removal step and cepstrum pre-whitening (SES-CPW and LES-CPW indicators). Performance on the signal after the synchronous average removal step only are also reported for comparison (SES-SA and LES-SA indicators). The indicators were configured such as to allow to separate the two fault frequencies, but allowing some slippage through one-percent width analysis bands. Only the first harmonic of the outer race fault frequency was considered. With reference to section 3, care was taken when performing the OT steps in the SA removal and when resampling the signal to order domain before the cepstum pre-whitening.

Table 2 – Bearing 1 and 2 characteristic fault frequencies in order domain.

Parameter	Description [dimension]	Bearing 1 Value	Bearing 2 Value
BSF	Roller Spin Frequency [Hz]	3.64	3.67
FTF	Fundamental Train Frequency [Hz]	0.44	0.45
$BPFO$	Ball Pass Frequency Outer race [Hz]	9.69	9.85
$BPFI$	Ball Pass Frequency Inner race [Hz]	12.31	12.15

Figure 2 shows the actual false alarm rate against the expected one for the SES and LES indicators for bearing 1 and bearing 2 on the whole collected healthy fleet data. It can be seen that the actual alarm rates for both the bearings do not match well the expected results in the case of the SA indicators. This is attributable to two main factors: first, the SA procedures do not remove all of the periodic components originally present in the signal, leaving some residual, biasing CS1 component; secondly, the spectrum of the measured vibration is far from resembling white noise, leading to the statistical thresholds computed using equations (8) and (11) being inaccurate. On the other hand, figures 2c and 2d show better results for the indicators computed after the cepstrum pre-whitening step. From the distribution check it can be seen that the LES-based indicator provides very accurate results for both the considered bearings: the actual false alarm rate agrees very well with the predicted one. Conversely, despite the spectral flattening and the SA removal, the SES indicator yields higher false alarm rate than expected. This fact might be explained by the SES statistics being affected by exogenous CS2 components, differently from those of the LES. In order to evaluate the detection performance of the devised statistical indicators, an anticipation over removal metrics is introduced. For each fault case, the removal anticipation (RA) achieved by HUMS alert and expressed in acquisitions number, was computed as a function of the actual false alarm rate on the fleet. Typically, the removal is triggered by an inspection of the particles captured in the chip detector and matching some criteria on size and composition. The RA figure allows to judge the achievable trade-off between detection capability and global false alarm rate performance of an indicator. Figures 3a and 3b show the results respectively for the fault cases 1 and 2, for the SES-SA and LES-SA outer race fault indicators, whereas figures 4a and 4b report the same results for the SES-CPW and LES-CPW outer race fault indicators. Concerning the fault cases 3 and 4, related to the ball bearing, the RA is constant for each considered indicator, despite pre-processing differences and equal respectively to 66 acquisitions and 34 acquisitions. This depends actually on two distinct facts: first, in the considered data-set, acquisitions for machine 3 begins already in a relatively advanced bearing degradation stage, where all indicators detect very clearly. Secondly, the second degradation produced very early, strong CS2 symptoms which were as well detectable in a robust manner from all of the considered indicators. Also, for the second ball bearing degradation (fault case 4), the HUMS alerting system was already deployed, and guaranteed the detection of the incipient bearing degradation, along with an optimized, planned maintenance intervention. Therefore, the time to removal metrics is not representative of the anticipation over the chip detector alert for case 4. For the roller bearing degradation cases, evidently the SES-CPW and LES-CPW achieve a better anticipation over removal for a given false alarm rate, implying their better performance in terms of early detection with respect to the SES-SA and LES-SA indicators in both the cases. The comparison between LES-CPW and SES-CPW shows that they perform very similarly for the first fault case (figure 4a), whereas the SES-CPW indicator shows better detection performance with respect to the LES-CPW in the second fault case (figure 4b).

4.3 Fault diagnosis stage

In this section, the diagnostic charts based on the cyclic coherence are shown for each detected bearing degradation. The defined operational procedure only requires to compute such quantities when an alert is raised, so to confirm the actual occurrence of a mechanical degradation. In figures 5 to 8, for fault cases respectively 1 to 4, the diagnostic charts and the associated LES and SES spectra, together with their statistical threshold computed for the 0.1 percent significance level according to equations (8) and (11), are shown in three conditions: before the beginning of the bearing degradation, during the bearing degradation and after gearbox replacement. In each case, the spalling manifests as a high-frequency excitation, at a cyclic order which is slightly different than that predicted from the theoretical calculations. Consequently, the actual loading conditions encountered in operations have an impact on the exact determination of the fault frequency, owing to the simplified kinematics assumptions being inadequate to describe the bearing dynamic behavior. By comparing figure 5 to figure 6, it can be seen that the symptoms of the spalling appears much more evident for the second fault case (compare, e.g., the statistical threshold to the value of the emerging peak in the two cases). Coherently, figures 3 and 4 show that the second fault case is predicted with a higher anticipation time from both the SES and LES indicators. This can be explained by looking at figure 1. It can be noticed from figure 1a and figure 1b that the shape of the surface degradation is consistently different in the two cases. In fault case number 2, the spalling area extends across the full span of the race, creating a slot. Conversely, in fault case number 1 the spalling area is restricted to part of the width of the race. It can be expected that for the second degradation

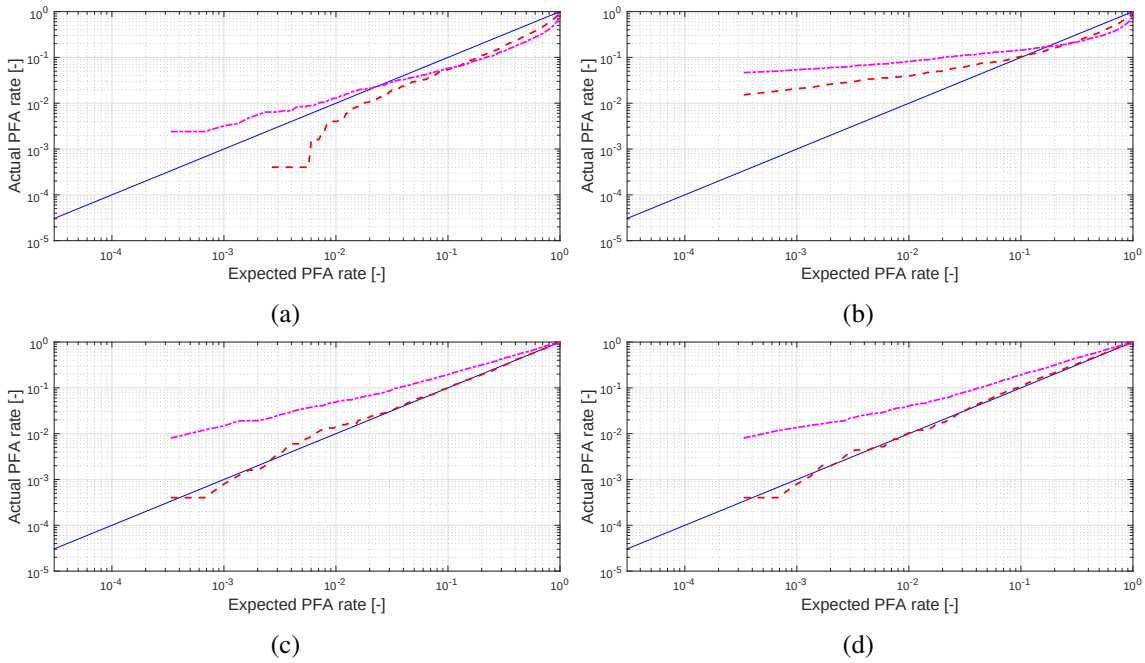


Figure 2 – Healthy fleet data, actual vs. expected probability of false alarm rate of the SES and LES outer race fault detection indicators for: a) Bearing 1, SA indicators; b) Bearing 2, SA indicators; c) Bearing 1, CPW indicators; d) Bearing 2, CPW indicators. Blue: theoretical relation; red dashed: LES indicator; magenta dot-dashed: SES indicator.

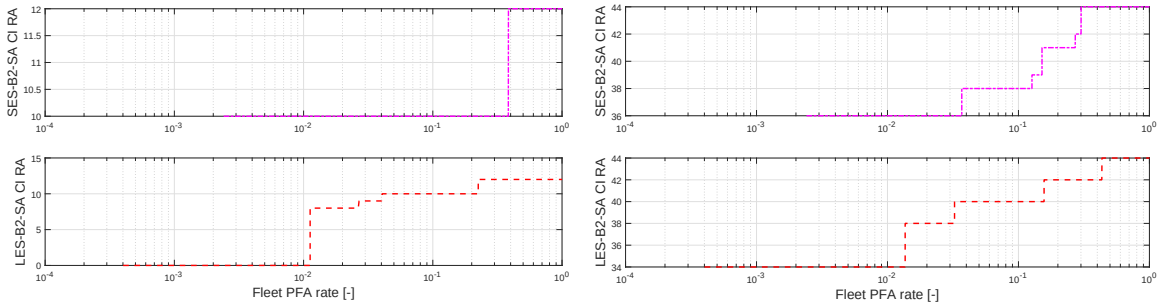


Figure 3 – Removal anticipation (RA) time in acquisitions vs. fleet false alarm rate for SES-SA (upper row, magenta dot-dashed) ad LES-SA (lower row, red dashed) indicators – a) Fault case number 1; b) Fault case number 2.

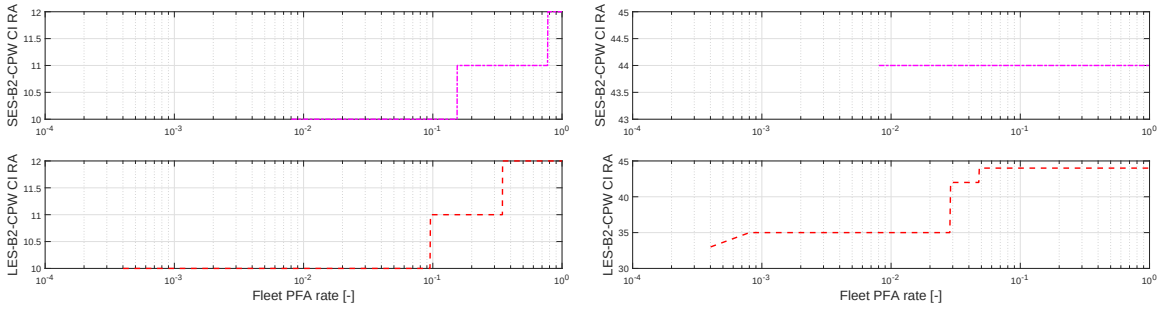


Figure 4 – Removal anticipation (RA) time in acquisitions vs. fleet false alarm rate for SES-CPW (upper row, magenta dot-dashed) ad LES-CPW (lower row, red dashed) indicators – a) Fault case number 1; b) Fault case number 2.

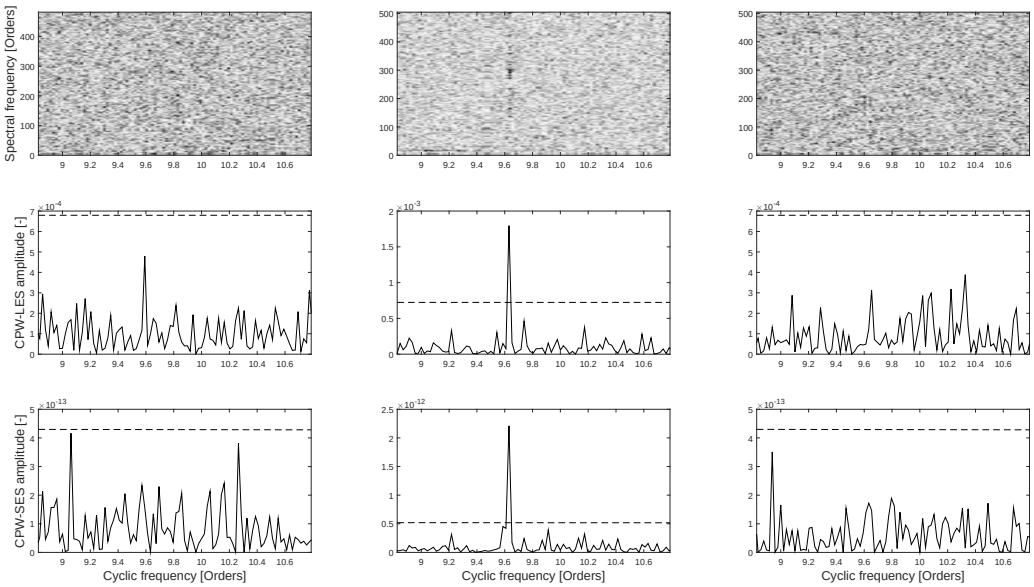


Figure 5 – Diagnostic charts for fault case 1. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

case, impacts will occur, exciting the bearing resonances and resulting in stronger CS2 symptoms with respect to those produced in case number 1, where the quasi-periodic change in the dynamic response is more likely due to the change of the load distribution within the contact line of the rolling elements rolling over the defective surface. As a matter of fact, the detection performance depends, among the other factors, from the evolution of the mechanical degradation. This does not appear to be the case for the ball bearing degradation, where the symptoms are of relatively comparable magnitude, despite the case number 3 presenting a more advanced degradation at the time of removal (figure 1). The fact that the contact taking place in ball bearings between the races and the elements can better be described as a point contact interaction could explain these results. In fact, any geometry of the degradation would almost surely provoke mechanical impacts between the elements and the race. The analysis of the diagnostic charts is an important step to confirm the alarms: the bearing fault signature presents itself as a clear cyclic excitation localized in the high spectral frequency band, and allows the analysts to reliably confirm whether the fault detection alert is effectively related to a mechanical fault. In fact, by observing the cyclic spectral coherence representation, it is possible in an operational scenario to rule out both the occurrence of false alarms and those false indications coming from corrupted measurements.

4.4 Impact of the operating conditions

In order to assess the sensitivity of the bearing monitoring indicators to the different operating conditions, the correlation of the indicator values to the torque and rotor speed values was studied. Figures 9 and 10 show such correlation for the LES-CPW outer race indicators related respectively to the roller (bearing 1) and to the ball bearing (bearing 2). Similar results were obtained for the other indicators, therefore only figures 9 and 10 were reported for brevity. As indicator value, the magnitude of the detected peak in the envelope spectrum was reported. This is a consistent indication, being the LES statistical distribution independent from the cyclic frequency. From the results, it can be observed that there is no significant correlation between the values of the indicator and the contextual parameter, whenever the indicator takes its nominal values. However, the higher values related to degraded conditions are mostly localized in correspondence of the high-torque region, for a rotor speed of around 97 percent of the nominal speed. This may indicate a slight sensitivity of the indicators to the operating conditions. However, it has to be pointed out that the majority of the recorded acquisitions occurred in the low rotor speed, high torque conditions. Being the acquisitions density in the region higher than in the rest of the operating spectrum, the likelihood of acquiring in those conditions during the progression of the fault is also higher. Finally, relatively high values of the indicator during the fault progression can be observed also in other regions of the operating spectrum, confirming its detection robustness to the different flying regimes of the helicopter.

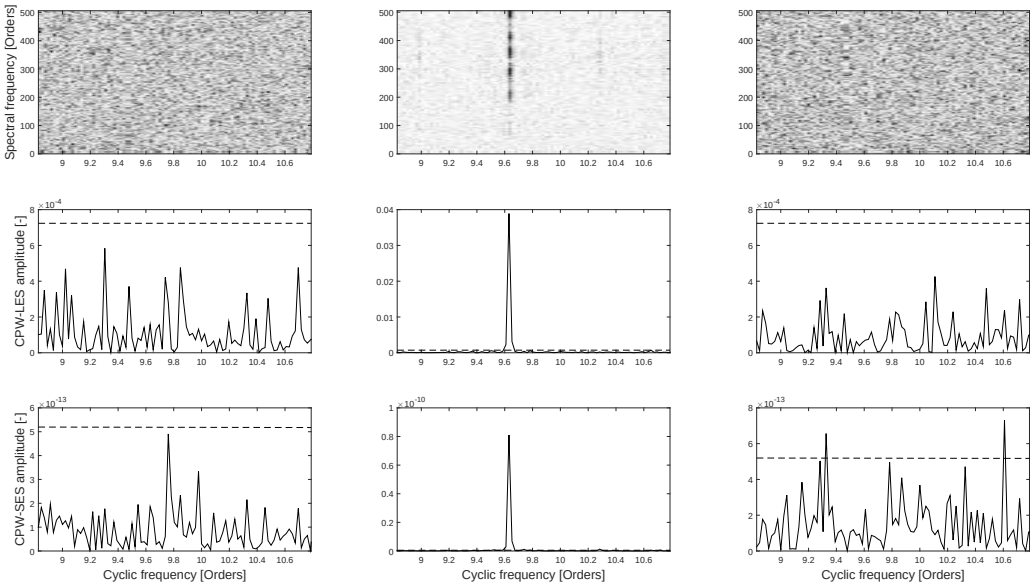


Figure 6 – Diagnostic charts for fault case 2. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

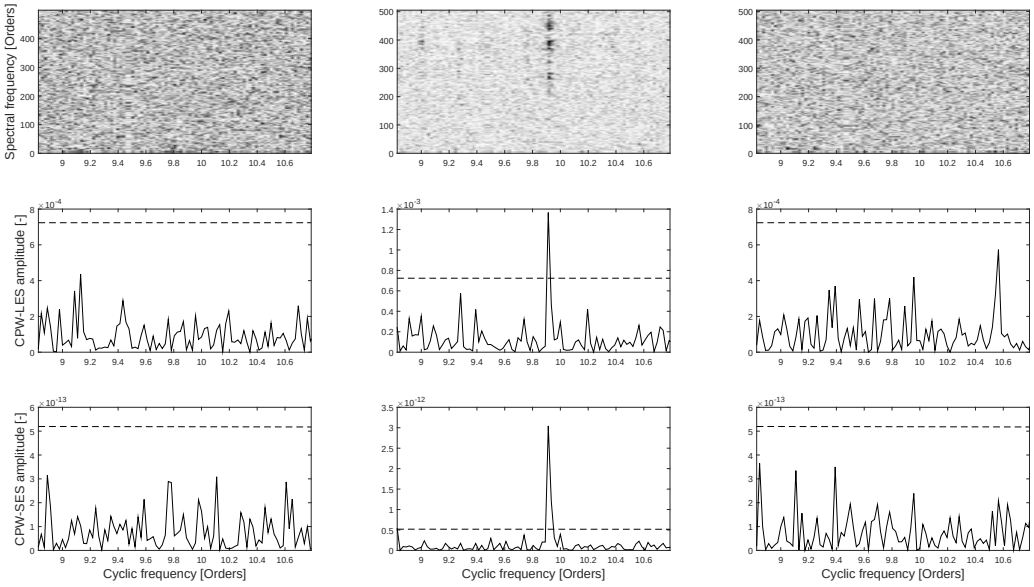


Figure 7 – Diagnostic charts for fault case 3. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

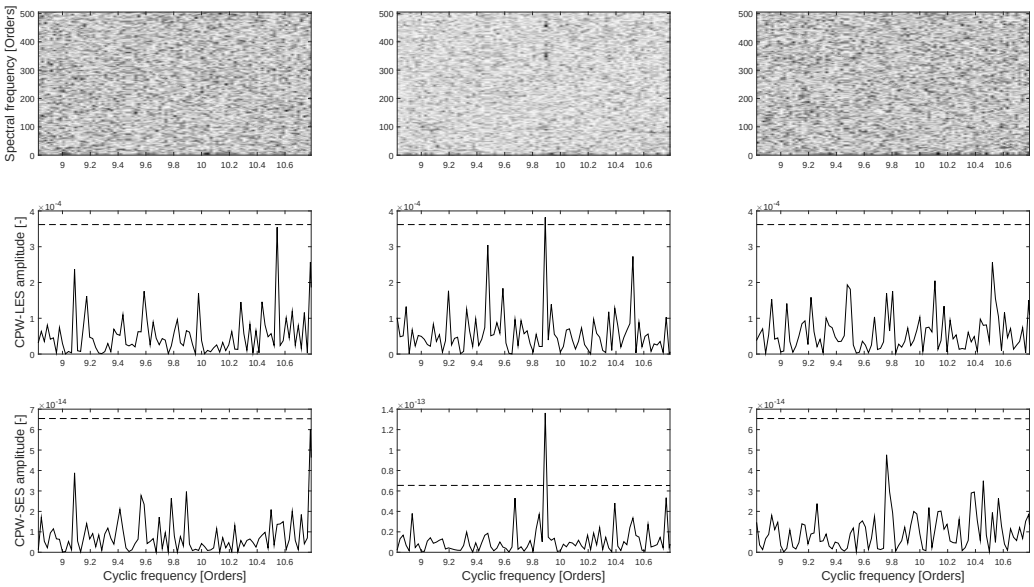


Figure 8 – Diagnostic charts for fault case 4. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

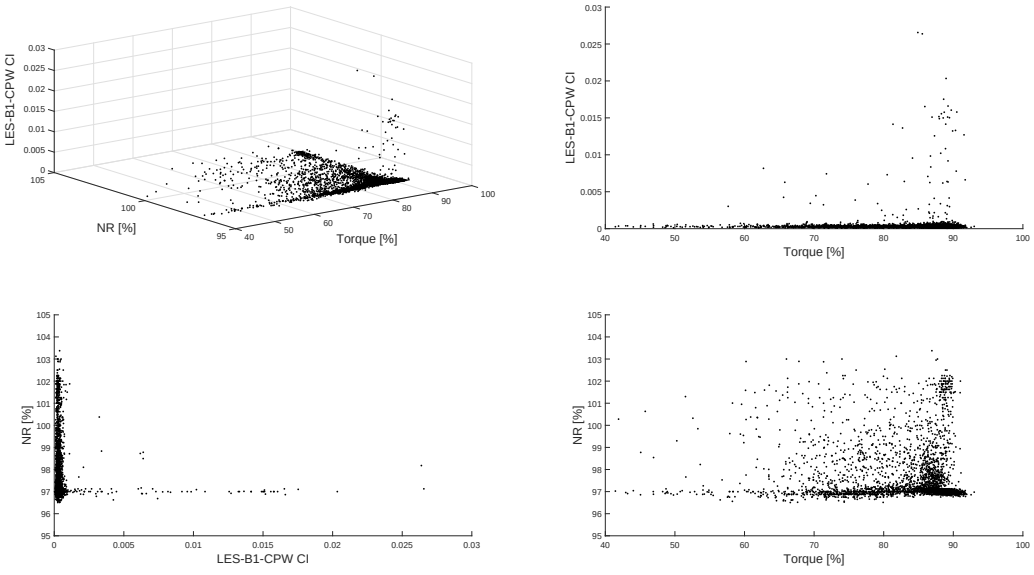


Figure 9 – Impact of the operating conditions on the bearing 1 outer race fault LES-CPW indicator values. Top-left: joint distribution of torque, rotor speed and indicator values; top-right: indicator values vs. torque; bottom-left: rotor speed vs. indicator values; bottom-right: rotor speed vs torque.

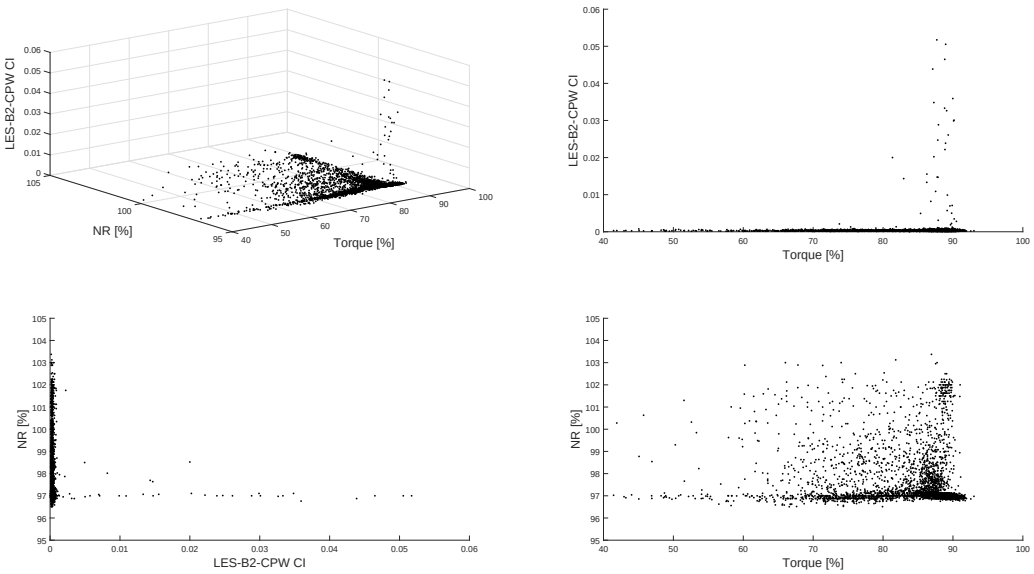


Figure 10 – Impact of the operating conditions on the bearing 2 outer race fault LES-CPW indicator values. Top-left: joint distribution of torque, rotor speed and indicator values; top-right: indicator values vs. torque; bottom-left: rotor speed vs. indicator values; bottom-right: rotor speed vs torque.

5 Conclusions

A predictive maintenance strategy for the monitoring of rolling element bearings in the context of helicopter operations was developed. The proposed methodology relies on a two-step fault detection and diagnosis process. The first step consists of devising reliable statistical indicators, allowing to attain a given false alarm rate. The second step consists of confirming eventual alarms through refined diagnostic analysis based on the estimation of the cyclic spectral coherence of the concerned signals. The devised procedure was validated on comprehensive, in-service helicopter fleet data set, comprising high-frequency acquisitions from fourteen machines flying according to several different profiles. Statistical indicators based on the logarithm envelope spectrum and on the squared envelope spectrum were compared in terms of both reliability in providing a given false alarm rate and ability to promptly anticipate four in-service bearing degradation cases. Two pre-processing algorithms based on synchronous average removal and cepstrum pre-whitening were considered, and some of their properties investigated. It was shown that indicators based on the logarithm envelope allow for a very fine tuning of the desired false alarm rate, together with providing acceptable detection performance in the considered cases. Conversely, squared envelope based indicators proved less reliable in actual operations. The results were shown to be consistent across the full range of considered operating conditions. The diagnostic step was shown to be able of accurately disclosing the faulty bearing signature, proving to be an effective discrimination mean to avoid unnecessary grounding of the machine in the occurrence of a HUMS alarm. At the same time, the burden of performing the diagnosis is restricted to those cases in which an alarm from the detection step actually occurs. The impact of different mechanical degradation shapes was also considered: it was shown that for the roller bearing outer race degradation cases, the fault shape has a significant impact on the HUMS detection performance. The developed approach allowed to deploy a statistically efficient, operationally valid procedure to monitor the rolling element bearings in helicopter mechanical transmissions within an in-service context, characterized by harsh mechanical environment, acquisition constraints and multiple different operating conditions. Remarkably, it could be used to deploy an effective, semi-automated monitoring for helicopter bearings which guarantees an effective condition based maintenance of the monitored components.

References

[1] A.R.S. Bramwell, D. Balmford, and G. Done. *Bramwell's helicopter dynamics*. Butterworth-Heinemann, 2001.

[2] A. Bayoumi et al. “Conditioned-Based Maintenance at USC-Part I: Integration of Maintenance Management Systems and Health Monitoring Systems through Historical Data Investigation”. In: *Proceedings of the AHS International Specialists Meeting on Condition Based Maintenance, Huntsville, AL*. 2008.

[3] B.D. Larder and M.W. Davis. “HUMS Condition Based Maintenance Credit Validation”. In: *American Helicopter Society 63rd Annual Forum*. 2007.

[4] B.D. Forrester. “Advanced vibration analysis techniques for fault detection and diagnosis in geared transmission systems”. PhD thesis. 1996.

[5] I. Howard. *A Review of Rolling Element Bearing Vibration Detection, Diagnosis and Prognosis*. Tech. rep. Defence Science and Technology Organization Canberra (Australia), 1994.

[6] R.J. Drago. *Fundamentals of Gear Design*. Butterworths, 1988.

[7] C. Capdessus, M. Sidahmed, and J.L. Lacoume. “Cyclostationary processes: application in gear faults early diagnosis”. In: *Mechanical systems and signal processing* 14.3 (2000), pp. 371–385.

[8] L. Bouillaut and M. Sidahmed. “Cyclostationary approach and bilinear approach: comparison, applications to early diagnosis for helicopter gearbox and classification method based on HOCS”. In: *Mechanical Systems and Signal Processing* 15.5 (2001), pp. 923–943.

[9] A.C. McCormick and A.K. Nandi. “Cyclostationarity in rotating machine vibrations”. In: *Mechanical systems and signal processing* 12.2 (1998), pp. 225–242.

[10] R.B. Randall, J. Antoni, and S. Chobsaard. “The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals”. In: *Mechanical systems and signal processing* 15.5 (2001), pp. 945–962.

[11] J. Antoni, J. Daniere, and F. Guillet. “Effective vibration analysis of IC engines using cyclostationarity. Part I: a methodology for condition monitoring”. In: *Journal of sound and vibration* 257.5 (2002), pp. 815–837.

[12] G.B. Giannakis. “Cyclostationary signal analysis”. In: *Digital Signal Processing Handbook* (1998), pp. 17–1.

[13] J. Antoni et al. “Cyclostationary modelling of rotating machine vibration signals”. In: *Mechanical systems and signal processing* 18.6 (2004), pp. 1285–1314.

[14] J. Antoni. “Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions”. In: *Journal of Sound and vibration* 304.3-5 (2007), pp. 497–529.

[15] J. Antoni. “Cyclic spectral analysis in practice”. In: *Mechanical Systems and Signal Processing* 21.2 (2007), pp. 597–630.

[16] J. Antoni, G. Xin, and N. Hamzaoui. “Fast computation of the spectral correlation”. In: *Mechanical Systems and Signal Processing* 92 (2017), pp. 248–277.

[17] P.D. McFadden and J.D. Smith. “Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review”. In: *Tribology international* 17.1 (1984), pp. 3–10.

[18] P. Borghesani and M.R. Shahriar. “Cyclostationary analysis with logarithmic variance stabilisation”. In: *Mechanical Systems and Signal Processing* 70 (2016), pp. 51–72.

[19] P. Borghesani et al. “Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals”. In: *Mechanical Systems and Signal Processing* 40.1 (2013), pp. 38–55.

[20] P. Borghesani and J. Antoni. “CS2 analysis in presence of non-Gaussian background noise–Effect on traditional estimators and resilience of log-envelope indicators”. In: *Mechanical Systems and Signal Processing* 90 (2017), pp. 378–398.

[21] P. Borghesani, P. Pennacchi, and S. Chatterton. “The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings”. In: *Mechanical Systems and Signal Processing* 43.1-2 (2014), pp. 25–43.

[22] J. Zeidler et al. “Adaptive enhancement of multiple sinusoids in uncorrelated noise”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.3 (1978), pp. 240–254.

[23] R.B. Randall and Y.J. Li. “Diagnostics of planetary gear bearings in the presence of gear vibrations”. In: *IMECHE Conference Transactions*. Vol. 5. Mechanical Engineering Publications. 1995, pp. 73–80.

[24] J. Antoni and R.B. Randall. “Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 89–101.

[25] J. Antoni and R.B. Randall. “Unsupervised noise cancellation for vibration signals: part II—a novel frequency-domain algorithm”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 103–117.

[26] P.D. McFadden. “A revised model for the extraction of periodic waveforms by time domain averaging”. In: *Mechanical Systems and Signal Processing* 1.1 (1987), pp. 83–95.

[27] R.B. Randall, N. Sawalhi, and M. Coats. “A comparison of methods for separation of deterministic and random signals”. In: *International Journal of Condition Monitoring* 1.1 (2011), pp. 11–19.

[28] N. Sawalhi and R.B. Randall. “Signal pre-whitening using cepstrum editing (liftering) to enhance fault detection in rolling element bearings”. In: *Proceedings of the 24th international congress on condition monitoring and diagnostic engineering management*. 2011, pp. 330–336.

[29] P. Borghesani et al. “Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions”. In: *Mechanical Systems and Signal Processing* 36.2 (2013), pp. 370–384.

[30] J. Antoni and R.B. Randall. “The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines”. In: *Mechanical Systems and Signal Processing* 20.2 (2006), pp. 308–331.

[31] V. Camerini et al. “Impact of pulse time uncertainty on synchronous average: statistical analysis and relevance to rotating machinery diagnosis”. In: *Mechanical Systems and Signal Processing* 129C (2019), pp. 308–336.

[32] P.D. McFadden. *Interpolation techniques for the time domain averaging of vibration data with application to helicopter gearbox monitoring*. Tech. rep. AERONAUTICAL RESEARCH LABS MELBOURNE (AUSTRALIA), 1986.

[33] R.B. Randall and J. Antoni. “Rolling element bearing diagnostics—A tutorial”. In: *Mechanical systems and signal processing* 25.2 (2011), pp. 485–520.

[34] P.D. McFadden and J.D. Smith. “Model for the vibration produced by a single point defect in a rolling element bearing”. In: *Journal of sound and vibration* 96.1 (1984), pp. 69–82.

[35] P.D. McFadden and J.D. Smith. “The vibration produced by multiple point defects in a rolling element bearing”. In: *Journal of sound and vibration* 98.2 (1985), pp. 263–273.

[36] B.J. Hamrock and W.J. Anderson. *Rolling element bearings*. Tech. rep. 1983.

[37] R.B. Randall. “Applications of spectral kurtosis in machine diagnostics and prognostics”. In: *Key Engineering Materials*. Vol. 293. Trans Tech Publ. 2005, pp. 21–32.

[38] N. Tandon and A. Choudhury. “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings”. In: *Tribology international* 32.8 (1999), pp. 469–480.

[39] J. Antoni and R.B. Randall. “A stochastic model for simulation and diagnostics of rolling element bearings with localized faults”. In: *Journal of vibration and acoustics* (2003).

[40] D. Ho and R.B. Randall. “Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals”. In: *Mechanical systems and signal processing* 14.5 (2000), pp. 763–788.

[41] J. Antoni and R.B. Randall. “Differential diagnosis of gear and bearing faults”. In: *Journal of Vibration and Acoustics* 124.2 (2002), pp. 165–171.

[42] S. Kass, A. Raad, and J. Antoni. “Self-running Fault Diagnosis Method for Rolling Element Bearing”. In: *Mechanism, Machine, Robotics and Mechatronics Sciences*. Springer, 2019, pp. 127–140.

[43] J. Antoni. “Cyclostationarity by examples”. In: *Mechanical Systems and Signal Processing* 23.4 (2009), pp. 987–1036.

A new indicator designed from the spectral coherence, proposition and application to bearing diagnosis

Souhayb KASS[1,2], Amani RAAD[2], Jérôme Antoni[1]

[1] Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon, LVA EA677, F-69621 Villeurbanne, France, e-mail : ***jerome.antoni@insa-lyon.fr***

[2] Engineering faculty, Lebanese University, Tripoli, Lebanon

1 Abstract

In vibration-based diagnosis of rolling element bearings, the complexity of the signals requires an expert to use advanced signal processing tools and to interpret the results based on his/her experience. Recently, a few autonomous methods have been proposed to alleviate the demand on the user's expertise, yet they have been mainly focused on fault detection. They ideally track certain properties in the signal, whose occurrence is correlated with the symptom of a fault. This paper follows a similar direction but with wider objectives: it aims to develop an indicator that is sensitive to both non-stationarity, non-Gaussianity and to the modification of the acoustic signature of the vibratory signal. The indicator is based on the recently developed Fast Spectral Coherence, a key tool of the theory of second-order cyclostationary processes. It condenses the whole information initially displayed in three dimensions into a scalar. it initially addresses the case where the faults frequencies are unknown. In addition, the proposed indicator is able to return information for different levels of damages in both stationary and non-stationary operating conditions. A new pre-processing step is provided to ensure an efficient and constant statistical threshold. The proposed indicator is intended to be used in an autonomous process without the need for visual analysis and human interpretation. The proposed indicator is compared with a recent indicator based on the Envelop Spectrum, in terms of classification and detection performance. Several applications using real and benchmarked data eventually illustrate the capability for self-running diagnosis.

Keywords: *spectral coherence, Gaussianity, Stationarity, Acoustic signature, Indicator.*

2 Introduction

Roller bearings (REBs) are one of the essential components of rotating machines, hence the demand for their efficient and reliable condition monitoring (CM). Condition monitoring ensures maximum production, prevents accidents and serious damage and helps to detect failures at an early stage by keeping the system in good conditions. Over the last few decades, considerable research has been conducted on diagnostics based on REB vibrations and acoustics signals leading to the development of some condition's indicators. Therefore, several strategies have been adopted. This is derived from the fact that the occurrence of numerous faults induces changes in signal characteristics that can be described as (i) a deviation from Gaussianity, and/or (ii) a shift in the statistical behavior of the signal from stationary to non-stationary, accompanied by (iii) a change in the machine's acoustic signature.

The crest factor [1], the peak-to-peak, the entropy [2]–[4], the form factor, the third-order central moment (skewness) and the fourth-order central moment (kurtosis) [5], [6], or any higher-order moments or cumulants [7], [8] are all typical examples of the non-Gaussianity measure. They are dedicated to characterizing the non-Gaussian behavior in the form of impulsivity of machine signals. The most traditional and probably the most widely used is the kurtosis, or its combination with other indicators. In the past, they have been used mainly because of their simple calculation formulas, and their relatively

short computation times. This is despite its uncorrelated values with the fault symptoms in numerous situations, reported in many studies[9], [10]. This argument has become obsolete thanks to modern computing capabilities.
The roughness indicator, as traditionally calculated in the time domain from the Aures model [11], is an example of the psychoacoustic parameters used to monitor the existence of a fault based on an alteration of the machine's acoustic signature. It aims to mimic the ability of the human auditory system to detect high-frequency modulation, as evidenced by the faulty rotating machines. The researchers have developed various mathematical models [12]–[15] to estimate roughness but none of them have been normalized.
The degree of cyclostationarity [16] and the indices of cyclostationarity [17] based on the 2nd-order cyclical cumulants or any higher-order cyclostationarity indices based on higher-order cyclical cumulants [18] are typical examples of measures of the non-stationarity that characterizes the cyclostationarity introduced by the fault existence. Despite its importance in diagnosis, the real-time use of CS indices can be hindered in practice by its high cost, especially in real-time applications [19].

This paper aims to fill in these gaps by proposing an indicator sensitive to both non-stationarity, non-Gaussianity and to the modification of the acoustic signature of the vibratory signal. To do so, the frequency domain instead of time domain is used since it better extracts cyclic repetition from a signal produced by a repetitive fault and also reduces the noise impact. The cyclostationary framework is then our subject of interest since it has been reported that rotating machine signals are cyclostationary [20]. Advantage is taken of the availability of a recently proposed fast algorithm to calculate the spectral coherence [19], on which the proposed conditioning indicator is based. The spectral correlation is a three-dimensional distribution of all modulation patterns existing in a signal as a function of the carrier frequency in Hertz and the modulation frequency (also called cyclic frequency) in machine order, which generalizes the SC to nonstationary operations. It is thus considered optimal for revealing bearing fault signatures under stationary and nonstationary speed regimes [21], [22].
The idea is to condensate the whole information initially displayed in the spectral coherence into a scalar after an appropriate weighting performed to select the audible frequencies range from about 20 Hz to 20 kHz and the audible modulations range from about 15 Hz to 200 Hz. This selection can be easily made using the weight $\omega_{k,p}$ designed as a bandpass filter, used to accentuate or reduce certain frequency components in order to model the bandpass characteristic of the roughness on the modulation frequency. This approaching, the spectral coherence which is sensitive to both non-stationarity and non-Gaussianity, also becomes sensitive to the acoustic signature of the vibratory signal. A new preprocessing step is provided in order to eliminate any possible bias (as typically produced by transient disturbances in the signal or the presence of unexplained nonstationarities) in the estimated spectral coherence. This original step is necessary to produce a pivotal statistic by forcing the spectral coherence to have a constant probability distribution with respect to the dual frequencies plan.
The proposed indicator is statistically consistent, i.e. its variance converges to zero when the signal length increases. On the opposite, Aures' roughness does not involve any time average and is therefore prone to significant estimation errors. Contrary to the kurtosis, the proposed indicator separates impulsivity from non-stationarity, allowing the identification of the type of deviation from normality. A non-nonparametric hypothesis test is also provided in order for this indicator to be credible and possibly implemented in an automated monitoring system. The capacity of the proposed indicator is validated on real data and benchmarked with the kurtosis to extract meaningful conclusions. It is found to return higher performance in terms of detecting faulty bearings.

3 Indicator

3.1 Preliminary steps

This steps briefly resumes the statistical methodology proposed by *kass et al.*[22] to design the test statistics. The starting point is to describe the health of the system under investigation by two alternative hypotheses, H_0 and H_1, which correspond respectively to the healthy and the faulty states, respectively. The principle is to consider the spectral coherence, $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, as the random quantity of interest rather than its squared magnitude.

In principle, Under H_0, the random field $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, seen as a function of the two frequency variables α_l and f_k, can be shown to have zero probability of being nil at any position $(\alpha_l\ ,\ f_k)$ even though having small values. Under the alternative hypothesis H_1, the difference is that the random field will have higher magnitudes along parallel lines, discretely located at cyclic frequencies associated with the fault frequencies. The objective is to keep only these values and to zero all the other ones. To do so, a statistical threshold, defined as a high percentile, is needed to differentiate between information and background noise. Hence, the presence of a possible bias (as typically produced by transient disturbances in the signal or the presence of unexplained nonstationarities) in the $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ compromises the efficiency of the latter threshold. As result, the noise baseline is not uniformly distributed along the frequencies axis. It is therefore impossible to establish a fixed threshold to distinguish between information and noise. To correct this situation, the following empirical steps are proposed. The first step is to standardize the EES in order to force it to have a constant probability distribution with respect to the cyclic order α, under the null hypothesis H_0. In principle, The transformation reads

$$\gamma_X^{(2)}(\alpha_l\ ,\ f_k)\ = \frac{\gamma_X^{(1)}(\alpha_l\ ,\ f_k) - \mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}}{\sqrt{\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\} - \mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}^2}} \tag{1}$$

where $\mathrm{E}\{\cdots|H_0\}$ stands for the ensemble averaging operator taken under H_0. One issue is to replace $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}$ and $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\}$ in the above equation by estimates obtained from a realization of $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ which may either pertain to H_0 or H_1. Since the difference in $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ under the null and alternative hypotheses is essentially marked by the presence of parallel lines, it is proposed to estimate $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}$ from a running median of $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, called $\mu_{MED}(\alpha_k)$, and $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\}$ from the running median of the absolute deviation, called $\sigma_{MAD}(\alpha_k)$. The rationale for using a running median is to leave unaffected informative peaks in the spectral coherence. Therefore, the $\gamma_X^{(2)}(\alpha_l\ ,\ f_k)$ reads

$$\gamma_X^{(2)}(\alpha_l\ ,\ f_k)\ =\ \frac{\gamma_X^{(1)}(\alpha_l\ ,\ f_k) - \mu_{MED}(\alpha_l, f_k)}{\sigma_{MAD}(\alpha_l, f_k)} \tag{2}$$

which returns a “pivotal” statistics (i.e. whose probability distribution does not depend on any unknown parameter). Briefly, the presence of the possible bias is firstly removed by subtracting a running median, then the results are standardized by dividing it with a running median of the absolute deviation. These two steps can be though as a normalization of the H_1 statistics with respect to the H_0 statistics.

3.2 Proposition

The first step is to preserve the informative values, i. e. the spectral lines parallel to the frequency axis f_k along α_l, expected during the fault’s existence i. e. under H_1.To do so, this step consists of setting all non-significant peaks in $\gamma_X^{(2)}(\alpha_l\ ,\ f_k)$ to zero which are found below a given threshold. A reasonable choice is

to define the threshold as a high percentile $\mathrm{p_c}$ (e.g. $p_c = 0.9$)[23], which means that only the $100(1 - p_c)\%$ highest values will be kept. Thanks to the standardization step, the threshold is constant over the full frequency plane $(\alpha_l\ ,\ f_k)$. And the $\gamma_X^{(3)}(\alpha_l\ ,\ f_k)$ is expressed as

$$\gamma_X^{(3)}(\alpha_l\ ,\ f_k) = \gamma_X^{(2)}(\alpha_l\ ,\ f_k).\mathbb{I}_{\left[\gamma_X^{(2)}(\alpha_l\ ,\ f_k) > p_c\right]} \tag{3}$$

where the symbol $\mathbb{I}_{\left[\gamma_X^{(2)}(\alpha_l\ ,\ f_k) > p_c\right]}$ denotes the indicator function defined on the frequencies plane $(\alpha_l\ ,\ f_k)$ having the value 1 for all elements of $(\alpha_l\ ,\ f_k)$ satisfying the condition $\gamma_X^{(2)}(\alpha_l\ ,\ f_k) > p_c$ and the value zero otherwise.

This is a crucial step as the next step involves modelling the bandpass characteristics of the roughness by accentuating or reducing certain frequency components, thus preventing the occurrence of misleading peaks.

The next step is to performs an appropriate weighting of the $\gamma_X^{(3)}(\alpha_l\ ,\ f_k)$ so as to select the audible frequencies range from about 20 Hz to 20 kHz and the audible modulations range from about 15 Hz to 200 Hz in order to approach the roughness measurement proposed by Aures. This selection can be easily realized based on the weight $\omega_{k,p}$ designed as a bandpass filter, used to accentuate or reduce specific frequency components in order to model the roughness bandpass characteristic on modulation frequencies. In other words, $\omega_{k,p}$ resembles the distribution matrix of the weighting functions for each *Bark* channel. The weighting of $\gamma_X^{(3)}(\alpha_l\ ,\ f_k)$ is achieved as follows:

$$\gamma_X^{(4)}(\alpha_l\ ,\ f_k) = \gamma_X^{(3)}(\alpha_l\ ,\ f_k).\omega_{k,p} \tag{4}$$

The third step is divided into two sub-steps, the first is to integrating the $\gamma_X^{(4)}(\alpha_l\ ,\ f_k)$ over the cyclical frequency axis α_l, which condenses the whole information initially displayed in three dimensions into a two-dimensional representation,

$$I_X^{(5)}(\ f_k) = \frac{1}{F_1}\left[\sum_{l \in F_1} \gamma_X^{(4)}(\alpha_l\ ,\ f_k)\right] \tag{5}$$

While the second sub-step consists of dividing the frequency axis into a *Bark* filter bank to estimate the modulation depth per auditory channel which is spaced by 1 *Bark* representing a psychoacoustic scale for the bandwidths of the hearing filters. The latter is a frequency band established by Zwiker [24], it is divided into 24 critical bands ranging from 0 to 15500 Hz.

As a final step, the roughness dependence with respect to the carrier frequency is introduced into the model by multiplying $I_X^{(5)}(\ f_k)$ by a weighting function $g(f_k)$ with factors ranging from 0.6 to 1.1 in accordance with the dependency of the roughness to the carrier frequency of the amplitude modulated tones. The values of the weighting function with respect to the channel number are shown in Figure 1.

Finally, the proposed indicator I_R is obtained by integrating $I_X^{(5)}(\ f_k)$ over the frequency axis f_k.

$$I_R = \frac{1}{F_2}\sum_{k \in F_2} I_X^{(5)}(\ f_k)\ .\ g(f_k) \tag{6}$$

By analogy with the result given by the connection between the kurtosis and the sum of the squared envelope spectrum [25], it was shown in [23] that the proposed indicator might be interpreted as a kurtosis, yet sensitive only to cyclostationary components.

The next section describes how the fault detection will be done using a statistical hypothesis test using the proposed indicator. In this context, alternative strategies to statistical testing can also be used. For example, the proposed indicator can also be used as an input parameter for an SVM classifier (machine

vector support) or a neural network. This statement is based on the observation of the results obtained by applying the proposed indicator to the various databases.

3.3 Hypothesis testing, design

The interpretation of the proposed indicator could vary from one application to another, depending on several parameters (such as the noise level related to transient perturbations in the signal or to the presence of unexplained non-stationarity as well as the vibration level and the interfering contribution of other second-order components emitted by other sources). In the majority of the literature [16]–[18], the provided methods give their results as scalar. The latter indicates the presence of a fault in some applications while in another, and for the same value, the fault will be considered as absent. This is why a threshold is needed for decision-making. To do so, two thresholds are provided in this paper. The lower threshold I_L is defined as the indicator value of the randomized version of the vibration signal under which the fault is absolutely absent. In detail, I_L is equal to the indicator value when the signal is randomized. The randomization of the signal is defined as a circular permutation of its elements, it can be performed using the MATLAB function called "randperm". The latter returns a new version of the signal containing a random permutation of its values. On the other hand, the link between the kurtosis and the spectral correlation makes it possible to reach the upper threshold, I_U, beyond which the fault presence is declared with high certainty.

After the I_U is calculated, a comparison is then made with the proposed indicator to detect the fault presence. If I_R has a value greater than I_U, the fault exists.

The null hypothesis test relative to our case originating from the comparison between I_R and I_U can be written as:

“Reject the null hypothesis H0 if:

$$I_R \geq I_U \tag{7}$$

where $I_U \geq 2 \times I_L$.

It easily allows performing a statistical test: according to the decision rule, any value of I_R that is greater than the I_U-threshold will indicate that the signature of the fault is detected. The proof of proposition Eq. (7) is based on observing that under the null hypothesis test H_0 the quantity asymptotically follows a nonparametric distribution that has a constant bias and variance all over the cyclic order axis. It also remembers that this test is true almost everywhere.

The complete flow diagram for the algorithm described in this section is shown in Fig.3

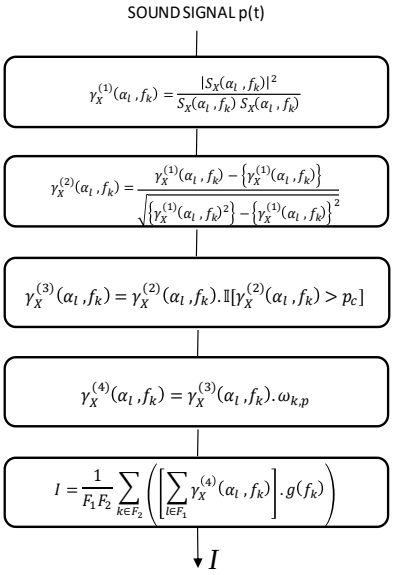


Figure 1.Complete algorithm flowchart

It is worth noting that I_R is very similar to the Aures' roughness measure used in psychoacoustics. One difference is that Aures' roughness is based on a decomposition of the signal through a Bark filter bank whereas a narrow-band decomposition is used in this paper, yet this is more or less transparent after integration over the frequency plane $(\alpha_l\ ,\ f_k)$. Another difference is that Aures' roughness does not involve any time average and is therefore prone to significant estimation errors. On the contrary, the indicators introduced in this work are statistically" consistent" (i.e. their variances converge to zero when the signal length increases).

4 Experimental Validation

The ability of any method in detecting a bearing fault must be validated on real signals. In the present paper, four benchmarks are used. The first is provided by the Case Western Reserve University (CRWU) bearing data center [26], while the second is an industrial database provided by SOMFY-Cluse. These databases are widely used to test new algorithms by comparing their efficiency with existing techniques [27]. The CRWU's database provides multiple fault types, i.e. rolling element, cage inner-race, and outer-race fault, and it is used to illustrate the proposed method and to compare the proposed algorithm to those existing in the literature. The industrial database is used to illustrate the diagnosis of bearings in a real industrial world signal.

4.1 Algorithm illustration and comparison with kurtosis

To illustrate the proposed algorithm, we consider a real industrial signal. This analyzed signal is provided by Somfy and includes an industrial fault. It is provided as supplementary material of the article. This may be used as a general source of bench mark data for research on diagnosis of industrial faults under constant speed operation. The comparing the results of the proposed indicator with those given by kurtosis. The duration of signals is 20 s with a sampling frequency of 50 kHz.

As explained above, the first begins with the calculation of the fast estimator of the spectral coherence for the resampled time domain signal $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$. In what follows, the window length in the Fast-OFSC is set to $Nw = 2^9$ in order to achieve a frequency resolution of about 100 Hz and the cyclic range $\alpha_{max} = 750$ Hz. The next step is to standardize $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ in order to force it to have a constant probability

distribution with respect to the frequency plane, under H_0. The statistical threshold will be defined as a high P_c percentile ($P_c = 0.9$), which means that only the $100(1 - P_c)\%$ of the highest values will be preserved. It is noteworthy that this method perfectly preserves the diagnostic information that nicely appears with a significant overrun of the 0,1% statistical threshold. This signal is easily diagnosable and it should, therefore, be considered as a preliminary test for the proposed algorithm. The visual inspection of the spectral coherences presented in Figure 2 (a) and (b) shows a series of symptomatic pulses at the fault frequency - spectral lines parallel to the *f-axis* discretely located at cyclic frequencies associated with the fault frequencies- as expected by the model given by Somfy during under H_1. In Figure 2 (b) it is obvious that some frequency components disappear from the $\gamma_X^{(1)}(\alpha_l\,,\,f_k)$ while retaining only informative peaks and eliminating noise-related components found below the chosen threshold.

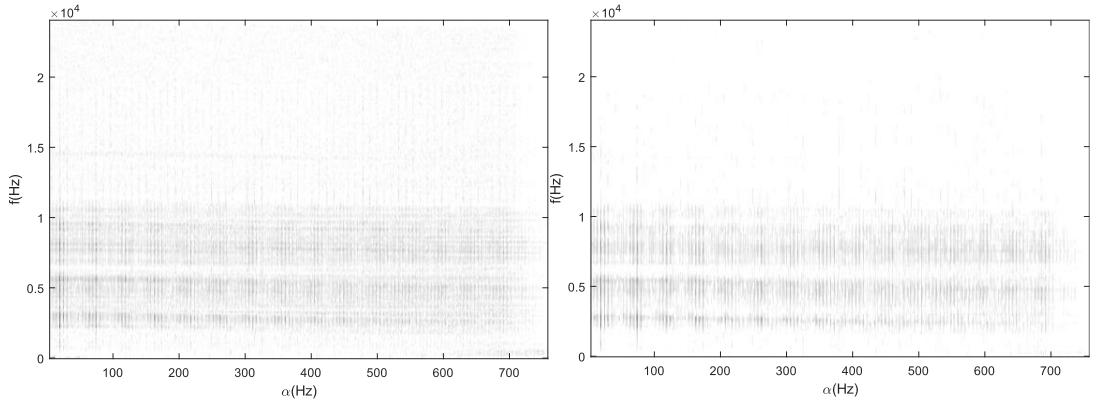


Figure 2. For the raw time signal a) the $\gamma_X^{(1)}(\alpha_l\,,\,f_k)$, and b) the $\gamma_X^{(3)}(\alpha_l\,,\,f_k)$

The audible frequencies and modulations range are selected by using the weight $\omega_{k,p}$, shown in Figure 3(a), so as to select the audible frequencies range from about 20 Hz to 20 KHz and the audible modulations range from about 15 Hz to 200 Hz. The weighted version of spectral coherence $\gamma_X^{(4)}(\alpha_l\,,\,f_k)$ is presented in Figure 3 (b). As shown in Figure 3 (b), certain frequency components are emphasized or reduced so as to model the band-pass characteristic of the roughness over the modulation frequency.

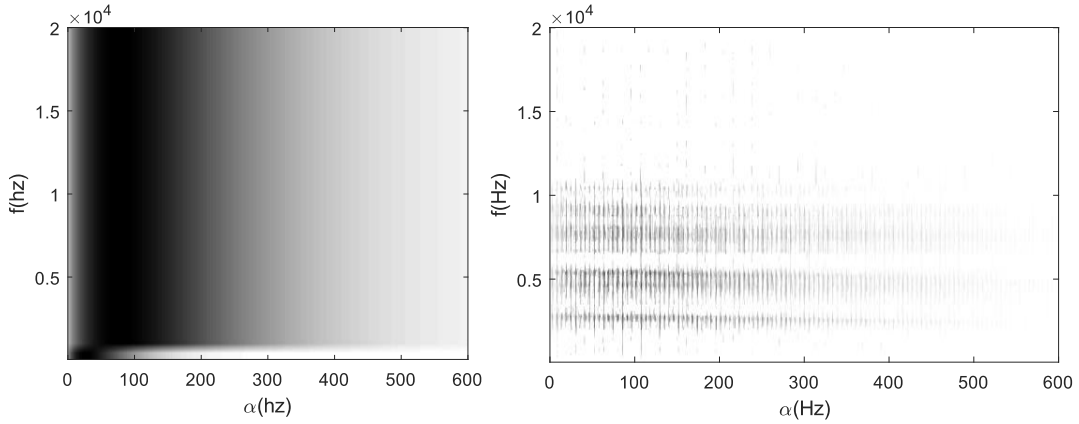


Figure 3 (a) The weight $\omega_{k,p}$ used to select the audible frequencies and modulations ranges, (b) the $\gamma_X^{(4)}(\alpha_l\,,\,f_k)$.

In the next step and as mentioned in section 3.2, the integration of $\gamma_X^{(4)}(\alpha_l\,,\,f_k)$ over α_k will be performed. The two-dimensional representation $I_X^{(5)}(\,f_k)$ which condenses the three-dimensional information is shown in Figure 4 (a). Then, the modulation-depth per auditory channel is estimated by

dividing the f-axis into the 24 bands of Bark filter. $I_X^{(5)}(f_k)$ is then multiplied by a weighting function $g(f_k)$. Finally, the roughness indicator I_R is obtained by integrating the weighted $I_X^{(5)}(f_k)$ over the f-axis. The latter has in this case a value of 2.2624.

The signal is now randomized in order to calculate the upper and the lower threshold. Figure 4 (a) et (b) shows both the raw time signal and its normalized version. When both signals are visually inspected, it is evident that cyclostationary symptoms are lost when the signal is randomized, which is consistent with the auditory test performed using MATLAB's sound function (). More precisely, when performing a hearing test, the original signal exhibits periodic behavior that is produced with each cycle. This periodic symptom is no longer heard after the randomization of the signal. The value of the indicator in this case $I_L = 0.27109$ is negligible compared to the case of existing fault, matching both the visual and the hearing inspection. I_R is equal to about 10 times I_L. According to the decision rule provided by Eq. (7), the fault existence is reported.

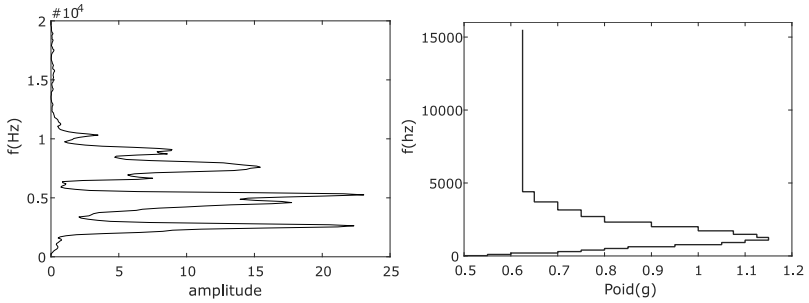


Figure 4. (a) the integration of $\gamma_X^{(4)}(\alpha_l, f_k)$ over α_l, (b) the weighting function g

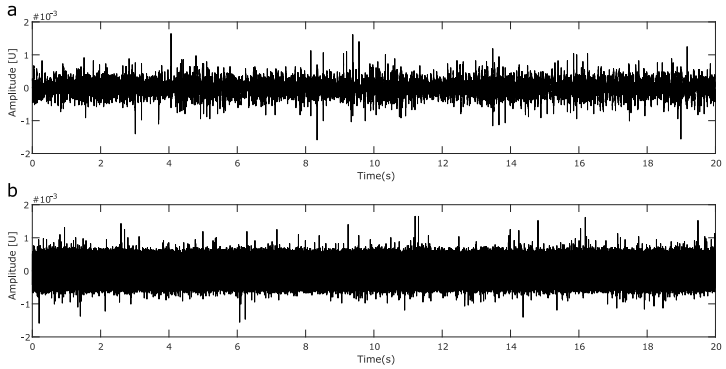


Figure 5. (a) the raw time signal and (b) its randomized version

Surprisingly, and contrary to both visual and auditory inspection tests, the kurtosis of the original and randomized signal gives a very low value of 3.0088 reporting the fault absence in both cases.

After an appropriate filtration of the raw time signal, the kurtosis value is now 20.8589, corresponding to a very high value -7 times the value of a normal case - indicating the failure's presence. The success of kurtosis after proper filtering highlights its limitation when analyzing a signal with a low signal-to-noise ratio and at the same time demonstrating the superiority of the indicator designed from a cyclostationary method for the detection of fault symptoms.

Now the filtered signal is randomized to study kurtosis response in this case and presents another superiority of the proposed indicator over kurtosis. In detail, the same kurtosis value for both the randomized signal and the filtered signal is obtained indicating that the fault is detected in these cases. The obtained results are expected since kurtosis is defined as follows

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \quad (8)$$

This equation shows that regardless the sequence of this signal summation, results will be the same since the permutation disorganizes only the sequence of the original signal. The kurtosis in this case cannot therefore indicate whether the signal is cyclostationary or stationary but not Gaussian. On the other hand, the proposed indicator can perform this distinction giving a cyclic roughness value of 2.2424 for the filtered signal and 0.25819 for its randomized version.

4.2 Performance Evaluation in the CWRU database

The performance of the proposed indicators is now evaluated on the bearing signals provided by the CWRU database. The CWRU database has been used in many references (e.g. [19], [22], [27], [28]) and can be considered as a reference to test newly proposed algorithms and compare them against the state-of-the-art. The experimental setup consists of a 1.4914 kW, reliance electric motor driving a shaft on which a torque transducer and encoder are mounted. Torque is applied to the shaft via a dynamometer and electronic control system. Four types of vibration signals are collected (normal, ball fault, inner-race fault, and outer-race fault), acquired by accelerometer sensors under different operating loads and speeds. More details about the test bench as well as the description of its vibration signals can be found in the reference source [27]. In this study, the drive end data-set category with sampling frequency 48 kHz have been analyzed. Information for all 64 data sets used are shown in table 1. The capacity of the proposed indicator is evaluated using different faults types. The $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ parameters are as given in the previous section.

Table 1. The 48K drive end bearing faults data sets used.

Fault types	Data sets name
Inner-race	110,111,112,174,176,177,14,215,217
Outer-race (centered)	135,136,137,138,201,203,204,238,239,240,241
Normal data	97,98,99,100

Table 2 and table 4 collect the results of the proposed method. Included in these tables are the I_R as well as the *kurtosis* values, of each raw time signal and its randomized version.

As shown in Table 3, for the original signals, I_R is close to 0.15, while for its randomized version, it is about 0.1 ($I_L = 0.1$). In all these cases and in according to the decision rule in equation (7), the fault is declared missed. The kurtosis in these cases is approximately 3. From Table 3, nearly the same increasing or decreasing behavior of the proposed indicator values are detected compared to the values provided by kurtosis and by the roughness indicator provided in commercial psychoacoustic software. In all these cases and in according to the decision rule in equation (7), the fault is declared presented. According to the obtained results, all faults detected by the human visual inspection of [27] in the inner ring and outer ring are also detected by the proposed indicator. It is clearly proven that the distinction between healthy and defective bearings can be made using these indicators. Unfortunately, given that the proposed indicator provides overlapping values when applied to the different types of bearing faults, the existence of the fault can be detected but not identified.

In conclusion, the objective has been achieved and the proposed indicator can identify the fault even if its frequencies are unknown.

Table 2. analysis results of the healthy bearing; Kurtosis, I_R

	Raw signal	Randomized signal

Dataset	*kurtosis*	I_R	*kurtosis*	I_L
100	2.9572	0.1354	2.9572	0.0999
97	2.7642	0.1471	2.7642	0.1011
98	2.9306	0.1422	2.9306	0.1002
99	2.9306	0.1422	2.9306	0.0977

Table 3. analysis results of the faulty bearing; Kurtosis, I_R

Inner-race faults				
	Raw signal		Randomized signal	
Dataset	*kurtosis*	I_R	*kurtosis*	I_L
110	7.36862	0.494821	7.36862	0.101008
111	7.51043	0.467891	7.51043	0.103501
112	6.82779	0.439782	6.82779	0.104584
174	11.6689	0.504829	11.6689	0.10279
176	20.1511	0.861307	20.151	0.117898
177	14.9400	0.823425	14.940	0.114356
214	3.93724	0.507309	3.93724	0.100236
215	3.65185	0.444887	3.65185	0.099826
217	3.65185	0.444887	3.65185	0.09756
Outer-race faults (Centered)				
	Raw signal		Randomized signal	
Dataset	*kurtosis*	I_R	*kurtosis*	I_L
135	6.7415	0.5032	6.741461	0.1047
136	6.8648	0.5293	6.864769	0.1027
137	6.9370	0.5295	6.937026	0.1017
138	7.3917	0.5154	7.391694	0.1030
201	3.7676	0.2039	3.767617	0.1031
203	6.0172	0.2108	6.017181	0.1015
204	2.9993	0.2104	2.999284	0.0996
238	20.7910	2.3374	20.79105	0.1169
239	20.5436	2.7322	20.54365	0.1148
240	21.0265	0.5032	21.02651	0.1205
241	19.7818	0.5293	19.78182	0.1199

5 Conclusion

This paper aims introduces an autonomous method of bearing diagnosis. It relies on the introduction of a new scalar indicator. The indicator results from a post-processing of the spectral coherence, as computed by the fast algorithm.

The factors that are likely to impede the autonomous diagnosis have been addressed; a new standardization of the estimated spectral coherence to remove any possible bias and frequency dependence in the estimation variance. The method comes with a robust hypothesis test, which is crucial for decision making.

The proposed method has been validated on several databases, where it has been checked to be able to systematically replace both the human intervention or the classical conditioning indicator to efficiently complete the diagnosis of bearings.

6 References

[1] J. Igba, K. Alemzadeh, C. Durugbo, and E. T. Eiriksson, “Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes,” *Renew. Energy*, vol. 91, pp. 90–106,

2016.

[2] R. Yan and R. X. Gao, “Approximate entropy as a diagnostic tool for machine health monitoring,” *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 824–839, 2007.

[3] H. Hong and M. Liang, “Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform,” *J. Sound Vib.*, vol. 320, no. 1–2, pp. 452–468, 2009.

[4] J. Antoni, “The infogram: Entropic evidence of the signature of repetitive transients,” *Mech. Syst. Signal Process.*, vol. 74, pp. 73–94, 2016.

[5] A. Rai and S. H. Upadhyay, “A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings,” *Tribol. Int.*, vol. 96, pp. 289–306, 2016.

[6] C. Pachaud, R. Salvetat, and C. Fray, “Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces,” *Mech. Syst. Signal Process.*, vol. 11, no. 6, pp. 903–916, 1997.

[7] M. Cocconcelli, G. Curcurú, and R. Rubini, “Statistical evidence of central moment as fault indicators in ball bearing diagnostics,” in *The International Conference Surveillance 9*, 2017.

[8] H. R. Martin, “Statistical moment analysis as a means of surface damage detection,” in *Proceedings of the 7th international modal analysis conference*, 1989, vol. 1, pp. 1016–1021.

[9] J. Antoni and P. Borghesani, “A statistical methodology for the design of condition indicators,” *Mech. Syst. Signal Process.*, vol. 114, pp. 290–327, 2019.

[10] R. B. Randall, J. Antoni, and K. Gryllias, “Alternatives to kurtosis as an indicator of rolling element bearing faults,” in *Proceedings of ISMA2016 International Conference on Noise and Vibration Engineering*, 2016, pp. 2503–2516.

[11] W. Aures, “Ein berechnungsverfahren der rauhigkeit,” *Acta Acust. united with Acust.*, vol. 58, no. 5, pp. 268–281, 1985.

[12] P. Daniel and R. Weber, “Psychoacoustical roughness: Implementation of an optimized model,” *Acta Acust. united with Acust.*, vol. 83, no. 1, pp. 113–123, 1997.

[13] P. N. Vassilakis, “Auditory roughness as a means of musical expression,” *Sel. Reports Ethnomusicol.*, vol. 12, no. 119–144, p. 122, 2005.

[14] M. Leman, “Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM),” in *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), December*, 2000, pp. 7–9.

[15] R. Duisters, “The modeling of auditory roughness for signals with temporally asymmetric envelopes.” Master’s thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2005.

[16] G. D. Živanović and W. A. Gardner, “Degrees of cyclostationarity and their application to signal

detection and estimation," *Signal Processing*, vol. 22, no. 3, pp. 287–297, 1991.

[17] P. PRIEUR and G. D'URSO, "Des indices de cyclostationnarité pour la surveillance des engrenages," in *15° Colloque sur le traitement du signal et des images, FRA, 1995*, 1995.

[18] A. Raad, J. Antoni, and M. Sidahmed, "Indicators of cyclostationarity: Theory and application to gear fault monitoring," *Mech. Syst. Signal Process.*, vol. 22, no. 3, pp. 574–587, 2008.

[19] J. Antoni, G. Xin, and N. Hamzaoui, "Fast computation of the spectral correlation," *Mech. Syst. Signal Process.*, vol. 92, pp. 248–277, 2017.

[20] J. Antoni, "Cyclic spectral analysis in practice," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 597–630, 2007.

[21] D. Abboud, "Vibration-based condition monitoring of rotating machines in nonstationary regime," 2015.

[22] S. Kass, A. Raad, and J. Antoni, "Self-running bearing diagnosis based on scalar indicator using fast order frequency spectral coherence," *Measurement*, vol. 138, pp. 467–484, 2019.

[23] J. Antoni, P. Borghesani, S. Kass, A. Raad, and K. Gryllias, "Methodologies for designing new condition indicators," in *Proceedings of ISMA 2018-International Conference on Noise and Vibration Engineering and USD 2018-International Conference on Uncertainty in Structural Dynamics*, 2013, pp. 883–890.

[24] E. Zwicker and H. Fastl, "Psychoacoustics: Facts and Models. 1990." Springer-Verlag.

[25] P. Borghesani, P. Pennacchi, and S. Chatterton, "The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 43, no. 1–2, pp. 25–43, 2014.

[26] "The Case Western Reserve University Bearing Data Center Website." [Online]. Available: https://csegroups.case.edu/bearingdatacenter/.

[27] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, 2015.

[28] W. Li, M. Qiu, Z. Zhu, F. Jiang, and G. Zhou, "Fault diagnosis of rolling element bearings with a spectrum searching method," *Meas. Sci. Technol.*, vol. 28, no. 9, p. 95008, 2017.

Multi band integration on the cyclostationary bivariable methods for bearing diagnostics.

Alexandre MAURICIO[1,2], Konstantinos GRYLLIAS[1,2]

[1]KU Leuven, Faculty of Engineering Science, Department of Mechanical Engineering, Division PMA, Celestijnenlaan 300 B, B-3001, Heverlee, Belgium

[2]Dynamics of Mechanical and Mechatronic Systems, Flanders Make
alex.ricardomauricio@kuleuven.be

Abstract

Rolling element bearings are critical parts of rotating machinery, as they support the loads applied to the rotating components. Therefore, continuous monitoring of the health state of the operational bearings is applied in order to detect early damages before any unexpected breakdown of the rotating machinery occurs. Bearing diagnostics is a field of intensive research, focusing nowadays mainly in complicated machinery (e.g. planetary gearboxes, multi-stage gearboxes, etc.) operating under varying conditions (e.g. varying speed and load), as they still provide challenges in terms of accuracy and time of detection/diagnosis. One of the most common methods for bearings diagnostics is the Envelope Analysis. A filter is usually applied around an excited frequency band (by impulsive damage) and the signal is enveloped, thus obtaining the Squared Envelope Spectrum. For the detection of the filtering frequency band, several band selection tools have been proposed in the past that extract the optimal band in a semi-autonomous or fully autonomous manner. The most widely used tool for band selection is the Kurtogram, where the band that returns the highest Spectral Kurtosis value is selected as the optimal band for demodulation. However, as the bearing damage may excite several frequency bands simultaneously, band-pass filtering around only one frequency band may not be sufficient for the detection of the bearing fault under the presence of noise. One proposed method to circumvent this case is to filter around several bands that carry the Signal of Interest (bearing damage signature). Recently, multi-band filtering based on the Autogram feature values, used as a pre-step in order to extract the Combined Squared Envelope Spectrum (CSES) has been presented, providing better detection performance of faulty bearings compared to the extraction of the SES after filtering over a single optimal band returned by the Autogram. Recently, a particular interest had been target to the Cyclic Spectral Correlation (CSC) and to the derived methods, due to their effectiveness in describing second-order cyclostationary signals. One of such methods is the Cyclic Spectral Coherence (CSCoh) which is a normalized version of the CSC bivariable map. Both methods are represented in the frequency-frequency domain. It has been shown that the integration of the bivariable functions over discrete spectral frequency bands is analogous to band-pass filtering. The IESFOgram has been proposed as a band selection tool, based on either the CSC or CSCoh, in order to extract the optimal frequency band. The integration on the frequency band of the bivariable map further enhances the detectability of faulty bearings on the resulting Improved Envelope Spectrum (IES). However, the method has been proposed with the integration of one single band. In this paper the method is extended towards the extraction of the Combined Improved Envelope Spectrum (CIES), performing a multi-band integration of the bivariable map around multiple resonant frequencies that are carriers of the bearing damage signature. The proposed method is applied, tested and evaluated on experimental data and the results are compared with other state-of-the-art band-selection tools.

1 Introduction

Rolling element bearings are critical components of rotating machinery and their failure can cause sudden breakdown of the system, leading to time-loss and increased costs. Condition monitoring is the field where rotating machinery is analysed, including bearings and gears and damages that may be present on the structures can be detected. Therefore, maintenance and faulty component repair can be performed before breakdown. The

diagnostics of bearings continues to be a challenge however, as their signatures are usually masked under noise and other stronger component signatures (e.g. gears). Specifically, condition monitoring of complex machinery has seen increased research, due to their wide application on critical mechanisms and to their high difficulty to diagnose caused by their many components signatures.

One of the most well established methods is the Envelope Analysis, where the signal is demodulated after band-pass filtering around the resonant frequencies excited by the damage impulses, obtaining in the end a filtered Squared Envelope Spectrum (SES). The main idea is to obtain an optimal filter band which presents a high Signal-to-Noise ratio (SNR) leading to a SES after demodulation where the fault harmonics are enhanced [1, 2]. The selection of this frequency band for demodulation is a frequency and continuous topic present in the field. The main reason is because some sort filtering processing is common to most of condition monitoring applications, and the band can either be selected by engineering knowledge, or by a methodology that selects the band in a (semi-)automated manner. The most widely used of these band selections tools is the Fast Kurtogram (FK) [3], which is an automated band selection tool based on the maximum kurtosis level. Aside from this tool there are other band selection tools that have been developed to obtain the SES. The Optimised Spectral Kurtosis (OSK) [4] selects the band with the maximum kurtosis as well, while retaining a narrow bandwidth in order to by-pass electro-magnetic interference noise on the signals. The Sparsogram [5] is based on the sparsity level on different bands based on the wavelet-packet, and the Infogram is utilizes the negentropy as a feature to detect the impulsive bands of the signal for demodulation. Moshrefzadeh and Fasana proposed the Autogram [6], a tool also based on the maximum kurtosis, but unlike the FK, it is calculated from the unbiased autocorrelation of the squared envelope of the demodulated signals. Instead of a classical filter (e.g. Butterworth filter), the undecimated wavelet packet transform (MODWPT) is used instead to split the signal in a series of frequency bands. The band of the autocorrelated squared envelope with the highest Kurtosis is selected as the optimal one, and is shown to have higher diagnosis performance than other stat-of-the-art band selections tools. One conclusion of the method is that band selection tools select only one node as the optimal, and often other unused nodes may contain useful information that is negleted. As such, they introduce the concept of multi-band integration to the Autogram, where several filtered SES, corresponding to the highest kurtosis of each level, are all combined into one spectrum denominated as Combined Squared Envelope Spectrum (CSES).

The Cyclic Spectral Correlation (CSC) and the Cyclic Spectral Coherence (CSCoh) have been proposed in the last two decades as an alternative for the SES-based methods [7, 8, 9]. The main advantage of this method falls on its ability to reveal hidden periodicities of second-order cyclostationarity, like bearing signals that are masked under stronger signals. They are represented in bi-variable maps in the frequency-frequency domain, from which its spectral axis can be integrated to obtain either the Enhanced Envelope Spectrum (EES) or the Improved Envelope Spectrum (IES). These spectra, the EES and IES, have been seen to improve the detection of cyclostationary faulty signals. However, to obtain the optimal band of demodulation for the CSC or CSCoh, its bi-variable map needs to be analysed in order to select the optimal band for integration along the spectral axis. The IESFOgram [11] as been previously proposed as a band selection tool to be applied on the bi-variable maps of CSC or CSCoh, in order to take advantage of its good perfomance in extracting the cyclistationry information of the signals. It also displays a color-mapped 1/3 binary tree like the FK and is seen to provide an optimal band of integration resulting in an IES allowing the detection of the fault frequency harmonics. However, as information of the damage can be present in other bands beside the optimal band, the authors propose an approach to combine the spectra of different bands into one Combined Improved Envelope Spectrum (CIES).

The objective of this paper is the proposal of extending the IESFOgram methodolgy by adding the information of other bands into a combined spectrum, the CIES. The methodology is validated on real signals of two datasets (one with roller bearing damage under variable speed and load conditions [10] and one from a planetary gearbox with electromagnetic interference [4]). Furthermore, the performance of the methodology is compared with the band pass filtering selection based on the Fast Kurtogram-based SES and the Autogram-based CSES. Figure 1 depicts a scheme of the used methodologies. The rest of the paper is outlined as follows. In Section 2, the background theory deemed needed for the application of the proposed method is detailed. In Section 3 the proposed methodology itself is presented. In Section 4, the methodology is tested, validated and compared with state of the art methodologies. The paper closes in Section 5 with some conclusions.

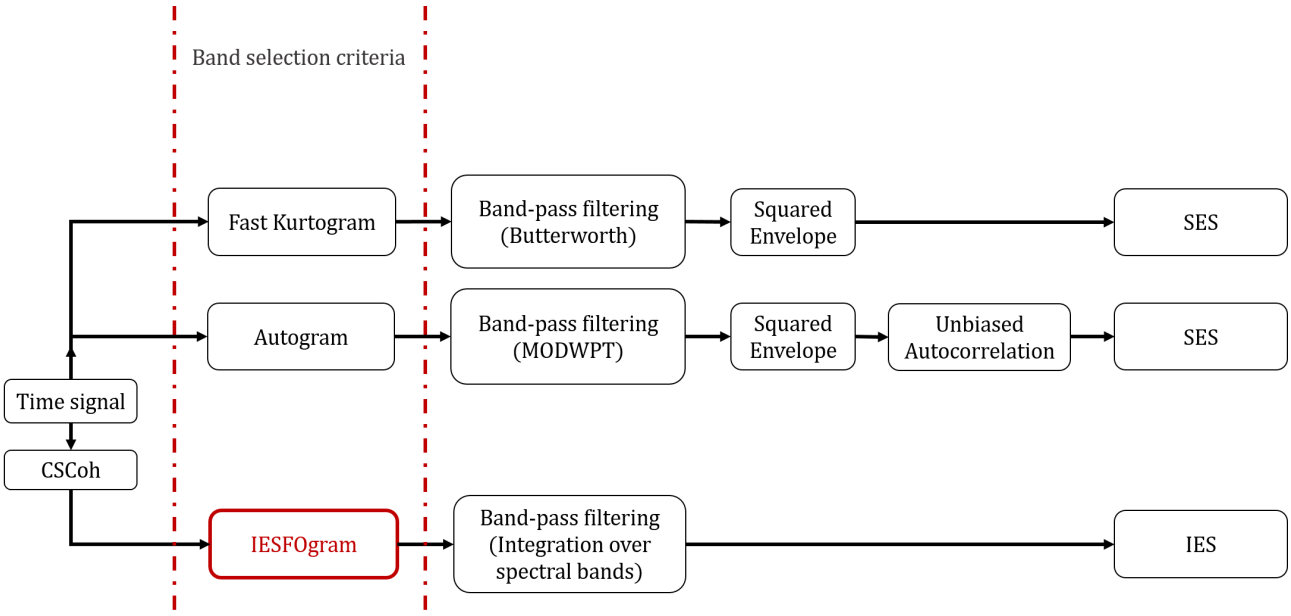


Figure 1: Schematic description of diagnostic methods based on Fast Kurtogram, Autogram, and IESFOgram

2 Cyclostationary signals

Rotating mechanical components are likely to generate cyclic transient signatures which are periodic in nature if the rotational speed is kept constant during the acquisition of signals. These signals often carry information on the health of its components, and signal processing and feature extraction are widely used in order to track the health condition of its components. Following the cyclostationary theory, the signals of interest acquired from rotating machinery can be defined into two orders of cyclostationary signals. Signals of first order of cyclostationarity (CS1) are signals whose first-order statistical moment is a periodic function of T that complies with the condition of Eq. 1.

$$C_{1x}(t) = \mathbb{E}\{x(t)\} = C_{1x}(t+T) \tag{1}$$

where $\mathbb{E}$ denotes the ensemble averaging operator, and t stands for time. In rotating machinery, CS1 vibrations signals are periodic waveforms related to components phase-locked with the rotor speed(*e.g.* shaft misalignment, spalling on meshing gears, etc). A second-order cyclostationary (CS2) signal is a signal whose second order statistical moment is periodic [13]. In particular, if its autocorrelation function is periodic with period T as described in Eq. 2.

$$C_{2x}(t,\tau) = \mathbb{E}\{x(t)x(t-\tau)^*\} = C_{2x}(t+T,\tau) \tag{2}$$

where τ corresponds to the time-lag variable. Bearing vibration signals are often described as CS2, due to having a hidden periodicity related to the shaft speed. Finally, an nth- order cyclostationary (CSn) is a signal whose nth-order statistical moment is periodic, but signal with higher order than CS2 are not taken into account, as CS1 and CS2 describe well the signals of interest generated by rotating machinery.

The Cyclic Spectral Correlation (CSC) is a tool in which the CS1 and CS2 signals are well described in the frequency-frequency domain. The method is represented as a distribution function of two frequency variable: the *cyclic frequency* α linked to the modulation; and the *spectral frequency* f linked to the carrier signal. The tool can be described also as the correlation distribution of the carrier and modulation frequencies of the signatures present in the signals, defined in Eq. 3.

$$CSC(\alpha, f) = \lim_{W\to\infty} \frac{1}{W}\mathbb{E}\{\mathscr{F}_W[x(t)]\mathscr{F}_W[x(t+\tau)]^*\} \tag{3}$$

where $\mathscr{F}_W[x(t)]$ stands the Fourier transform of the signal $x(t)$ over a finite time duration of W. Processing the CSC results in the bi-variable map which reveals the hidden modulations, making it a robust tool for

detecting the cyclostationarity in vibration signals [8, 9].

In order to minimize uneven distributions, a whitening operation can be applied to the CSC. This extended tool, named the Cyclic Spectral Coherence (CSCoh), describes the spectral correlations in normalized values between 0 and 1, and is defined as in Eq. 4 :

$$CSCoh(\alpha, f) = \frac{CSC_x(\alpha, f)}{\sqrt{CSC_x(0, f)CSC_x(0, f+\alpha)}} \tag{4}$$

Both the CSC and the CSCoh bi-variable maps can be integrated along the spectral frequency axis in order to obtain a regular spectrum, resulting in a one dimension spectrum function of the $cyclicfrequency\alpha$. The band of spectral frequencies to be integrated can be defined as the full available band, from zero to the Nyquist frequency, resulting in a spectrum that exhibits all modulations present in the signal. On the other, the band can be defined as the one that maximizes the cyclic characteristic frequency of interest while minimizing the background noise and the other frequency components that may mask the frequency of interest. In this manner, the integration over a specific band on the bi-variable map can improve the detection rate of the characteristic frequency related to the present damage on the signal. The resulting spectrum is then named Improved Envelope Spectrum (IES) and it is obtained from the frequency-frequency domain according to Eq. 5 :

$$IES(\alpha) = \frac{1}{F_2 - F_1} \int_{F_1}^{F_2} |CSCoh_x(\alpha, f)| df \tag{5}$$

3 Proposed methodology

Diagnosis using the bi-variable maps requires a deep understanding of the map in order to exploit its information. Analysis of one dimensional spectra is far more widely applied in the academia and industry and easier to anaylise. It has been seen that integration of the bi-variable function along its spectral variable results in a one dimensional spectrum, which would be a good tool on itself for diagnostics purposes. On the other hand, the diagnostics information could still be masked under the noise and other components signatures. Integration of the specific band that carries the signal of interest can further enhance the spectrum and increase the performance in the detection of the frequencies of interest. The detection of the optimal band of integration on the bi-variable map is not always straight forward.

The Improved Envelope Spectrum via Feature Optimization-gram (IESFOgram)[11] is one such band selection tool to be applied in the bi-variable map as placed in the scheme depicted Fig. 1.

The proposed method tries to optimize a Normalized Diagnostic Feature (NDF) based on the cyclic characteristics of interest (e.g. rolling element bearing characteristic fault frequencies/orders) on the demodulated spectrum resulting from the integration of the bi-variable map. The method is thought to be general enough to be applied to either the CSC or the CSCoh. The scheme representing the IESFOgram procedure and the extraction of the NDF is shown in Fig. 2, and step-by-step details for its extraction as described as follows.

Step 1: In the first step, the bi-variable map is extracted from the signal. The estimators of the $CSC(\alpha, f)$ can be based on the Averaged Cyclic Periodogram, Cyclic Modulation Spectrum or any other numerical method [14] to extract the CSC bi-variable map previously described in Eq. 3. The CSC can also be in its normalized version $CSCoh(\alpha, f)$. The user can define on its own discretion which method to use in order to obtain the bi-variable map $CSC(\alpha, f)$. The reader is forwarded to the references [15, 16], suggested as providers of the numerical implementation of the CSC in the Order-Frequency domain, and the reference [8] to the Frequency-Frequency domain CSC, if it is applied to the order tracked signal as a function of angle results in the Order-Order domain CSC.

Step 2: The next step consists in dividing the map along the spectral axis f according to the $1/3$-binary tree that is also applied to the Fast Kurtogram [3]. Each is defined by a series with a decreasing bandwidth bw and incremental steps of center frequency cf which define the upper and lower limit f_1 and f_2 described in the integration of Eq. 5. Each band is then integrated and results in a demodulated spectrum $IES_{cf,bw}(\alpha)$.

Step 3: From each processed $IES_{cf,bw}(\alpha)$, one Diagnostic Feature $DF(cf, bw)$ is extracted. This feature is based on the cyclic fault frequency/order of interest. Therefore, to calculate this feature, as well as the IESFOgram, this cyclic component needs to inserted into the method as input. The feature $DF(n)$ is defined as the sum of the N-harmonics of the characteristic fault frequency/order α_{fault} normalized by the noise level estimated in a bandwidth $2 \times f_b$, as described in Eq. 6.

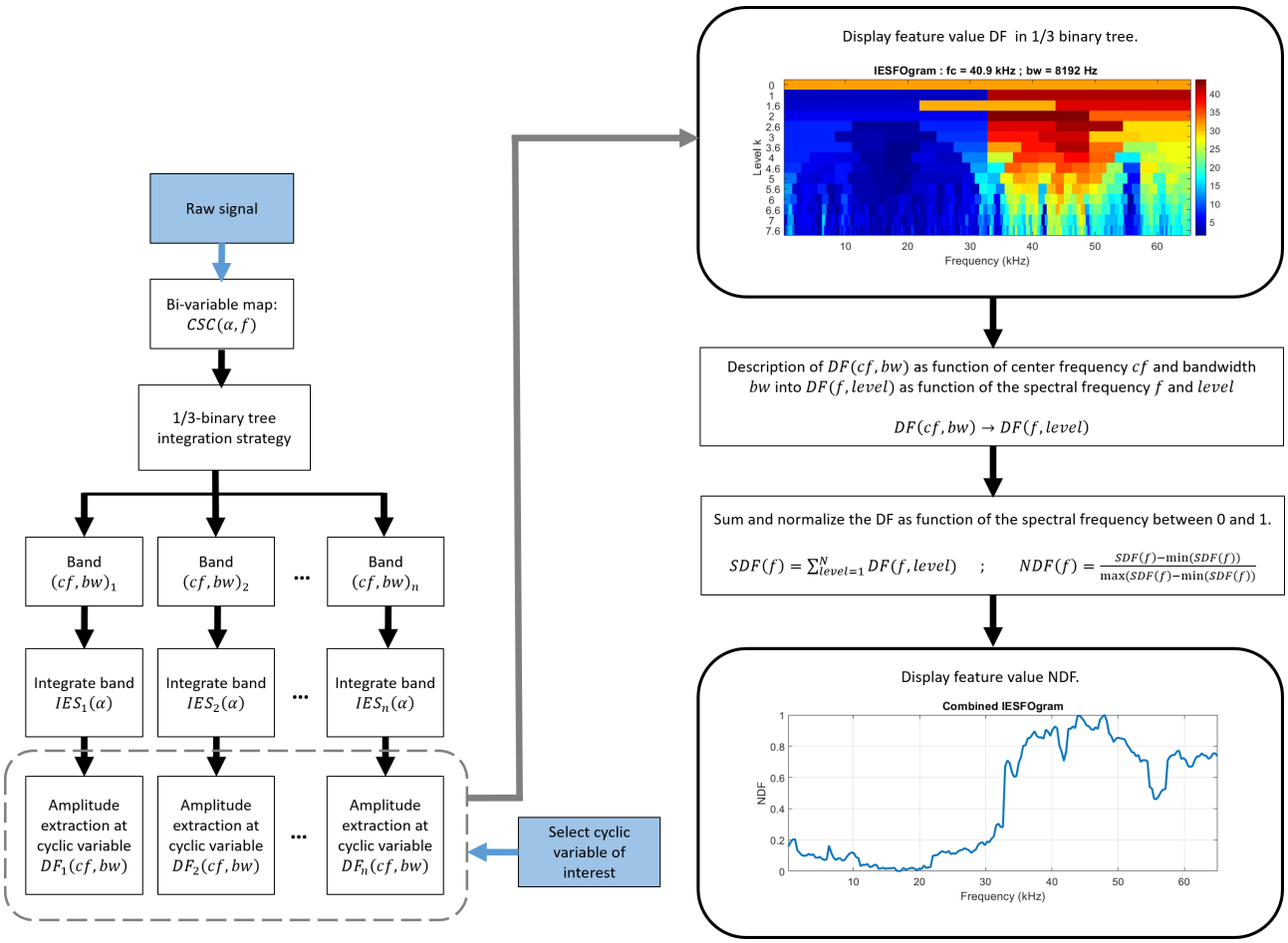


Figure 2: Schematic description of the IESFOgram procedure for extraction of CIES.

$$DF(cf, bw) = \sum_{k=1}^{N} \frac{IES_{cf,bw}(k \times \alpha_{fault})}{\frac{1}{2f_b}\left[\int_{kf_{fault}-f_b}^{kf_{fault}+f_b} IES_{cf,bw}(\alpha)d\alpha - IES_{cf,bw}(k \times \alpha_{fault})\right]} \tag{6}$$

The normalization procedure is important to be taken into account. This is due to some bands having high peak values of noise, and the direct absolute value at the fault frequencies can be higher than at the optimal band. Normalizing with the background noise level at the peaks solves this problem, making high values of the DF to correspond to bands where the frequency peaks of interest are present.

Step 4: The objective of this step is to find the the weight value to be applied on each band for integration of the bi-variable. To quantify the presence of a cyclic component in each band, the library of features $DF(cf, bw)$ is used. The higher the value of DF, the higher the presence of the component of interest is present. Thus, the optimal band OB is identified as the arguments which maximizes $DF(cf, bw)$, as described in Eq. 7.

$$OB = \arg\max_{cf,bw}[\, DF(cf, bw)\,] \tag{7}$$

Colormap presentation of the values of DF as function of (cf, bw) in a 1/3-binary tree is called the IESFOgram, and its maximum value corresponds to the selected optimal band for integration.

Step 6: The representation of $DF(cf, bw)$ as function of center frequency cf and bandwidth bw is transformed to a representation function of the spectral frequency f and the $level$, as $DF(f, level)$. To be used as a proper weight on the bi-variable map, the $DF(cf, bw)$ is summed along its level ($SDF(cf, bw)$), and then normalized between 0 and 1 ($NDF(cf, bw)$):

$$SDF(f) = \sum_{level=1}^{N} [\, DF(f, level)\,] \tag{8}$$

$$NDF(f) = \frac{SDF - min(SDF)}{max(SDF) - min(SDF)} \tag{9}$$

Step 7: The final step is to integrate the bi-variable map along the spectral axis on and weight each band with normalized diagnostic feature $NDF(f)$ in order to obtain the Combined Improved Envelope Spectrum $CIES(\alpha)$. This step can be considered not be part of the IESFOgram procedure, but as the extraction of the CIES with highest Signal-to-Noise Ratio (SNR) for diagnostic purposes.

$$CIES(\alpha) = \sum_{f=0}^{fs/2} CSCoh(\alpha, f) * NDF(f) \tag{10}$$

To define if the extracted amplitude value on the spectra of the cyclic component are statistically relevant, a threshold is also calculated, and visualized on all spectra of this paper. The threshold is the same as the one presented by the authors in [13], based on 3 times the Moving Absolute Deviation (MAD) of its spectra. The window defined on all spectra corresponds to the total number of samples of each corresponding spectrum divided by 2^7. All values above the threshold are considered to be statistically relevant for detection of the frequency.

4 Experimental application and results

In order to test and validate the proposed methodology, vibration data captured from two separate test rigs are used as case studies. One dataset corresponds to damage on the roller of a bearing under variable speed and load conditions. The second dataset corresponds vibration signals with damage in the outer race of the bearing with high electromagnetic interference present in the signals.

4.1 Case 1 - Roller damage under variable speed and load conditions

The first studied case of this paper is performed on the rolling bearing testing rig developed at the DIRG lab of Politecnico di Torino, where a high-speed spindle drives a shaft supported by a couple of identical bearings. Different damage conditions have been imposed on one of the bearings, and accelerations have been recorded in different positions and directions, under various load and speed conditions. The evaluation of the local damage on a roller has also been evaluated by monitoring the bearing under the same speed and load conditions for about 230 hours. A selection of the acquired records can be downloaded from `ftp://ftp.polito.it/people/DIRG_BearingData/` [10] .

The test rig is composed of three bearings (B1, B2 and B3) mounted inline on the output shaft of a high-speed spindle motor and a precision sledge applying radial load on the bearing on the middle, as depicted in Fig. 3.

The speed of the spindle is set through the control panel of an inverter but can not be actively controlled: not only the spindle has no keyphasor transducer or tachometer to detect its actual speed but also there is no feedback to the controller of the inverter. As a direct consequence, the actual speed of the shaft is always lower than the ideal one and the difference increases with the applied load. A static load cell allows for measuring the resulting force, whose direction is purely radial.

The main geometrical properties of the three bearings, specifically manufactured for this high speed aeronautical applications, are listed in Table 1.

Bearing reference	Pitch diameter D (*mm*)	Rollers diameter d (*mm*)	Rolling elements
*B*1 & *B*3	40.5	9.0	10
*B*2	54.0	8.0	16

Table 1: Main properties of the roller bearings [10].

The dataset corresponding to the variable conditions has 7 cases: three with indentation on the inner ring of the bearing; three with indentation on a roller of the bearing; and one with no damage (healthy). Each damage corresponds to indentation damages diameter of $450, 250$ and $150\mu m$. The case depicted in this paper

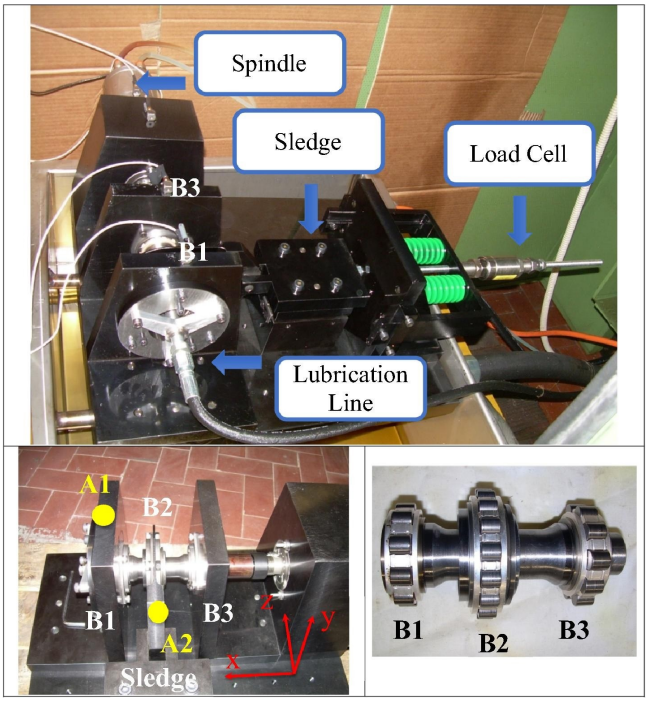


Figure 3: Test rig (up) general view, (left) position of the two accelerometers, (right) shaft with the three accelerometers [10].

corresponds to the damage on the roller with a diameter of $150\mu m$. Each case contains signals acquired during 20 seconds at 51.2 kHz under speeds of: 100, 200, 300, 400 and 500 revolutions per second. Furthermore, the acquired signals are under 4 radial load conditions: 1000 N, 1400 N, 1800 N, and no load. The damaged bearing corresponds to bearing B1, and its characteristic frequencies under the different speed conditions are described in Table 2.

Motor speed (Hz)	FTF (Hz)	2xBSF (Hz)	BPFO (Hz)	BPFI (Hz)
100	38.9	427.8	388.9	611.1
200	77.8	855.6	777.8	1222.2
300	116.7	1283.3	1166.7	1833.3
400	155.6	1711.1	1555.6	2444.4
500	194.4	2138.9	1944.4	3055.6

Table 2: Characteristic frequencies of bearing B1 under different steady speed conditions.

From the two triaxial accelerometers, the signals used to diagnose the roller damage are on the radial direction of the accelerometer mounted on the damaged bearing B1, or in other word, the radial output of accelerometer A1 is used.

The speed reference is provided along with the signals, but the description details that the speed is in reality lower than the provided one. Upon analyzing the signal in the bivariable map from the CSCoh, it is defined that the shaft speed frequency is clear at around 90% of the theoretical one and can be extracted from the spectra under null applied load. Figure 4 shows the CSCoh map exemplifiying the clear shaft frequency at 288 Hz, for the case of speed 300 Hz and a load of 0 N.

When load is applied to the bearing, the FTF harmonics of bearing B1 become the prominent ones, as can be seen from the CSCoh map in Fig. 5 for the case of speed 300 Hz and a radial load of 1000 N.

These peaks at 90% of the given speed and the FTF were used as reference for the real speed of the test rig. With this, the frequencies at the 2xBSF can be correctly defined to determine if they are indeed present in the spectra.

The frequency at 2xBSF and its two next harmonics are then used as inputs to calculate the IESFOgram, and the weight of the different bands that will enhance the peak extraction at those frequencies, as shown in Fig. 6.

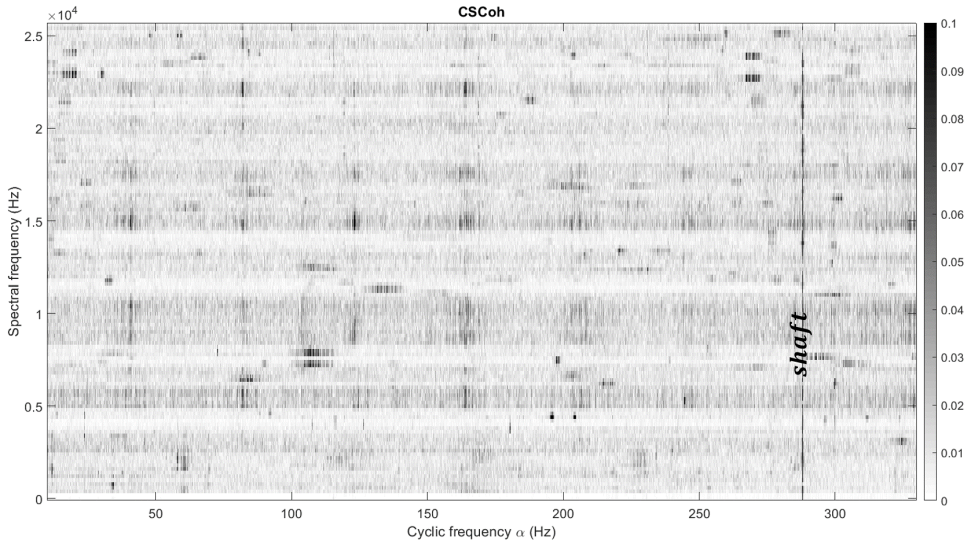


Figure 4: CSCoh bi-variable map around the first 3 FTF harmonics of the damaged bearing and shaft frequency speed of signal under no radial load.

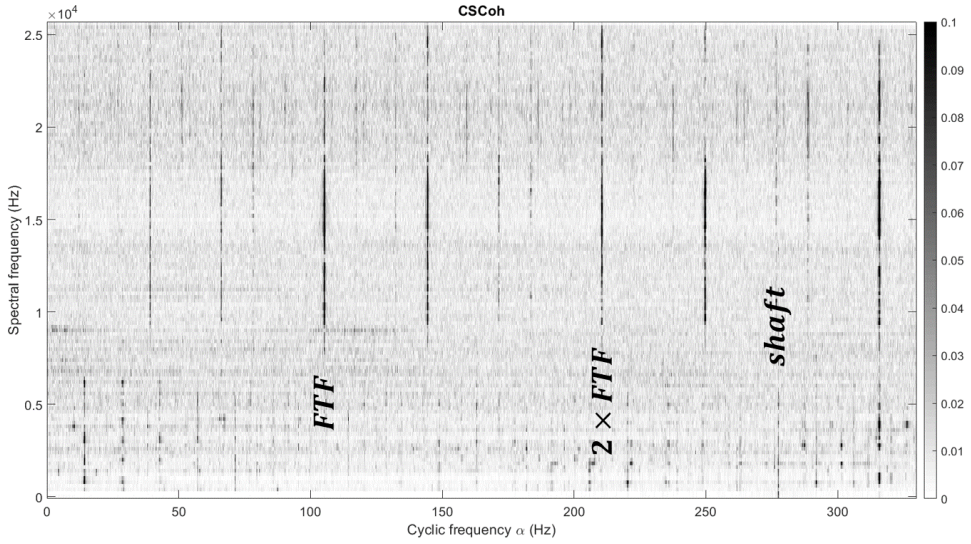


Figure 5: CSCoh bi-variable map around the first 3 FTF harmonics of the damaged bearing and shaft frequency speed of signal under 1000 N of radial load.

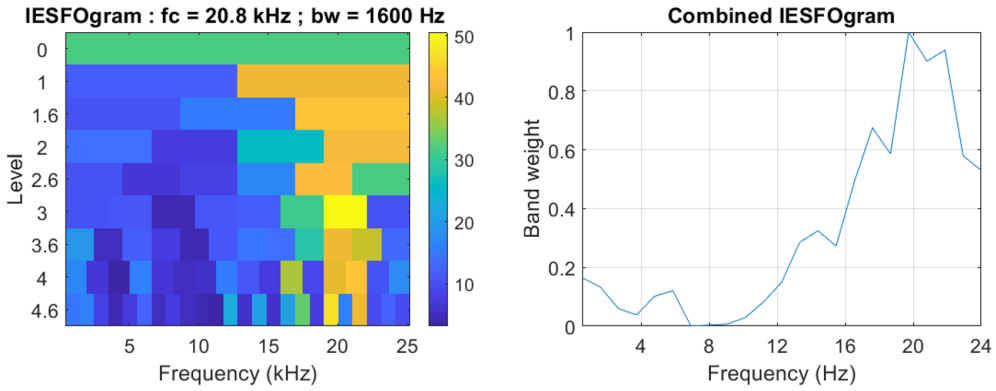


Figure 6: Signal under 300 Hz speed and 0 N load: (left) IESFOgram, (right) Combined IESFOgram.

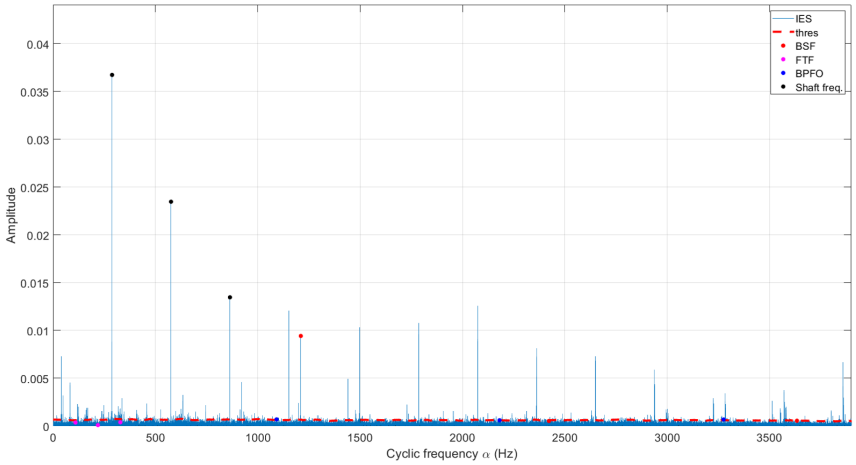


Figure 7: CIES for the signal under 300 Hz speed and 0 N load.

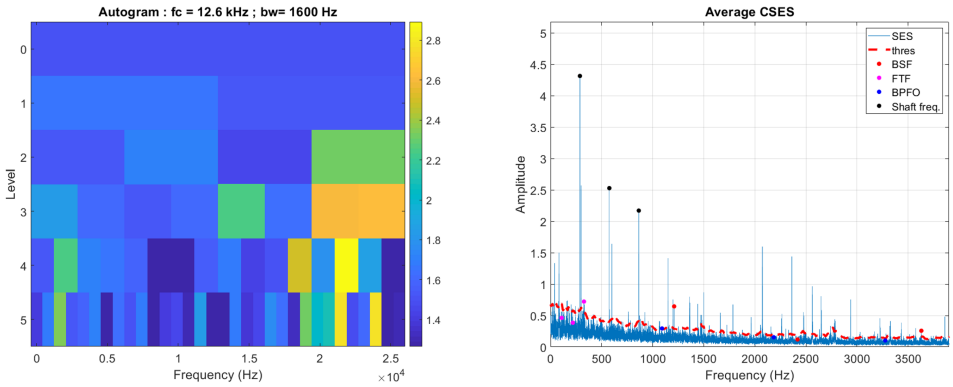


Figure 8: CSES for the signal under 300 Hz speed and 0 N load.

The IESFOgram shows high feature values at high frequencies, mainly around 20 kHz. Applying the resulting weighted function (Combined IESFOgram) to the integration of the CSCoh, the 2xBSF can be clearly extracted from the CIES seen in Fig. 7.

The Autogram applied to the same signal has a similar result, with high feature values for around high frequency bands. The Combined SES based on the Autogram is successful in extracting the peak at the 2xBSF above the threshold, along with the shaft frequency harmonics, as shown in Fig. 8.

The Fast Kurtogram detects the center frequency 10.1 kHz with a bandwidth of 1067 Hz with the highest kurtosis. The resulting SES based on the Fast Kurtogram does not provide any valuable diagnostic information, as illustrated by Fig. 9.

Following to the case of radial load of 1000 N at 300 Hz of shaft speed, the IESFOgram shows high feature values around 12 kHz. Applying the weighted function based on the IESFOgram, the 2xBSF and its harmonics are clearly detected from the Combined IES seen in Fig. 10.

For this case, the radial increases the impulses due to the roller damage, making it easier to detect the fault frequencies. The harmonics of the FTF that modulate the roller characteristic frequency also became dominant under the radial load.

The Autogram applied to the same signal has a similar result, with high feature values for around high frequency bands. The CSES based on the Autogram is also successful in extracting the peak at the 2xBSF above the threshold, as well as the FTF harmonics, as shown in Fig. 11.

In the cases with high load applied, the FTF and 2xBSF harmonics are so dominant on the signals that both the FK-based SES as well as the classical SES with no filtering detect these above the threshold. As a final remark, for the other speeds, the same pattern is found, where in high radial loads the damage related frequencies are dominant while in the no load cases they are masked in the noise level. For avoiding redundancy

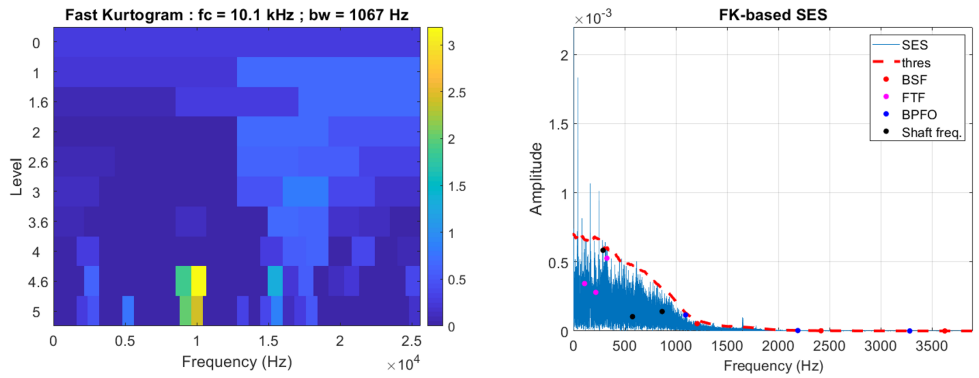


Figure 9: FK-based SES for the signal under 300 Hz speed and 0 N load.

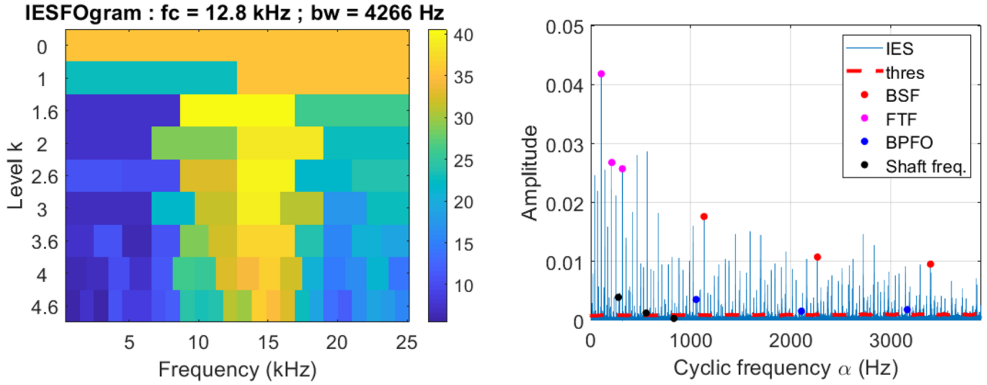


Figure 10: (left) IESFOgram and (right) CIES for the signal under 300 Hz speed and 1000 N of radial load.

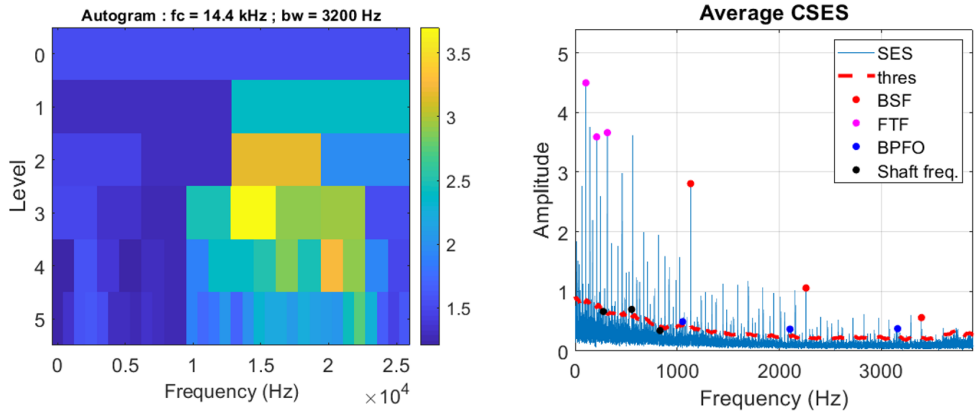


Figure 11: (left) Autogram and (right) CSES for the signal under 300 Hz speed and 1000 N of radial load.

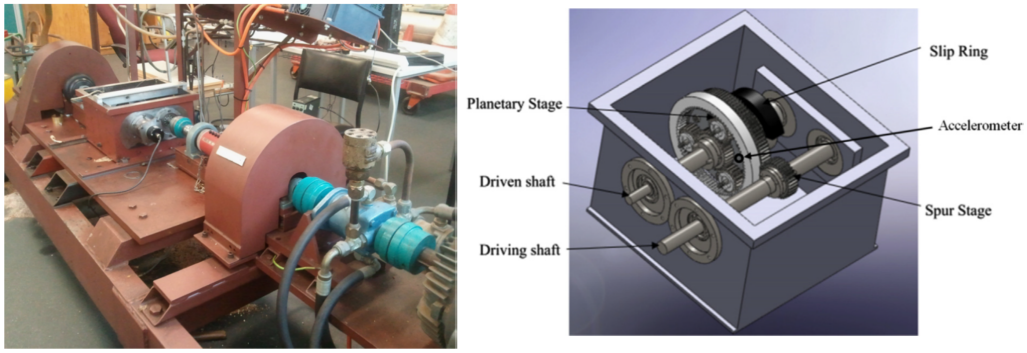


Figure 12: Test rig (left) inside view, (right) schematic representation [4].

on the results, these figure results are not shown here.

4.2 Case 2 - Outer race damage under electromagnetic interference (EMI)

The following case vibration data was acquired from the planetary gearbox test rig with applied torque is presented in Fig. 12. The gearbox torque is provided by a hydraulic system driven by a three-phase induction motor. A torque transducer is attached to measure the applied torque on the gear set while the speed of the driving shaft is controlled by a VFD. The gear ratio of the planetary stage is 1:3 (speed up) consisted of a 40 tooth sun gear, three 20 tooth planetary gears and a 80 tooth ring gear. The planet carrier is the input of the planetary stage, the sun gear is the output while the ring gear is fixed. The overall transmission ratio of the test rig is approximately 1:1 as an initial 90:32 reduction stage is attached. An accelerometer is mounted on the planet carrier to measure acceleration in the axial direction focusing towards the investigation of internal vibration measurements. The vibration signal is finally transmitted to the signal conditioner by the use of a slip ring. Faults have been seeded in the inner race and the outer races of the planet gear bearings using spark erosion. needle roller bearings are used, containing 15 rollers of 2 mm diameters and a pitch diameter of 18 mm. The depth of the faults is 0.4 mm while the width is 1.2 mm for the outer race and 1.0 mm for the inner race respectively. The measurements have been realised for each type of defect (inner and outer race) at a constant input shaft speed of 6 Hz for three torque loads 30, 50 and 70 Nm. The sampling frequency has been selected equal to 131,072 Hz, the switching frequency of the VFD was set at 14 kHz, and the control frequency of the VFD was 24 Hz (giving a nominal 6 Hz shaft speed for the 8-pole motor). The PWM carrier and the PWM message are equal respectively to 14000 Hz and 24 Hz. Based on the geometry of the bearing and its speed, the BPFO is equal to 55 Hz and the BPFI is equal to 69 Hz, but only the BPFO case is demonstrated in this paper.

Applying the Autogram to the signal with outer race damage and with a applied torque of 50 Nm, the band with the maximum Kurtosis with center frequency of 44 kHz and a bandwidth of 2048 Hz is selected as the optimal. This is the correct band with the carrier of the outer race damage. However, the impulsive bands of the carrier of the EMI have also high values of kurtosis, and the resulting CSES shows the peaks at noise level and bellow the threshold, as shown in Fig. 13.

This is the principal obstacle in detecting bearing damage with EMI noise, as both signatures have a impulsive nature which are represented with high kurtosis. The Fast Kurtogram selects the bands related to EMI as the ones with the highest kurtosis, as it as been concluded on this signal by the authors in [4, 12]. In this case, the SES based only on one band with the maximum kurtosis level of the Autogram would provide better performance.

The IESFOgram and its CIES are presented in Fig. 14, and the peaks of the BPFO are seen to be detect with clarity, well above the threshold. The case of BPFI not shown here the IESFOgram extracts the peaks of BPFI, however the CSES version also allows detection of fault.

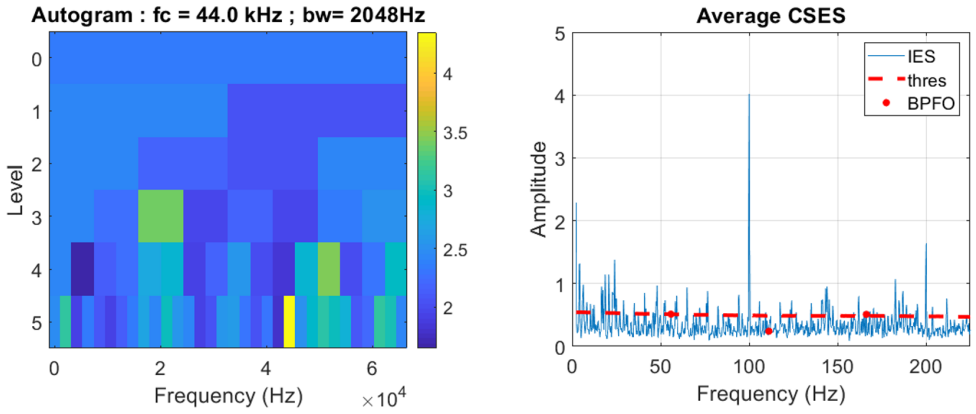


Figure 13: (left) Autogram and (right) CSES for the signal under 50 Nm of torque.

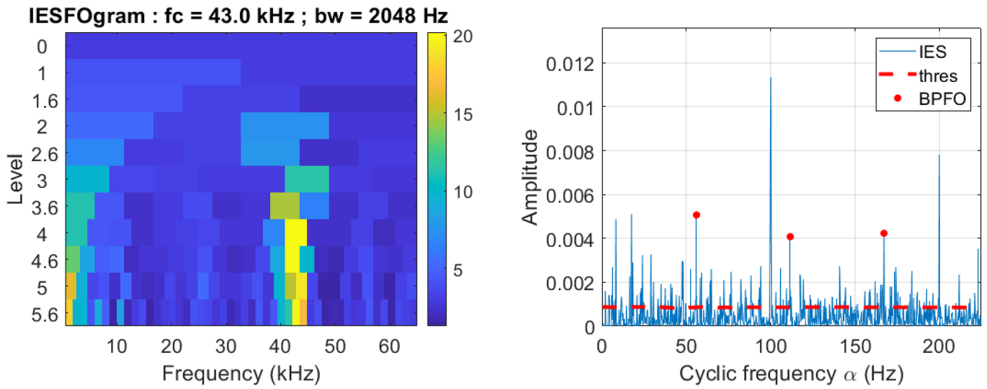


Figure 14: (left) IESFOgram and (right) CIES for the signal under 50 Nm of torque.

5 Conclusion

This paper proposes a new method for demodulation on multi-bands of frequencies using the bi-variable maps based on the Cyclic Spectral Correlation and Coherence. Initially, the IESFOgram related to the damaged bearing characteristic frequency is calculated in order to extract the spectral frequency bands that have high information content on the fault. The Normalized Diagnostic Feature as a function of frequency is defined as the normalized sum of feature values of the IESFOgram. The last step is to perform a multi band integration with the enhanced bands with high Normalized Diagnostic Feature values, and finally obtaining the Combined Improved Envelope Spectrum that allows a correct bearing diagnosis.

The methodology shows good performance in detecting the characteristic frequencies when they are present on the signals, by enhancing the bands with more relevant information on the fault and performing a multi band integration with higher weigthed values on those bands. The method was compared mainly with the Combined Squared Envelope Spectrum based on the Autogram, which also showed higher performance than the Fast Kurtogram on the detection of the bearing characteristic frequencies by also performing a multi band filtering procedure. When the signals contain a high impulsive signal with the same nature as a bearing damage, such as electromagnetic interference, the CSES-based on the Autogram shows to keep the peak frequencies of interest bellow the noise level in case of high impulsive noise spread over several bands. The CIES shows to be able to extract clearly the fault frequency peaks above the noise level while ignoring the bands dominated by the impulsive noise. This is shown to be the case because the IESFOgram is targeted at the frequencies of interest, while the Autogram is blind. Indeed, the main disadvantage of the IESFOgram compared to the Autogram and the Fast Kurtogram is that it is not a blind method.

The methology was tested and validated on two experimental datasets: the first under variable speed and load condition, and the second on signals with high electromagnetic interference contamination. The results show the method can diagnose with confidence bearing damages on vibration signals.

Acknowledgements

The authors would like to thank Ali Moshrefzadeh and Alessandro Fasana for the sharing the roller bearing damage vibration data and the Autogram publicly, Jerome Antoni for sharing the Fast Kurtogram publicly, and Wade Smith and Robert Randall for providing vibration signals with electromagnetic interference. K. Gryllias gratefully acknowledges the Research Fund KU Leuven.

References

[1] Randall, R., Antoni, J., *Rolling Element Bearing Diagnostics - A Tutorial*, Mechanical Systems and Signal Processing, 2012, pp.485-520 (25).

[2] Feldman, M., *Hilbert transform applications in mechanical vibration*, John Wiley and Sons, 2011.

[3] Antoni, J., *Fast computation of the kurtogram for the detection of transient faults*, Mechanical Systems and Signal Processing, 2007, pp.108-124 (21).

[4] Smith, W., Fan, Z., Pheng, Z., Li, H., Randall, R., *Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference*, Mechanical Systems and Signal Processing, 2016, pp.371-394 (75).

[5] Tse, P. W., *The sparsogram: A new and effective method for extracting bearing fault features*, Prognostics and System Health Managment Confernece, 2011, pp.1-10).

[6] Moshrefzadeh, A., Fasana, A., *The Autogram: An effective approach for selecting optimal demodulation band in rolling element bearings diagnosis*, Mechanical Systems and Signal Processing, 2017, pp.294-318 (105).

[7] Randall, R., Antoni, J., Chobsaard, S., *The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals*, Mechanical Systems and Signal Processing, 2001, pp.945-962 (15).

[8] Antoni, J., Xin, G., Hamzaoui, N., *Fast computation of the spectral correlation*, Mechanical Systems and Signal Processing, 2017, pp.248-277 (92).

[9] Antoni, J., *Cyclic spectral analysis in practice*, Mechanical Systems and Signal Processing, 2007, pp.597-630 (21).

[10] Daga, A. P., Fasana, A., Marchesiello, S., Garibaldi, L., *The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data*, Mechanical Systems and Signal Processing, 2019, pp.252-273 (120).

[11] Mauricio, A., Smith, W., Qi, J., Randall, R., Gryllias, K., *Cyclo-non-stationarity based bearings diagnostics of planetary bearings*, CMMNO, Santander, Spain, 2018, pp.1-10.

[12] Mauricio, A., Qi, J., Sarrazin, M., Janssens, K., Smith, W., Randall, R., Gryllias, K., *Cyclostationary-based bearing diagnostics under electromagnetic interference*, ICSV25, Hiroshima, Japan, 2018, pp.1-8.

[13] Antoni, J., Bonnardot, F., Raad, A., El Badaoui, M., *Cyclostationary modelling of rotating machine vibration signals*, Mechanical Systems and Signal Processing, 2004, pp.1285-1314 (18).

[14] Borghesani, P., Antoni, J., *A faster algorithm for the calculation of the fast spectral correlation*, Mechanical Systems and Signal Processing, 2018, pp.113-118 (111).

[15] Abboud, D., Baudin, S., Antoni, J., Remond, D., Eltabach, M., Sauvage, O., *The spectral analysis of cyclo-non-stationary signals*, Mechanical Systems and Signal Processing, 2016, pp.280-300 (75).

[16] Abboud, D., Antoni, J., *Order-frequency analysis of machine signals*, Mechanical Systems and Signal Processing, 2017, pp.229-258 (87).

Data Mining Classification & Machine Learning methods

Deep Learning Protocol for Condition Monitoring & Fault Identification in a Rotor-Bearing System from Raw Time-Domain Data

Sonkul Nikhil Avinash Singh Jasdeep Vyas Nalinaksh. S.

Department of Mechanical Engineering
Indian Institute of Technology, Kanpur 208016, India
nasonkul@gmail.com, vyas@iitk.ac.in

Abstract

A Deep Learning protocol is developed for identification of typical faults occurring in rotating machinery. In past Support Vector Machines (SVMs), Clustering, Artificial Neural Networks (ANNs) and other algorithms have been used for this purpose. However, these algorithms require the raw time-domain data, from the sensors on the machine, to be first processed and handcrafted into parameters like Fast Fourier Transform (FFT) coefficients, Statistical Moments, etc. before being fed as inputs. ANN with back-propagation is a popular algorithm in many domains. It however suffers from the vanishing gradient problem and not adequate enough when subjected to raw sensor response as input. Deep Learning or Deep Neural Network is understood as a form of neural network with a large number of layers. Convolutional Neural Network (CNN) architecture is commonly used in a deep neural network for image recognition. A Deep Learning CNN architecture has been developed, employing the analogy of an RGB image, to directly work upon the raw time-domain signals obtained from sensors on a rotor-bearing system. The analogous RGB channels are vibration data from different sub-systems of the complete rotor-bearing assembly. The Deep Learning Network effectively recognizes all kinds of faults that were investigated.

1 Introduction

Data collection, feature extraction and fault identification are typical steps in fault diagnosis of a mechanical system. Conventional approaches in fault diagnosis extract features from time and frequency domain of raw signals. Statistical parameters and Fast Fourier Transform coefficients are widely used features. Machine learning techniques are used with these handcrafted features to identify faults. Fuzzy Logic, Wavelets, Clustering, Decision Trees, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are some of the techniques used in past for fault identifications [1]–[7]. Statistical moments of raw time domain signal and its derivatives were used as features of ANN by Vyas [7] for fault diagnosis of rotor-bearing system. Efforts to automate the feature extraction using deep learning techniques are being made in recent studies [8]. Janssensa [9] used Discrete Fourier Transform coefficients as input to Convolutional Neural Network (CNN) for bearing fault classification. Features learnt from CNN network combined with time domain features were used by Xie [10] to train SVM model. Guo [11] trained a CNN network with input as continuous wavelet transform scalogram of rotor machinery.

ANNs are not adequate to abstract features from raw time domain data due to large dimension of the data and vanishing gradients. The method proposed in this study, directly works upon raw time domain data eliminating the pre-requisite of extracting features.

Deep Neural Network, like CNN is capable of learning abstract features from large and multi-dimensional data like images and audio signals. CNN preserves the topology of the input and has lesser number of learning parameters than a neural network of the same depth, which also makes learning faster.

2 Methodology and Training Data

CNN is chosen as primary model in this study. As a multi sensor model is necessary to identify faults in large systems, a Multi-channel Convolution Neural Network (McCNN) is incorporated in the proposed model.

2.1 Network Architecture

Input layer, Convolution layer, Pooling layer and Fully Connected layer form the building blocks of a CNN. The architecture is described in Figure 2.1. The input layer is a multi-dimensional array like an RGB image. Convolution is the dot product of a Kernel with a part of input or previous layer. The topology of a Kernel is smaller than previous layer. Kernels are weight parameters of CNN and they traverse over the input space and generate a feature map. A Kernel of size m x n after convoluting over input of size M x N will result into feature map of size (M-m+1) x (N-n+1). Number of Kernels and Size of Kernel are the hyperparameters for particular convolutional layer. The Feature Map is then activated by an Activation Layer. ReLU Activation Function is used in CNN as it reduces vanishing gradients. The Feature Map is sometimes padded with zeroes to control the topology of further feature maps. Batch Normalization Layer is added between Convolution and ReLU Activation Layer. It normalizes the feature map after Convolution layer, firstly by subtracting the mini-batch mean from each of its inputs and then dividing by their standard deviation for each channel; and secondly by scaling the new obtained featured map by γ and then shifting by β, where γ and β are learnable parameters. Batch Normalization accelerates deep network training [12]. Pooling layer summarizes the response over a neighbourhood. It reduces the output size and makes features invariant to small input noise. Max Pooling, Average Pooling are widely used pooling operations. The size of the region is the hyperparameter. Finally, softmax function is used for fault classification.

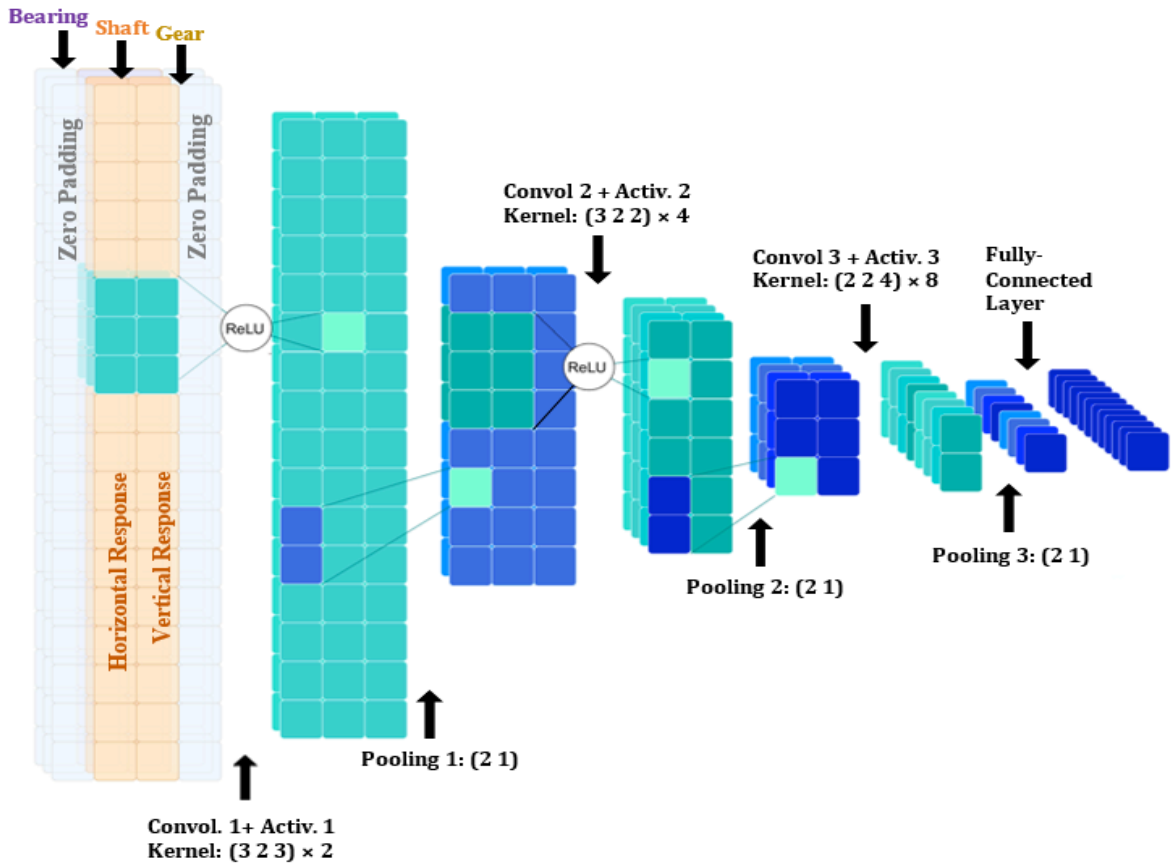


Figure 2.1: CNN Architecture: Input matrix is made with responses from three sub-systems corresponding to three channels with two sensors response (horizontal & vertical) in each channel.

2.2 Experiment Setup & Training Data

Experiments were performed by Jasdeep Singh [13] on a rotor test rig (Machine Fault Simulator Figure 2.2), which consists of a shaft supported in two roller bearings and driven by a DC motor. A flexible coupling is used to connect rotor shaft to that of motor. At one end of the shaft there is a sheave, which is connected to a reciprocating mechanism through a belt drive and a gearbox. Two discs mounted on the shaft.

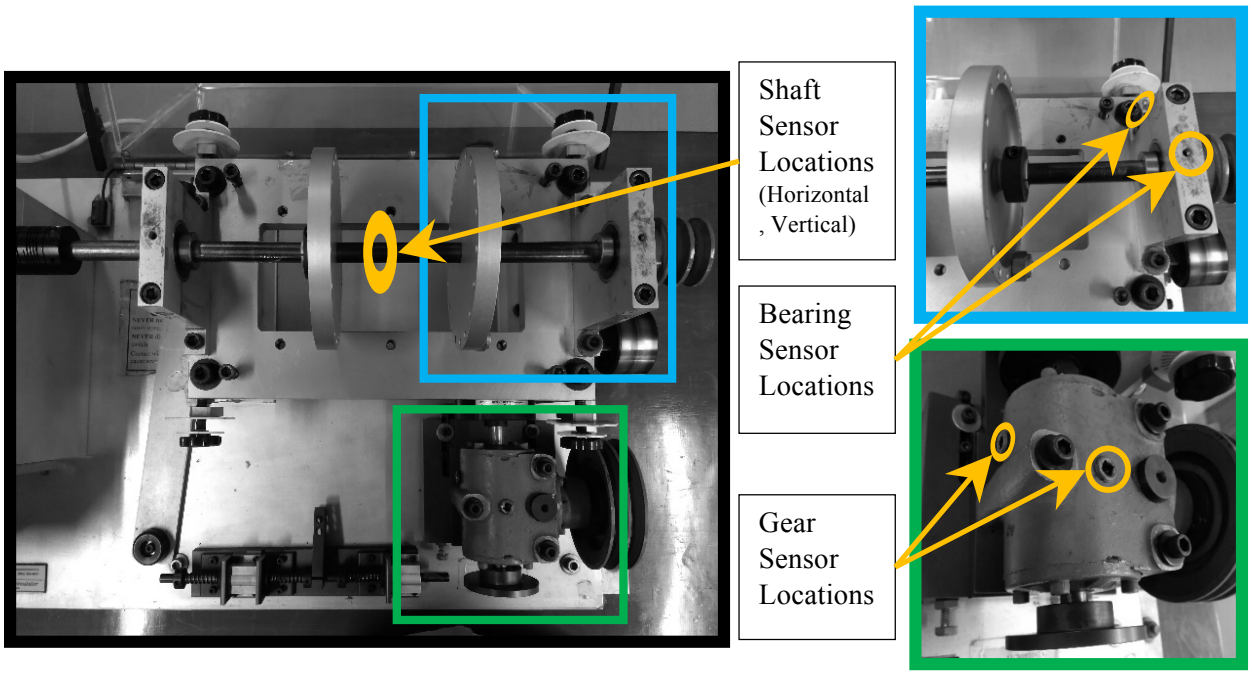


Figure 2.2: The rotor system (Machine fault Simulator) and sensor locations

The set-up comprises subsystems - (i) Gearbox (ii) Bearing and (iii) Shaft-Rotor Disc. Accelerometers were used at (i), (ii) and Proximity pickups at (iii). Each location has two sensors oriented mutually perpendicular to each other (horizontal & vertical). Data was collected at 2560 samples/sec. For an experiment data is collected for duration of 1.6 sec i.e. 4096 sample points in time domain. Variety of faults (Table 2.1) were introduced at a constant operating frequency of 40 Hz. Typical vibrations sensed for one of the faults introduced in the rotor are shown in Figure 2.3. Twenty such signals were collected for each sensor for the fault introduced.

	Faults (Set A)
1	No Fault with Attached Belt
2	Loose Belt
3	Tight Belt
4	Missing Tooth
5	Loose Gear
6	Unbalance 6.82 gram
7	Eccentric Rotor
8	Cocked Rotor
9	Bearing Outer Race Defect
10	Bearing Inner Race Defect
11	Ball Spin Fault
12	Combined Bearing Fault

	Faults (Set B) *Combined with missing tooth*
13	Unbalance 6.82 gram
14	Eccentric Rotor
15	Cocked Rotor
16	Bearing Outer Race Defect
17	Bearing Inner Race Defect
18	Ball Spin Fault
19	Combined Bearing Fault
20	Loose Gear

Table 2.1: Type of faults in dataset (Set A). Extra combined faults and (Set A) form (Set B)

Twelve faults (Set A) are described in Table 2.1. Set B is a combination of Set A and combined faults. Set A has 240 and Set B has 400 experimental responses. The datasets are divided between - training (60%), validation (20%) and test (20%) datasets.

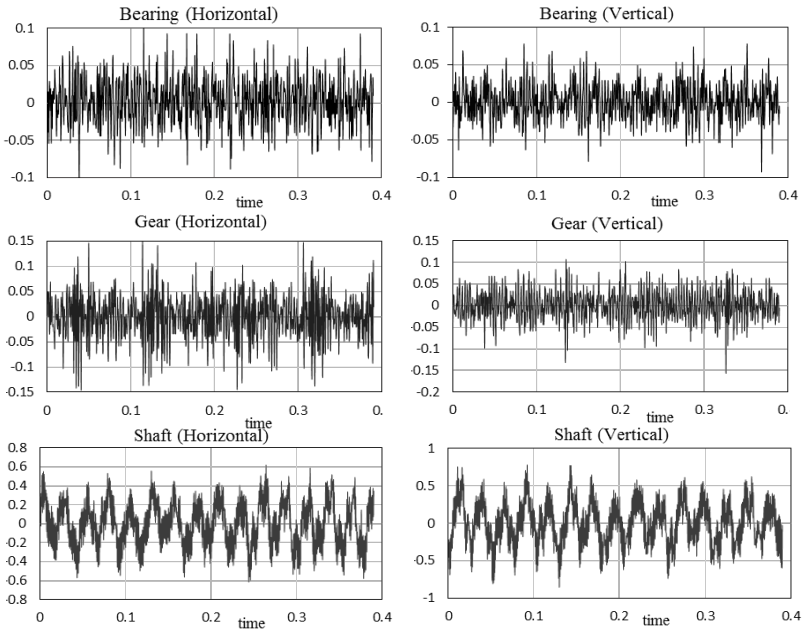


Figure 2.3: Typical sensor data

2.3 Network Topology

As described earlier (Figure 2.1), the response of all six sensors is mapped between normalised values of 0 and 1, on a McCNN, a multi-channel input similar to an RGB image. The two (horizontal and vertical) sensors of a sub-system comprise a channel and three such sub-systems complete the multi-channel input to the CNN. The topology of input, therefore is, 4096x2x3. Each channel is analogous to an RGB channel of a colored image. Each channel is padded with zero columns on both sides as shown in Figure 2.1. This input is passed down to further layers as listed in Table 2.2.

#	*Layer*	*Kernel Size/ Pooling Region*	*# of Kernels*	*Padding*
1	Input	-	-	-
2	Convolution 1	40 × 2	8	Yes
3	Batch Normalization + ReLU	-	-	-
4	Pooling 1	3 × 1	-	-
5	Convolution 2	20 × 2	16	Yes
6	Batch Normalization + ReLU	-	-	-
7	Pooling 2	4 × 1	-	-
8	Convolution 3	4 × 2	32	No
9	Batch Normalization + ReLU	-		-
10	Pooling 3	3 × 1	-	-
11	Convolution 4	4 × 2	32	No
12	Batch Normalization + ReLU	-	-	-
13	Pooling 4	2 × 1	-	-
14	Convolution 5	2 × 2	32	No
15	Batch Normalization + ReLU	-	-	-
16	Pooling 5	2 × 1	-	-
17	Convolution 6	2 × 1	32	No
18	Batch Normalization + ReLU	-	-	-
19	Pooling 6	2 × 1	-	-
20	Fully Connected	12 Neurons	-	-

Table 2.2: CNN architecture and hyperparameters

Network training is performed with stochastic gradient descent algorithm with momentum. The network was limited to a maximum of 800 iterations. Learning rate was 0.01. Mini-batch size was kept as 64 for Set A training and 128 for Set B. Both, max and average pooling were used in training for comparison.

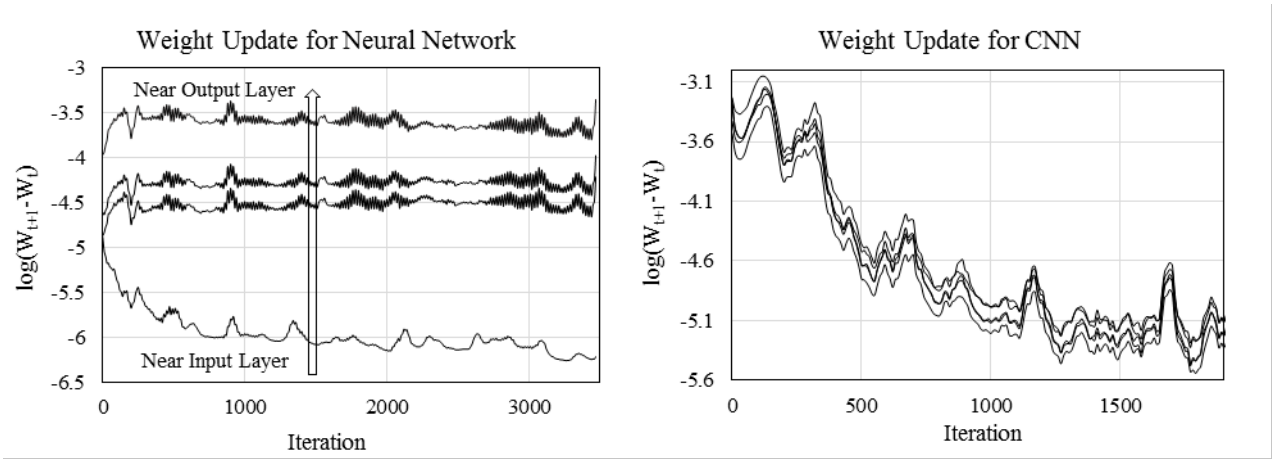
3 Training and Validation

Training of McCNN is performed, both, on Set A and Set B. The architecture classified all training, validation and test dataset successfully for both Set A (individual faults) and Set B (combined faults). The accuracy increases as we increase number of layers hence depth of network, (Table 3.1). Variation of accuracy with pooling function is also studied. Average Pooling network required less number of layers in comparison to Max Pooling to achieve zero error.

Network Depth	*Set A (Test Accuracy %)*		*Set B (Test Accuracy %)*	
	(Max. Pooling)	(Avg. Pooling)	(Max. Pooling)	(Avg. Pooling)
Layer [1 to 4] + 20	0.00	39.00	5.00	33.00
Layer [1 to 7] + 20	50.00	29.00	23.00	16.00
Layer [1 to 10] + 20	81.00	95.00	97.00	92.00
Layer [1 to 13] + 20	95.00	100.00	98.00	98.00
Layer [1 to 16] + 20	100.00	100.00	98.00	100.00
Layer [1 to 19] + 20	-	-	100.00	100.00

Table 3.1: Accuracy variation with depth and pooling function

A comparison between traditional Neural Network and McCNN is also made to study the weights updation with training iteration. From Figure 3.1 it can be seen that traditional Neural Network shows the effect of vanishing gradient, while in McCNN the weight updation of layers distant from output layer are of nearly same order compare to proximate layers. Complex and large systems can be diagnosed in McCNN with significant depth without facing vanishing gradient.



(a) traditional ANN (b) McCNN

Figure 3.1: Weight updation

It is seen that better results are obtained when training is carried out with data from all sub-systems (row number 4 in Table 3.2) in comparison to those obtained through training by isolated data from an individual sub-system (row numbers 1-3 in Table 3.2). This observation is visible for Set B, which contains more and combined faults, which increase the complexity of task, and underlines the fact that for fairly large and complex system we need sensor data from multiple critical locations.

Network Input	*Input Layer Topology*	*% Error (Set A)*			*% Error (Set B)*		
		Training	Validation	Test	Training	Validation	Test
Bearing	4096 × 2 × 1	0.00	0.00	0.00	0.00	7.50	3.75
Gear	4096 × 2 × 1	0.00	0.00	0.00	0.00	1.25	1.25
Shaft	4096 × 2 × 1	0.00	0.00	0.00	0.00	0.00	1.25
Bearing + Gear + Shaft	4096 × 2 × 3	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.2: Training with input from individual Sub-Systems and all sub-systems combined

Figure 3.2, shows features obtained after each ReLU activation layer, after they are transformed to two dimensions using t-SNE dimensionality reduction algorithm. The fault classification process can be seen to be consolidated at each ReLU layer where fault data points are getting clustered more compactly and distant from other faults. The distinction and separation between faults increases as we move toward output layer.

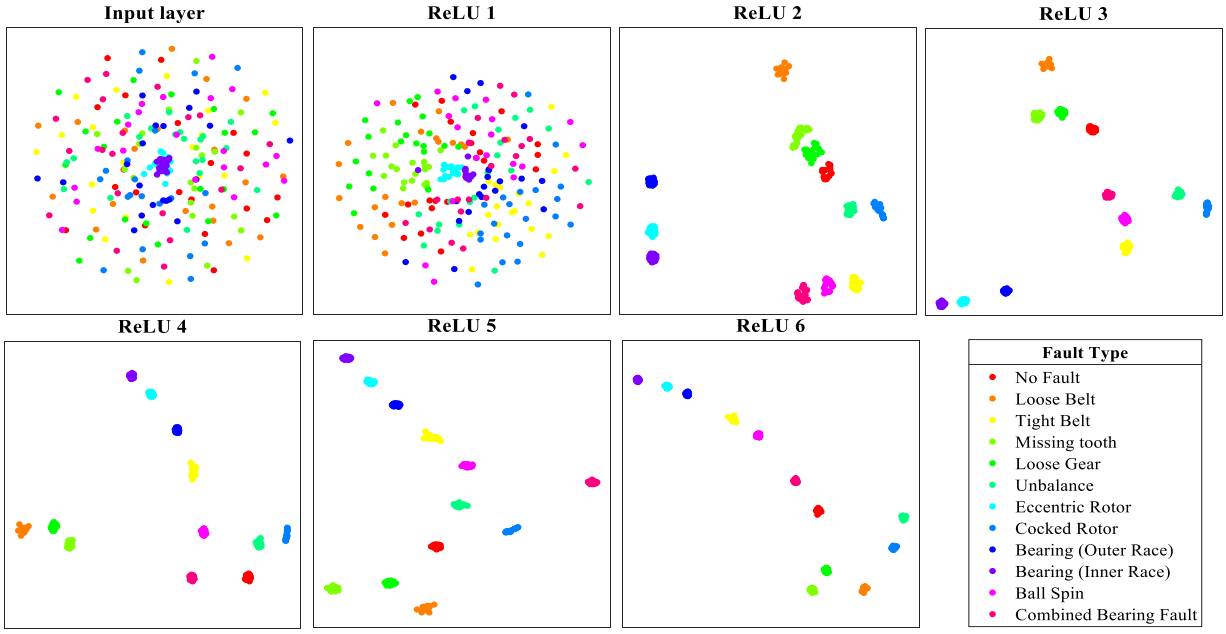


Figure 3.2: Feature map after each ReLU layer, plotted t-SNE (Set A).

4 Conclusion

The present study introduces an approach for fault identification of a rotor-bearing system using Convolutional Neural Network. The need to extract features like statistical parameter or to convert time domain data to FFT domain for training is eliminated in present study. Raw time domain sensor response is used without any pre-processing or feature engineering for constructing input layer. A generic method of fault diagnosis for big complex systems with distant and multiple subsystems is presented, which merges sensor responses from demanding subsystems into distinct channels of input matrix to form the input layer of McCNN architecture. Patterns in raw time domain data, difficult to be comprehended manually, are perceived by the McCNN network.

References

[1] M. Boumahdi, J. P. Dron, S. Rechak, and O. Cousinard, *On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree,* Expert Syst. Appl., vol. 37, no. 8, pp. 5887–5894, 2010.

[2] Y. Lei, Z. He, Y. Zi, and X. Chen, *New clustering algorithm-based fault diagnosis using compensation distance evaluation technique,* Mech. Syst. Signal Process., vol. 22, no. 2, pp. 419–435, 2008.

[3] R. Yan, R. X. Gao, and X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications,* Signal Processing, vol. 96, no. PART A, pp. 1–15, 2014.

[4] M. Systems, *Objective Machinery Fault Diagnosis Using Fuzzy Logic,* Mech. Syst. Signal Process., vol. 12, pp. 855–862, 1998.

[5] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, *Rolling element bearing fault detection in industrial environments based on a K-means clustering approach,* Expert Syst. Appl., vol. 38, no. 3, pp. 2888–2911, 2011.

[6] L. B. Jack and A. K. Nandi, *Support vector machines for detection and characterization of rolling element bearing faults,* Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci., vol. 215, no. 9, pp. 1065–1074, 2001.

[7] N. S. Vyas and D. Satishkumar, *Artificial neural network design for fault identification in a rotor-bearing system,* Mech. Mach. Theory, vol. 36, no. 2, pp. 157–175, 2001.

[8] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, *Deep Learning and Its Applications to Machine Health Monitoring: A Survey,* vol. 14, no. 8, pp. 1–14, 2016.

[9] O. Janssens et al. *Convolutional Neural Network Based Fault Detection for Rotating Machinery,* J. Sound Vib., vol. 377, pp. 331–345, 2016.

[10] Y. Xie and T. Zhang, *Fault Diagnosis for Rotating Machinery Based on Convolutional Neural Network and Empirical Mode Decomposition,* Shock Vib., vol. 2017, pp. 1–12, 2017.

[11] S. Guo, T. Yang, W. Gao, and C. Zhang, *A novel fault diagnosis method for rotating machinery based on a convolutional neural network,* Sensors (Switzerland), vol. 18, no. 5, 2018.

[12] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,* In ICML 2015.

[13] J. Singh, *Condition Monitoring of Geared Rotor Systems using Neural Networks,* Indian Institute of Technology Kanpur, 2006.

WIND TURBINE GEARBOXES FAULT DETECTION THROUGH ON-SITE MEASUREMENTS AND VIBRATION SIGNAL PROCESSING

Francesco Castellani[1], Davide Astolfi[1], Luigi Garibaldi[2], Alessandro Paolo Daga[2]

[1]University of Perugia, Department of Engineering, Perugia (Italy)
[2]Politecnico di Torino, Department of Mechanical and Aerospace Engineering, Torino (Italy)
francesco.castellani@unipg.it, davide.astolfi@unipg.it,
luigi.garibaldi@polito.it, alessandro.daga@polito.it

Abstract

Condition monitoring of gear-based mechanical systems undergoing non-stationary operation conditions is in general very challenging. In particular, this issue is remarkable as regards wind turbine technology because most of the modern wind turbines are geared and gearbox damages account for at least the 20% of unavailability time. For this reason, wind turbines are often equipped with condition monitoring systems (CMS), processing vibration measurements collected at meaningful subcomponents of the gearbox. In this work, a novel approach for the diagnosis of gearbox damages is proposed: the turning point is that vibration measurements are collected at the tower, instead that at the gearbox and can be performed also for machine not provided with specific CMS. This implies that measurement campaigns are quite easily performed and repeatable, also for wind turbine practitioners, and that there is no impact on wind turbine operation and power production. A test case study is discussed: it deals with a wind farm owned by Renvico, featuring 6 wind turbine with 2 MW of rated power each. The vibration measurements at two wind turbines suspected to be damaged and at reference wind turbines are processed through a multivariate Novelty Detection algorithm in the feature space. The application of this algorithm is justified by univariate statistical tests on the time-domain features selected and by a visual inspection of the dataset via Principal Component Analysis. Finally, the novelty indices based on such time-domain features, computed from the accelerometric signals acquired inside the turbine tower, prove to be suitable to highlight a damaged condition in the wind-turbine gearbox, which can be then successfully monitored.

1 Introduction

The diagnosis of gears and bearings faults of gearbox systems [1] is a very important topic, especially if the gear-based mechanical system of interest undergoes non-stationary operation conditions.

The technology of most of the modern wind turbines is based on the transformation of the slow rotor rotational speed (order of 10 revolutions per minute) into the fast generator rotational speed through a gearbox. It is estimated [2] that the unavailability time of a large wind turbine operating in an industrial wind farm is of the order of the 3% and at least the 20% of this quantity is due to gearbox damages. For this reason, therefore, the improvement in gearbox condition monitoring is a crucial step for the target of 100% availability of wind turbines. Therefore, commonly, megawatt-scale wind turbines are equipped with condition monitoring systems, elaborating the vibration measurements collected at meaningful sub-components of the gearbox.

Nevertheless, in the wind energy practitioners community, gearbox vibration data are often under-exploited because of the complexity of the analysis techniques that are required in order to interpret them. Therefore, often it happens that oil particle counting and operation data analysis (especially temperatures, as in [3]) are employed as condition monitoring techniques, despite they provide a late stage fault diagnosis, with respect to vibration analysis.

Therefore, two can be important direction targets as regards wind turbine gearbox condition monitoring through vibration analysis: on one hand, the precision of the diagnosis and on the other hand the simplicity of

the methods. As regards the former aspect, there are several recent studies. In [4], data mining algorithms and statistical methods are applied to analyze the jerk data obtained from monitoring the gearbox of a wind turbine: the failed stages of the gearbox are identified in time-domain analysis and frequency-domain analysis. In [5], the proposed techniques are based on three models (signal correlation, extreme vibration, and RMS intensity) and have been validated with a time-domain data-driven approach using condition monitoring data of wind turbines in operation. The results of that study support that monitoring RMS and extreme values serves as a leading indicator for early detection. In [6], the focus is on separating the bearing fault signals from masking signals coming from drivetrain elements like gears or shafts. The separation is based on the assumption that signal components of gears or shafts are deterministic and appear as clear peaks in the frequency spectrum, whereas bearing signals are stochastic due to random jitter on their fundamental period and can be classified as cyclo-stationary [7]. In [8], order analysis is individuated as a useful technique for condition monitoring the planetary stage of wind turbine gearbox. The approach takes advantage of angular resampling to achieve cyclo-stationary vibration signals and lessen the effects due to speed changes. In [9], the objective is condition monitoring of the planetary stage of wind turbine gearboxes: the proposed technique is resampling vibration measurements from time to angular domain, identification of the expected spectral signature for proper residual signal calculation and filtering of any frequency component not related to the planetary stage.

On the grounds of this brief literature survey, it arises that the techniques for the analysis of cyclo-stationary signals are the most employed for an accurate condition monitoring. The type and the quality of data that are requested for this kind of analysis confines the subject mainly to the scientific community and at present discourages the collaboration between industry and academia. Actually, the commercial condition monitoring systems, that are mostly adopted in most operating wind turbines exploited at the industrial level, record vibration measurements only when some trigger events occur and, most of all, don't stock the raw data (commonly, Fourier transforms and-or simple statistical indicators are stocked).

On these grounds, there is a growing demand of vibration-based gearbox condition monitoring techniques that could be easily repeatable, without impacting on the wind turbines operation (i.e. without intruding in the gearbox), and whose interpretation could be sufficiently simple and powerful. One remarkable study by this point of view is [10], where sound and vibration measurements collected at the wind turbine towers are employed for condition monitoring of generators. Tower vibration signals are analyzed using Empirical Mode Decomposition (EMD) and the outcomes are correlated with the vibration signals acquired directly from the generator bearings. It is shown that the generator bearing fault signatures are present in the vibrations from the tower.

This study is devoted to the test case of two multi-megawatt wind turbines sited in Italy, owned by Renvico (a company managing around 340 MW of wind turbines in Italy and France, www.renvicoenergy.com). The wind turbines are not equipped with gearbox condition monitoring systems and they have been diagnosed of gearbox damages (of different severity) through the analysis of oil particle counting. Before the gearboxes replacement intervention, a measurement campaign has been conducted by the University of Perugia. The idea is measuring vibrations at the tower: the measurements are collected on the target damaged wind turbines and on one (or more) reference undamaged wind turbines. Subsequently, the data are processed through a multivariate Novelty Detection algorithm in the feature space. The application of this algorithm is supported by statistical analysis on the time-domain features selected. Finally, the novelty indexes based on such time-domain features prove to be suitable to diagnose a damaged condition. It should be noticed that the obtained results allow distinguishing between the two target wind turbines and the corresponding different severity of the gearbox damages.

The manuscript is organized as follows: in Section 2, the test case wind farm, the measurement techniques and equipment and the obtained data sets are described. Section 3 is devoted to the data analysis, feature extraction and results discussion. Finally, in Section 4 some concluding remarks and further directions of this study are indicated.

2 The on-site measurements and the data sets

The wind farm is composed of six multi-megawatt wind turbines and it is sited in southern Italy. The layout of the wind farm is reported in Figure 1, where the damaged wind turbines are indicated in red. The lowest inter-turbine distance on site is of the order of 7 rotor diameters.

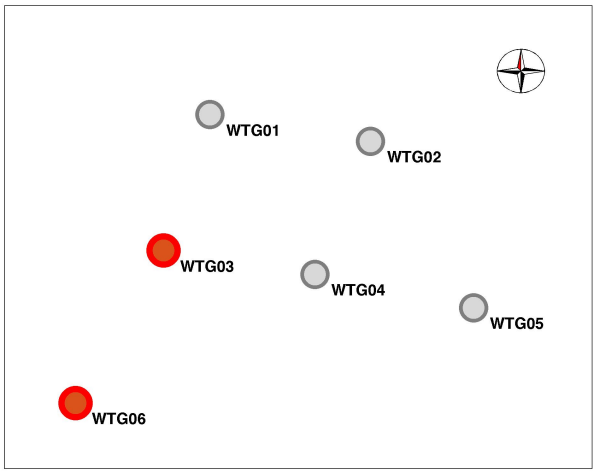


Figure 1: The layout of the wind farm. In red, the damages wind turbines (WTG03 and WTG06) are indicated.

It should be noticed that the damages to WTG03 and WTG06 have different level of severity: actually, the damage at WTG06 was detected through oil particle counting some days before the measurement campaign, while the damage at WTG03 can be considered at incipient stage.

The measurements are conducted as follows: accelerometers are mounted inside the tower of the wind turbine. They measure the longitudinal (x-axis) and transversal (y-axis) vibrations, as displayed in Figure 2. An overall set of four accelerometers (respectively two on the superior level 7 m above ground and two at the inferior level 2 m above ground) and a microphone (on the inferior level) were used for the acquisition. Each acquisition therefore consists of 4 channels sampled at 12.8 kHz for 2 minutes.

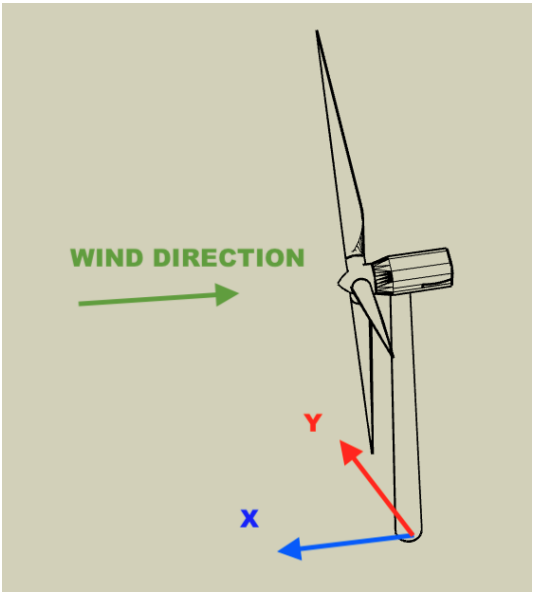


Figure 2: Definition of the reference frame for the longitudinal and the transversal directions.

Operational data have been provided by the wind turbine manufacturer in real time during the measurement campaign, with a sampling time of the order of the second. These have been used to assess the similarity of the wind and operation conditions at different wind turbines at the same time.

The vibration time series have been organized as indicated in Tables 1 and 2. The WTG01 time series are not labelled with more details (as for example the recording time) because they can be interchanged and the

results of the following analysis don't sensibly change.

TS number	Wind turbine	Wind turbine status	Use
1	WTG01	healthy	reference - calibration
2	WTG01	healthy	reference - calibration
3	WTG01	healthy	validation
4	WTG03	damaged	validation

Table 1: The data set for WTG03 damage detection

TS number	Wind turbine	Wind turbine status	Use
1	WTG01	healthy	reference - calibration
2	WTG01	healthy	reference - calibration
3	WTG01	healthy	validation
$\bar{4}$	WTG06	damaged	validation

Table 2: The data set for WTG06 damage detection

The information regarding the state of health of the wind turbine must be extracted from these data. In order to highlight it, some features can be extracted from the raw data set. Obviously, the choice of these characteristic parameters is strongly affecting the ability to perform a damage detection, so that they should be selected wisely. A simple choice is to use common time-domain statistics such as root mean square, skewness, kurtosis, peak value and crest factor (peak/RMS). These are usually quite sensitive to the operational and environmental conditions and are very fast to compute [11]. To ensure the statistical significance of the results, many measurement points are necessary. These features will be then extracted on short, independent (no overlap) chunks of the original signals. In particular, each acquisition is divided in 100 sub-parts on which the five features are computed. The considered data sets X results then to be a $n \cdot d$ matrix, where $n = 20$ is the number of channel and feature combinations, while $d = 400$ is the number of samples from the 4 acquisitions of Tables 1 and 2 placed one after the other.

3 Analysis and results

The results about the feature extraction are reported in Figures 3 and 4. The samples 0-200 are referred to the training data set for the wind turbine WTG01, the samples 201-300 are referred to the validation data set for the wind turbine WTG01 and, finally, the samples 301-400 are referred to the validation data set for the wind turbine WTG03 (WTG06, respectively). In the Figures, the training - calibration data set is separated from the validation data set by a black line. The validation data set for the damaged wind turbine is separated from the rest of the data sets through a red line.

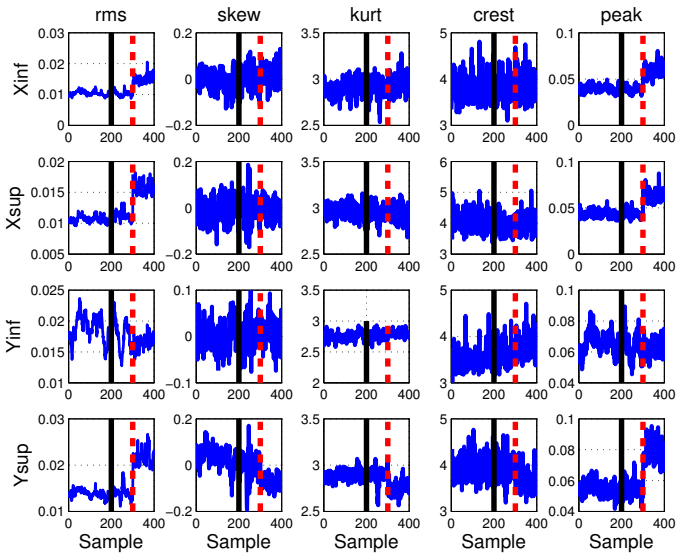


Figure 3: The extracted features for the data sets in Table 1.

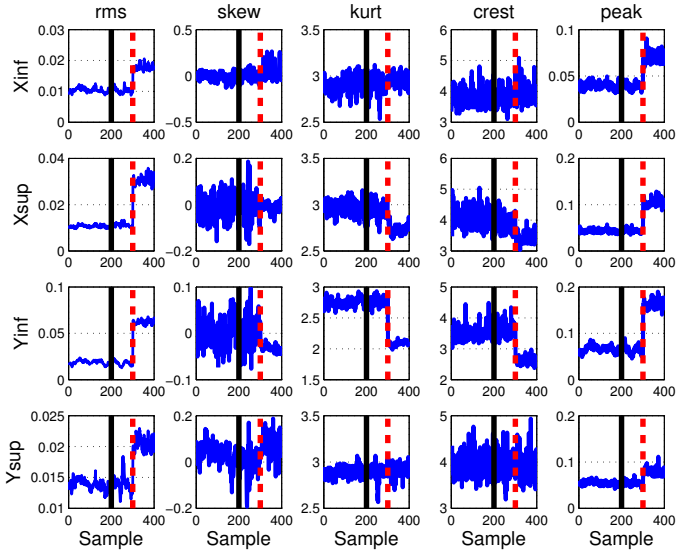


Figure 4: The extracted features for the data sets in Table 2.

A statistical approach is used in this paper to test if some diagnostic information can be obtained from the data, basically assessing the goodness of the selected features. The study starts with a univariate Analysis Of Variance (ANOVA), able to infer from the data the hypothesis that no statistical difference is detected among the groups, meaning that all the groups come from the same distribution.

The ANOVA is a statistical tool to test the omnibus (variance based) null hypothesis H_0: all the considered groups populations come from the same distribution, meaning that no significant difference is detectable. This hypothesis will be accepted or rejected according to a statistical summary $\hat{F}$ which, under the assumptions of independence, normality and homoscedasticity of the original data, follows a Fisher distribution:

$$\hat{F} = \frac{\frac{\sigma^2_{bg}}{G-1}}{\frac{\sigma^2_{wg}}{N-G}} \simeq F(G-1, N-G), \tag{1}$$

where

$$\sigma^2_{bg} = \sum_{j=1}^{G} \frac{n_j}{N} (\bar{y} - \mu_j)^2, \tag{2}$$

$$\sigma_{wg}^2 = \frac{1}{N}\sum_{j=1}^{G}\sum_{i=1}^{n_j}(\bar{y}_{ij}-\mu_j)^2\,, \tag{3}$$

with G being the number of groups of size n_j, N being the global number of samples with overall average $\bar{y}$, σ_{bg}^2 being the variance between the groups, σ_{wg}^2 being the variance within the groups (basically the average of the variance computed in each group) [12, 13]. The null hypothesis H_0 will be accepted with a confidence level $1-\alpha$ if the summary $\hat{F}$ is less extreme than a critical value $F^{\alpha}(G-1,N-G)$. A corresponding p-value can also be computed: it coincides with the probability of the summary to be more extreme than the observed $\hat{F}$, assuming H_0 to be true. If this value is less than α (typically, 5%), H_0 is rejected. The concepts of critical value and p-value are summarized in Figure 5.

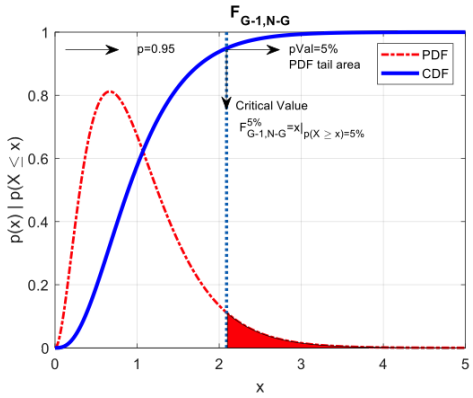


Figure 5: $F(G-1,N-G)$ distribution, with highlighted the 5% critical value and the concept of p-value.

In this analysis, the data sets are divided in 2 groups: the healthy one contains the first 300 samples (time series 1 to 3), while the last 100 samples, coming from the damaged and turbines WTG03 and WTG06 (time series 4 and $\bar{4}$), are labelled as damaged. The assumption of normality can be considered verified with enough confidence. The same does not hold for the homoscedasticity (equal variance in the different groups), but the ANOVA is commonly considered robust to such violations, so that the trustworthiness of the results will not be affected. It is relevant to point out that in this case, which uses 2 groups only, the ANOVA reduces to a Studentâs t-test. Furthermore, the ANOVA is a univariate technique, so it will be repeated per each channel and feature combination (20 times). The results are reported in Tables 3 and 4.

Feature / Channel	Xinf	Xsup	Yinf	Ysup
RMS	$< 10^{-32}$	$< 10^{-32}$	$9\cdot 10^{-13}$	$< 10^{-32}$
Skewness	$< 10^{-8}$	0.4	0.4	$< 10^{-32}$
Kurtosis	$< 10^{-12}$	0.6	$< 10^{-12}$	$< 10^{-32}$
Crest	0.005	0.8	$< 10^{-19}$	$< 10^{-22}$
Peak	$< 10^{-32}$	$< 10^{-32}$	0.01	$< 10^{-32}$

Table 3: ANOVA p-values for the data sets in Table 1. The red cells are used to highlight the acceptance of H_0 (p-value$>$ 5%), which implies a more difficult damage detection.

Feature / Channel	Xinf	Xsup	Yinf	Ysup
RMS	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$
Skewness	$< 10^{-13}$	0.2	$< 10^{-32}$	$< 10^{-32}$
Kurtosis	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	0.01
Crest	0.001	$< 10^{-32}$	$< 10^{-32}$	0.6
Peak	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$

Table 4: ANOVA p-values for the data sets in Table 2. The red cells are used to highlight the acceptance of H_0 (p-value$>$ 5%), which implies a more difficult damage detection.

The Principal Component Analysis (PCA) is a technique widely used in multivariate statistics, in particular for the purpose of allowing the visualization of multi-dimensional data sets using projections on the first 2 or 3 principal components. This dimension reduction is not really advisable for diagnostic purposes, as the condition-information may, in principle, be hidden in the neglected principal components, making the detection more challenging. In any case, it is used in this analysis as a qualitative visualization of the data set under a different point of view, resulting from the transform produced by the technique. The PCA uses an orthogonal space transform to convert a set of correlated quantities into the uncorrelated variables called principal components. This transform is basically a rotation of the space in such a way that the first principal component will explain the largest possible variance, while each succeeding component will show the highest possible variance under the constraint of orthogonality with the preceding ones. This is usually accomplished by eigenvalue decomposition of the data covariance matrix, often after mean centering.

The PCA transform has been applied to the reference data set: the statistical features matrix extracted from the WTG01 time series 1 and 2 of Tables 1 and 2. Subsequently, the validation data sets have been separately projected to the space generated by the first two principal components of the reference data set. The results are reported in Figures 6 and 7, from which it arises that the data set of WTG06 is more easily distinguishable with respect to the calibration data set than the data set of WTG03. As regards Figure 7, the indication is that the visual inspection based on the first two principal components can be sufficient for detecting an anomaly.

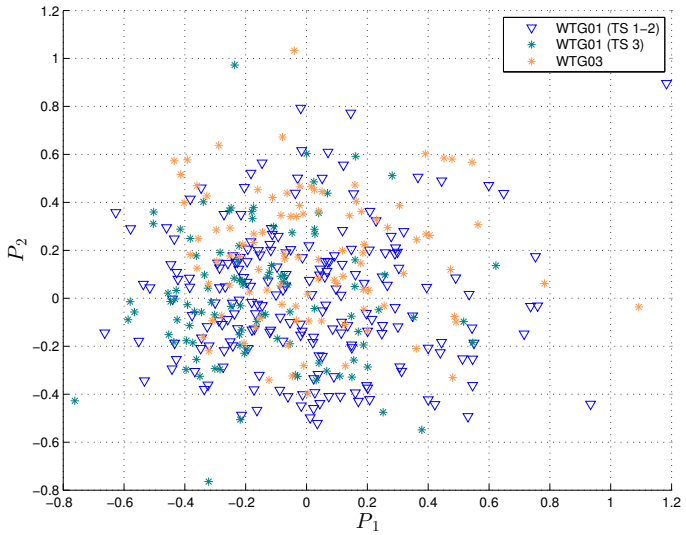


Figure 6: Projection of the data from Figure 3 to the space generated by the two principal components of the calibration data set (first 200 samples in Figure 3)

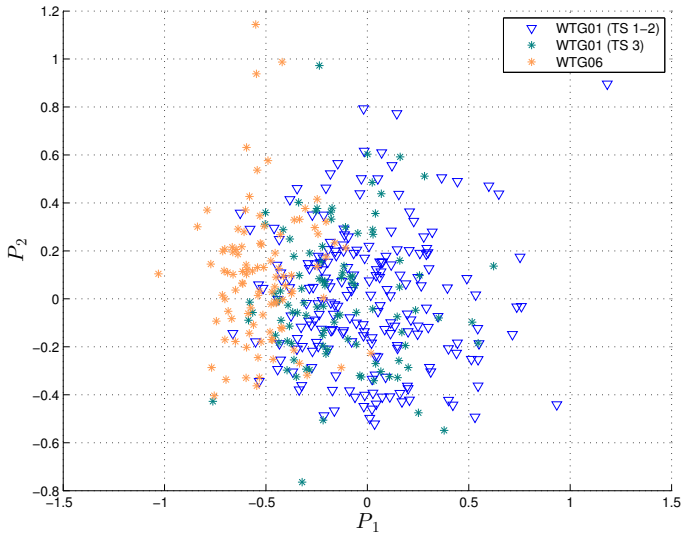


Figure 7: Projection of the data from Figure 4 to the space generated by the two principal components of the calibration data set (first 200 samples in Figure 4)

In statistics, the detection of anomalies can be performed pointwise, looking for the degree of discordance of each sample in a data set. A discordant measure is commonly defined outlier, when, being inconsistent with the others, is believed to be generated by an alternate mechanism. The judgment on discordance will depend on a measure of distance from the reference distribution, usually called Novelty Index (*NI*) on which a threshold can be defined [14]. The Mahalanobis distance is the optimal candidate for evaluating discordance in a multi-dimensional space, because it is non-dimensional and scale-invariant, and takes into account the correlations of the data set. The Mahalanobis distance between one measurement y (possibly multi-dimensional) and the x distribution, whose covariance matrix is S, is given by

$$d_M(y) = \sqrt{(y-\bar{x})\,S^{-1}\,(y-\bar{x})}. \tag{4}$$

In the following, the reference x distribution is selected as the statistical features matrix extracted from the WTG01 time series 1 and 2 of Tables 1 and 2. The target y is selected as the statistical features matrix extracted from respectively time series 3 (WTG01), 4 (WTG03), $\bar{4}$ (WTG06).

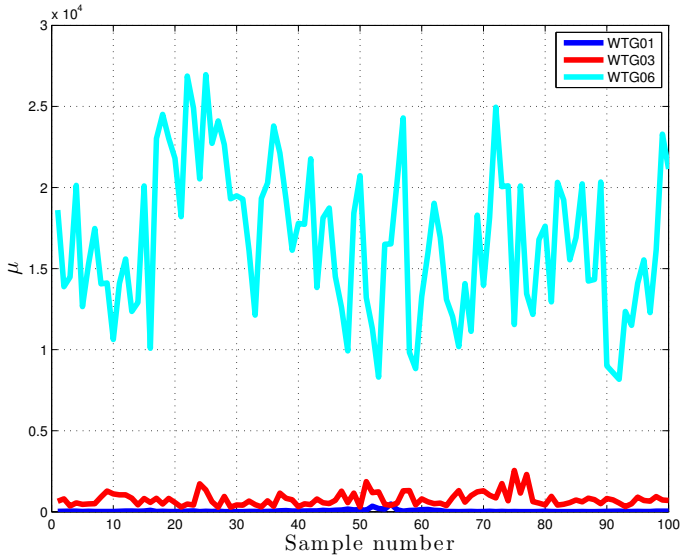


Figure 8: The Mahalanobis distance μ with respect to the calibration WTG01 data set: WTG01, WTG03, WTG06.

From Figure 8, it is possible to clearly distinguish between wind turbine WTG01 and wind turbines WTG03 and especially WTG06. The Mahalanobis distance therefore qualifies to be particularly responsive for novelty detection issues.

4 Conclusions

A novel approach for damage detection of a wind turbine gearbox was proposed in this study. One main novelty is that the accelerometric acquisitions were performed inside the tower of the wind turbines of interest because, despite the distance with respect to the gearbox, it is easily accessible by the turbine practitioners without shutting down the wind turbine. This measurement technique is a distinctive part of the outcome of the present work. One reference healthy wind turbine and two wind turbinse affected by different damage severity have been selected as test cases for the measurement campaigns proposed in the present study.

Subsequently, a Novelty detection procedure was set up, based on the calculation and the elaboration of common time domain features like RMS, Skewness, Kurtosis, Crest factor and Peak value. The analysis started with an ANOVA and a PCA, two fundamental tools in univariate and multivariate statistics. Both techniques proved that the damages can be detected. Finally, the Mahalanobis Novelty detection showed optimal results in detecting the possible damage, given the large margin which separates the supposedly damaged wind turbines from the healthy wind turbine. This algorithm also proved to be a good unsupervised damage detection technique considering the quickness, the simplicity and the full independence from human interaction, which makes it suitable for real time implementation. Overall, the whole gearbox vibration monitoring methodology can be considered validated by the test. The simple, non-invasive measurement system composed of just 2 biaxial accelerometers placed in accessible locations at 2 levels inside the tower of the wind turbine, together with the Novelty detection algorithm applied on the common time-domain features extracted, demonstrated indeed to provide a robust monitoring system, which can be easily integrated in existing installations.

This system can, in principle, enable to monitor also the damage evolution in time, establishing the foundations for further works on prognostics: this is supported by the responsiveness of the proposed methods (especially the Mahalanobis distance analysis) with respect to the severity of the damages (Figure 8). The straightforward further direction of the present work is therefore the analysis of the evolution in time of the same test case.

Acknowledgements

The authors would like to thank Ludovico Terzi, technology manager of the Renvico company. This research was partially funded by Fondazione âCassa di Risparmio di Perugiaâ through the research projects OPTOWIND (Operational Performance and Technical Optimization of WIND turbines) and WIND4EV (WIND turbine technology EVolution FOR lifecycle optimization).

References

[1] J. Rafiee, F. Arvani, A. Harifi, & M.H. Sadeghi, *Intelligent condition monitoring of a gearbox using artificial neural network*, Mechanical systems and signal processing, 21(4), 1746-1754.

[2] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. Tameghe, & G. Ekemb, *Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges*, Energies, 7(4), 2595-2630.

[3] D. Astolfi, F. Castellani, & L. Terzi, *Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm*, Diagnostyka, 15.

[4] Z. Zhang, A. Verma, & A. Kusiak, *Fault analysis and condition monitoring of the wind turbine gearbox*, IEEE transactions on energy conversion, 27(2), 526-535.

[5] J. Igba, K. Alemzadeh, C. Durugbo, & E.T. Eiriksson, *Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes*, Renewable Energy, 91, 90-106.

[6] C. Peeters, P. Guillaume, & J. Helsen, *Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy*, Renewable Energy, 116, 74-87.

[7] J. Antoni, *Cyclic spectral analysis in practice*, Mechanical Systems and Signal Processing, 21(2), 597-630.

[8] F. Yanhui, L. Jiawei, Q. Yingning, Y. Wenxian, & D. Infield, *Study on order analysis for condition monitoring wind turbine gearbox*, 3rd Renewable Power Generation Conference (RPG 2014).

[9] G.A. Skrimpas, T. Ursin, C. Sweeney, K. Marhadi, N. Mijatovic, & J. Holboell, *Residual signal feature extraction for gearbox planetary stage fault detection*, Wind Energy, 20(8), 1389-1404.

[10] E. Mollasalehi, D. Wood., & Q. Sun, *Indicative fault diagnosis of wind turbine generator bearings using tower sound and vibration*, Energies, 10(11), 1853.

[11] A. P. Daga, L. Garibaldi, A. Fasana, & S. Marchesiello, *ANOVA and other statistical tools for bearing damage detection*, International Conference Surveillance 9, Fez (Morocco), 22-24 May 2017.

[12] H. Scheffe, *The analysis of variance*, John Wiley Sons, 1959.

[13] H. Sahai, & M.I. Ageel, *The analysis of variance: fixed, random and mixed models*, Springer Science Business Media, 2012

[14] K. Worden, G. Manson, & N.R. Fieller, *Damage detection using outlier analysis*, Journal of Sound and Vibration, 229(3), 647-667.

Gears and bearings defaults: from classification to diagnosis using machine learning for SURVISHNO Conference 2019

Sylvain BARCET[1], Valentin BARON[2], Alexandre CARBONELLI[1]

[1]VIBRATEC, F-69130 Ecully, France
[2]MICRODB, F-69130 Ecully, France

sylvain.barcet@vibratec.fr
valentin.baron@micro-db.fr
alexandre.carbonelli@vibratec.fr

Abstract

Gears and bearings are more and more used in every industrial area mainly due to their strong reliability. Nevertheless as every mechanical transmission system, failures appear during time life. It induces critical damage, time cost for maintenance services to repair the fault potentially on duty. A wide part of work in the scientific community already provides a large quantity of features to follow health status of these systems (e.g., RMS, kurtosis, crest factor, FM0) in order to detect the fault as soon as possible.

Since few years, methods developed in signal post-processing are coupled with Machine Learning (ML). ML allows ability to detect novelty or fault based on a trained algorithm. According to the literature [1], to identify the type of damage, a supervised algorithm is needed. Consequently an accurate diagnosis implies labelled data which are often difficult to obtain practically.

The aim of this paper is to provide keys, based on our knowledge about features in Structural Health Monitoring (SHM), to get higher information level in classification by adding a qualitative analysis (type of damage) without label or information about the type of fault.

Work carries on a measurement database. The assumption is made about two classes “healthy” / “faulty” using a supervised algorithm. The contribution of our work brings a new step in the default analysis by adding a probability for a defect case to be identified. Indeed, by combining some sensitive features selected for their relevance to describe a type of fault, a probability to have this particular default can be given. This classification is tested against three fault classes: bearing, gear generalized, gear localized.

Results show that a probability for having bearing fault can be identified using this method contrary to the gear generalized and localized fault which are more complex to characterize. This new step enables to help maintenance services to focus more efficiently on the incriminated faulty part of the system, inducing a reduction of time to repair for maintenance services, a shorter out of order time leading to a significant productivity gain.

Introduction

Gears are used in a huge quantity of mechanical systems. As a consequence, monitoring their possible faults the most accurately possible is a major issue in the field of Structural Health Monitoring (SHM) as they can provoke critical damages.

With the increasing use of ML techniques, different algorithms have emerged to deal with this problem. Support Vector Machine, Neural Network, Random Forest are examples of ML methods currently used to classify faulty and healthy sample.

Looking at industrial maintenance services requirements, the needs in terms of monitoring may be resumed as:

1- Find efficient condition indicators (CI) to monitor their systems,
2- Use ML algorithms to allow a continuous monitoring and an high efficiency of faulty detection,
3- Have a minimal cost and time to repair the faulty equipment.

The first and the second point are already addressed in literature. This paper proposes to industrials a method to complete their process by the third part: a qualitative analysis of fault, leading to a reduction of cost and time for maintenance services.

Presentation of the study case

PHM Society proposes a challenge for monitoring and fault detection. They provide a measurement database (measured on a test bench). Students, researchers and companies can participate. Each one proposes their own method to classify the given database. This work is based on the database provided for the 2009 challenge. Figure 1 presents the test bench used to build this measurement database.

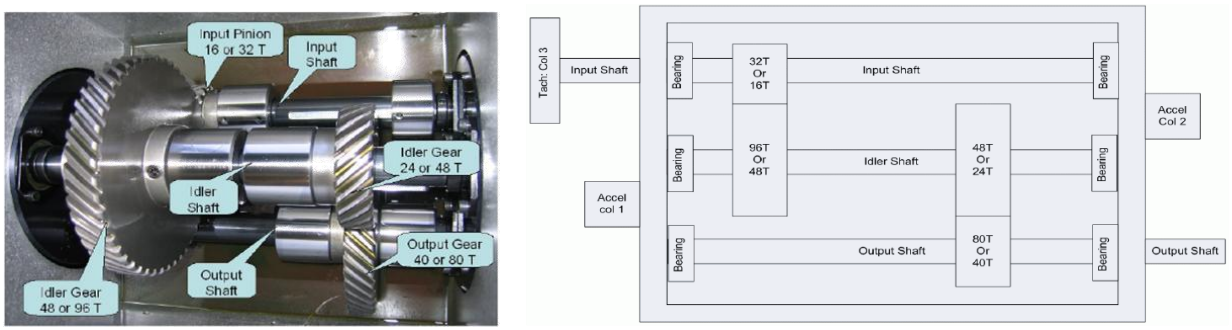


Figure 1: Presentation of the test bench used for measurements coming from PHM Society 2009 Challenge

This test bench is built with two gear stages mounted on three shafts with six bearings. Two gear geometries are used: one using spur gears, the other one using helical gears.

Table 1 presents the gear parameters for both configurations.

Shaft	Gear	Spur gear	Helical gear
Input shaft	input gear	32	16
Idler shaft	1st idler gear	96	48
Idler shaft	2nd idler gear	48	24
Output shaft	output gear	80	40

Table 1 : Gear description for the two geometries: spur and helical

Whatever the gear geometry, the gear ratio between each gear stage is the same, leading to keep the same global gear ratio. Thus, from input to output the gear reduction ratio is 5 to 1 reduction:

$$gear\ ratio = \frac{16}{48} \times \frac{24}{40} = \frac{1}{5}. \quad (1)$$

The instrumentation is composed of a limited number of sensors with two accelerometers mounted on the housing and one tachometer on the input shaft (see Figure 1). The tachometer delivers 10 tops by rotation. The sampling frequency is the same for the three sensors and fixed at 66666.67 Hz. Each sample of the database is composed of three raw data columns of one second length.

Several configurations are listed such as:
- 14 gear cases:
  - 8 configurations on spur gear [7 faulty and 1 healthy],
  - 6 configurations on helical gear [5 faulty and 1 healthy],
- 5 rotational speeds [30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz],
- 2 load cases [high, low].

This leads to 140 different configurations. Each configuration is repeated four times to give at least 560 measurement inputs.

Method

The methodology developed and exposed here is composed of three main steps:
- CI computing,
- Classification using ML,
- Qualitative analysis using relevant selected features for each chosen default.

Condition indicators computing

Figure 2 and Figure 3 present the post-processing used to build a matrix with all indicators.

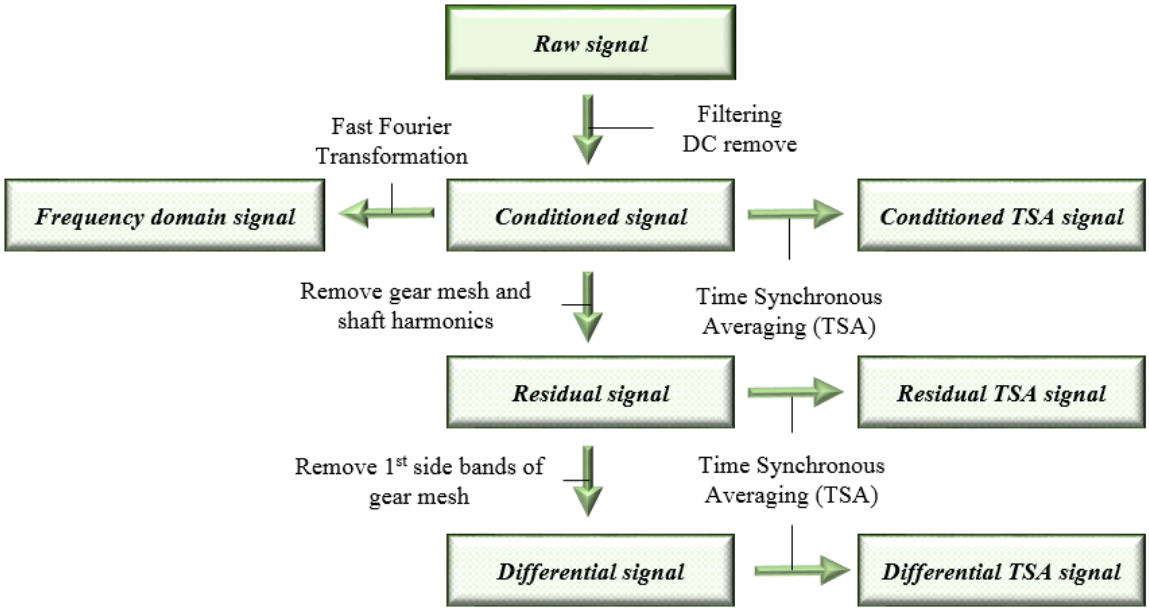


Figure 2 : Signal post-processing used to access to the different needed types of signals

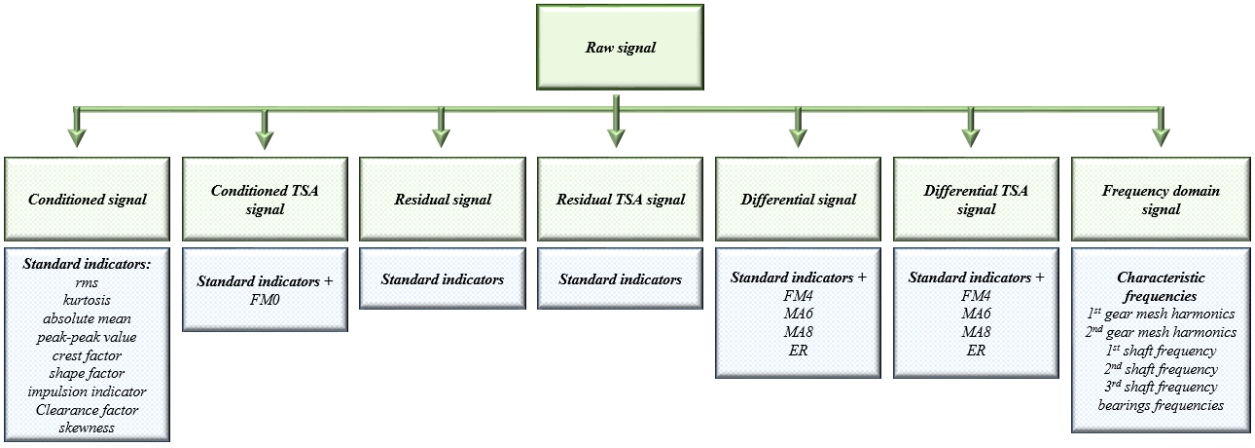


Figure 3 : List of indicators calculated on each signal

From raw signals, based on literature on bearings and gearboxes monitoring, some CI are extracted. Taking into account the difficulty to add a lot of different sensors, most of usual CI used in industry are directly computed from accelerometers raw signals (conditioned signal) such as RMS, kurtosis, crest factor. Based on [3], some new indicators, using residual or differential signals, requiring a tachometer information allow to increase strongly the efficiency of fault detection.

In order to complete this list of CI, some other indicators are computed from the frequency domain, such as gear mesh harmonics, shaft rotational frequencies and specifics bearings frequencies.

All these frequencies are calculated and added in the panel of features using the following equations:

$$Ball\ Pass\ Frequency\ of\ Inner\ ring: BPFI = \frac{N_b}{2}\ f\ \left(1 + \frac{D_b}{D_p}\cos(\alpha)\right), \qquad (2)$$

$$Ball\ Pass\ Frequency\ of\ Outer\ ring: BPFO = \frac{N_b}{2}\ f\ \left(1 - \frac{D_b}{D_p}\cos(\alpha)\right), \qquad (3)$$

$$Ball\ Spin\ Frequency: BSF = \frac{D_p}{2\ D_b}\ f\ \left(1 - \left(\frac{D_b}{D_p}\right)^2 (\cos(\alpha))^2\right), \qquad (4)$$

with f the number of revolutions per second, D_b the ball diameter, N_b the ball number, D_p the pitch diameter and α the contact angle.

$$Gear\ mesh\ frequency = f_{sh(i)} \times n_{teeth(i)}, \qquad (5)$$

with $f_{sh(i)}$ the rotational speed frequency of the shaft "i" and $n_{teeth(i)}$ the number of teeth on the gear mounted on the shaft "i".

Considering the two accelerometers, a total of 298 indicators are extracted.

Classification using Machine Learning

Based on all our indicators, the second step of the methodology consists in the classification of the 560 signals in two classes: healthy or faulty. Although the labels are not given in the database from the PHM Society, in [2], authors give the label of all healthy samples they have classified. Based on this, an approach using supervised algorithms is possible for the classification.

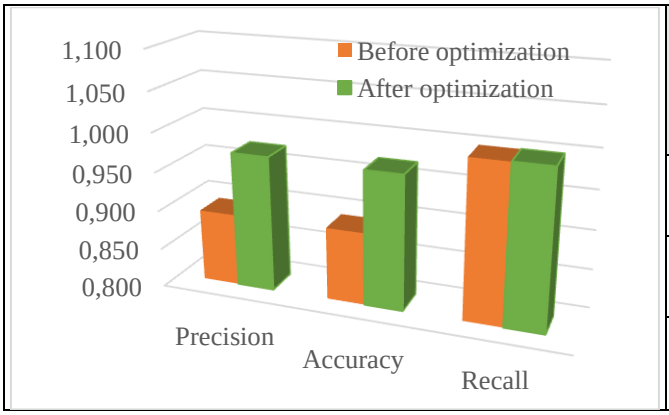
A wide range of supervised algorithms exists in the literature. Among them, a classification is proposed using four of them implemented within the Scikit-learn Python module [5]:

- Nearest Neighbors classifier (KNN) [6],
- Random Forest classifier (RFC) [7],
- Support Vector classifier (SVC) [8],
- Multi Layer Perceptron Classifier (MLPC) [9].

Each algorithm works with a specific method to classify. SVC determines a boundary between the two classes using only the data of each class which are close one to the other, the so-called support vectors. KNN looks at the same class nearest samples of a particular observation to build the boundary between classes. RFC builds decision trees and it combines them together to give its final classification. Finally MLPC relies on a trained neural network to decide that class the tested observation belongs to.

The considered input signals for this step are the 560 one second duration signals provided by the PHM database. So an observation for the following of the contribution refers to a vector of dimension 298 gathering all the features computed from one of these 560 signals. These 560 signals are splitted in two categories: a train set and a test set. 80 % of the database is used for the train set (448 signals) and 20 % for the test set (112 signals).

The efficiency of these algorithms without any optimization is around 90% of good classification. An efficient solution to increase the performance of a ML algorithm is to optimize some hyper parameters. Table 2 presents results before and after optimization on SVC algorithm.



SVC algorithm	Before optimization	After optimization
Accuracy	0.893	0.973
Precision	0.893	0.971
Recall	1.000	1.000

Table 2 : Results of Support Vector Classifier before and after hyper parameters optimization

The optimization phase enables to increase significantly the accuracy of the considered algorithm from 89.3 % to 97.3 %. After an optimization of the hyper parameters on every algorithm, each individual algorithm gives interesting results with a percentage of good classification between 90 % and 97.1 %.

Finally a method to increase strongly the performances relies on combining different algorithms with different approaches. This method called “ensemble learning” reduces individual weaknesses of each algorithm. In the present case, the precision reaches 100 % using ensemble learning on the three best algorithms, namely RFC, SVC and MLPC.

Qualitative analysis

The classification in healthy or faulty cases enables to alarm maintenance service immediately after or ideally a short time before the failure. However, no information is given about the type of fault.

The methodology proposed in this work consists in adding a qualitative information in the classification.

The work presented here proposes a method to estimate the probability for a faulty observation to correspond to one of these three types of faults:

- Bearing faults,
- Gear generalized faults,
- Gear localized faults.

The wide quantity of indicators to monitor bearings and gearboxes given in the first section does not describe the same input signals' features. Consequently they are not equally sensitive to the different types of faults [4].

Where RMS represents the energy of the signal leading to an accurate indication about the general state of the complete system, the peak-peak value is in opposition, very sensitive to any localized phenomenon on the signal, enabling to discriminate a localized fault such as crack on tooth or more critical case like a missing tooth. A solution based on the physics described by each indicator is investigated to predict the type of fault.

3.3.1 Feature selection

The selection proposed is composed of:

- 70 indicators for generalized faults:
    - Rms on TSA signal:
        - Rms on conditioned TSA signal on the 1st shaft, 2nd shaft and 3rd shaft,
        - Rms on residual TSA signal on the 1st shaft, 2nd shaft and 3rd shaft,
        - Rms on differential TSA signal on the 1st shaft, 2nd shaft and 3rd shaft.
    - Absolute mean:
        - Absolute mean on conditioned signal,
        - Absolute mean on conditioned TSA signal on the 1st shaft, 2nd shaft and 3rd shaft,
        - Absolute mean on residual signal,
        - Absolute mean on residual TSA signal on the 1st shaft, 2nd shaft and 3rd shaft,
        - Absolute mean on differential signal,
        - Absolute mean on differential TSA signal on the 1st shaft, 2nd shaft and 3rd shaft.
    - MA6:
        - MA6 on differential signal,
        - MA6 on differential TSA signal on the 1st shaft, 2nd shaft and 3rd shaft.
    - Gear mesh harmonics:
        - 1st gear mesh frequency, harmonics 1 to 5,
        - 2nd gear mesh frequency, harmonics 1 to 5.
- 14 indicators for localized faults:
    - FM0:
        - FM0 on conditioned TSA signal on the 1st shaft, 2nd shaft and 3rd shaft.
    - Peak-peak value on residual signal:
        - Peak-peak value on residual signal,
        - Peak-peak value on residual TSA signal on the 1st shaft, 2nd shaft and 3rd shaft.
- 18 indicators for bearing faults:
    - BPFO:
        - BPFO on 1st shaft, 2nd shaft and 3rd shaft.
    - BPFI:
        - BPFI on 1st shaft, 2nd shaft and 3rd shaft.
    - BSF:

- BSF on 1st shaft, 2nd shaft and 3rd shaft.

From the 298 indicators, a total of 102 indicators are extracted for this qualitative analysis, they are summarized in Table 3.

Type of fault	Type of CI	Number of indicators (for 2 accelerometers)
Generalized	RMS on residual signals	18
	MA6	8
	Absolute mean	24
	Gear mesh harmonics	20
Localized	FM0	6
	peak-peak value on residual signal	8
Bearing	BPFO	6
	BPFI	6
	BSF	6

Table 3: List of CI used for the qualitative analysis

3.3.2 Characterization method

Once feature selection is performed according to their relevance for each of the investigated fault, the method to highlight the cases associated with a particular fault relies on the Principal Component Analysis (PCA) method [10]. The goal of PCA is to find, in a point cloud, the direction on which the projection of its point cloud has a maximum variance. This direction is called the first principal component and the next ones are the orthogonal directions of this first one that again explain the maximum variance. In this contribution each feature selection for each fault gives a point cloud of high dimension:

- Dimension 70 for the gear generalized faults,
- Dimension 14 for the gear localized faults,
- Dimension 18 for the bearing faults.

Consequently the PCA is used in this case to perform a dimensionality reduction in order to be able to represent in a two-dimensional space data which came from these high-dimensional spaces. From this representation the observations which represent a particular fault are expected to be significantly far away from the healthy and the other default cases.

Therefore once the PCA is performed on each case, the probability density function (PDF) derived from the healthy cases is estimated. According to isolines corresponding to specific probability values three categories are differentiated:

- A category with a low probability to have the studied fault type whose data are close to the healthy ones after PCA,
- A category with a medium probability to have this fault which are a little bit further from the healthy data,
- A category with a high probability to have this fault which are far away from the healthy data.

Results and analysis

4.1.1 Bearing defaults

Once the 18 features which should represent well the bearing faults are selected, the PCA is performed and Figure 4 is obtained.

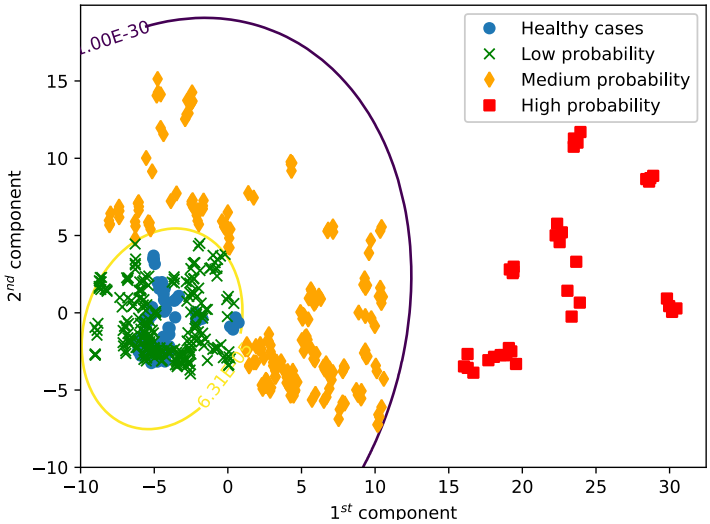


Figure 4: PCA result with the bearing faults feature selection. Yellow and purple lines represent respectively the $10^{-4.2}$ and 10^{-30} isolines of the probability density function derived from the healthy data. Healthy cases are given by blue dots, cases with a low / medium / high probability to have a bearing faults are given by green crosses / orange diamonds / red squares.

It clearly shows a distinct group further than the 10^{-30} probability isoline that gives candidates for having bearing fault. Then in between the two probability isolines the observations are classified as possible to have bearing fault. Finally data which are lying among the healthy cases are not likely to have bearing fault at all.

Consequently the feature selection made beforehand has allowed a qualitative analysis of the data. Indeed, thanks to their position compared to the healthy cases some observations can be classified within a bearing fault category with a given probability.

4.1.2 Localized and generalized defaults

Concerning gear localized and generalized faults, the results are more difficult to interpret as it is shown in Figure 5 with the PCA result of the localized fault case and in Figure 6 with the generalized fault case.

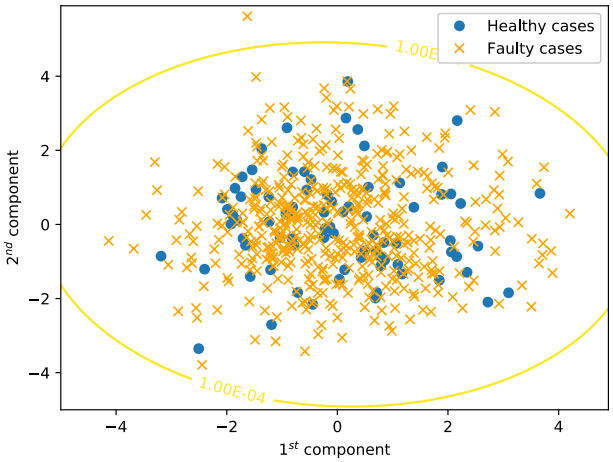


Figure 5: PCA result with the gear localized fault feature selection. The yellow line represents the 10^{-4} isoline of the probability density function derived from the healthy data. Healthy cases are given by blue dots and faulty cases are given by orange crosses.

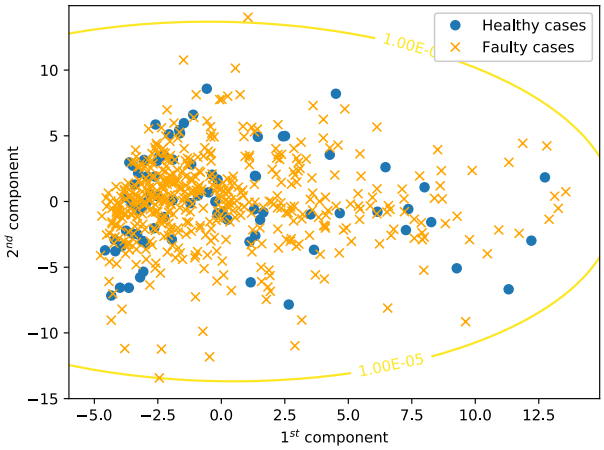


Figure 6: PCA result with the gear generalized fault feature selection. The yellow line represents the 10^{-5} isoline of the probability density function derived from the healthy data. Healthy cases are given by blue dots and faulty cases are given by orange crosses.

In these cases the feature selection has not allowed to distinguish different categories within the fault cases. It means that the selected features are not a relevant set enough to predict both gear localized or generalized faults. Some work is in progress to find better sets of features in order to be able to reproduce the results obtained with the bearing fault identification.

Conclusion

This work presents the full method to diagnose and follow the healthy condition of a rotating equipment: from the conditions indicators to the classification. From 298 relevant conditions indicators obtained from accelerometer signals, a first supervised classification step enables to decide whether or not a considered observation is healthy or faulty using ensemble learning with three combined ML algorithms. To add a qualitative analysis of the faulty cases, a new step is performed using principal component analysis on a feature subset of the 298 ones. Three different faults are studied, bearing faults and gear localized and generalized faults, bringing three different subset of features of dimension 18, 14 and 70, respectively. The dimensionality reduction obtained using the principal component analysis allows to represent in a two-dimensional space, corresponding to the two first principal components, both healthy and faulty cases. From this representation a probability density function of the healthy cases is determined and the faulty cases can be marked as having a low or high probability to indeed, having this specific fault from their position compared to this function.

Results show that bearing faults can be identified using this methodology with three different categories highlighted: low, medium and high probability for having this fault type. However the feature selection made for the gear localized and generalized faults have not allowed to distinguish clearly between healthy and faulty cases and consequently there is no specific observation which can be identified with these fault types.

Work is in progress on the feature selection to be able to reproduce the bearing fault results and other methods are investigated to perform a better dimensionality reduction such as manifold learning techniques. These two ideas could bring the missing part to be able to give a complete qualitative assessment of the fault types.

References

[1] K. Worden and al., *The fundamental axioms of structural health monitoring*, 2007

[2] H. Al-Atat, and al., *A systematic methodology for gearbox health assessment and fault classification*, Journal of Prognostics and Health Management, 2011

[3] A. S. Sait, Y. I. Sharaf-Eldeen, *A Review of Gearbox Condition Monitoring Based on vibration Analysis Techniques Diagnostics and Prognostics*, 2011

[4] V. Sharmaa, A. Pareya*, *A review of gear fault diagnosis using various condition indicators, 12th International Conference on Vibration Problems*, ICOVP 2015

[5] F. Pedregosa et al., *Scikit-learn : Machine Learning in Python*, JMLR 12, pp. 2825-2830, 2011.

[6] R. Duda et al., *Pattern Classification*, 2nd Edition, John Wiley & Sons, 2012.

[7] L. Breiman, *Random Forests, Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001 January 01.

[8] C. Cortes and V. Vapnik, *Support-vector networks, Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995 September 01.

[9] D. Rumelhart et al., *Learning representations by back-propagating errors*, Nature, vol. 323, no. 6088, pp. 533-536, 1986 October.

[10] I. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer Series in Statistics, 2002.

Vibration Feature for Detecting Eccentric Workpiece/Runout Faults During Continuous Gear Grinding Processes

Agusmian Partogi OMPUSUNGGU[1], Yann VONDERSCHER[2], Daniel MOTL[2]

[1]Flanders Make, Gaston Geenslaan 8, 3001 Leuven, Belgium
agusmian.ompusunggu@flandersmake.be

[2]VCST Industrial Products, Schurhovenveld 3 025, 3800 Sint-Truiden, Belgium
yann.vonderscher@gmail.com, daniel.motl@vcst.be

Abstract

Continuous gear grinding is a well-established and widely used process in the industry for large-scale production gears. It offers an economic/efficient process for finishing gears, which shapes the micro-geometry of the gear tooth flank and improves its surface quality. The resulting quality of ground gears depends on several factors, namely the tool performance, the machine stability as well as the correct clamping/positioning of the workpiece. The grinding step is very crucial since it has a direct impact on the operating quality of gears and in particular on the running noise behaviour of the end product. The potential of online vibration based gear grinding monitoring has been explored and demonstrated in the previous work [1] as a means of quality control that could lead to the overall reduction of production losses and to the prevention of sending defective parts to customers. A number of features which could be used to monitor the grinding processes and to identify a specific type of defects have been proposed & experimentally validated to some extent. The types of faults include *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. However, a further investigation on a new test campaign revealed that none of the features developed in [1] was sensitive and robust enough to detect eccentric workpieces during the grinding process. It is worth mentioning here that an eccentric workpiece fault is unlikely to happen, but it is analogous to a runout on the incoming workpiece quality. In this paper, a qualitative model to predict the vibration signature due to eccentric workpieces/runouts is developed and discussed. Based on the qualitative understanding, a novel feature to detect eccentric workpieces/runouts during gear grinding processes based on vibration signals has been developed. The newly developed feature has been validated on real vibration signals captured during the emulation of process malfunctions on an industrial gear grinding machine. The experimental results show that the novel feature is sensitive and robust for detecting workpiece eccentricities of about 40 microns. It is also shown in this study that the feature is insensitive to other types of gear grinding faults, which is important for diagnostics/root-cause analysis purposes.

1. Introduction

A new industrial revolution, known as Industry 4.0, has been taking place worldwide for less than a decade. However, Europe is still at the beginning of this era and is currently in an effort to reindustrialise and to increase total value added from the manufacturing sector to a targeted 20%. The EU supports industrial change through its industrial policy and through research and infrastructure funding in the frames of H2020. The key success to Industry 4.0 is digitalisation and data transformation into business insight, namely how to gather, filter, analyse, store and retrieve data and extract useful information from it. The information is then subsequently shared and used to drive, control and monitor processes. This direction is supported by the continuous need for lower production cost, higher quality, more flexibility, better safety and more environmentally friendly, as well as the ever-increasing per currency computing power of microprocessors and the ever-decreasing cost of sensors and measuring platforms.

One concrete example of Industry 4.0 applications is the monitoring of manufacturing processes for improved efficiency and productivity. The monitoring application in manufacturing can contribute to the machine health monitoring, to the tool condition monitoring (e.g. wear, breakage etc), to the workpiece inspection and quality (e.g. geometry) as well as to the process monitoring [2]. As a result, the manufacturing processes monitoring could from one side reduce the downtime and possible repair costs, and from the other side, be used as a product quality control, leading to the overall reduction of production losses and to the prevention of sending defective products to customers. Since manufacturing processes and equipment are

numerous and vastly different, a plethora of sensors and signal processing tools have been proposed in the literature being usual adapted to the specific application. Among other approaches, online vibration monitoring seemed to be promising and could be used in order to monitor manufacturing processes and consequently accurately, on time and online malfunctions and defects can be identified, detected and diagnosed.

In gear manufacturing processes, continuous generating gear grinding is a well-established and commonly used process in the industry for large-scale gears production because it offers an economic/efficient process for finishing gears. This process is a dominating hard-fine finishing process, especially in the field of automotive gears [3]. Due to its high process efficiency, continuous generating gear grinding has replaced other grinding processes such as profile grinding in batch production of small- and middle-sized gears. Figure 1 illustrates a continuous generating gear grinding process that involves three movements, namely 1) Infeed – along with the tooth depth, 2) feed – along with the workpiece axis and 3) shift – along with the grinding wheel axis. The three movements serve as the process parameters determining the final quality of gears.

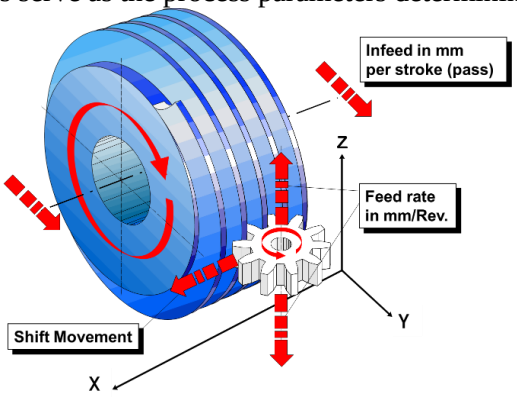


Figure 1: Illustration of a continuous generating gear grinding [4]

Continuous generating gear grinding is characterized by a high stock removal rate and is thus suited for high productivity batch processing. One of the main challenges is the determination of the abrasive/grinding forces due to their significant influence on the dynamics of the grinding process. During the grinding, there are multiple points of contact between the grinding worm and the gear, however, the number of these contacts changes continuously. This involves also a continuous change of the excitation forces. Thus, optimizing the cutting forces can lead to an increased quality of ground gears and a minimised wear behaviour of the grinding worm. Despite its wide industrial application, the knowledge of the generating grinding process is limited. The process design is based on experience along with time- and cost-intensive trials and research is based mostly on empirical studies. To maintain the high quality of ground gears, a reliable and robust online monitoring solution for gear grinding processes is therefore necessary.

The potential of online vibration based gear grinding monitoring has been explored and demonstrated in the previous work [1] as a means of quality control that could lead to the overall reduction of production losses and to the prevention of sending defective products to customers. A number of vibration features which could be used to monitor the grinding processes and to identify a specific type of defects were proposed & experimentally validated to some extent. The types of gear grinding faults include *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. However, a further investigation on a new test campaign revealed that the previously developed features [1] were not sensitive and robust enough to detect eccentric workpiece during the grinding process.

To fill the gap discussed earlier, the main focus of this study is therefore on the development of a novel feature that is reliable and robust for detecting eccentric workpieces/runout faults during gear grinding processes based on vibration signals measurement on the grinding machines. Some quality parameters of the ground gears were also measured that can be used as objective measures for the quality assessment. The remainder of the paper is organised as follows. In Section 2, an introduction to eccentric workpiece/runout is presented. In Section 3, qualitative modelling to predict the vibration signature of eccentric workpiece/runout is presented. In Section 4, the signal processing and feature extraction algorithm is presented. In Section 5, the

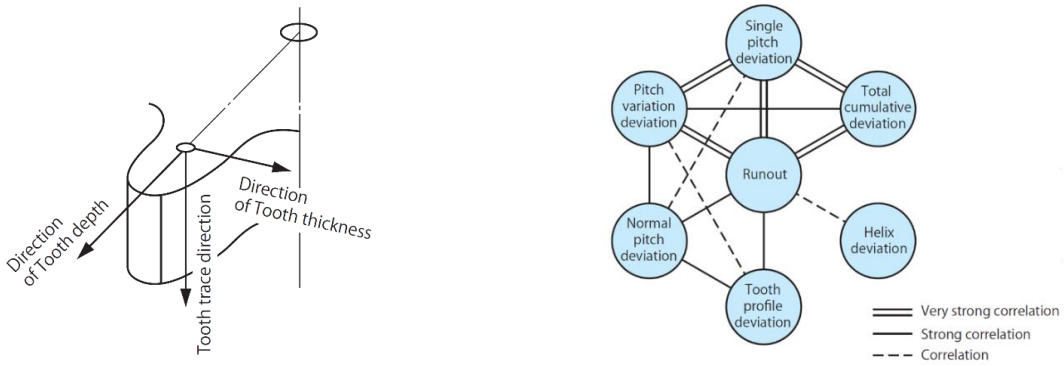
experimental setup and the design of experiments are discussed. In Section 6, the results of the data analysis are discussed. Finally, some conclusions and recommendations for future work are summarised in section 7.

2. Eccentric Workpiece/Runout as A Type of Gear Errors

Gear errors can be classified into individual and composite errors. As shown in Figure 2(a), an individual error is, in fact, a three-dimensional (3D) error occurring in the direction of:

1) Tooth depth referring to the shape of tooth profile and length of tooth depth.
2) Tooth trace referring to the inclination and unevenness of tooth trace.
3) Tooth thickness referring to the thickness of tooth and tooth space.

These three types of individual errors are measured by taking apart a three-dimensional error into a two-dimensional error. However, these individual errors are correlated and the extent of correlation differs between the methods of production and measurement. In general, the correlations of these individual errors are shown in Figure 2(b) [5].



(a) (b)

Figure 2: (a) Theoretical three-dimensional deviations, (b) Correlation with an individual deviation of ground spur gears [5]

Runout is one of the gear errors and defined as the amount of off-centre measured between gear and the axis. It is a characteristic of gear quality that results in an effective centre distance variation. Figure 2(b) shows how the runout deviation matters for gear quality and in fact affects every other characteristic of gear quality, such as involute or tooth form, index or pitch variation, lead or tooth alignment variation, etc. Therefore, a good practice for the manufacturing or inspection/monitoring of gears requires the control of runout. Runout results in accumulated pitch variation, and this causes non-uniform motions, which affect the function of gears [5]. Although runout is a radial phenomenon, the resulting accumulated pitch variation affects the tangential functionality that causes transmission error. In gear quality metrology, the runout error Fr is measured by indicating the position of a pin or ball inserted in each tooth space around the gear and taking the largest difference as schematically shown in Figure 3.

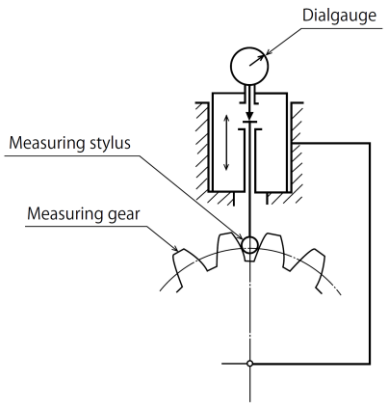


Figure 3: Measurement of runout [5].

When a workpiece is not centred on the spindle rotating axis as illustrated in Figure 4 (i.e. eccentric workpiece), the grinding of the eccentric workpiece will lead to a variation in the infeed penetration (i.e. the direction of tooth depth) once per gear revolution. Eventually, this variation will result in a runout error on the workpiece as will be shown later in the gear quality measurement results discussed in Section 6. Note that an eccentric workpiece fault is unlikely to happen, but a runout deviation might occur on incoming workpiece quality prior to grinding process.

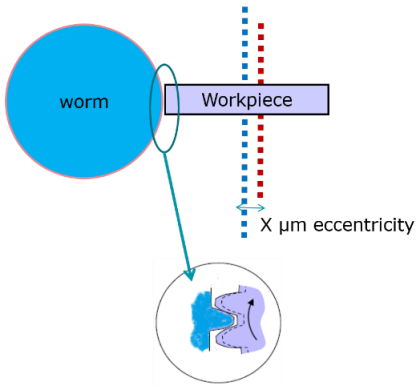


Figure 4: Illustration of an eccentric workpiece fault [1].

3. Qualitative Prediction of Grinding Vibration Signature/Feature Resulting From Eccentric Workpiece/Runout Faults

Figure 5 illustrates the kinematic model between a grinding worm and an eccentric workpiece. Without loss of generality, let assume that the workpiece is mounted on a rigid shaft, while the grinding wheel is mounted on a flexible shaft. Let r_1 be the radius of workpiece and r_2 be the radius of the grinding wheel. For a concentric workpiece as illustrated by the dashed circles in the figure, the rotation centre of the workpiece O coincides with the centre of the workpiece geometry and the rotation centre of the grinding wheel is at point $\hat{O}$. However, for an eccentric workpiece, the centre of the geometry O_1 does not coincide with the rotation centre O and the distance between the point O_1 and O determines the eccentricity level ϵ. As a result of the eccentricity on the workpiece, the rotation centre of the grinding worm (which is coinciding with the centre of the geometry) shifts from $\hat{O}$ to O_2.

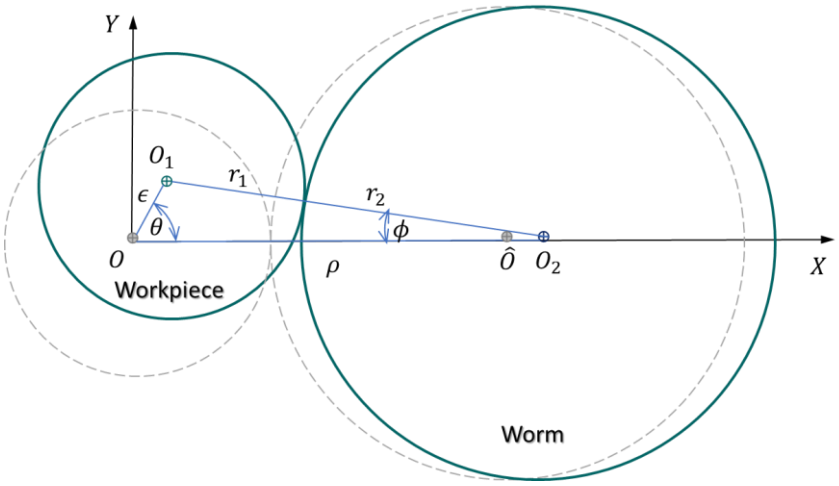


Figure 5: Kinematic model of a grinding wheel and an eccentric workpiece (gear).

With the aid of Figure 5, the distance between the rotation centres of the workpiece and the grinding wheel ρ can be derived as follows:

$$\rho = \epsilon \cos(\theta) + R \cos(\phi) \tag{1}$$

With $R = r_1 + r_2$, and the angle θ denoting the angular displacement of the workpiece which can be determined by integrating the angular speed of the workpiece over time $\theta = \int_o^t \omega_g \, dt$. As the angular speed is typically constant, thus $\theta = \omega t$. The angle ϕ can be inferred by using the sine rule as follows:

$$\frac{\epsilon}{\sin(\phi)} = \frac{R}{\sin(\theta)} \tag{2}$$

Eq (2) can be further rearranged as follow:

$$\phi = \arcsin\big(\delta \sin\big(\omega_g t\big)\big) \tag{3}$$

With $\delta = \frac{\epsilon}{R}$. In practice, the eccentricity is a lot smaller than the sum of the workpiece and the grinding wheel radii, $\epsilon \ll R \ \rightarrow \frac{\epsilon}{R} = \delta \approx 0$, so the angle ϕ is approximately zero, $\phi \ \approx 0$.
Hence, the distance ρ can now be approximated as follows:

$$\rho \approx R \times \big[1 + \delta \cos\big(\omega_g t\big)\big] \tag{4}$$

As theoretically and experimentally shown in [3,6], the amplitude of the gear grinding force in the normal direction F_n is proportional to the grinding depth a_e, the grinding width b_D and the feed rate v_w. Mathematically, the grinding force amplitude F_n can be expressed as follows:

$$F_n = C \times a_e^{f_a} \times b_D^{f_b} \times v_w^{f_v} \tag{5}$$

where C, f_a, f_b and f_v are constants which should be determined experimentally.

Eq. (4) shows that the distance ρ changes in time with the amplitude modulated with the same frequency as the workpiece angular speed. This suggests that the grinding depth (infeed penetration) a_e also varies in time with the same frequency as the workpiece angular speed. This implies that:

$$a_e \propto \rho \propto \big[1 + \delta \cos\big(\omega_g t\big)\big] \tag{6}$$

Hence, by substituting Eq. (6) to (5) and by assuming the other terms are constant, the gear grinding force amplitude can be written as follows:

$$F_n \propto \big[1 + \delta \cos\big(\omega_g t\big)\big] = \hat{C} \ \times \big[1 + \delta \cos\big(\omega_g t\big)\big]^{f_a} \tag{7}$$

where $\hat{C}$ is a constant. Since $\delta \ll 1$, Eq. (7) can thus be approximated by using the Taylor expansion as follows:

$$F_n = \hat{C} \ \times \big[1 + f_a \times \delta \cos\big(\omega_g t\big)\big] \tag{8}$$

It is obvious from Eq. (8) that the normal force amplitude F_n is modulated by the rotational speed of the workpiece $\omega_g = 2\pi f_g$.

According to [7], the dynamic mesh forces between two meshing gears have spectral content at (at least) the harmonics of mesh frequency. By assuming the grinding wheel and the workpiece have similar properties as two meshing gears, the dynamic mesh forces between the grinding wheel and workpiece can be expressed as follows:

$$F_{mesh}(t) = \sum_{k=1}^{K} F_{nk}(t) \sin(2\pi k f_m t + \psi_k(t)) \tag{9}$$

With $f_m = z_g \times \omega_g = z_w \times \omega_w$ denoting the gear meshing frequency between the grinding wheel and the workpiece, the index $k = 1, 2, \ldots, K$ denoting the harmonic order of the meshing frequency, $F_{nk}(t)$ denoting the normal force amplitude at the harmonic order k and $\psi_k(t)$ is the phase which probably varies in time. By substituting Eq. (8) into (9), one can show that the latter can be rewritten as follows:

$$F_{mesh}(t) = \sum_{k=1}^{K} A_k \underbrace{\left[1 + \varepsilon_k \sin(2\pi f_g)\right]}_{\text{Related to gear eccentricity}} \times \sin(2\pi k f_m t + \psi_k(t)) \quad (10)$$

Eq. (10) suggests that workpiece eccentricity/runout errors modulate the gear mesh force amplitude with the frequency the same as the workpiece rotational speed. Hence, one can expect that the resulting vibration signals will also contain some amplitude modulation with the frequency of the workpiece rotational speed around the gear mesh frequency (i.e. the carrier frequency) and possibly the higher order harmonics.

4. Proposed Signal Processing and Feature Extraction Algorithm

In practice, depending on the application, gear grinding process is repeated for a given set of gears (multiple stroke passes). For such applications, it is important to localise the measured vibration signal corresponding to each stroke pass. This way, the physical effects of different passes on the measured vibration signal that are not relevant to the fault detections are eliminated.

During the first pass, the eccentricity will generate a variation of the amount of stock removal around one rotation of the gear. But during the following passes, the amount of stock removal will be constant all around the gear because the first pass will have corrected the variation of stock with respect to the eccentricity. Hence, only the first stroke pass (Pass#1) is relevant for detecting an eccentric workpiece during grinding processes

Based on the aforementioned reasoning and the qualitative model described in Section 3, the workflow of the signal processing and feature extraction for eccentric workpiece/runout fault detection has been established as visualized in Figure 6. The details of each step are described in the subsequent paragraphs.

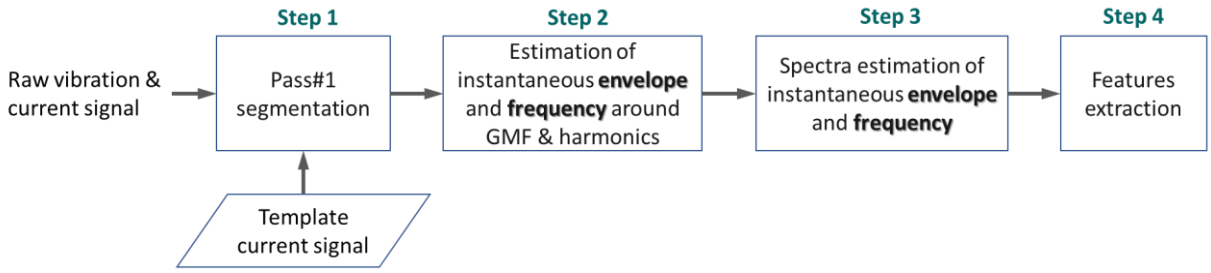


Figure 6: Workflow of the signal processing and features extraction for eccentric workpiece/runout faults detection.

Step 1

As shown in Figure 6, the ingredients to complete this first step are (*i*) the vibration signal recorded on a continuous gear grinding process which is synchronously acquired together with the worm motor current signal and (*ii*) a template current signal corresponding to Pass#1 that needs to be pre-determined by expert supervision. Pass#1 is determined by using the worm motor current signal $I_w(t)$ and the template current signal $I_t(t)$. As the current signal of the worm motor is typically noisy, low-pass filtering on this signal is recommended. The key process in this step is to align the measured current with template current signals. The alignment is attained by first computing the cross-correlation of the two current signals. Later on, the time delay is computed by detecting the lag corresponding to the peak in the resulting cross-correlated signal as described in [8]. Subsequently, the time delay is used to align the two current signals. Once aligned, the sample indices corresponding to Pass#1 can be determined. Finally, the sample indices are used to segment the measured vibration signal corresponding to Pass#1. Note that the segmentation of the signal can be further refined, for example, by chopping off the segmented signal around the maximum current signal. The workflow of this step is schematically illustrated in Figure 7.

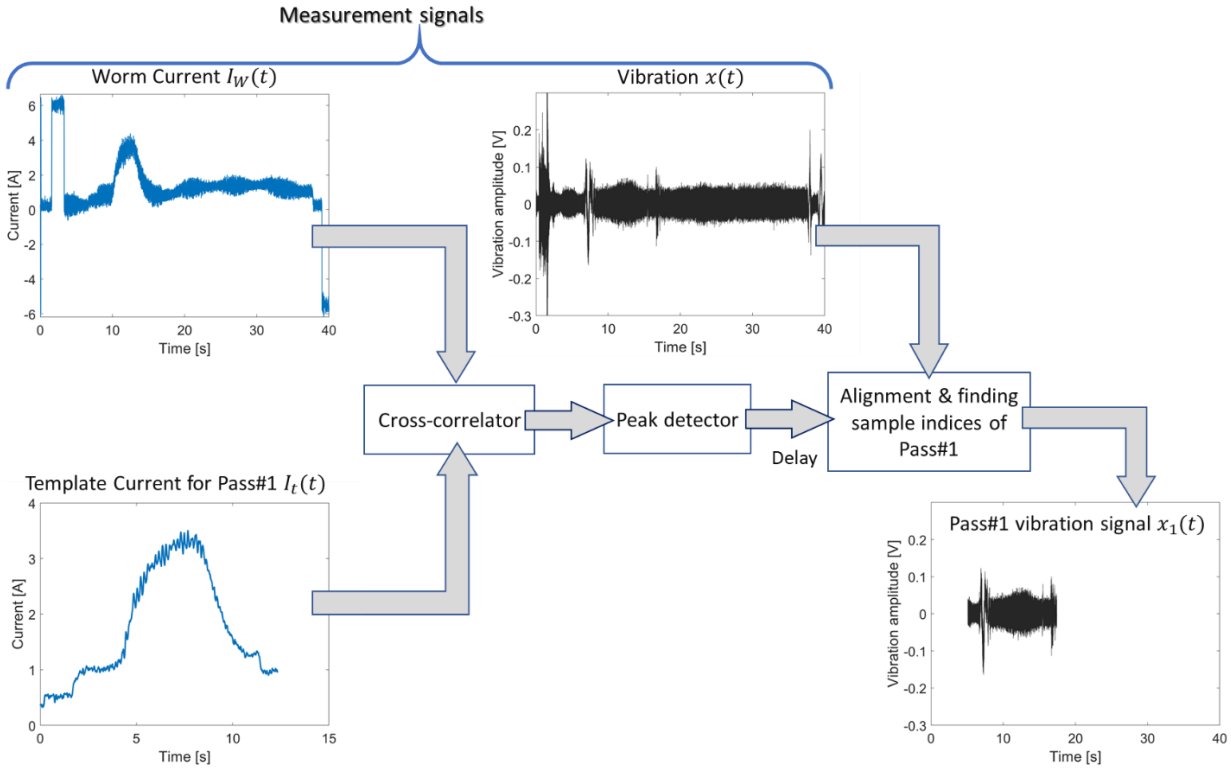


Figure 7: Block-diagram of the vibration signal segmentation of Pass#1.

Step 2

Once the raw vibration signal corresponding to Pass#1 has been segmented following the procedure in Step 1, the segmented signal is then fed into a band-pass filter. The central frequency of the band-pass filter is at the gear mesh frequency (GMF) between the workpiece and the worm, while the bandwidth of the filter is fixed about 4 times of the rotational speed of the workpiece.

The instantaneous envelope (IE) and the instantaneous frequency (IF) of each band-pass filtered signal are estimated using the Hilbert Transformation [9]. For an arbitrary signal $x(t)$, its Hilbert transform is defined as

$$H[x(t)] = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{x(\tau)}{t-\tau}d\tau \tag{11}$$

And the analytical signal $z(t)$ of $x(t)$ is determined as

$$z(t) = x(t) + jH[x(t)] = A(t)e^{j\phi(t)} \tag{12}$$

In this case, the IE $A(t)$ is given by

$$A(t) = |z(t)| = \sqrt{x^2(t) + H^2[x(t)]} \tag{13}$$

And the instantaneous phase $\phi(t)$ is calculated as follows

$$\phi(t) = \arg[z(t)] = \arctan\left[\frac{H[x(t)]}{x(t)}\right] \tag{14}$$

Furthermore, the IF $\upsilon(t)$ can be computed by taking the time derivative of the instantaneous phase

$$\upsilon(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{15}$$

Step 3

The frequency spectra of the IE and IF are determined by applying the Fourier transform to $A(t)$ and $v(t)$ as described in the following equations:

$$X_{IE}(f) = \mathcal{F}[A(t)] = \int_{-\infty}^{+\infty} A(t)\, e^{-j2\pi ft} dt \tag{16}$$

$$X_{IF}(f) = \mathcal{F}[v(t)] = \int_{-\infty}^{+\infty} v(t)\, e^{-j2\pi ft} dt \tag{17}$$

With $X_{IE}(f)$ and $X_{IF}(f)$ denoting the frequency spectra of the IE and IF respectively.

Step 4

The IE and IF features are defined as the average magnitude of their corresponding spectrum within a given frequency range ($[f_l \;\; f_u]$). Based on this definition, the IE and IF features can be mathematically expressed as the following equations:

$$\text{IE Feature} = \frac{1}{f_u - f_l} \int_{f_l}^{f_u} X_{IE}\, df \tag{18}$$

$$\text{IF Feature} = \frac{1}{f_u - f_l} \int_{f_l}^{f_u} X_{IF}\, df \tag{19}$$

The frequency range ($[f_l \;\; f_u]$) should be selected such that a few harmonics of the *workpiece rotational speed* are included. In this study, f_l is set at 2 Hz and f_u is set at 20 Hz.

5. Experimental Study

5.1 Experimental Setup

An industrial grinding machine depicted in Figure 8(a) has been used to emulate different grinding faults, namely *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. Two triaxial accelerometers were mounted on the machine, the first one on the grinding worm holder and the second one on the base of the workpiece spindle. From the previous investigation [1], it can be concluded that one accelerometer mounted on the grinding worm holder is sufficient for gear grinding monitoring purposes, which is the focus of this study. Note that the axes orientations of the triaxial accelerometer in this study are different from the ones in the previous study [1]. The accelerometer axes orientations of this study are the following:

- X-axis: *Parallel to the worm axis*
- Y-axis: *Direction of Infeed* (and small component in direction of Feed)
- Z-axis: *Upward, the direction of Feed* (and small component in direction of Infeed)

(a) (b)

Figure 8: (a) Experimental setup of the grinding machine, (b) Emulation of an eccentric workpiece.

5.2 Design of Experiment

In this study, three types of gear grinding faults are considered, namely *i*) *high feed rate*, *ii*) *high infeed* and *iii*) *eccentric workpiece*. Moreover, two severity levels of each fault type are considered as summarised in Table 1. The eccentric workpiece is emulated by mounting the clamping arbour eccentrically as shown by Figure 8(b). To seed the eccentric fault, the clamping arbor was mounted with an eccentricity. This can be done by playing with 3 centring bolts at the base of the clamping arbor. The actual eccentricity was checked using a dial indicator as shown in the photograph.

Test Classification	Settings
Baseline	Grinding with the nominal (optimised) parameters
Feed 1	The feed rate is 75% higher than the nominal feed rate
Infeed 1	The infeed is 80% higher than the nominal infeed
Excenter 1	Nominal grinding parameters. The workpiece is mounted eccentrically at 20 µm
Feed 2	Feedrate is 150% higher than the nominal feed rate
Infeed 2	The infeed is 150% higher than the nominal infeed
Excenter 2	Nominal grinding parameters. The workpiece is mounted eccentrically at 40 µm

Table 1. The settings of each test condition

For each type of fault, five measurements are captured sequentially by grinding five gears (workpieces) in a row. Prior to performing tests in a faulty state, five measurements are realised by grinding five gears with the nominal parameters to set up the baselines. Moreover, the grinding wheel is redressed prior to the baseline tests. The complete test sequence is illustrated in Figure 9, so in total there are 60 tests.

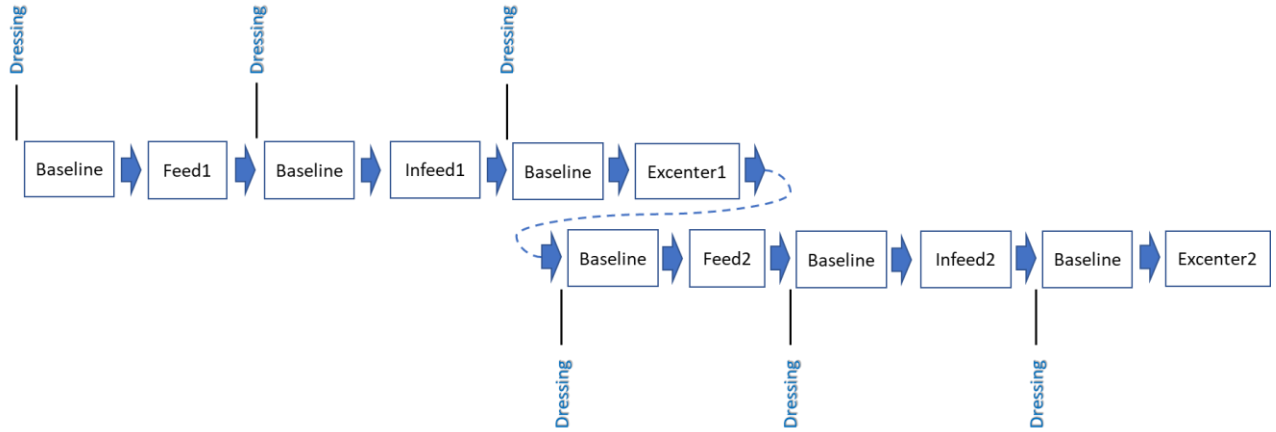


Figure 9: Overview of the test sequence

The grinding program is realised in two-stroke passes. As mentioned earlier in the previous section that only the first pass (Pass#1) is relevant for detecting eccentric workpieces/runouts during grinding processes. The duration of the signal is approximately about 40 seconds and the sampling frequency was selected at 25.6 kHz.

6. Results and Discussion

After completing the test campaign, a number of gear quality parameters of individual ground gear were measured using a dedicated gear metrology machine. For the analysis of this particular eccentric workpiece/runout fault test data, only the runout error Fr is relevant as discussed earlier. Figure 10 shows the boxplot of the measured runout error for each test class. As seen in the figure, only the *Excenter#2* (i.e. eccentricity of 40 µm) test class leads to significantly increased runout errors. Moreover, one can clearly see that the *Excenter#2* test class is very well separated with the *Baseline* test class and the other test classes including *Feed#1*, *Infeed#1*, *Excenter#1*, *Feed#2*, and *Infeed#2*. Notably, the other fault types (high infeed and high feed rate) do not lead to significantly increased runout errors.

Despite the fact that all the gears are ground with the nominal (optimised) parameter in the *Baseline* tests, there still exists some residual runout error on the ground gears. As can be seen in the figure, the averaged runout error of the gears in the baseline tests is about 8 µm. The residual runout errors could be caused by existing runout errors on the incoming workpiece due to imperfect pre-grinding processes.

Although the eccentricity introduced on the workpiece in the *Excenter#2* tests is about 40 µm, the resulting averaged runout error is about 65 µm, which is higher than the workpiece eccentricity. This suggests that the grinding processes seem to add more runouts on the ground gears. On the other hand, the averaged runout error resulting from the *Excenter#1* test class is about 10 µm, which is lower than the workpiece eccentricity of 20 µm. It is not clear yet till now, why for the *Excenter#1* test the resulting runout error is lower than the workpiece eccentricity. A possible explanation is that the emulation method for introducing an eccentric workpiece fault of 20 µm was not accurate enough, see Figure 8(b).

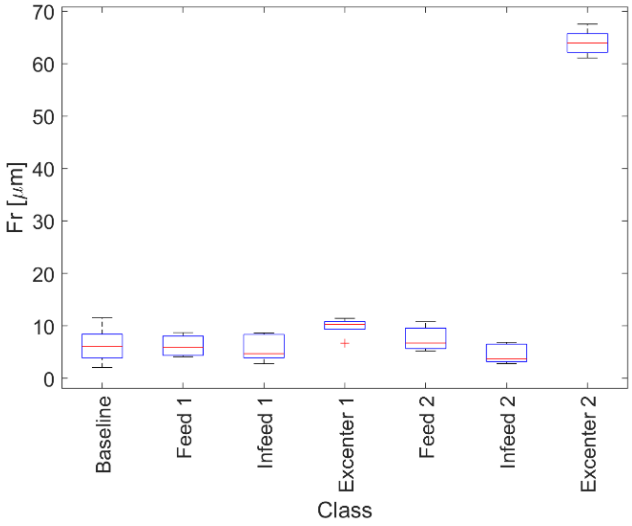


Figure 10: The measured runout error Fr of different test classes.

The measured vibration data of each test class has been processed and analysed using the algorithms proposed in Section 4. Initially, the raw vibration signal is segmented according to the event corresponding to Pass#1. The segmentation steps of the vibration signal corresponding to Pass#1 from the raw vibration signal have been visualised in Figure 7. Figure 11 shows the resulting signals of a *Baseline* test and an eccentric workpiece test (*Excenter#2*) after applying the Pass#1 segmentation algorithm. Visually, it is seen that the overall vibration amplitude of the eccentric workpiece test is slightly higher than that of the *Baseline* test. However, at this point, it is still difficult to justify the difference between the two signals.

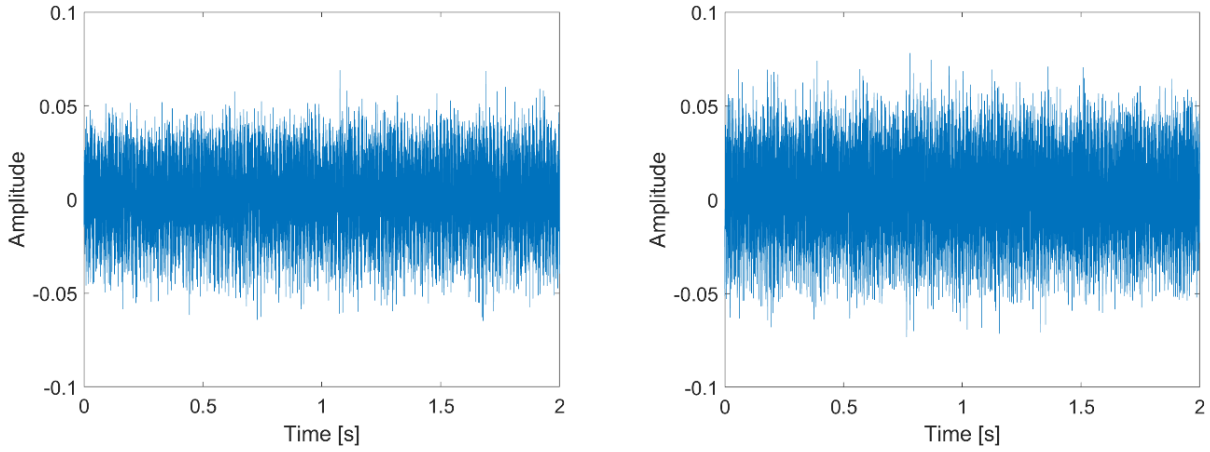
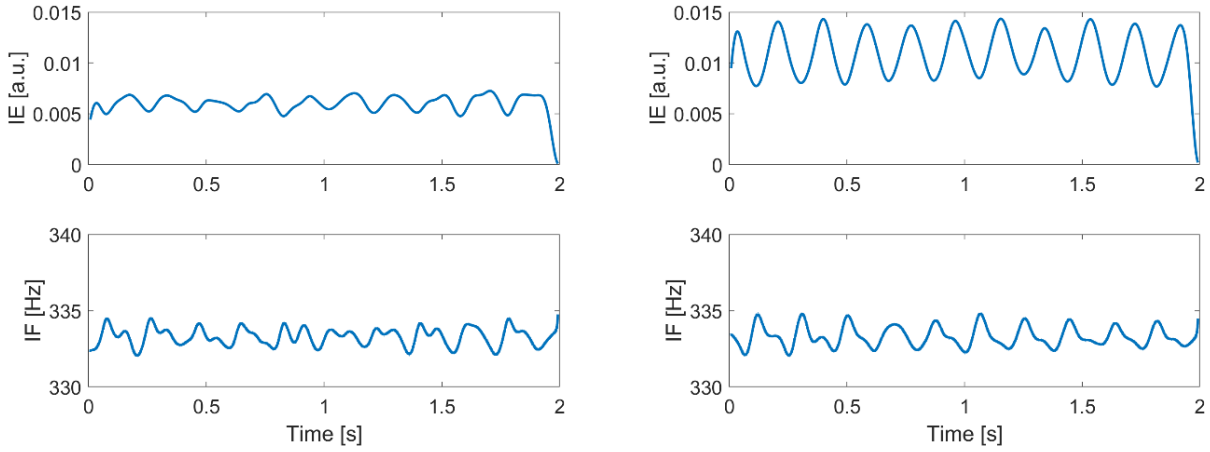


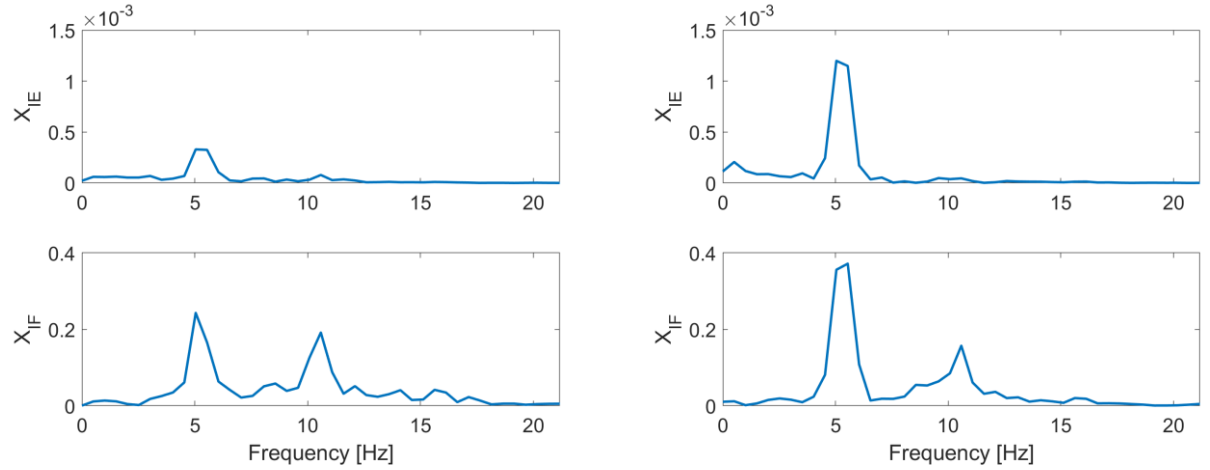
Figure 11: Typical vibration signals segmented from Pass#1 measured in the *infeed direction*/Y-axis of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test.

Once the vibration signal corresponding to Pass#1 has been segmented, the instantaneous envelope (IE) and instantaneous frequency (IF) are then estimated using Eq. (13) and (15), respectively. Figure 12 shows the resulting IE and IF signals calculated from the segmented signals of the baseline and eccentric workpiece test. As seen in the figure, the IE varies periodically in time and the IF around the fundamental gear mesh (GMF), $\approx$ 333 Hz, also varies periodically in time. It becomes more apparent from the figure that the proposed signal processing steps for computing the IE and IF help us to highlight the difference between the eccentric workpiece (*Excenter#2*) test and the *Baseline* test. Furthermore, the IE and IF are transformed into the frequency domain as shown in Figure 13. The difference between the two test conditions is even more highlighted by comparing the magnitudes of the IE and IF spectra, respectively X_{IE} and X_{IF}, around the workpiece rotational speed and the magnitudes of the higher order harmonics.



(a) (b)

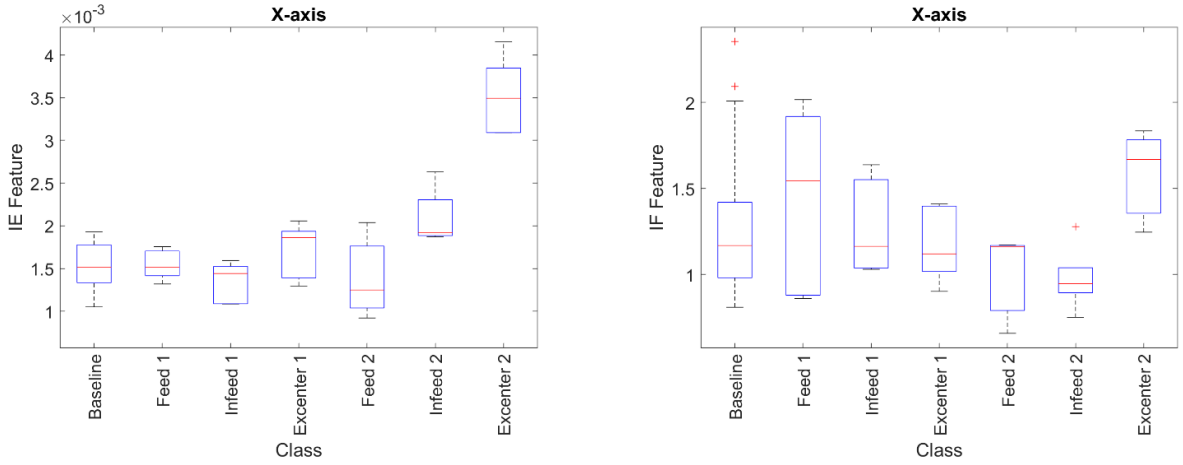
Figure 12: The estimated IE and IF of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test.



(a) (b)

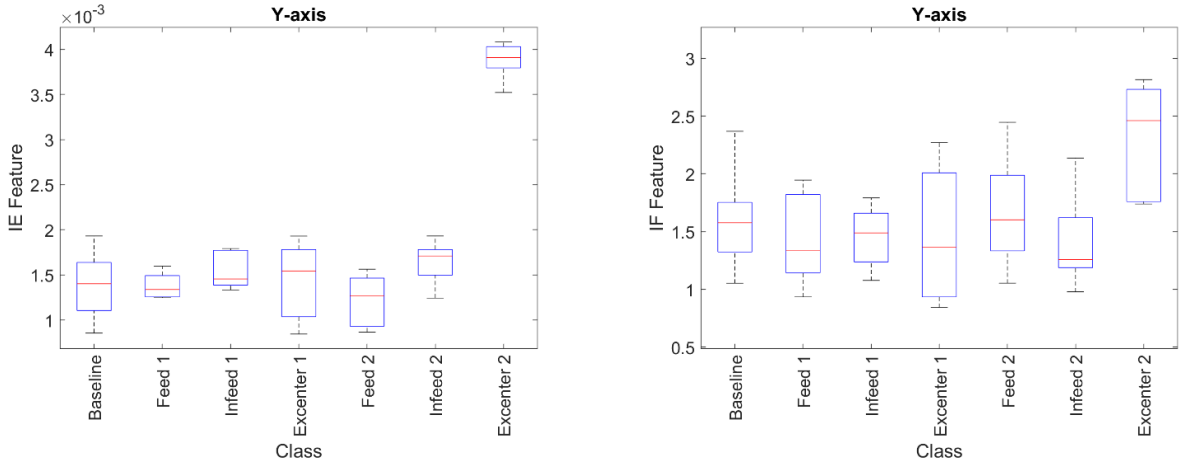
Figure 13: The frequency spectra of IE and IF of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test. Note that the fundamental frequency of 5 Hz corresponds to the workpiece (gear) rotational speed f_g.

The final step is to compute the IE and IF features from the frequency spectra based on Eqs. (18) and (19). Figure 14, Figure 15 and Figure 16 show the IE and IF features extracted from the X-axis, Y-axis and Z-axis signal, respectively. As seen in the figures, in general, the IE feature shows a strong correlation with the measured runout error shown in Figure 10. On the other hand, the IF feature shows a weak correlation with the measured runout error. Notably, the strongest correlation of the IE feature with the runout error can be observed on the measurement data in Y-axis, which is in the *infeed* direction. This is expected since the workpiece eccentricity/runout errors affect significantly the grinding forces in the infeed direction, as discussed earlier in Section 3. A rather weak correlation between the IE feature and the runout error can be observed on the measurement data of Z-axis, which is in the *feed* direction.



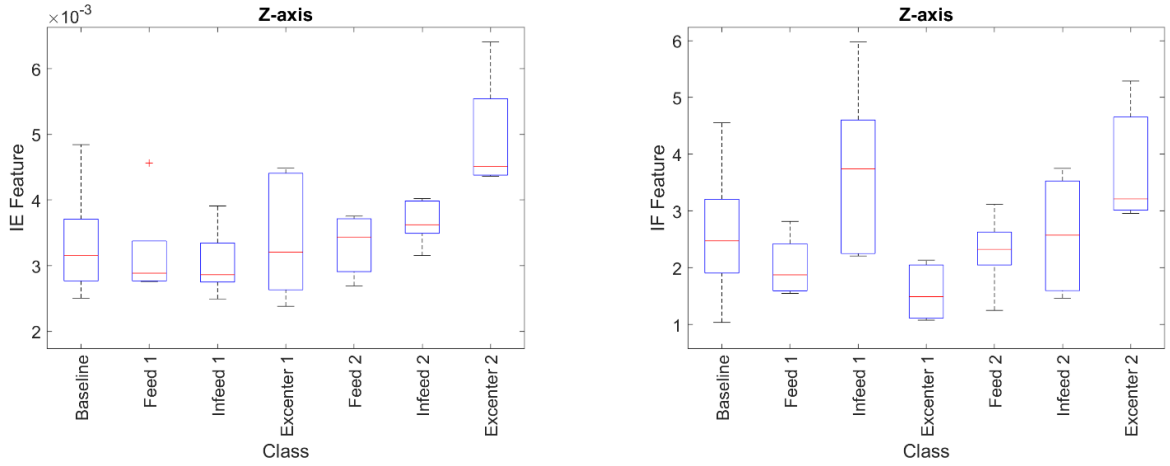
(a) (b)

Figure 14: (a) The IE feature and (b) the IF feature of the **X-axis** signal from all test conditions.



(a) (b)

Figure 15: (a) The IE feature and (b) the IF feature of the **Y-axis** signal from all test conditions.



(a) (b)

Figure 16: (a) The IE feature and (b) the IF feature of the **Z-axis** signal from all test conditions.

7. Conclusions and Outlook

Eccentric workpiece/runout error is one of the gear errors affecting the operating quality of gears and in particular the running noise behaviour of the end product. For gear manufactures, it is important to detect this fault to prevent sending defective gears to customers. A qualitative model developed in this paper has enabled us to derive a novel vibration feature for detecting eccentric workpiece/runouts. The feature extracted from the vibration signals measured at different test conditions/other fault types shows that it is sensitive and robust for detecting workpiece eccentricity of 40 microns. It is also shown in this study that the feature is insensitive to other types of gear grinding faults, which is important for diagnostics/root-cause analysis purposes. Furthermore, it is also shown in the study that the most sensitive measurement direction to detect workpiece eccentricity/runout errors is in the *infeed* direction.

Future work will be on the investigation of the minimum workpiece eccentricity/runout error that can still be detected by the proposed vibration feature during gear grinding processes.

Acknowledgements

This research was supported by both Flanders Make vzw, the strategic research centre for the manufacturing industry, and VLAIO in the frame of the VIBMON_ICON project.

References

[1] K. Gryllias, B. Kilundu, S. Devos, Y. Vonderscher, B. Vandewal, *Condition monitoring of gear grinding processes*, Surveillance 9, Fez (Morocco), 2017.

[2] B, Karpuschewski and I. Inasaki, Monitoring Systems for Grinding Processes, *Springer London*, 2006, 42,83-107.

[3] C. Brecher, M. Brumm, F. Hubner, *Approach for the calculation of cutting forces in generating gear grinding*, 9th CIRP Conference on Intelligent Computation in Manufacturing Engineering, Procedia CIRP 33, pp. 287 – 292, 2015.

[4] https://slideplayer.com/slide/6077265/ (Accessed on 9th of May 2019)

[5] https://www.kggear.co.jp/en/gear-technical-data (Accessed on 9th of May 2019)

[6] B.W. Kruszynski, S. Midera, Forces in Generating Gear Grinding – Theoretical and Experimental Approach, Annals of the CIRP Vol. 47, 1998.

[7] R.G. Parker, S.M. Vijayakar, T. Imajo, *Non-linear dynamic response of a spur gear pair: Modelling and Experimental Comparisons*, Journal of Sound and Vibration, 273(3), pp. 435-455, 2000.

[8] J. Ianniello, *Time delay estimation via cross-correlation in the presence of large estimation errors*, in IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 30, no. 6, pp. 998-1003, 1982.

[9] M. Feldman, Hilbert Transform Applications in Mechanical Vibration, *John Wiley & Sons*, Ltd., 2011.

Toward the quality prognostic of an aircraft engine workpiece in Inconel Alloy 625: case study and proposed system architecture

Antoine PROTEAU[1], Marc THOMAS[1] and Antoine TAHAN[1]

[1]École de technologie supérieure, Montréal, QC, Canada
antoine.proteau.1@ens.etsmtl.ca

Abstract

Manufacturing companies are under a constant pressure due to multiple factors: new competition, disruptive innovations, cost reduction request, etc. To survive, they must strive to innovate and adapt their business model to improve their productivity. Recent developments based on the concept of Industry 4.0 such as big data, new communication protocols and artificial intelligence provide several new avenues to explore. In the specific context of machining, we are working toward the development of a system capable of making the prognostic of the quality (in terms of dimensional conformance) of a workpiece in real time while it is being manufactured. The goal of this paper is to showcase a prototype of the data acquisition aspect of this system and a case study presenting our first results. This case study has been conducted at our industrial partner facility (Quebec, Canada) and is based on the manufacturing of an aircraft component made from Inconel alloy 625 (AMS5666). The proposed prototype is a data acquisition system installed on a 5 axis CNC machines (GROB model G352) used to acquire and to contextualize the vibration signal obtained from the CNC machine sensor. The contextualization of the data is a key component for future work regarding the development of a prognostic system based on supervised machine learning algorithms. In the end, this paper depicts the system architecture as well as its interactions between the multiple systems and software already in place at our industrial partner. This paper also shows preliminary results describing the relationship between the workpiece quality (in terms of respect toward the dimensional requirements) and the extracted features from the sensors signals. We conclude that it is now possible to do the diagnostic of a cutting operation. Additionally, with the same information we show that it is possible to quickly do the general diagnostic of the health state of the machine. Future work regarding this project will include data acquisition from a wider range of products (i.e. different shapes, materials, processes, etc.) and the development of a machine learning based prognostic model.

1 Introduction

Fuelled by the rapid evolution and introduction of new technologies and new philosophies such as Industry 4.0, the manufacturing industry is quickly transforming. This new manufacturing era brings a lot of possibilities to an industry that is under constant pressure for cost reduction and better quality caused by a global competition [1]. In the specific context of machining; automation and methodologies such as lean manufacturing were the go to solutions to decrease process cost and improve quality output. However, in this new age, possibilities brought by artificial intelligence, more affordable technologies such as sensing technologies and collaborative robotic offer new improvements directions.

In this context, the objective of our research project is to see if we can connect the operational information of a machining process to the physical phenomenon happening during the machining of a workpiece on a CNC machine in order to be able to predict the quality of this workpiece in real-time. Thus, the objective of this paper is to propose a data acquisition system architecture based on the prototype we built, showcase that it is now possible to do the diagnostics of a cutting operation with this system and that we are now able to put in relationship the quality, in terms of the conformity towards a workpiece's G&DT specifications, and the physical phenomenon happening during the machining process.

In a general manufacturing context, attempts have been made to try to predict the quality of a production process. For instance, Wang [2] tried to predict the quality of a chemical batch process operation. However,

their results are based on simulated data and not industrial data such as what we propose. Closer to the machining industry, through our exploration of the literature we have not yet found authors who have proposed a methodology to predict the quality of a whole machining process and the produced workpiece. Nevertheless, we can find articles related to the prognostic of some aspect of a machining process such as predicting the surface roughness. In that context, Benardos, Vosniakos [3] propose a review of the works that have been done in that domain and more recently, Balamurugamohanraj et al. [4] used a machine learning approach and data from an accelerometer to predict the surface roughness in terms of its Ra value.

Even though we have not found many publications with industrial application of prognostic methodology related to the quality of a workpiece, we clearly see an interest for the concept of prognostic in the manufacturing industry. Reviews and publications by authors such as Vogl et al. [5], Wang [6], Peng et al. [7],Lee et al. [8] are all dedicated to the state of the prognostic concept or the proposal of a framework related to manufacturing. Thus, we are not the only one with interest in applying these concepts to a manufacturing context. Still, one of the biggest challenge to the industrial application of such concept and the development of prognostic methodologies is the access to data of good quality and in sufficient quantity. The foremost challenge is addressed in this article.

We also see that, in our research domain, the interest related to applying prognostic methodologies is strong in fields related to tool wear prediction and condition-based maintenance. For instance, Proteau et al. [9] showed that it is possible to predict the tool wear with a Long Short-Term Memory (LSTM) neural network. Balan, Epureanu [10, 11] and Aghazadeh et al. [12] also proposed different methodologies based on artificial intelligence approaches to monitor and predict the cutting tool condition. Related to condition-based maintenance, Waqar, Demetgul [13] and Aydin, Guldamlasioglu [14] also suggested methodologies based on artificial intelligence to predict the state of an equipment or a component (e.g. bearing, gears, etc.).

To improve the state of this research domain and to make a step toward the industrial application of prognostic methodologies, this paper will present our most recent work to show that it is now possible to put the workpiece quality in relationship with the physical phenomenon happening during the machining process. We also want to prove that we are now making a step forward to go from being able to diagnose a cutting operation toward being able to predict the quality of that process. To do so, this article is structured as follows: section 2 will introduce our research partner as well as our research environment and equipment. Then, in section 3, we propose a data acquisition (DAQ) system architecture and describe the dataset built. In section 4, we present our signal processing methodologies and the different features that we extracted from the acquired signals. In section 5, we detail our results and show that it is now possible to do the diagnostic of a cutting operation as well as working toward the prognostic of the quality of a workpiece. Finally, in section 6, we make our conclusions.

2 Research environment

To pursue this research project, we are collaborating with an industrial partner: APN Inc.[1] APN is a leader of the machining industry as well as at the forefront of the Industry 4.0 movement in Quebec, Canada. They are specialized in the machining of complex products in exotic material (i.e. titanium, Inconel, etc.) for the aerospace and high-tech industry.

In our research context, our work was done on a 5 axis CNC machine made by GROB, model G352 (see Figure 1). This machine was acquired in 2017.

[1] More information here : http://apnglobal.ca/en/apn/

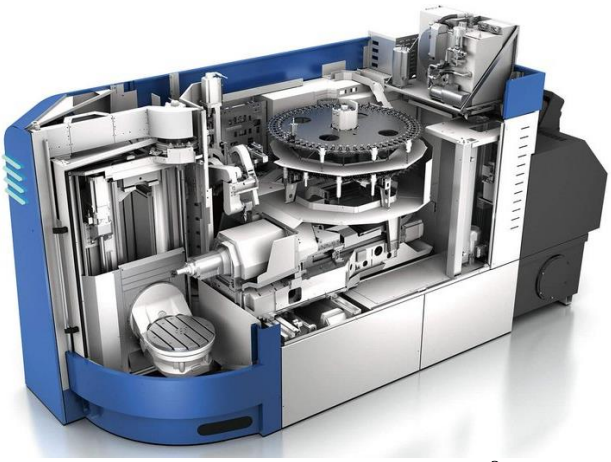


Figure 1 GROB G352 CNC machine[2]

It is also important to state that, to be able to acquire a vibration signal, we worked with the GROB employees to have access to the accelerometer already installed into the machine's spindle. Thus, the signal was acquired through an IFM VSA004[3] accelerometer on which the signal was amplified with a Phoenix Contact signal conditioner model MACX MCR-UI-IU[4]. From Figure 2, we can see where the accelerometer was installed by the manufacturer (as indicated by the bubble #1). This information was provided by the GROB documentation available at APN Inc. In the next section, we present our DAQ system architecture.

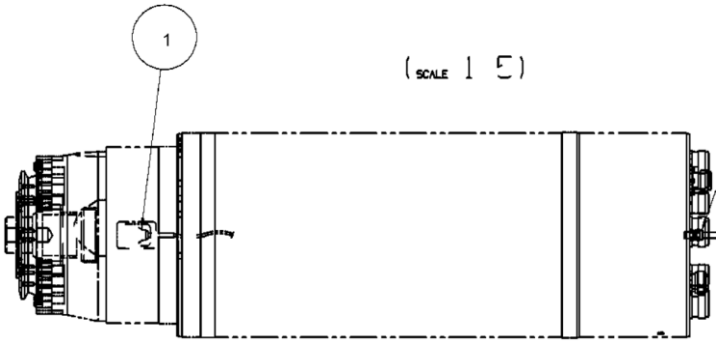


Figure 2 Accelerometer location in the spindle

3 Data acquisition system architecture

One of our hypothesis is that, in order to be able to predict the quality of a workpiece based on the physical information of the CNC machine, we must contextualize the signals acquired from sensors. Therefore, we developed a data acquisition system to automatically execute this operation. Figure 3 shows the contextualization of the data in terms of its relationships.

[2] Picture source : https://www.grobgroup.com/en/products/product-range/universal-machining-centers/milling-centers/g350/
[3] Specifications : https://www.ifm.com/ca/en/product/VSA004
[4] Specifications : https://www.phoenixcontact.com/online/portal/ca?uri=pxc-oc-itemdetail:pid=2811446&library=caen&tab=1

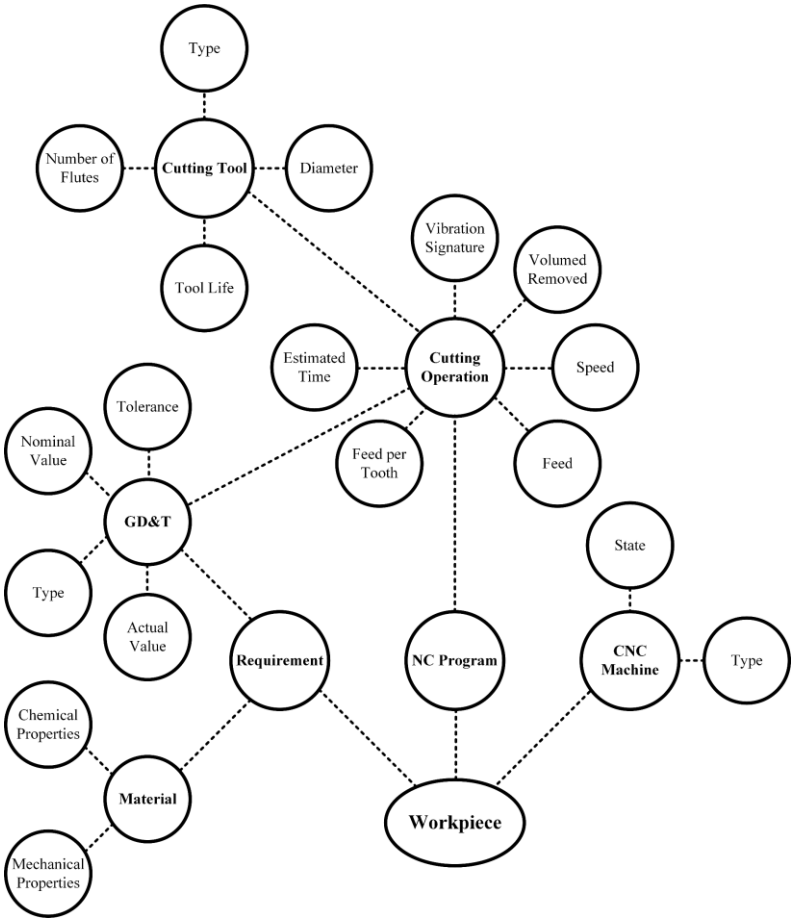


Figure 3 Relationships between data sources

Whit this figure, we can see that, through this system, it is possible to create a relationship that goes from the workpiece requirements (including the actual measurements made on a finished workpiece) up to the vibration signature of a specific cutting operation. This means that at every moment during the machining process, we can know which cutting operation was being executed, its vibration signature, what was the cutting tool and its cutting parameters as well as which GD&T was influenced. To achieve these relationships, multiple data sources must be integrated. Table 1 shows the source of each data types.

Data types	**Sources**
Workpiece Requirements	APN Quality System (From the technical drawing)
Actual Measurements	APN Quality System
Material Properties	APN Quality Documents System
CNC Machine information	Machine controller through OPC Protocol
Cutting Operation	CAM Software
Cutting Tool	CAM Software
Vibration signal	Accelerometer and National Instruments Card

Table 1 Data sources by data types

To automatically integrate these data sources and create the relationships between the data, we developed an acquisition system that had to take into account the state of the machine (cutting or not, on idle, etc.). To illustrate our acquisition system, Figure 4 shows an overview of the acquisition process. On this figure, we can also see the isometric view of the workpiece as well as the quality data flow.

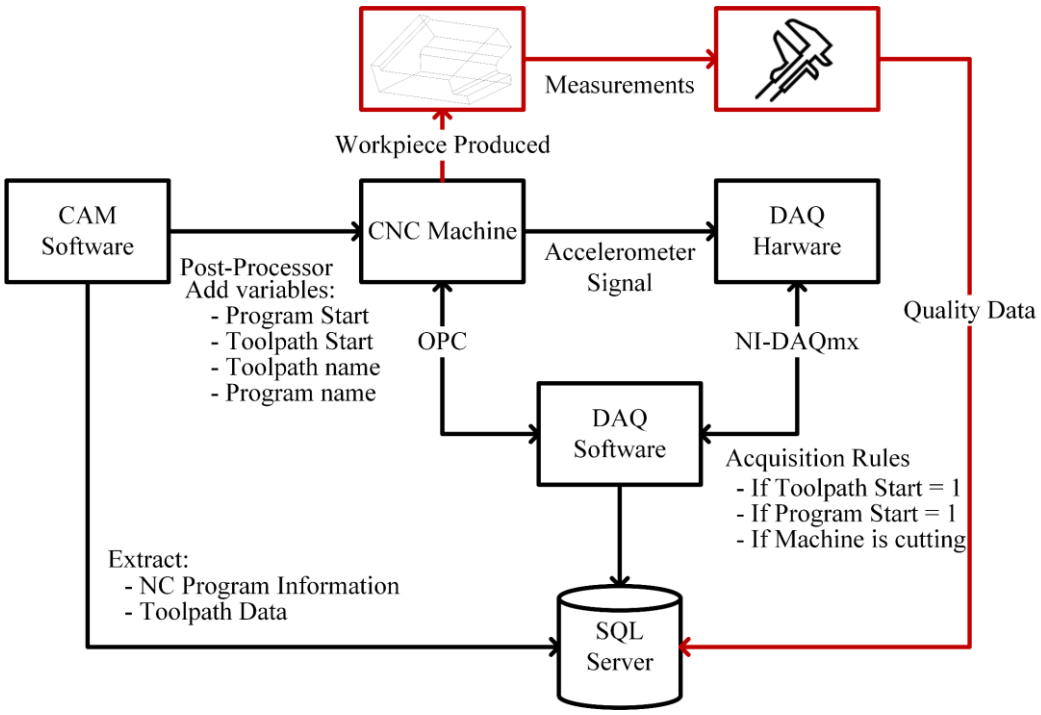


Figure 4 DAQ System architecture

We worked with our industrial partner to modify their post-processor program to add four variables: when the NC Program start/stop, when a cutting operation start/stop, the cutting operation name and the NC Program name for reference. This modification allowed us to control the behaviour of the National Instruments data acquisition card by sampling the signal only when the machine was actually cutting the workpiece. However, since we are in a production environment, different machine states can also arise: the machine is in idle during a cutting operation, an alarm is raised, etc. Therefore, in our acquisition rules, we added some logic based on variables extracted in real-time from the CNC machine controller through the OPC communication standard. The reader can refer himself to the Siemens Sinumerik 840D SL documentation for a complete list of all available variables.

During the acquisition process, the raw signal is thus contextualized and attached to the current workpiece and the current cutting operation being executed on the CNC machine. This, consequently, gives us a contextualized vibration signal suited to model the relationships between the physical and operational data (the model input) and the actual measurement in terms of GD&T (the model output). The next section presents an overview of the data collected.

3.1 Dataset overview

To prove the concept and functionalities of our proposed system, we conducted a first acquisition process. The acquisition was made during the machining of an Inconel 625 (AMS5666) workpiece intended for the aerospace industry. During our acquisition process, we were able to cover the entirety of the machining process which means:

- 22 different cutting tools;
- 140 cutting operations of multiple types;
- 135 GD&T to be respected for a workpiece to be considered conform.

The acquired data covers five finished workpieces which represents approximately 13 hours of machining process. Unfortunately, due to industrial constraints, we were not able to gather more workpiece. However, this information is sufficient to prove our concept and start our analysis. In this article, we focused on one specific operation where measurements were made for every workpiece to showcase our results. Information related to this operation can be found in Table 2. Information regarding the cutting tool used during the cutting

operation can be found in Table 3. The cutting tool was new at the beginning of the machining process and was not changed during the machining of the five workpieces. For visualization, Figure 5 shows the cutting operation strategy obtained from the CAM software and Figure 6 shows the difference between the finished workpiece and the raw material used. Due to confidentiality, the 3D model shown in this paper have been redesign to showcase the overall shape of the workpiece and not the actual geometry.

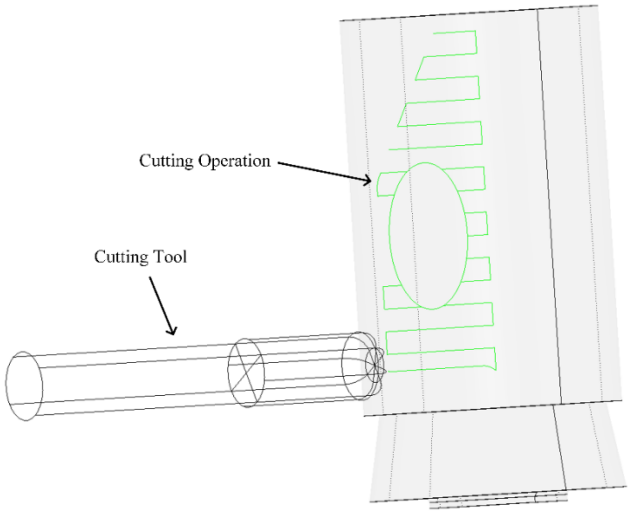


Figure 5 Cutting operation *OP_510*

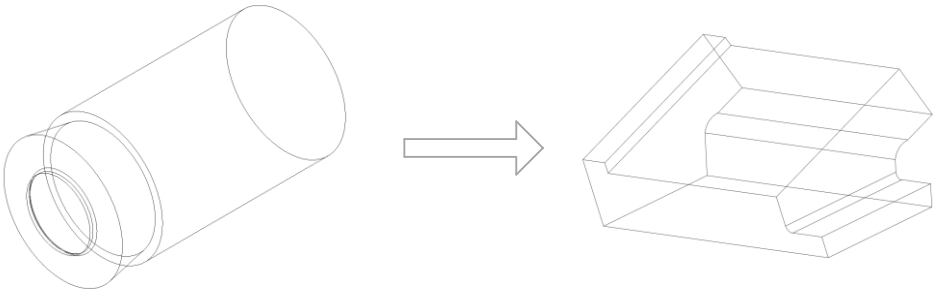


Figure 6 Difference between the raw material and the finished workpiece

Name	Type	Spindle speed [RPM]	Feed [mm/s]	Est. cutting time [min]	Volume removed [cm^3]	Coolant pressure [MPa]
OP_510	Face Milling	1047	233.934	2.5	0.078	6

Table 2 Cutting operation information

Name	Type	Diameter [mm]	Radius [mm]	Number of flutes
EMR0.500R0.125L4LG1.100	Radius End Mill	12.7	3.175	4

Table 3 Cutting tool information

By connecting our system to the quality system of APN, we are able to associate the cutting operation with the specifications (GD&T) influenced by that operation. These associations are made by expert employees at APN. Thus, for this project, we assumed that the associations are good. Through these associations, we know that the *OP_510* operation influences the specification #15. Details about this specification is found in Table 4.

Number	GD&T type	Minimum value [µm]	Maximum value [µm]	Severity	Illustration [µm]	Inspection tool used
15	Flatness	0	25.4	Critical	25.4	CMM

Table 4 Details of the GD&T specification #15

The interpretation of this type of GD&T was made according to the standard Y14.5 [15]. The reader can refer to ASME [15] for further details.

For the five workpieces, Figure 7 shows each actual measurements made by the operator after each workpiece was produced.

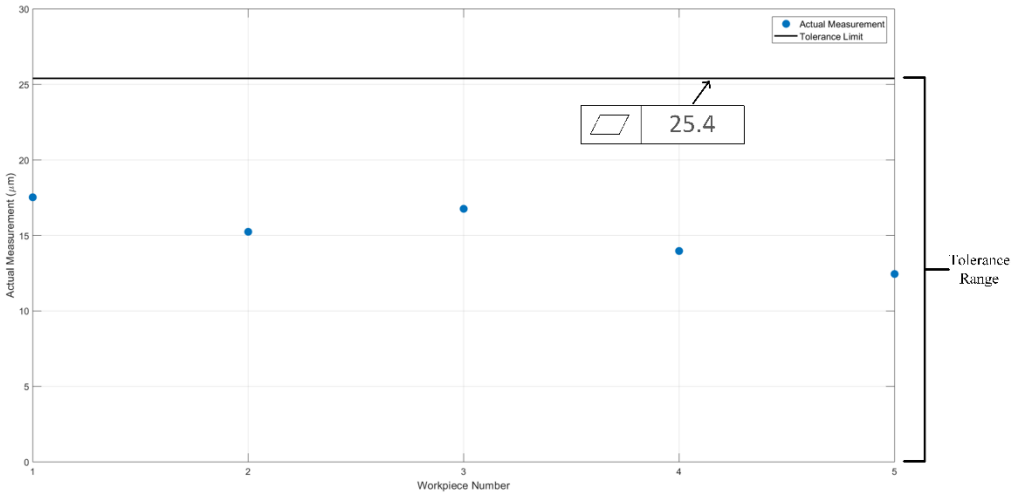


Figure 7 Actual measurement per workpiece for the specification #15

The next section will present the signal processing methodology applied to the data acquired.

4 Signal processing methodology

We have shown in the previous section that we can now acquire a signal that is well contextualized. The objective of this section is to describe our signal processing methodology in order to be able to do the diagnostic of the machining process as well as the production equipment itself; taking a step toward a predictive methodology.

Our methodology is segmented in two sections: a time domain methodology and a frequency domain methodology. Once acquired, each signal sampled file is cleaned and has several features extracted. The signal features used are described below and are chosen according to the work of Lei et al. [16], Elattar et al. [17] and Abellan-Nebot, Romero Subirón [18].

4.1 Time domain

To describe the signal in the time domain, we used equation (1) to equation (5) which refer respectively to the Root Mean Square (RMS), the Kurtosis (K), the Peak value ($Peak$), the Peak-to-Peak value (PTP) and the Crest Factor (CF).

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{1}$$

$$K = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{\frac{1}{N}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^2} \tag{2}$$

$$Peak = \max(x) \tag{3}$$

$$PTP = \max(x) - \min(x) \tag{4}$$

$$CF = \frac{Peak}{RMS} \tag{5}$$

Where x is a signal of N samples, x_i is the value of i^{th} sample and $\overline{x}$ is the average of x.

4.2 Frequency domain

In the frequency domain, we are interested in following the evolution, through time, of the cutting tool frequency (A_{CT}). This is because the degradation of the tool is one of the major cause of the degradation of a machining process. To do so, we used equation (6). This equation is similar to the RMS equation in the sense that we could interpret its result as being the energy content of the signal around a specific frequency.

$$A_{CT} = \sqrt{\sum_{i=a}^{b} A_i^2} \tag{6}$$

Where A_i is the amplitude of the signal at the i^{th} frequency and a and b are the two corners of the window. In our case, we assigned $a = 64$ Hz and $b = 75$ Hz, which correspond to ±5 Hz around the frequency of interest (f_i). We are also interested in making the same measurement at the different harmonics of the cutting tool frequency, thus we also applied the same equation at the 2nd, 3rd, 4th and 5th harmonics ($2f_i$, $3f_i$, $4f_i$ and $5f_i$).

To improve the physical meaning of this feature we use the work of Proteau et al. [9]. In Proteau et al. [9], the authors proposed an adaptation of the specific cutting energy (SCE, k_c) metric first established by Debongnie [19]. In their work, the authors used their version of the SCE to show that it can adequately represents tool wear degradation. Their version is defined by equation (7).

$$k_c = \frac{P_{Tool}[\mathrm{W}]}{Q[\mathrm{cm^3 s^{-1}}]} = \frac{[\mathrm{J s^{-1}}]}{[\mathrm{cm^3 s^{-1}}]} = [\mathrm{J cm^{-3}}] \tag{7}$$

Where P_{Tool} is the power [W] consumed by the cutting tool and Q is the material removal rate express in $\mathrm{cm^3 s^{-1}}$.

SCE is therefore defined in terms of the energy required to remove and keep a certain rate of material removal in a specific material (aluminum, Inconel, etc.). The reader can refer himself to Proteau et al. [9] for the details. Since we do not have the actual power transmitted to the cutting tool, we can estimate this value by using the energy contained in the signal at the frequency related to the cutting tool (A_{CT}). For the same material and a constant material removal rate, k_c should be constant. In case of tool wear, k_c increase through time. The next section will present the results of our analysis.

5 Results and discussion

We first present the results of our analysis in the time domain. Figure 8 shows the values of the RMS, $Peak$ and PTP values through time for each workpiece.

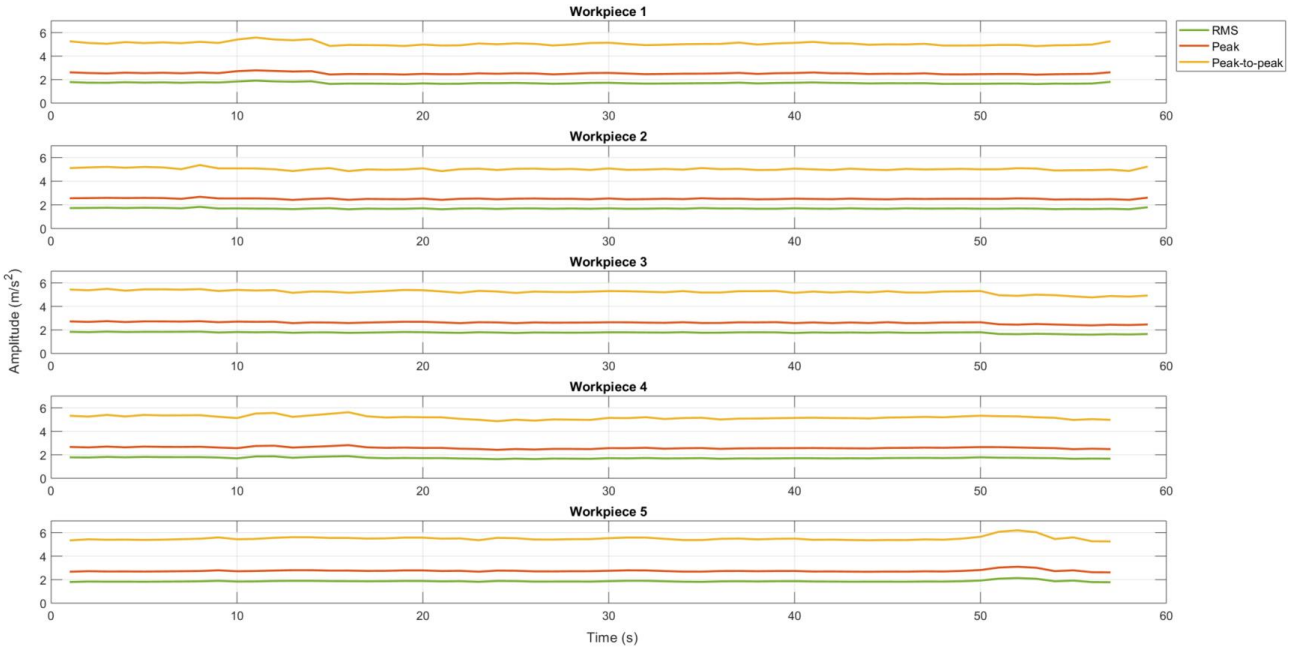


Figure 8 RMS, $Peak$ and PTP values through time per workpiece

Then, Figure 9 presents the average values for the RMS, $Peak$ and PTP value per workpiece. We also included the evolution of the actual measurement per workpiece for the specification #15 to see if there is a direct relationship between the evolutions of the two phenomenon that could be visually witnessed.

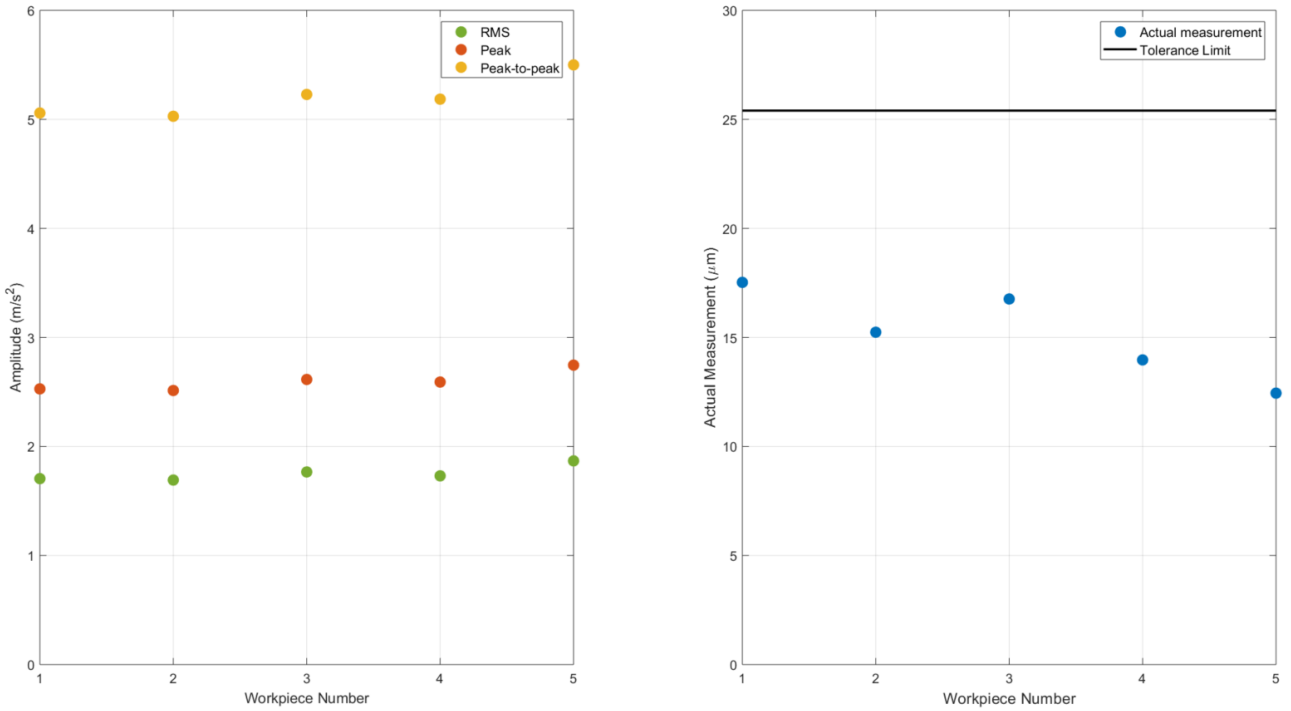


Figure 9 RMS, $Peak$ and PTP values per workpiece and actual measurement per workpiece

We did the same analysis with the kurtosis and the crest factor. Results are shown in Figure 10 and Figure 11.

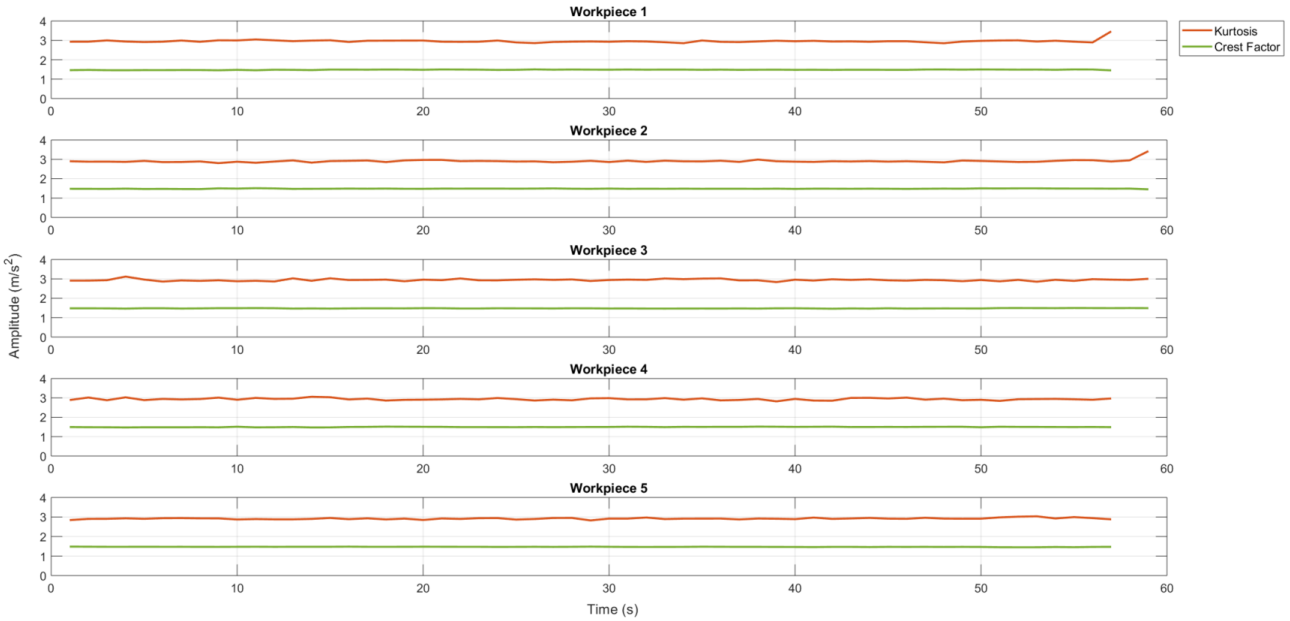


Figure 10 K and CF values through time per workpiece

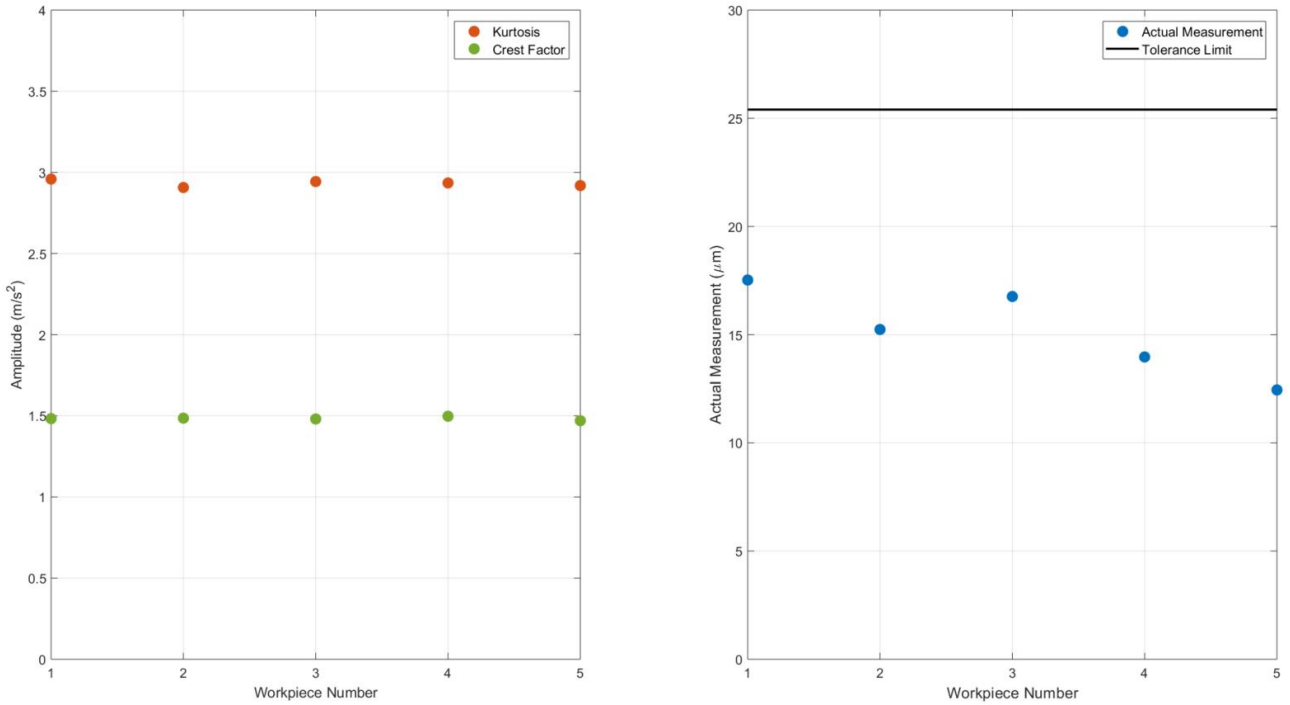


Figure 11 K and CF values per workpiece and actual measurement per workpiece

We then did the same analysis in the frequency domain. Results of the SCE values for each harmonics are shown in Figure 12 and Figure 13.

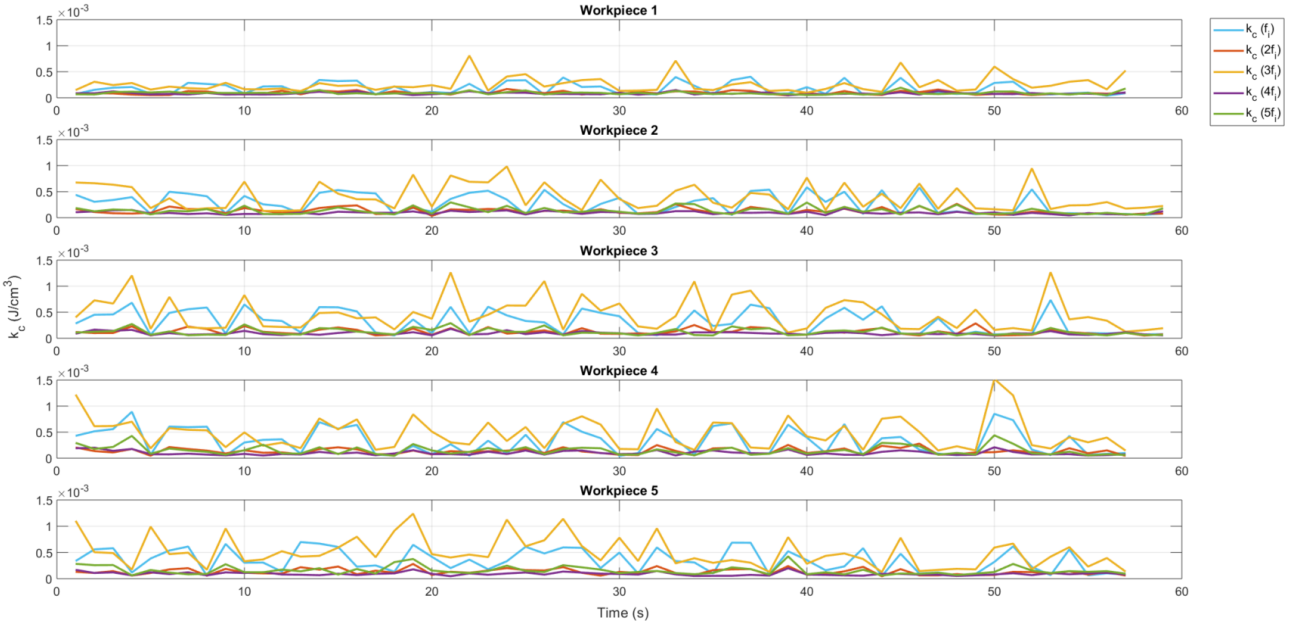


Figure 12 k_c values for each harmonics of f_i through time per workpiece

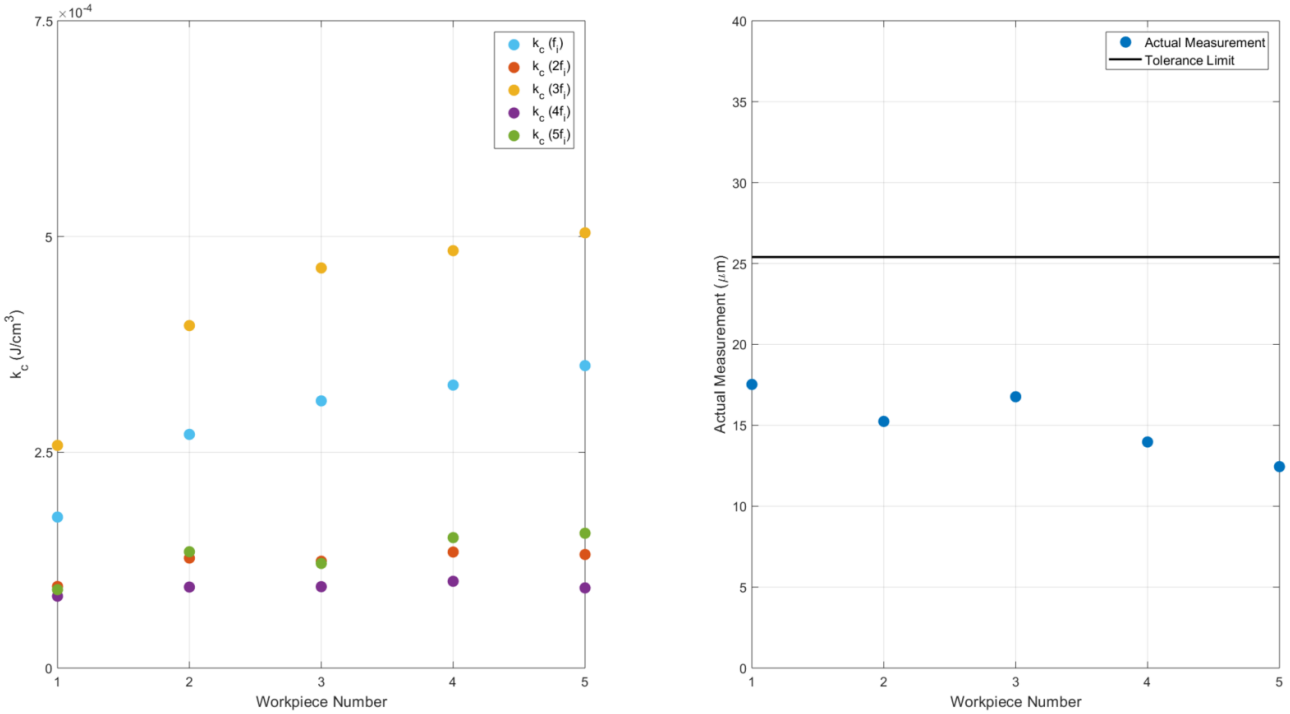


Figure 13 k_c values for each harmonics of f_i per workpiece and actual measurement per workpiece

Finally, we also did a time-frequency analysis where we looked at the frequency domain of the signal through time. Figure 14 presents the spectrogram we obtained. The white lines represent the separation between each workpiece; starting to the left with workpiece 1 up to the right with workpiece 5. In a) we gave the spectrogram for the frequencies between 0 and 400 Hz and in b) for the frequencies 400 to 1000 Hz. Most frequencies of interest are located in the range of 0 to 1 kHz.

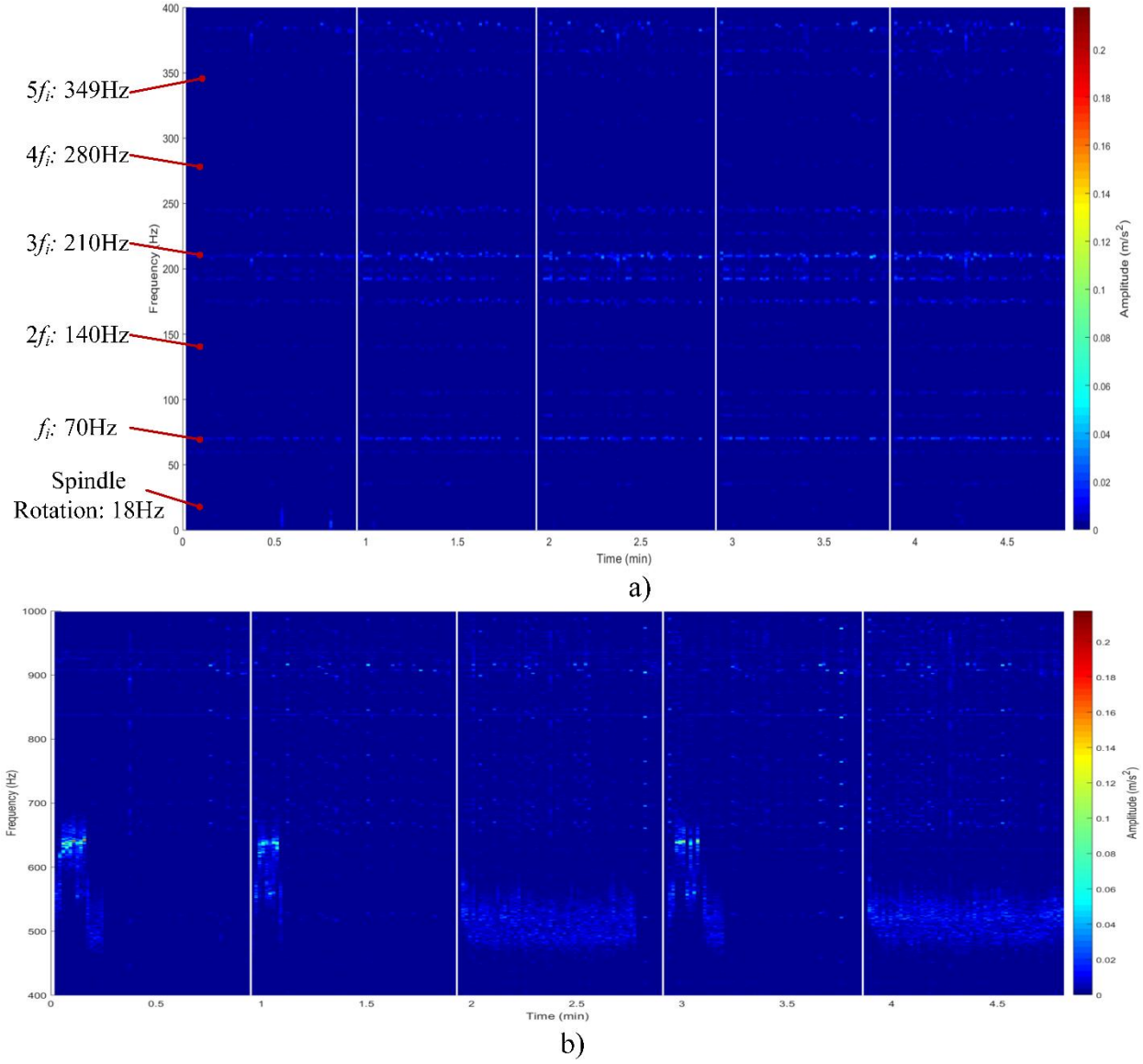


Figure 14 Time-frequency analysis for all the workpiece

From Figure 8, we can see that, for all workpieces, most values are comprised between an amplitude of 0 and 6 m/s^2; with some peaks during the machining process of each workpiece. However, we cannot clearly see that there was either a degradation or an improvement regarding the machining process in a part-to-part point of view. Also, when we look at the average values per part in Figure 9, we cannot clearly state that there is a direct relationship between the actual measurement and the evolution of the RMS, $Peak$ or PTP.

We could conclude the same thing regarding the evolution of the K and CF through time with the results shown in Figure 10 and Figure 11. However, it is interesting to also use these results to do the diagnostic of the machine health state. Based on these results, we could conclude that the machine is in a good health state. From Thomas [20], a kurtosis value around a value of 3 means a random signal, hence a machine in good health where no spike or impact were recorded. Values higher than that would start to indicate that impacts are being recorded. This is also supported by the values of the crest factors which are low and near the value indicating a good condition ($CF = 1.41$). We also tried to look at the fundamental frequencies related to the bearings installed in the spindle. However, since the machine and its component are relatively new, the amplitudes related to the typical fault (FTF, BPFI, etc.) do not stand out. This could indicate that they are in good condition and that their signal is lost in the noise of the machine during the machining process. A diagnostic when the machine is not cutting could probably help us identify with better accuracy these frequencies.

From Figure 12, we can see that the variation in terms of amplitude seems to increase between the workpiece 1 and workpiece 5. This would seems to be consistent with the claim of Proteau et al. [9] that the energy required to keep a material removal rate is increasing with tool wear. When we look at Figure 13, we can see that the values are increasing with every workpiece; for the first and third harmonics (f_i and $3f_i$). The 2nd, 4th and 5th harmonics ($2f_i$, $4f_i$ and $5f_i$) seems to have a low amplitude throughout the data we collected.

Moreover, when we look at the scale of the amplitude of the data, we can see that they are pretty low. This is somewhat counter intuitive to our belief. We believed that because the Inconel 625 is a very hard and difficult material to work with, we would have seen very high amplitudes due to the force required to remove the material. It was not possible to get the exact depth of cut used in this operation, therefore, maybe the engineers responsible of this product at APN used a very low depth of cut parameter in order to create less friction between the material and the cutting tool in order to facilitate the machining process.

When we look at the spectrogram shown in Figure 14, we can quickly see that the overall frequencies' amplitudes are consistent with our previous claim; it is low across most frequencies. We can still detect some frequencies of interest such as the spindle rotation (1047 RPM = 17.45 Hz), the cutting tool frequency (with 4 flutes: 69.8 Hz) and its harmonics (139.6 Hz, 209.4 Hz, 279.2 Hz and 349 Hz).

Aside from these specific frequencies, this low amplitude claim seems to hold true except for some spontaneous peak between 600 and 700 Hz. In fact, if we look at the graph in b), we see a phenomenon where we have not yet found the source. No video recording was made during the acquisition process. This kind of data would surely help us to correlate such phenomenon with actual events during the machining process. Additionally, this phenomenon is not consistent across all workpiece. Cutting parameter and overall machining strategy were not changed between the workpieces, hence we would have expected a similar pattern for each workpiece. However, we can denote two patterns: one related with workpiece 1, 2 and 4 and the other with the workpiece 3 and 5. The peaks in amplitude are also related to the first pattern for workpiece 1, 2 and 4.

In a diagnostic point of view, our conclusion related to the kurtosis and crest factor values seems to hold in the frequency domain since we do not seem to detect traces of impact during the utilization of the equipment.

The objective of this paper was to demonstrate that we can now have access to data allowing us to describe and diagnose a machining process and its cutting operations as well as making a step toward being able to do the prognostic of the overall quality of a workpiece. With the results shown in this section we can conclude that the propose data acquisition architecture enable us now to adequately contextualize, in real-time and automatically, signals acquired through sensing devices. However, when we look at the final objective of our project; that is the prognostic of the quality of a workpiece in term of the respect of its GD&T requirements, we have not seen a clear linear relationship or pattern between the cutting operation vibration signal and the evolution of the actual measurement of the specification #15 neither in the time and frequency domain. Throughout this article, we have been looking at one operation influencing the specification #15; in fact, there is a total of 11 cutting operations influencing this specific requirement. In other words, we conjuncture that a clear linear relationship cannot be establish between only one operation and the evolution of a specific requirement. On the contrary, it is maybe the "sum" or sequence of all these operations that could influence the conformity of a workpiece specific requirements. In other words, all the variations across all these operations could explain the evolution of a specification. Consequently, we believe that only through a machine learning approach we could be able to predict the quality of a workpiece. Our strategy to apply such approach to this research project still hold to this point.

6 Conclusion

To conclude this article, we wanted to showcase our data acquisition system architecture and demonstrate that we can now adequately contextualize a vibration signal to better do the diagnostic of a cutting operation to, in the end, facilitate the development of a prognostic methodology for the quality of a workpiece. We

believe that we have successfully achieve these objectives by showing multiple results related to the cutting operation *OP_510*. However, we have not yet been able to showcase a linear relationship between the vibration signal of this operation and the evolution of the quality of the workpiece. The use of a machine learning approach could probably help us achieve this objective. Further work in order to close the gap between our current status and our final objective to predict the quality of a workpiece will include adding sensors to the GROB CNC machine: a tri-axial accelerometer, an acoustic emission sensor, current and voltage sensors to the motor of the spindle as well as the ones of the three main working axis of the CNC machine and try to apply cyclostationarity analysis based on the work of Lamraoui et al. [21]. We will also expand our system capacity to have it works in a more autonomous way and we will finally use and apply multiple machine learning approaches to perform sensors fusion and the actual prediction of a workpiece quality.

Acknowledgments

The authors would like to thank APN Inc.: their owners, Jean and Yves and all their employees related to this project for their support and time in helping our research team. Special thanks to Alain, Jonathan, Michel, David, Daniel, Raphael, Maxim and Mathieu for the many questions they answered and their help to make this project happen. We would also like to acknowledge the financial support of the *Fonds de Recherche du Québec – Nature et Technologies* (FRQNT) to this project through Grant #257668.

References

1. Kohler D, Weisz J-D (2016) Industrie 4.0 *Les défis de la transformation numérique du modèle industriel allemand*. France
2. Wang D (2011) *Robust Data-Driven Modeling Approach for Real-Time Final Product Quality Prediction in Batch Process Operation*. IEEE Transactions on Industrial Informatics 7 (2):371-377
3. Benardos PG, Vosniakos GC (2003) *Predicting surface roughness in machining: a review*. International Journal of Machine Tools and Manufacture 43 (8):833-844
4. Balamurugamohanraj G, Vijaiyendiran K, Mohanaraman P, Sugumaran V (2016) *Prediction of Surface Roughness Based on Machining Condition and Tool Condition in Boring Stainless Steel-304*. Indian Journal of Science and Technology 9 (47):6
5. Vogl GW, Weiss BA, Helu M (2016) *A review of diagnostic and prognostic capabilities and best practices for manufacturing*. Journal of Intelligent Manufacturing:1-17
6. Wang K-S (2013) *Towards zero-defect manufacturing (ZDM)—a data mining approach*. Advances in Manufacturing 1 (1):62-74
7. Peng Y, Dong M, Zuo MJ (2010) *Current status of machine prognostics in condition-based maintenance: a review*. The International Journal of Advanced Manufacturing Technology 50 (1):297-313
8. Lee J, Jin C, Bagheri B (2017) *Cyber physical systems for predictive production systems*. Production Engineering 11 (2):155-165
9. Proteau A, Tahan A, Thomas M (2019) *Specific cutting energy: a physical measurement for representing tool wear*. The International Journal of Advanced Manufacturing Technology:1-10
10. Balan GC, Epureanu A (2008) *The monitoring of the turning tool wear process using an artificial neural network. Part 1: The experimental set-up and experimental results*. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222 (10):1241-1252
11. Balan GC, Epureanu A (2008) *The monitoring of the turning tool wear process using an artificial neural network. Part 2: The data processing and the use of artificial neural network on monitoring of the tool wear*. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222 (10):1253-1262
12. Aghazadeh F, Tahan A, Thomas M (2018) *Tool Condition Monitoring Using Spectral Subtraction Algorithm and Artificial Intelligence Methods in Milling Process*. International Journal of Mechanical Engineering and Robotics Research 7 (1):30-34
13. Waqar T, Demetgul M (2016) *Thermal analysis MLP neural network based fault diagnosis on worm gears*. Measurement 86:56-66
14. Aydin O, Guldamlasioglu S *Using LSTM networks to predict engine condition on large scale data processing framework*. In: 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE), 8-10 April 2017 2017. pp 281-285

15. ASME (2009) *Dimensioning and Tolerancing; Engineering Drawing and related documentation practices*. American Society of Mechanical Engineers, New York, NY
16. Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) *Machinery health prognostics: A systematic review from data acquisition to RUL prediction*. Mechanical Systems and Signal Processing 104:799-834
17. Elattar HM, Elminir HK, Riad A (2016) *Prognostics: a literature review*. Complex & Intelligent Systems 2 (2):125-154
18. Abellan-Nebot JV, Romero Subirón F (2010) *A review of machining monitoring systems based on artificial intelligence process models*. The International Journal of Advanced Manufacturing Technology 47 (1):237-257
19. Debongnie J-F (2006) *Usinage*. Editions du CEFAL,
20. Thomas M (2011) *Fiabilité, maintenance prédictive et vibration des machines*. Presses de l'Université du Québec, Montréal
21. Lamraoui M, Thomas M, El Badaoui M, Girardin F *Cyclostationarity analysis of instantaneous angular speeds for monitoring chatter in high speed milling*. In: IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, Montreal, Qc, Canada, 25-28 Oct. 2012 2012. pp 3868-3873

Fault prognosis of planetary gearbox using acoustic emission and genetic algorithm: a case study

Félix Leaman[1], Cristián Molina Vicuña[2], Ralph Baltes[1] and Elisabeth Clausen[1]

[1]Institute for Advanced Mining Technologies, RWTH Aachen University
Wüllnerstr. 2, D-52062 Aachen, Germany
{fleaman, rbaltes, eclausen}@amt.rwth-aachen.de
[2]Laboratorio de Vibraciones Mecánicas, Departamento de Ingeniería Mecánica,
Universidad de Concepción, Edmundo Larenas s/n, Concepción, Chile
crimolin@udec.cl

Abstract

One of the most important aspects of machine fault prognosis is the selection of sensors and features to represent the degradation process of a faulty component. Several approaches in the literature have used features extracted from vibration signals to estimate the future condition based on time series forecasting. Another technology that has been used increasingly for this task are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. On gearboxes some studies have shown that the AE technology can be used effectively for fault diagnosis, but its use for fault prognosis is still a relative new field of research that offers encouraging opportunities. One downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature, which may lead to difficulties during the forecasting of an AE-based feature. Thus, in this study a novel feature based on a relative counting of the AE bursts is proposed and tested with data from a planetary gearbox with a ring gear fault. The proposed feature reduces the influence of the temperature on the generation of AE when it is compared to the counting based on a fixed amplitude threshold. Therefore, it can then be more suitable for fault prognosis than traditional AE counting. In this case study the forecasting of the proposed feature is carried out using an artificial neural network (ANN), whose hyperparameters were selected using a genetic algorithm. The results are promising and constitute a basis for further research.

1 Introduction

Machine diagnosis is defined as the examination of symptoms to detect and determine the nature of faults or failures, while machine prognosis is defined as the evaluation of the actual condition of a machine, forecast of the future condition and estimate the remaining useful lifetime (RUL) before a catastrophic failure based on the symptoms of faults [1, 2]. As it is well established for machine diagnosis, the type of fault that is going to be examined must be correctly represented by the monitoring data. A feature is a specific representation of the data in a lower dimension, which can be easier to relate to a specific fault. For instance, a misalignment can be examined using the vibration data and a feature related to the amplitude of the rotational speed harmonics can represent it in a better way than the RMS value. For prognosis purposes the same principle applies, but rather than only represent a fault the features should also be capable to represent the degradation process of the faulty component [3].

Prognosis models can be categorized into experience-based models, physical-based models or data-driven models. Experienced-based models use component failure history to predict the future condition. Physical-based models are based on physical principles of the failure process, which are modelled by mathematical equations. Data-driven models use historical monitoring data to generate a model capable of predicting the future condition of the machine. Within data-driven models artificial neural networks (ANN), hidden Markov and support vector machine models have been used for feature forecasting [2, 4, 5].

Condition monitoring data from vibration sensors have been extensively used for diagnosis purposes and to a lesser extent for prognosis purposes. Another type of sensor that has been used increasingly for these tasks

are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. AE is defined as elastic stress waves generated by rapid release of strain energy due to changes in the internal structure of a material [6]. Many sources of AE are damage-related such as: crack initiation and growth, crack opening and closure, dislocation movement and others [7]. One important feature to represent the AE data consists of the rate of generated AE bursts. This is adressed by many studies that have for example correlated the rate of AE bursts with crack propagation [8, 9].

In the case of gearboxes the AE technology has been effectively used for fault diagnosis [10, 11]. For the case of prognosis it is a relatively new field of research that offers encouraging opportunities [12]. It has been reported that during cyclic loading of a single tooth, the amplitude and energy of the AE bursts increase intermittently because of the accumulated damage and crack propagation [13]. Thus, it makes sense to evaluate the rate of AE bursts as a possible feature suitable for fault prognosis in gearboxes. However, one downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature [14], which may lead to difficulties during the forecasting of an AE-based feature. Therefore, in this study a novel feature based on a relative counting of the AE bursts that minimizes the effect of temperature is proposed and tested with data from a planetary gearbox with a ring gear fault. Then, a data-driven model is generated with an ANN and the progression of the feature is predicted.

2 Experiment

The test bench consisted of a planetary gearbox connected to an electric motor on its input side (high speed) and to a generator on its output side (low speed). The gearbox had the following number of teeth: $Z_s = 18$ for the sun gear, $Z_r = 72$ for the ring gear and $Z_p = 26$ for each of the planet gears. The number of planet gears was $N_p = 3$. The outside diameter of the ring gear was 144 mm. The speed reduction ratio was 1:5. A localized fault was produced in one tooth of the ring gear using a grinding tool. An AE sensor VS375-M was fixed under the gearbox case using a magnetic holder. The surface under the sensor was previously polished and copper paste was applied between the sensor and the surface to improve the acoustic coupling. The test bench and the fault can be seen in Figure 1.

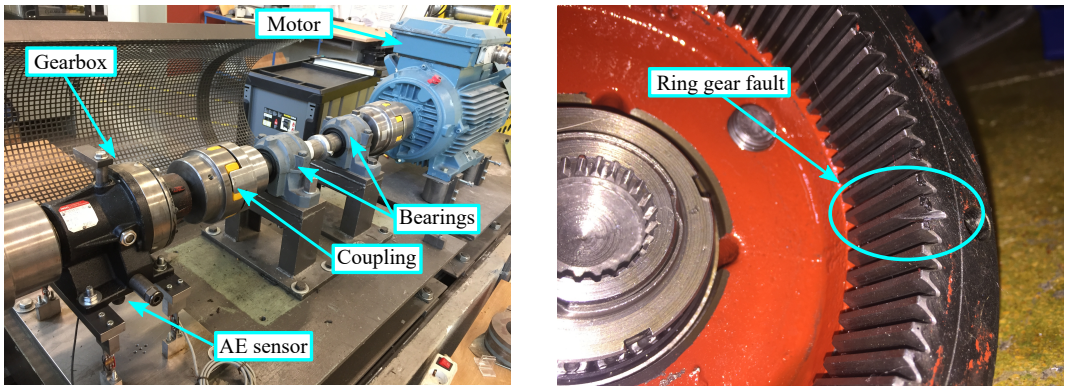


Figure 1: Planetary gearbox test bench (a) and ring gear fault (b)

AE signals were measured using an input rotational speed of 1300 r/min and a load of 112 Nm in the output shaft. Thus, the sun gear rotating frequency was $f_s = 21.67$ Hz and the carrier rotating frequency was $f_c = 4.33$ Hz. For all measurements a sampling frequency of 1 MHz and acquisition time of 5 s for the AE signals were used. The AE signals were measured continuously in eight measurement campaigns (MCs), of which seven consisted of approximately 75 min of continuous operation and one of approximately 120 min. During this eight MCs neither planetary gearbox nor sensor were disassembled. Aditionally in each MC the same instrumentation for the measurement of the AE signals was employed. Summing the MCs, the planetary gearbox cumulated approximately 10.5 h of intermitent operation. After the 8^{th} MC the planetary gearbox was disassembled for inspection, which revealed no visual growth of the fault. Despite this observation, damage progression at a microscopic or subsurface level could not be discarded.

The external temperature of the planetary gearbox case was used to represent the temperature of the oil. It was measured every two minutes for each MC. The result is shown in Figure 2, where the end of each MC is indicated with a dashed line. As expected the temperature increases during each MC. The rate of temperature increase is higher at the beginning of each MC and is then progressively reduced.

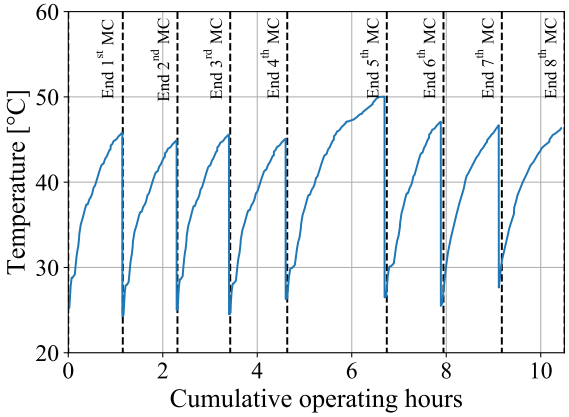


Figure 2: Planetary gearbox temperature during the tests.

As discussed in the previous section, the temperature affects the generation of AE in gearboxes. Therefore, the detection of AE bursts is expected to be affected by this phenomena. In the next section this topic is adressed.

3 Counting of AE Bursts

3.1 Fixed counting of AE bursts

The traditional approach to detect AE bursts consists of the definition of a fixed amplitude threshold T_a. When the amplitude of the signal exceeds this value an AE burst is detected (see Figure 3). The end of the AE burst is determined when after a hit definition time (HDT) there are no more threshold crossings. The beginning of the next AE burst can only be detected when a hit lockout time (HLT) has elapsed. The value for the threshold T_a is established empirically after inspection of the AE signals.

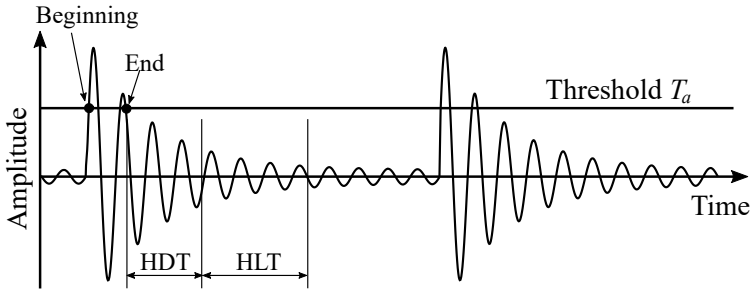


Figure 3: AE burst detection with fixed amplitude threshold

For this case study a threshold $T_a = 0.8$ mV was selected together with a HDT and a HLT of 300 μs each. The number of detected AE bursts is counted every 5 seconds in each signal, so the obtained result expresses the rate of detected AE bursts every 5 seconds. In order to minimize small fluctuations and concentrate the analysis on the global trend, a moving average filter of width 60 s was applied to the rate of AE bursts. Figure 4 shows the result of this analysis with fixed counting approach. It is observed that the rate of AE bursts rises at the beginning of each MC and then decreases as the MC continues. Although the same behavior occurs in each MC, the maximal rates of AE bursts have different values for different MCs. Moreover, the maximal values

reached do not take place necessarily as the accummulated operating hours increase. However, when only the last parts of each MC are analyzed, it is observed that the rate of AE bursts increases as the hours of operation cumulate.

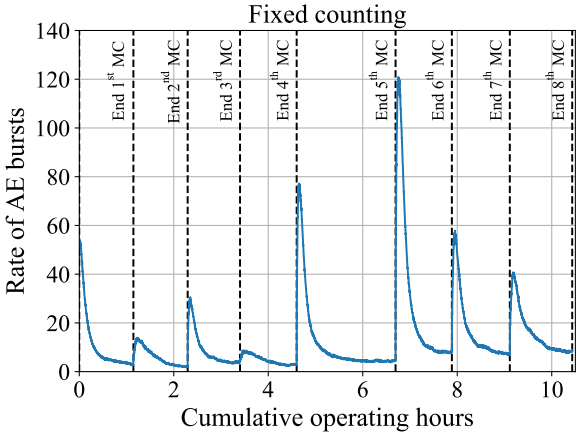


Figure 4: Rate of AE bursts during tests calculated with a fixed counting

The aforementioned observations make the fixed counting approach inappropiate for prognosis purposes in this study case, since the obtained rate of AE bursts has very low monotonicity. Therefore, it is required to explore other approaches as follows in the next subsection.

3.2 Relative counting of AE bursts

In this subsection an approach to make a relative counting of the AE bursts is proposed. The relative counting consists of taking into account only the AE bursts that should take place when the fault of the ring gear interacts with each of the planet gears. Knowing the theoretical temporal positions of the bursts originated by the interaction between fault and planet gears within a signal requires the use of an encoder that provides angular position. However, this requirement can be avoided by considering only the spacing between the detected bursts instead of their absolute position. Thus, the relative counting will be maximum if in the signal there are detected bursts equally spaced at the inverse of the ring gear fault frequency f_{fr}. This corresponds to the frequency at which the planet gears interact with the fault and corresponds to the rotational frequency of the planet carrier f_c times the number of planets N_p as expressed in equation 1:

$$f_{fr} = f_c \cdot N_p \tag{1}$$

Mathematically the relative counting can be achieved by taking the maximal cross-correlation between a binary signal $\mathring{x}(t)$ that represents where the bursts are actually detected, and another binary signal $\mathring{z}(t)$ that represents where the AE bursts should take place, if they are always originated by the interaction between the damaged ring gear tooth and the planet gears.

Figure 5 shows schematically the path of each planet gear across the fault and the originated AE bursts. Some of them overpass the threshold T_a and thus are detected. Accordingly, the binary signal $\mathring{x}(t)$ is constructed with its high value at the position of the detected bursts and its low value everywhere else. On the other hand, the binary signal $\mathring{z}(t)$ is constructed with its high values equally spaced at the inverse of the ring gear fault frequency f_{fr}. Thanks to the cross-correlation it is not necessary that the high values of the binary signal $\mathring{z}(t)$ are in phase with the path of the planet gears across the fault. By taking the maximal cross-correlation the phase difference can be corrected. Therefore, only the spacing between the high values of the binary signal $\mathring{z}(t)$ are considered in its construction.

The proposed relative counting gives values between zero and one. A value one means that all expected bursts have a maximal amplitude higher than the threshold. Analogously, a value zero means that none of the expected bursts have an amplitude higher than the threshold. This is schematically illustrated in Figure 6. Here, the binary signal $\mathring{x}_1(t)$ has a maximal cross-correlation of 1 with the binary signal $\mathring{z}(t)$, since they can match

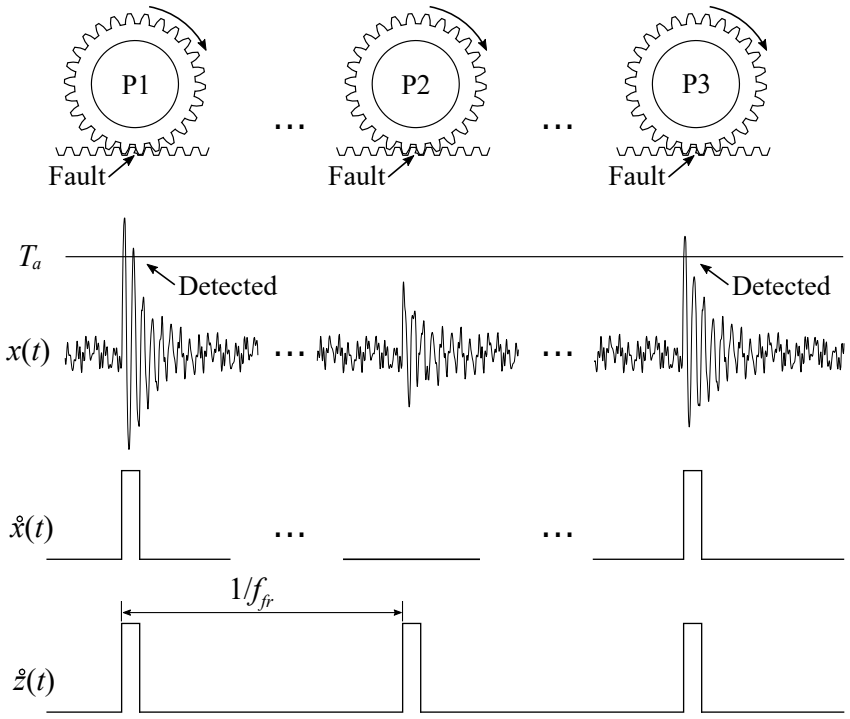


Figure 5: Binary signals representing actual $\mathring{x}(t)$ and hypothetic $\mathring{z}(t)$ time location of AE bursts

in all their high values. The binary signal $\mathring{x}_2(t)$ has a maximal cross-correlation of 0.5 with $\mathring{z}(t)$, since they can match in two out of four high values. The binary signal $\mathring{x}_3(t)$ has a maximal cross-correlation of 0.25 with $\mathring{z}(t)$, since they only can match in one out of four high values. Finally, the binary signal $\mathring{x}_4(t)$ has a maximal cross-correlation of 0 with $\mathring{z}(t)$ since it does not have any high values. Notice that this figure points out the aforementioned observation that $\mathring{z}(t)$ does not have to be in phase with the signals $\mathring{x}(t)$.

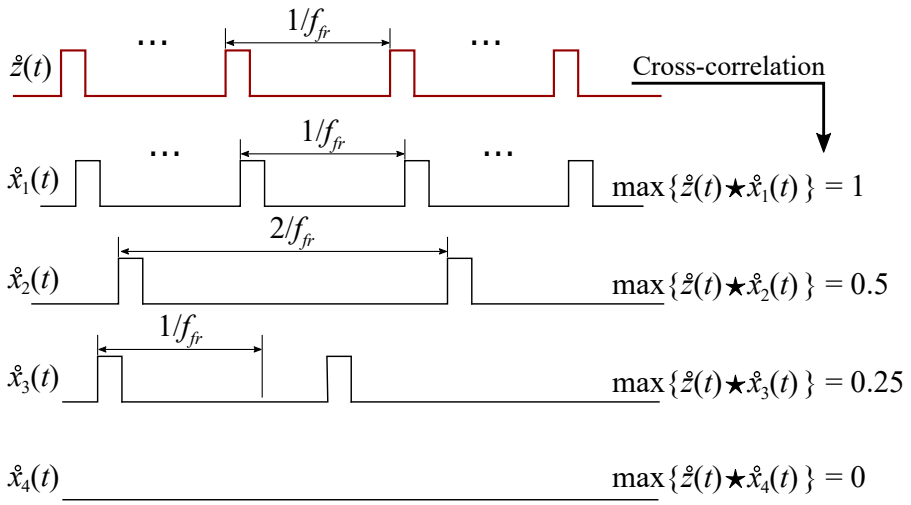


Figure 6: Relative counting of AE bursts based on cross-correlation with the signal $\mathring{z}(t)$

For the analysis of the case study the same values for T_a, HDT and HLT used for the fixed counting are now used for the relative counting. In the same way, a moving average filter of width 60 s is also employed to eliminate too small fluctuations in the rate of AE bursts. Figure 7 shows the result of this analysis with the relative counting approach. It is observed that the rate of AE bursts (now represented by the maximal cross-correlation) is much more monotonous than for the fixed counting approach. The large fluctuations due to the beginning of each MC are also considerably reduced.

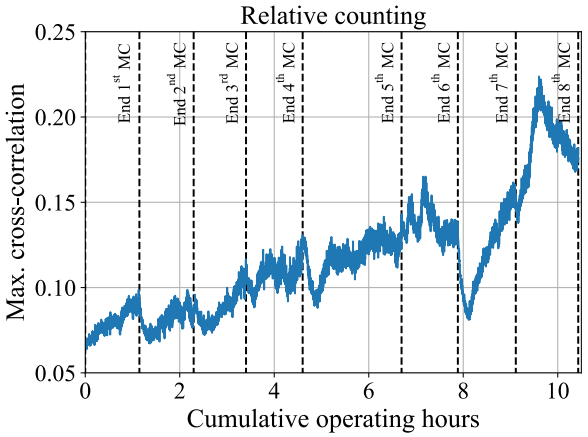


Figure 7: Maximal cross-correlation during tests calculated with a relative counting of AE bursts

In account of this observations the rate of AE bursts calculated with the relative counting approach is much more suitable to be uses as feature for failure prognosis. The forecasting of this feature is addressed in the following section.

4 Prognosis

4.1 Time series forecasting with an ANN

There are multiple possible architectures to build an ANN suited for forecasting a time series. One of the most straightforward architectures consists of using the observations x_0 to x_n from a time series $\{x\}$ as input features to predict the observation x_{n+1}. In the following prediction step (PS) the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} and so on. Another possible architecture consists of using the observations x_0 to x_n to predict the observations x_{n+1} to x_{n+m}. In the following PS the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} to x_{n+m+1} and so on. Thus, although m observations are predicted in each PS, only the first of them is used as input feature to make the prediction in the next PS. One variation of such an architecture consists of using only the last predicted observation rather than the first one to make new predictions. By using these types of architectures a higher stability in the forecasting is expected. The training and test of such ANN are illustrated schematically in Figure 8. It shows how the training examples (TE) are constructed from a time series considering values $n = 4$ and $m = 3$. Each of the TE is used to iteratively adjust the weights of the ANN using the backpropagation algorithm. After the ANN has already been trained with all the TE, it is now capable to forecast the time series in successive PS and its performance can be evaluated with the test data.

During the building of the ANN the dataset is divided into two groups: training and test. Within the training group a datasubset is left aside to validate the results of the training process. One key objective of the training process is to find the most suited hyperparameters for the ANN. They include: number of nodes in input and output layers (n and m, respectively), number of nodes per hidden layer, regularization term α, number of mini-batches and number of epochs. More information regarding the hyperparameters can be found in [15]. Initially the most suited values for the hyperparameters are selected using grid search. This consists of selecting possible values for each hyperparameter and calculating which combination of hyperparameters achieves the best performance overall. For each combination of hyperparameters an ANN is trained without including the validation data. Then, the forecasting performance of the ANN is calculated in the data used for its training and then in the validation data. This division is carried out with the aim of minimizing the bias and variance errors of the ANN model as much as possible. A performance calculated only in the data used for training can cause higher variance since the model could over fit the training data. On the other hand, a performance calculated only in the validation data can cause higher bias since the model could under fit the training data. In this study a global performance is calculated as the average between the training and validation performances.

The performance of the model is measured with the use of the mean squared error (MSE) between prediction

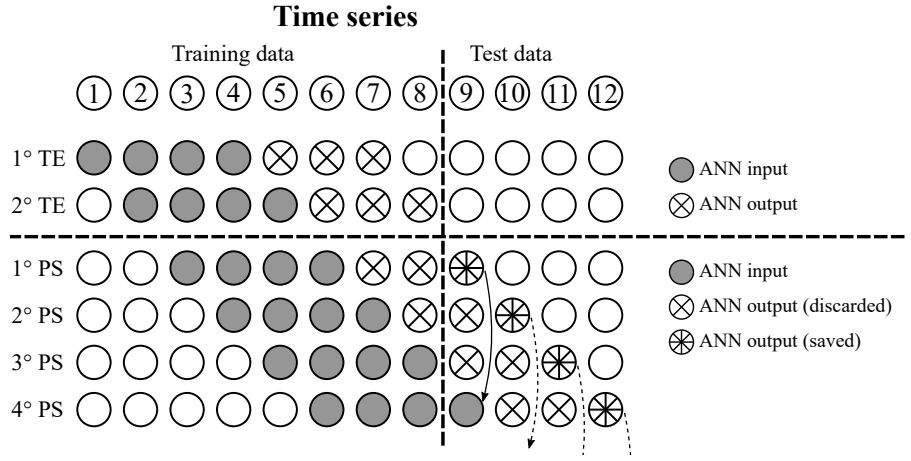


Figure 8: Training and prediction scheme of the ANN for time series forecasting

and real data. The MSE is calculated as follows in equation 2:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{2}$$

, where x_i represents the i-th observation of the time series $\{x\}$ and $\hat{x}_i$ represents the i-th observation of the predicted time series $\{\hat{x}\}$.

Due to computational restrictions the grid search can only consider a limited number of values for the hyperparameters. One way to increase the efficiency of the search is to set an initial group of values for the hyperparameters and determine the best combinations. After that, the search continues but considering only variations of the already determined best combinations. This can be achieved by using a genetic algorithm, which is explained in the following section.

4.2 Genetic algorithm

The genetic algorithms are computational procedures to find a sufficiently good solution to an optimization problem. They are inspired by the mechanism of natural selection and emulate biological processes such as mutation, crossover and selection. In this study a genetic algorithm is employed to guide consecutive grid searches and find the best possible combination of hyperparameters for the ANN model. First, an initial population of hyperparameters is evaluated using a fitness function. The best combination of hyperparameters are called *parents* and are used to generate *children* through crossover and mutation techniques. The crossover combines information from parents to generate the children, while the mutation consists of random alteration of the information traspased to the children. In this study, the information corresponds to the hyperparameters for the ANN model. The parents together with the children constitute a new *generation*. This new generation is once again evaluated with the fitness function to create new parents and new children in an iterative process. The procedure is repeated until a stoppage criterion is met. When this is accomplished, the best solution in the last generation is selected as the best solution of the whole iterative process. Two possible stoppage criteria are: a fixed number of generations to iterate or a fixed number of consecutive generations for which the best fitness score does not improve more than a pre-defined percentage.

Figure 9 shows schematically how the genetic algorithm works. Let us consider four hyperparameters to optimize: α, β, γ and δ. The initial population contains five possible combinations of hyperparameters, which are evaluated with the fitness function. The best two combinations are selected as parents and are used to generate three children. The employed crossover corresponds to a single-point crossover, which randomly selects at which single point (star marker) the information from the parents is divided. Then, in the mutation some information of the children is randomly altered. One important aspect is that the percentage of mutation must not be so high as to completely vanish the result of the crossover.

In this study single-point crossover, maximum 10% of mutation and a total of 5 generations are considered.

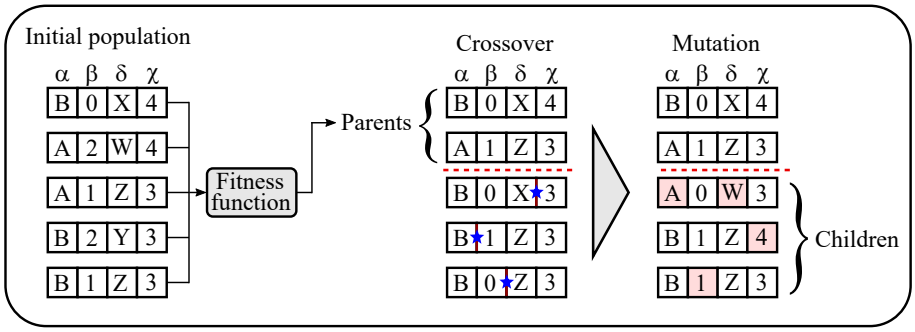


Figure 9: Schema of the genetic algorithm for hyperparameters optimization

The employed fitness function corresponds to the calculation of a fitness score (FS) that is the inverse of the MSE as it is shown in equation 3. Thus, higher values of the FS correspond to best performance of the model.

$$FS = \frac{1}{MSE} \tag{3}$$

4.3 Results

Table 1 shows the hyperparameters considered for the initial population. The values for the number of nodes in the input layer n and in the output layer m expressed as percentages refer to a fraction of the data used to train the ANN models. Two configurations for the hidden layers are initially tested: one hidden layer with a number of nodes equal to the mean value between n and m, and two hidden layers with number of nodes equal to $0.75(n+m)$ for the first layer and $0.25(n+m)$ for the second layer.

Hyperparameter	Values
n	50%; 32%; 16%
m	13%; 6%; One
Nodes per layer	$0.5(n+m)$; $0.75(n+m)$ - $0.25(n+m)$
α	1.e-1; 1.e-3
Mini-batches	32; 128
Epochs	Until 100

Table 1: Initial hyperparameters considered to generate ANN models

The total of 7512 observations of the relative counting of AE bursts was divided into 80% for training and 20% for test. For the process of hyperparameters estimation 25% of the training data was left aside for validation. Table 2 shows the fitness scores for the initial population and 5th generation during the search for the most suited hyperparameters. The $FS_{training}$ is highly improved from the initial population until the 5th generation, while the $FS_{validation}$ only decreases marginally. As a consequence, the overall $FS_{average}$ is highly increased. Table 3 shows the corresponding best values for the hyperparameters for the initial population and 5th generation that achieve the aforementioned fitness scores. It is observed that the number of nodes in each layer tends to increase. So does the regularization term α, while the number of epochs decreases. The number of mini-batches remains approximately constant.

Best score	For initial population	For 5th generation
$FS_{training}$	6.12e4	1.83e5
$FS_{validation}$	2.25e3	2.00e3
$FS_{average}$	3.17e4	9.25e4

Table 2: Fitness scores of the best models from the initial population and 5th generation in validation phase

Figure 10 shows the result of the forecasting for the maximal cross-correlation with the best hyperparameters of the initial population in (a) and for the 5th generation in (b). It is observed that for the initial population

Hyperparameter	Best values for initial population	Best values for 5th generation
n	50%	63.481%
m	13%	15.053%
Nodes per layer	1836 - 1002	1858 - 1014
α	1.0e-1	1.3085e-1
Mini-batches	32	33
Epochs	64	21

Table 3: Best combination of hyperparameters for the initial population and 5th generation

the prediction (test) deviates considerably from the real data. However, for the 5th generation the prediction is very close to the real data, which confirms the improvement produced by the genetic algorithm. This is also confirmed by the calculation of the fitness scores in Table 4. The FS calculated on the training data indicates how good the model fits the training data. Thus, it can be only considered as an estimation of the performance of the ANN model. The FS calculated on the test data indicates how good the model can predict unseen data. Both scores should be as high as possible and have a as low as possible difference between them. It is also expected that $FS_{training}$ is higher than FS_{test}, since the model should perform better on data that it has already seen. For the initial population both fitness scores $FS_{training}$ and FS_{test} have low values. Moreover, FS_{test} is higher than $FS_{training}$. For the 5th generation this deviation is corrected and both fitness scores reach much higher values.

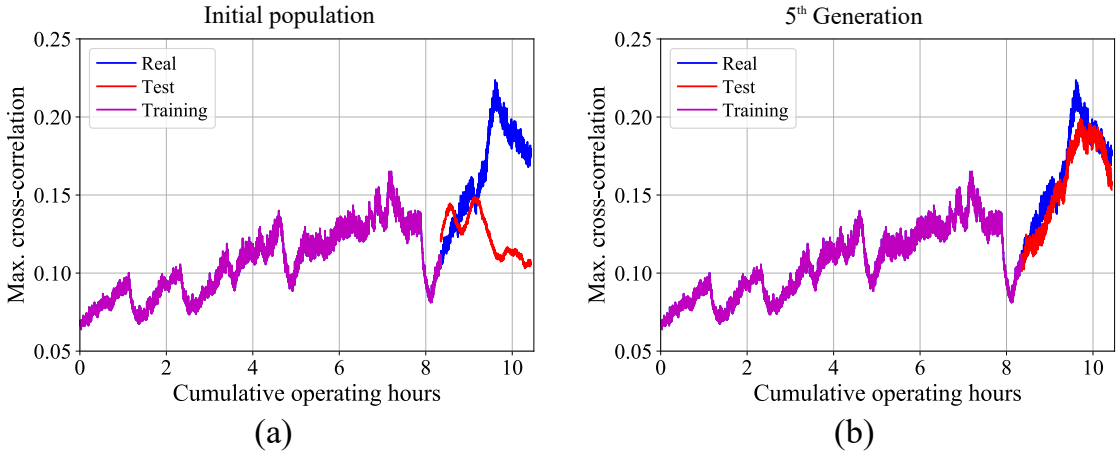


Figure 10: Forecasting of the maximal cross-correlation for the initial population (a) and 5th generation (b)

Best score	For initial population	For 5th generation
$FS_{training}$	9.39e1	3.64e4
FS_{test}	3.06e2	6.37e3

Table 4: Model fitness scores from the initial population and 5th generation in the testing phase

5 Conclusions

A novel approach for counting of AE bursts in gearboxes was presented and compared with the traditional approach with measurements from a planetary gearbox. In conditions where the gearbox accummulated several hours of intermittent operation with a localized fault in the ring gear, the proposed approach was able to provide a rate of AE bursts represented by a cross-correlation value suitable for prognosis purposes. The forecasting was carried out with an ANN, whose performance was enhanced through a genetic algorithm for the selection of its hyperparameters. With basis on the obtained results the authors conclude that:

- In presence of a localized fault in the ring gear of a planetary gearbox, the rate of detected AE bursts increases as the hours of operation cumulate.

- For intermittent operation the approach to detect AE bursts using a fixed amplitude threshold provides a rate of AE bursts highly affected by the temperature, which is not suitable for prognosis purposes.
- In comparison the proposed approach for a relative counting of the AE bursts provides a rate of AE bursts with higher monotonicity, and thus it is more suitable for prognosis.
- The rate of AE bursts, in this case represented by a cross-correlation value, can be effectively predicted with the use of an ANN for time series forecasting if its hyperparameters are correctly selected.
- The proposed way for selection of hyperparameters was an initial grid search followed by a genetic algorithm for hyperparameters optimization, which was capable to provide good results.

References

[1] W. Caesarendra, G. Niu, B. Yang *Expert Systems with Applications Machine condition prognosis based on sequential Monte Carlo method.*, Expert Systems with Applications, 37, 2010, pp. 2412-2420

[2] Y. Peng, M. Dong *A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction*, Mechanical Systems and Signal Processing, 25, 2011, pp. 237-252

[3] X. Zhang, J. Kang, J.S. Zhao *Features for Fault Diagnosis and Prognosis of Gearbox*, Chemical Engineering Transactions, 33, 2013, pp. 1027-1032

[4] W.Q. Wang, M.F. Golnaraghi, F. Ismail, *Prognosis of machine health condition using neuro-fuzzy systems*, Mechanical Systems and Signal Processing, 18, 2004, pp. 813-831.

[5] J. Kang, X. Zhang, J. Zhao, et al., *Gearbox fault prognosis based on CHMM and SVM*, In: International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2012.

[6] R.K. Miller, P. McIntire *Non-destructive testing handbook: Volume 5 acoustic emission testing*, Publisher: American Society for Non-destructive Testing, 1987.

[7] M. Huang, L. Jiang, P. Liaw, et al. *Using acoustic emission in fatigue and fracture materials research.*, JOM-e: Research Summary, 50(11), 1998, pp. 1-12.

[8] C. Barile, C. Casavola, G. Pappalettera *Fatigue Damage Monitoring by Means of Acoustic Emission and Thermography in Ti Grade 5 Specimens*, Procedia Engineering, 114, 2015, pp. 487-492.

[9] P. Mazal, F. Vlasic, V. Koula *Use of Acoustic Emission Method for Identification of Fatigue Micro-cracks Creation*, Procedia Engineering, 133, 2015, pp. 379-388.

[10] A.B. Novoa, C.M. Vicuna *New aspects concerning the generation of acoustic emissions in spur gears, the influence of operating conditions and gear defects in planetary gearboxes*, Insight, 58(1), 2016, pp. 18-27.

[11] D.S. Gu, J.G. Kim, Y.S. An, et al. *Detection of faults in gearboxes using acoustic emission signal*, Journal of Mechanical Science and Technology, 25(5), 2011, pp. 1279-1286.

[12] D. Mba *Prognostic Opportunities Offered by Acoustic Emission for Monitoring Bearings and Gearboxes.*, In: 12th International Congress on Sound and Vibration, 2005.

[13] A. Singh, D.R. Houser, S. Vijayakar *Detecting Gear Tooth Breakage Using Acoustic Emission: a Feasibility and Sensor Placement Study*, Journal of Mechanical Design, 121, 1999, pp. 587-593.

[14] C.M. Vicuna *Effects of operating conditions on the Acoustic Emissions (AE) from planetary gearboxes*, Applied Acoustics 77, 2014, pp. 150-158.

[15] I. Loshchilov, F. Hutter *Online Batch Selection for Faster Training of Neural Networks*, In: International Conference on Learning Representations (ICLR), 2016, Workshop Track.

Rotating Machine Diagnosis using Acoustic Imaging and Artificial Intelligence
SURVISHNO Conference 2019

Abdelhakim Darraz[1,2], Jérôme Antoni[3], Pierre Mollon[2], Gabriel Kirié[2], Jean-Hugh Thomas[1], Anthony Larcher[4]

[1] Université du Mans/ENSIM, LAUM, UMR CNRS 6613, F-72085 Le Mans, France
[2] VALEO Equipements Electriques Moteur, F-38070 Saint-Quentin-Fallavier, France
[3] Univ Lyon, INSA Lyon, Laboratoire Vibration Acoustique, F-69621 Villeurbanne, France
[4] Université du Mans/ENSIM, LIUM, EA 4023, F-72085 Le Mans, France

Corresponding author: abdelhakim.darraz@valeo.com

Abstract

Mass production of quality equipment in the automotive industry requires controls throughout the production line. These controls are done through monitoring and validation tools for both production and finished products. The use of signal processing methods, applied to acoustic and vibratory recordings collected during the operating cycle, aims to ensure that they are in good working order, to maintain them and to guarantee the quality of the service provided by a manufacturer to its customers. However, sometimes the techniques used do not reach the expected performance, which of course depends on the defect to be recognized but also on the conditions under which the measurements were made. This paper introduces a preliminary study on machine diagnosis by combining both signal processing methods and artificial intelligence. This work is dedicated to develop a system which allows us to measure several high frequency channels and to transmit them to a computer via USB interface.

1 Introduction:

In a production environment many parts are unfairly detected as defective when monitoring is based on indicators from the literature. The causes of these errors are often related to the not conducive noisy environment to such a diagnosis by records sensitive to disturbance. Moreover, from one production site to another, it is not possible to apply the same default detection thresholds because of a different environment involving a variation of the structures and of the product frequency responses. Therefore, today it remains difficult to do a relevant diagnosis in a noisy environment and particularly on non-stationary signals. The aim of the study is to improve this diagnosis by first using a microphone antenna and then operating an artificial intelligence process on a database acquired on production benches. The microphone array leads to the provision of a spatial map of the acoustic field generated by the monitored system. An acoustic imaging approach allows the addition of a new spatial dimension in the data representation. The preliminary study presented consists in differentiating several states of the system to be monitored from the simultaneous exploitation of information expressed in the time-frequency-space domain.

1.1 Rotating Machine

The system considered for the study is a starter. A starter is an electric motor used to rotate a thermal engine until the self-combustion in the engine takes over. It is composed by several subassemblies such as a planetary gear reducer, a drive shaft, an armature permitting under the effect of a magnetic field the rotation of the motor and a current transmission subassembly allowing the current to flow to the armature by brushes.

Each subset radiates its own acoustic signal which, by interactions, contributes to the overall signal emitted by this motor. These interactions associated to the resonance phenomena in the transient operating phases of the machine make the diagnosis more difficult. Indeed, this diagnosis allows to focus the acoustic measurement on the rotating machine by a beamforming process while freeing of the disturbing sources coming from other directions. First results of comparisons will be presented.

There are three states of the electric motor to recognize:

- Healthy state
- Armature imbalance
- Gear defect

Figure 1: Electric motor

1.2 Classification

Automatic classification is a branch of artificial intelligence. Artificial intelligence, commonly known as AI, is defined as the set of theories and techniques used to make machines capable of simulating human intelligence [1]. This field was born in the middle of the 20th century with the development of computer science and the ambition to create machines with the ability to think in a similar way in their functioning to the human mind in terms of perception, understanding and even in taking decision.

Automatic classification is used in all pattern recognition systems. A pattern in the broad sense is an object of very varied nature. It can be a bar code, a face, a fingerprint and more generally a digital data suite that will constitute the signature of the belonging of an object to a family.

Indeed, these systems allow the algorithmic characterization of objects and consist of assigning an object to a class or category based on prior learning.

Learning is the process of constructing a general model based on particular observations of the real world in order to predict a behaviour or a decision in front of new unseen data. The second definition of learning, taking more the sense of training, consists of improving the performance of the model in a progressive way by being confronting to the exercise of an activity.

This idea of improving the accuracy of models through training is easy to perceive for humans since the ability of a person to perform a certain task is often judged by his experience in the field.

2 Experimentation

2.1 Simple sensor analysis

2.1.1 Machine learning

Supervised Machine Learning requires expert know-how in the intended field application. Indeed, it is a question of being able to label the samples correctly on the one hand but also to define relevant indicators to characterize samples regarding the classification to be carried out in the sense that these indicators must be representative for class distinction.

Here the chosen features are indicators coming from signal processing such as:

- RMS value
- Peak to peak
- Kurtosis
- Partial levels around kinematic frequencies
- Global level
- Energy in one-third octave band

In order not to bias the classification for the algorithms used in this part, a step of standardization of these different parameters is necessary. This consists in subtracting a value of a parameter by the average of the values of this parameter and then by dividing the obtained value by the standard deviation of the parameter. For most classification algorithms it is also important to use the same number of samples per class, i.e. to have a balanced dataset. Indeed, Bayesian algorithms introduce the probability of belonging to a class in the calculation of the conditional probability that a sample belongs to a class knowing its characteristics.

In this part we are primarily interested in a diagnosis of the electric motor based on a single microphone recording. The setup consists on the analysis of a signal coming from a single omnidirectional microphone located at 50 centimeters from the electric motor. The dataset is composed by 25 starters per class.

	Bayesian	Neural Network	Decision Tree	Support Vector Machine	Random Forest
True Healthy	24	21	21	21	25
False Healthy	2	5	2	7	1
True Imbalance	25	23	25	21	25
False Imbalance	0	1	1	2	0
True Gear	23	22	22	22	24
False Gear	1	3	4	2	0
# Error	3	9	7	11	1
Performance	96%	88%	90.67%	85.33%	**98.67%**

Table 1: Classification results

The classification mode used for the results obtained above is the so-called cross-validation mode. Cross-validation consists in dividing the population into N groups of constant size. Then, N-1 groups are randomly selected and used to learn and create a model while the Nth group is used as a test population. This step is repeated N times so that all samples are used for learning and testing in order to create N models.

The global performance is an average of the N classification results created. Cross-validation thus makes it possible to obtain a faithful and global performance of the tested algorithm but has the disadvantage of non-negligible cost in terms of computing time.

Table 1 indicates a very good fault recognition performance for all algorithms, particularly for Bayesian and Random Forest, which respectively achieved 96% and 98.67% of recognition. These results confirm the interest of using such methods for machine diagnosis. However, these measurements were carried out in a healthy environment without external disturbance.

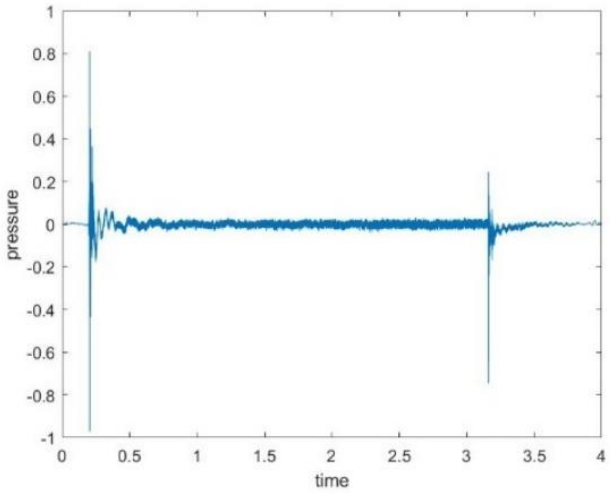


Figure 2: Acoustic pressure signal of the electric motor in unload condition

In the first case, the Mel Frequency Cepstrum Coefficients (MFCC) are extracted from the time signal. This extraction of the coefficients is developed using the Fast Fourier Transform (FFT) and the Discrete Cosine Transform (DCT) on the Mel scale which is a perceptual parametric model. These are the most used criteria in Automatic Speech Recognition (ASR) systems.

First, the signal is split into N windows of a few milliseconds with overlapping (usually in ASR the length of the window is 25 ms and the overlap is 10 ms). A Hamming window is then applied to the signal to limit the spectral distortion (appearance of parasitic high frequencies) related to the overlap and the disturbance at the beginning and the end of the window before passing to the frequency domain via the FFT.

A conversion of the frequency scale f to the Mel scale is performed according to

$$mel(f) = 2595 * \log(1 + \frac{f}{700}). \quad (1)$$

A triangular response filter bank with variable frequency bandwidth is applied to simulate the response of the human ear in the best possible way. This band variation represents the capacity of the human being to be able to easily distinguish two near frequencies at low frequencies than at high frequencies. For each triangular filter, the sum of the energies is calculated so we get as many coefficients as filters.

Finally, we convert the logarithmic spectrum of Mel obtained to the time domain with the DCT, and then we usually keep the first 12 coefficients for each of the windows.

The coefficients are calculated according to

$$C_k = \sum_{i=1}^{N} \log(E_i) * \cos[\frac{\pi k}{N}(i - 0.5)] \quad (2)$$

- N: number of filters.
- E_i: energy calculated with the ith filter.
- C_k: kth Mel Frequency Ceptsrum Coefficient.

Short windows are used for the transient operation phases of the motor while longer windows are used on the stationary phase (this phase is not used in its entire duration; an average and a standard deviation are calculated to take into account any variation of the remaining signal). In addition, to add more information and thus improve the signal recognition, the differential coefficients Δ and acceleration ΔΔ are implemented. These

coefficients, calculated directly from the MFCC coefficients taking respectively the derivative of the first and the second order, make it possible to consider the dynamics of the signal.

According to [2], the addition of these differential coefficients and accelerations increases the recognition by about 20%. At least 702 features were extracted from each time signal. Here the dataset is reduced to 15 starters per class.

	Bayesian	Decision Tree	Support Vector Machine	Random Forest
True Healthy	13	11	11	14
False Healthy	7	7	4	3
True Imbalance	14	14	15	14
False Imbalance	0	2	2	0
True Gear	9	9	10	13
False Gear	2	2	3	1
# Error	9	11	9	4
Performance	80%	75.56%	80%	**91.11%**

Table 2: Classification results using MFCCs

Table 2 shows the performance achieved by algorithms. Random Forest remains more efficient than the other algorithms with a recognition performance of 91.11%. However, the overall performance achieved with this approach is less than the performance obtained previously in Table 1.

2.1.2 Deep Learning

The aim here is not to extract previously mentioned indicators, but rather to provide complete acoustic data to the algorithms. Spectrogram as input was studied.

The classification consists in identifying the class of each starter by recognizing the spectrogram of the acquired signal. For this study, an image processing approach based on a pre-learnt Convolutional Neural Network (CNN) AlexNet is used. AlexNet was developed at Toronto university [3] for the ImageNet LSVRC-2010 contest, a competition for which it is proposed to classify 1.2 million images into 1000 different classes. The authors achieved a winning top-5 test error rate of 15.3% for the ILSVRC-2012 contest.

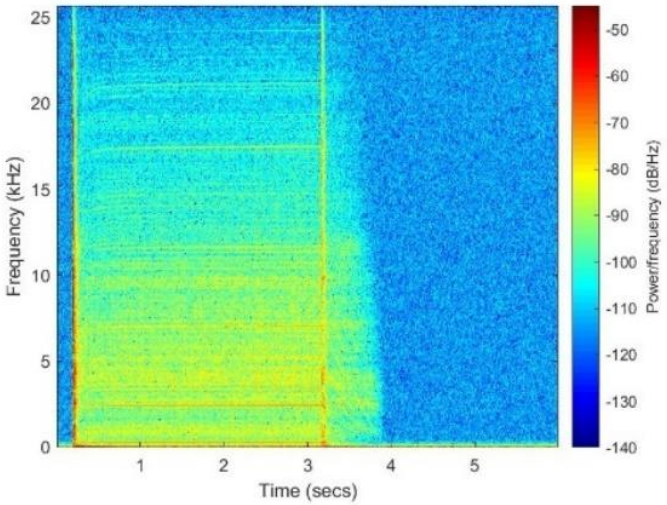


Figure 3: Acoustic signal spectrogram of the electric motor in unload condition

Transfer learning consists in using a trained Neural Network to extract features, which achieved very good performance in recognition on a specific field with thousands of samples for the learning phase, and to transfer the knowledge learnt in that field to another field. This is what we have done here, AlexNet has been trained on 1 million images and performs very well on object recognition in images.

This allows us to extract relevant features for image recognition even if we only have a low amount of data. A modification of the final layers and a training phase are necessary in order to adapt the classification to our field by changing the number of outputs and the weights of each neuron.

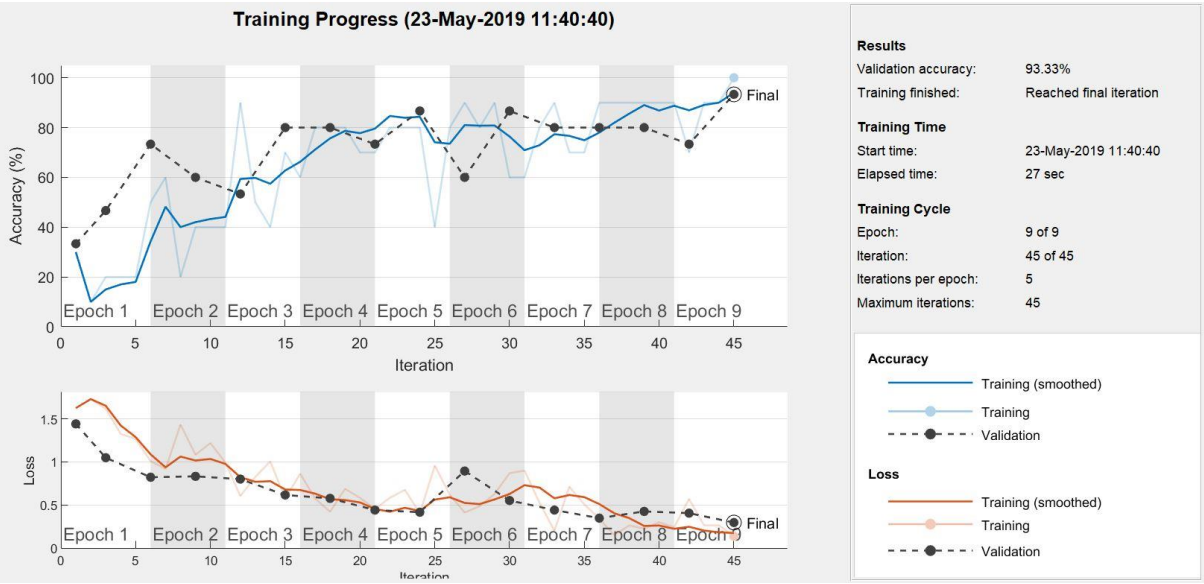


Figure 4: Acoustic pressure by time of the electric motor in unload condition

We achieved a good validation accuracy of 93.33% with this method compared to results obtained by MFCCs. However, it is possible that the model overfits the data because of a large number of parameters (60 million features) compared to the number of samples. In statistics, overfitting corresponds to a model or an analysis which fits perfectly with a dataset. In machine learning, overfitting is one phenomenon to avoid because these models perform very well on the data used to create the model but cannot generalize on new data which means low performances.

2.2 Acoustic localization

Unfortunately, it is not possible to obtain such a recognition rate in a production environment. Indeed, the random ambient noise generates ambiguity perturbing the recognition of defects. This is why we have chosen to introduce acoustical imaging to our work in order to locate the sound sources in a space, focus the measurement towards a privileged direction and thus filter the signal emitted by the electric motor from the ambient noise.

An acoustic antenna represents a distribution of microphones according to a specific geometry and number of sensors. The resulting geometry formed by the repartition of these microphones as well as their number condition the performance of the system, which is characterized by the frequency range of use, the resolution and the capacity of localization. For this project, we made the choice of digital MEMS (Micro electro-Mechanical Systems).

This technology introduced for the first time in 1967 [4] corresponds to miniaturized sensors or actuators which couple several physical principles including mechanics and electronics. The miniaturization allowed by this technology responds to a growing need related to congestion, sensitivity, mass production and to the complexification of systems by allowing the addition of new functions in a non-intrusive way.

The method used to represent the acoustic field is a Beamforming-based method.

2.2.1 Acquisition System

The acquisition system is composed of MEMS microphones. The working of these systems remains faithful to those of conventional microphones because in a general way the physical laws governing the different domains are unchanged. A MEMS consists of a fixed substrate (a semiconductor generally made of silicon) and a moving part.

In the case of the MEMS microphone, the moving part is represented by a membrane. A conductor measures the impedance variation induced by the deformation of the membrane and this allows to obtain a minimum size while not skimping on the performance of the sensor.

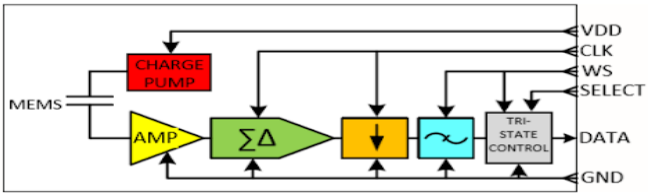


Figure 5: Block diagram of the I2S digital mic

- The charge pump (in red) allows the supply voltage to be raised to a level necessary to polarize the transducer.

- The MEMS is responsible for converting the measured pressure into a voltage.

- The amplifier (in yellow) stores the voltage from the MEMS and amplifies the signal.

- The Sigma Delta converter (in green) converts the analog signal in memory into a pulse-density modulation (PDM) signal on 1-bit resolution.

- The decimator (in orange) converts the PDM signal into a 24-bit pulse coded modulation (PCM) signal by down-sampling with a factor of 64.

- A low pass filter (in blue) removes the remaining high frequency components.

- The three-state mode (in gray) makes it possible to associate two MEMS on the same I2S line for a stereo recording.

Pre-assembled I2S digital MEMS microphones with welds on PCB plate already made were chosen for convenience because the size of the microphone involves a high precision in the welding process.

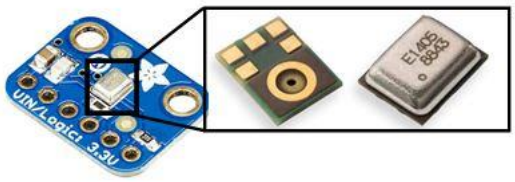


Figure 6: I2S MEMS microphone

An I2S output format MEMS implies that the microphone integrates a large part of the acquisition chain and a MiniDSP acquisition board is used to measure up to 8 synchronous I2S channels by board.

2.2.2 Beamforming

Beamforming [5] is a signal processing tool used in the field of antennas for the directional signal transmission or reception. The main principle of this method is to combine the elements of a sensor array in such a way that in a particular direction signals interfere constructively whereas in other directions the interferences are destructives. It is a question of measuring and applying a delay on the signals acquired by the microphones.

Beamforming is used with radio or sound waves and has many applications in radar, sonar, seismology and acoustics. It allows a representation of a wave field by an estimate of the directions of arrival.

Let an antenna formed by 10 microphones (red stars in FIG. 6), a sinusoidal source (green circle in FIG. 6) with a frequency of 2500 Hz and a white noise (blue circle in FIG. 6). The sources are both placed 0.20 meters from the plane of the antenna and spaced from each other by 0.05 meters.

Through a Beamforming algorithm, we are able to distinguish these two sources.

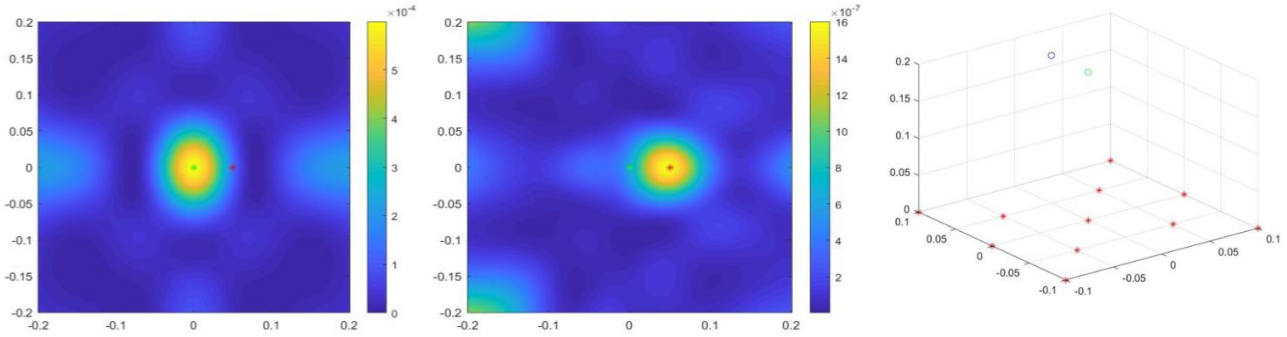


Figure 7: Measurement of the pressure field at 2500 Hz on the left, at a different frequency in the middle

Using an antenna makes the diagnosis less sensitive to disturbance and thus more reliable. We can from this observation reconstruct the acoustic signal coming only from the direction of the source in order to apply the diagnosis to a non-noisy signal. Therefore, we hope to find at least the results obtained in Table 1. An improvement in these results will then be considered from a more advanced imaging diagnosis.

Conclusion and further works

The results introduced above show that the use of classification algorithm is relevant to diagnose electric motors. However, even if these results seem good it is not possible to consider an implementation of these tools in a production bench because 1.33% of False Positive misclassification, achieved with Random Forest in 2.1.1, on a production of 5000 starters per day first means that 67 electric motors are misclassified and in this particular case 67 electric motors are in a bad condition but classified as healthy could be sent to customers.

We will first focus the work on collecting data with the microphone array in a production environment and then improve the recognition by using classification algorithms developed on time signals and spectrograms. By using algorithms near from our application field and adding spatial informations with Beamforming, we hope increase the recognition tasks.

References

[1] C. Bishop, Pattern Recognition and Machine Learning, New York, Springer, 2006.

[2] Hanson, Brian, and Ted Applebaum. "Robust speaker-independent word recognition using static, dynamic and acceleration features: Experiments with Lombard and noisy speech." Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on. IEEE, 1990.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks", Pages 1097-1105 NIPS'12, 2012.

[4] H.C.Nathanson, et al., "The Resonant Gate Transistor", IEEE Trans. Electron Devices, March 1967, vol. 14, 1967.

[5] J.J. Christensen, J. Hald, “Beamforming”, Brüel & Kjaer Technical Review N°1, 1-48, 2004.

[6] J.-H. Thomas, Diagnostic d’un système non stationnaire à partir d’une approche de reconnaissance des formes floue. Application au diagnostic de cliquetis moteur, Traitement du Signal, 16 (3), 203-216, 1999.

[7] M. Mein, Surveillance acoustique de moteurs diesel, thèse de doctorat de l’Université du Maine, Le Mans, 1997.

Macroscopic-Microscopic Attention in LSTM Networks based on fusion Features for prediction of bearing remaining life

QIN Yi[1,2] XIANG SHENG[1,2] HAORAO YAN[1,2]

[1] College of Mechanical Engineering, Chongqing University, Chongqing 400044, China
qy_808@aliyun.com
a1121623518@1636.com
[2] State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China
18334729055@163.com

Abstract

In the mechanical transmission system, the bearing is one of the most widely used transmission components. The failure of the bearing will cause serious accident and huge economic loss. Therefore, the remaining life prediction of the bearing is greatly important. In order to predict the remaining life of the bearing, a prediction method combining macro-micro attention, long-short-term memory neural network and isometric mapping is proposed. First, some typical time-domain and frequency-domain characteristics of the vibration signal are calculated .respectively, such as the the mean frequency, the absolute mean value, the standard deviation, the RMS and so on. Then, the principal component of these characteristics is extracted by the isometric mapping method. The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

1 Introduction

Bearing is widely used in mechanical equipment and is one of the most universally used mechanical parts[1]. Under the complex working condition and environment, bearings are easily subject to failures, which may result in the catastrophe of the machine running and even threaten the personal safety [2]. Bearing remaining life prediction is beneficial to determine the equipment maintenance time reasonably, improve the production efficiency, reduce the accident rate, and prevent the sudden accidents, which is significant for engineering production [3].Recently, with the rapid developments in sensing, signal processing and artificial intelligence technology, Prognostics and Health Management (PHM) technologies based on data (experience) have gradually become the mainstream solution either in fault diagnosis or remaining useful life（RUL）estimation, instead of physics-based methods which can be expensive and tedious to develop[4]. Neural network is one of the most advanced models for the study of sequence classification and prediction in data driven methods. Zhao et al. [5] introduced the application of deep learning in machine health monitoring system, mainly from the perspective of automatic encoder (AE) and its variants, restricted boltzmann machine and its variants, including deep belief network (DBN) and deep boltzmann machine (DBM), convolutional neural network (CNN) and recursive neural network (RNN). Qin et al. [6] proposed a model for fault diagnosis of wind turbines' gearboxes based on deep belief networks (DBNs) with improved logistic sigmoid units via extracting impulsive features. Gebraeel et al. [7] used the amplitude of the bearing vibration decay signal as the criterion to assess the life, and used the BP neural network to predict the life.

Based on particular design, RNN is suitable for processing timing-related information [8]. However, RNN also has its own drawbacks, e.g. the excessive recursion time is indirectly equivalent to increasing the depth of neural

network and the training time; the vanishing gradient problem usually [9]. In order to solve these problems, long short-term memory (LSTM) was proposed by Hochreiter and Schmidhuber in 1997 [10]. It can avoid the long-term dependence problem of RNN, thus it has been widely used. Yuan et al. [11] investigated three RNN models including vanilla RNN, LSTM and GRU models for fault diagnosis and prognostics of aero engine. They found that these advanced RNN model based on LSTM and GRU performed better than the convetional RNN via a number of experiments. Elsheikh et al. [12] proposed the bidirectional handshake network which solved the problem that bidirectional LSTM could not be well utilized in the prediction field, and proved the superiority of the method in the life experiment of turbine engines.

Attention mechanism can be regarded as a kind of contribution screening of information which improves the efficiency of neural network by selecting key information for processing. Attention-based recurrent networks have been successfully applied to a wide variety of tasks, such as handwriting synthesis[13], machine translation[14], image caption generation[15] and visual object classification[16]. In the prediction field, attention-based LSTM is getting more and more attention. Ran et al. [17] substituted a tree structure with attention mechanism for the unfolding way of standard LSTM to construct the depth of LSTM and modeling long-term dependence for travel time prediction. Fernando et al. [18] combined two kinds of attention and used LSTM for human trajectory prediction and abnormal event detection. Filtering key information can reduce computing resources, but it can also cause some degree of information loss. Differential treatment of input data according to the screening of attention mechanism can not only reflect the focus to important information, but also retain useful information as far as possible.The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

2 Related work

2.1 LSTM

The LSTM neural network is a special RNN neural network which also consists of an input layer, a hidden layer, and an output layer. The difference lies in using an LSTM structure that includes the input gate, the output gate, the forget gate, and the memory cell as the hidden layer, as shown in Figure. 1.

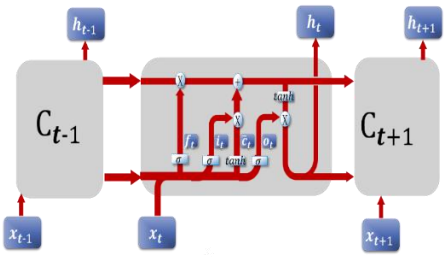


Figure. 1 Hidden layer structure of LSTM neural network

The forget gate is used to determine whether to keep the historical information stored in the current memory cell. If the door is opened, the historical information stored in the current memory cell is retained, otherwise the historical information is forgotten. The input gate is used to determine whether to allow the input layer information to enter the current memory cell. The open door allows the input layer signal to enter, and the closed door does not allow. The output gate is used to determine whether to output the current input layer signal to the next layer, the open door allows signal output and the closed door does not allow.

An LSTM network computes a mapping from an input sequence $\mathbf{x} = (x_1, x_2, \ldots x_n)$ to an output sequence $\mathbf{y} = (y_1, y_2, \ldots y_n)$ by calculating the network unit activations using the following equations iteratively from t = 1 to T:

$$
\begin{aligned}
\mathbf{i}_t &= \sigma\left(\mathbf{w}_{ix}\mathbf{x}_t + \mathbf{w}_{im}\mathbf{h}_{t-1} + \mathbf{b}_i\right) \\
\mathbf{f}_t &= \sigma\left(\mathbf{w}_{fx}\mathbf{x}_t + \mathbf{w}_{fm}\mathbf{h}_{t-1} + \mathbf{b}_f\right) \\
\mathbf{c}_t &= \mathbf{f}_t \quad \mathbf{c}_{t-1} + \mathbf{i}_t \quad \tanh\left(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{cm}\mathbf{h}_{t-1} + \mathbf{b}_c\right) \\
\mathbf{o}_t &= \sigma\left(\mathbf{w}_{ox}\mathbf{x}_t + \mathbf{w}_{om}\mathbf{h}_{t-1} + \mathbf{b}_o\right) \\
\mathbf{h}_t &= \mathbf{o}_t \quad \tanh\left(\mathbf{c}_t\right)
\end{aligned} \tag{1}
$$

where, $\mathbf{i}$ is the input gate, $\mathbf{o}$ is the output gate, $\mathbf{f}$ is the forgetting gate, $\mathbf{c}$ includes cell activation vectors, and $\mathbf{h}$ is the memory cell outputs. $\mathbf{w}$ represents the weight matrix (for example, the weight $\mathbf{w}_{ix}$ matrix representing the input $\mathbf{x}$ to the input gate), and $\mathbf{b}$ represents the threshold ($\mathbf{b}_i$ which is the threshold of the input gate). σ is the sigmoid activation function, $\tanh$ is the tanh activation function, represents dot product.

3 Macroscopic-Microscopic Attention in LSTM

The structure of MMALSTM neural network is in Figure. 2. The number of input cells and the number of output cells in MMALSTM are respectively set as 60 and 1, and the learning rate is set as 0.05. The number of hidden layer cells is set as 17 in this study. The initialization method of neural network employs the standard initialization. And the recurrent neural network based on MMA is deduced as follows.

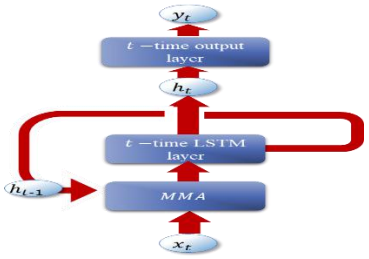


Figure. 2 structure of MMALSTM neural network

Firstly, we deal with the data matrix and calculate its macro and micro attention coefficient by the macro -micro attention mechanism. The input information $\mathbf{X}_t = \lceil \mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_t \rceil$ at time t is defined as $\mathbf{x}_t = \lceil x_{t,1} \quad x_{t,2} \quad \cdots \quad x_{t,n} \rceil$ and the recurrent information at time t-1 is defined as $\mathbf{h}_{t-1} = \lceil h_{t-1,1} \quad h_{t-1,2} \quad \cdots \quad h_{t-1,m} \rceil$. Macro attention mechanism is to process the data in the whole time interval by attention mechanism. Micro attention mechanism is to process the data $\mathbf{x}_t = \lceil x_{t,1} \quad x_{t,2} \quad \cdots \quad x_{t,n} \rceil$ and $\mathbf{h}_{t-1} = \lceil h_{t-1,1} \quad h_{t-1,2} \quad \cdots \quad h_{t-1,m} \rceil$ in each time instant by attention mechanism. Thus the operation of attention mechanism on data in the whole time dimension and each time dimension is called macro-micro attention mechanism. In this paper, the macro -micro attention mechanism of input matrix and the micro attention mechanism of recurrent matrix are processed (The recurrent data in the whole time dimension is not known before input into the network, so it cannot be processed at the macro level). In the training process, $\bar{x}_{T+1}$ is set as query vector q_M at the macro level, $x_{t+1.n}$ is set as query vector q_m at the micro level. In the prediction process, $\bar{x}_T$, $x_{t.n}$ represents q_M and q_m separately. And then the attention coefficients are calculated by:

$$
\chi_i = \frac{\exp\left(S\left(\bar{x}_i, \mathrm{q}\right)\right)}{\sum_{j=1}^{T} \exp\left(S\left(\bar{x}_j, \mathrm{q}\right)\right)} \tag{2}
$$

$$
\alpha_i = \frac{\exp\left(s\left(x_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n} \exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m} \exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)} \tag{3}
$$

$$
\lambda_i = \frac{\exp\left(s\left(h_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n} \exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m} \exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)} \tag{4}
$$

where α_i , λ_i are micro attention coefficients corresponding to input data and recurrent data respectively, χ_i is macro attention coefficients corresponding to the whole input data, $\overline{x}_i$ represents the mean value of $\mathbf{x}_i$ .And

$$S(\overline{x}_i, q) = \frac{\overline{x}_i q}{\sqrt{t}} \tag{5}$$

$$s(x_i, q) = \frac{x_{t,i}^T q}{\sqrt{n+m}} \tag{6}$$

$$s(h_i, q) = \frac{h_{t-1,i}^T q}{\sqrt{n+m}} \tag{7}$$

Secondly, according to the macro -micro attention coefficient and Eq. (4), the weight of input data is amplified in real time at multiple levels. And the weight of recursive data is amplified in real time based on the micro attention coefficient.

$$\begin{aligned}
\mathbf{w}'_{ix} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ix} \\
\mathbf{w}'_{ox} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ox} \\
\mathbf{w}'_{fx} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{fx} \\
\mathbf{w}'_{ih} &= (1+\lambda_i)\mathbf{w}_{ih} \\
\mathbf{w}'_{oh} &= (1+\lambda_i)\mathbf{w}_{oh} \\
\mathbf{w}'_{fh} &= (1+\lambda_i)\mathbf{w}_{fh}
\end{aligned} \tag{8}$$

From the above, the flow chart of MMA, MA, ma are depicted in the follow figures. Figure. 3 represents the flowchart of MMA, Figure. 4 (a)represents the flowchart of MA, Figure. 5(b) represents the flowchart of ma.

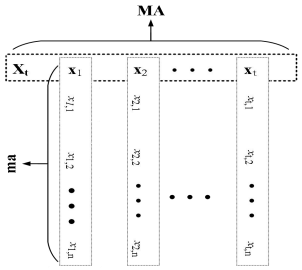
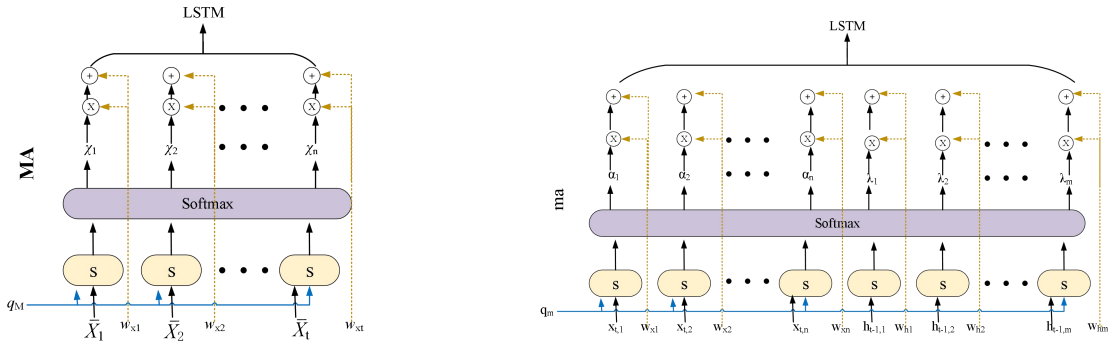


Figure. 3 The flowchart of MMA



(a) (b)

Figure. 4 The flowchart of MA and ma.

With the weight amplified based on MMA, a new variant of LSTM is proposed, which is named as MMALSTM. Via Eq. (1), we can derive the calculation formula of MMALSTM as follows:

$$
\begin{aligned}
\mathbf{i}_t &= \sigma\left(\mathbf{w}'_{ix}\mathbf{x}_t + \mathbf{w}'_{ih}h_{t-1} + \mathbf{b}_i\right)\\
\mathbf{f}_t &= \sigma\left(\mathbf{w}'_{fx}\mathbf{x}_t + \mathbf{w}'_{fh}h_{t-1} + \mathbf{b}_f\right)\\
\mathbf{c}_t &= \mathbf{f}_t \quad \mathbf{c}_{t-1} + \mathbf{i}_t \quad \tanh\left(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{r}_{t-1} + \mathbf{b}_c\right)\\
\mathbf{o}_t &= \sigma\left(\mathbf{w}'_{ox}\mathbf{x}_t + \mathbf{w}'_{oh}h_{t-1} + \mathbf{b}_o\right)\\
\mathbf{h}_t &= \mathbf{o}_t \quad \tanh\left(\mathbf{c}_t\right)\\
\mathbf{y}_t &= \mathrm{g}\left(\mathbf{w}_{yh}h_t + \mathbf{b}_y\right)
\end{aligned}
\tag{9}
$$

where g is the linear activation function.

4 Prediction method based on MMALSTM

In summary, the steps of the proposed method for predicting the bearing remaining life based on the combination of the isometric mapping algorithm [19] and MMALSTM neural network are given below:

1. During the bearing life cycle, collect the bearing vibration samples according to the interval Ts between two adjacent samples. Suppose that the number of sample is n.

2. Calculate the 5 kinds of time-frequency characteristics （time-domain :mean absolute difference，standard deviation and root-mean-square, frequency-domain: mean of frequency distribution, envelope spectrum: mean of frequency distribution）of these vibration samples separately, then the eigenvalue matrix X of the size $n\times 5$ can be obtained.

3. Select the eigenvalue matrix $X1$ of the $n1$ samples from X as the training matrix.

4. The training matrix $X1$ and original matrix X are respectively processed by the ISOMAP algorithm and Savitzky-Golay filtering, and the obtained eigenvectors with the largest eigenvalues after filtering $V1=\left(v1_1, v1_2, \ldots, v1_{n1}\right)^T$ and $V=\left(v_1, v_2, \ldots, v_n\right)^T$ are used as their principal components respectively. The matrix X and the vector V are only used to verify the validity of this method, which don't participate in neural network training and prediction.

5. Since the size of the matrix X is larger than that of the matrix $X1$ and both the sums of V and $V1$ are equal to zero according to the characteristic of ISOMAP algorithm, the two vectors may have different starting value, even though their trends are same. Therefore, it is necessary to unify them. By minimizing

$$
E=\sum_{i=1}^{n1}\left(v1_i - av_i - b\right)^2 \tag{10}
$$

a and b can be computed, and then all the elements of V are unified through the following equation

$$
v_i^{'} = av_i + b \tag{11}
$$

6. Linearly normalize the vector $V1$ to obtain the normalized vector $W=\left(w_1, w_2, \ldots, w_{n1}\right)^T$.

7. Reconstruct matrix U:

$$
U=\begin{bmatrix} w_1 & w_2 & \cdots & w_{n1-p} \\ w_2 & w_3 & \cdots & w_{n1-p+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p+1} & w_{p+2} & \cdots & w_{n1} \end{bmatrix}=\begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \vdots \\ \mathbf{u_{p+1}} \end{bmatrix} \tag{12}
$$

where p is the cell number of input layer and

$$
\mathbf{u}_i=\begin{bmatrix} w_i & w_{i+1}\cdots & w_{n1-p+i-1} \end{bmatrix} \tag{13}
$$

8. Set the first p vectors of the matrix U as the input of the MMALSTM neural network and the last vectors as the output respectively, then train the MMALSTM neural network.

9. With the trained MMALSTM neural network, the mapping function f for prediction is defined. By inputting the last p vectors of the matrix U into the trained MMALSTM, the output $\overline{u}_{p+2}$ at the first prediction time (FPT) can be calculated as

$$
\overline{\mathbf{u}}_{p+2}=f\left(\mathbf{u}_2, \mathbf{u}_3 \cdots, \mathbf{u}_{p+1}\right) \tag{14}
$$

Then U is updated by

$$U=[\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_{p+1} \quad \bar{\mathbf{u}}_{p+2}]^T \tag{15}$$

10. Repeating step 9 several times, we can obtain the predicted data series $\bar{u}_i$.

$$\begin{aligned}
\bar{\mathbf{u}}_{p+3} &= f\left(\mathbf{u}_3,\mathbf{u}_4\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2}\right)\\
\bar{\mathbf{u}}_{p+4} &= f\left(\mathbf{u}_4,\mathbf{u}_5\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2},\bar{\mathbf{u}}_{p+3}\right)\\
\bar{\mathbf{u}}_{p+5} &= f\left(\mathbf{u}_5,\mathbf{u}_6\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2},\bar{\mathbf{u}}_{p+3},\bar{\mathbf{u}}_{p+4}\right)\\
&\vdots\\
\bar{\mathbf{u}}_i &= f\left(\mathbf{u}_{i-p},\mathbf{u}_{i-p+1}\cdots,\mathbf{u}_{p+1},\bar{\mathbf{u}}_{p+2}\cdots\bar{\mathbf{u}}_{p+(i-p-2)},\bar{\mathbf{u}}_{p+(i-p-1)}\right)
\end{aligned} \tag{16}$$

When the predicted characteristic data $\bar{w}_{n1-p+i-1}$ ($\bar{w}_{n1-p+i-1}\in\bar{\mathbf{u}}_i$) exceeds a preset threshold, the bearing remaining life can be calculated by $(i-n1)\times T_s$. Then we anti-normalize the predicted data series and get $\bar{v}_i$ ($i=n1+1,n1+2,\cdots,n$), which can be compared with the actual vector $V'=\left(v'_{n1+1},v'_{p+2},\ldots,v'_n\right)^T$ to verify the validity of this method.

5 Experimental signal analysis

The experimental data come from PRONOSTIA in the IEEE PHM 2012 Data Challenge [20]. The platform mainly contains three major parts: a rotatory part, a degradation generation part, and a signal acquisition part. To accelerate the degradation of bearing, radial load force is applied with a controllable shaft speed. Two accelerated sensors perpendicular to each other are installed on the key position for the test bearing. The sampling frequency is 25 600 Hz. Each sample has a duration of 0.1 s, which means each sample has 2560 points. The record interval is 10 s. The test is ceased once the amplitude of the collected signal surpasses a certain level to prevent damage. Our proposed method is applied on the data set bearing1-1, which means that our method currently only considers the operation condition with constant speed and load. The run-to-failure bearing1-1 contains 2803 samples. Moreover, many samples were acquired under the steady stage, and their characteristics were almost the same, which were of little significance to use these data for prediction. Therefore, only the last 1750 samples were used for bearing life prediction.

The characteristic matrix of 1650 experimental samples is used as the training matrix to predict the 100 characteristic points corresponding to the next 100 time instants. Via the proposed method, the training, predictive and actual curves are illustrated in Figure. 7. As can be seen from Figure. 5, the trend of the predicted curve is not close to actual curves but both have down trend. It follows that the predicted value only exceed the threshold when the MMALSTM neural network predict a large number of points. In such case, the predicted sampling points are 1736, which is 14×10s different from the actual life. In this study, to assess the performance of the proposed approach, an error index is defined as

$$Er=\frac{Rul-\overline{Rul}}{Rul}\times 100\% \tag{17}$$

where Rul denotes the actual remaining useful life, and Rul denotes the predictive remaining useful life. It is easy to note that the prediction error Er is 9.3%.

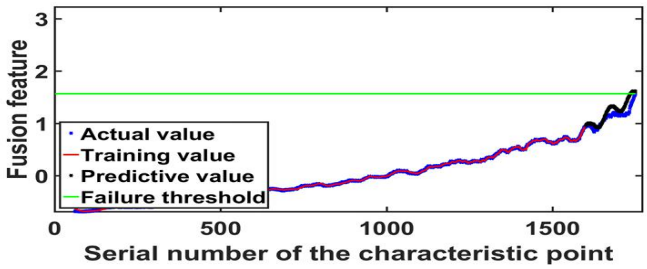


Figure. 5 Failure threshold, training curve, predicted curve and actual curve for 370 experimental samples of real bearing.

Then we add the number of known characteristic points to test the predictive power of the model, each group was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .6.

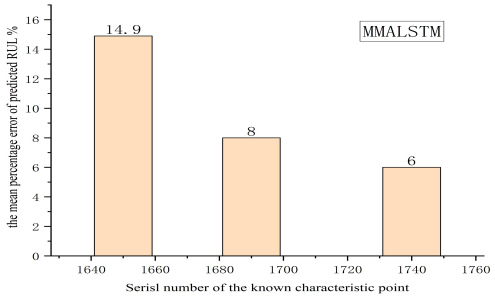


Figure. 6 The obtained prediction errors under the various numbers of the known characteristic point of data set .

Next, given 1690 characteristic points, MMALSTM, LSTM, deep LSTM (DLSTM) and LSTM with a projection (LSTMP) are used for comparison under the same conditions, each method was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .7.

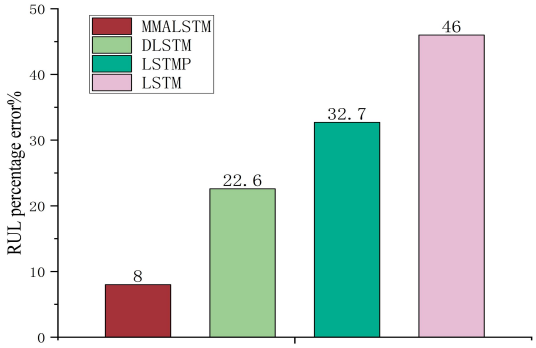


Figure. 10 Comparison of bearing remaining useful life prediction results obtained by the four types of LSTMs.
From the experimental results, it can be concluded that MMALSTM has not higher accuracy, while the four types of LSTMs do not have the satisfactory performance when fewer samples are used. And the closer the prediction point is to the failure point, the more accurate the prediction of bearing remaining life will be. Therefore, the proposed method is more suitable for bearing life prediction with enough samples.

6 Conclusion

Modern manufacturing requires high reliability and high efficiency, which makes bearing health analysis and residual life prediction an increasingly important research field. Advances in networked manufacturing and AI-oriented big data analysis provide new opportunities for bearing health analysis and residual life prediction. In this paper, isometric mapping algorithm is used to fuse 3 time-domain features,1 frequency-domain features and 1 envelope spectrum features of bearing vibration signal into a new feature. Due to the information characteristics of life data, the input weight of neural network and the weight of recurrent layer are all multi-amplified in real time by partial processing based on MMA. We use the improved long and short term neural network(MMALSTM) to study the bearing degradation model and predict the residual service life. The performance of the network model is verified by the data obtained from the bearing lifetime experiment and compared with LSTM, DLSTM and LSTMP. Numerical experiment results show that the proposed method not only has a higher accuracy of predicted. This will not only have certain reference significance and help for the online detection of bearing residual life, but also have great significance for the determination of equipment maintenance time and the prevention of unexpected accidents in engineering production.

In the future we will plan to study the performance of the proposed depth structure model in the field of bearing residual life prediction. Furthermore, the prediction of the residual life of bearings under variable working conditions by the deep learning model is also worth studying and working on

references

[1]Yi Qin, Jingqiang Zou, Baoping Tang, Yi Wang, Haizhou Chen. Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary, IEEE Transactions on Industrial Informatics, 2019, DOI 10.1109/TII.2019.2909305

[2]Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 7, pp. 1693-1702, 2017.

[3]L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 8, pp. 1887-1899, 2018.

[4] X. Han, Y. F. Zhu. Prediction of Residual Fatigue Life of In-service Gears[J]. Journal of Wanxi University, 2003, 19(2): 85-87.

[5]R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

[6]Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," IEEE Transactions on Industrial Electronics, vol. 66, no. 5, pp. 3814-3824, 2019.

[7]Gebraeel N, Lawley M, Liu R, et al. Residual life predictions from vibration-based degradation signals: a neural network approach[J]. IEEE Transactions on Industrial Electronics, 2004, 51(3):694-700.

[8]Graves A, Mohamed A R, Hinton G. Speech Recognition with Deep Recurrent Neural Networks[J]. 2013, 38(2003):6645-6649.

[9]Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 2002, 5(2):157-166.

[10]Hochreiter S, Schmidhuber J. Long short-term memory.[J]. Neural Computation, 1997, 9(8):1735-1780.

[11]M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," in Aircraft Utility Systems (AUS), IEEE International Conference on, 2016, pp. 135-140: IEEE.

[12] A. Elsheikh, S. Yacout, M.-S. Ouali, Bidirectional handshaking LSTM for remaining useful life prediction, Neurocomputing, 323 (2019) 148-156.

[12]K. B. Charbonneau and O. Shouno, "Neural Trajectory Analysis of Recurrent Neural Network In Handwriting Synthesis," arXiv preprint arXiv:1804.04890, 2018.

[14]H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," Neurocomputing, vol. 284, pp. 171-176, 2018.

[15]P. Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077-6086.

[16]L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," IEEE transactions on cybernetics, no. 99, pp. 1-12, 2018.

[17]X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-Based Method with Attention Mechanism for Travel Time Prediction," Sensors, vol. 19, no. 4, p. 861, 2019.

[18]T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," Neural networks, vol. 108, pp. 466-478, 2018.

[19]A. Najafi, A. Joudaki, and E. Fatemizadeh, "Nonlinear dimensionality reduction via path-based isometric mapping," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 7, pp. 1452-1464, 2016.

[20]P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in IEEE International Conference on Prognostics and Health Management, PHM'12., 2012, pp. 1-8: IEEE Catalog Number: CPF12PHM-CDR.

Milling Diagnosis using Machine Learning Approaches

Dominique KNITTEL[1,2], Hamid MAKICH[1], Mohammed NOUARI[1]

[1]LEM3 (Laboratoire d'étude des microstructures et de mécanique des matériaux), site INSIC, 27 rue d'Hellieule, 88100 Saint Dié des Vosges, France
[2]University of Strasbourg, Faculty of Physics and Engineering, 3 - 5 rue de l'Université, 67000 Strasbourg, France
Corresponding author : knittel@unistra.fr

Abstract

The manufacturing industries more and more pay close attention to artificial intelligence (AI). For example, smart monitoring and diagnosis, real time evaluation and optimization of the whole production and raw materials management can be improved by using machine learning and big data tools.
In this work, a smart milling diagnosis has been developed for composite sandwich structures based on honeycomb core. The use of such material has grown considerably in recent years, especially in the aeronautic, aerospace, sporting and automotive industries. But the precise milling of such material presents many difficulties.
The objective of this work is to develop a data-driven industrial surface quality diagnosis for the milling of honeycomb material, by using supervised machine learning methods. In this approach cutting forces are online measured in order to predict the resulting surface flatness.
The developed diagnosis tool can also be applied to the milling of other materials (metal, polymer, …).

1 Introduction

The manufacturing industries more and more pay close attention to artificial intelligence (AI). For example, smart monitoring and diagnosis, real time evaluation and optimization of the whole production and raw materials management can be improved by using machine learning and big data tools [1]. An accurate milling process implies a high quality of the obtained material surface (roughness, flatness) [2]. With the involvement of AI-based algorithms, milling process is expected to be more accurate during complex operations.

T. Mikołajczyk *et al*. developed an Artificial Neuronal Network (ANN) for tool-life prediction in machining with a high level of accuracy, especially in the range of high wear levels, which meets the industrial requirements [3].

D. Pimenov *et al*. evaluated and predicted the surface's roughness through artificial intelligence algorithms (random forest, standard Multilayer perceptron) [4]: in their investigation the obtained performance depends on the parameters contained in the dataset.

M. Correa *et al.* compared the performances of Bayesian networks (BN) and artificial neural networks for quality detection in a machining process [5]. Even ANN models are often used to predict surface quality in machining processes, they preferred BNs for their significant representation capability and for the fast model building.

The work of C. Zhang *et al*. [21] focused on monitoring the condition and life of the cutting tool in dry milling environment. From de-noised vibration signal they extract some relevant features such as the root mean square, the skewness and the kurtosis in both time and time-frequency domain. Based on Neuro-Fuzzy Network (NFN), they implemented a tool wear prediction model which performs the best,

with the smallest Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) compared with Back Propagation Neural Network (BPNN) and Radial Basis Function Network (RBFN) algorithms.

Z. Rui *et al*. [24] implemented a hybrid approach combining handcrafted feature design with automatic feature learning for machine health monitoring: local feature-based gated recurrent unit (LFGRU) networks. By comparison with some other methods such as the Support Vector Machine (SVM), the k-nearest neighbor (kNN), they verified the effectiveness and robustness of the proposed LFGRU model for tool wear prediction.

D. Wu *et al*. [25] have worked on cloud-based machine learning for tool wear prediction in milling. The research was about the development of a novel approach for machinery prognostics using a cloud-based random forest algorithm. Their experimental result have shown that despite the fact that random Forests give the best accuracy for large dataset, parallel random forest algorithm has the best ratio training time/accuracy. Future more, they will predict tool wear with other machine learning algorithms such as support vector machines as well as to make a comparison with their actual algorithms.

For machining result prediction, similar algorithms could be used but the recurrent problem is how to increase the accuracy of those algorithms. K. Javed. *et al.* [26] have worked on an enabling health monitoring approach based on vibration data for accurate prognostics. They have shown that prognostic efficiency is closely related to the extracted features and by the same way proposed a method for enabling features that can lead to simple and accurate prognostics.

K. Durmus [27], by using neuronal networks, worked on the prediction and the control of surface roughness in CNC lathe using artificial neural network. His study has concluded that artificial neural network (ANN) can produce an accurate relationship between cutting parameters and surface roughness. Based on the ANN training model, he could find the best machining parameters for obtaining a desired surface roughness.

By also using neuronal artificial neural network M. Azlan [28] has developed a surface roughness prediction models for end milling machining, in the logic to find the best ANN network structure for surface roughness prediction.

Another approach consists to measure and analyze the drive power (for example by current measuring) [31], which is not applicable in our experiment. In this paper few artificial intelligence methods are tested: random forest (RF), standard Multilayer perceptrons (MLP), Regression Trees, and radial-based functions.

In our work, a smart milling diagnosis has been developed for composite sandwich structures based on honey-comb core. The use of such material has grown considerably in recent years, especially in the aeronautic, aerospace, sporting and automotive industries. Recent development projects for Airbus A380 or Boeing 787 confirm the in-creased use of the honeycomb material. But the precise milling of such material presents many difficulties.

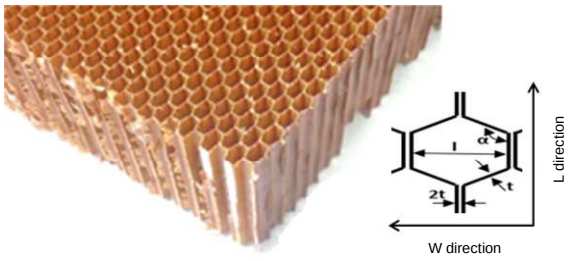
The objective of this work is to develop a data-driven industrial surface quality diagnosis for the milling of honeycomb material, by using supervised machine learning methods. Therefore, cutting forces are online measured in order to predict the resulting surface flatness.

2 Description of the Experiments

2.1 Workpiece material and tools

The workpiece material studied in this investigation is Nomex® honeycomb cores with thin cell walls. It is produced from aramid fiber dipped in phenolic resin (Fig. 1).

Honeycomb designation: **A10-72-5OX**



Density [Kg/m3]	Cell size l [mm]	Wall size t [mm]	Angle α [°]
72	5	0.08	100

Figure 1: *Nomex*® honeycomb cores and the main geometrical characteristics

The honeycomb cores consist of continuous corrugated ribbons of thin foil bonded together in the longitudinal direction. The aim of such a process is to create a structure allowing lightness and stiffness together thanks to the hexagonal geometry of formed cells. Figure 1 illustrates the geometric characteristics of the honeycomb core. The use of honeycomb material in sandwich composite is limited by the fragility of each wall of the honeycomb, which influences the quality of obtained surfaces after machining [7, 8, 9].

The Nomex® honeycomb machining presents several defects related to its composite nature (uncut fiber, tearing of the walls), the cutting conditions and to the alveolar geometry of the structure which causes vibration on the different components of the cutting effort [10].

It is clear that the use of ordinary cutting tools and also the mechanical and geometrical characteristics of honeycomb cores will have a crucial effect on machinability and on the quality of the resulting surface [11]. In fact, ordinary cutting tools for machining honey-comb core produce generally tearing of fibers and delamination of cell structures. Subsequently, these cause a reduction of bond strength between the skin and the honey-comb core, and thus a weaker joint for composite sandwich structures.

In our study, the used milling cutter is provided from our industry partner, the EVATEC Tools Company. As shown in figure 2, the used EVATEC tool is a combined specific tool with two parts designed to surfacing/dressing machining operation. The first part is a cutter body made of high speed steel with 16 mm in diameter and having ten helixes with chip breaker. This tool part is designated by Hogger. The second part is a circular cut-ting blade made of tungsten carbide with a diameter of 18.3 mm and having a rake angle of 22° and a flank angle of 2.5°. These two parts are mechanically linked to each other with a clamping screw.

Figure 2: Milling cutter used for Nomex® honeycomb core ‘‘CZ10’’.

2.2 Milling experiments

All experimental milling tests illustrated in this paper were carried out on a three-axis vertical machining center Realmeca® RV-8.

For assessing the performance of the machining process of Nomex® honeycomb core we monitored and measured the cutting forces generated during cutting, by using the Kistler dynamometer model 9129AA. The Kistler table is mounted below the Nomex sample in order to measure the three components of the machining force as shown in figure 3. During the measurements, the x-axis of the dynamometer is aligned with the feed direction of the milling machine and the longitudinal direction of the workpiece (parallel to core ribbons and the direction of honeycomb double wall). The three orthogonal components of machining force (Fx, Fy and Fz) were measured according to figure 3 using the Kistler table.

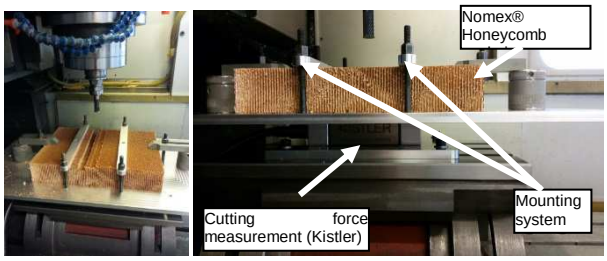


Figure 3: Experimental test setup

The milling experiment conditions are summarized in table 1. Four different speeds (high and low speeds) and four feed values were selected.

Spindle speed (rpm)	2 000	10 000	15 000	23 000
Feed rate (mm/min)	150	1 000	1 500	3 000

Table 1 : Milling experiment conditions

Two main modes of surface damage are observed (Figure 4): uncut aramid fibers along the machined surface and tearing of the walls. The appearance of the uncut fibers is a machining defect specific to the composite material which depends on the type of the fibers and their orientation. The tearing of Nomex® paper, linked to the cellular appearance of the honeycomb structure, occurs under the shear loading effect [28, 29].

Uncut fibers are observed on Figure 4 -a and -c. It is well known that the surface quality is of high importance for the use of the Nomex® honeycomb in sandwich materials. The machining defects cause a reduction of bond strength between the skin and the honeycomb core, and thus a weaker joint for composite sandwich structures.

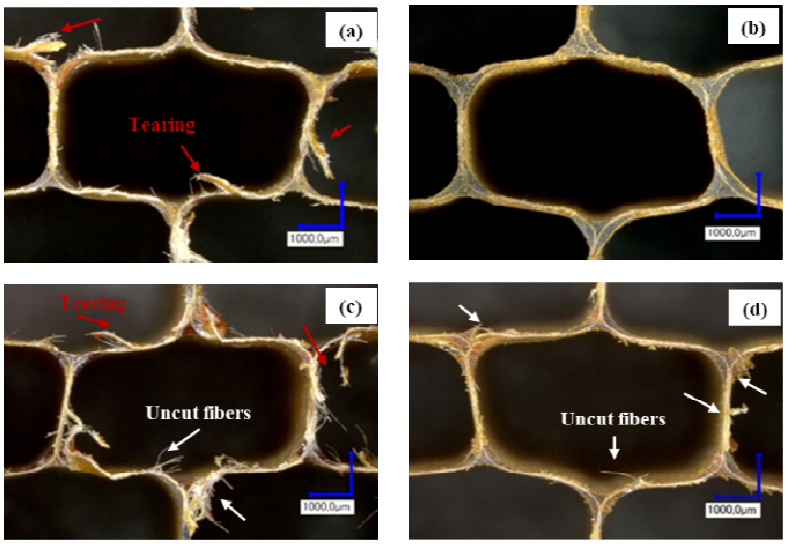


Figure 4: Obtained honeycomb machining surfaces.
The case (b) represents the best milling result

2.3 Measured signals

Many milling experiences have been made in our study. For example, figure 5 shows the milling forces measured for honeycomb at 2000 rpm spindle speed and 3000 mm/min feed rate.

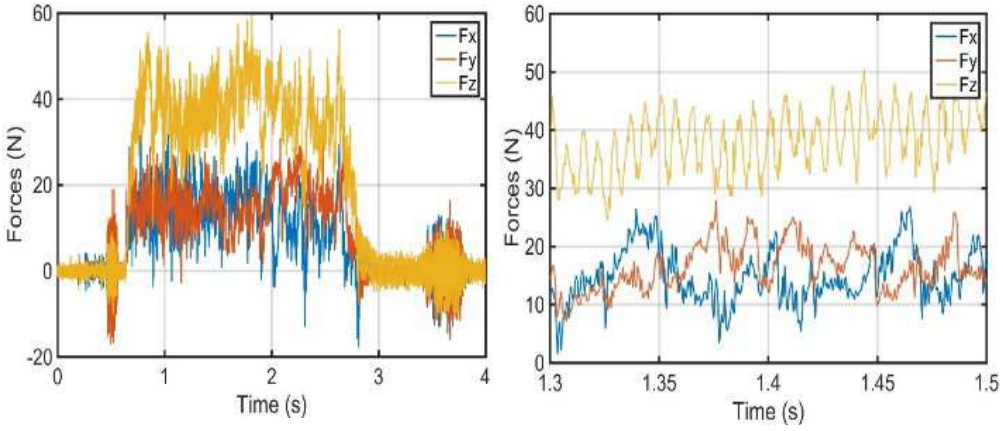


Figure 5: Milling force measurements for 2000 rpm spindle speed and 3000 mm/min feed rate:
(a) during all process; (b) during 0.2s (zoom)

Cutting forces are in the order of a few Newtons, they do not exceed 60 Newtons. Generally, the force in vertical direction (Fz) is quite small, thus, it is advised that to keeping vertical forces small in milling composite due to the delamination issue. In our case, the vertical cutting force component is greater than other forces components which can be attributed to the mechanical properties of the honeycomb structure where the honeycomb structure is characterized by a better out-of-plane compression behavior than its tensile and shear strength. The evolution of cutting forces shows significant oscillations. These oscillations are caused by vacuum in the cells of the honeycomb and the angle between the cutting direction and the honeycomb cell wall direction.

Figure 6 shows the obtained evolution of the surface quality (flatness) for various combinations of cutting conditions (spindle speed and feed rate). The defect of shape is higher for low speeds. Thus, for

high feed rates that exceed the 1500 mm/min, the unevenness exceeds 500 µm which characterizes the severe tearing of the honeycomb walls.

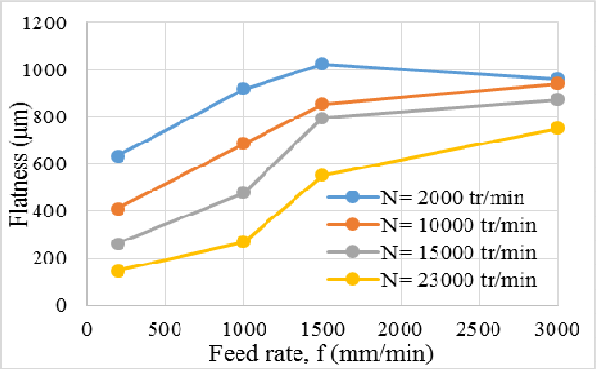


Figure 6: Effect of cutting parameters on surface flatness

Given the low level of cutting forces, the quality of the obtained machined surface allows to establish criteria for determining the machinability of the honeycomb structures. The appearance of the uncut fibers is a machining defect specific to the composite material which depends on the type of the fibers and their orientation. The tearing of Nomex® paper, linked to the cellular appearance of the honeycomb structure, occurs under the effect of shear loading [5, 12].

Alternatively a surface response [30] could have been built in order to predict the milling surface quality. But close milling parameters (such as splindle speed, feed rate, depth of cut) can lead to different results, depending on the material, the quality of the machining tool, etc.

Therefore, in our approach supervised machine learning techniques (with labeled measurements for the model training) are used. These tools need the construction of features associated with the measurements.

3 Milling diagnosis using machine learning techniques

Machine learning techniques can be separated mainly in two categories [17, 23]:

-Unsupervised approaches: based only on input data (data are unlabeled). The goal is to find groups and structures in the data set, in order to classify new observations (measurements) into the different groups.

- Supervised approaches: based on input and output data (Now the data are labelled.

The raw data (measurements) are firstly filtered, with low pass filters in order to eliminate high frequency noises, and labeled (“obtained signals for good surface quality”, “obtained signals for bad surface quality”). Then the features are calculated offline or online.

All the experiments are then split into two groups: 75% for the machine learning model training, 25% for the obtained model evaluation also called test phase in the literature (another percentage can be chosen, for example 60% - 40%, depending on the number of experiments). This can be made randomly, but the ratio “good surface quality” and “bad surface quality” must be kept in each group.

3.1 Features calculation

The features are calculated in the time domain and the frequency domain [6, 13] from the raw signal represented on figure 7, in steady state behaviour. In fact, transient zones (that means when the cutting tool entries or exits the honeycomb core) are not taken into account.

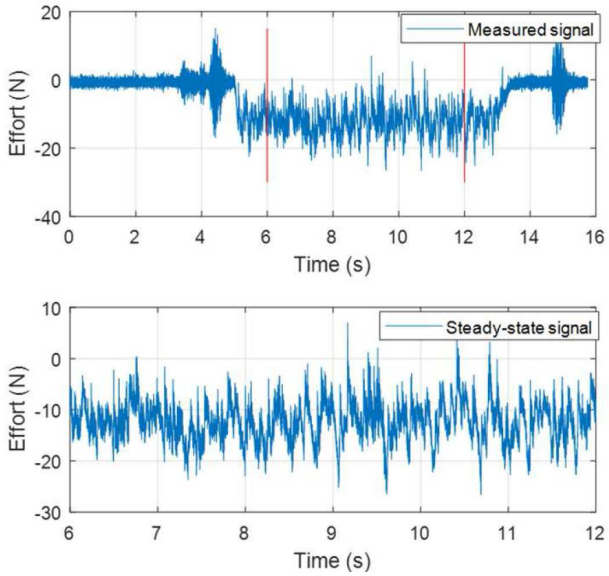


Figure 7: Measured milling force in time domain:
(a) total data plot, (b) signal during steady-state phase

After a first data processing (low pass filtering), firstly 19 features are calculated in the time domain for the measured milling force signal (for example : maximum, minimum, amplitude range, median value, maximum of the absolute value of the signal, interquartile range, average value of the signal, energy of the signal, Skewness, Kurtosis, Shannon entropy, …)

Secondly another 19 features are calculated in frequency domain in a similar way for the measured milling force signal. Therefore, the Fast Fourier transform (FFT) of the signal has been calculated.

All the calculated features (in time and frequency domains) are normalized and stored in a table whose lines and columns respectively represent the experimental number (also called instance) and the associated feature values. The description of the used features are indicated in [32]

The reduction of the features is then be made by using PCA (Principal Component Analysis) [18].

3.2 Labeled data

From the evaluation of the effect of the cutting parameters on surface flatness result, we defined two classes of surface quality applied to the output data of each observation (see table 2)

Label	Flatness (µm)	Qualitative value
'A‘	0 – 600	Best surface quality
‘B'	600 – …	Worst surface quality

Table 2: Label table for the experimental observations

3.3 Applied supervised learning algorithms

In this work, several classification algorithms have been implemented in the Matlab software environment [20, 21]: k-nearest neighbor (kNN), Decision trees (DT), Support Vector Machine (SVM). The different machine learning algorithms (with their adapted tuning parameters) are applied to the normalized labeled training data set (75% of the total experiments). The obtained trained models are then tested on the labeled test data set (25% of the total experiments). The objective is to find again the labels of the test data set: table 3 shows the obtained accuracy result of each algorithm.

Algorithms	Accuracy
KNN	83.4%
KNN k=2	81.3%
Weighted KNN k=2	83.4%
Chebychev KNN k=2	87.5%
Tree	99%
Pruned tree	66.67%
Linear SVM	83.4%
Gaussian SVM	66.67%

Table 3: Prediction error for the normalized data set

We used some news experimental data set in order to evaluate the performance of the trained model. The goal is to predict online (during milling) the surface quality. The results are presented here for the trained model by using the linear SVM classifier algorithm:

	Predicted class		
Actual class	A	B	
A	TP = 83%	FN = 17%	100%
B	FP = 0%	TN = 100%	100%

(TP: true positive rate; FN: false negative rate; FP: false positive rate; TN: true negative rate)

Table 4: Performance of the prediction using SVM classifier

The class B was the best predicted class. Despite the fact that linear SVM algorithm lost in performance for data set with large predictors (i.e. large number of features), it has been an accurate algorithm with a good prediction rate and the lowest training time.

4 Conclusion

The milling's performance is qualified by evaluating the roughness or the flatness of the resulted surface. In this work, different supervised machine learning algorithms have been implemented and compared. To do this, features were firstly calculated from measured milling forces and then each Artificial Intelligence (AI) based model has been trained by the labeled set of features. From the prediction results, SVM algorithm seems to be a good efficient diagnosis algorithm in this application of honeycomb material milling. The developed diagnosis approach can also be applied to the milling of other material.

References

[1] Beskri A., Mejri H., Mehdi K. and Rigal J.F., *Systèmes de surveillance automatique en usinage: Moyens et méthodes* (Automatic monitoring systems in machining: tools and methods), French Mechanics Congress 2013.

[2] IFPM-Formation, *Usinage: Tournage Fraisage* (IFPM courses : Machining: Milling Turning), September 2015

[3] Mikołajczyk T., Nowicki K., Bustillo A. and Pimenov D., *Predicting tool life in turning operations using neural networks and image processing*, Mechanical Systems and Signal Processing, 104: 503-513, 2018.

[4] Pimenov D. Bustillo A. and Mikołajczyk T., *Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth*, Journal of Intelligent Manufacturing, 29(5): 1045-1061, 2018.

[5] Correa M. Bielza C. and Pamies-Teixeira J., *Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process*, International journal of Expert Systems with Applications, 36: 7270-7279, 2009.

[6] A.I.H. Committee, ASM Handbook Volume 16: Machining, ASM International, 1989.

[7] J. Kindinger, Lightweight structural cores, ASM Handbook Met. Composites, 21: 2001.

[8] D. Gay, Matériaux composites (Composite materials), Hermes, 2015.

[9] R. Carl, *Three-dimensional honeycomb core machining apparatus and method*, US Pat. App. 13/707,670. 1, 2012.

[10] Jaafar M., Atlati S., Makich H., Nouari M., Moufki A. and Julliere B., *A 3D FE modeling of machining process of Nomex® honeycomb core: influence of the cell structure behaviour and specific tool geometry*, Procedia CIRP, 58: 505-510, 2017.

[11] Mendoza M. et al., *Development of a new milling cutter for aluminum honeycomb*, Int. J. Mach. Tool Des. Res. 23: 81-91, 1983.

[12] Tchoutouo H., *Adhesiveless honeycomb sandwich structure with carbon graphite prepreg for primary structural application: a comparative study to the use of adhesive film*, master thesis, Wichita State University, May 2012.

[13] Rion J. Leterrier Y. and Manson E., *Prediction of the adhesive fillet size for skin to honeycomb core bonding in ultra-light sandwich structures*, Compos. Part A, 39: 1547-1555, 2008.

[14] Agusmian P. Ooijevaar T. and Kilundu B., *Automated bearing fault diagnostics with cost effective vibration sensor*, WCEAM/VETOMAC, Conference, 2017

[15] Madhusudana CK, Budati S., Gangadhar N., Kumar H. and Narendranath S., *Fault diagnosis studies of face milling cutter using machine learning approach*, Journal of Low Frequency Noise, Vibration and Active Control, 35(2): 128-138, 2016

[16] Gopal S. and Kishore K., *Normalization: A Preprocessing stage*, CSE & IT department, VSSUT, Burla, India 2015.

[17] The MathWorks Inc., *Mastering Machine Learning A Step-by-Step Guide with MATLAB*, Matlab Ebook section 1 to 4, ©2018

[18] Vidal R., Ma Y., Sastry S.S., *Generalized Principal Component Analysis*, Springer 2016.

[19] Yang F., Yun Z., Haiyu Q., Dequn L., Huamin Z. and Jürgen L., *Analysis of feature extracting ability for cutting state monitoring using deep belief networks*, Conference on Modelling of Machining Operations (CIRP), 31: 29-34, 2015.

[20] Chen G. Wei X., Yan R. and Yuqing Z., *Numerical control machine tool fault diagnosis using hybrid stationary subspace analysis and least squares support vector machine with a single sensor*, MDPI conference, 7: 346-358, 2017

[21] Zhang C., Yao X., Zhang J. and Jin H., *Tool Condition Monitoring and Remaining Useful Life Prognostic Based on aWireless Sensor in Dry Milling Operations*, MDPI conference, 16: 795-815, 2016.

[22] Kubat M., *An introduction to machine learning*, Second Edition, Springer, 2017

[23] Shalev-Shwartz S. and Ben-David S., *Understanding Machine Learning: From Theory to Algorithms*, Published by Cambridge University Press, 2014.

[24] Rui Z., Dongzhe W., Ruqiang Y., Kezhi M., Fei S. and Jinjiang Wang, *Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks*, IEEE Transactions on Industrial Electronics, 65(2): p. 1539-1548, February 2018.

[25] Wu D., Jennings C., Terpenny J. and Kumara S., *Cloud-Based Machine Learning for Predictive Analytics: Tool Wear Prediction in Milling*, IEEE International Conference on Big Data, p. 2062-2069, 2016.

[26] Javed K., Gouriveau R., Zerhounic N., and Nectoux P., *Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics*, IEEE Transactions on Industrial Electronics, 62(1): 647-656, January 2015.

[27] Durmus Karayel, *Prediction and control of surface roughness in CNC lathe using artificial neural network*, journal of materials processing technology 209: 3125–3137, 2009.

[28] Azlan Mohd Zain, Habibollah Haron and Safian Sharif, *Prediction of surface roughness in the end milling machining using Artificial Neural Network*, Expert Systems with Applications 37: 1755–1768, 2010.

[29] Codjo L., Jaafar M., Makich H., Knittel D. and Nouari M., *Milling diagnosis using machine learning techniques toward Industry 4.0*, 29th International Workshop on Principles of Diagnosis DX'18, Warsaw, August 2018.

[30] Myers R., Montgomery D., Anderson-Cook C, "Response surface methodology", Wiley 2016

[31] D. Yu. Pimenov, A. Bustillo, T. Mikolajczyk, *Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth*, Journal of Intelligent Manufacturing, June 2018, Volume 29, Issue 5, pp 1045–1061

[32] Codjo L., Jaafar M., Makich H., Knittel D. and Nouari M., *Honeycomb Core Milling Diagnosis using Machine Learning in the Industry 4.0 Framework*, IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Torino Italy, September 2018

Passive control of vibrations

Experimental identification of the corrective effect of a non-circular pulley: application to timing belt drive dynamics

Sébastien PASSOS[1,2], Lionel MANIN[2], Didier REMOND[2], Olivier SAUVAGE[1], Laurent ROTA[1] and Etienne BESNIER[2]

[1]Groupe PSA, Scientific and Future Technologies Department, OpenLab Vibro-Acoustic-Tribology@Lyon, F-78943 Vélizy-Villacoublay, France

[2]Univ Lyon, INSA-Lyon, LaMCoS, CNRS UMR5259, F-69621 Villeurbanne, France

sebastien.passos@mpsa.com
sebastien.passos@insa-lyon.fr

Abstract

The work presented in this paper aims at showing experimentally how an oval pulley can generate a corrective effect on a timing belt drive subjected to a 2nd order periodic excitation (fluctuating load torque). For that purpose, a simplified timing belt drive of a 4-cylinder car engine is reproduced on a test stand. The oval pulley is mounted on the crankshaft axis. Experiments are conducted for different phasing angles of the oval pulley at several driving speeds. Pulley angular vibrations and span tension fluctuations are monitored. The results obtained from these experiments are analyzed in the angular and angular frequency domains. The results are compared with experiments performed on a reference case (equivalent circular crankshaft pulley). The study focuses on the effect of the oval pulley phasing angle on the amplitude of the 2nd order harmonics governing the angular response of the transmission.

1 Introduction

In an automotive engine, the Timing Belt Drive (TBD) is a key component in charge of synchronizing the valve train system and the crankshaft (figure 1). This synchronization constitutes one of the essential functions of the engine. It ensures the correct timing of the valves opening and closing with respect to the piston movement.

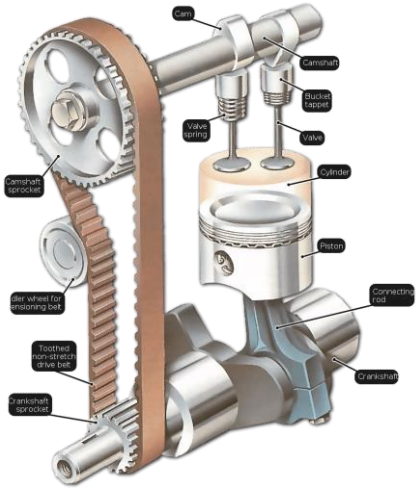


Figure 1: Principle of the valve train system of a car engine with a timing belt drive

In running conditions, TBD are exposed to a very harsh vibratory environment. Various excitation sources such as crankshaft acyclism and fluctuating load torques applied to driven pulleys, generate vibratory

phenomena that may affect the TBD dynamic performances and life. In particular, two phenomena require careful monitoring and need to be minimized:

- Rotational vibrations of the transmission axes.
- Tension force fluctuations in the belt spans.

If mishandled, these two phenomena may induce increased component fatigue, power losses, noise and in the most extreme case can cause desynchronization of the TBD that could result in engine failure (piston-valve clash).

In response to ever-stricter requirements for engine efficiency and reliability, car manufacturers now commonly use TBD comprising innovative pulleys with Non-Circular (NC) pitch profile. When rotating, a NC pulley causes periodic elongations of its adjacent belt spans. Hence, it behaves like an exciter that may generate a corrective rotational excitation able to counteract the other excitation sources acting on the TBD. It is now known that for optimal profile shape and phasing in the transmission, the use of a NC pulley can improve considerably the vibratory performances of a transmission [1,2,3,4]. Nevertheless, determining the optimal design parameters of a non-circular pulley remains hard to accomplish. To achieve this, it is important to clearly understand and identify the impact of such pulleys on the dynamic behaviour of TBD.

The literature on this subject is relatively poor. Most works concern kinematics and quasi-static analyses as reported in [5,6]. In recent papers, Zhu et al. [2] and Passos et al. [3,4] propose numerical models developed to predict the dynamic behaviour of a transmission comprising NC pulleys. Using these models, the authors perform numerical studies that show how an oval crankshaft pulley can significantly reduce the rotational vibrations of a camshaft pulley in a four-cylinder engine (2^{nd} order periodic camshaft load torque). The models are based on a discrete approach (also called 0D/1D approach) similar to that implemented by Hwang et al. [7] for poly-V belt transmissions comprising circular pulleys only. In his model, Hwang considers the belt spans as linear spring-damper elements connected to the pulleys represented by rotational inertias. In [8], Passos et al. conducted an experimental investigation on a basic transmission comprising two pulleys. Two configurations of the basic transmission were studied: one with an oval pulley mounted on the driving axis and the other with an equivalent circular pulley. The experiments were done with no excitation source (constant driving speed and constant load torque applied to the driven pulley). Instantaneous angular speed and acceleration of the driven pulley, transmission error and transmitted torques obtained for the two configurations of the transmission were analyzed and compared. The comparison provided a rigorous description of the proper effect of the oval pulley on the rotational dynamics of the transmission.

The present work aims at completing the studies presented in [3,4,8]. It shows experimentally how an oval pulley can generate a corrective effect able to counteract an excitation source such as a 2^{nd} order periodic (H2) load torque. For that purpose, the oval pulley is mounted on the driving axis of a transmission whose architecture is similar to that of a 4-cylinder car engine timing belt drive. Usually, in this type of transmissions the camshaft pulley is subjected to a fluctuating load torque having a 2^{nd} order periodicity. The test stand is presented in section 2. Experiments are conducted for different phasing angles of the oval pulley and at several driving speeds. Angular vibrations and span tension fluctuations are monitored. The results obtained from these experiments are analyzed and compared with those of a reference case (equivalent circular crankshaft pulley) in section 3.

2 Test stand

2.1 Studied transmission

The studied transmission is represented in figure 3 and its geometrical characteristics are given in table 1. The transmission has been designed in order to reproduce a simplified TBD of a 4-cylinder car engine. It comprises four pulleys: one oval crankshaft pulley (driving pulley), one camshaft pulley and one idler pulley on each side of the transmission. All the pulleys are circular except the crankshaft pulley that has an elliptical pitch profile. Such a pulley has two design parameters: its eccentricity e and its phasing angle ϕ_0 corresponding to the initial orientation angle between the oval pulley major axis and the y-axis (figure 3). In the present study, the eccentricity is imposed (marketed component) and the phasing angle is freely adjustable.

$$\phi_0 = \theta_{CS}(t = 0) \tag{1}$$

Three levels of stationary crankshaft speeds are considered ($\omega_{CS} \in \{600 ; 1500 ; 3000 \text{ rpm}\}$). The camshaft pulley is subjected to a fluctuating load torque C_{CAM} having a 2^{nd} order periodicity with respect to

the crankshaft rotation θ_{CS}. The measured load torque is plotted versus the crankshaft angle for the three different driving speeds in figure 2 (a, c, e). The angular frequency content of the torque is shown in figure 2 (b, d, f). Whatever the driving speed, the torque is governed by a dominant 2nd order harmonic and secondary even harmonics of 4th, 6th and 8th orders. The torque is also modulated by harmonics of 0.5th and 1st orders due to geometry faults and misalignments affecting slightly the measurement. When the crankshaft runs at 3000 rpm, order 9.5 appears due to dynamic effects occurring in the mechanical device that generates the load torque (see section 2.2).

	Coordinates of the pulley centers		**Pitch profiles of the pulleys**
	X (cm)	**Y (cm)**	
Crankshaft	0	0	Ellipse: – Major axis $a = R_{CS} + e$ – Minor axis $b = R_{CS} - e$ with $\begin{cases} R_{CS} = 3.19\ cm \\ e = 0.07\ cm \end{cases}$
Idler (Tight side)	17.3	-2.8	Circle of radius $R_{IT} = 3\ cm$
Camshaft	39.3	0	Circle of radius $R_{CAM} = 2R_{CS}$
Idler (Slack side)	16.9	4.7	Circle of radius $R_{IT} = R_{IS}$

Table 1: Geometrical characteristics of the transmission

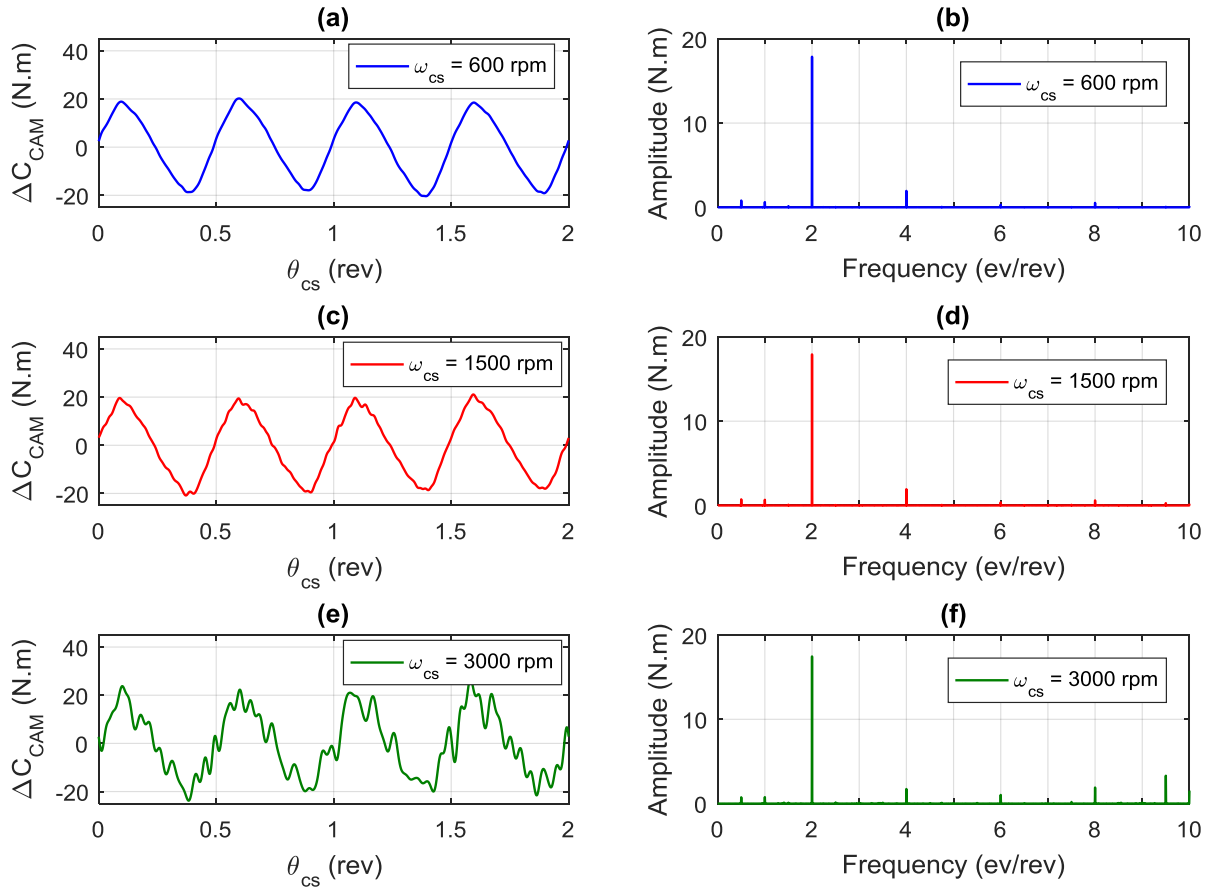


Figure 2: Variations of camshaft load torque in the angular (a, c, e) and angular frequency (b, d, f) domains

2.2 Global architecture

The global architecture of the test stand is illustrated in figures 3 and 4. It comprises two rotary shafts (one driving and one driven shaft) and two supporting parts for idler pulleys. The distances between the pulleys are freely adjustable, enabling a custom setting of the transmission geometry.

The crankshaft pulley is mounted on the driving axis coupled to a speed-controlled electric motor. The driven axis that supports the camshaft pulley of the studied transmission is coupled to the camshaft of a cylinder head fully equipped. It enables applying a real camshaft load torque to the transmission (see figure 2). This load torque has already been discussed in section 2.1.

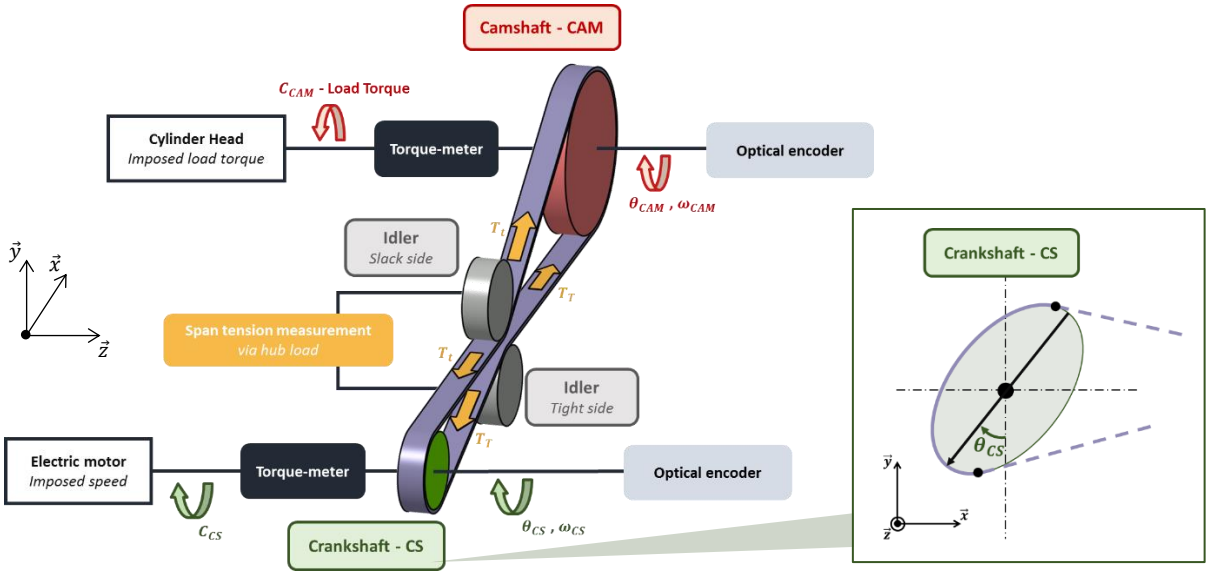
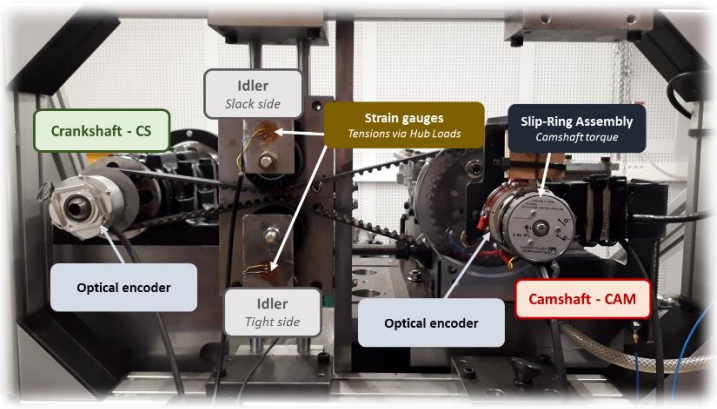
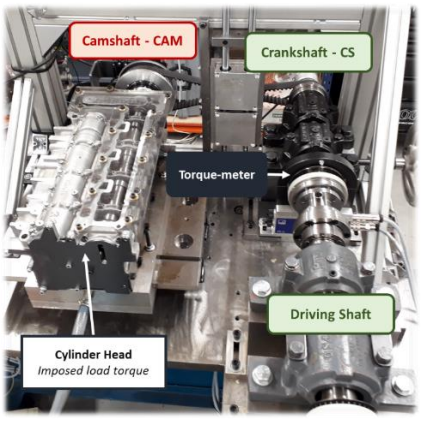


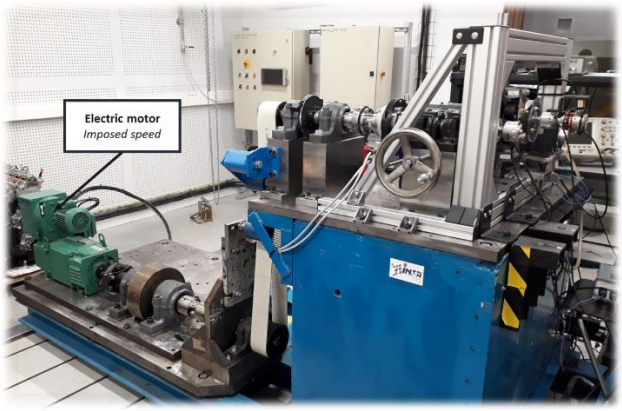
Figure 3: Scheme of the global test stand architecture



(a)



(b)



(c)

Figure 4: Face (a), top (b) and global (c) views of the test stand

2.3 Measurement devices

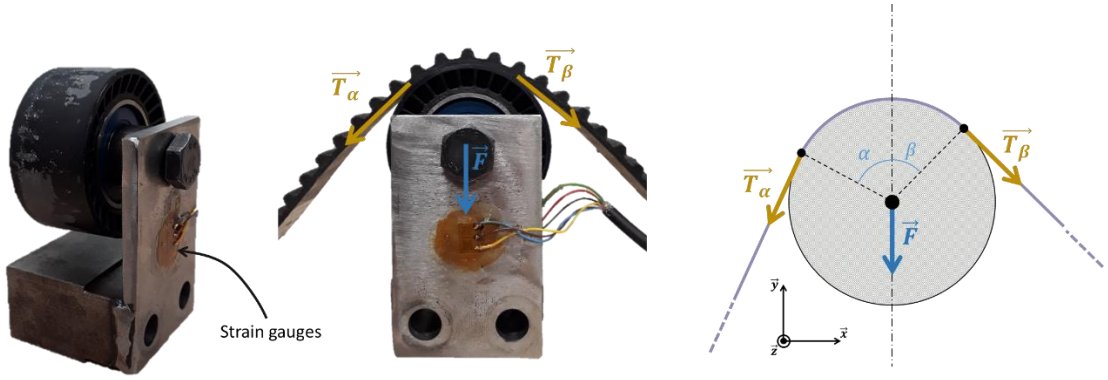
The measurement system is very similar to that used for the works presented in [8]. The system comprises usual devices employed for experimental investigations on gear transmission error and belt transmissions [9,10,11]. It enables measuring all the quantities depicting the rotational dynamics of the transmission.

Optical encoders with a resolution of 2500 pulses per revolution are mounted on the driving and driven axes for measuring their rotation angles and speeds. Torques transmitted by the driving and the driven axes are monitored. On the driving axis, a torque meter with a range of 200 N.m is placed before the driving pulley. On the driven axis, the camshaft is equipped with strain gauges for measuring the resistive load torque. The electrical connection for gauges is provided by means of a slip-ring assembly. The supporting parts used for idler pulleys are equipped with strain gauges so that to measure y-axis component of the hub load applied onto the idler axis. The belt span tension forces can be deduced from the hub load as illustrated in figure 5. On both sides of an idler, belt tensions can be considered equal ($T_\alpha = T_\beta = T$) and then:

$$T = \frac{F}{\sin\alpha + \sin\beta} \tag{2}$$

where F is the y-axis component of the hub load, T is the tension force in both belt spans and α and β are the respective orientation angles of the spans.

The data acquisition system is custom made within a N. I. PXI frame including counter boards for the optical encoders (pulse timing method [9]) and classical data acquisition boards for the other sensors.



(a) (b)

Figure 5: Idler supporting part used to measure the span tensions (a) and principle diagram (b)

3 Experimental analysis

The result analyses are performed by observing the following quantities in the angular and the angular frequency domains:

- Instantaneous angular speed variations for the camshaft pulley ($\Delta\omega_{CAM}$).
- Tension force variations in the tight and the slack sides (ΔT_T and ΔT_t)

These quantities are first studied for a reference case for which the crankshaft pulley is circular (section 3.1). Then, the transmission equipped with the oval crankshaft pulley is considered (section 3.2). A comparison study is conducted showing how the oval pulley affects the levels of angular vibrations and tension force fluctuations depending on its phasing angle and the driving speed.

3.1 Reference case: circular crankshaft pulley

The fluctuations and the frequency content of the camshaft angular speed are shown in figure 6. For all driving speeds, the variations are periodic and dominated by a 2nd order harmonic due to the dominant 2nd order harmonic governing the load torque. The amplitude of the H2 harmonic fluctuation is about 4.5 rpm when the driving speed is 600 rpm and rises to approximately 15 rpm when the crankshaft speed reaches 1500 and 3000 rpm. One can see also that the frequency spectrum contains secondary harmonics mainly due to the other camshaft torque harmonics (even orders and 9.5th order for ω_s = 3000 rpm) but also to some disturbing excitations such as resonances of the test stand and 0.5th and 1st order modulations due to misalignment and geometry faults on the crankshaft and camshaft pulleys.

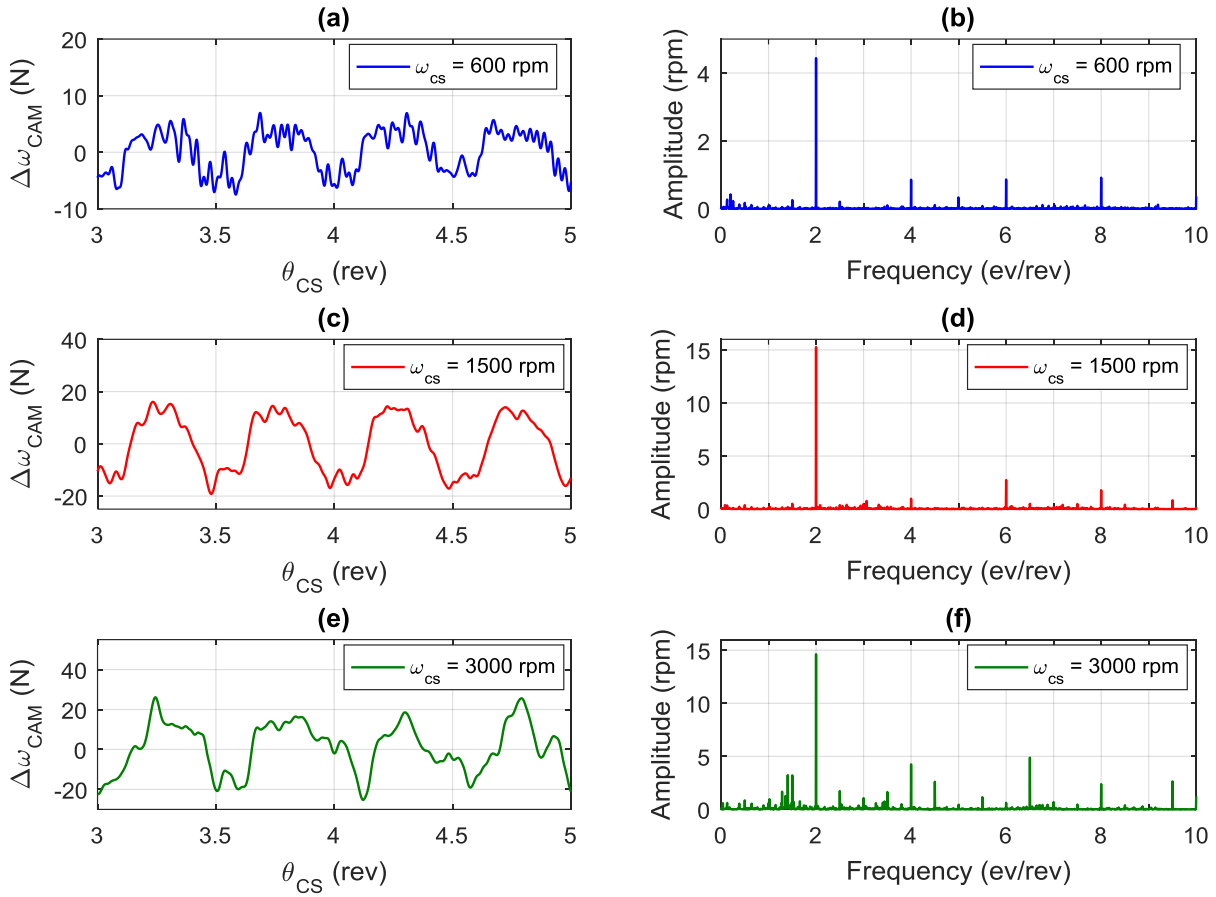


Figure 6: Camshaft angular speed variations in the angular (a, c, e) and angular frequency (b, d, f) domains

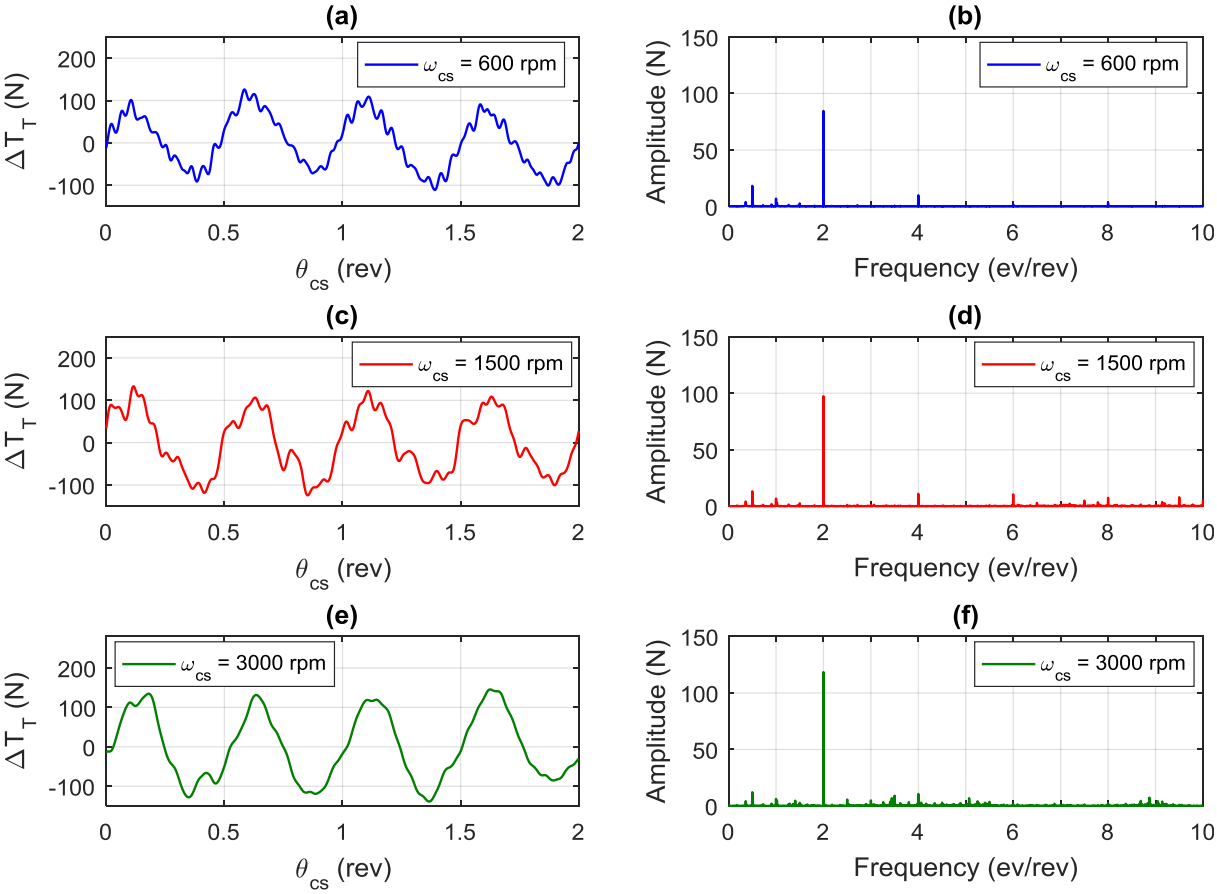


Figure 7: Tight span tension variations in the angular (a, c, e) and angular frequency (b, d, f) domains

Figure 7 and 8 show the variations and the frequency content of the belt tensions in the tight and slack spans of the transmission. The amplitude of the dominant H2 harmonic slightly increases with the driving speed. The H2 amplitude is of course always higher in the tight span than in the slack span. Moreover, one can note that the tension fluctuations in the tight and slack spans are out-of-phase, which is another common result for belt transmission subjected to fluctuating load torque.

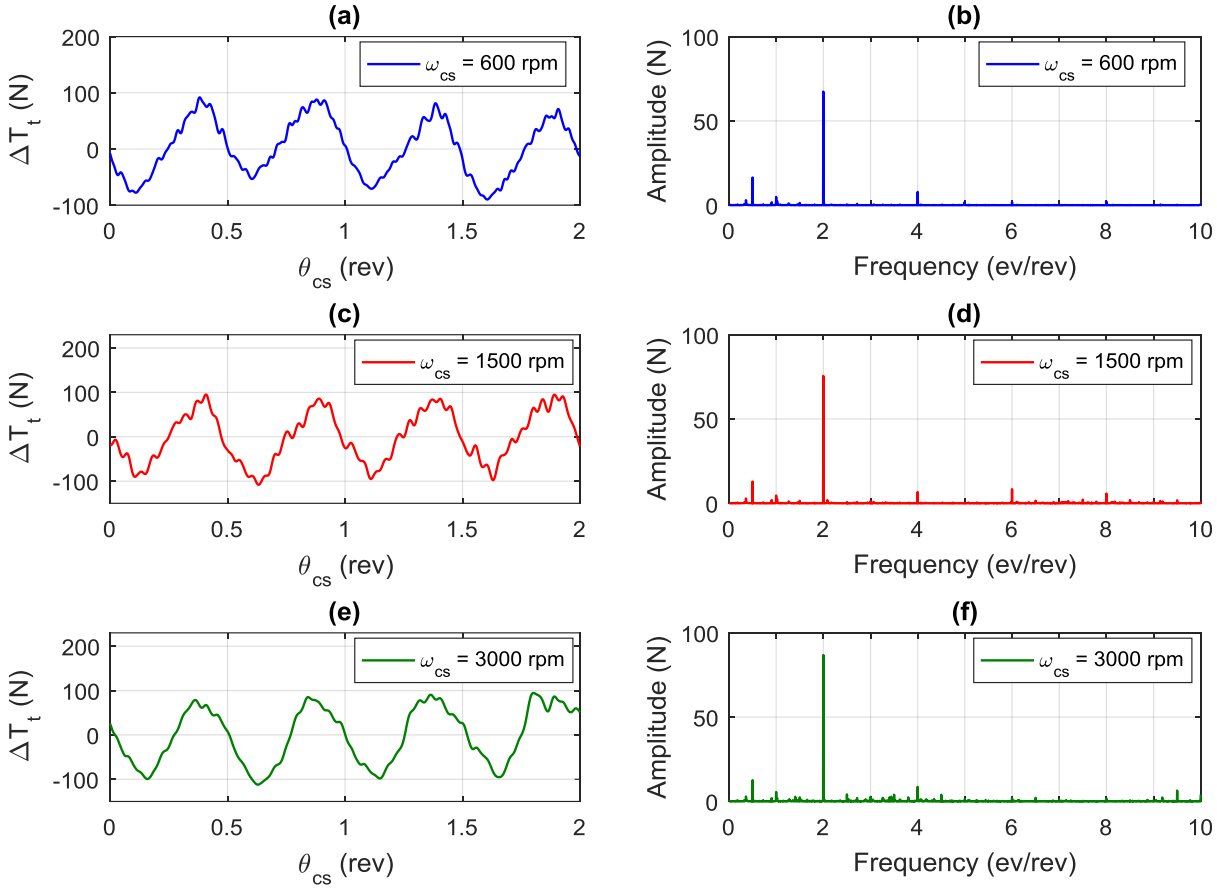


Figure 8: Slack span tension variations in the angular (a, c, e) and angular frequency (b, d, f) domains

3.2 Oval crankshaft pulley: study of the corrective effect

Results obtained for the reference case (section 3.1) have shown that the variations of the camshaft speed and the belt span tensions caused by the camshaft load torque are strongly dominated by a 2nd order harmonic (H2) for all the driving speeds. Thus, in the following, the analyses focus only on the amplitude of H2 harmonics. A parametric study has been performed in order to show how the oval pulley phasing angle impacts the H2 amplitude. Considering that an oval pulley is symmetric along its major axis, the experiments are only performed for a phasing angle ranging between 0 and 180°.

Figures 9, 10 and 11 show respectively the results obtained for the three driving speeds 600 ; 1500 and 3000 rpm. Each figure comprises three graphs on which the amplitudes of the H2 harmonics that govern the camshaft speed and tension forces in the tight and slack spans are respectively plotted versus the phasing angle of the crankshaft oval pulley. On these graphs, the H2 amplitude evolution with the phasing angle is represented with a red solid line marked with circles. Horizontal blue solid lines represent the amplitude of the H2 harmonics in the reference case (circular crankshaft pulley). Green and red hatched areas correspond to phasing angle ranges for which the oval pulley respectively involves a reduction or an increase of the H2 harmonic.

When the crankshaft runs at 600 rpm (figure 9), a speed fluctuation reduction can be obtained for a phasing angle range [0,40°]∪[150,180°] with a maximum reduction of nearly 82% for a phasing angle around 0°. In contrast, for a phasing angle in the range]40,150°[the amplitude is higher than in the reference case with a critical increase of 75% for a phasing angle equal to 90°. H2 amplitude of the tension forces in tight and slack sides are significantly reduced for non-overlapping phasing angle ranges, respectively [97,177°] and [0,82°]. Amplitudes of tension force variations are less impacted by the phasing angle of the oval pulley. For both spans, the maximum reduction and increase ratios are the same, respectively about 24 and 26%.

When the crankshaft runs at 1500 rpm (figure 10), the phasing angle ranges corresponding to an increase and a reduction of fluctuations are quite the same as for a driving speed equal to 600 rpm. The reduction and increase ratios are of the same order of magnitude and are obtained for very similar optimal and critical phasing angle values.

For a driving speed of 3000 rpm (figure 11), the angle ranges and the corresponding reduction and increase ratios differ. The speed variations are reduced when the phasing angle belongs to the range $[0,30°]\cup[170,180°]$. The optimal and critical phasing angles are respectively 10° and 95° with a maximum reduction and increase ratios of 60% and 166% respectively. The tension force variations in the tight span are reduced for a phasing angle in the range $[0,22°]\cup[137°,180°]$. The maximum reduction is about 50% when the phasing angle is around 10° and there is a maximum increase of nearly 67% for an angle of 95°. The tension force fluctuations induced in the slack span are reduced for a phasing angle comprised in the range $[0,45°]\cup[158,180°]$ with a minimum amplitude (-30%) and a maximum amplitude (+39%) for phasing angles of 10° and 105° respectively. It can be seen that if the phasing angle belongs to the range $[0,22°]\cup[170°,180°]$, the fluctuations are reduced for the camshaft speed and the tension forces in both spans simultaneously. In particular, for a phasing angle equal to 10° all the fluctuations are significantly reduced.

In the light of these results, one can finally note that it is possible to reach a compromise with a phasing angle around 0°. Whatever the driving speed, such a value ensures a strong reduction of the camshaft speed variations and limits the tensions variations in both spans.

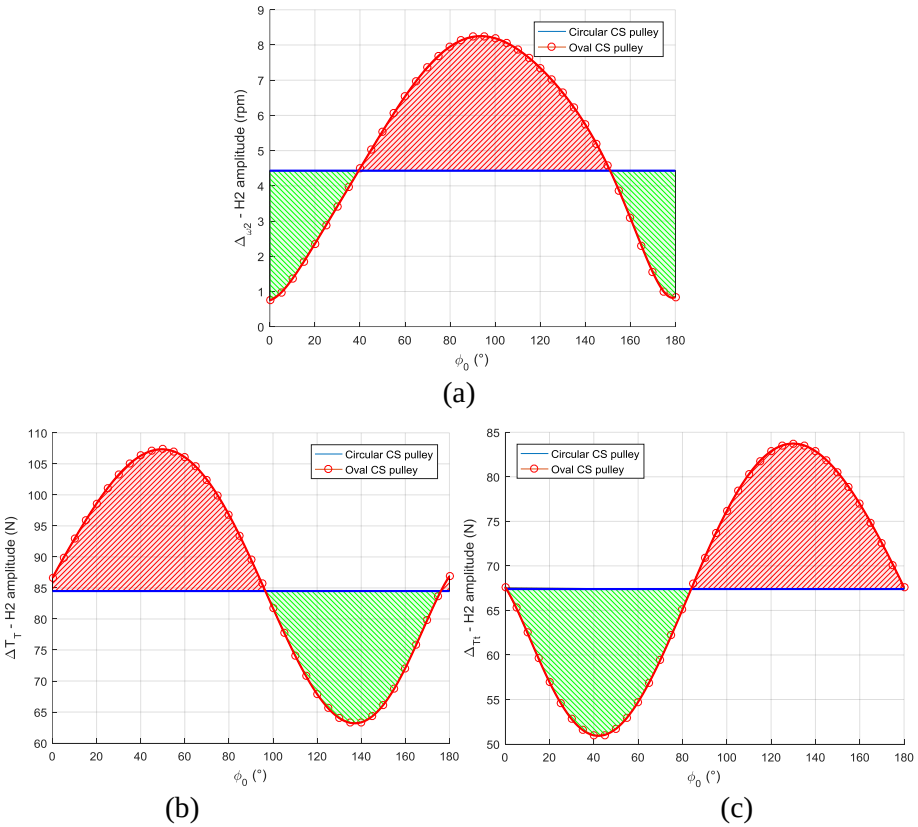
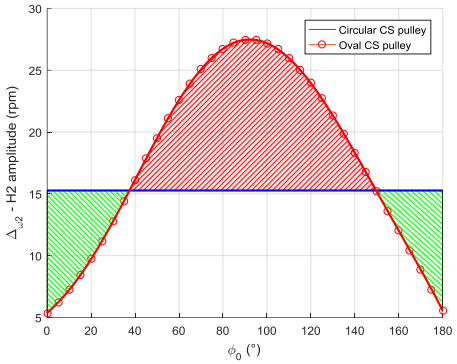
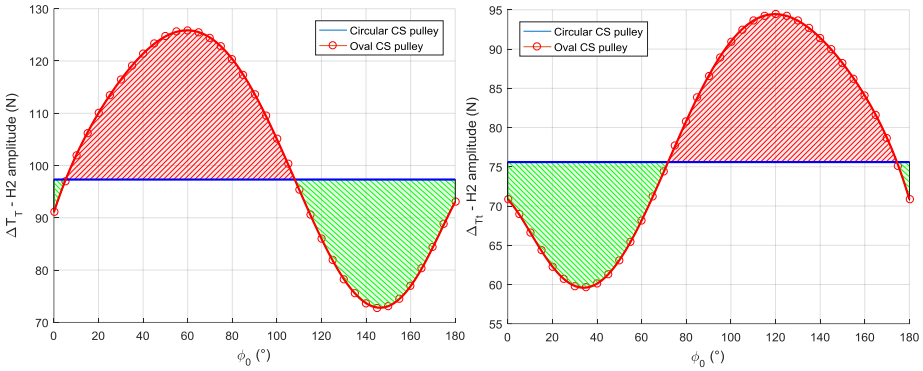


Figure 9: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at 600 rpm

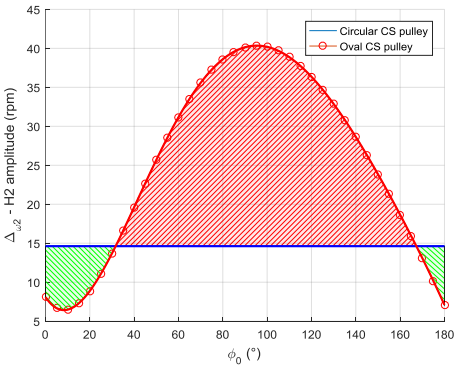


(a)

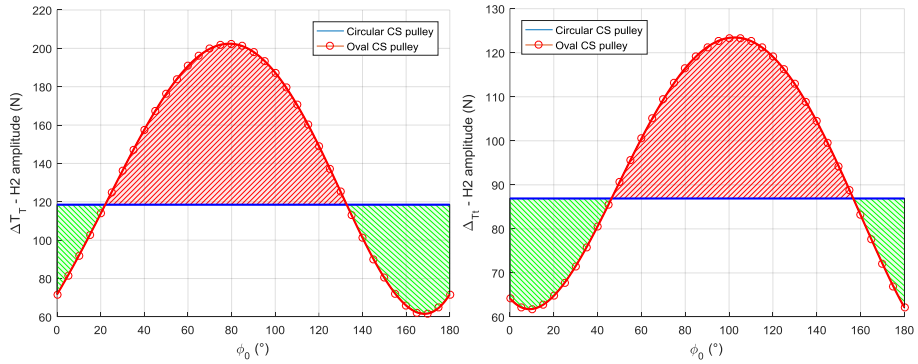


(b) (c)

Figure 10: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at 1500 rpm



(a)



(b) (c)

Figure 11: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at 3000 rpm

4 Conclusion

The experimental results presented in this paper demonstrate to what extent, for a well-chosen phasing angle, an oval pulley has a corrective effect enabling a reduction of the H2 fluctuations affecting the camshaft angular speed and the span tension forces. These results also show how much, for wrong phasing angles, the pulley can degrade the dynamic behaviour of the transmission inducing a strong increase of speed and tension fluctuations. This is why transmissions comprising NC pulleys must be designed carefully.

In addition, one can note that the impact of an oval pulley depends on the driving speed, which make the design of the transmissions more difficult. As already discussed in previous works [3,4], it is not so easy to reduce significantly and simultaneously speed and tension force fluctuations in belt spans using a NC pulley only.

For completing the present work, it could be meaningful to extend the researches to more complex systems. Thus, future works could involve transmissions subjected to other excitation sources such as acyclism and/or transmissions comprising a dynamic tensioner. It could be interesting as well to study the case of a NC pulley having a different profile shape adapted to the correction of harmonics of other order. Also, the effect of NC pulleys on TBD acoustic radiation still remains to investigate (impact on belt meshing noise, span transverse vibrations, …).

References

[1] GAJEWSKI, W., *Synchronous drive apparatus and methods*, U.S. Patent No 7.044.875, Litens Automotive, 2006.

[2] ZHU, H., HU, Y., ZHU, W. D., *Dynamic response of a front end accessory drive system and parameter optimization for vibration reduction via genetic algorithm*, Journal of Vibration and Control, Volume 24, Issue 11, pp. 2201-2220, 2018.

[3] PASSOS, S., MANIN, L., SAUVAGE, O., ROTA, L., DELATTRE, B., REMOND, D., *Investigation on the timing belt drive with a non-circular pulley*, Conference Proceedings - International Gear Conference, Lyon – Villeurbanne, France, Volume 2, pp. 1186-1198, 2018.

[4] PASSOS, S., MANIN, L., SAUVAGE, O., ROTA, L., REMOND, D., *Rotational dynamics of timing belt drives comprising non-circular pulleys*, SIA International Conference and Exhibition on Simulation, Paris – Saclay, France, 2019.

[5] ZHENG, E., JIA, F., SHA, H., WANG, S., *Non-circular belt transmission design of mechanical press*, Mechanism and Machine Theory, Volume 57, pp. 126-138, 2012.

[6] ENDO, G., YAMADA, H., YAJIMA, A., OGATA, M., HIROSE, S., *A passive weight compensation mechanism with a non-circular pulley and a spring*, IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, pp. 3843-3848, 2010.

[7] HWANG, S. J., PERKINS, N. C., ULSOY, A. G., MECKSTROTH, R. J., *Rotational response and slip prediction of serpentine belt drive systems*, J. Vib. Acoust. , Volume 116, Issue 1, pp. 71-78, 1994.

[8] PASSOS S., MANIN, L., REMOND, D, SAUVAGE, O., ROTA, L., BESNIER, E., *Experimental investigation on timing belt drives dynamics with non-circular pulley*, ISMA International Conference on Noise and Vibration, Leuven, Belgium, 2018.

[9] REMOND, D., *Practical performances of high-speed measurement of gear transmission error or torsional vibrations with optical encoders*, Measurement Science and Technology, IOP Publishing, pp. 347-353, 1998.

[10] MAKITA, K., KAGOTANI, M., UEDA, H., KOYAMA, T., *Influence of idler on error transmission in synchronous belt drives (under transmission force)*, ASME J. Mech. Des., Volume 125, Issue 2, pp. 404-410, 2003.

[11] MANIN, L., MICHON, G., REMOND, D., DUFOUR, R., *From transmission error measurement to pulley-belt slip determination in serpentine belt drives: influence of tensioner and belt characteristics*, Mechanism and Machine Theory, Volume 44, Issue 4, pp. 813-821, 2009.

Robust optimization of nonlinear energy sinks used for dynamic instabilities mitigation of an uncertain friction system

Cherif SNOUN, Baptiste BERGEOT, Sébastien BERGER

INSA Centre Val de Loire, Université d'Orléans,Université de Tours, LaMé EA 7494,
3 Rue de la Chocolaterie, CS 23410, 41034 Blois Cedex, France

Abstract

In this paper, robust optimization tool is proposed for nonlinear energy sinks used for the mitigation of friction-induced vibrations due to mode coupling instability in braking systems. The study is based on a mechanical system which is composed of two NESs coupled to the well-known two-degrees-of-freedom Hultèn's model. In such an unstable system coupled with NES, it is usual to observe a discontinuity in the steady-state amplitude profiles which separates the parameters space into two parts which contain mitigated and unmitigated regimes respectively. We developed a methodology based on Multi-Element generalized Polynomial Chaos to identify this discontinuity which allows us to determine the Propensity of the system to undergo a Harmless Steady-State Regime (PHSSR). The objective of this work is, therefore, to maximize the value of PHSSR to obtain an optimal design of the NESs. For that, several stochastic optimization problems are presented taking into account the dispersion of the uncertain parameters.

Keywords: Brake squeal noise, Friction-induced vibrations, Nonlinear energy Sink, Uncertainties, Robust optimization

1 Introduction

Dynamic instability also named friction-induced vibrations is one of the important problem that the friction system can confront. This phenomenon can appear by the generation of Limit Cycle Oscillation (LCO) induced by dry friction. They are explained, in most studies of self-excited systems, by the coupling of the tangential and normal modes [1, 2, 3]. As a way to model these dynamic instabilities related to friction, the well-known two-degrees-of-freedom Hultèn's model [4, 5] has been widely used and it is also considered in this paper.

The NES is a nonlinear spring mass damper with a strong cubic stiffness. It can adapt itself to the main system that it is attached to without being tuned to a specific frequency. Its operation is based on the concept of Targeted Energy Transfer (TET) which has become an important passive control technique for reducing or eliminating unwanted vibrations [6, 7].

In this paper, two ungrounded NESs are used in order to mitigate or eliminate mode coupling instability in the Hultèn's model. In [8], the authors classified the steady-state response regimes in two main regimes related to the dispersion of some uncertain parameters: the first are the mitigated regimes and the second are the unmitigated regimes. As usual in the context instability mitigation by means of NESs, the LCO amplitude profile presents a discontinuity between these regimes which makes the NES potentially very sensitive to uncertainty. For that, with a low computational cost, the aim is to maximize the performance of the NES (i.e. minimize the region parameter of unmitigated regimes), by optimizing its design parameters taking into account uncertain parameters.

NES has been extensively optimized in a deterministic context (see e.g. [9]), but there is very little work performing optimization of NES under uncertainties in general [10] and for mitigation of self-excited vibrations in particular [11]. In the latter, the authors have developed a methodology of optimization under uncertainties of a NES attached to a two-degree-of-freedom airfoil. In the present work, a robust optimization tool based on the Multi-Element generalized Polynomial Chaos (ME-gPC) is developed.

The article is constructed as follows: in Section 2, the two degrees-of-freedom Hultèn's model coupled to two NESs is presented. In Section 3, the steady-state of the system as well as the discontinued profile are introduced. The optimization formulation under uncertainties is formulated in Section 4. Section 5 describes

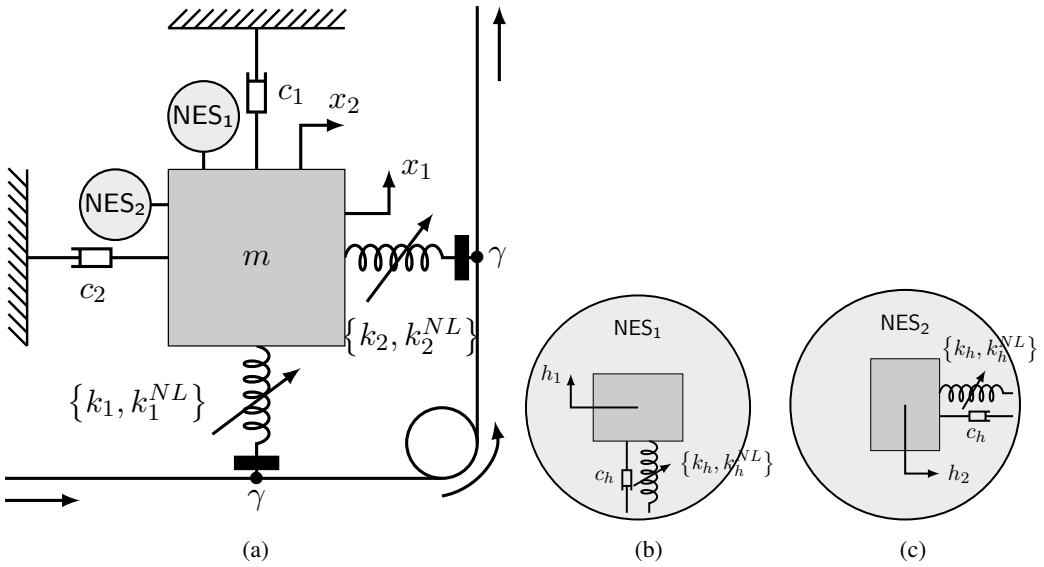


Figure 1: (a) Mechanical system with NESs. (b) Zoom on the NES_1. (c) Zoom on the NES_2.

the polynomial chaos theory and the used stochastic optimization algorithm. The results of the optimization methods are presented in Section 6. Finally, conclusions are given in Section 7.

2 The mechanical system

The mechanical system considered in this work is composed by the two degrees-of-freedom (DOF) Hultèn's model [4, 5], which represents the primary system, coupled to two identical NESs with masses m_h, damping coefficients c_h and cubic stiffnesses k_h^{NL}. The NESs are attached on the primary system in an ungrounded configuration as shown in Fig. 1.

The equations which describe the mechanical system are given by :

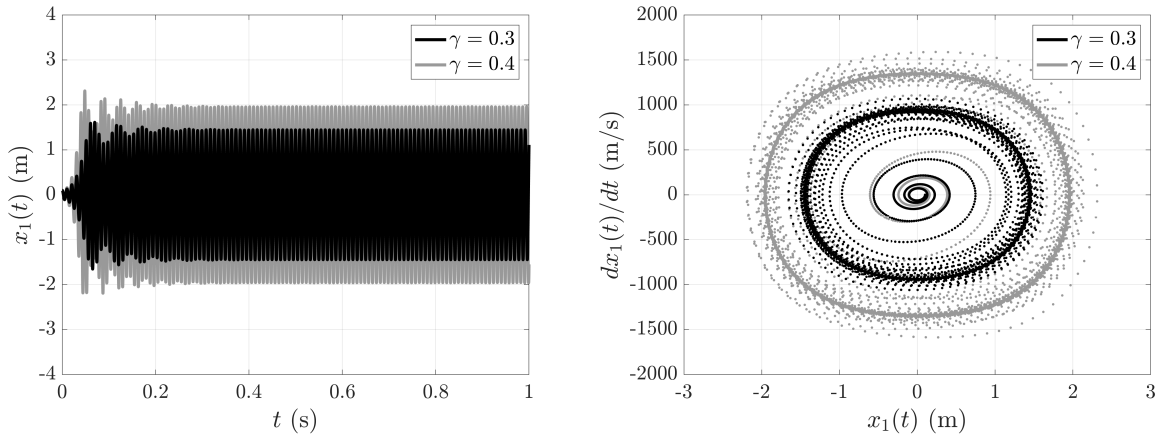
$$\frac{d^2x_1}{dt^2}+\eta_1\omega_1\frac{dx_1}{dt}+\omega_1^2x_1-\gamma\omega_2^2x_2+\varphi_1x_1^3-\gamma\varphi_2x_2^3+ \mu\omega_1\left(\frac{dx_1}{dt}-\frac{dh_1}{dt}\right)+\xi_h(x_1-h_1)+\varphi_h(x_1-h_1)^3=0 \tag{1a}$$

$$\varepsilon\frac{d^2h_1}{dt^2}+\mu\omega_1\left(\frac{dh_1}{dt}-\frac{dx_1}{dt}\right)+\xi_h(h_1-x_1)+\varphi_h(h_1-x_1)^3=0 \tag{1b}$$

$$\frac{d^2x_2}{dt^2}+\eta_2\omega_2\frac{dx_2}{dt}+\omega_2^2x_2+\gamma\omega_1^2x_1+\gamma\varphi_1x_1^3+\varphi_2x_2^3+ \mu\omega_1\left(\frac{dx_2}{dt}-\frac{dh_2}{dt}\right)+\xi_h(x_2-h_2)^+\varphi_h(x_2-h_2)^3=0 \tag{1c}$$

$$\varepsilon\frac{d^2h_2}{dt^2}+\mu\omega_1\left(\frac{dh_2}{dt}-\frac{dx_2}{dt}\right)+\xi_h(h_2-x_2)+\varphi_h(h_2-x_2)^3=0, \tag{1d}$$

where $h_1(t)$ and $h_2(t)$ (respectively $x_1(t)$ and $x_2(t)$) represent the NESs displacements (respectively the displacements of the primary system), $\eta_i=c_i/\sqrt{mk_i}$, $\omega_i=\sqrt{k_i/m}$, $\varphi_i=k_i^{NL}/m$ (with $i=1,2$), $\varepsilon=m_h/m$ assuming $0.01<\varepsilon<0.1$, $\xi_h=k_h/m$, $\mu=c_h/\sqrt{mk_1}$ and $\varphi_h=k_h^{NL}/m$.



(a) (b)

Figure 2: Numerical integration in the direction x_1 of the Hult'en's model without NESs for two values of the friction coefficient: (a) Times series; (b) Representation in the phase plane.

3 Preliminary results

3.1 Vibratory levels and possible steady-state regimes

A example of the numerical simulation of the system without NESs is plotted in Fig. 2 for two different values of the friction coefficient: $\gamma = 0.3$ and $\gamma = 0.4$. The other parameters are

$$\begin{array}{cccc} & \omega_1 = 2\pi 100, & \omega_2 = 2\pi 85, & \\ \eta_1 = 0.02, & \eta_2 = 0.06, & \varphi_1 = 10^5, & \varphi_2 = 0, \\ \varepsilon = 0.05, & \xi_h = 0.001, & \mu = 0.02, & \varphi_h = 1.4 \cdot 10^5. \end{array} \tag{2}$$

Moreover, the initial conditions are small perturbation of the trivial equilibrium position: $x_1(0) = x_2(0) = 0$, $\dot{x}_1(0) = \dot{x}_2(0) = 10$. When $\gamma = 0.4$, Limit Cycle Oscillations (LCO) are observed. We focus our analysis on the capacity of the NESs to suppress or mitigate these LCOs.

Four main types of steady-state regimes can be generated when two NESs is attached on the: complete suppression of the instability, mitigation through Periodic Response (PR), mitigation through Strongly Modulated Response (SMR) or no mitigation.

An illustration of these regimes is shown in Fig. 3 which plots the displacements $x_1(t)$ with respect to times with and without the NESs. Hereafter, mitigated regimes referred to complete suppression, PR and SMR.

3.2 The objective function

We define the amplitude A_1^{wNES} of the variables x_1 of the coupled system Eq. (1) and within a steady-state regime as

$$A_1^{wNES} = \frac{\max\left[x_1^{SSR}(t)\right] - \min\left[x_1^{SSR}(t)\right]}{2}, \tag{3}$$

where $x_1^{SSR}(t)$ is the times series of the variables x_1 obtained from the numerical integration of the coupled system Eq. (1) within the steady-state regime[1].

The amplitudes A_1^{wNES} and A_1^{woNES} (i.e. the amplitude of the system without NESs) are plotted as a function of the friction coefficient γ in Fig. 4 for the set of parameters Eq. (2).

The figure highlights a jump (or discontinuity) in the amplitude profile A_1^{wNES}. This discontinuity corresponds, when γ increases, to the transition from SMR to no suppression regimes and separates mitigated regimes and unmitigated regimes.The value of γ at the jump is called mitigation limit (with respect to γ) and denoted γ_{ml}.

[1]The calculation was performed on 4 seconds in order to be sure that the steady-state has been reached.

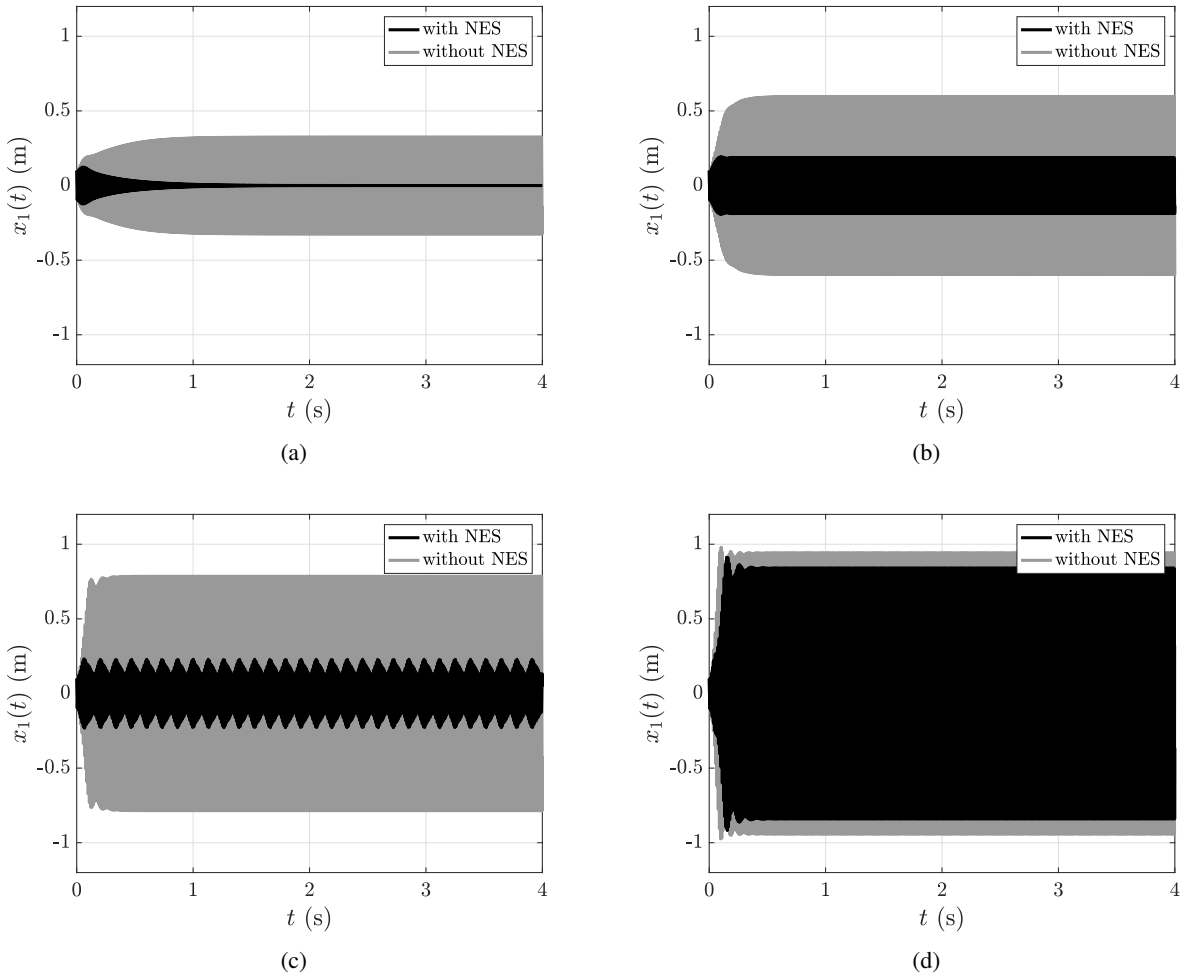


Figure 3: Comparison between time serie $x_1(t)$ resulting from the numerical integration of the braking system with and without NESs. (a) Complete suppression, $\gamma = 0.16$; (b) Mitigation: PR, $\gamma = 0.18$; (c) Mitigation: SMR, $\gamma = 0.2$; (d) No suppression, $\gamma = 0.22$. The set of parameters Eq. (2) is used.

Then the Propensity of the system to undergo a Harmless Steady-State Regime (PHSSR) of the oscillation by the NESs is estimated. We compute a set of S_{total} samples of the uncertain parameters within the considered uncertain space and following their distribution law. Then, the PHSSR is defined as follows

$$\text{PHSSR} = \frac{S_{\text{HSSR}}}{S_{\text{total}}} \times 100 \tag{4}$$

where S_{HSSR} is the number of samples within the region of uncertain parameters space in which the LCO of the system is mitigated or the system is stable.

The PHSSR represents the objective function of the study.

4 Optimization Formulation under uncertainties

The sensitivity of the friction coefficient is such that the steady state of the mechanical system is discontinuous and presents a jump. This jump induces areas in which the efficiency of the NES is either high or low. In order to obtain a robust design of NES which is insensitive to the dispersion of the uncertain parameters, a stochastic optimization problem is considered as follow:

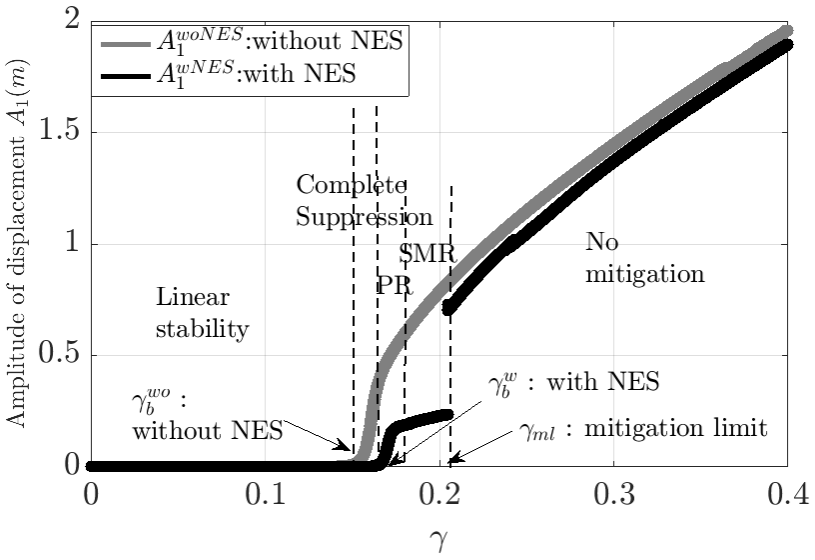


Figure 4: Amplitudes A_1^{wNES} and A_1^{woNES} as a function of the friction coefficient γ.

$$\begin{aligned} &\text{Maximize} && PHSSR\left(A_1^{wNES}(x_d,x_u)\right) \\ &\text{Subject to} && x_d^{(min)} \leqslant x_d \leqslant x_d^{(max)} \end{aligned} \tag{5a}$$

(5b)

where

- x_d represents the design variables of the NES with lower bounds $x_d^{(min)}$ and upper bounds $x_d^{(max)}$.
- x_u represent the uncertain parameters of the primary system.

5 Optimization Algorithm

In order to solve an optimization problem under uncertainties, the determination of the LCO of the mechanical system need to use a stochastic method to take into account the uncertain parameters. In this work, the Multi-Element generalized Polynomial Chaos is used as an alternative approach to reduce the computational cost of the traditional methods (Monte Carlo or deterministic simulations).

5.1 Polynomial Chaos theory

5.1.1 generalized Polynomial Chaos

The gPC theory [12, 13] allows to express a random process $X(\xi)$ called also the Quantity of Interest (QoI) with a truncated orthogonal polynomial function series such as

$$X(\xi) \approx \sum_{j=0}^{N_p} \bar{x}_j \phi_j(\xi), \tag{6}$$

where:

- $\bar{x}_j$ are the gPC coefficients of the stochastic process $X(\xi)$,
- $\phi_j(\xi)$ are orthogonal polynomial functions,
- $N_p = \frac{(p+r)!}{p!r!} - 1$ where p is the order of the gPC and r is the number of the uncertain parameters,
- $\xi(\xi_1,...,\xi_r)$ is a vector of r independent random variables within$[-1,1]^r$.

5.1.2 Multi-Element Generalized Polynomial Chaos (ME-gPC)

The ME-gPC consists to split ξ into a collection of m non-intersecting elements and to approximate the stochastic process $X(\xi)$ using the gPC in each element [14]. It is given by

$$X(\xi) \approx \sum_{k=1}^{m} X_k(\bar{\xi}^k) J_k, \tag{7}$$

where

- $X_k(\bar{\xi}^k) = \sum_{j=1}^{N_p} \bar{x}_{k,j} \phi_j(\bar{\xi}^k)$ where $\bar{\xi}^k$ is an uniform random variables corresponding to the k^{th} element,
- J_k is the element size.

We denote by $\sigma^2_{p,k}$ the variance of the QoI estimated in the k^{th} element directly from the gPC coefficients as:

$$\sigma^2_{p,k} = \frac{1}{2^r} \sum_{j=1}^{N_p} \bar{x}^2_{k,j} \langle \phi^2_j \rangle. \tag{8}$$

5.2 Proposed algorithm

The ME-gPC metamodel would not be efficient in its initial configuration due to the presence of discontinuities. In fact, the goal is not to obtain an accurate representation of the response of the system, but to be able to locate the discontinuity. Figure 5 shows the different steps of the ME-gPC based method to detect the jump. The method consists to split the stochastic parameter space into two equal non-intersecting elements and to evaluate in each element the amplitude of displacement of the system. Then, the verification of the presence or not of discontinuity is addressed according to the comparison of the variance $\sigma^2_{p,k}$ to a chosen threshold θ_1. The process stops either if the element size reaches a threshold

$$J_{kmin} = \theta_2 \, J_{k0}, \tag{9}$$

where J_{k0} represents the size of the initial element or if the number N' of numerical simulations required with the ME-gPC based method is smaller than the typical number N of simulations required with the reference method. The mitigation limit is, therefore approximated by the upper bound of the last found element and then the robust evaluation of the PHSSR is calculated.

6 Results

In this work, the friction coefficient γ is considred as the uncertain parameter (x_u) and the damping rate of the NES μ as the design variables (x_d).

6.1 Reference study

In this section, the reference is evaluated with the deterministic model (1). In the first case, 50 deterministic samples of the damping rate of the NES ($\mu \in [0, 0.04]$) are used. For each sample of μ, 1 000 samples of the friction coefficient ($\gamma \in [0, 0.4]$) are used in order to determine the mitigation limit $\gamma_m l$. The number of simulations is therefore equal to 50 000 ($50 \times 1\,000$). Fig. 6(a) shows the variation of the mitigation limit as a function of μ. Because of the presence of a plateau, we realize that the number of samples of γ is not sufficient to determine the optimal design value of μ. For that, 10 added deterministic samples of the damping rate of the NES ($\mu \in [0.02, 0.03]$) with now 10 000 samples of the friction coefficient γ to determine the mitigation limit. The total number of simulations required is now equal to 150 000 ($50\,000 + 10 \times 10\,000$). Fig. 6(b) shows the new variation of the mitigation limit as a function of μ. In this case, the maximum (i.e. optimal) value of γ_{ml} is 0.2004 which gives a maximum value of PHSSR equals to 50.1%. The corresponding optimal design value for the damping rate is $\mu^{\text{Ref}}_{\text{opt}} = 0.024$.

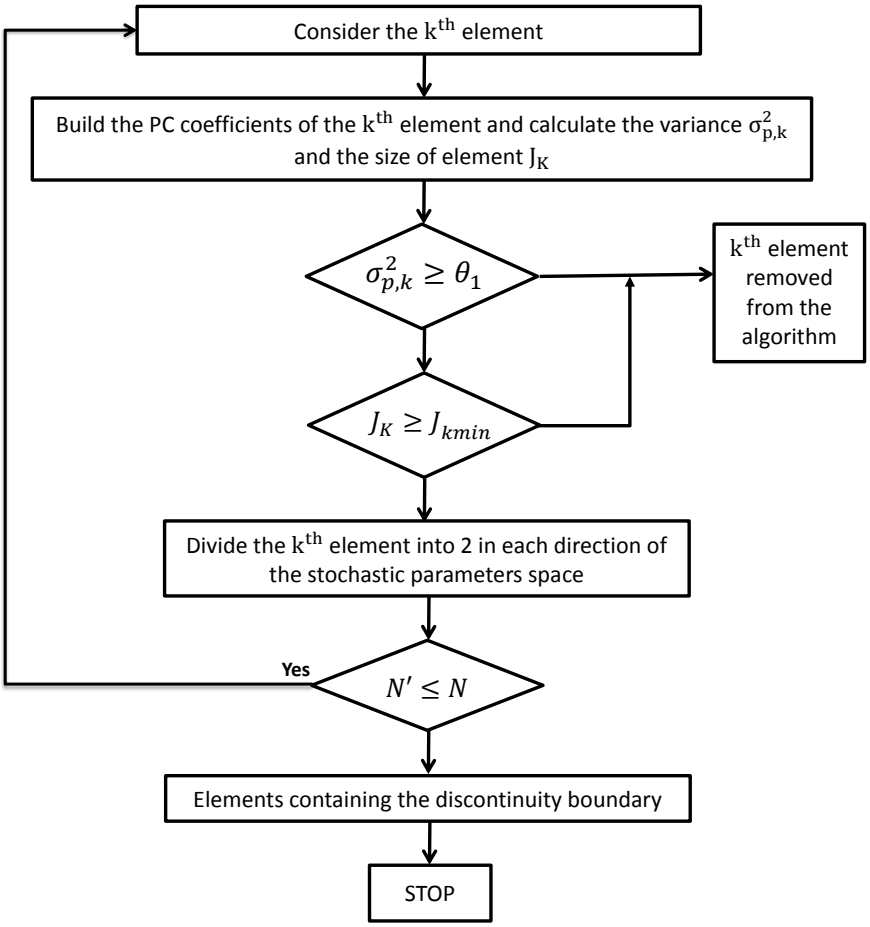


Figure 5: Algorithm of the ME-gPC based method to identify the mitigation limit.

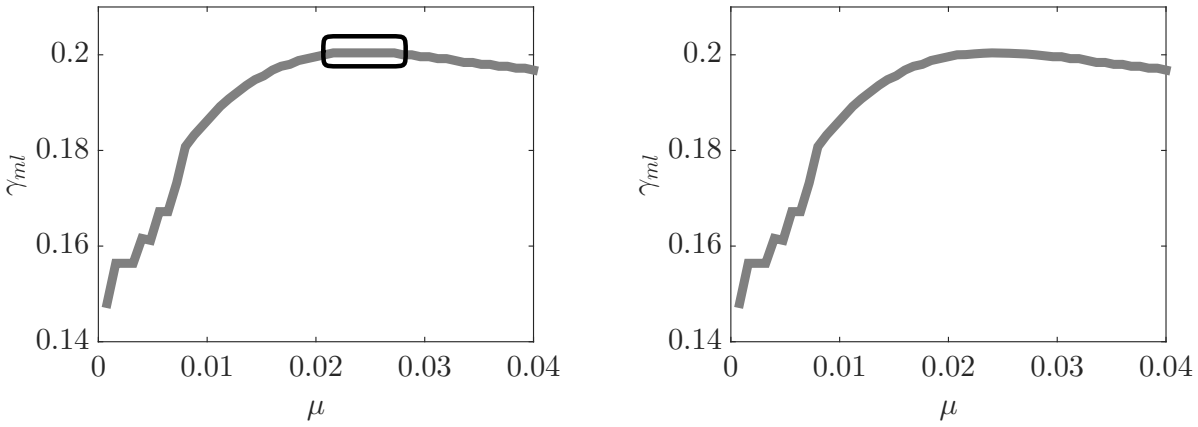


Figure 6: Variation of the mitigation limit as a function of the damping ratio of the NES (μ) by the deterministic method. (a) 50 samples of $\mu \in [0, 0.04]$; (b) 10 added samples of $\mu \in [0.02, 0.03]$.

6.2 Optimization using the ME-gPC based method

In this section, the algorithm presented in Fig. 5 is applied to estimate A_1^{wNES} (the considered QoI) and then the mitigation limit and the PHSSR. The gPC order is $p = 1$ and the threshold $\theta_1 = 2.10^{-3}\%$. Fig. 7 shows an example, with a given value of μ, of the algorithm described in Section 5.2. As described above, the mitigation limit is approximated by the upper bound of the last found element. Now, 50 deterministic samples of the damping rate of the NES ($\mu \in [0, 0.04]$) are used. The samples of the friction coefficient γ are built

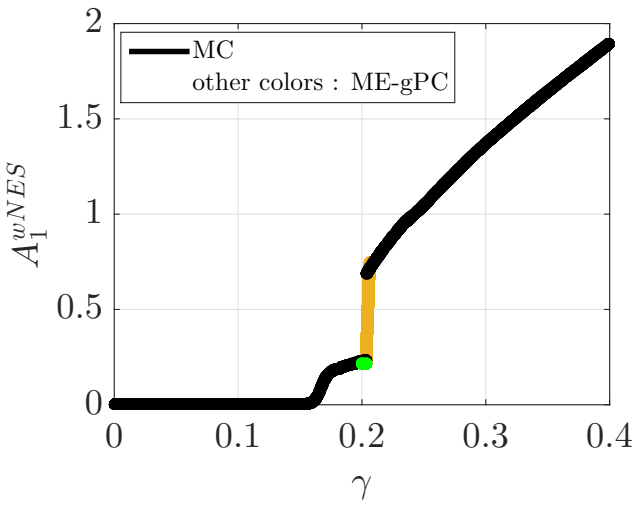


Figure 7: Last step of the ME-gPC based method to identify of the mitigation limit (γ_{ml}).

using the Latin Hypercube Samples (LHS) method. Fig. 8(a) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 1\%$. In this case, the number of simulations is equal to 6547. The optimal value of μ is in the interval $I_{\text{opt},1} = [0.016, 0.032]$. In order to reduce the size of I_{opt} the threshold θ_2 and the mitigation limit is determined within $I_{\text{opt},1}$. Fig. 8(b) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 0.1\%$. The added number of simulations is equal to 2766 and the optimal value of μ is in the interval $I_{\text{opt},2} = [0.019, 0.026]$. Fig. 8(c) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 0.05\%$ used within $I_{\text{opt},2}$. In this case, the added number of simulations is equal to 1395 and the optimal value of μ is now in the interval $I_{\text{opt},2} = [0.0208, 0.0248] = 0.0228 \pm 8\%$. This interval has a midpoint equal to 0.0228 which gives a relative error equal to 5% compared to the reference. The total number of simulations is equal to 10708 which represents 14 times less than the reference.

7 Conclusion

In this paper, the optimization of two NESs attached to the two degrees-of-freedom (DOF) Hultèn's model is presented. It accounts the discontinuity presented in the steady-state profiles using a stochastic optimization method. The results show the efficiency of the ME-gPC based method to detect the discontinuity and therefore the so-called mitigation limit. This allows us to provide an optimal design of the NESs with a low computational cost compare to that the deterministic optimization reference method.

References

[1] J.T. Oden and J.A.C. Martins. Models and computational methods for dynamic friction phenomena. *Computer Methods in Applied Mechanics and Engineering*, 52(1):527–634, 1985.

[2] Guillaume Fritz, Jean-Jacques Sinou, Jean-Marc Duffal, and Louis Jézéquel. Investigation of the relationship between damping and mode-coupling patterns in case of brake squeal. *Journal of Sound and Vibration*, 307(3):591–609, 2007.

[3] B. Hervé, J.-J. Sinou, H. Mahé, and L. Jézéquel. Analysis of squeal noise and mode coupling instabilities including damping and gyroscopic effects. *European Journal of Mechanics - A/Solids*, 27(2):141–160, mar 2008.

[4] J. Hultén. Friction phenomena related to drum brake squeal instabilities. In *ASME Design Engineering Technical Conferences, Sacramento, CA*, 1997.

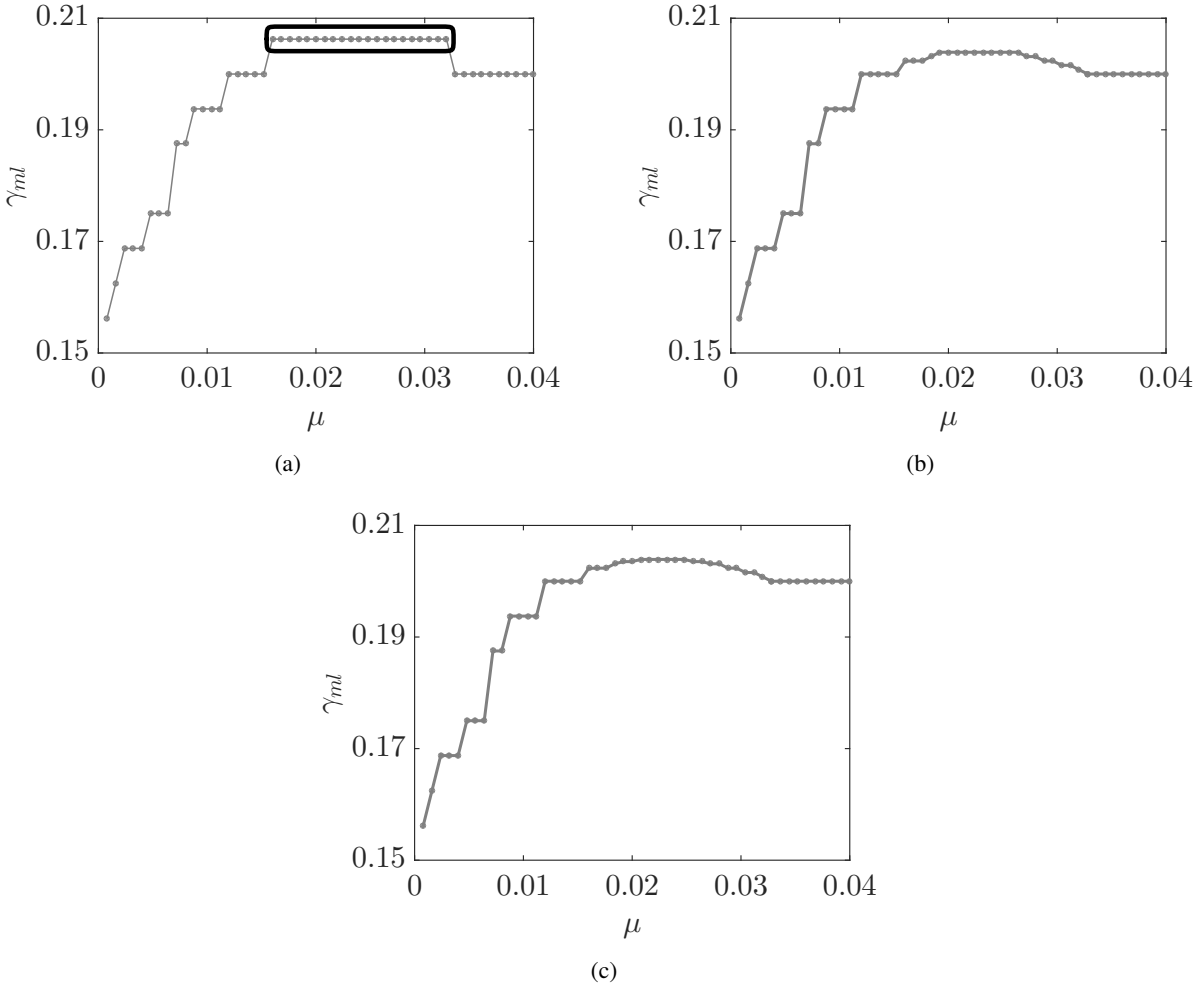


Figure 8: Variation of the mitigation limit as a function of the damping ratio of the NES using the ME-gPC based method. (a) $\theta_2 = 1\%$; (b) $\theta_2 = 0.1\%$; (c) $\theta_2 = 0.05\%$.

[5] J. Hultén. Brake squeal - a self-exciting mechanism with constant friction. In *SAE Truck and Bus Meeting, Detroit, Mi, USA*, 1993.

[6] A.F. Vakakis and O.V. Gendelman. Energy pumping in nonlinear mechanical oscillators: Part II - Resonance capture. *Journal of Applied Mechanics*, 68:42–48, 2001.

[7] A. F. Vakatis, O. V. Gendelman, L. A. Bergman, D. M. McFarland, G. Kerschen, and Y. S. Lee. *Nonlinear Targeted Energy Transfer in Mechanical and Structural Systems*. Springer-Verlag, Berlin, New York, 2008.

[8] B Bergeot, S Berger, and S Bellizzi. Mode coupling instability mitigation in friction systems by means of nonlinear energy sinks : numerical highlighting and local stability analysis. *Journal of Vibration and Control*, 24(15):3487–3511, 2017.

[9] Tuan Anh Nguyen and Stéphane Pernot. Design criteria for optimally tuned nonlinear energy sinks—part 1: transient regime. *Nonlinear Dynamics*, 69(1-2):1–19, nov 2011.

[10] Ethan Boroson, Samy Missoum, Pierre-Olivier Mattei, and Christophe Vergez. Optimization under uncertainty of parallel nonlinear energy sinks. *Journal of Sound and Vibration*, 394:451 – 464, 2017.

[11] Bharath Pidaparthi and Samy Missoum. Stochastic optimization of nonlinear energy sinks for the mitigation of limit cycle oscillations. *AIAA Journal*, pages 1–11, jan 2019.

[12] Roger G. Ghanem and Pol D. Spanos. Stochastic finite elements: A spectral approach. 1991.

[13] R.H. Cameron and W.T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[14] Xiaoliang Wan and George Em Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*, 209(2):617 – 642, 2005.

Energy exchange between a nonlinear absorber and a pendulum under parametric excitation

Gabriel HUREL[1], Alireza TURE SAVADKOOHI[1], Claude-Henri LAMARQUE[1]

[1]Univ. Lyon, ENTPE, LTDS UMR CNRS 5513, F-69518 Vaulx-en-Velin, France
gabriel.hurel@entpe.fr

Abstract

The studied system is a planar pendulum coupled with a nonlinear absorber and parametrically excited at its basis. The dynamical equations are treated with a multiple scale method. At fast time scale, a slow invariant manifold represents the asymptotic behavior. At slow time scale, the equilibrium points and their stability are investigated. Several phase portraits complete the analysis of the dynamical behavior of the system. Finally, numerical examples are given to confirm analytic predictions.

1 Introduction

Vibrations may be problematic in mechanical systems. They can provoke abnormal wear, noise or discomfort especially in case of transportation. Some devices have been designed in order to control these vibrations. Frahm [1] proposed a tuned mass damper i.e. a spring mass device coupled with the main system and able to reduce the energy of one mode. Later on, Roberson [2] showed that a nonlinear behavior of the control device can be more efficient. Since then, several nonlinear absorbers have been designed such as nonlinear tuned vibration absorber (NTVA) [3] or the nonlinear energy sink (NES) [4, 5]. The latter is purely nonlinear i.e. there is no linear term in the restoring force function. Whereas the tuned mass damper is efficient only for one mode, the NES can be used on a wider range of frequency.
Here, the system is a pendulum subject to a parametric excitation corresponding to the vertical displacement of its rotation axis. It corresponds to many industrial system, in particular to a rope-way vehicle excited by the movement of the cable. Matsuhisa et al. [6] designed several linear and nonlinear tuned mass damper in order to control the oscillations of a pendulum. Song [7] analyzed a parametrically excited pendulum used as a nonlinear absorber with an harmonic balance method. Hurel et al. [8, 9] studied a NES coupled to a two-dof pendulum excited by a generalized force with a multiple scale method. Here, the dynamical behavior of the system is also analyses with a multiple scale method.
In the section 2, the system is presented and dynamic equations are written. These equations are analyzed at two different scales of time in section 3. Then, in section 4, two numerical examples are given to illustrate analytic developments. Finally, the paper is concluded in section 5.

2 Description of the studied system

2.1 Main system

The main system is a pendulum in the plan $(\vec{e}_x,\ \vec{e}_y)$ attached at the point P by a hinged joint characterized by a viscous damping coefficient C_φ as seen on Fig. 1. Its mass, moment of inertia and center of mass are noted respectively M, J and G. The length L is the distance between the points P and G. The pendulum rotates around the point P with an angle φ. A gravitational field of magnitude g and direction $-\vec{e}_y$ exists.

2.2 Nonlinear absorber

In order to control the oscillations of the pendulum, a nonlinear absorber is coupled to the main system at a distance a from the point P. The mass m of the absorber is very small compared to the mass of the main system.

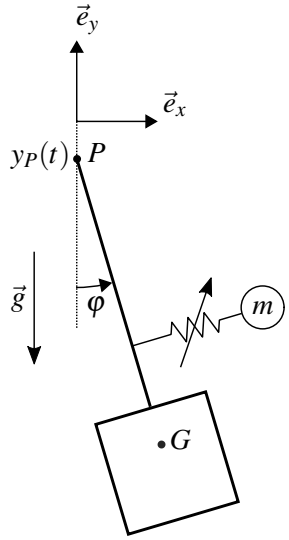


Figure 1 – Parametrically excited pendulum coupled with a nonlinear energy sink.

The ratio of mass is called ε:

$$\varepsilon = \frac{m}{M} \ll 1 \tag{1}$$

The nonlinear force function of the absorber reads:

$$s(u) = Ku^3 + C_u\dot{u} \tag{2}$$

where u is the relative displacement between m and the attached point with the main system and C_u is a viscous damping coefficient.

2.3 Parametric excitation

The main system is subject to a parametric excitation: an imposed vertical displacement of the point P called $y_P(t)$. We assume the displacement small (order of ε) and periodic with a frequency Ω. It can be written as Fourier series:

$$y_P(t) = \varepsilon \sum_{n\in\mathbb{Z}} y_n e^{in\Omega t} \tag{3}$$

where i is the complex number such as $i^2 = -1$.

2.4 Dynamical equations

The coordinates of the center of mass G and the mass of the absorber m read:

$$\begin{cases} x_G = L\sin(\varphi) \\ y_G = y_P - L\cos(\varphi) \end{cases}, \qquad \begin{cases} x_m = a\sin(\varphi) + u\cos(\varphi) \\ y_m = y_P - a\cos(\varphi) + u\sin(\varphi) \end{cases} \tag{4}$$

The kinetic $\mathscr{K}$ and potential $\mathscr{U}$ energies of the system became:

$$\mathscr{K} = \frac{1}{2}J\dot{\varphi}^2 + \frac{1}{2}M\left(\dot{x}_G^2 + \dot{y}_G^2\right) + \frac{1}{2}m\left(\dot{x}_m^2 + \dot{y}_m^2\right) \tag{5}$$

$$\mathscr{U} = Mgy_G + mgy_m + \frac{1}{4}Ku^4 \tag{6}$$

The non-conservative internal forces of the system are:

$$F_\varphi = C_\varphi\dot{\varphi} \tag{7}$$

$$F_u = C_u\dot{u} \tag{8}$$

We deduct from Eqs. 4, 5, 6, 7 and 8 with the Lagrange equations, the dynamic equations of the system:

$$\begin{cases} \left[L+j+\varepsilon(a^2+u^2)\right]\ddot{\varphi}+\varepsilon a\ddot{u}+\varepsilon c_\varphi\dot{\varphi}+2\varepsilon\dot{\varphi}\dot{u}u+\left[L\sin(\varphi)+\varepsilon(a\sin(\varphi)+u\cos(\varphi))\right](g+\ddot{y}_P)=0 \\ \varepsilon\left[a\ddot{\varphi}+\ddot{u}-\dot{\varphi}^2u+c_u\dot{u}+(g+\ddot{y}_P)\sin(\varphi)\right]+ku^3=0 \end{cases} \tag{9}$$

where $j=\dfrac{J}{M}$, $c_\varphi=\dfrac{C_\varphi}{M}$, $c_u=\dfrac{C_u}{m}$ and $k=\dfrac{K}{M}$. The natural frequency ω_0 of the main system alone at small angle reads:

$$\omega_0=\sqrt{\frac{Lg}{j+L^2}} \tag{10}$$

3 Asymptotic behavior

We use a multiple scale method to understand the behavior of the system at several scales of time. To this end, the time t is broken down in several scales τ_n, thanks to the small parameter ε:

$$\tau_n=\varepsilon^n t,\ n\in\mathbb{Z} \tag{11}$$

The derivative operator can be redefined:

$$\frac{\mathrm{d}}{\mathrm{d}t}=\sum_{n\in\mathbb{Z}}\varepsilon^n\frac{\partial}{\partial\tau_n} \tag{12}$$

We assume the angle φ and the displacement u are small. A change of scale can be performed:

$$\varphi=\sqrt{\varepsilon}\overline{\varphi} \tag{13}$$

$$u=\sqrt{\varepsilon}\overline{u} \tag{14}$$

Then the complex variables of Manevitch [10] are introduced:

$$\Phi\mathrm{e}^{i\Omega t}=\dot{\overline{\varphi}}+i\Omega\overline{\varphi} \tag{15}$$

$$U\mathrm{e}^{i\Omega t}=\dot{\overline{u}}+i\Omega\overline{u} \tag{16}$$

In the following development, we keep only the first harmonics thanks to a Galerkin method. This is carried out for an arbitrary function of the system $h(\tau_0,\tau_1,\tau_2,...)$ via:

$$H=\frac{\Omega}{2\pi}\int_0^{\frac{2\pi}{\Omega}}h(\tau_0,\tau_1,\tau_2,...)\mathrm{e}^{-i\Omega\tau_0}\mathrm{d}\tau_0 \tag{17}$$

We assume the frequency of the first harmonic of the excitation Ω is closed to ω_0:

$$\Omega-\omega_0=\sigma\varepsilon \tag{18}$$

3.1 Slow time scale τ_0

At fast time scale τ_0, the Eqs. 9 of the system yield to:

$$\frac{\partial\Phi}{\partial\tau_0}=0 \tag{19}$$

$$\frac{\partial U}{\partial\tau_0}+i\frac{a\omega_0^2-g}{2\omega_0}\Phi+\frac{i\omega_0+c_u}{2}U-i\frac{3k}{8\omega_0^3}|U|U^2=0 \tag{20}$$

We conclude from the Eq. 19 that the amplitude Φ is independent of fast time τ_0. We are looking for the asymptotic state of the system at fast time: $\tau_0\longrightarrow\infty$ and $\frac{\partial U}{\partial\tau_0}=0$. By writing the complex variables in the polar form $\Phi=N_\varphi\mathrm{e}^{i\delta_\varphi}$ and $U=N_u\mathrm{e}^{i\delta_u}$, the Eq. 20 gives:

$$(a\omega_0^2-g)^2N_\varphi^2=\left(\frac{3k}{4\omega_0^2}N_u^3-\omega_0^2N_u\right)^2+c_u^2\omega_0^2N_u^2 \tag{21}$$

The Eq. 21 describes the slow invariant manifold of the system (SIM). It is showed on Fig. 2 with the following parameters: $k = 0.15\,\mathrm{m^{-2}\,s^{-2}}$, $j = 10\,\mathrm{m}$, $L = 1\,\mathrm{m}$, $a = 1\,\mathrm{m}$, $c_u = 0.1\,\mathrm{s^{-1}}$ and $g = 9.81\,\mathrm{m\,s^{-2}}$. By following the method described by Ture Savadkoohi et al. [11] we find the following singular points:

$$N_{u1,2} = \frac{2\omega_0^{3/2}\sqrt{2\omega_0 \pm \sqrt{\omega_0^2 - 3c_u^2}}}{3\sqrt{k}} \tag{22}$$

The zone of the SIM between these singular points is unstable.

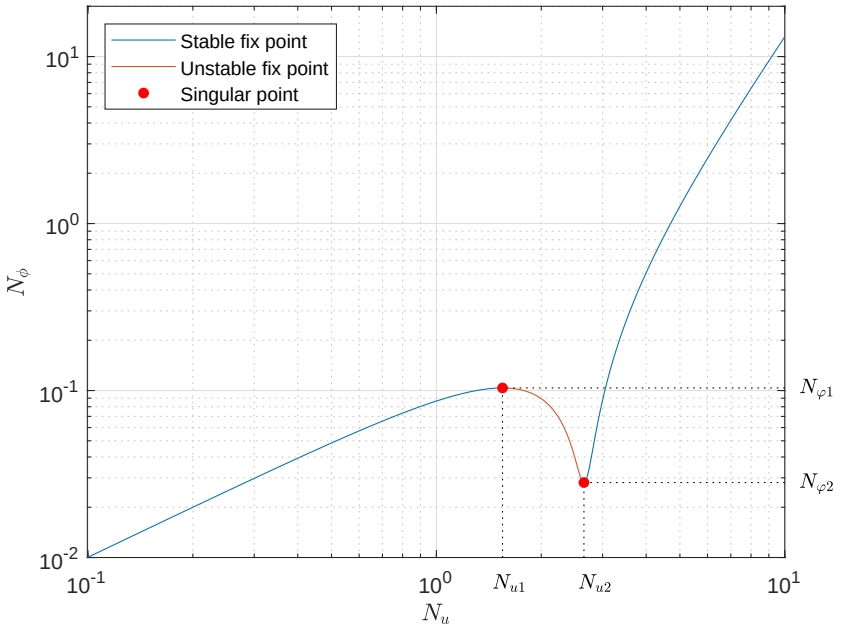


Figure 2 – Slow Invariant Manifold of the system with stable and unstable zones and singular points.

3.2 Fast time scale τ_1

We study now the system at fast time scale around the SIM. The analysis of Eqs. 9 gives:

$$Lg\frac{\partial \Phi}{\partial \tau_1} + \left(i\frac{2\sigma Lg + a\omega_0(a\omega_0^2 - g)}{2} + \frac{c_\varphi \omega_0^2}{2}\right)\Phi + i\frac{\omega_0(a\omega_0^2 - g)}{2}U - 2iy_2 L\omega_0^3\Phi^\star + i\frac{Lg|\Phi|^2\Phi}{16\omega_0} = 0 \tag{23}$$

The complex variable S can be expressed as a function of U thanks to the Eq. 20 of the SIM:

$$\Phi = \frac{U}{a\omega_0^2 - g}\left(\frac{3k|U|^2}{4\omega_0^2} - \omega_0^2 + 4ic_u\omega_0\right) \tag{24}$$

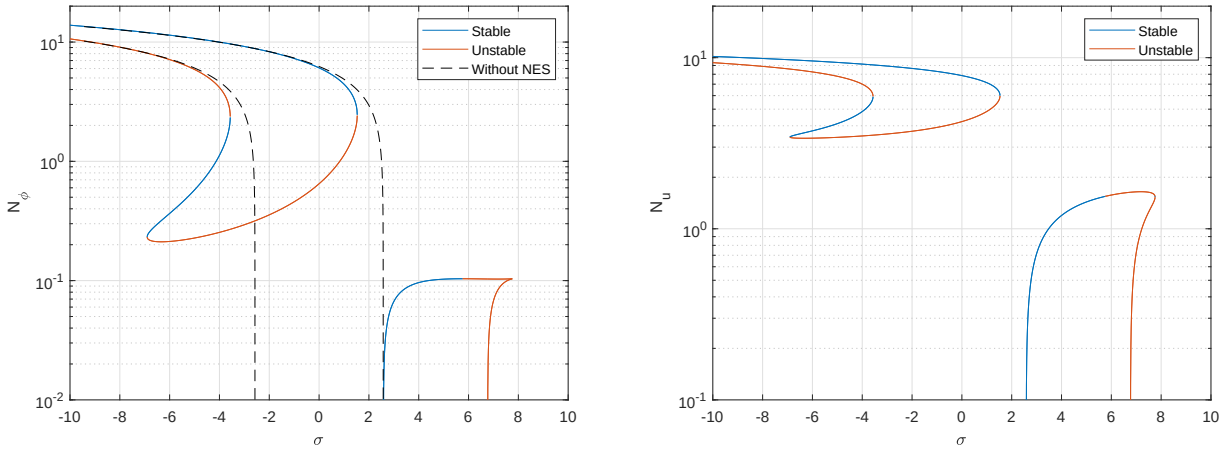
To find he equilibrium points, we consider no variation of Φ at fast time i.e. $\frac{\partial \Phi}{\partial \tau_1} = 0$. By replacing Eq. 24 in Eq. 23 and by taking the norm, we obtain a polynomial of degree 9 in N_u^2. The solutions are obtained numerically and represented on Fig. 3 as a function of σ with $c_\varphi = 50\,\mathrm{m\,s^{-1}}$ and $y_2 = 20\,\mathrm{m}$.

The stability of the equilibrium points is determined by a perturbation method with Eqs. 23 and 24:

$$\delta_u \to \delta_u + \Delta\delta_u, \qquad N_u \to N_u + \Delta N_u \tag{25}$$

After linearisation, we can write:

$$\mathbf{A}\begin{bmatrix} \frac{\partial \Delta\delta_u}{\partial \tau_1} \\ \frac{\partial \Delta N_u}{\partial \tau_1} \end{bmatrix} = \mathbf{B}\begin{bmatrix} \Delta\delta_u \\ \Delta N_u \end{bmatrix} \tag{26}$$



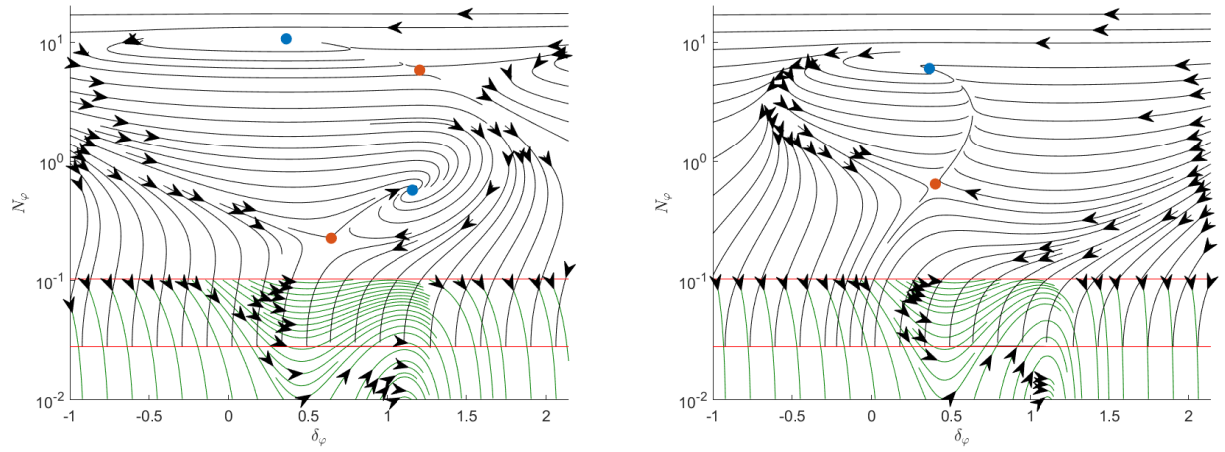
(a) N_φ (b) N_u

Figure 3 – Equilibrium point of the system with NES compared to the equilibrium points of the system without NES (dashed line) as a function of σ.

where $\mathbf{A}$ and $\mathbf{B}$ are matrices. The stability depends on the signs of the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$. On the Fig. 3, the stability of the equilibrium points is represented with blue (stable) and red (unstable) colors. Note that $N_u = N_\varphi = 0$ is also an equilibrium point for every σ but is not represented because of the logarithmic scale of the graph. Its stability depends on δ_u.

3.3 Phase portrait

The knowledge of the equilibrium points and their stability is not enough to predict the behavior of the system. To complete the analysis, phase portraits are computed and plotted on Fig. 4 for two values of σ. In each case, for a starting point of the system with $N_\varphi(0) < 0.2\,\mathrm{s}^{-1}$, the system will evolute to stay below the second singular point i.e. $\exists t_1, \forall t > t_1\ N_\varphi(t) \leq N_{\varphi 1}$. This fact is illustrated with numerical examples in the next section.



(a) $\sigma = -5\,\mathrm{Hz}$ (b) $\sigma = 0\,\mathrm{Hz}$

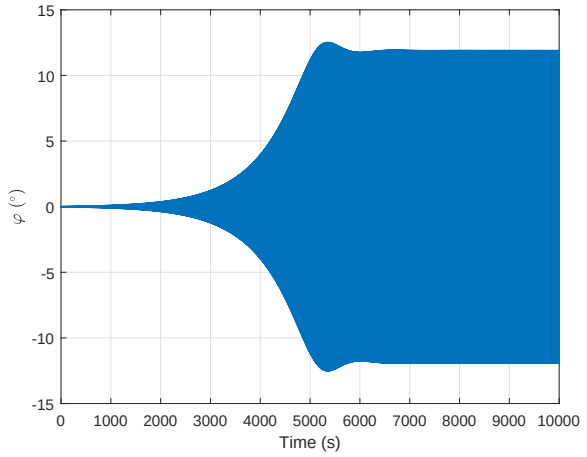
Figure 4 – Phase portrait of the system for two values of σ. A blue point is a stable equilibrium point whereas a red point is an unstable equilibrium point. Red lines correspond to singular points $N_{\varphi 1}$ and $N_{\varphi 2}$. The first stable zone of the SIM ($N_u < N_{u1}$) is represented by green curves and the second stable zone ($N_u > N_{u2}$) by black curves. The equilibrium point $N_\varphi = 0$ is not visible because of the logarithmic scale.

4 Numerical simulations

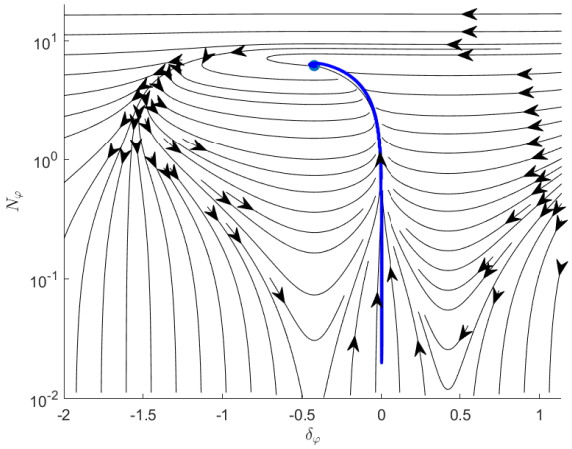
In order to illustrate the efficiency of the NES, two cases are shown with $\sigma = 0\,\text{Hz}$:

- case 1: the pendulum without NES with a very low initial amplitude $N_\varphi(0) = 0.1\,\text{s}^{-1}$ but with a particular phase $\delta_\varphi(0) = 0\,\text{rad}$.
- case 2: the pendulum with NES with a relatively high initial amplitude $N_\varphi(0) = 0.4\,\text{s}^{-1}$ and an arbitrary phase $\delta_\varphi(0) = 1\,\text{rad}$.

The results are shown on figure 5 and 6 with $\varepsilon = 10^{-2}$. First, we note a good agreement between numerical calculations and analytic phase portraits. In the first case, despite the initial amplitude is low, the system moves toward an equilibrium point with high amplitude ($N_\varphi = 0.8$). In the second case, the system goes to the equilibrium point with zero amplitude while the initial condition was higher than in the previous case. This is true for any initial phase angle $\delta_\varphi(0)$.

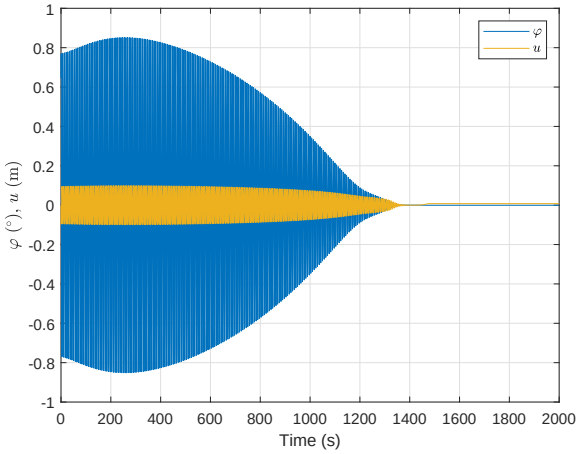


(a) φ as a function of time.

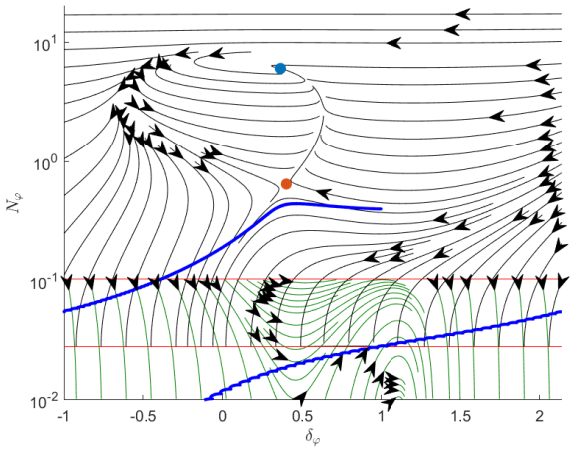


(b) Phase portrait and numerical result.

Figure 5 – Case 1: numerical result without NES, $N_\varphi(0) = 0.1\,\text{s}^{-1}$, $\delta_\varphi(0) = 0\,\text{rad}$



(a) φ and u as functions of time.



(b) Phase portrait and numerical result.

Figure 6 – Case 2: numerical result with NES, $N_\varphi(0) = 0.4\,\text{s}^{-1}$, $\delta_\varphi(0) = 1\,\text{rad}$

5 Conclusion

After a presentation of a system of a pendulum coupled with a nonlinear energy sink under parametric excitation, the dynamic equations are written thanks to the Lagrange equations. The analysis with a multiple scale method at fast time shows a slow invariant manifold of the system. At the next order, the equations give the equilibrium points and their stability. They are traced as a function of the frequency of excitation. To better predict the behavior of the system, phase portraits are drawn at different values of frequency. They show that the nonlinear energy sink can help the system to stay below a value instead of going to an equilibrium point with very high amplitude. This result is illustrated by two numerical examples. The first one shows that even if the initial amplitude of the angle of the pendulum without absorber is very low, the system can reach very high amplitude. As shown in the second example, this does not happen with a nonlinear absorber.

Acknowledgements

The authors would like to thank “La Région Auvergne-Rhône-Alpes” for supporting this work in the frame of the CALIPSO project 17 010971 01 - 15 713.

References

[1] H. Frahm, “Device for damping vibrations of bodies.,” Apr. 1911. US Patent 989,958.

[2] R. E. Roberson, “Synthesis of a nonlinear dynamic vibration absorber,” *Journal of the Franklin Institute*, vol. 254, pp. 205–220, Sept. 1952.

[3] G. Habib and G. Kerschen, “A principle of similarity for nonlinear vibration absorbers,” *Physica D: Nonlinear Phenomena*, vol. 332, pp. 1–8, Oct. 2016.

[4] O. Gendelman, L. I. Manevitch, A. F. Vakakis, and R. M’Closkey, “Energy Pumping in Nonlinear Mechanical Oscillators: Part I-Dynamics of the Underlying Hamiltonian Systems,” *Journal of Applied Mechanics*, vol. 68, no. 1, p. 34, 2001.

[5] A. F. Vakakis and O. Gendelman, “Energy Pumping in Nonlinear Mechanical Oscillators: Part II. Resonance Capture,” *Journal of Applied Mechanics*, vol. 68, no. 1, p. 42, 2001.

[6] H. Matsuhisa and M. Yasuda, “Dynamic vibration absorber for pendulum type structure,” Oct. 1995. US Patent 989,958.

[7] Y. Song, H. Sato, Y. Iwata, and T. Komatsuzaki, “The response of a dynamic vibration absorber system with a parametrically excited pendulum,” *Journal of Sound and Vibration*, vol. 259, pp. 747–759, Jan. 2003.

[8] G. Hurel, A. Ture Savadkoohi, and C.-H. Lamarque, “Nonlinear vibratory energy exchanges between a two degrees-of-freedom pendulum and a nonlinear absorber.,” *Journal of Engineering Mechanics*, 2019. (in press).

[9] G. Hurel, A. Ture Savadkoohi, and C.-H. Lamarque, “Passive control of a two degrees-of-freedom pendulum by a non-smooth absorber,” *Nonlinear Dynamics*, 2019. (in press).

[10] L. I. Manevitch, “The Description of Localized Normal Modes in a Chain of Nonlinear Coupled Oscillators Using Complex Variables,” *Nonlinear Dynamics*, vol. 25, pp. 95–109, July 2001.

[11] A. Ture Savadkoohi, C.-H. Lamarque, M. Weiss, B. Vaurigaud, and S. Charlemagne, “Analysis of the 1:1 resonant energy exchanges between coupled oscillators with rheologies,” *Nonlinear Dynamics*, vol. 86, pp. 2145–2159, Dec. 2016.

Smart Structures

Hybrid crankshaft control for the reduction of torsional vibrations and rotational irregularities

Guillaume PAILLOT[1], Simon CHESNÉ[1], Didier RÉMOND[1]

[1]Univ Lyon, INSA-Lyon, CNRS UMR5259, LaMCoS, F-69621, France
guillaume.paillot@insa-lyon.fr

Abstract

Rotative systems do not have a constant revolution speed. The problem is even more significant in internal combustion engines, where the crankshaft is submitted to a torque that is far from being constant over a revolution period. This matter causes unacceptable noises in the the gearbox and fatigues the shaft, and it can be even further amplified by the shaft dynamics. This well-known problem is already tackled by numerous passive systems, and some active devices are being introduced to further enhance their capabilities. However, these ones are complex, not always fail-safe, need a control unit and consume high levels of energy.
In this paper, the authors introduce a hybrid self-fed damper for the so-called rotational irregularities, based on a tuned massed damper controled through an electromagnetic coupling by the irregular behaviour itself. For a better description, the whole behaviour of this non-stationary process is governed with an angular approach.

1 Introduction

In an internal combustion engine, the impulse initiated by the combustion of gas within the cylinders is able to put the crankshaft into a rotating movement, thus transmitting a torque to the wheels through the drivetrain. However, the movement of the piston is not uniform, and the transmitted torque is far from being constant at a steady engine regime [1]. Indeed, the important pressure variations on the piston as well as the influence of the alternating mass inertia create a periodic oscillation around an average value in torque. This creates a similar oscillation in angular acceleration, speed, and position, which is commonly referred to as rotational irregularities.
This oscillation tends to generate unacceptable noises, such as rattling noise in the gearbox, and fatigues the drivetrain parts, shortening the operating lifetime. For these reasons, such irregularities need to be reduced, which is traditionally done by a flywheel located between the crankshaft and the gearbox, while a viscous Lanchester damper on the front end reduces the torsional magnitude through energy dissipation. However, due to the downsizing of engines for fuel-consumption reasons, the simple flywheel is not enough anymore for acceptable noise levels.
This rotational irregularity phenomenon can be further amplified when the critical frequency of the crankshaft is a multiple of the current engine average speed. Here again, downsizing is a problem as a crankshaft length reduction increases the critical frequency, enabling a larger spectrum of engine speeds to excite the resonance. The presence of the viscous damper ensures that the crankshaft does not immediately break, but at the cost of a dissipated power up to 1kW for heavy duty vehicles.

For those reasons, there is a continuous research for more efficient solutions tackling the two presented issues : mitigation of the torsional levels in service and especially at the critical frequencies, and reduction of the dissipated energy in the damper. A large amount of passive solutions have been introduced over the two last decades in order to decrease the magnitude of torsional motions, typical examples being the Double (or Triple) Mass Flywheel ([2],[3]), variable inertia flywheels ([4],[5]), planetary gear systems ([6]), or Centrifugal Pendulum Vibrations Absorbers ([7],[8],[9]). These latter rely on the creation of an extra degree of freedom, enabling an "escape way" during resonance. However, its behaviour is still dependent on the steady engine regime, complicating a tuning process[10].

Some active concepts have also been proposed ([11],[12],[13],[14]), as the control possibilities they offer greatly enhance the capabilities of such dampers. Their limitation is the need for a power source.
Regarding the second issue, a promising way would be to collect the otherwise dissipated energy, and to store it so that it can be used by other devices in the vehicle. Several possibilities exist for the conversion, based on piezoelectricity - R-EHVTA [15] or pendulum harvester [16] - or on electromagnetic coupling, with some concepts developped over the last years, such as [17] or [18].

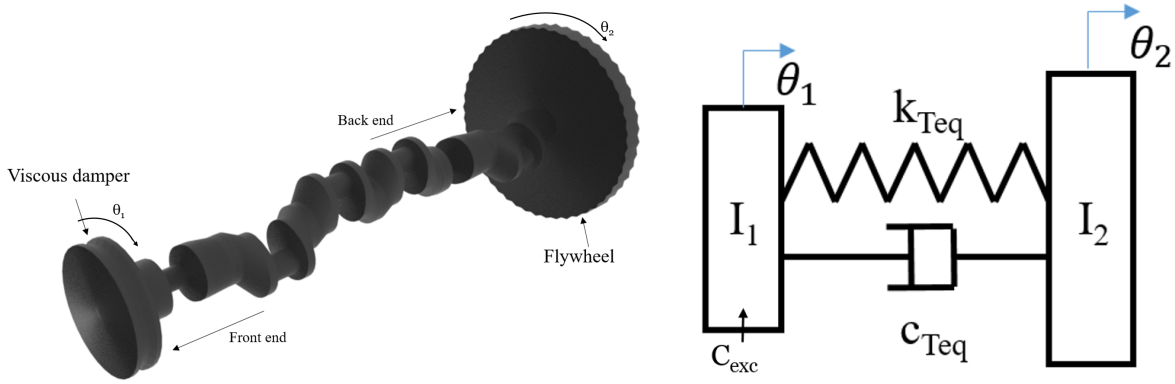
In this paper, the authors introduce a concept for a new self-fed rotational hybrid damper, combining an electromagnetic harvester with a hybrid Tuned Mass Damper (hTMD), that has already proved its capabilities for translational motions applications [19]. It is based on the generation of a counteracting torque by a set of permanent magnets and coils, similar to what happens in active dampers, coupled with the principle of the TMD, efficient against resonance issues. The TMD embeds a permanent magnet that moves in a coil. Both movements will change the local magnetic fields and generate an induced current in a common circuit, acting on each other.
In order to tackle non-stationary operating conditions, the present model involves angular and time approaches to describe the cyclic excitations and the frequency resonances.

2 Model building, important concepts and assumptions

The present concept can be applied to any rotating shaft submitted to rotational irregularities, no matter their profile. However in this paper we consider a 6-cylinder crankshaft, each with a cubic capacity of 1,3L.

2.1 Mechanical model

The crankshaft is first reduced to a 2 DoF-in-torsion part, as only the first torsional mode is of interest in this academic first application. As described in Fig.1, a regime-dependant cyclic excitation C_{exc} is introduced on the front degree of freedom. This excitation is the sum of the the torque produced by the explosion in the cylinders and of inertial effects. It is further described in the next section.
The measures are conducted on both ends, the second being the output of the shaft, connected to the gearbox. On this one, a resistive torque C_c assumed constant is applied, so as to balance the system and enable a steady speed. The actual interest is indeed the difference between both angles $\theta_2 - \theta_1$ for a constant engine regime. Numerical values of the introduced parameters are to be found in Table.1.



(a) Full crankshaft (b) Reduced model

Figure 1 – Model reduction

This model is governed by the system of equations :

$$\begin{cases} I_1\ddot{\theta}_1 + c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2) = C_{exc} \\ I_2\ddot{\theta}_2 + c_{Teq}(\dot{\theta}_2 - \dot{\theta}_1) + k_{Teq}(\theta_2 - \theta_1) = -C_c \end{cases} \tag{1}$$

Parameter	Value
I_1	0.23 kg.m^2
I_2	1.65 kg.m^2
k_{Teq}	524 114 Nm/rad
c_{Teq}	100 Nms/rad

Table 1 – Parameters of the 2-DoF model

In a perfect engine, the natural frequency is excited only by the orders that are multiples of the number of cylinders, N_{cyl}, devided by two. To that extent, only the orders multiple of 3 are here taken into account. According to Table 2, as our engine speed range is from 600RPM to 2600RPM and the natural frequency is 256,8 Hz, we can restrict the analysis up to the 24th order : higher orders will not excite the natural frequency .

Order	Engine steady regime
3	-
6	2569 RPM
9	1713 RPM
12	1285 RPM
15	1028 RPM
18	857 RPM
21	734 RPM
24	642 RPM

Table 2 – The investigated orders and the corresponding engine regime at resonance

2.2 Excitation torque profile and non-stationary conditions

In a 4-stroke internal combustion engine, a complete period is accomplished in two revolutions, each stroke corresponding to half a revolution. Among them, only the firing stroke produces a useful torque, with a non-constant force over the revolution. However, the presence of the crank also causes it to be highly dependant on the instantaneous position α, as it is depicted in Fig.2. Noting λ the ratio of the rod length L_B to the crank radius R, F_{bg} the projection on the rod of the gas-induced force and d its lever arm, a geometric projection of $C_g = -F_{bg}.d$ brings :

$$C_g(\alpha) = -F_g(\alpha) \cdot R \cdot \sin(\alpha) \cdot \left(1 + \frac{\cos(\alpha)}{\sqrt{\lambda^2 - \sin(\alpha)^2}}\right) \tag{2}$$

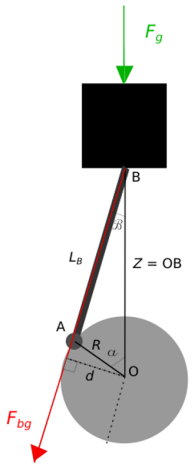


Figure 2 – Crankshaft, piston and rod

The force F_g, created by the explosion in the piston, is also dependant on the engine steady regime. In this expression, a constant part is the engine torque, and the oscillations around it are a first source of rotational irregularities.

The second source originates from the motions of the masses around and above the axis : the sum of the forces created by their inertias is not 0. Here again, there is a strong dependence on the instantaneous angular position, and also on the instantaneous angular speed. We note m_{alt} the mass considered in a translational motion when the shaft rotates. Assumed that the inertias do not change with speed and position :

$$\begin{cases} F_i = -m_{alt}.\dfrac{d^2Z}{dt^2} \\ Z = R\left(\cos(\alpha) + \sqrt{\lambda^2 - \sin(\alpha)^2}\right) \\ C_i = -F_i.R.\left(cos(\alpha) + \sqrt{\lambda^2 - sin(\alpha)^2}\right)\dfrac{sin\alpha}{\sqrt{\lambda^2 - sin(\alpha)^2}} \end{cases} \tag{3}$$

Which provides :

$$C_i = -m_{alt}.R^2\left(\left(sin(\alpha) + \frac{sin(2\alpha)}{2\sqrt{\lambda^2 - sin(\alpha)^2}}\right)\ddot{\alpha} + \left(cos(\alpha) + \frac{\lambda^2 cos(2\alpha) + sin(\alpha)^4}{(\lambda^2 - sin(\alpha)^2)^{3/2}}\right)\dot{\alpha}^2\right)sin(\alpha)\left(1 + \frac{cos(\alpha)}{\sqrt{\lambda^2 - sin(\alpha)^2}}\right) \tag{4}$$

The final excitation torque C_{exc} is the sum of the inertial torque and the oscillating gas torque. Its process is non-stationnary, though the simplifying stationary assumption $\alpha = \omega.t$ with a constant ω is often made. However, the instantaneous speed does not remain constant, since the input torque itself is harmonic : the angular acceleration is not 0, and its magnitude is directly linked to C_{exc}. In order to prevent any hidden behaviour to disappear in the analysis, and as the angular position θ_1 appears as the natural variable for such a rotating system, we use this variable instead of time for the resolution. The counterpart here is the introduction in the equations of a non-linear behaviour when the variable is swapped, as there is now a division by the instantaneous angular speed : $\frac{dX}{d\theta_1} = \frac{1}{\dot{\theta}_1}\frac{dX}{dt}$.

2.3 TMD model

First addition to the previous 2-DoF model is the mass damper, tuned on its resonance frequency according to Den Hartog's principles [20]. It consists in a magnet clamped on a mass, linked to the rotating shaft with a spring set at a certain distance R_{TMD} of the rotation axis. The resulting torsional stiffness k_{TMD} depends on the tuning. Around this magnet, a coil is set, clamped to the shaft, in such a way that the TMD can translate within.

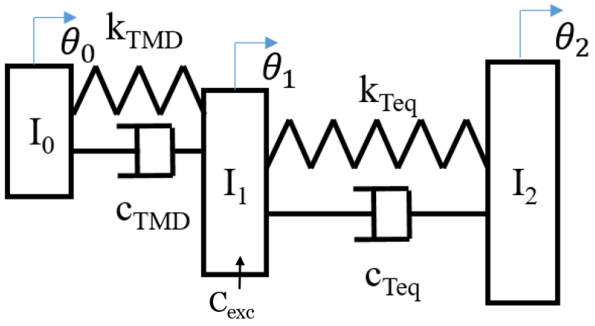


Figure 3 – Reduced model of crankshaft including a TMD in front position

However, as we want to control the torsional behaviour, determined by the difference of the angles, both degrees of freedom are potential candidates to host the TMD set, but the two will not act the same.
The front end has the greatest amplitude during vibration due to its lower inertia, and the lower the inertia, the lower the efficient modal inertia I_{eq}. At a given R_{TMD}, a lower mass for the TMD is then required, as the ratio of its inertia I_0 to I_{eq} is the key parameter of the tuning. To that extent, better performance can be achieved for

Parameter	Value
I_0	0.013 kg.m^2
μ	0.05
k_{TMD}	29 617 Nm/rad
c_{TMD}	7.36 Nms/rad
R_{TMD}	0.2 m
m_{TMD}	0.325 kg

Table 3 – Parameters of the 3-DoF model

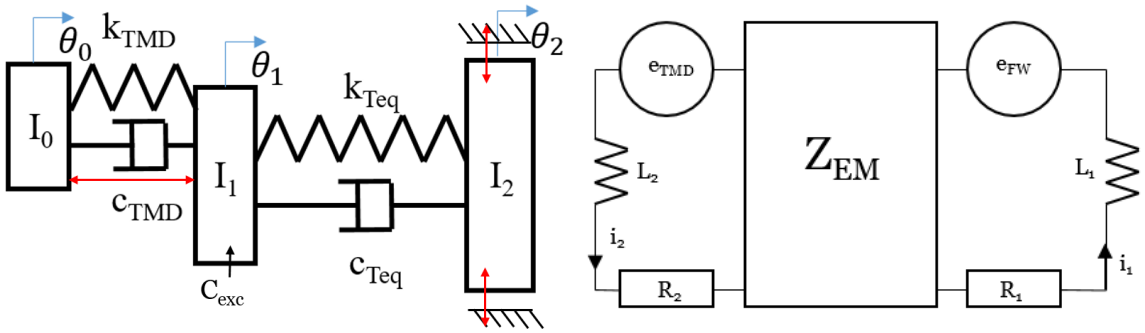
a TMD on the front end and in the rest of the paper, this integration possibility is chosen, described in Fig.3. Its absolute angle around the rotation axis is parameterized by θ_0.
To bring what happens to the TMD in a more practical frame, let us define a relative variable for the new degree of freedom : $x = R_{TMD}(\theta_0 - \theta_1)$. This variable translates the torsional motion of the TMD around the axis of rotation in regard to the disc to which it is clamped. The values for all introduced parameters can be found in Table 3. The inertia ratio μ is chosen to 0.05, as this value allows a realistic mass and a realistic distance to the shaft to be chosen.

2.4 Electromagnetic model

The electromagnetic model actually consists in two parts. One has already been introduced in subsection 2.3, and describes the interaction of the magnet on the TMD and its surrounding coil. Indeed, any motion of the TMD creates an induced current in the coil, and conversely, any current circulating in the coil creates a magnetic field interacting with the magnet.

The second part is located around the flywheel. It consists in a set of coils and another of magnets, one being embedded in the flywheel and rotating with the shaft, the other being fixed in the vehicle, all around the flywheel. Again, moving a magnet close to a coil will create a current, itself generating a magnetic field that will slow down the motion, according to Lenz's law, and eventually damp the rotational irregularities. We note ϕ the magnetic flux.

The red arrows in Fig.4a show the different electromagnetic interactions that occur. The two parts are linked to another through an impedance Z_{EM}, as to be seen in Fig.4b. What this impedance covers depends on the interaction scenario.



(a) Reduced model with EM interactions

(b) Electromagnetic circuit

Figure 4 – Electromagnetic model

With the electromotive forces e_{FW} at the flywheel and e_{TMD} at the TMD, defined as:

$$\begin{cases} e_{FW} = -\dfrac{d\phi}{dt} \\ e_{TMD} = -T_2\dot{x} \end{cases} \tag{5}$$

and under the assumption that $Z_{EM} = 0$, the total model can thus be described with a system of five equations, three mechanical equations and two electrical ones :

$$\begin{cases} I_0\ddot{x} + (1+\frac{I_0}{I_1})(c_{TMD}\dot{x} + k_{TMD}x) - \frac{I_0}{I_1}R_{TMD}(c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2)) = -\frac{I_0}{I_1}R_{TMD}.C_{exc} + R_{TMD}^2 T_2 i_2 (1+\frac{I_0}{I_1}) \\ I_1\ddot{\theta}_1 + c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2) - \frac{c_{TMD}}{R_{TMD}}\dot{x} - \frac{k_{TMD}}{R_{TMD}}x = C_{exc} - R_{TMD}^2 T_2 i_2 \\ I_2\ddot{\theta}_2 + c_{Teq}(\dot{\theta}_2 - \dot{\theta}_1) + k_{Teq}(\theta_2 - \theta_1) = -C_c + T_1.R_{ext} i_1 \\ L_1\frac{di_1}{dt} + R_1 i_1 = -pT_1\dot{\theta}_2 \\ L_2\frac{di_2}{dt} + R_2 i_2 = -T_2\dot{x} \end{cases} \tag{6}$$

Where T_2 is the coupling factor depicting the electromagnetic interaction between the TMD and the surrounding coil [21], p is a scalar and T_1 is a coupling function linked to $\frac{d\phi}{dt}$. Both are further described in the next subsection.
In the case where the TMD is actually located on the flywheel, or depending on Z_{EM}, these equations can be slightly different.

2.5 The T_1 function and p parameter

In subsection 2.2, we have stated that an angular approach is better suited to describe the rotation of the shaft and the irregularities that occur. Moreover, Faraday's law states that the electromotive force is proportional to the time derivative of the magnetic flux, which itself is then a function of the instantaneous angular position, here measured at the flywheel. Changing the variable t to θ_2 provides $e_{FW} = -\dot{\theta}_2\frac{d\phi}{d\theta_2}$.
On the other hand, the countering torque C_{count} created at the flywheel by the current also depends on the location of the various coils and magnets. It is able to mitigate the rotational irregularities when it and the excitation torque C_{ext} have the same main harmonics, but with a 180° phase shift. Assuming this torque and $\frac{d\phi}{d\theta_2}$ have a same sine shape involving the same harmonics, the introduced T_1 function aims to bridge the gap between them. To that extent, the torque exerted when a number n of coils and a number m of magnets interact during the revolution can be described using the expression :

$$T_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} K_{i,j} m_j \mu_0 \frac{N_i}{L_{sol,i}} sin(\beta_{i,j}) \tag{7}$$

and the relationship :

$$C_{count} = T_1.R_{ext} i_1 \tag{8}$$

In (7), N_i is the number of turns in solenoid i, $L_{sol,i}$ the length of the same solenoid, m_j the magnetic moment of magnet j, and K is a matrix describing the coupling coefficient between coil i and magnet j. $\beta_{i,j}$ is a local variable standing for the angular position θ_2 when a magnet interacts with a coil.
With this description, we also have the relationship :

$$e_{FW} = -p\dot{\theta}_2 T_1 \tag{9}$$

the description of the harmonics being included in T_1, except for their amplitude. To that extent, in order to keep realistic values for the electromotive force e_{FW}, we introduce an additional parameter, p, which purpose is to adjust the magnitude of the harmonics.

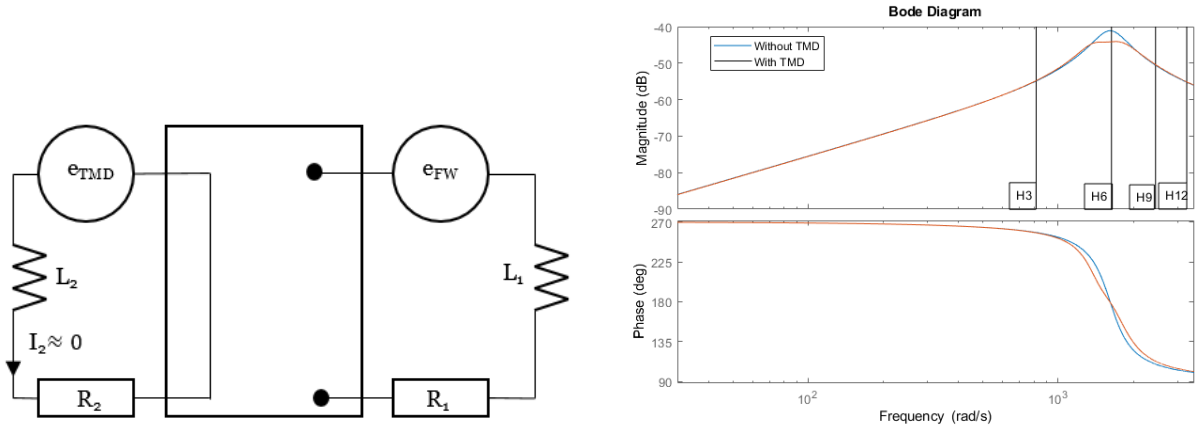
3 Model results

3.1 Efficiency of the TMD

To display the efficiency of the concept in our model, let us first consider a decoupled electromagnetic circuit, as modeled in Fig.5a. With a low value for T_2, the created current is very low, and the natural motion of the TMD is almost not impeded. Regarding the electromagnetic setup around the flywheel, there is also

little impact on the overall rotation as the countering torque C_{count} is four orders of magnitude lower than the excitation torque C_{exc}. Such configuration is equivalent to actually no circuit at all, and enables to measure the effect of a passive TMD.

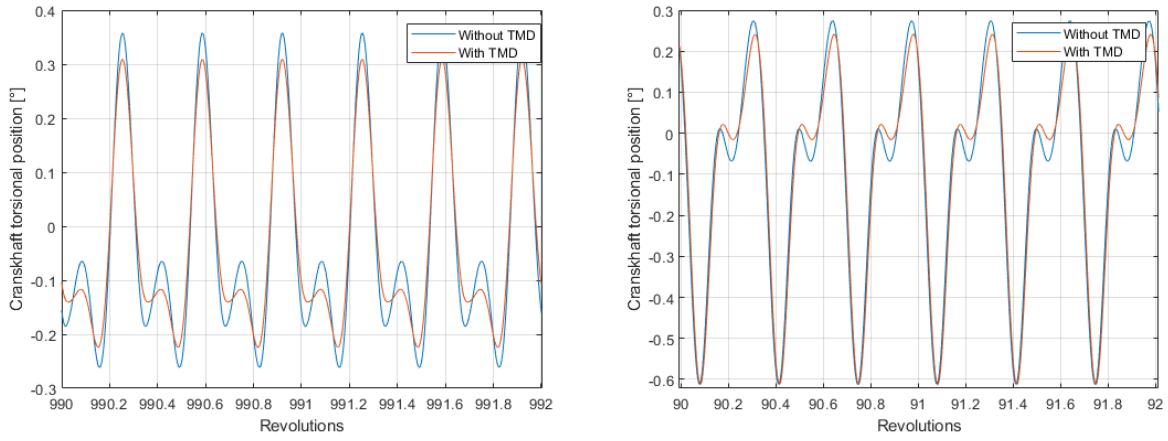
As we can see on the Bode diagram in Fig.5b, the TMD flattens the peak at the natural frequency of the torsional mode. The tuning parameters used are those detailed in Table.3. This enables a reduction of 15% the amplitude in torsion, as displayed in Fig.6a for the response at 2600RPM. At this regime, the 6^{th} order excites a frequency which is very close to the natural frequency. At 2200RPM, situation displayed in Fig.6b, the frequencies excited are further from the natural frequency, and the damping is smaller, even though still present.



(a) Decoupled electromagnetic circuit

(b) Bode diagram with the harmonics at 2600RPM

Figure 5 – Passive TMD scenario



(a) At 2600RPM

(b) At 2200RPM

Figure 6 – Typical magnitude reductions

3.2 Coupling effect in the hybrid damper

Now let us consider a direct hookup between the two constituents of the electromagnetic circuit, a situation depicted in Fig.7. The equations in (6) must then be changed, as it is now imposed $i_1 = i_2$. The parameters used are listed in Table.4. In this situation, we still have $Z_{EM} = 0$

Parameter	Value
L_1	555mH
L_2	185mH
R_1	15.51Ω
R_2	5.17Ω
T_2	0.8
p	0.25
$T_1(\theta)$	$0.01.sin(6\theta)$

Table 4 – Parameters for the coupled model

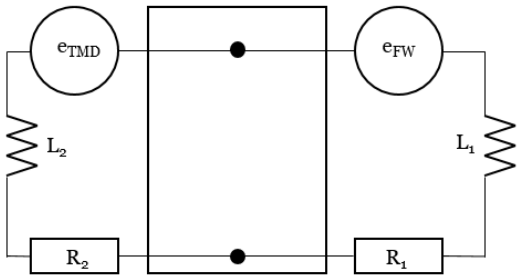
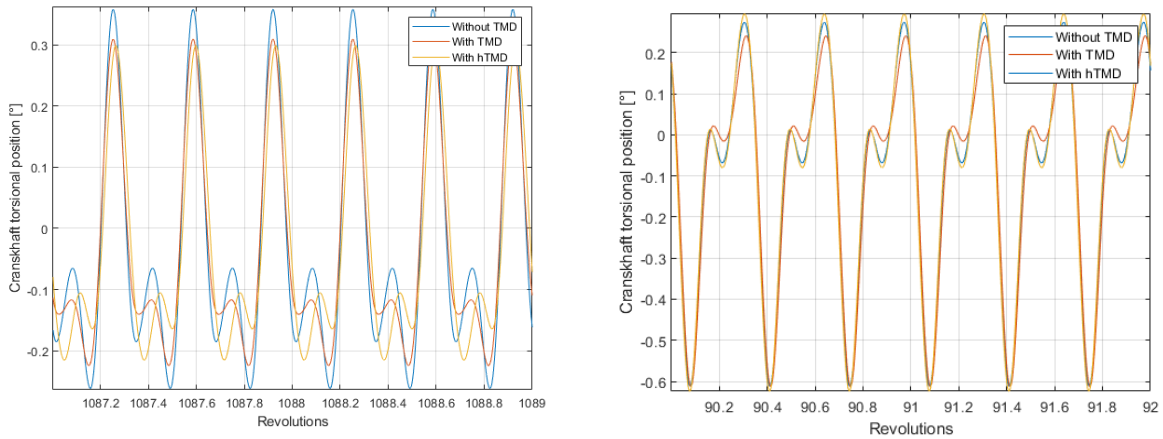


Figure 7 – Direct hookup electromagnetic circuit



(a) At 2600 RPM

(b) At 2200 RPM

Figure 8 – Comparison with the previous case

As we can see in Fig.8a, at 2600RPM in that particular situation, the damping is slightly better, even though the values considered are not tremendous. However, this results must be handled carefully, since no phase effect has yet been taken into account in T_1, effects that could make the hTMD ineffective or even worsen the situation, which is what appears at 2200RPM in Fig.8b. Indeed, the hookup is equivalent to a modification of c_{TMD} and - to a lesser extent - of k_{TMD}. A modification of the damping factor can thus be advantageous or detrimental, and what happens is highly dependent on the involved harmonics and phases, and on the hookup configuration. To that extent, it appears necessary to further investigate on the effect of each harmonic's phase and on the configurations for Z_{EM}, in order to properly damp the torsional mode in all cases.

Discussion - Conclusions

In this paper, a new concept for a hybrid mass damper against rotational irregularities has been introduced and described. The self-fed property avoids any exterior source of power, but also complicates the tuning of the system due to the instantanous angular position and the steady regime dependence. The interaction between the two parts of the electromagnetic circuit is not yet fully understood, whereas this is a major requirement to have them work together properly. Indeed, it has been seen in 3.2 that a direct hookup can provide unexpected behaviour if the phases are not carefully handled.
Further work will also be devoted to a check of the assumptions that are made. In particular, it is necessary to prove the possibility to approximate the $\frac{d\phi}{d\theta_2}$ function to a sine with given harmonics. There is also no guarantee that any harmonic signal can be approximated by at least one particular coil-and-magnet spatial layout. The tayloring is quite straightforward when only one or two harmonics are involved, though here 8 different harmonics are being investigated. However a truncation to the fundamental or to a predominant harmonic could be sufficient to efficiently damp the rotational irregularities. In all cases, further work on the damping possibilities of such a hTMD is expected, in order to present its capabilities in comparison to a fully passive TMD.
The hookup of the full electromagnetic circuit in a real engine also matters. It has been stated in 2.3 that the TMD is more efficient when mounted on the lower inertia, however the rest of the electromagnetic circuit is necessarily mounted on the flywheel, which is the other end of the crankshaft. In this configuration, it is nessary to embody the wires along the shaft in the rotating referential. The choice is left to equip the flywheel with the TMD instead, in another configuration that provides a better integration rate at the cost of a lower performance.

Acknowledgements

This work has been supported by the chair 'Solutions for the future of Road Freight Transport', jointly created by INSA Lyon LaMCoS and Volvo Group.

References

[1] J.-L. Ligier and E. Baron, *Acyclisme et vibrations: applications aux moteurs thermiques et aux transmissions. 1, Notions de base*. Paris, France: Ed. Technip, 2002, ISBN: 978-2-7108-0820-6.

[2] U. Schaper, O. Sawodny, T. Mahl, and U. Blessing, "Modeling and torque estimation of an automotive dual mass flywheel", in *2009 American Control Conference*, St. Louis, MO, USA: IEEE, 2009, pp. 1207–1212, ISBN: 978-1-4244-4523-3. DOI: 10.1109/ACC.2009.5160136.

[3] L. Wramner, V. Berbyuk, and H. Johansson, "Vibration dynamics in non-linear dual mass flywheels for heavy-duty trucks", in *Proceedings of ISMA2018 International Conference on Noise and Vibration Engineering*, 2018, pp. 1935–1947.

[4] X. Dong, "Magneto-rheological variable inertia flywheel", *Smart Materials and Structures*, vol. 27, no. 11, Jul. 2018.

[5] L. L. Schumacher, "Controllable inertia flywheel", U.S. Patent 4995282A, Feb. 26, 1991.

[6] A. Suryanarayana, "Engine dynamics and torsion vibration reduction", Master Thesis, Chalmers, 2015, 52 pp.

[7] D. E. Newland, "Developments in the design of centrifugal pendulum vibration absorbers", presented at the ICSV 24, 2017, p. 8.

[8] A. Kooy, Ed., *Isolation is the key*, Solving the Powertrain Puzzle: 10th Schaeffler Symposium April 3/4, 2014, Springer Vieweg, 2014, ISBN: 978-3-658-06194-4.

[9] M. Pfabe and C. Woernle, "Reduction of periodic torsional vibration using centrifugal pendulum vibration absorbers", *PAMM*, vol. 9, no. 1, pp. 285–286, 2009, ISSN: 1617-7061. DOI: 10.1002/pamm.200910116.

[10] K. Nishimura, T. Ikeda, and Y. Harata, "Localization phenomena in torsional rotating shaft systems with multiple centrifugal pendulum vibration absorbers", *Nonlinear Dynamics*, vol. 83, no. 3, pp. 1705–1726, Feb. 1, 2016, ISSN: 1573-269X. DOI: 10.1007/s11071-015-2441-2.

[11] J. Pfleghaar and B. Lohmann, "The electrical dual mass flywheel -an efficient active damping system", *IFAC Proceedings Volumes*, 7th IFAC Symposium on Advances in Automotive Control, vol. 46, no. 21, pp. 483–488, Jan. 1, 2013, ISSN: 1474-6670. DOI: 10.3182/20130904-4-JP-2042.00046.

[12] S. Schwunk, "Method for vibration damping of a drive train by means of an electric machine", U.S. Patent 20170334448A1, Nov. 23, 2017.

[13] A. Gruendl, B. Hoffmann, U. Masberg, T. Pels, and K.-P. Zeyen, "System zur aktiven dämpfung von drehungleichförmigkeiten der kurbelwelle eines verbrennungsmotors oder einer damit gekoppelten welle", European pat. 0847485B1, Jun. 23, 1999.

[14] F. Knopf and M. Steidl, "Verfahren zum aktiven isolieren eines antriebsstrangs von drehschwingungen einer welle einer maschine, insbesondere einer kurbelwelle einer hubkolbenmaschine, und eine entsprechende anordnung zur durchführung des verfahrens", pat. WO2014118246A1, Aug. 7, 2014.

[15] F. Infante, W. Kaal, S. Perfetto, and S. Herold, "Design development of rotational energy harvesting vibration absorber (r-EHTVA)", Sep. 10, 2018. DOI: 10.1115/SMASIS2018-7902.

[16] L. Gu and C. Livermore, "Pendulum-driven passive self-tuning energy harvester for rotating applications", *Applied Physics Letters*, p. 4,

[17] S. Troebst, "Dämpfungssystem mit Energierückgewinnung für ein Kraftfahrzeug und Antriebsvorrichtung mit einem derartigen Dämpfungssystem", European pat. 2910813A2, Aug. 26, 2015.

[18] B. Gunn, S. Theodossiades, and S. J. Rothberg, "A nonlinear concept of electromagnetic energy harvester for rotational applications", *Journal of Vibration and Acoustics*, Nov. 19, 2018, ISSN: 1048-9002. DOI: 10.1115/1.4042040.

[19] S. Chesne and C. Colette, "A simple hybridization of active and passive mass dampers", presented at the ISMA2016, KU Leuven, Sep. 2016.

[20] J. P. Den Hartog, *Mechanical vibrations*, 4th. New York, NY : McGraw-Hill, 1956, 446 pp.

[21] A. Preumont and B. Mokrani, *Electromagnetic and piezoelectric transducers*.

Semi-active Torsional Vibrations Control of a Rotor Using a Smart ER Dynamic Absorber

Yulan Sun and Marc Thomas

Department of Mechanical Engineering, ETS,
1100, Notre-Dame West Street, Montreal, Quebec, (H3C 1K3), CANADA
marc.thomas@etsmtl.ca

Abstract

Torsional rotor vibration is always very difficult to control since the implementation of a control system is not an easy task while the machine is rotating. Excessive torsional vibration can lead to failures of many mechanical components. A common method for controlling vibrations involves the use of dynamic absorbers. Due to their variable properties, smart materials can be used to increase the frequency range in order to control vibration. This article is concerned with the application of Electrorheological Fluids (ERF) to the reduction of torsional vibrations of a rotor by controlling damping and stiffness of a rotational dynamic absorber. A cylindrical type of electrorheological (ER) torsional absorber is designed and manufactured according to the required damping force level and the critical velocity of the rotor system. This paper presents torsional vibration control performance of a smart ER dynamic absorber using a bang-bang (ON-OFF) control strategy. The experimental results very closely approximate the simulation results. These results show that the ER dynamic absorber exhibits very good performances in terms of reducing the torsional vibration of rotor system.

Key words: Electrorheological fluids, torsional vibrations, dynamic absorber, Bang-Bang control, smart material.

1. Introduction

Torsional rotor vibration is always very difficult to control since the implementation of a control system is not an easy task while the machine is rotating. Excessive torsional vibration can lead to failures of many mechanical components. Furthermore, in many industrial applications, torsional vibration problems may not be apparent until a failure occurs. For example, it is the case when the machine contains a gearbox which causes the torsional vibration to cross-couple into lateral vibrations [1]. A common method to control vibrations involves the use of dynamic absorbers [2]. However, when the physical parameters of dynamic absorbers are constant, their frequency range efficiency is tight and thus not really suitable for variable speeds systems. To remedy this, a damping could be added to the absorber, but this would lead to a loss of efficiency at the main frequency and increases the transmissibility on a large frequency range. On the other hand, due to their variable properties, smart materials may be used to increase the frequency range in order to control vibration. Electrorheological (ER) fluids are attractive materials that undergo very fast reversible changes in their rheological properties (mainly their yield stress and effective viscosity) upon the application of electric fields [3]. This fluid may potentially be applied in many industrial areas [4]. They may be used to control vibrations [5], to control valves [6] or rotor squeeze film dampers [7]. ER fluid may also be applied to control active suspensions [8] or seismic vibrations [9], through the application of

high electric fields. Previous studies conducted by the authors have allowed the development of a new very efficient Electro-Rheological (ER) fluid [10]. In this paper, we propose the design of a new dynamic torsional absorber by using ER fluid for a rotor system. The smart absorber is controlled by varying the electric field applied to the fluid.

2. ER smart dynamic torsional absorber

2.1 Model of ER fluid

In general, most ER fluid exhibit Bingham rheological behaviors and shear-thinning. In previous studies, the authors developed a model call the Quasi-Bingham model to explain this complex phenomenon (Sun *et al*., 2009, 2010), which is:

$$\begin{aligned}
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} \quad or \quad \tau = \eta_\infty \dot{\gamma} && when \quad E = 0 \\
&\tau = \tau_{yQB} + \eta_0 \dot{\gamma} && when \quad E \neq 0 \quad \dot{\gamma} \le \dot{\gamma}_1 \\
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} + (\eta_0 - \eta_\infty)\frac{\dot{\gamma}}{1+(t\dot{\gamma})^n} && when \quad E \neq 0 \quad \dot{\gamma}_1 \le \dot{\gamma} \le \dot{\gamma}_2 \\
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} && when \quad E \neq 0 \quad \dot{\gamma} \ge \dot{\gamma}_2
\end{aligned} \tag{1}$$

where τ is the shear stress (Pa); $\dot{\gamma}$ is the shear rate; τ_{yQB} is the elastic yield stress (Pa) which is dependent on the strain and on the applied electric field E.

According to the dielectric loss model of Hao *et al*. [11], the yield stress τ_{yQB} is a function of E^2; η_0 is the zero shear viscosity; it is defined as the value at a very low shear rate and is a function of E; η_∞ is the infinite shear viscosity; it is defined as the value at very high shear rate; the parameter "n" is known as the Cross Rate Constant [12]. It is a dimensionless factor and is a measurement of the degree of dependence of viscosity on shear rate in the shear-thinning region; "t" is known as the Cross Time variable and has a dimension of time. The reciprocal, 1/t, gives us a critical shear rate that is a useful indicator of the onset shear rate for shear thinning; $\dot{\gamma}_1, \dot{\gamma}_2$ are two critical shear rates. We can obtain the values by using the two equations:

$$1+\left(t\dot{\gamma}_1\right)^n \approx 1; \quad \dot{\gamma}_2 \gg 1/t \tag{2}$$

This model can explain very well the rheological behaviors. An empirical equation (equation (3) of stresses was developed for a mixture called ETSERF40-20 as an example of the application of the Quasi-Bingham model. This equation is very useful for modeling a control system with the ER fluid [13].

$$\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} + (\eta_0 - \eta_\infty)\frac{\dot{\gamma}}{1+(t\dot{\gamma})^n} \tag{3}$$

$$\tau \quad = 27.04\times10^{-9}E^2 + 0.218\dot{\gamma} + 29.4\times10^{-3}E\frac{\dot{\gamma}}{1+t\dot{\gamma}} \tag{4}$$

The electric field E is expressed in kV/mm.
Equations 3 and 4 may be replaced by Equation (5) accordingly with the following relationships:

$$\tau \quad = \alpha E^2 + \eta_\infty \dot{\gamma} + \beta E \frac{\dot{\gamma}}{1 + t\dot{\gamma}} \tag{5}$$

$$\tau_{yQB} = \alpha E^2 ; \eta_0 = \beta E + \eta_\infty \tag{6}$$

The α , β and η_∞ are intrinsic values of the ER fluid to be experimentallv determined. The field-dependent yield stresses of these ER fluids were experimentally obtained by $\tau_{yQB} = 27.04 \times 10^{-3} E^2$ Pa. The viscosity dynamic was experimentally obtained by $\eta_0 = 29.4E + 0.218$ for fluid ETSERF40-20. This rheological model allows for the exploration of the suitability of ER fluids to control the torsional vibrations of rotors through simulations.

2.2 ER torsional absorber design

A cylindrical type of ER torsional absorber is designed and manufactured according to the required damping force level and a critical velocity of a rotor system. Figure 1 describes the prototype of the proposed ER torsional absorber, which consists in an outer cylinder for the rotor and an inner cylinder for the absorber with the ER fluid enclosed between both cylinders. The ER fluid is composed of diatomite (called ETSERF40-20) mixed into silicone oil and is used to study the effect of ER fluids on the dynamic absorber [10] , [13].

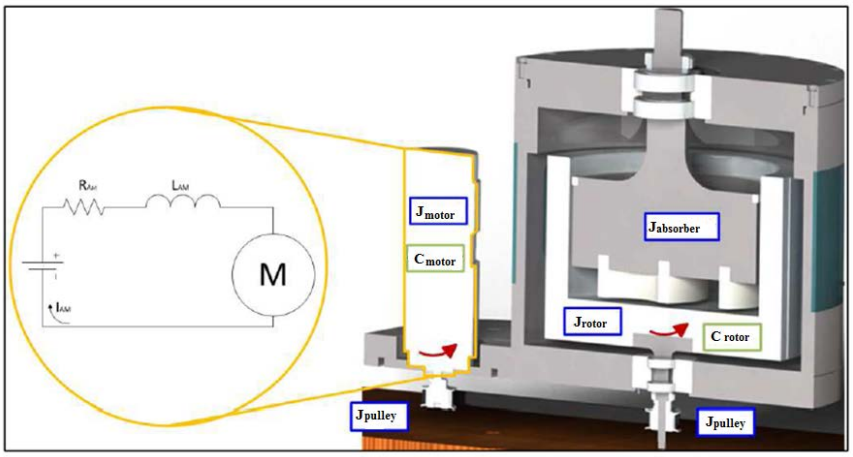


Figure 1. Overview of the torsional dynamic absorber

The positive (+) voltage is connected to inner cylinder and the negative voltage (-) connected to the outer cylinder. The proposed ER absorber is detailed accordingly with the following design parameters; electrode length (H)= 62.5mm and electrode gap (e)= 4.4mm. In the absence of electric field, the ER absorber produces a damping force only due to the fluid resistance when rotating. This damping (called C_2) of ER fluid can be calculated from equation (7) [13]. It is inherent, and is not influenced by the electric field.

$$C_2 = \frac{2\pi R^3 H \eta_\infty}{e} \tag{7}$$

R is the internal radius of outer cylinder, η_∞ is the viscosity at infinite velocity that is intrinsic value of the ER fluid to be experimentally determined. Applied to the considered fluid ETSERF40-20, we found C_2 = 0.0092 N.m.s/rad.

If a certain level of the electric field is applied to the ER absorber, the ER absorber produces an additional damping force. This damping force can be continuously tuned by controlling the intensity of the electric field. This system is then called semi-active.

2.3 Mechanical model of the smart system

2.3.1 Dynamic mode of a rotor system with the ER absorber

The vibration of the initial system may be described as a single degree-of-freedom (SDOF) damped rotating system, composed of the rotational inertia J_1 = 0.116 kg.m^2, the torsional stiffness K_1 = 801.51 N.m/rad, and the torsional damping C_1 =2.192 N.m.s/rad. The theoretical natural frequency of this primary system is 13.23 Hz with a damping rate of 11.37%. It is the amplitude close to this frequency that must be controlled.

This vibration control method in this system consists in adding another SDOF system, which is the torsional dynamic absorber. The smart ER dynamic absorber is composed of a rotational inertia J_2 = 0.0265kgm^2 driven by the ER fluid, which can thus be modeled by a torsional spring K_2, a linear viscous dashpot C_2 and a viscous dashpot C_{ER}. The resulting mechanical system is a two DOF system which is illustrated in Figure 2.

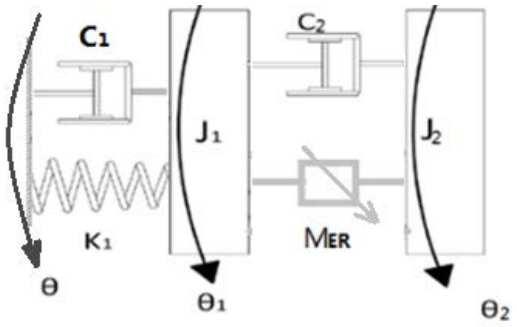


Figure 2. ER torsional dynamic absorber

The torque M_{ER} which is due to the effect of the ER fluids is composed of a nonlinear torsional stiffness torque and a nonlinear viscous torque, which depend on the electric field. It is only active when the electric field is applied. It can be expressed as:

$$M_{ER} = C_{ER}(\dot{\theta}_1 - \dot{\theta}_2) + K_{ER}(\theta_1 - \theta_2) \tag{8}$$

The total torque of the proposed ER absorber can be obtained as $M_F = C_2\dot{\theta} + M_{ER}$, where C_2 is the effective damping due to the fluid viscosity, θ is the excitation angular displacement, and M_{ER} is the field-dependent damping torque which is tunable as a function of the applied electric field.

The rotor dynamic system could be described as follows:

$$\begin{aligned} J_1\ddot{\theta}_1 &= -M_{ER} - C_1(\dot{\theta}_1 - \dot{\theta}) - K_1(\theta_1 - \theta) - C_2(\dot{\theta}_1 - \dot{\theta}_2) \\ J_2\ddot{\theta}_2 &= -C_2(\dot{\theta}_2 - \dot{\theta}_1) + M_{ER} \end{aligned} \tag{9}$$

2.3.2 Modeling of the system by state space formulation

Equation 9 can be expressed as the following equations:

$$\begin{aligned} &J_1\ddot{\theta}_1 + K_1\theta_1 + C_1\dot{\theta}_1 + C_2(\dot{\theta}_1 - \dot{\theta}_2) = K_1\theta + C_1\dot{\theta} - M_{ER} \\ &J_2\ddot{\theta}_2 + C_2(\dot{\theta}_2 - \dot{\theta}_1) = M_{ER} \\ &\ddot{\theta}_1 = \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - C_1\dot{\theta}_1 - C_2(\dot{\theta}_1 - \dot{\theta}_2)) \\ &\ddot{\theta}_2 = \frac{1}{J_2}(M_{ER} - C_2(\dot{\theta}_2 - \dot{\theta}_1)) \end{aligned} \tag{10}$$

By modifying the variables as follows:

$$\begin{aligned} &x_1 = \theta_1 - h_0\theta; \\ &x_2 = \dot{x}_1 - h_1\theta = \dot{\theta}_1 - h_0\dot{\theta} - h_1\theta \\ &x_3 = \theta_2; \\ &x_4 = \dot{x}_3 = \dot{\theta}_2 \end{aligned} \tag{11}$$

We can write:

$$\theta_1 = x_1 + h_0\theta; \dot{x}_1 = x_2 + h_1\theta \tag{12}$$

We obtain:

$$\dot{x}_1 = x_2 + h_1\theta \tag{13}$$

$$\dot{x}_2 = \ddot{\theta}_1 - h_0\ddot{\theta} - h_1\dot{\theta} = \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - C_1\dot{\theta}_1 - C_2(\dot{\theta}_1 - \dot{\theta}_2)) - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - (C_1 + C_2)\dot{\theta}_1 + C_2\dot{\theta}_2) - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}\left\{K_1\theta + C_1\dot{\theta} - M_{ER} - K_1(x_1 + h_0\theta) - (C_1 + C_2)(x_2 + h_0\dot{\theta} + h_1\theta) + C_2x_4\right\} - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}\{K_1 - K_1h_0 - (C_1 + C_2)h_1\}\theta + \left\{\frac{1}{J_1}[C_1 - (C_1 + C_2)h_0] - h_1\right\}\dot{\theta} - h_0\ddot{\theta}$$

$$+ \frac{1}{J_1}\{-M_{ER} - K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\}$$

Take $h_0 = 0$ and $\frac{1}{J_1}[C_1 - (C_1 + C_2)h_0] - h_1 = 0$, we can obtain $h_1 = \frac{C_1}{J_1}$

We have:

$$\begin{aligned} &\dot{x}_1 = x_2 + h_1\theta = x_2 + \frac{C_1}{J_1}\theta \\ &\dot{x}_2 = \frac{1}{J_1}\{K_1 - K_1h_0 - (C_1 + C_2)h_1\}\theta + \left\{\frac{1}{J_1}[C_1 - (C_1 + C_2)h_0] - h_1\right\}\dot{\theta} - h_0\ddot{\theta} \\ &+ \frac{1}{J_1}\{-M_{ER} - K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\} \\ &= \frac{1}{J_1}\left\{K_1 - (C_1 + C_2)\frac{C_1}{J_1}\right\}\theta - \frac{1}{J_1}M_{ER} + \frac{1}{J_1}\{-K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\} \\ &\dot{x}_3 = x_4 \end{aligned} \tag{14}$$

$$\dot{x}_4 = \frac{C_1 C_2}{J_1 J_2}\theta + \frac{1}{J_2} M_{ER} + \frac{1}{J_2}\{C_2 x_2 - C_2 x_4\}$$

By regrouping these equations under a matrix form, we obtain:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{K_1}{J_1} & -\frac{(C_1+C_2)}{J_1} & 0 & \frac{C_2}{J_1} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{C_2}{J_2} & 0 & -\frac{C_2}{J_2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} \frac{C_1}{J_1} & 0 \\ \frac{1}{J_1}\left\{K_1 - (C_1+C_2)\frac{C_1}{J_1}\right\} & -\frac{1}{J_1} \\ 0 & 0 \\ \frac{C_1 C_2}{J_1 J_2} & \frac{1}{J_2} \end{pmatrix} \begin{pmatrix} \theta \\ M_{ER} \end{pmatrix} \tag{15}$$

$$x = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^T$$

$$\dot{x} = \begin{pmatrix} \dot{x}_1 & \dot{x}_2 & \dot{x}_3 & \dot{x}_4 \end{pmatrix}^T$$

$$u = \begin{pmatrix} \theta & M_{ER} \end{pmatrix}^T$$

By calling these matrices as:

$$y = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \tag{16}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{K_1}{J_1} & -\frac{(C_1+C_2)}{J_1} & 0 & \frac{C_2}{J_1} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{C_2}{J_2} & 0 & -\frac{C_2}{J_2} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{C_1}{J_1} & 0 \\ \frac{1}{J_1}\left\{K_1 - (C_1+C_2)\frac{C_1}{J_1}\right\} & -\frac{1}{J_1} \\ 0 & 0 \\ \frac{C_1 C_2}{J_1 J_2} & \frac{1}{J_2} \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$$

We obtain the classic state space formulation:

$$\dot{x} = Ax + Bu$$
$$y = Cx$$

We consider a simple and linear case, in order to establish an ON-OFF control. The block diagram is shown in Fig. 3.

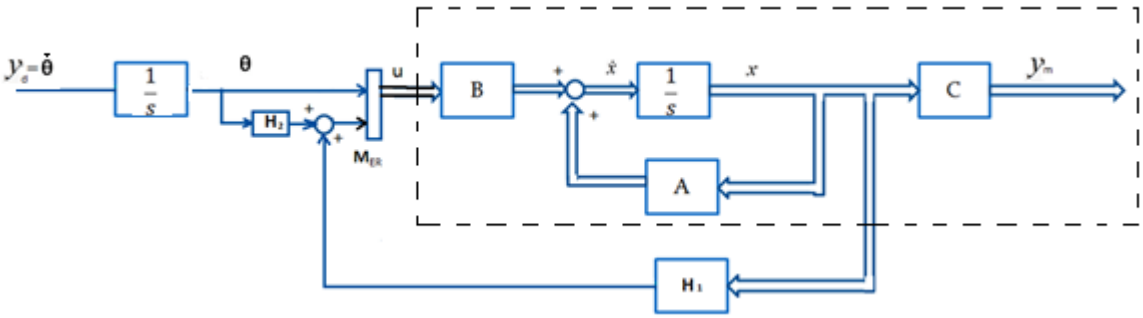


Figure 3. Block diagram du system

By defining

$$H_1 = \begin{pmatrix} K_{ER} & C_{ER} & -K_{ER} & -C_{ER} \end{pmatrix}; \quad H_2 = \frac{C_{ER}C_1}{J_1}; \tag{17}$$

The controllable torque may be expressed as:

$$\begin{aligned} M_{ER} &= K_{ER}(x_1 - x_3) + C_{ER}(x_2 - x_4 + \frac{C_1}{J_1}\theta) \\ M_{ER} &= K_{ER}x_1 - K_{ER}x_3 + C_{ER}x_2 - C_{ER}x_4 + \frac{C_{ER}C_1}{J_1}\theta = H_1 x + H_2 \theta \end{aligned} \tag{18}$$

H_1 *and* H_2 can be controlled by the electrical field by checking the difference of the angular velocity between $y_d = \dot{\theta}$ and $y_m = \dot{\theta}_m$. If the difference of the absolute value of is greater than zero, we must apply the electric field; otherwise we set the electric field to zero.

3. Numerical simulation

The control system with an Dynamic absorber (SERDA) has been numerically simulated by using Simulink (figure 4), in order to observe the reaction of the system, we use a linear chirp signal which is sine wave whose frequency varies linearly with time (figure 5), the amplitude is 1. We select E = 0 and 1 kv/mm in this simulation, the results is showed in the figure 6. We can see that amplitude of the response varies with the frequency; the amplitude is very small when the frequency is down 13Hz or more than 13Hz, but it arrives the maximum value, when the frequency is around 13Hz without electric fields and with the electric field, the amplitude decreases to about half of the original amplitude when electric fields applied is 1 kv/mm.

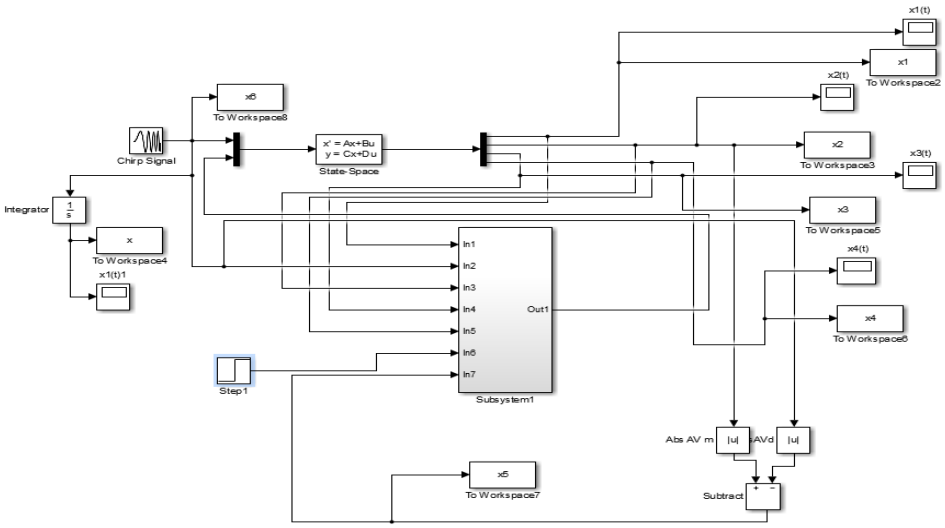


Figure 4. Block Simulink

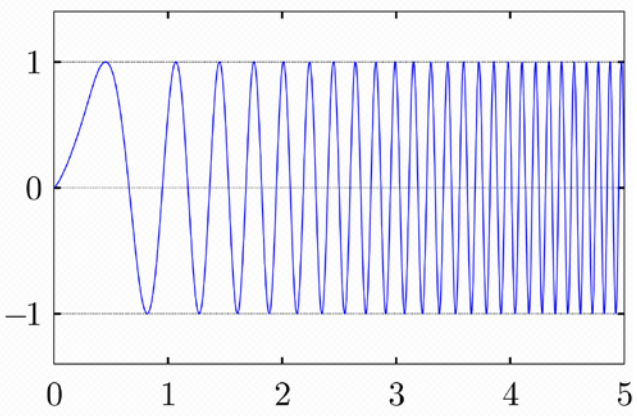


Figure 5. Linear chirp waveform

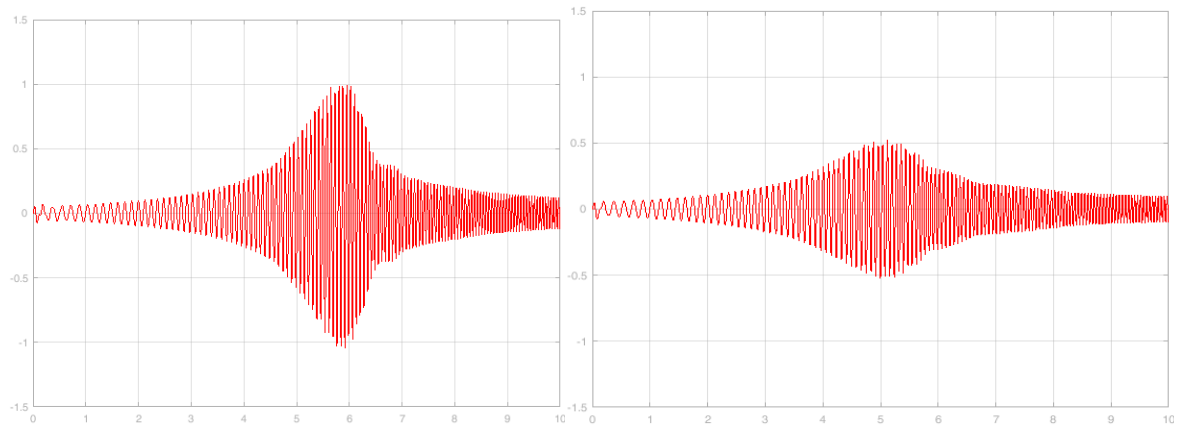


Figure 6.Reponse of system under a chirp signal obtained by simulation (left figure: E=0kV/mm; right figure : E-1kV/mm)

4. Experimental verification

In this system, the two cylinders J_1 and J_2 serve as electrodes for the ER effect in this study. The oscillation caused by the torsional vibration is measured by a triaxial accelerometer placed in the radial position as shown in Figure.6.

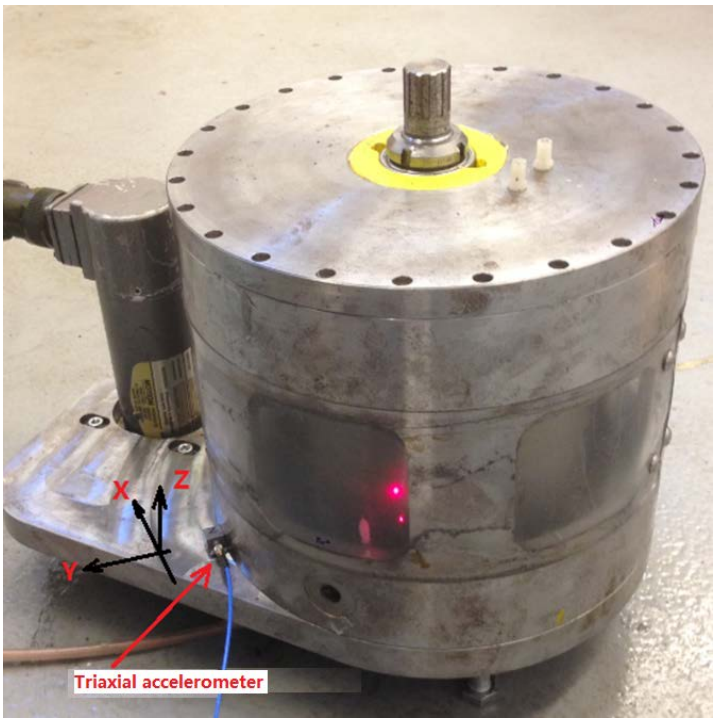


Figure 7. Electrorheological torsional absorber

We clearly see that vibration control performance in figures 8, 9 and 10, especially at 13 Hz, since the natural torsional frequency of the system is designed for this value as 13.24Hz, the critical angular velocity is 794 RPM. The two peaks as shown in Figure 8 are two natural frequencies, one is of flexible vibration, and the other is of vibration torsional. In fact, this absorber is designed only to attenuate the torsional vibration to determine the effects of the fluid ER in torsional vibration. Clearly, the torsional vibrations are attenuated in the range of 12.5 to 15Hz, but the peak which is at 12Hz does not change anything, it shows that this absorber is effective for torsional vibration, but not for flexible vibration. It can be noticed that the results of simulation are very comparable with the experimental results.

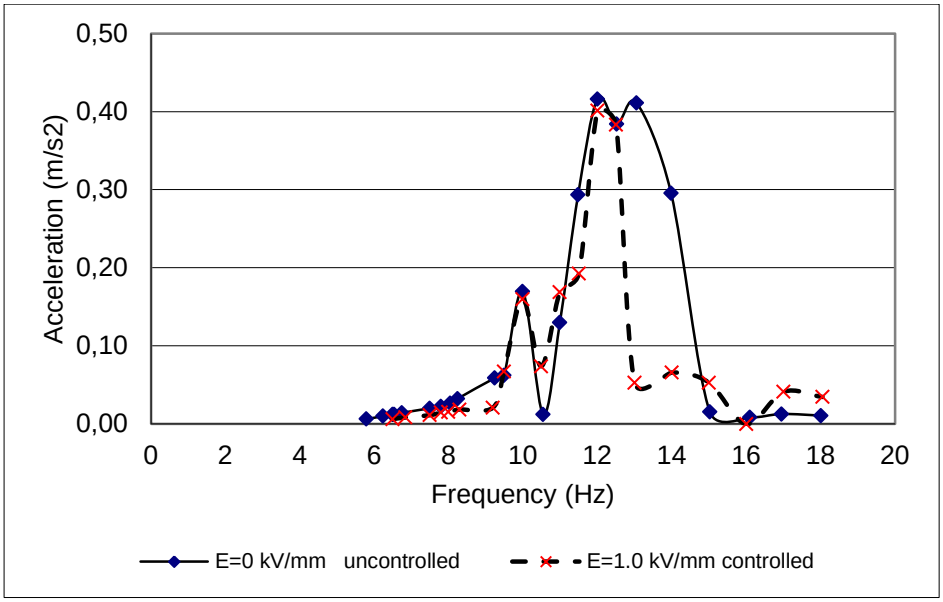


Figure 8.Response acceleration of system under different electrical field

The measurement results in time and frequency domain are shown in figure 9 and 10 for different applied electric fields. The amplitude of sinusoidal vibration decreases when the applied field increases, but not for the high electric field as E = 1.36 kv / mm since there is an optimal value of electric fields that applies on ER fluid to attenuate the torsional vibrations when a simple On-Off control strategy is chosen to realize this control system.

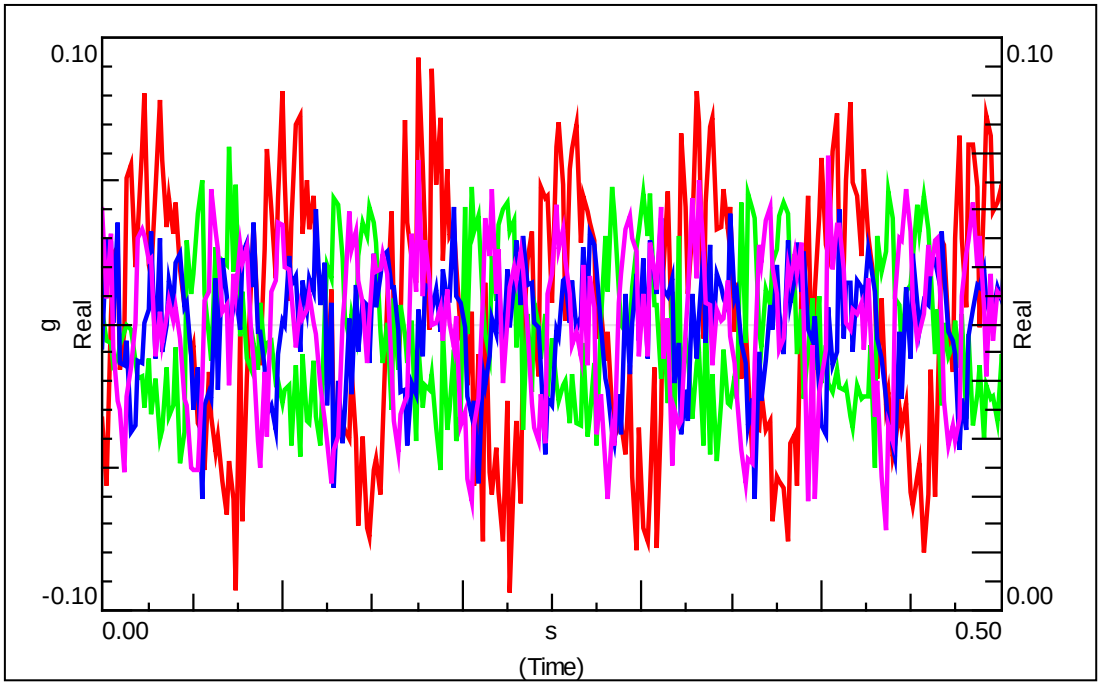


Figure 9.Temporal response of system when the rotor rotates in 780 RPM under different electrical field (red: E=0kV/mm; green: E=0.68 kV/mm; blue: E=1kV/mm; pink: E=1.36kV/mm)

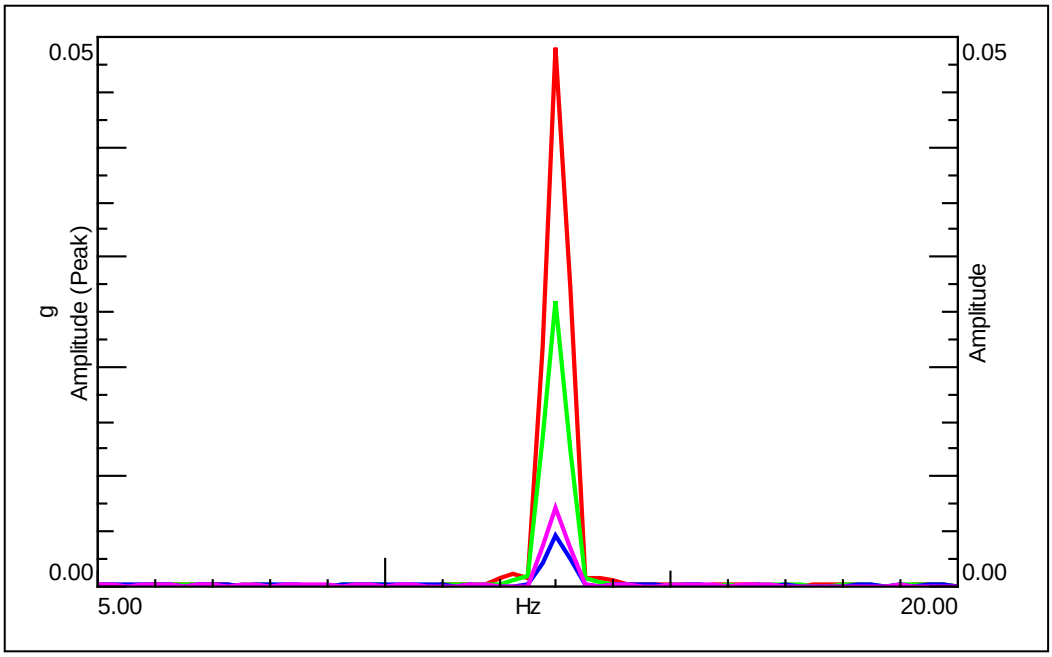


Figure 10. Frequency response of system when the rotor rotates in 780 RPM under different electrical field (red: E=0kV/mm; green: E=0.68 kV/mm; blue: E=1kV/mm; pink: E=1.36kV/mm)

5. Conclusion

This article is concerned with the application of Electrorheological Fluids (ERF) to the reduction of torsional vibrations of a rotor by controlling damping and stiffness of a rotational dynamic absorber. A cylindrical type of electrorheological (ER) torsional absorber was designed and manufactured according to the required damping force level and the critical velocity of the rotor system. This paper presents torsional vibration control performance of a smart ER dynamic absorber using a bang-bang (ON-OFF) control strategy. The absorber efficiency is measured, and the results show that the ER dynamic absorber exhibits very good performances in terms of reducing the torsional vibration of rotor system. The experimental results very closely approximate the simulation results. By use of controlled damping one can reduce (minimize) unwanted torsional vibration not only in theory but also in practice, this study shown that ER fluid are useful smart materials to realize a controllable system.

References

[1] Troy; Hill, Charles L. (2009). *Prevention Of Torsional Vibration Problems In Reciprocating Machinery*. Texas A&M University. Turbomachinery Laboratories.

[2] Den Hartog J.P. (2008), *Mechanical vibrations*, Crastre Press, ISBN 1443725366, 496 p.

[3] Winslow W.M., 1949. *Induced vibration of suspensions.* Journal of applied physics. **20**(12): p. 1137-1140.

[4] Block H. and Kelly J.P., 1988. *Review Article Electro-rheology.* Journal of Physics D: Applied Physics. **21**: p. 1661-1667.

[5] Stanway R., Sprostonz J.L., and El-Wahedz A.K., 1996. *Applications of electro-rheological fluids in vibration control: a survey*. Smart Materials and Structures. 5: p. 464-482.

[6] Abed A., Bouzidane A., Thomas M. and Zahloul H., December 2016, *Performance characteristics of a three-pad hydrostatic squeeze film damper compensated with new electrorheological valve restrictors*. Journal of Engineering Tribology (vols 208-210) 1994-1996.

[7] Bouzidane A. and Thomas M., July 2013. *Nonlinear Dynamic Analysis of a Flexible Shaft Supported by Smart Hydrostatic Squeeze Film Dampers*, Journal of Tribology (ASME), 135(3), 031701-1, 031701-9.

[8] Choi S. B., Y. M. Han, and all, 2007. *Field Test on Vibration Control of Vehicle Suspension System Featuring ER Shock Absorbers*, Electrorheological Fluids and Magnetorhelogical Suspensions, pp. 496-503 (2007)

[9] Symans M. D., M. C. Constantinou, 1999. *Semi-active control systems for seismic protection of structures: a state-of-the-art review*, Engineering Structures 21 (1999) 469–487

[10] Sun Y., Thomas M. and Masounave J., 2004. *Investigation of Dynamic Behavior of Electrorheological Fluids by Experimental Design*, Proceedings of the 7th Cansmart International workshop on Smart Materials and structures, Montreal, pp 153-162.

[11] Hao T., A. Kawai, and F. Ikazaki, 2000. *The Yield Stress Equation for the Electrorheological Fluids*. Langmuir, 16: pp. 3058-3066.

[12] Cross, M.M., 1965. *Rheology of Non Newtonian Fluids: A New Flow Equation for Pseudoplastic Systems*. Journal Colloid Science. 20 pp.417-437.

[13] Sun Y., Thomas M. and Masounave J., (2010). *A quasi-Bingham model for predicting electrorheological fluid behaviour*, Multidiscipline Modeling in Materials and Structures. **6** (1), 141-165.

Shunted piezoelectrical flextensionnal suspension for vibration insulation

K. BILLON [1], N. MONTCOUDIOL [1], A. AUBRY [2], R. PASCUAL [2], F. MOSCA [2], F. JEAN [2], C. PEZERAT[3], C. BRACAULT [3], S. CHESNÉ [1]

[1]Univ Lyon, INSA-Lyon, CNRS UMR5259, LaMCoS, F-69621, France
simon.chesne@insa-lyon.fr

[2]PYTHEAS Technology, 100 Impasse des Houllières, ZA Le Pontet, F-13590, France

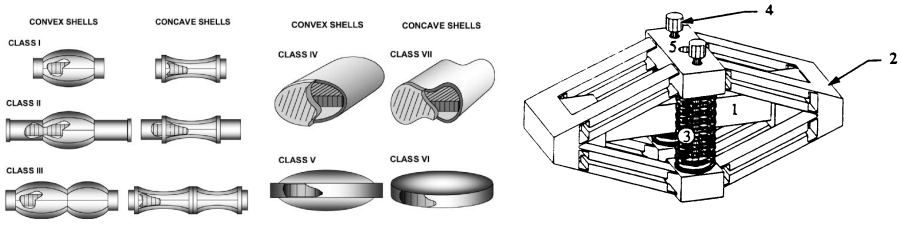
[3]Laboratoire d'Acoustique de l'Université du Mans - UMR CNRS 6613, F-72085, France

Abstract

The objective of the PyDAMP project is to develop a hybrid mechanical suspension to reduce the vibrations transmission on a wide frequency band. The undesired vibrations are generated by small electric motors (few kilograms). A suspension with piezoelectric pillar developped by PYTHEAS Technology is compared to a conventional viscoelastic suspension in terms of performances in the audible frequency range. The principle and design of the piezoelectrical suspension are approached through an electromechanical model and a finite element model. The electromechanical coupling of the transducer allows the introduction of mechanical damping and electric damping with different shunts based on resistor and negative capacitance.

1 The piezoelectrical suspension: principle and conception

The concept of suspension is inspired by a Class IV flextensionnel transducer [1, 2, 3] (figures 1). Flextensional transducers are a class of mechanical amplifiers composed of an active part, usually piezoelectric (bars, discs, rings), or magnetostrictive, and a shell that radiates in the surrounding fluid[4, 5, 6].



(a) (b)

Figure 1: a) Sketch of various classes of flextensional transducers [1, 2]. b) Sketch of a Class IV flextensional piezoelectric transducer. [3]

A finite element study has been achieved to ensure the validity of the concept in terms of maximum admissible Von Mises stress, maximum displacement and modes shapes. Figure 2 shows the CAD view of the piezoelectric suspension.

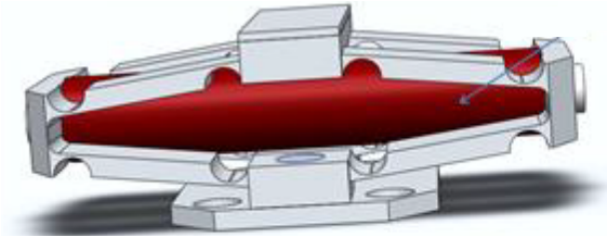


Figure 2: CAD view of the piezoelectric suspension.

2 Electrical model and simulations

An electromechanical model of the piezoelectrical suspension has been developed. Mechanical elements are converted in electrical components and an equivalent electrical circuit can be found. The simulation and the shunt optimisation are facilitated with only one physic, taking into account the whole dynamic behaviour of the piezoeletrical suspension. Figure 3 shows the schema of the piezoelectrical suspension.

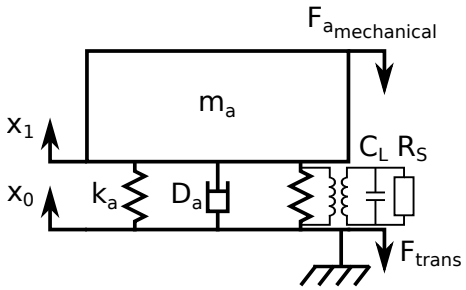


Figure 3: Schema of the piezolelectrical suspension.

The equivalent electrical circuit of the piezoelectrical suspension with mechanical excitation and resistor shunt is shown on figure 4.

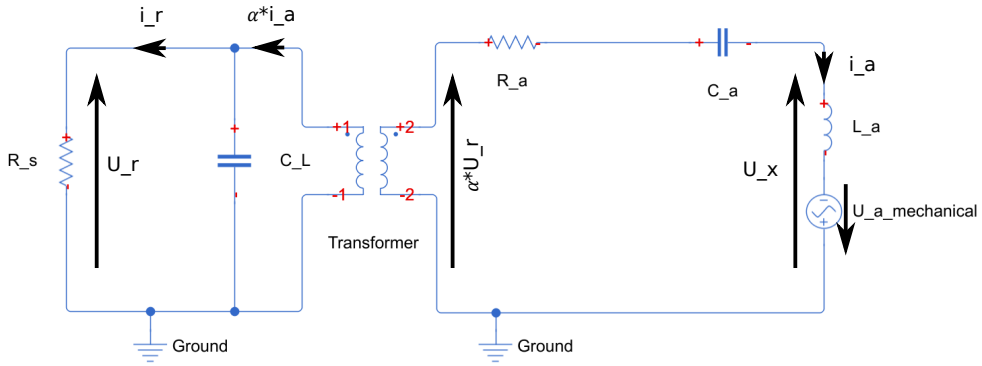


Figure 4: Equivalent electrical circuit of the piezoelectrical suspension.

Figure 5 shows the displacement transmissibility functions in open-circuited, in short-circuited conditions, with resistor shunt and with negative capacitance shunt. The damping (ξ) in open-circuited, in short-circuited conditions is similar, equal to 0.3%. As expected, the surtension is reduced by the resistor shunt, the damping (ξ) is equal to 1.6%. The negative capacitance increases the performances of the suspension to reach 2.9% of damping.

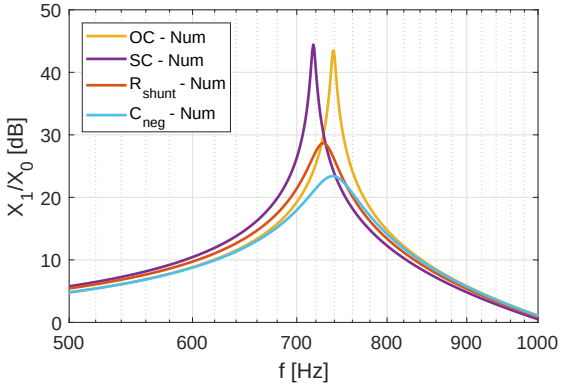


Figure 5: Transmissibility functions.

3 Comparison with a conventional viscoelastic suspension

A comparison with a conventional viscoelastic suspension is done. The conventional viscoelastic suspension is a classical spring (k_a), mass (m_a), damper (D_a). Two damping values are tested and compared with the piezoelectrical suspension namely $\xi = 0.2\%$ and 2.9%. Figures 6 show the displacement transmissibility functions with negative capacitance shunt (blue curves) compared to conventional viscoelastic suspension with damping equal to $\xi = 2.9\%$ (green curves) and 0.2% (black curves). Figures 6 show the comparison with equivalent damping. Differences can be observe in high frequencies from 10 kHz where the slope moves from $-40\ dB/dec$ to $-20\ dB/dec$ for the conventional viscoelastic suspension. For the piezoelectric suspension, the change appears a decade later. For audible perturbation between 10 kHz and 20 kHz, in this case the difference is between 2 dB and 6 dB. Figures 6 show the comparison where the slope moves from $-40\ dB/dec$ to $-20\ dB/dec$ at the same frequency. Differences can be observe around 700 Hz, for the conventionnal suspension, the maximum transmission level is twice as high compared to the piezoelectrical suspension.

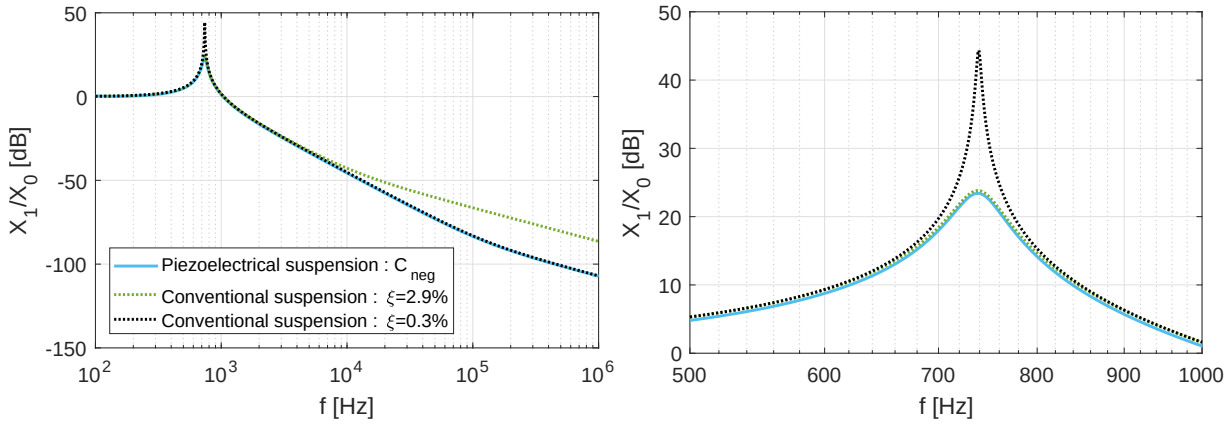


Figure 6: a) Comparison between the transmissibility functions of a conventional viscoelastic suspension and the piezoelectrical suspension. b) Zoom.

4 Experimental setup

Figure 7 shows the experimental setup. Two accelerometers are used to obtain the transmissibility functions. The mechanical excitation is provided by a shaker. The piezoelectrical suspension is glued on the shaker. Different electrical conditions can be applied on the stack. Experimental results are

under post-processing.

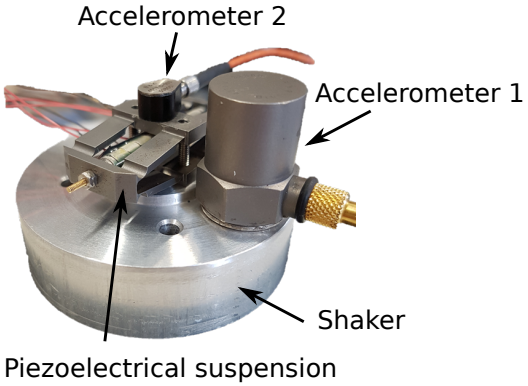


Figure 7: Experimental setup

5 Conclusion

A suspension with piezoelectric pillar developed by PYTHEAS Technology has been studied. An electromechanical model and a finite element model allowed the conception and the design of the suspension. A resistor shunt optimisation has been performed and good results were observed in order to reduce the surtension in the displacement transmissibility. Performances have been increased using a negative capacitance shunt. A comparison with conventional suspension has been conducted. For the piezoelectric suspension, no static problems are observed and there is no creep behaviour. The slope modification from $-40\ dB/dec$ to $-20\ dB/dec$ appears a decade later and is not sensitive to damping. It could be important for audible applications.

References

[1] J. Butler and C. Sherman, *Transducers and arrays for underwater sound*, Springer, 2016.

[2] D. Jones and N. Abboud, "Performance analysis of a low-frequency barrel-stave flextensional projector," in *Proceeding of ONR Transducer Materials and Transducers Workshop*, pp. 1–4, 1996.

[3] N. Hagood, J. Aldrich, and A. V. Flotow, *Design of passive piezoelectric damping for space structures*, National Aeronautics and Space Administration, Langley Research Center, 1994.

[4] B. Damiri, *Modèle analytique pour l'analyse du transducteur flextensionnel de classe IV dans l'air*. PhD thesis, Lille 1, 1996.

[5] G. Brigham and L. Royster, "Present status in the design of flextensional underwater acoustic transducers," *The Journal of the Acoustical Society of America* **46**(1A), pp. 92–92, 1969.

[6] L. Royster, "Predicting the effective dynamic coupling coefficient of the flextensional transducer (preliminary design considerations)," *The Journal of the Acoustical Society of America* **38**(5), pp. 879–880, 1965.

Exploring periodicity and dispersion diagrams in muffler design

Vinícius D. Lima[1], Belisário N. Huallpa[2] and José R. F. Arruda[1]

[1]University of Campinas, Rua Mendeleyev, 200, Campinas, SP, Brazil
vinidiaslima@gmail.com, arruda@fem.unicamp.br
[2]Universidade Federal de Lavras, Lavras, Aquenta Sol, MG, Brazil
belisario.nina@deg.ufla.br

Abstract

Periodic waveguides may be analyzed using dispersion diagrams, which plot the wavenumbers as functions of frequency. Imaginary wavenumbers mean propagation is not possible and, therefore, normal modes cannot build up. Muffler design has traditionally explored periodicity, but usually not using dispersion analysis. In this work, we show how to model one-dimensional acoustic waveguides with plane wave assumption using spectral elements (SE), how to obtain dispersion diagrams and, using semi-infinite elements, transmission loss from an SE model. The technological challenge consists in opening band gaps at low frequencies with a limited size muffler, and SE models are handy for low cost parameter optimization. For arbitrary shapes, this work uses scaled SE models or, alternatively, a state-space formulation recently developed by the authors. Additive manufacturing is an enabling technology for the implementation of the designed mufflers. In this work, we show experimental results for simple periodic mufflers built using 3D printing. The proposed simulation methodology is simple and can be used for quick design of 3D-printed polymer mufflers.

1 Introduction

Acoustic mufflers are usually characterized by their insertion loss (IL) or transmission loss (TL) [1]. The latter is a more absolute characterization, as it does not depend upon the acoustic impedance of the system connected to the muffler. Mufflers can be reactive, active of hybrid depending on the sound attenuation mechanism. Reactive mufflers do not need to dissipate energy and attenuate using acoustic impedance discontinuities and acoustic resonators that cause destructive interference. Active mufflers attenuate propagating sound by energy dissipation. Mufflers may combine these two mechanisms. The design of reactive sound mufflers usually start from a chosen geometry combining Helmholtz resonators, quarter length tubes and expansion chambers and adjusting their parameters to maximize the TL at desired frequency ranges.

In recent years, inspired by the scientific advances in photonics [2], researchers started using wave dispersion diagrams of periodic elastic systems to investigate the existence of frequency stop bands created by destructive interference (Bragg scattering) and local resonance due to resonators. This is exactly the effect desired in reactive mufflers. Therefore, in this work, we investigate the use of dispersion diagrams to design periodic reactive mufflers.

Dispersion diagrams are usually obtained using the Plane Wave Expansion (PWE) method [3]. Otherwise, given a transfer matrix of a periodic cell, the Floquet-Bloch theorem can be used to obtain the wavenumbers as a function of frequency. In this work we derive a spectral acoustic element to model the linear acoustics of ducts using the plane wave assumption. It is straightforward to derive a semi-infinite acoustic duct element. Using finite and semi-infinite spectral elements a model can be easily built to compute the TL and simulate the forced responses.

We first introduce the acoustic spectral elements, show how to assemble these elements, how to reduce the global matrix and transform it into a transfer matrix, and, finally, how to compute the dispersion diagram. Then

we show how to assemble a duct system consisting of source, uniform duct, periodic muffler and semi-infinite duct that allows computing the TL. We show that with a small number of periodic cells a high TL is achieved. Finally, experimental results obtained in an impedance tube for a 3D-printed periodic muffler are shown.

2 Analytical Formulation

2.1 Acoustic spectral element

The derivation of the acoustic duct spectral element is analogous to the spectral element for elementary rods [4]. Assuming plane wave propagation, the spectral element for an acoustic waveguide starts from the one-dimensional non-dissipative wave equation [5]

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1}$$

where p is the acoustic pressure at a position x along the waveguide and c is the speed of sound. In the frequency domain, Eq. 1 becomes an ordinary differential equation

$$\frac{\partial^2 \hat{p}}{\partial x^2} + \frac{\omega^2}{c^2}\hat{p} = 0 \tag{2}$$

Equation 2 has a wave-based solution form given by:

$$\hat{p} = C_1 e^{-ikx} + C_2 e^{-ik(L-x)} \tag{3}$$

where k is the wavenumber. For low sound levels, the acoustic pressure and the particle velocity u are related by the linear Euler's equation [5]

$$\rho\frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x} \tag{4}$$

where ρ is the mass density of the fluid. Replacing the solution given by Eq. 3 into Eq. 4 and rearranging in terms of the volume acceleration $\hat{v} = S\partial u/\partial t$ yields

$$\hat{v} = \frac{S}{\rho c}(-C_1 e^{-ikx} + C_2 e^{-ik(L-x)}) \tag{5}$$

where S is the cross-sectional area.

Considering a straight acoustic waveguide with length L, constants C_1 and C_2 can be determined for given boundary conditions at $x = 0$ and $x = L$. Finding the solution for these constants yields the following matrix system [4] [6]:

$$\begin{bmatrix}\hat{V}_l\\ \hat{V}_r\end{bmatrix} = \frac{ikS}{\rho(1-e^{-2ikL})}\begin{bmatrix}1+e^{-2ikL} & -2e^{-ikL}\\ -2e^{-ikL} & 1+e^{-2ikL}\end{bmatrix}\begin{bmatrix}\hat{p}_l\\ \hat{p}_r\end{bmatrix} = \begin{bmatrix}D_{ll} & D_{lr}\\ D_{rl} & D_{rr}\end{bmatrix}\begin{bmatrix}\hat{p}_l\\ \hat{p}_r\end{bmatrix} = \mathbf{D}\begin{bmatrix}\hat{p}_l\\ \hat{p}_r\end{bmatrix} \tag{6}$$

where $\mathbf{D}$ is the spectral element matrix for the acoustic waveguide.

2.2 Dispersion diagram

An acoustic tube with periodically varying cross section ca be called a phononic crystal [2]. The band structure of a phononic crystal may be represented by a dispersion diagram, i.e., a plot of the wavenumber versus frequency (or vice-versa). A transfer matrix relates the acoustic pressure and the volume acceleration at the left and right ends of a periodic cell as

$$\begin{bmatrix}\hat{p}_r\\ -\hat{V}_r\end{bmatrix} = \mathbf{T}\begin{bmatrix}\hat{p}_l\\ \hat{V}_l\end{bmatrix} = \begin{bmatrix}-D_{ll}D_{lr}^{-1} & -D_{lr}^{-1}\\ D_{rl} - D_{ll}D_{rr}D_{lr}^{-1} & -D_{rr}D_{lr}^{-1}\end{bmatrix}\begin{bmatrix}\hat{p}_l\\ \hat{V}_l\end{bmatrix} \tag{7}$$

The transfer matrix of a periodic system with n cells can be obtained from the matrices of each cell $\mathbf{T_1}, \mathbf{T_2}, ..., \mathbf{T_{n-1}}, \mathbf{T_n}$ as

$$\mathbf{T} = \mathbf{T_n}\mathbf{T_{n-1}}...\mathbf{T_2}\mathbf{T_1} \tag{8}$$

The Floquet-Bloch theorem states the relation between the state vector at the right and left boundaries of the unit cell as

$$\begin{bmatrix} \hat{p}_r \\ -\hat{V}_r \end{bmatrix} = e^{-ikL} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} \tag{9}$$

Equations 7 and 9 yield the following eigenproblem:

$$\mathbf{T} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} = e^{-ikL} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} = \lambda \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} \tag{10}$$

The solution of the eigenproblem in Eq. 10 gives the normalized wavenumber kL for each angular frequency ω using the relation

$$kL = i\ln(\lambda) \tag{11}$$

2.3 Transmission loss

In this section we recall the two-load method for computing the TL with an impedance tube. The loudspeaker is positioned at the left extremity, and measurements are performed with two different terminations at the right end. The sample is positioned at the central position, with two microphones measuring the pressures at each side. With the positions of the microphones and the pressures, the amplitudes of the transmitted and reflected waves can be determined by

$$A = i\sqrt{G_{rr}} \frac{H_{1r}e^{ikx_2} - H_{2r}e^{ikx_1}}{2\sin k(x_1 - x_2)} \tag{12a}$$

$$B = i\sqrt{G_{rr}} \frac{H_{2r}e^{-ikx_1} - H_{1r}e^{-ikx_2}}{2\sin k(x_1 - x_2)} \tag{12b}$$

$$C = i\sqrt{G_{rr}} \frac{H_{3r}e^{ikx_4} - H_{4r}e^{ikx_3}}{2\sin k(x_3 - x_4)} \tag{12c}$$

$$D = i\sqrt{G_{rr}} \frac{H_{4r}e^{-ikx_3} - H_{3r}e^{-ikx_4}}{2\sin k(x_3 - x_4)} \tag{12d}$$

where G_{rr} is the autospectrum of the loudspeaker signal (there is no need to use the actual loudspeaker velocity, the voltage input signal is sufficient) and H_{nr} and x_n are the frequency response functions (FRF) for input r and output pressure measured with microphones $n = 1,2,3,4$. The pressures and particle velocities at the sample extremities are given by

$$p_0 = A + B \tag{13a}$$

$$u_0 = \frac{(A - B)}{\rho c} \tag{13b}$$

$$p_d = Ce^{-jkd} + De^{-jkd} \tag{13c}$$

$$u_d = \frac{Ce^{-jkd} - De^{-jkd}}{\rho c} \tag{13d}$$

Denoting with subscripts a and b the measurements with two different tube terminations (e.g. open and closed), the transfer matrix of the sample can be determined by

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \frac{1}{p_{da}v_{db} - p_{db}v_{da}} \begin{bmatrix} p_{0a}v_{db} - p_{0b}v_{da} & p_{0b}p_{da} - p_{0a}p_{db} \\ v_{0a}v_{db} - v_{0b}v_{da} & p_{da}v_{0b} - p_{db}v_{0a} \end{bmatrix} \tag{14}$$

The transmission loss of the sample is obtained by:

$$TL = 20\log_{10}\left|\frac{1}{2}\left(T_{11} + \frac{T_{12}}{\rho c} + \rho c T_{21} + T_{22}\right)\right| \tag{15}$$

In the cases where the sample is symmetric, another two conditions can be set, and only one load is necessary to determine the transfer matrix. These conditions are the reciprocity and symmetry given by

$$T_{11} = T_{22} \tag{16a}$$

$$T_{11}T_{22} - T_{12}T_{21} = 1 \tag{16b}$$

With the conditions of Eqs. 16 the transfer matrix is determined by:

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \frac{1}{p_0 v_d + p_d v_0} \begin{bmatrix} p_d v_d + p_0 v_0 & p_0^2 - p_d^2 \\ v_0^2 - v_d^2 & p_d v_d + p_0 v_0 \end{bmatrix} \tag{17}$$

where subscripts a and b are no longer necessary.

2.4 Insertion loss

The insertion loss is the difference between the acoustic power transmitted with and without the muffler, denoted by the subscripts 2 and 1, respectively

$$IL = L_{W_1} - L_{W_2} = 10 \log \left(\frac{W_1}{W_2} \right) \tag{18}$$

If the temperature gradient is negligible and the source pressure is constant [1] it becomes simply

$$IL = 20 \log \left| \frac{p_1}{p_2} \right| \tag{19}$$

where p_1 is the pressure at the outlet without the muffler and p_2 with the muffler.

3 Numerical results

Using the spectral element previously derived, the periodic muffler can be modeled by assembling the element matrices by imposing continuity at the element connections. Two connected elements are assembled as

$$\mathbf{D} = \begin{bmatrix} D_{ll}^{(1)} & D_{lr}^{(1)} & 0 \\ D_{rl}^{(1)} & D_{rr}^{(1)} + D_{ll}^{(2)} & D_{lr}^{(2)} \\ 0 & D_{rl}^{(2)} & D_{rr}^{(2)} \end{bmatrix} \tag{20}$$

After assembling all the elements composing the periodic cell, the resulting matrix can be condensed to the nodes at the two ends, resulting in a 2 x 2 matrix

$$\bar{\mathbf{D}} = \begin{bmatrix} D_{11} - D_{21} D_{22}^{-1} D_{12} & -D_{23} D_{22}^{-1} D_{12} \\ -D_{21} D_{22}^{-1} D_{32} & D_{33} - D_{32} D_{22}^{-1} D_{23} \end{bmatrix} \tag{21}$$

which can be transformed into a transfer matrix using Eq. 6 , in terms of $\bar{\mathbf{D}}$

$$\mathbf{T} = \begin{bmatrix} -\bar{D}_{12}^{-1} \bar{D}_{11} & \bar{D}_{12}^{-1} \\ -\bar{D}_{21} + \bar{D}_{22} \bar{D}_{12}^{-1} \bar{D}_{11} & -\bar{D}_{12}^{-1} \bar{D}_{22} \end{bmatrix} \tag{22}$$

From the transfer matrix of the periodic cell the dispersion relation is computed as previously described. It is shown in Fig. 2. Furthermore, using Eq. 15, the TL can be computed, as shown in Fig. 1. To compute the FRFs for a volume acceleration input at one end and pressure outputs at the two ends matrix $\bar{\mathbf{D}}$ is inverted and multiplied by vector $\hat{\mathbf{V}} = [1\ 0]^T$.

4 Experimental results

The two loads used in the impedance tube were a rigid and a nearly anechoic termination, implemented using absorptive foam in the open tube end. Figure 4 shows the experimental setup.

The sample is a periodic muffler built using 3D printing (polyamide and selective laser sintering). Figure 5 shows the sample manufactured for the experiment.

Measurements were made with a commercial impedance tube with four-microphones. The loudspeaker was driven with a Gaussian noise signal approximately white (flat spectrum) in the frequency range from 0 Hz to 5272 Hz. A single roving microphone was used to avoid the phase calibration required in the four-microphone measurement [7].

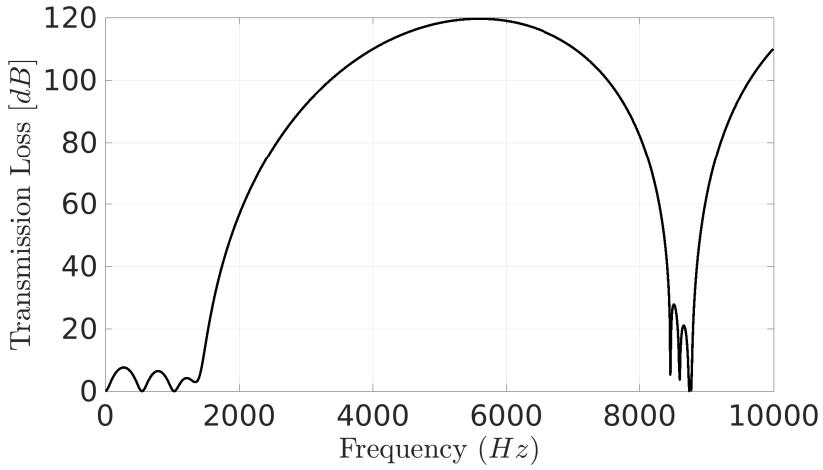


Figure 1: Simulated TL.

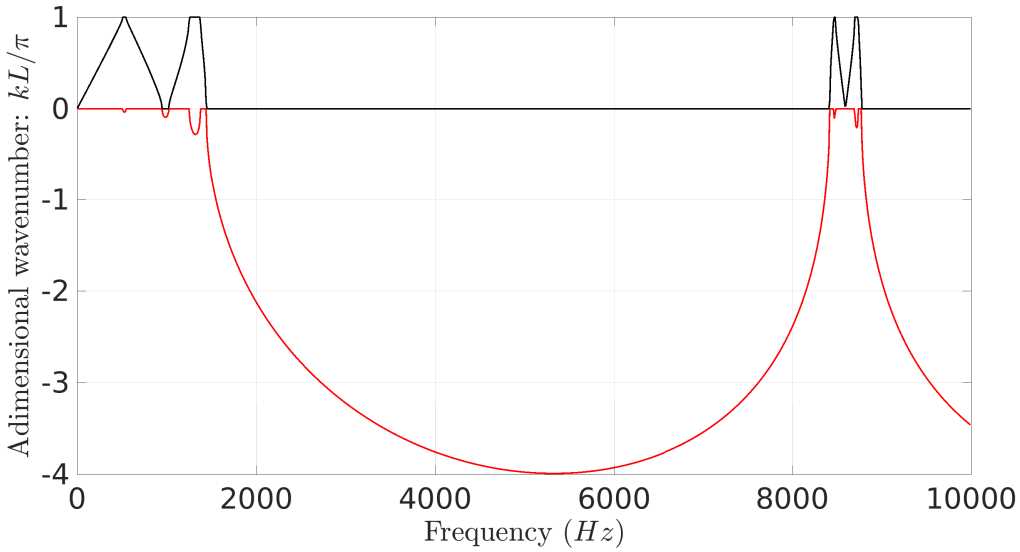


Figure 2: Simulated dispersion diagram.

Figure 6 shows a comparison between simulated and experimental results for the TL. The one-load method, allowed by the symmetric muffler geometry in our case, yielded better results. Experimental background noise did not allow to obtain TL values above approximately 20 dB.

Figure 7 shown a comparison between the predicted and measured IL. The measurements are less noisy, but again background noise did not allow obtaining IL values larger than 20 dB. Measurements inside an anechoic chamber are under way to try to get measurements with lower background noise in the case of the open tube.

5 Conclusions

We have shown how to use dispersion diagrams to predict band gaps in periodic mufflers, which result in large transmission losses. Using acoustic spectral elements it is straightforward to obtain the muffler transfer matrix, from which the dispersion diagram and TL can be computed. For arbitrary shapes, the periodic cell can be computed by discretizing the cell with uniform spectral elements. Otherwise, as shown by the authors [8], a state-space formulation and Riccati-type equation for the acoustic impedance can be used. Optimized mufflers can be constructed using 3D printing. Different optimization strategies are under investigation by the authors. The tools shown in this work set a framework for periodic muffler design optimization using dispersion relations.

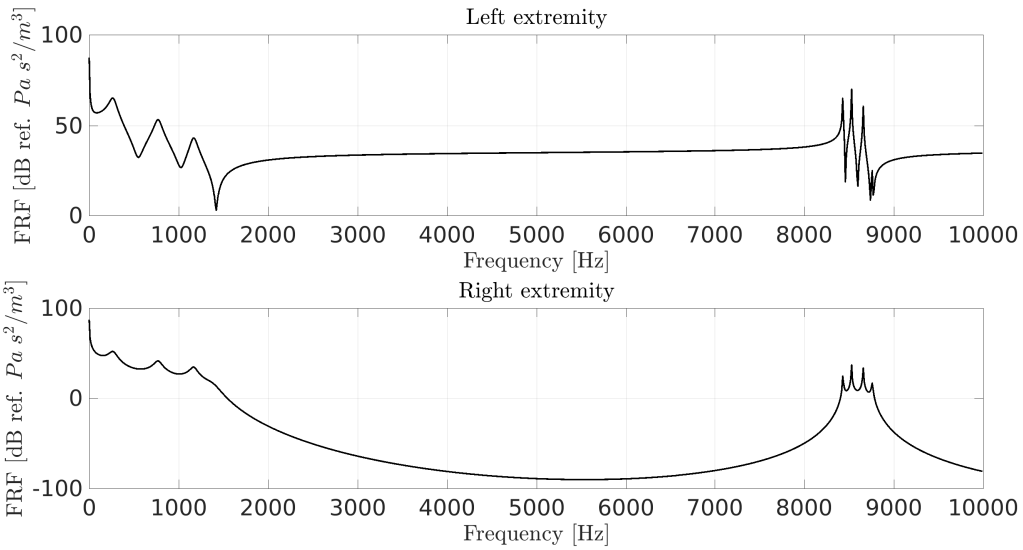


Figure 3: Simulated FRFs.

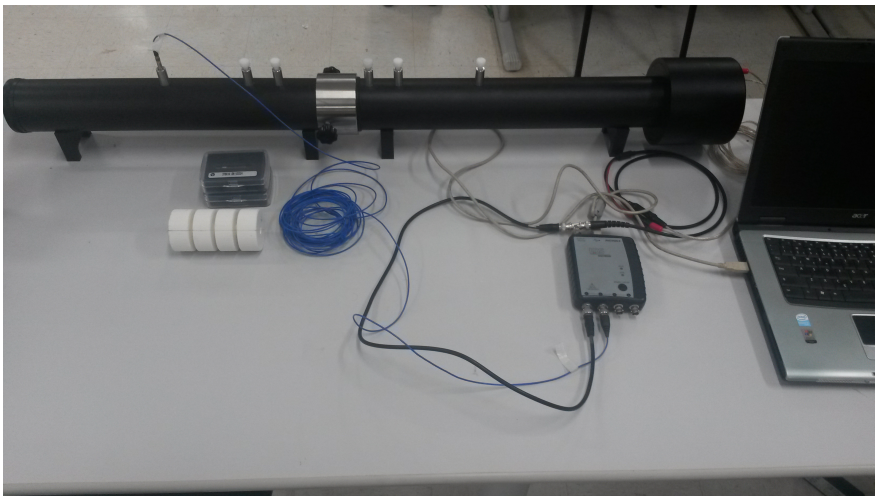


Figure 4: Experimental setup.

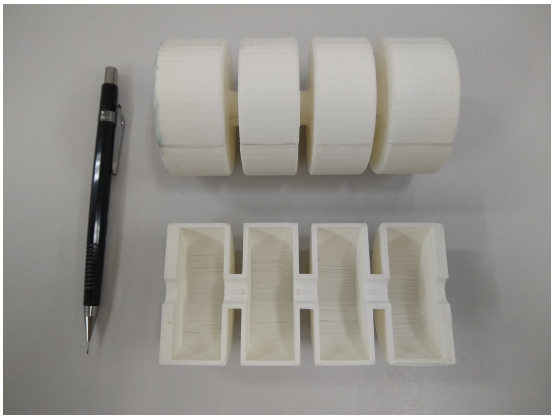


Figure 5: Manufactured muffler.

Acknowledgements

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superiorl (CAPES Finance Code 001) and the São Paulo Research Foundation (FAPESP 2018/15894-0) for the financial support.

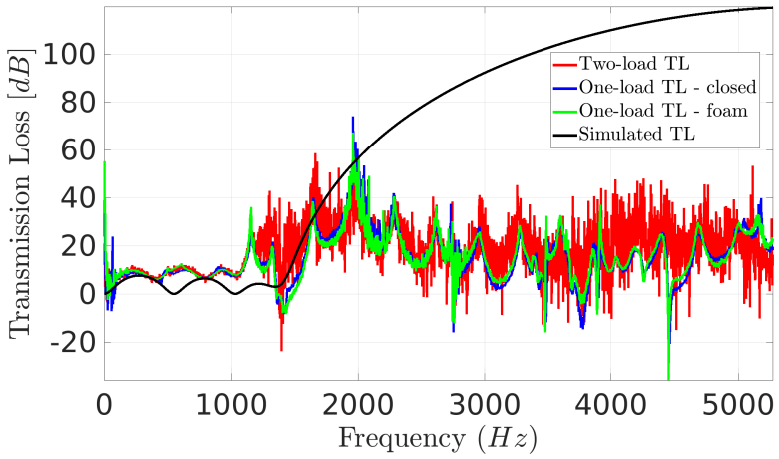


Figure 6: Simulated and experimental transmission loss.

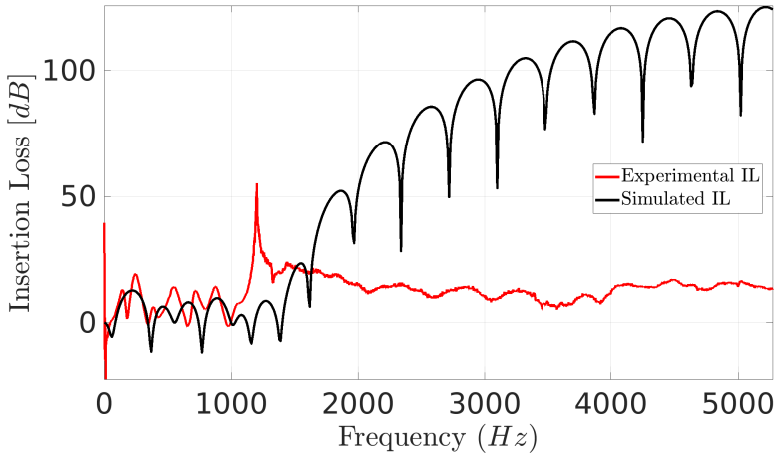


Figure 7: Simulated and experimental insertion loss.

References

[1] M.L. Munjal, *Acoustic of ducts and mufflers with application to exhaust and ventilation system design*, John Wiley and Sons, Canada, 1987.

[2] M.I. Hussein, M.J. Leamy and M. Ruzzene, *Dynamics of phononic materials and structures: Historical origins, recent progress, and future outlook*, Applied Mechanics Reviews, American Society of Mechanical Engineers, 2014, V. 66, n. 4, pp.1-4.

[3] M.M. Sigalas and E.N. Economou, *Elastic and acoustic wave band structure*, Journal of Sound and Vibration, 1992, V. 158, n. 2, pp. 337-382.

[4] J.F. Doyle, *Wave Propagation in Structures: Spectral Analysis Using Fast Discrete Fourier Transforms*, Mechanical Engineering Series, Springer New York, USA, 1997.

[5] L.E. Kinsler, A.R. Frey, A.B. Coppens and J.V. Sanders, *Fundamentals of acoustics*, John Wiley and Sons, USA, 2000.

[6] M. I. N. Rosa, V. D. Lima, J. M. C. Santos and J. R. F. Arruda, *Numerical and Experimental Investigation of Interface Modes in Periodic Acoustic Waveguides*, Proceedings of the International Conference on Structural Engineering Dynamics, Viana do Castelo, Portugal, 24-26 June 2019, 10 pages.

[7] B.H. Song and J.S. Bolton, *A transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials*, Journal of the Acoustical Society of America, 2000, V. 107, n. 3, pp.1131-1152,

[8] G.F.C.A. Assis, D. Beli, E.J.P. Miranda Jr., J.F. Camino, J.M.C. dos Santos and J.R.F. Arruda, *Wave Propagation in Structures: Spectral Analysis Using Fast Discrete Fourier Transforms*, International Journal of Mechanical Sciences (submitted), 2019.

Wednesday

Angular approaches

Angular vibration on-site measurements and application to torsional analysis on industrial cases

François COMBET, Louis BOISSELEAU

DYNAE, rue Condorcet, 38090 Villefontaine, France
francois.combet@eiffage.com

Abstract

The measurement and analysis of torsional phenomena remains an uncommon and challenging task to perform in the industry. The measurement of the torque via strain gauges provides explicit results but can be difficult to implement on-site and is generally limited to a low rotation speed. The motor current measurement is easier to perform but does not always reflect the torque instantaneous variations. The measurement of the Instantaneous Angular Speed (IAS) presents an interesting alternative as several techniques exist which are relatively easy to install on-site. This is classically performed using optical encoders or magnetic pickup sensors. A lesser known technique is based on Laser Torsional Vibrometry (LTV), using parallel beam laser vibrometers, which has the advantage of being totally non-intrusive. One difficulty however of IAS based diagnosis techniques is then to interpret the IAS amplitude, due to the lack of rules and criteria in this domain, on contrary to translational vibration. Advantages of these different techniques for IAS on-site measurements are discussed, and applications are then presented on industrial cases: the test and certification of the coupling of a fuel injection pump and the torsional analysis of the flexible coupling of a Diesel-generator group.

1 Introduction

The measurement and analysis of torsional phenomena remains an uncommon and challenging task to perform in the industry. The measurement of the transmitted torque via strain gauges provides explicit results but can be difficult to implement on-site and is generally limited to a low rotation speed. The motor current measurement is an easier technique to perform but was shown to not always reflect the torque instantaneous variations, due to the filtering effect of the shaft line inertia on the higher frequency torque & speed variations. The measurement of the instantaneous angular speed (IAS) presents an interesting alternative as several techniques exist which are relatively easy to install on-site. This is classically performed using optical encoders or magnetic pickup sensors. The analysis of the IAS variations was shown to allow the detection of roller bearing faults [3-4]. Another technique is based on Laser Torsional Vibrometer (LTV), using parallel beam laser vibrometers, which has the advantage of being totally non-intrusive [5-6].

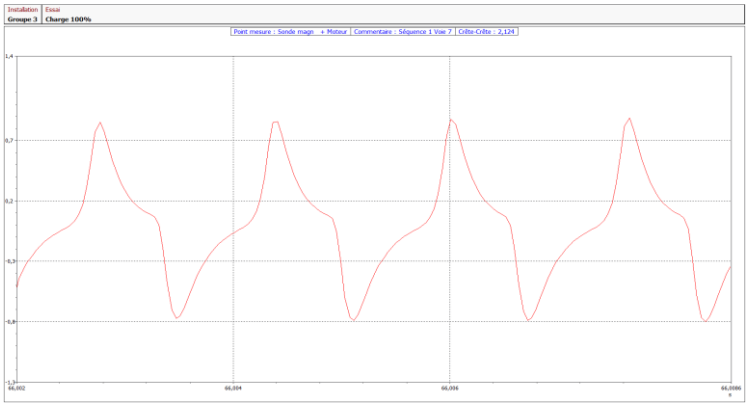
We first discuss the advantages of different techniques for IAS on-site measurements. We then present their application on industrial cases: first the test and certification of the coupling of a fuel injection pump, then the torsional analysis of a flexible coupling of Diesel-generator group.

2 Angular vibration measurement techniques

2.1 Use of impulse sensors (optical encoders & magnetic pickups)

The use of impulse sensors for IAS measurement from strip bands or optical encoders is quite common and well documented. The computation of the IAS can then be performed by the elapsed time method [8].

The IAS can also be measured by a magnetic pickup. In this case the computation is preferably done by frequency demodulation, as the measured signal is non-sinusoidal and modulated in amplitude (depending on the rotation speed).

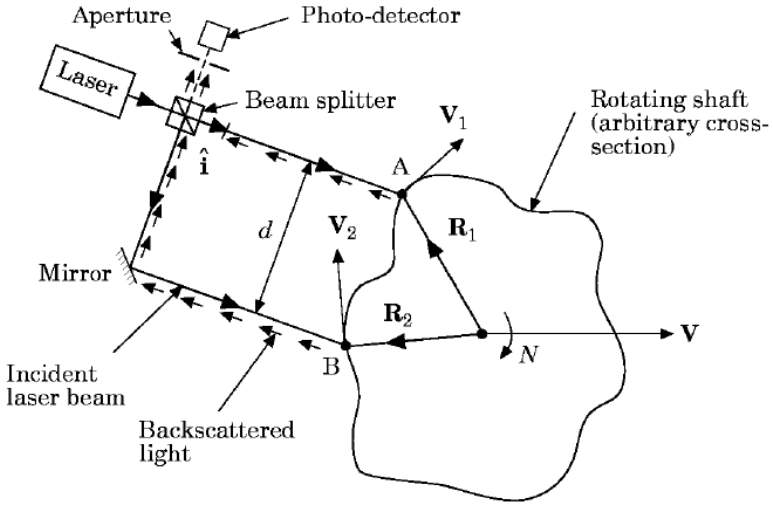


Signal measured by a magnetic pickup

2.2 The Laser Torsional Vibrometer

The Laser Torsional Vibrometer (LTV) is another technique that offers *in situ* measurement, thus avoiding machinery downtime and can function on rotating components of arbitrary shape. The optical geometry used makes the instrument insensitive to solid body oscillation of the target or operator as well as to the cross-sectional shape of the component [5].

The laser beam with wavelength λ is divided into two equal intensity parallel beams separated by distance *d*:



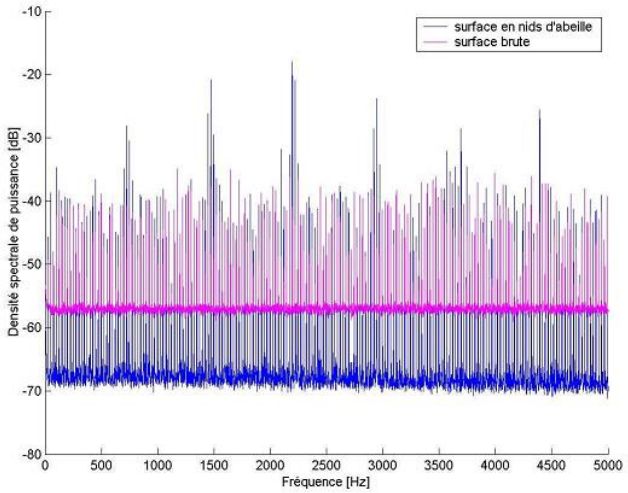
The laser torsional vibrometer optical geometry (picture taken from [5])

The ‘beat’ frequency between the two backscattered light beams received by the photo-detector corresponds to a Doppler frequency *fD*, which is directly proportional to the speed of rotation Ω of the target component:

$$fD=(2d/\lambda)\Omega$$

the fluctuating part of which is the IAS variation. The frequency response of the instrument is dictated by that of the demodulation system used and the usual bandwidth of practical interest is up to 10 kHz.

In practice however this technique is limited by the Signal / Noise ratio of the signal, which is influenced by a run-out phenomenon induced by the target surface roughness. This creates a speckle pattern periodicity at the shaft rotation. For use at very low levels of torsional vibration the speckle pattern periodicity can be attenuated by modulating the spatial position of the incident laser beams from side to side in a random manner.



Train of harmonic components induced by the run-out phenomenon

It should be noted that the setting and calibration of the LTV must be performed *in situ*, i.e. while the shaft is rotating, at least at a low speed.

Note that the LTV also offers the possibility of successive IAS measurement on different sections of the shaft line while the equipment is running. If a phase reference (tachymeter) is used, a torsional Operational Deformed Shape (ODS) of the shaft line can thus be performed.

2.3 In search of criteria and guiding rules for IAS assessment

One difficulty when dealing with angular vibrations is the lack of criterion in the literature in order to assess the maximum allowed amplitude of the measured IAS.

For combustion engines there is an industry regulation for the allowable magnitude of the Peak-Peak twist angle of the crankshaft, which must be below **0.4° PP** [7]. Some constructors have also defined their own criteria, depending on their experience. Note that Pr Nerubenko underlines in [7] that a *torsional vibration* is an angular vibration which implies a *twist* of the shaft, and must not be confounded with the *fluctuation of the rotation speed*, or the RPM non-regularity, which is another dynamical problem. The latter may be defined by the coefficient:

$$C = \Omega\_Peak\text{-}Peak / \Omega\_mean$$

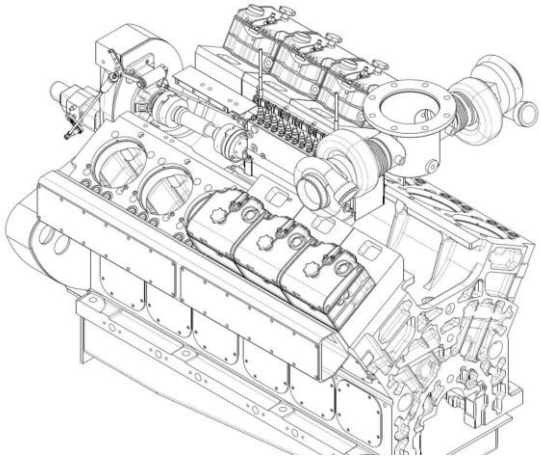
For combustion engines the guide numbers for coefficient *C* are in the range **0.6–1.2%**. Apart from the combustion engine industry we cannot seem to find any other criterion for the angular vibrations.

3 Industrial case studies

We present here applications of IAS measurement techniques to the torsional analysis of industrial equipments on two case studies.

3.1 Certification of a fuel injection pump coupling

This first application deals with a fuel injection pump of a 12V Diesel engine used on a ship. After installation of a new type of injection pump, the coupling of the pump happened to repeatedly break and was changed after only a few hours of service. It was then decided to test other types of coupling and to perform torsional vibration measurements, as a torsional resonance was suspected to be the cause of the damage.

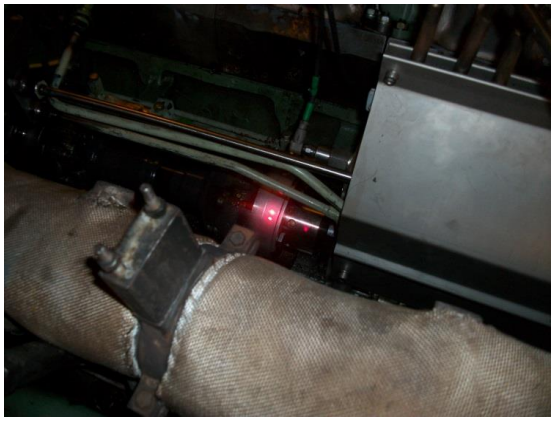


Injection pump installed in the V of the engine

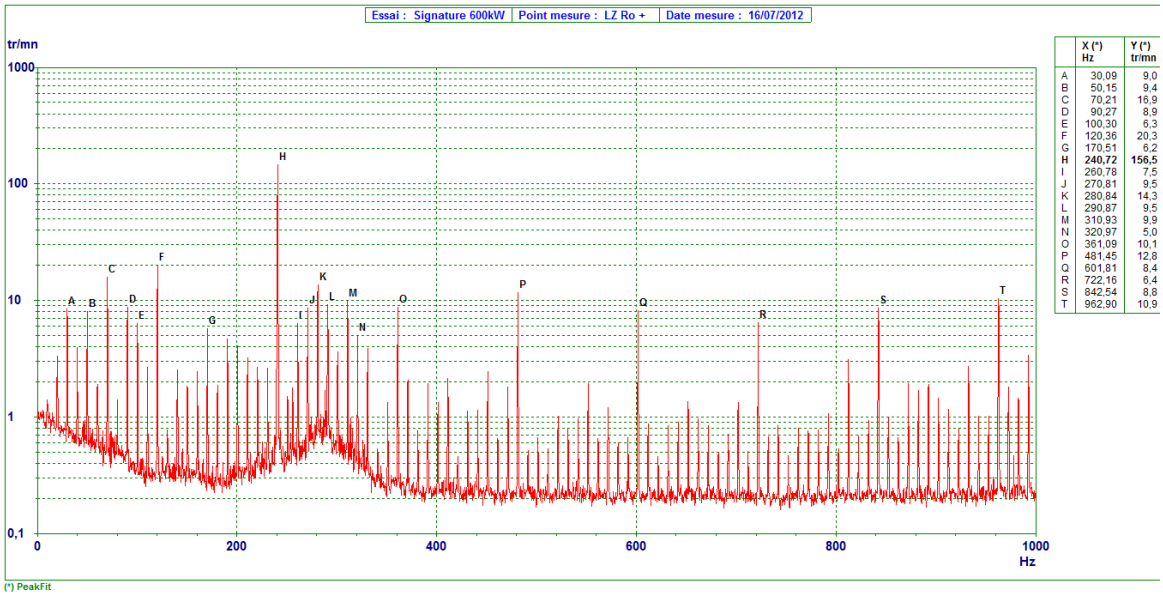
The pump is installed in the ‘V” of the engine and is run by a cardan coupling, which is run itself by the pinion cascade of the distribution. The methodology was as follow:

- Measurement with the initial coupling on the Diesel engine in order to qualify the existing system,
- Measurement with different couplings on a test bench and selection of that showing the best results,
- Measurement with the selected coupling on the ship for validation.

The measurements with the initial coupling were performed with the LTV device aiming the coupling of the injection pump, which is running at 600rpm:



The spectrum of the instantaneous angular speed (IAS) signal is showing an elevation of the background noise around 280Hz, which seems to amplify the spectral component at 2X the injection frequency (*Finj*=120Hz):

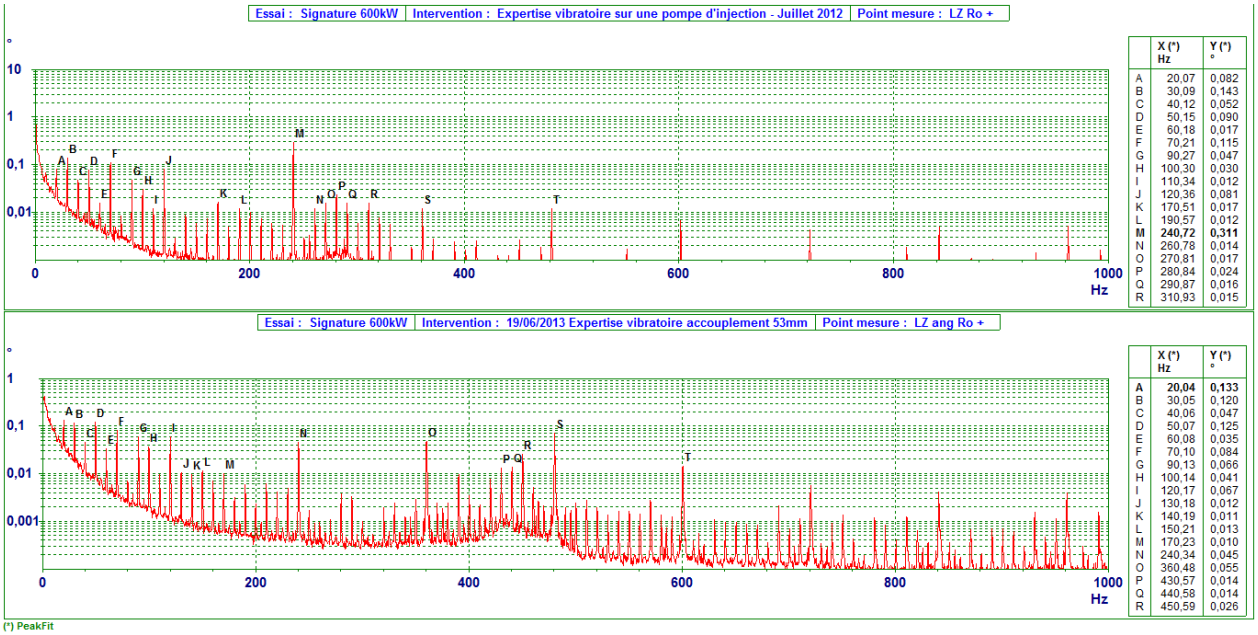


Spectrum of the IAS

The amplitude of the IAS variation at 2X Finj is 156.5rpm Peak-Peak, corresponding to an angular variation of 0.31° 0-P. A limit value was defined empirically for any component of the integrated IAS spectrum, at **0.25° 0-P**.

The injection pump was then tested on a test bench with different types of couplings (1 cardan & 3 disk couplings with different diameters). The best results were obtained with a 53mm diameter disk-coupling.

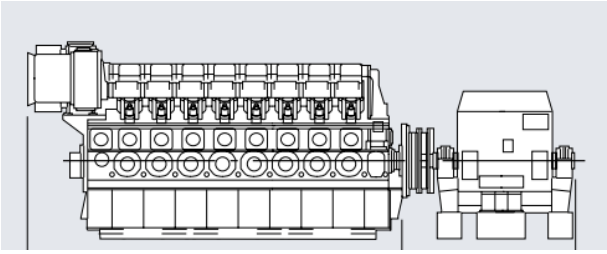
This coupling was then mounted on the engine on board for an endurance test. The amplitudes of the IAS spectral components are significantly lower compared to the original coupling and below the criterion of 0.25° 0-P. They also remain stable after 4 months of service.



Comparison of the angular vibration spectra (in ° 0-P) with the original (top) and new coupling (bottom)

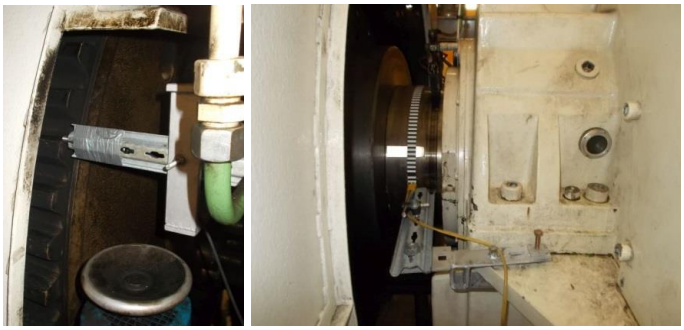
3.2 Torsional analysis of a flexible coupling

This second case study deals with the torsional analysis of a coupling of a high power Diesel group. The equipment is constituted by an 18V Diesel engine driving a generator via a flexible coupling. The shaft line of the group is running at 500rpm.



Diesel engine and generator group

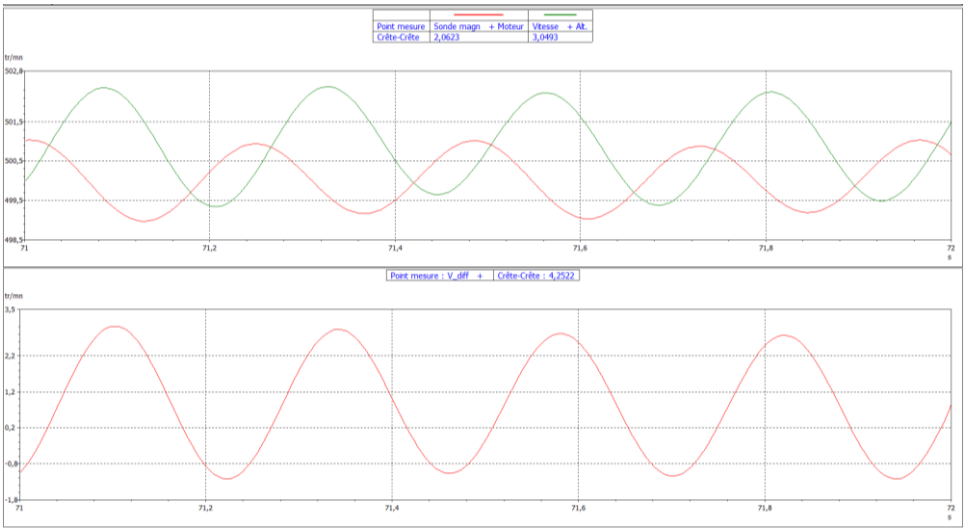
After a sudden and destructive damage of the coupling, it was decided to perform IAS measurement on each side of the coupling in order to analyse its torsional behaviour in service. On the generator side the IAS was measured by using an optical sensor and a strip band with 125 pulses. On the engine side we used a magnetic pickup placed near the toothed wheel of the engine (74 teeth).



Speed measurements on the engine and on the generator

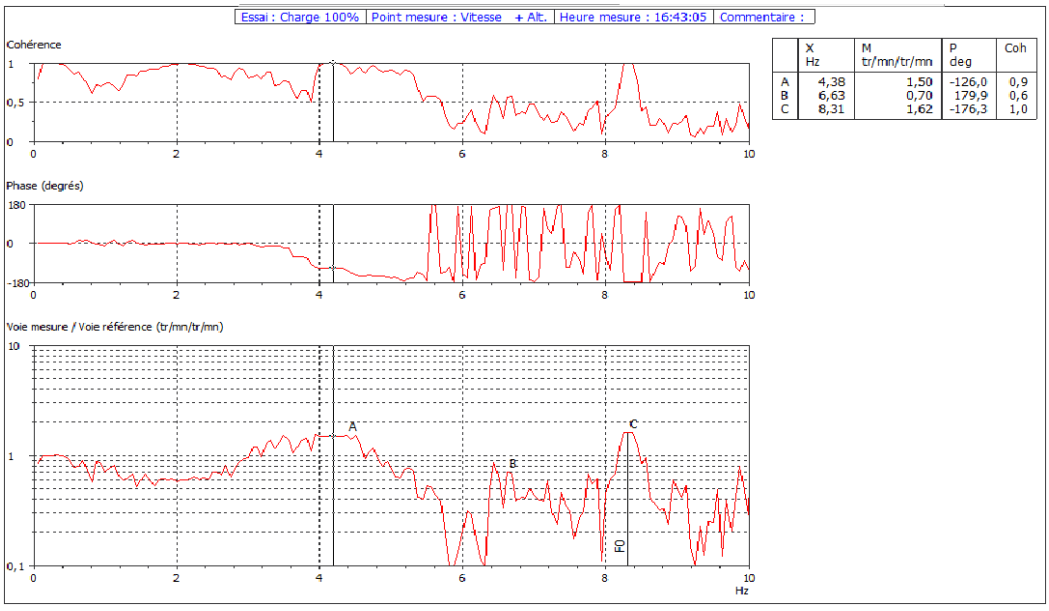
Note that the imperfect junction of the strip band stuck on the generator shaft is inducing periodic spikes at the rotation frequency. These spikes can then be suppressed by a dedicated algorithm. Besides, an advantage of these spikes is to provide a 1X/rev phase reference on the generator shaft.

The analysis of the IAS when the group is running at full load reveals an IAS fluctuation at the combustion cycle frequency of the Diesel engine (i.e. half the rotation frequency) which is present on each side of the coupling. Moreover we observe that this component is phased out and amplified from the motor side to the generator side. The associated speed fluctuation due to this component is about 2rpm PP on the motor side and 3rpm PP on the generator side. This corresponds to a coefficient *C* of speed non-regularity of 0.4% and 0.6% respectively, which seems acceptable. However the fluctuation of the differential speed between the engine & generator is higher (4.25rpm PP), and the corresponding twist of the coupling is about 1° PP.



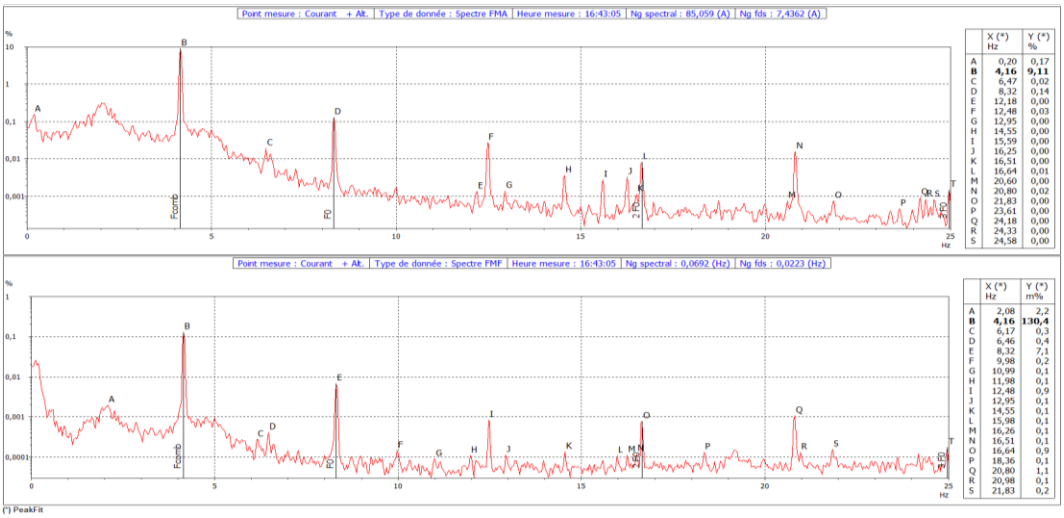
Filtered IAS on the motor & generator sides and differential speed showing a fluctuation at the combustion cycle frequency

The amplification and the phasing of the IAS component at the cycle frequency seem to indicate the proximity of a torsional modal frequency of the shaft line that may amplify the cycle frequency component at 4.17Hz. This seems to be confirmed by the following Frequency Response Function (FRF) calculated between the IAS of the engine and of the generator, which shows a 90° phase lag at a frequency slightly under 4Hz.



FRF calculated between the engine & generator IAS

A measurement of the electrical current was also performed on the generator and shows a strong amplitude modulation at the cycle frequency (9%). Thus the observed IAS fluctuation at the cycle frequency is inducing a torque modulation and so a modulation of the generated electrical power. This modulation is then inducing a torsional fatigue of the elements of the shaft line, especially of the coupling.



Analysis of the generator current amplitude & frequency modulation functions at full load

We also note the presence of an amplification of the background noise around 2Hz on the current amplitude modulation spectrum, which may indicate a torsional modal frequency of the shaft line. This modal frequency is also observed by a transient oscillation on the IAS signals at the moment of the electrical coupling of the generator. Thus it seems to involve the electromagnetic forces of the generator.

We attempted to perform a torsional modelling of the shaft line as in [9], however some parameters still remain unknown: the inertia of the engine and also the torsional stiffness due to the electromagnetic forces of the generator (which seems difficult to estimate from its electrical properties). Further tests are yet to be performed on this group, e.g. testing a flexible coupling with a lower torsional stiffness in order to reduce the incriminated torsional modal frequency close to the combustion cycle frequency.

4 Conclusion

We attempted here to show the interest of different IAS measurement techniques for the torsional analysis on industrial case studies. The use of a specific technique will depend on the application and on the possibilities on site. Advantage of the VRL is to be totally non-intrusive and to avoid machine downtime; however its setting must be performed while the equipment is running.

We also showed the advantage of combining IAS measurement with other measurements, such as the electrical current or the transmitted torque, in order to obtain a full understanding of the torsional behaviour of the equipment.

References

[1] J-L Vasselin, Practical methodologies for on-site measurements of torsional natural frequencies – application to industrial cases, Surveillance 7 international conference (2013).

[2] F Combet, Gear fault diagnosis by motor current analysis – application to industrial cases, Surveillance 7 int. conf. (2013).

[3] L. Renaudin & al, Natural roller bearing fault detection by angular measurement of true instantaneous angular speed, Mechanical Systems and Signal Processing 24 (2010) 1998–2011.

[4] Adeline Bourdon & al, Reconstruction of angular speed variations in the angular domain to diagnose and quantify taper roller bearing outer race fault, Mechanical Systems and Signal Processing 120 (2019) 1–15.

[5] The laser torsional vibrometer: a step forward in rotting machinery diagnostics, N. A. HALLIWELL, Journal of Sound and Vibration (1996) 190(3), 399-418.

[6] B. J. Halkon & S. J. Rothberg, Angular (pitch and yaw) vibration measurements directly from rotors using laser vibrometry, Mech. Systems & Signal Proc. 46 (2014).

[7] G. Nerubenko, Tuned Torsional Vibration Damper, SAE International 2013-01-1897.

[8] H. André & al, Precision of the IAS monitoring system based on the elapsed time method in the spectral domain, Mechanical Systems and Signal Processing 44 (2014) 14–30.

[9] H. S. Han & al, Parametric study to identify the cause of high torsional vibration of the propulsion shaft in the ship, Engineering Failure Analysis 59 (2016) 334–346.

Towards the use of hybrid models for diagnosis and prognosis in turbomachinery health management

P Stephan HEYNS[1], Brian ELLIS[1], David H DIAMOND[1], Ronald G DU TOIT[1], Ronnie SCHEEPERS[2]

[1]Centre for Asset Integrity Management, University of Pretoria, Pretoria, South Africa
stephan.heyns@up.ac.za

[2]Research Testing and Development, Eskom, Rosherville, Johannesburg.

Abstract

Turbomachinery failures are often caused by the dynamic behaviour of rotating blades. The implications of such failures may be serious and practical blade condition monitoring methodologies are therefore becoming important. In this work the idea of hybrid models that rely partly on data driven blade tip timing and partly on physics based finite element models is explored. The work is founded on a Bayesian linear regression based blade tip timing technique which is combined with a stochastic finite element model. This stochastic hybrid approach is applied for the identification and classification of turbomachine blade damage. For the purposes of demonstration, discrete damage is incrementally introduced to a simplified test blade of an experimental rotor setup. The damage identification and classification processes are further used to determine whether a specified damage threshold has been reached, therefore providing sufficient evidence to schedule maintenance.

1. Introduction

Failures of turbomachinery are often caused by the dynamic behaviour of rotating blades. The financial and production implications of such failures may be very significant and appropriate blade condition monitoring methodologies are therefore of critical importance [1,2]. Blade tip timing (BTT) is a non-intrusive measurement technique for online measurement of turbomachine vibration. Essentially it senses when a blade passes a number of proximity probes distributed circumferentially and mounted radially through the turbomachine casing above the row of rotor blades being measured, to determine the blade tip time of arrival. This can be linked to the blade vibration by employing an accurate measure of the once per revolution reference signal. The technique is non-intrusive and online monitoring is possible.

BTT is therefore often regarded as a feasible technique to track the condition of turbomachine blades and prevent unexpected and catastrophic failures. The processing of BTT data to find the associated vibration characteristics is however not trivial. In addition, these vibration characteristics are difficult to validate, therefore resulting in great uncertainty of the reliability of BTT techniques. To deal with the uncertainties of the method, various new concepts have been introduced [2,3,4]. These ideas deal primarily with diagnosis. Techniques for prognosis to assist with maintenance decision making is however becoming more important. Mishra et al. [5] explored a range of techniques of interest to accomplish this through the use of hybrid models that merge physics based and data driven approaches into a unified approach.

In this work the idea of hybrid models is pursued further in the context of turbomachinery blades, by proposing an approach comprising a stochastic finite element model (FEM) based modal analysis and a Bayesian linear regression (BLR) based BTT technique. The use of this stochastic hybrid approach is demonstrated for the identification and classification of turbomachine blade damage. For the purposes of this demonstration, discrete damage is incrementally introduced to a simplified test blade of an experimental rotor setup. The damage identification and classification processes are further used to determine whether a damage threshold has been reached, therefore providing sufficient evidence to schedule maintenance. It is shown that the proposed stochastic hybrid approach may offer benefits for practical implementation. This paper highlights

some aspects from of a comprehensive study by Du Toit, Diamond and Heyns [6] as well as some additional results that further corroborates the approach.

2. Hybrid methodology

Figure 1 outlines the proposed hybrid approach for the identification and classification of damage in a turbomachine blade. The process firstly comprises a data driven blade tip timing analysis to track the relative change in natural frequency of a blade to identify and infer the extent of blade damage at a given time. A blade damage threshold is established through finite element analysis. The BTT analysis is repeated continuously until this threshold is reached. Once this threshold is reached, the operation of the turbomachine should be stopped and relevant maintenance be conducted. Subsequently the blade natural frequency and amplitude determined from the blade tip timing results are clustered using predetermined mean values as initial cluster centres. This clustering then allows the severity of the damage to be classified.

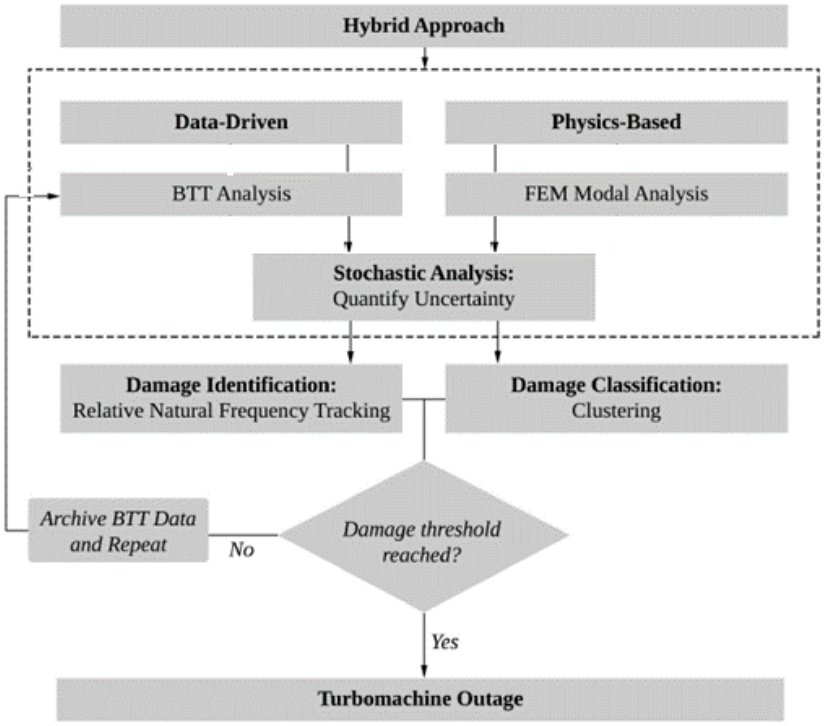


Figure 1: Schematic overview of the proposed hybrid methodology

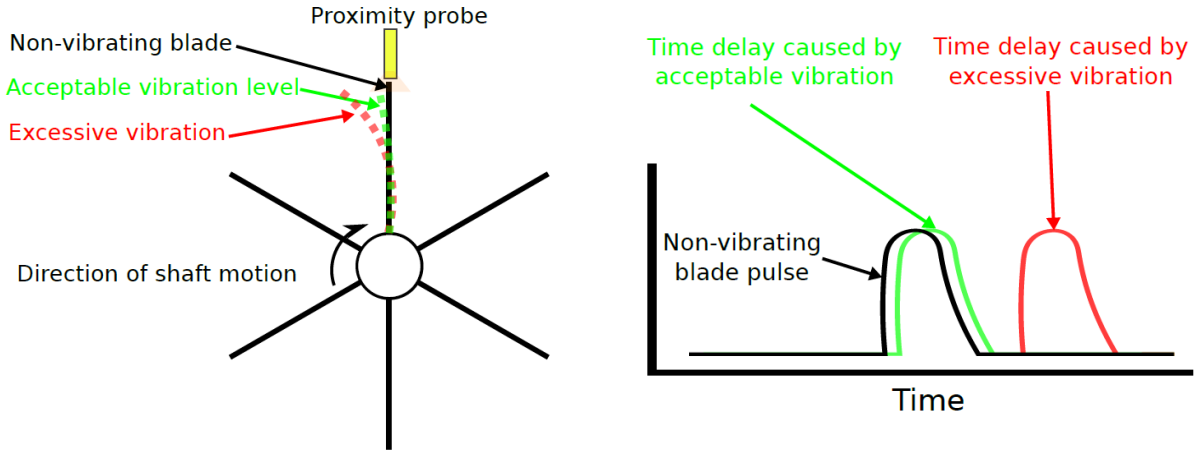
The rationale behind adopting such a hybrid approach is to:

- Use finite element analysis (FEA) to establish a baseline for comparison before BTT tests are performed.
- Project expected blade condition before it becomes possible from the BTT measurements.
- Correct for real blade behaviour and aspects not considered in the FEA.
- Enhance remaining useful life estimation.

The current methodology comprises two analysis types, namely damage identification in which the relative change in natural frequency is tracked to identify and infer the extent of the blade damage, and damage classification in which the natural frequency and blade response amplitudes derived from the BBT results are clustered using predetermined mean values as initial cluster centres to determine the severity of the blade damage. A damage threshold is determined based on FEA.

3. Data-driven investigation

The basic idea behind blade tip timing is to measure the time of arrival (ToA) of blades passing a proximity probe in the casing. This data is then analysed to determine the vibrational state of the blade. The ToA is dependent on the rotational speed of the shaft. However, a vibrating blade will either arrive earlier or later than expected at the proximity probe (see figure 2). Extracting the blade vibration from the ToA measurements can be complex in practice. Many algorithms have been developed for this purpose. Most of these algorithms are based on so-called *indirect* methods in which the maximum amplitude and corresponding frequency at resonance is determined during transient operating conditions and only one or two proximity probes are required, or *direct* methods in which the maximum amplitude at each rotational speed is determined during steady state and at least four probes are normally used.



(a) (b)

Figure 2: Blade tip timing (a) Rotor blade during operation (b) Proximity probe output

In this work we utilise a direct BTT method based on statistical inference [2]. The technique employs Bayesian linear regression (BLR) which offers a number of practical advantages: The tip displacement amplitude and phase with respect to the angular reference signal may be determined at each measured rotational speed. Firstly, this results in a detailed picture of these parameters and their changes over the operating domain. Secondly the processed data considers the whole range of inferred solutions, therefore resulting in a stochastic solution. Lastly, the stochastic nature of the processed data allows one to establish confidence intervals for amplitude and phase and renders the approach noise tolerant.

The BTT technique employed here assumes a single-degree-of-freedom (SDOF) model for the blade vibration. Equation (1) describes the blade tip displacement x at a specific time t, rotor order EO and circular frequency ω:

$$x_i(t) = A_i \cdot \cos(\omega \cdot t_i) + B_i \cdot \sin(\omega \cdot t_i) + C_i \quad \text{where} \quad \omega = EO \cdot \Omega \tag{1}$$

BLR is used to infer the values of the constants A, B and C as probabilistic quantities. The equations are solved for each revolution i at a corresponding shaft speed Ω, with EO also inferred from a probabilistic approach. The parameter set **x** is solved for each revolution and forms part of a multivariate normal distribution with associated mean μ_i and covariance matrix Σ_i as per equation 2.

$$\mathbf{x}_i = \begin{pmatrix} A_i \\ B_i \\ C_i \end{pmatrix} \quad \boldsymbol{\mu}_i = \begin{pmatrix} \mu_{A_i} \\ \mu_{B_i} \\ \mu_{C_i} \end{pmatrix} \quad \boldsymbol{\Sigma}_i = \begin{pmatrix} \Sigma_{AA_i} & \Sigma_{AB_i} & \Sigma_{AC_i} \\ \Sigma_{BA_i} & \Sigma_{BB_i} & \Sigma_{BC_i} \\ \Sigma_{CA_i} & \Sigma_{CB_i} & \Sigma_{CC_i} \end{pmatrix} \tag{2}$$

These random values are substituted in equations (3) to quantify the amplitude and phase:

$$\widehat{A}_i = \sqrt{A_i^2 + B_i^2} \qquad (3)$$
$$\phi_i = \arctan\left(\frac{B_i}{A_i}\right)$$

The maximum amplitude and the associated natural frequency is indicated by equation (4)

$$f_{n_{\widehat{A}}} = f(\vartheta) \quad \text{where} \quad \vartheta = \max_{i \in N} \widehat{A}_i \qquad (4)$$

while the natural frequency can also be derived from the phase angle

$$f_{n_{\hat{\phi}}} = f(\varrho) \quad \text{where} \quad \varrho = \frac{f_u + f_l}{2} \quad \text{subject to} \quad \Delta\hat{\phi} \geqslant \pi \qquad (5)$$

For this work we used an experimental setup (see figures 3 and 4) comprising a rotor assembly, and excitation mechanism, sensors, a data acquisitioning and signal generation system. Labview was used to generate the signals for controlling the motor speed.

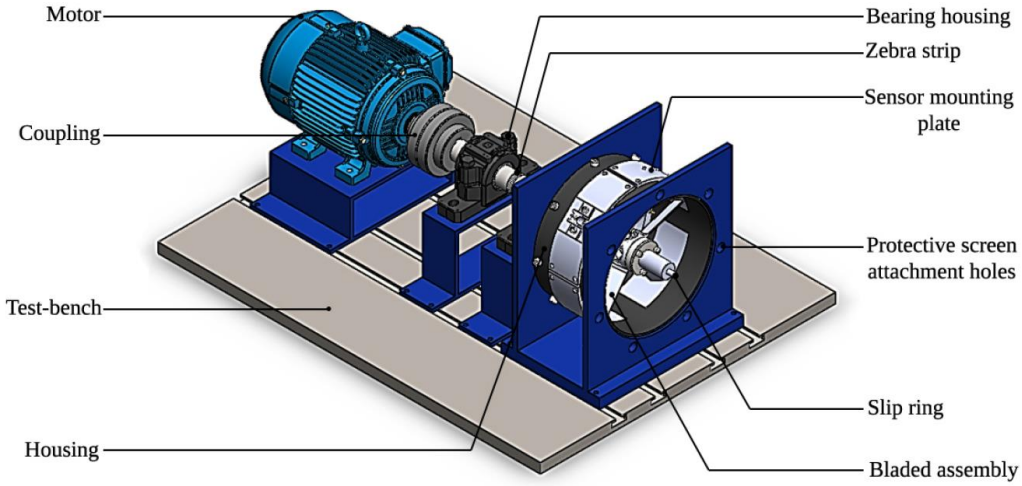


Figure 3: Schematic of the experimental setup

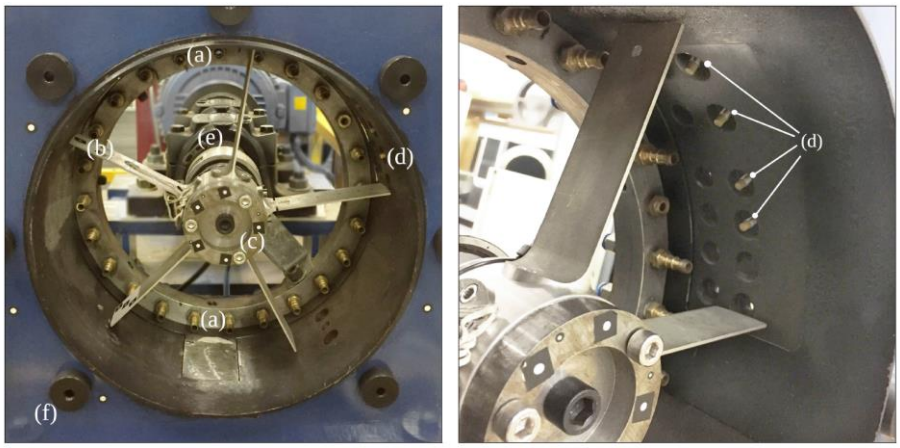


Figure 4: Photos of the rotor and casing

The bladed assembly and casing setup comprises (a) compressed air supply nozzles for blade excitation (b) a bladed assembly with 5 aluminium blades (c) a central hub with a slip ring arrangement for validation

measurements using strain gauges (d) four irregularly spaced eddy current probes and (5) a shaft connected to a motor with a variable speed drive.

In this study damage was artificially introduced to a single blade (blade 2) of width 40 mm in twelve discrete steps ranging from 0 mm to 8.61 mm. Damage was classified in three ranges: non-severe (damage increments 1 to 6, ranging from 0 to 1.81 mm), mid-severity (damage increments 7 to 9, ranging from 3.11 mm to 5.60 mm) and severe (damage increments 10 to 12, ranging from 6,97 mm to 8.61 mm).

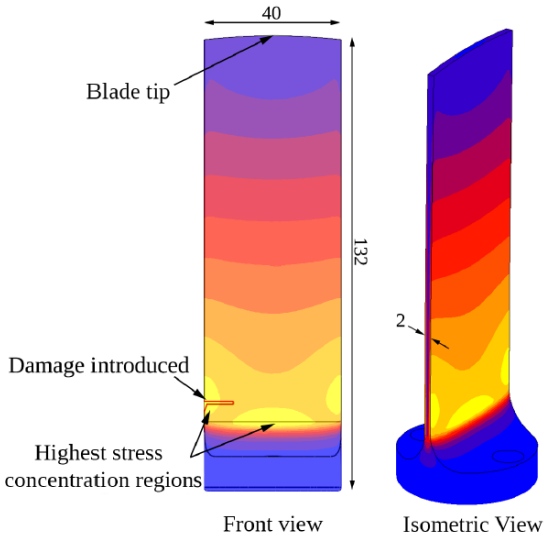


Figure 5: Damage introduced on blade 2, relative to the areas of highest stress concentration for mode 1, as determined from a FEA

The rotor shaft speed was subsequently ramped up from 1195 r/min to 1330 r/min in order to pass through the blade resonant frequencies to allow for the estimation of the amplitude and phase parameters as indicated in equations 4 and 5. BTT tests were repeated six times for each of the 12 damage increments, i.e. a total of 72 tests were conducted. To allow comparison of the extracted blade phase results between the tests, all measurements were synchronised with respect to a specific shaft encoder pulse.

4. Physics based investigation

A 3D FEA was conducted to determine the likely blade resonances at specific operational speeds. Centrifugal loads were accounted for by applying angular velocities to all elements. Thermal loads were accounted for by varying the material properties. The FEA was intended to supplement the BTT results as outlined in sections 5 and 6.

Marc Mentat was used to perform the FEA with the Lanczos algorithm to compute the natural frequencies and amplitudes for the simplified blade. The damage was again introduced in 12 discrete stages corresponding to the damage in the experimental test blade. Slight variations in angle (approximately 10°) and size (length, width and height – about 12%) were introduced for each discrete stage to account for uncertainties in true crack and crack location measurements as well as material properties and centrifugal loads. 89885 ten-noded tetrahedral finite elements were used for the basic geometry. Samples of these parameters were randomly selected from a uniform distribution within the angle and dimension ranges. Aluminium was selected as material with uncertainty again being introduced by varying the material properties within 12% of the reference values.
Variation in these parameters resulted in variation in the natural frequencies for each test within a particular series of tests. This allowed a mean with its associated confidence intervals around the mean, to be determined for each damage condition.

5. Damage identification

The first part of the proposed hybrid procedure considers the tracking of the relative change in the blade natural frequency (see equation 6) to identify and infer the degree of blade damage. Tracking the relative natural frequency change aims to give a more general indication of the blade condition, due to differences in the responses determined from the BTT results and the FEA results, which might not necessarily capture all the important physics associated with the rotating blade response.

The procedure therefore relies on tracking the relative changes of the derived BTT natural frequency results to infer the extent of the blade damage. The FEA allows one to estimate the expected blade conditions (relative change in natural frequency) corresponding to a particular damage size that is considered critical. This process is illustrated in figure 6.

The change in natural frequency from a reference state Δf_{ni} (undamaged in this case) is first quantified. f_{n0} is the original natural frequency and f_{ni} is the current natural frequency.

$$\Delta f_{n_i} = \frac{f_{n_0} - f_{n_i}}{f_{n_0}} \times 100 \tag{6}$$

The aim is to determine if the blade damage threshold has been reached. The challenge however is of course that the actual crack size of the blades can in practice not be determined while the online BTT measurements are made, and that the extent of blade damage must therefore be inferred from the processed BTT results.

Figure 6 shows two iterations of the proposed identification procedure. The figure shows the results from tracking the BTT relative natural frequency based on amplitude and phase based results.

The process follows the following steps:

- Define a blade damage level that would justify physical inspection. This is calculated from the FEM based on the principles of fracture mechanics or fatigue analysis and leads to the identification of a Level 1 damage threshold as seen ion figure 6.
- The FEM modal analysis is performed stochastically at this discrete crack size in order to quantify uncertainty. The mean μ_{FEM1} and the standard deviation σ_{FEM1} corresponding to this damage increment is subsequently recorded.
- A new variable δ_{damage} is then defined as in equation 7. This parameter represents the difference between the relative change in the natural frequency from the BTT results (for a particular batch of tests) and the relative change in natural frequency of the FEM results at a particular discrete damage size, represented by Level 1:

$$\delta_{damage} = \Delta f_{n_{BTT}} - \Delta f_{n_{FEM}}(K) \tag{7}$$

K corresponds to the predetermined crack size in the FEM modal analysis. δ_{damage} is a normal probability distribution with both Δf_{nBTT} and $\Delta f_{nFEM}(K)$ having associated normal distributions.

- The calculation of the mean and the variance of δ_{damage} requires that a number of repetitive BTT tests must be performed and that the mean μ_{BTT} and the standard deviation σ_{BTT} be determined.
- The probability $P(\delta_{damage} \leq 0)$ is determined for a batch of BTT tests and the chosen Δf_{nFEM} (Level 1 in this case). This probability is found from the cumulative distribution function (CDF) of δ_{damage} with the associated mean and variance as shown below:

$$\delta_{damage} \sim \mathcal{N}(\mu_{BTT} - \mu_{FEM}, \sigma_{BTT}^2 + \sigma_{FEM}^2) \tag{8}$$

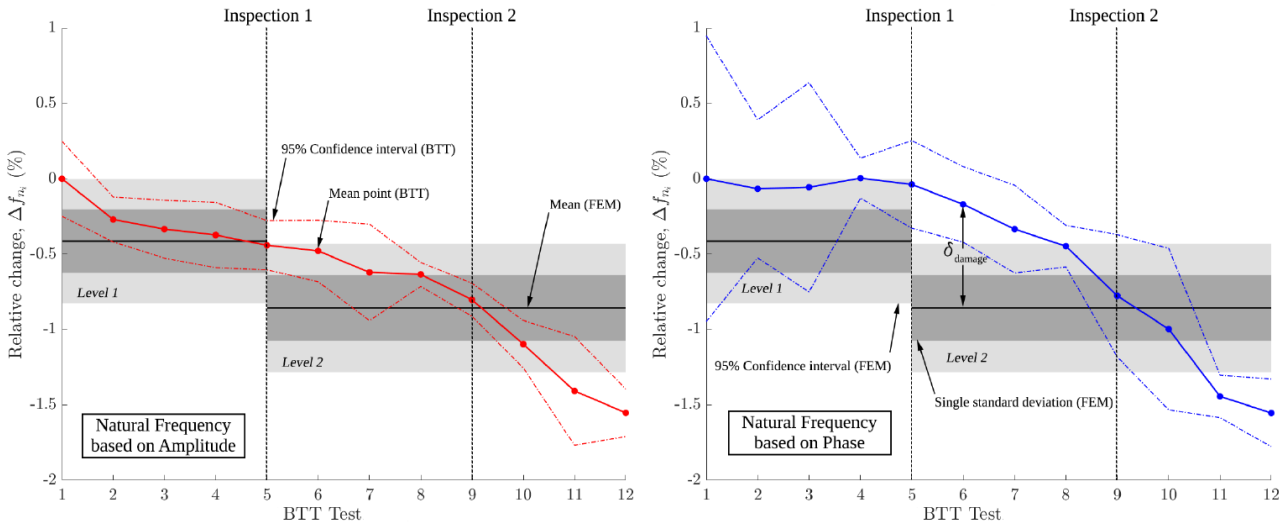


Figure 6: Damage identification at ambient temperature of 22°C

$P(\delta_{damage} \leq 0)$ is the probability that the relative change in natural frequency derived from the BTT measurements Δf_{nBTT} equals or exceeds the permissible relative change in natural frequency derived from the FEA, $\Delta f_{nFEM}(K)$. The damage threshold X_{dt} is based on a selected probability $P(\delta_{damage} \leq 0) > X_{dt}$

The user must decide what an acceptable probability would be to justify maintenance. Repetitive BTT tests are conducted until this probability value (damage threshold) is reached. It is important to note that the damage threshold X_{dt} may be selected to be conservative.

The above steps are repeated after every inspection or maintenance until a blade needs to be replaced. After each inspection a new damage level, based on the FEM results may be prescribed to determine a new blade damage threshold.

Figure 6 demonstrates two iterations of the damage identification process. In this diagram Level 1 corresponds to a relative discrete crack size of 3.95% and Level 2 corresponds to a relative discrete crack size of 9.675%. Both Levels 1 and 2 are arbitrarily selected here for demonstration purposes, but would in practice be based on physical evidence.

A very comprehensive set of experiments were also done at elevated temperatures of 41° and 98°C, using two 2 kW commercial heaters to heat the air stream through the machine. As a further example of the application of the procedure typical results for 98°C are also shown in figure 7. This figure confirms the general behaviour observed at ambient temperature (figure 6) and at 41°C (not shown here).

6. Damage classification

The damage identification process presented in section 5 relies on repetitive measurements to be available. The damage classification process described here however aims to allow the use of a single set of BTT measurements to determine the blade condition. A high level of confidence must therefore be attached to this measurement. To accomplish this a damage classification procedure that is based on the clustering of BTT natural frequency and amplitude values is investigated. Here we use the well-known K-means clustering technique, which assigns the observations with the nearest means to a certain cluster of data points. Physically this means that the vibrational characteristics of the blade from a specific BTT measurement, are assigned to an existing cluster of vibrational characteristics with the nearest mean. The aim of this cluster is to classify the

severity of the blade damage according to which group the vibrational characteristics are assigned to. This means that the damage severity may be established from these clusters.

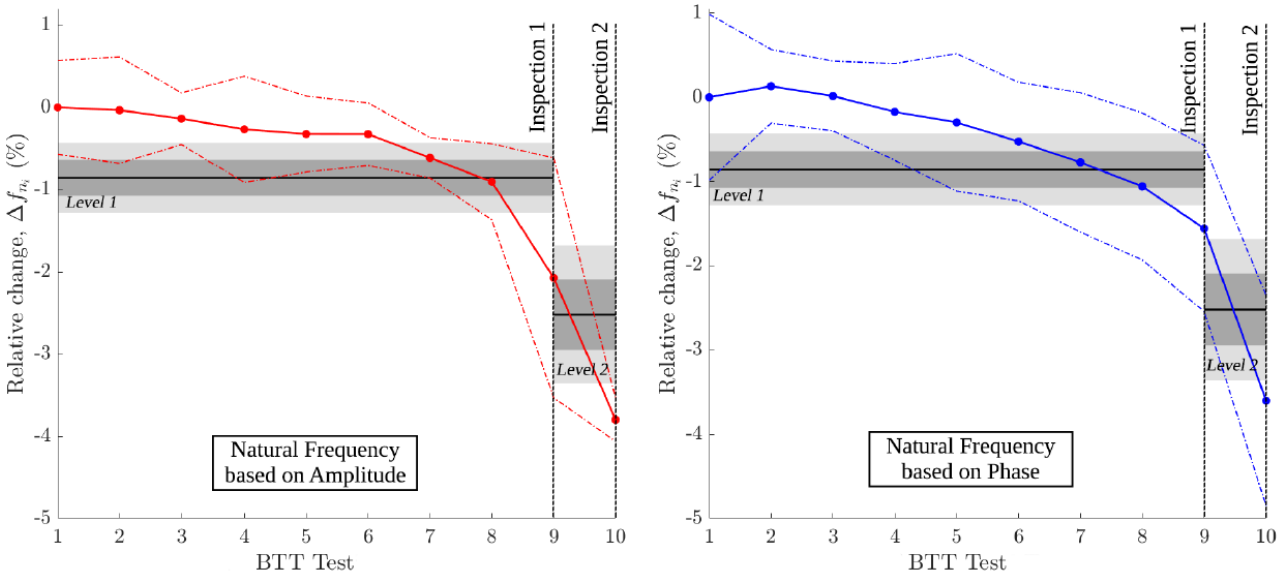


Figure 7: Damage identification at 98°C

The process follows the following steps:

- The amplitude and associated natural frequency of a specific blade are extracted from the BTT measurements.
- We propose clustering of the natural frequency and amplitude results using predetermined mean values. The predetermined values correspond to the natural frequency results from the FEA for mode 1. The blade amplitudes are however not computed from the FEA since this would require a computational fluid dynamic analysis. This is however a complex process and infeasible in practice, and is avoided here by simply assigning zero amplitudes to the initial clusters.
- In this work the data is partitioned in three clusters, namely non-severe damage, mid-severity damage and severe damage (see section 3). These partitions are not enforced on the BTT data before the K-means clustering commences. However after clustering is completed the individual natural frequency points and associated amplitudes are used to determine the accuracy of the final classification.
- The initial cluster centres correspond to zero amplitude values and the mean natural frequency values of the FEA results partitioned using the scheme mentioned above. This results in three amplitude and natural frequency combinations to be used as starting points for the centroids. These initial cluster centres enable individual BTT measurements to be classified to the closest partition.
- The BTT amplitude and natural frequency values are both normalised over the range 0 to 1 for use in the K-means clustering. The initial cluster centroids based on the FEA are also scaled using this scheme.
- The point-to-cluster-centroid distances are computed for all of the individual BTT points. These points are considered with no indication of which damage increment the points belong to. The overall averages of the points are calculated and the new centroid locations are allocated. As a result the BTT points are classified into the associated group of most likely Range of Damage (RoD) that the amplitude and natural frequency would represent. In figure 8 RoD I, RoD II and RoD III represent the new averaged clusters for the undamaged, middle damaged and largest damage increments respectively. The incorrectly classified points are also shown in figure 8.

Figure 8 shows the results of the K-means clustering implementation as part of the damage identification process. RoD I, RoD II and RoD III represent the new averaged clusters for the non-severe, mid-severity and severe damage classes respectively. Incorrectly classified points are indicated by a Roman numeral above the marker.

The overall classification accuracy is 78%. If only the RoD III case is considered, the accuracy would be 94%. However in one case an actual RoD III was incorrectly classified as a RoD I point. This is obviously dangerous and points to further research being required.

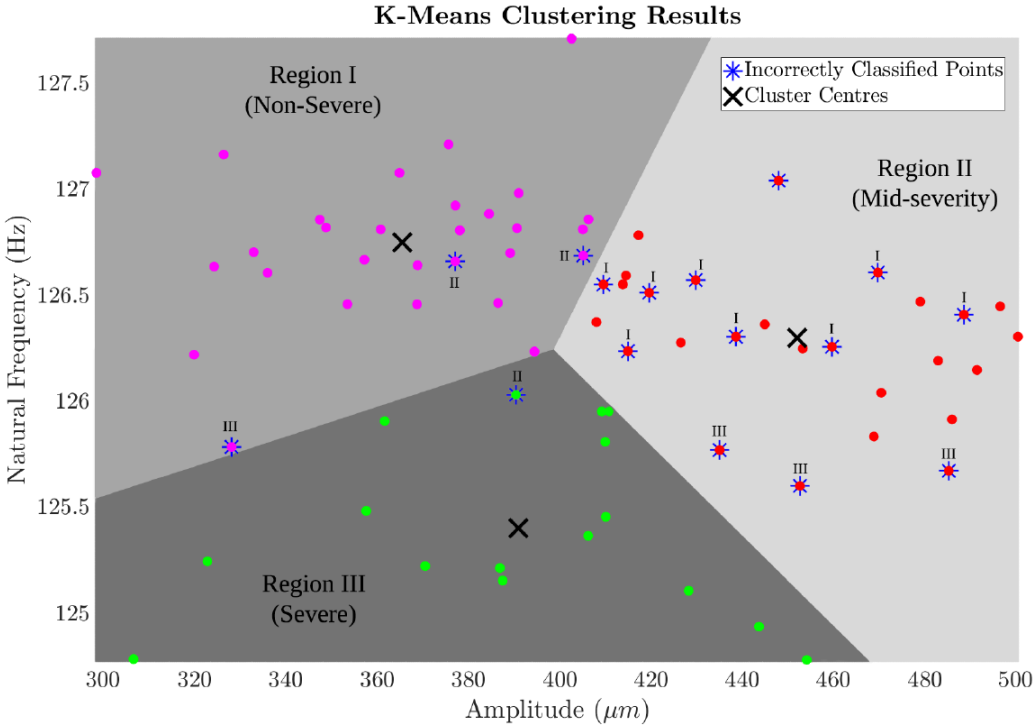


Figure 8: K-means clustering results

7. Conclusion

This work represents a first attempt to develop a hybrid methodology encompassing data-driven BTT and physics based finite element analysis for turbomachinery blade diagnostics and prognostics. This hybrid method uses the outputs of a BTT technique that is based on Bayesian linear regression and stochastic finite element analysis.

An experimental study was conducted on a simplified test rotor, with discretely introduced damage on a test blade.

The work led to a damage identification procedure based on the probability that the relative change in natural frequency of the BTT results is as large as that determined by the finite element modal analysis (at a chosen discrete damage size) projects it to be. This probabilistic damage identification process demonstrates the ability to infer the extent of blade damage.

A damage classification process is introduced to determine the blade condition using a single set of BTT measurements. To simplify the process, K-means clustering is used to classify the derived BTT amplitude and natural frequency values. The predetermined finite element analysis natural frequency results are used to

initiate clusters and cluster centroids. The clustering of the derived BTT vibrational characteristics to the nearest cluster centroid enables the severity of the blade damage to be classified. While 78% classification is achieved, it is however shown that the possibility exists of classifying a severely damaged blade as non-severely damaged. This is of course potentially dangerous in practice. Future research into the performance of alternative classification procedures is therefore required.

We do however believe that a useful step has been taken towards the use of hybrid models for diagnosis and prognosis in turbomachinery health management.

Acknowledgments

The authors gratefully acknowledge the Eskom Power Plant Engineering Institute for their support of the development of the test facilities that were used for this investigation.

References

[1] Gubran, A.A. & Sinha, J.K. Shaft instantaneous angular speed for blade vibration in rotating machine. Mechanical Systems and Signal Processing, 44(2014), pp.47-59.

[2] Diamond, D.H., Heyns, P.S. and Oberholster, A.J. A comparison between three blade tip timing algorithms for estimating synchronous blade vibration. 9th World Congress on Engineering Asset Management, WCEAM 2014, 28-31 October 2014.

[3] Diamond, D.H., Heyns, P.S. and Oberholster, A.J. Improved blade tip timing measurements during transient conditions in a state space model, Mechanical Systems and Signal Processing, 122(2019), pp.555-579.

[4] Diamond, D.H. and Heyns, P.S. A novel method for the design of proximity sensor configuration for rotor blade tip timing. Journal of Vibration and Acoustics, 140(2018), pp. 061003-1-8.

[5] Mishra, M., Saarti, J. Galar, D. and Leturiondo, U. Hybrid models for rotating machinery diagnosis and prognosis. Estimation of remaining useful life. Luleå University of Technology, 2014.

[6] Du Toit, R.G., Diamond, D.H. and Heyns, P.S. A stochastic hybrid blade tip timing approach for the identification and classification of turbomachine blade damage, Mechanical Systems and Signal Processing, 121(2019), pp.389-411.

Condition Monitoring

CM Base, a universal gateway to condition monitoring datasets

Cécile Capdessus[1], Frédéric Bonnardot[2]

[1]PRISME, Univ. Orléans, 21 Rue Loigny-la-Bataille 28000 Chartres, France
Cecile.capdessus@univ-orleans.fr
[2]Univ Lyon, UJM St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France
frederic.bonnardot@univ-st-etienne.fr

Abstract

Condition monitoring is a lively research domain, with hundreds of researchers sharing new techniques coping with complex problems, such as variability of the operating conditions, complex machinery monitoring, diagnostics, prognosis... Considering the many methods that have been proposed these last decades, it is striking how few of them have reached the industrial application. One of the reasons is that the intensive validation of these methods over real-life data required to ensure their reliability is often difficult to achieve. The results obtained over one dataset may not be easily reproduced over another one. Furthermore, can the comparison between two techniques be objective, when applied to different datasets?

Some laboratories and companies generously share the datasets recorded on their test benches or industrial devices. Unhappily, it is often difficult for them to know for certain who has worked on their data, which is both frustrating and a real problem when they report to their authority about the diffusion of the data, in order to get the subsidies required for the test bench exploitation. From the user's point of view, there can also be difficulties in finding the right datasets for a specific study among the jungle of all that is proposed on the web. How to find the appropriate dataset for a prognosis study, or a fatigue study, or the study of any specific kind of damage? Another problem can be met by those who wish to share their datasets with the scientific community but do not have the technical skills or staff to do so.

CM Base is a web portal that aims at facilitating the sharing of the data, by offering many a functionality, such as a list of all the existing datasets and test benches with all related papers, searching facility allowing to extract from the base all datasets related to a specific problem or papers related to a specific dataset.

1 Introduction

These last decades have seen a wide expansion of on line condition monitoring, due to various factors. First, the evolution of sensors, computing and communication technologies allow setting up condition monitoring systems more easily and at lower cost. Second, the need of such systems has increased. Indeed, in the present competitive international context, production companies cannot bear the consequences of unplanned production lines shutdown. Furthermore in many domains, products are lent rather than bought by the clients, or bought with some service package, including monitoring. The development of machine learning and Artificial Intelligence (AI) has also opened new fields for automated surveillance.

In this context, the need for reliable monitoring techniques has led industry as well as academic researchers to explore many possibilities, often based on the use of a set of sensors and the elaboration of relevant health indicators through the application of signal processing techniques. The elaboration and the testing of these methods require the production of experimental data.

Many laboratories have set test benches and recorded measure signals datasets and some of them chose to share these datasets with the scientific community through a web site. But it reveals difficult for them to get information about the diffusion of their dataset: for instance who it has been used by, the conferences and papers the obtained results have been published in. For their part, the users have some difficulties in finding the right dataset to test their technique or even to know about the available ones. Furthermore, some laboratories wish to share their datasets but lack computer knowledge or facility to do so.

CM Base is a database meant both to gather all the information about existing on line datasets and to help laboratories putting online their datasets. It can be accessed through a web site that allows inserting a new

dataset, inserting a paper related to an existing dataset, and browsing the basis in order to find some datasets relevant to a specific problematic with all related papers. In order for the basis to be as widely used as possible, the interface has been designed to simplify all procedures, which makes CM Base a very easy to use application.

In section 2, the need is analysed and a solution is designed, in section 3 the chosen technical solution is presented, in section 4 a user guide for the use of CM Base is given, and in section 5 a conclusion and some perspective are proposed.

2 Analysis of need

2.1 Reliability and reproducibility

Reliablility and reproducibility of the techniques have become one of the main topics in all scientific domains [1], peculiarly in all that is related to data processing and statistical analysis. Indeed, for a specific technique to be validated, it is not enough that it has been tested on one set of data. The quality of the obtained results can depend of some parameters related to the specific experimental setup, the location of the sensors, all things that are different from a test bench to another. Furthermore, real-life signals are yet another challenge. Techniques that have been tested on measure signals collected from a lab test bench can lack robustness when it comes to measures collected from an industrial device [2, 3]. It is thus of utmost importance that researchers might access to a wide set of different measures, in order to test the robustness of their proposed techniques and to specify in which conditions these can be applied. Researchers should thus be given access to a wide range of measure signals, collected on both lab test benches and real-life devices.

2.2 Comparison of techniques and measure signals

No indicator can be considered perfect for the monitoring of a system. Depending of its stage, the presence of some damage may produce on the measure signal alterations of a different nature. For instance, a localised defect on a roller bearing can be detected with techniques well adapted to transient signals, whereas once the damage has spread, such techniques will fail to put it at light. Furthermore, different kind of measure signals can be used to monitor a system: current, vibrations, temperature… For these reasons, most condition monitoring AI systems use both different kinds of measure signals and different kind of indicators. It is thus interesting to compare either the results obtained through one specific method on different datasets, or through different methods applied to a specific dataset. This can help defining how to combine these techniques and merge different indicators in order to ensure a reliable monitoring system. It would thus be interesting for the condition monitoring community to have access not only to different kind of measure signals, but also to the related bibliography.

2.3 Choice of an appropriate dataset

When addressing some specific topic that may be bearings, gearboxes, diagnosis, prognosis, wear, variable load, variable speed or any other relevant topic for condition monitoring, accessing to a measure dataset well adapted to this topic is not an easy task. All labs are not equipped so as to set up an experiment and while many datasets are now available on the web, it takes time to check whether they have the right profile. It would be convenient that all these datasets can be accessed through a browsing system that would allow getting straight access to the well-fitted ones. And then, for each dataset, it would be interesting to have also access to the related state of the art. On the top of it, it would ensure to the researchers who have already published on this dataset that any newcomer is informed of their work.

2.4 Storage and feedback

Many labs and even companies have chosen to share their datasets through their lab webpages or at the occasion of a conference contest [4]. In order to justify the maintaining of their test benches, they must prove to their organisation that sharing the data with the scientific community is profitable. This profit can be expressed in terms of international influence or appropriation of new techniques. The problem is that they often get no more than the number of times their data have been downloaded, or at best who asked for them. They rarely get information about what has really been done on their data. They sometimes discover by chance that their team is thanked in a paper for having shared the data. It would be interesting for them to get a list of all papers that have been published about techniques that were applied to their dataset. In other words, it would be useful to ty to each dataset the related bibliography. Furthermore, some researchers would like to share their data with the community but have no technical means to do so. It would be interesting to provide the community with a storage facility to share their dataset.

2.5 Specifications

The application should allow:

- Sharing a dataset ;
- Sharing the reference of a publication related to a dataset ;
- Browsing the base in order to get a list of all datasets relevant to:
    - A specific topic (diagnosis, prognosis…);
    - A specific device (gearbox, bearing, engine…);
    - A specific context (lab test bench, industrial recordings);
    - Specific operating conditions (constant or variable load or speed for instance);
- Getting the bibliography related to any dataset.

The access should be authorized only to registered visitors, so that the application must include some access management part. The interface must be easy to use, so that contributors do not lose time whenever they add a dataset or a paper, or browse the base.

3 Technical solution

3.1 Strategy of development

CM Base is accessed through a web site. This web site can be used in two ways: as a data navigator to search data by keywords and obtain associated bibliography, or as an editor to declare existing data and databases, associated articles and documents.

The catalogue of data, database, associated links, keywords and bibliography are stored in a relational database. The associated documents and pictures are directly stored on the server. CM Base is not a data storage server (like dropbox) that manages hundreds of MegaBytes of data, but a gate allowing access to the data. Only links to data and small document are stored into the database. The user can use a file-sharing solution such as Mendeley [5] to store his measure datasets, and then declare these data in CM Base to give both acces to the data and information about it: keywords, little document, link to files, …

The website was developed by using standard LAMP tools: Apache, MySql and Php and a Linux Debian web server hosted by Université Jean Monnet of St Etienne on a Virtual Machine.

In order to provide a long-term solution not sensitive to update we avoided using a web Content Management System (CMS) like Wordpress, Drupal, Joomla, … The solution was directly coded in php [6]. In the same way, instead of using css toolkits like bootstrap that require extra files and regular update, a more classical css formatting was used , based on flexbox.

A Model-View-Controller architecture [7,8] was used to organise the project. The database was designed through a Merise analysis [9] and is managed by sql requests in Php.

3.2 Database structure

The organisation of the database is illustrated by figure 1. Each box corresponds to a table, which is equivalent to a sheet from a spreadsheet. The name at the top is the name of the table/entity. The following elements are the fields, which are the equivalent to the column names in a spreadsheet. The interest of relational database is that it allows establishing some links between the tables. These links are symbolised by rounded boxes. For example, all fields related to a member, i.e. his member_ID, last name, e-mail, and so on, are stored in a Members Table/Entity. A link is established with the Papers Table in order to associate to each member the publications he introduced into CM Base. This link is created by adding a (table_ID, member_ID) pair to the IntroducedBy Table. More technical details about this representation could be found by searching literature on Merise method for Database Analysis.
The main tables are:

- The Members table, to store a list of people registered in CM Base,
- The Papers table that contains a list of papers stored in the database,
- The Datasets table, to store information about each test bench or data collected in the database,
- The Ressources table, in which are collected a list of files, links, pictures.

The links between tables enable us to find articles associated with datasets (About relation) or to make a link between type of Sensors and Datasets. The use of a database facilitates data management since all creation, updates of data are made by requests to MySql server, so that all the storage of the information is managed by Mysql.

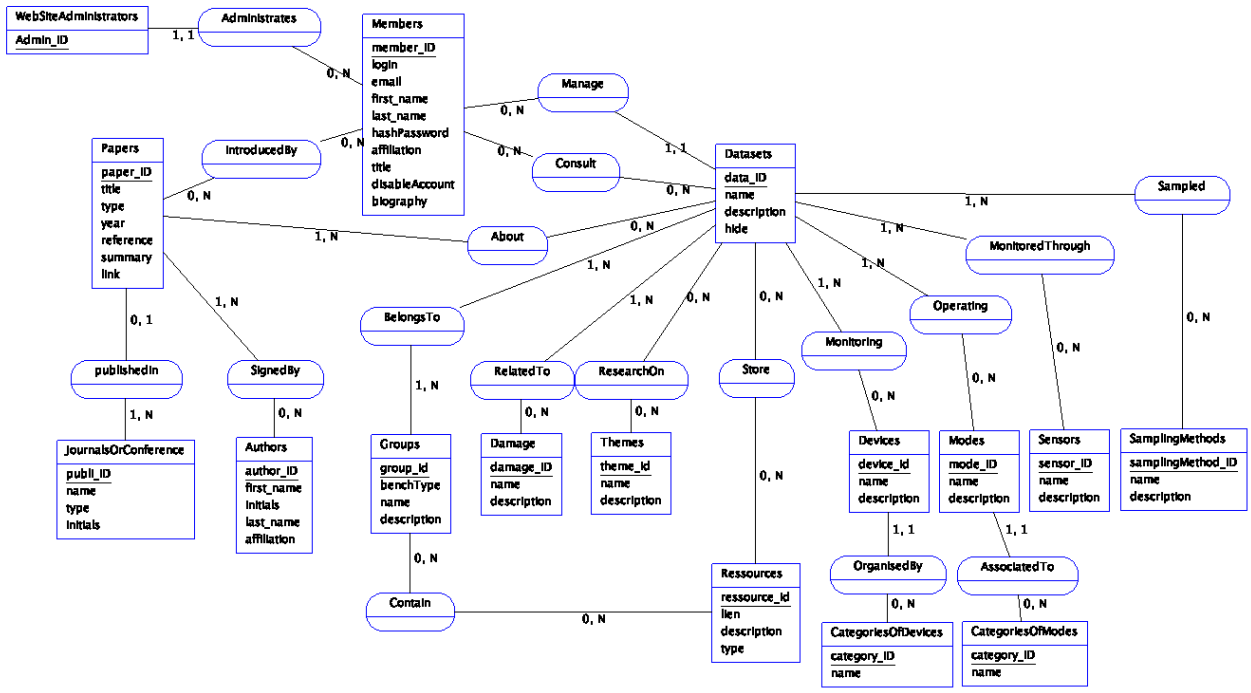


Figure 1: Database Conceptual Model

3.3 Organisation of the Web site

Access to the data is provided through a web site. A menu enables accessing to the following pages:

- “Search the base” to browse inside the collection of data,
- “Add my data” that enables the user to insert its own data into the database or made some links with other databases,
- “Add my papers” to declare papers linked with a dataset,
- “My profile” to change or give personal details,
- “My contributions” to list papers and data inserted by the logged user.

The next section will present some of the web pages as an example.

4 User guide

In this user guide are presented some of the functionality of CM Base. A more exhaustive user guide is available on CM Base web site.

4.1 First connection

Figure 2 shows the home of the database. An account must be created for the first connection by using request account link in the menu. The request account page asks some personal details, an e-mail, a biography and a password.
After connection, more options are available on the menu as shown in figure 3.

4.2 Search the base

The “search the base” web page lists the content of the dataset and enables to make search by using tags associated with the dataset (sensors, device, type of sampling, …). These tags are the same as those used for the dataset insertion (see “Add my data” section).

Figure 2: Home page

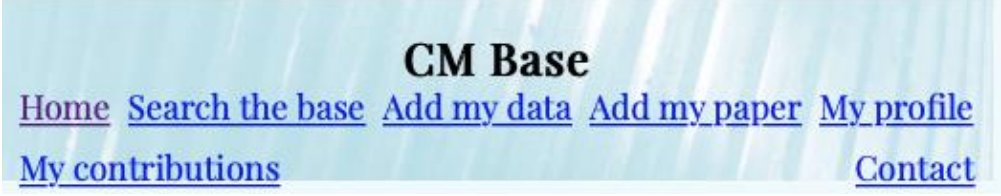
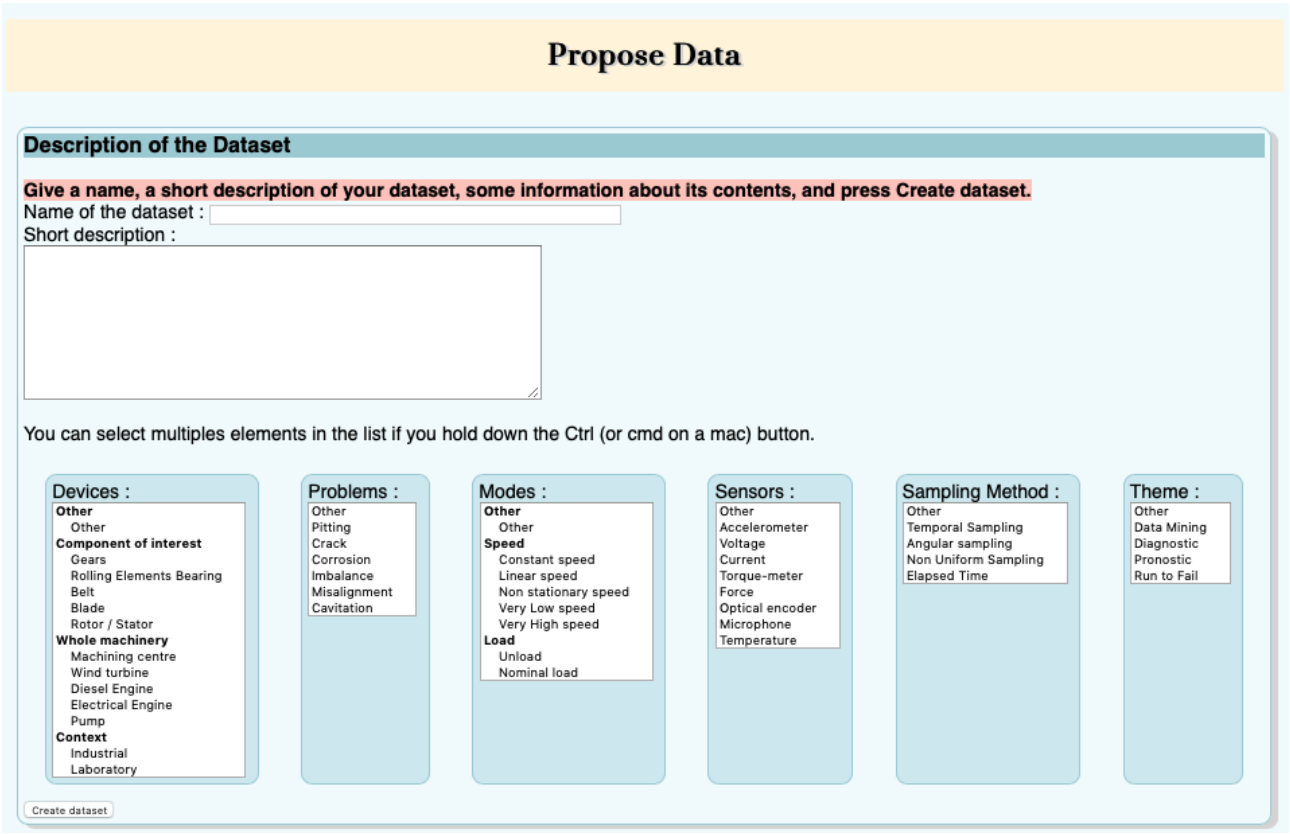


Figure 3: CM Base menu

4.3 Add my data

The menu item “add my data” allows inserting your data into CM Base and sharing them. The associated page is shown in Figure 4. At first, only the “Description of the dataset” part is displayed. Once this part is completed and validated by clicking on the “Create Dataset” button, the next elements appear.
The description of the dataset and main information sections display the same elements as those displayed in a dataset list or dataset search answer. Each element can be modified after submission if necessary.
The user should provide a dataset name, a short description (short enough to be displayed in a table of answer). A longer description can be provided as a pdf file in the main information file.
The six thematic lists are very important and should be filled with care, since the selected items will be used to perform sharper browsing through the base. One or more elements can be selected.
The “Devices” theme enables identifying the component of interest (gear, rolling elements bearing, …) within the dataset and the associated machinery. It is also necessary to give information about the damage, the sensors and sampling methods. The “Mode” section allows giving extra information on running conditions: speed, load, … Three different acquisition modes are available: diagnostic when it is necessary to find the type of defect, prognostic when it is necessary to estimate a lifetime duration, and, data mining when a large set of data with different states (health, damage, …) are provided.
The “Access to data” section gives three alternatives to provide data (it is possible to combine these options): first, the data can be obtained through an e-mail request to the associated contact, second, one or more links to a file on a server can be provided , and third, if a web page associated with the data already exists, it is possible to provide some web links.
Additional files like calibration information, document of the test bench, scripts to read the data can also be provided.



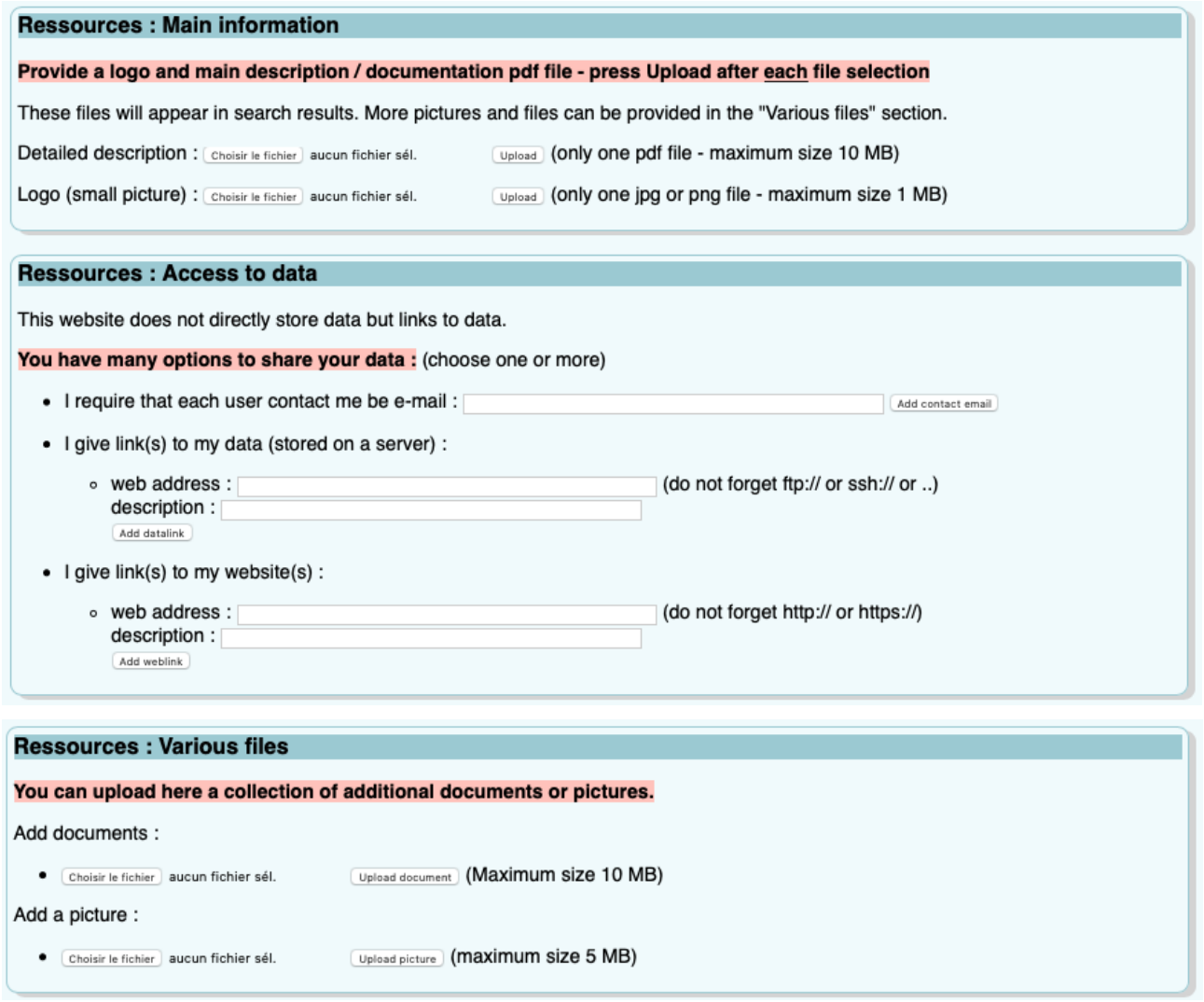


Figure 4: Add my data

4.4 Add my paper

Every user can declare paper associated to a dataset by using the “Add my paper” page. To do so, simply paste the citation of your paper in a format similar to bibliography style and click next. CM Base will analyse your entry and try to identify the author, the title, the journal and so on. The result of this analysis will be presented in a web page and you will be proposed to correct the values and to link your paper to a dataset.

4.5 My Profile

My profile web page allows entering or modifying your personal details, biography, password or delete your account.

4.6 My Contributions

This web page shows a list of submitted datasets as well as associated bibliography. It also gives the possibility to edit or delete the dataset. Since this database is moderated, the dataset should be accepted by the webmaster before being viewable on the web site.

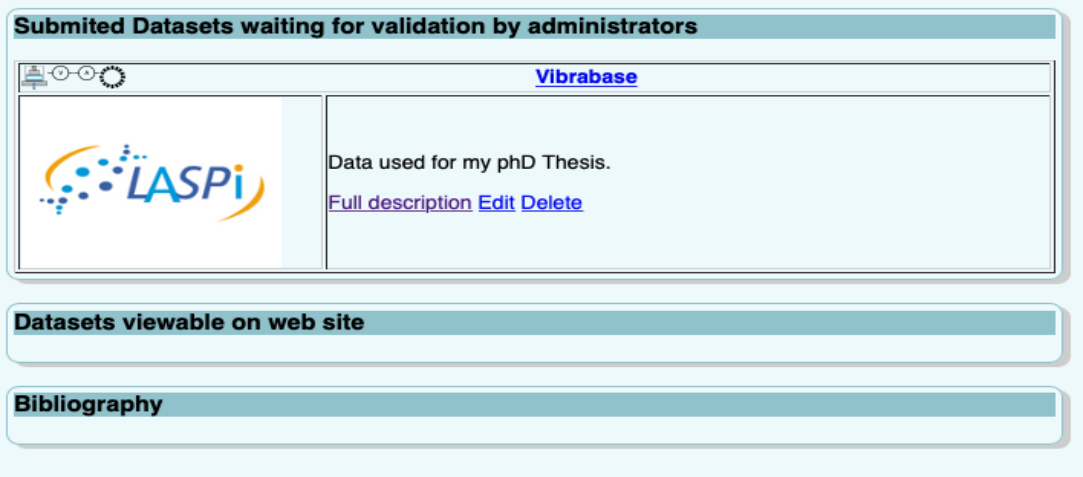


Figure 5: Add my data

5 Conclusion

By proposing a Condition Monitoring data base, CM Base, we aim at filling some gap in our research community. The proposed application will allow finding datasets well-fitted to a specific study, related publications, proposing new datasets and sharing publications. It was designed to be easy to use, and easy to maintain. The structure is such that some new functionality can be added if requested by the users. Some suggestions have already been made: auto alert messages, forums to debate about datasets and methods, challenges proposed by companies on hot topics, groups of datasets used for contest or teaching purpose, overview publications and so on. The interest of the application will depend on its success. The more datasets and papers will be referred to in the base the wider its use will spread. We thus count on your contributions and suggestions to improve the application.

References

[1] M. Baker, Is there a reproducibility crisis? A Nature survey lifts the lid on how researchers view the crisis rocking science and what they think will help, Nature, vol. 533, no. 7604, 2016, p. 452+.
[2] T. Barszcz, How to bridge academia and industry ? A practitioner's perspective, Keynote lecture, Surveillance 8, Roanne, France, October 20-21, 2015.
[3] R. Klein, *Automatic diagnostics of complex machinery*, Keynote lecture, Surveillance 8, Roanne, France, October 20-21, 2015.
[4] J. Antoni et al, Feedback on the Surveillance 8 challenge: Vibration-based diagnosis of a *Safran aircraft engine*, MSSP, Volume 97, December 2017, Pages 112-144.
[5] *https://data.mendeley.com/*
[6] PHP Notes for Professionals, Free Programming Books, goalKicker.com
[7] G. E. Krasner, S.T. Pope, A Cookbook for Using the Model-View-Controller User Interface Paradigm in Smalltalk-80, JOOP, September 1988.
[8] T. Reenskauag, *The original MVC reports*, thing-model-view-editor, 1979.
[9] J. Gabay, *Merise et ULM pour la modélisation des systèmes d'information*, Dunod, 5ème édition, 2004, ISBN 2 10 007205 6.

Experimental investigation of sensor mounting positions for localized faults detection of epicyclic gear sets

Yu GUO, Zhen LIU, Xing WU, Yinxin YU

Faculty of Mechanical and Electrical Engineering
Kunming University of Science and Technology
Kunming City, Yunnan, China 650500
kmgary@163.com

Abstract

In the literature reports on the vibration based localized faults detection of epicyclic gear sets, the vibration sensor is often mounted on the ring gear or the housing adjacent to the ring gear. However, this sensor mounting position is often too ideal to be utilized in applications. There are different structures of epicyclic gear sets widely used. It is a challenge for selecting a suitable sensor mounting position. In this paper, the sensor mounted on the input/output shaft bearing house for picking up the vibration is experimental investigated, and compared with that from the conventional sensor mounting position. Analysis results shown that the bearing house position can also be employed to expose the fault features of planet gears.

1 Introduction

Planetary gearboxes as important rotating mechanical transmission units are widely used in the wind turbines, helicopter, automobiles, and marine vehicles. However, due to rigorous working conditions in applications, and the wear and crack failures of planetary gear sets occurred frequently [1]. Therefore, the condition monitoring and faults diagnose of planetary gear sets become more and more important.

Compared with the vibration picked up from a fixed-axis gearbox, the vibration picked up from a planetary gearbox are more complicated. In a conventional single-stage planetary gearbox, several planet gears rotate around the sun gear and rotate around their own centers simultaneously, and each planet gear meshes with the sun gear and the ring gear at the same time. On the other hand, the epicyclic motions of planet gears make the mesh positions changed from time to time, which cause the vibration transfer paths between the gear mesh positions and the fix-mounted sensor position are time-varying. As a result, rich modulation phenomena can be observed in the picked up vibration from a planetary gearbox.

As well known that picking vibration is one of important steps in the condition monitoring and faults diagnosis. Generally, it is expected that the picked vibrations contain more significant fault features, which can reduce the requirement to the following signal processing procedures. Then, the sensor mounting position should be considered carefully at first. However, the sensor mounting position of the planetary gearbox is few discussed by literature. In most theoretical literature experimental studies, vibration sensors are often mounted on the ring gear or the housing adjacent to the ring gear [1-3]. However, this sensor mounting position is too ideal to be utilized in applications. Due to different structures of epicyclic gear sets. To some structures, e.g. the ring gear is rotating, the ring gear position is impossible to mount the sensor at all. It is a challenge for the vibration analysis based condition monitoring and faults diagnosis. In addition, part of the research is based on mounting the sensor inside the planetary gearbox for fault diagnosis. For example, mount the sensor on the planet carrier [4,5]. For the built-in sensor installation method, the sensor installation space in a planetary gearbox should be considered firstly, which will put forward new requirements on the design, manufacture and disassembly for the planetary gearbox. In other word, this method has no universal applicability. On the other hand, similar to the sensor installation method for the fixed-axis gearbox picking up the vibration, which mounted the sensor on the input/output shaft bearing housing, should be a feasible sensor mounting position

for planetary gearboxes. However, this mounting position has a problem that the vibration transfer path between the mesh point of the ring-plant gear pairs and the sensor is much longer comparing with the conventional ring gear related sensor mounting position. It is worth carrying out an investigation on the different sensor mounting positions for applications of the condition monitoring and faults diagnosis of planetary gearboxes. To address this issue, an experimental investigation has been carried out on a planetary gearbox test rig for the vibration based tooth-root crack faults detection. The well-known vibration separation technique and the synchronous averaging are utilized to extract the faults characteristics through the vibration picked up from different sensor mounting positions for a comparison. Experimental results show that the fault features contained in the observed vibration from bearing housing are weaker than that obtained from ring gear position. However, the bearing housing sensor mounting can also be utilized for the vibration based tooth faults detection by using the vibration separation and the synchronous averaging.

2 Planetary Gearbox Transmission

As shown in Fig. 1, a single-stage planetary gearbox is generally composed of a sun gear, several planet gears, a ring gear and a planet carrier [6]. Different from the conventional fixed-axis gearbox, the vibration transfer paths between the sensor and the meshing points are time-varying under the running condition of the planetary gearbox. Three structures of epicyclic gear sets commonly used are shown in Figs. 2(a)-(c), which correspond to the planetary gear set with the standstill ring gear, with the standstill sun gear, with the rotating sun gear and the rotating ring gear, respectively [7].

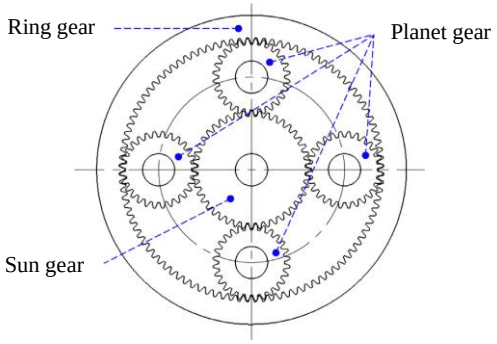


Figure 1: A planetary gear set with 4 planet gears.

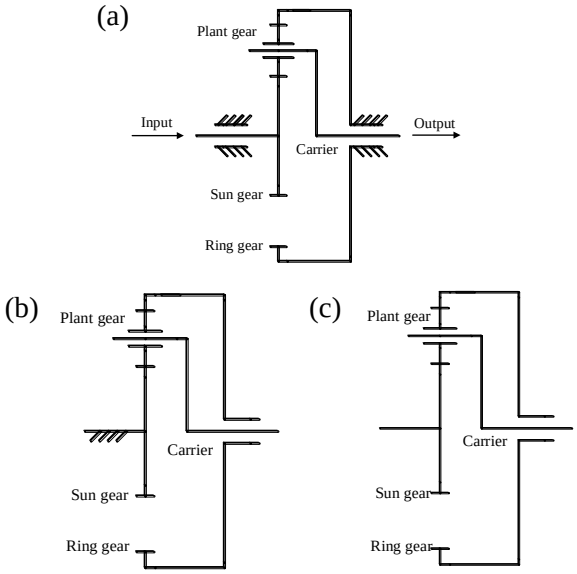


Figure 2 : Structures of planetary gear sets with: (a) standstill ring gear, (b) standstill sun gear, (c) rotating sun and ring gears

In literature reports of the vibration analysis of planetary gear sets, the sensor mounting position is often based on the structure with a standstill ring gear shown in Fig. 2(a). It is noted that the ring gears can also be rotating as the structures shown in Figs. 2(b) and (c), which rise challenges for selecting a suitable sensor mounting position. As well-known that the bearing housing utilized as the sensor mounting position for a fixed-axis gearbox is widely adopted in applications. However, the bearing housing sensor mounting for the vibration analysis of planetary gear sets is still an issue. It has two obviously drawbacks. Firstly, the vibration transfer path is much longer than that of the sensor mounted on the ring gear. Secondly, the interferences from adjacent bearings can lead the picked vibration much noisy. Then, the bearing housing sensor mounting is few reported in the literature. However, the bearing housing sensor mounting can be implemented in most applications of epicyclic gear sets. Then, it is worth investigating whether the bearing housing sensor mounting position is suit for the faults detection of epicyclic gear sets.

In this paper, experimental studies and vibration analysis for the planetary gearbox test rig with a tooth-crack of a planet gear have been carried out based on the vibration picked up by sensor 2 and 3 at the sensor mounting positions shown in Fig. 3, respectively.

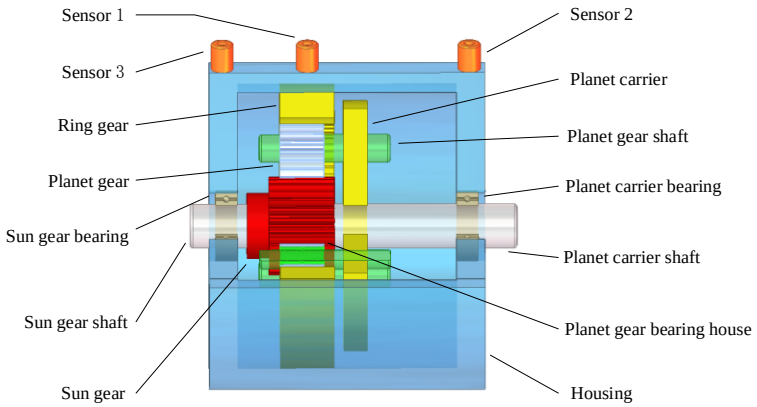


Figure 3: Installation location of sensors.

3 Briefs on Vibration Separation and the Synchronous Averaging

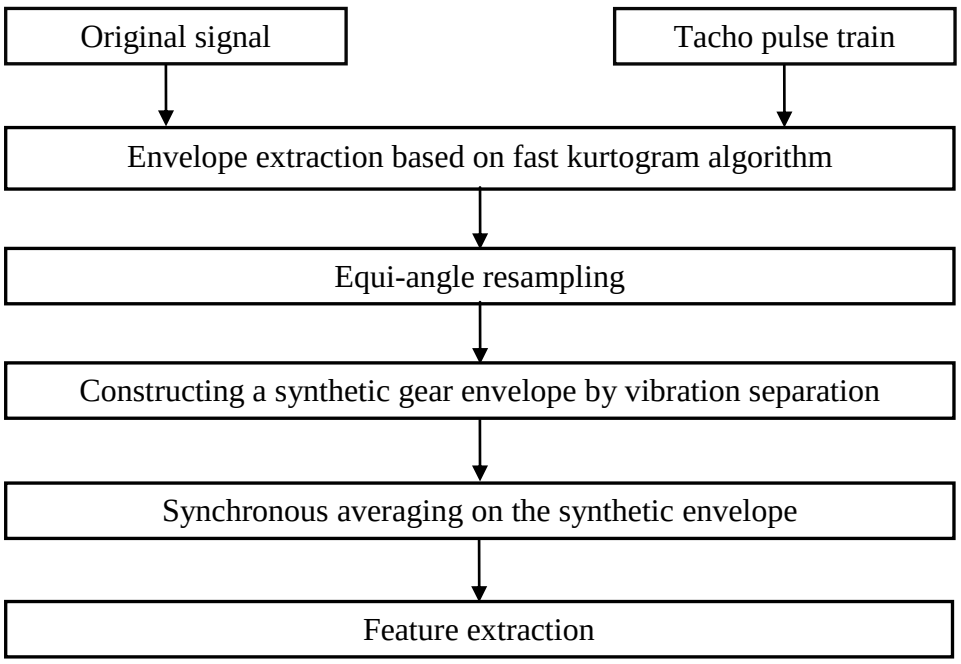


Figure 4 : Brief description of the method flow

As mentioned above, the vibration transfer path between the tooth mesh position and the sensor is time-varying under the condition that the sensor is mounted on the ring gear or the housing adjacent to the ring gear. It is worth mentioning that even though the length of vibration transfer paths seem to be a constant under the condition that the sensor is mounted on the input/output shaft bearing house, the vibration picked up by the sensor is also time-varying. The reason is that the changes of the meshing positions of planet gears make the picked vibration on the bearing house acting as a rotating vector. Therefore, the vibration separation [8,9] and the synchronous averaging [8] should be employed to eliminate the effects from the time-varying vibration transfer path. To extract the weak fault features of planetary gear sets, a combination scheme of envelope extraction, vibration separation and synchronous averaging has been proposed in [6] for the fault diagnosis of planetary gear sets recently. In this scheme, the envelope is demodulated to make the weak impulsive localized fault feature of the planetary gear sets prominent, the well-known vibration separation is utilized to eliminate the influence of time-varying transfer path, and the angular-domain synchronous averaging is used to reduce the influence of speed fluctuation and improve the signal-noise ratio (SNR). The schematic of the envelope-windowed vibration separation is shown in Fig. 4. The main steps are listed as follows [6].

(1) *Envelope extraction*. The fast kurtogram algorithm [10] is employed to extract the complex envelope from a maximum spectral kurtosis value determined demodulation frequency band.

(2) *Equi-angle resampling on the envelope*. Performing the equi-angle resampling scheme on the imaginary and the real parts of the complex envelope in the time domain, the complex envelope in the angular domain can be obtained.

(3) *Constructing a synthetic gear envelope by vibration separation*. The vibration separation technique is applied on the envelope in the angular domain. Then, a synthetic envelope is constructed according to the teeth mesh sequence of the interesting gear.

(4) *Synchronous averaging on the separated envelope.* The synchronous averaging is utilized on the synthetic envelope to remove the non-synchronous components.

(5) *Feature extraction*. The order envelope spectrum is calculated by the Fast Fourier Transform (FFT). Then, the fault feature can be exposed.

More details on the scheme of envelope windowed vibration separation and synchronous averaging can be found in [6].

4 Feature Frequencies of Planetary Gear Sets

Understanding the feature frequencies or orders of planetary gear sets is the core in the condition monitoring and faults diagnosis, by which we can explain the frequency lines in a spectrum reasonably and make a decision on the health status of the planetary gear set. Assuming f_c denotes the rotation frequency of the carrier, the meshing frequency f_m can be given as in [9] by

$$f_m = N_r f_c = N_p (f_p + f_c) = N_s (f_s - f_c) \tag{1}$$

where N_r is the teeth number of the ring gear, N_p is the teeth number of the planet gear, N_s is the teeth number of the sun gear, f_p denotes the absolute rotation frequency of the planet gear, and f_s represents the absolute rotation frequency of the sun gear. Using the carrier as the reference, the corresponding meshing order l_m can be calculated by

$$l_m = \frac{60 f_m}{n_c} = \frac{60 N_r f_c}{60 f_c} = N_r \tag{2}$$

where n_c is the carrier speed. The feature frequency of the planet gear with a tooth-root crack f_p^r is the rotation frequency of planet gear relative to the planet carrier, which is expressed as in [9] by

$$f_p^r = f_p + f_c = f_c (N_r / N_p) \tag{3}$$

And the feature order of the planet gear with a tooth-root crack can be given by

$$l_p^r = \frac{60 f_p^r}{n_c} = \frac{60 f_c (N_r / N_p)}{60 f_c} = N_r / N_p \tag{4}$$

5 Experimental verification

5.1 Experimental description

In the investigation, the experiments have been carried out on a planetary gearbox test rig for the vibration based tooth-root crack faults detection. The test rig is a 75 kW transmission system, which is driven by an AC drive motor with an adjustable speed range from 0 to 2500 rpm. The loading unit is an AC induction generator in tandem with a digital AC drive to regenerate the power back into the system. The test unit is a single-stage planetary gearbox (type: NGW 2K-H) as shown in Fig. 5 with the specific parameters listed in Table 1. The DH904 eddy current sensor is mounted on position 1 shown in Fig. 5 for picking up the tacho pulse train. Three DH112 acceleration sensors are mounted at positions 2, 3, and 4 respectively for picking up the vibrations from the planetary gearbox. A NI 9234 4-channel card is used for the data acquisition with sampling rate 51.2 kHz. It is worth mentioning that the planet carrier shaft is the input and the sun gear shaft is the output. The position 2 is on the sun gear bearing housing, the position 3 is on the housing adjacent to the ring gear, and the position 4 is on the planet carrier bearing housing. The data series used in this study are the vibrations picked up at positions 2 and 3 respectively. And the tacho pulse train is obtained by the eddy current sensor at position 1. In order to simulate the tooth-root crack fault of a planetary gear, a crack of about 4 mm is machined at the tooth root of a planetary gear by the wire cutting method as shown in Fig. 6.

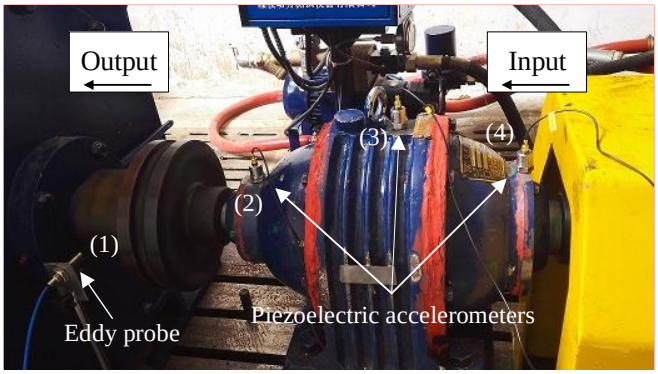


Figure 5 : Test rig of planetary gear transmission

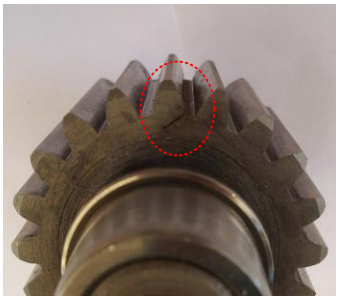


Figure 6 : Planet gear with tooth root crack

Gear	Sun gear	Planet gear	Ring gear
Number of teeth	28	20	71

Table 1: Parameters of planetary gearbox

In the experiment, the rotation speed of the input shaft is at about 1000 rpm. The characteristic orders of the planetary gear in the planetary gearbox can be calculated by Eqs. (2) and (4) in theory, which are listed in Table 2 by using the planet carrier as the reference.

Planetary gearbox meshing order l_m	71.00 ×
Planetary fault order l_p^r	3.55 ×
Planetary carrier frequency order l_c	1.00 ×
Planetary frequency frequency order l_p	2.55 ×

Table 2 : Characteristic orders of the planetary gearbox

5.2 Experimental data analysis

The waveforms in of the raw vibration picked up by the sensors mounted at positions 2 and 3 are shown in Figs. 7 and 8, respectively. The tacho pulse train obtained by the eddy probe at position 1 and the corresponding speed curve are shown in Figs. 9 and 10.

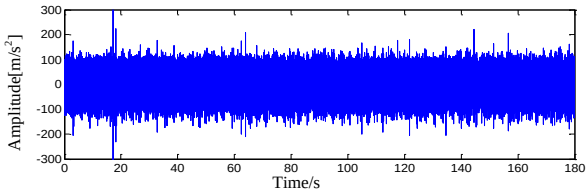


Figure 7 : Vibration waveform observed at position 2

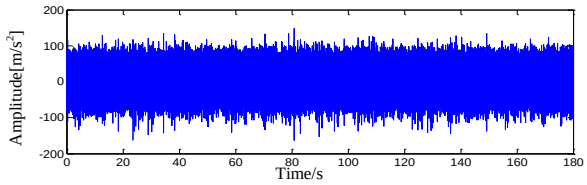


Figure 8: Vibration waveform observed at position 3

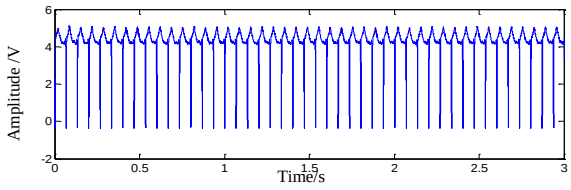


Figure 9: Speed pulse

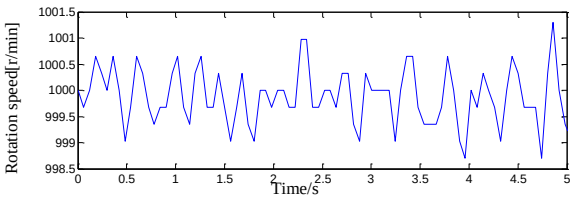


Figure 10 : Speed curve

Using the vibration separation and the synchronous averaging techniques, the order spectra of the vibrations picked up at the positions 2 and 3 are shown in Fig. 11, where the characteristic order (3.55×) and its harmonics related to the tooth-root crack of a planet gear order are exposed clearly. Experimental results show that the fault feature contained in the observed vibration from the bearing housing (position 2) is weaker than that obtained from the ring gear position (position 3). However, it is worth noting that the bearing housing sensor mounting position can also be utilized for the vibration based tooth faults detection by using the vibration separation and the synchronous averaging techniques.

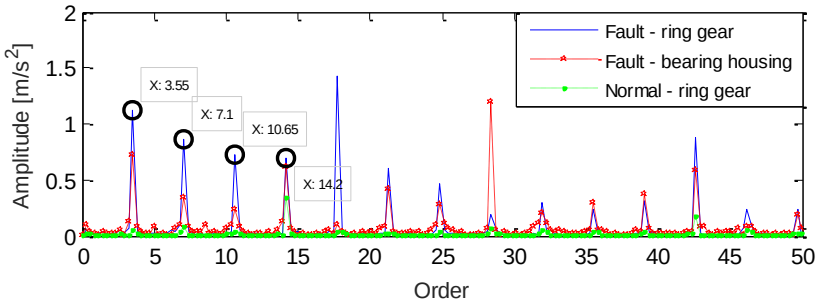


Figure 11 : Fault characteristic order of planet gear fault

6 Conclusion

In practical applications, different structures of epicyclical gear sets widely used. According to a specific structure of a planetary gear set, the sensors should be mounted on suitable positions to pick up vibration. The experimental investigation results show that the sensors mounted on the position of the bearing housing and mounted on the housing adjacent to the ring gear both can be utilized for the vibration based tooth faults detection by using the vibration separation and the synchronous averaging techniques. Moreover, the experimental results also show that the bearing housing sensor mounting position is also can be employed for the faults detection of planetary gear sets, but the fault feature is weaker than that obtained from ring gear position.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 51675251).

References

[1] Y. Guo, L. Zhao, X. Wu, et al., *Vibration separation technique based localized tooth fault detection of planetary gear sets: A tutorial*, Mechanical Systems and Signal Processing, 2019, 129: 130-147.

[2] Y. Lei, D. Han, J. Lin, et al., *Planetary gearbox fault diagnosis using an adaptive stochastic resonance method*, Mechanical Systems and Signal Processing, 2013, 38(1): 113-124.

[3] Z. Feng, M.J. Zuo, J. Qu, et al., *Joint amplitude and frequency demodulation analysis based on local mean decomposition for fault diagnosis of planetary gearboxes*, Mechanical Systems and Signal Processing, 2013, 40(1): 56-75.

[4] Z. Fan, H. Li, *A hybrid approach for fault diagnosis of planetary bearings using an internal vibration sensor*, Measurement, 2015, 64(384):71-80.

[5] W. Smith, L. Deshpande, R. Randall, et al., *Gear diagnostics in a planetary gearbox: a study using internal and external vibration signals*, International Journal of Condition Monitoring, 2013, 3(2):36-41.

[6] Y. Guo, L. Zhao, X. Wu, et al., *Tooth Root Crack Detection of Planet and Sun Gears Based on Resonance Demodulation and Vibration Separation*, IEEE Transactions on Instrumentation and Measurement, 2019. doi: 10.1109/TIM.2019.2893011.

[7] Y. Lei, J. Lin, M.J. Zuo, et al., *Condition monitoring and fault diagnosis of planetary gearboxes: A review*, Measurement, 2014, 48:292-305.

[8] P.D. McFadden, *A technique for calculating the time domain averages of the vibration of the individual planet gears and the sun gear in an epicyclic gearbox*, Journal of Sound and Vibration, 1991,144(l):163-172.

[9] P.D. Samuel, J.K. Conroy and D.J. Pines, *Planetary Transmission Diagnostics*, NASA/CR 2004-213068.

[10] J. Antoni, *Fast computation of the kurtogram for the detection of transient faults*, Mechanical Systems and Signal Processing, 2007, 21(1): 108-124.

Towards 3D AFM Using Multiple Vibration Modes

Rubin Eyal[1], Bucher Izhak[1], Davis Solomon[1]

[1]Dynamics Laboratory, Mechanical Engineering, Technion
Haifa 3200003, Israel
bucher@technion.ac.il

Abstract

Atomic force microscopy (AFM) is used for measuring nano-scale topographic features. By exciting a micro-cantilever with a sharp stylus at its tip, at or near resonance, a Frequency Modulated AFM (FM-AFM) device can sense the change of resonance frequency due to the change in tip-surface Van der Waals (VdW) potential. The topography is then retrieved from the relationship between the potential and the distance between stylus and the specimen. To improve the measurement speed and address complex geometries emerging in industrial microchip constructions, several enhancements are introduced. While most FM-AFM devices operate in a single vibrating mode, this article enhances existing sensing methods by extending to multidimensional sensing the resonance frequencies that are modulated by the topology, in several orthogonal vibration modes simultaneously. The latter opens new possibilities, e.g. to measure steep walls and trenches or other complex geometries. An Autoresonance (AR) control scheme for faster excitation, and fast frequency estimation algorithm were used for sensing several modes simultaneously, without the need to wait for steady state settling of the cantilever. The concept was tested on a large-scale experimental system, where VdW forces between tip to surface were replaced by magnetic forces, using a magnetic tip and ferromagnetic samples. Experimental results employ 3D relevant topographies such as inclined surfaces, steep walls and trenches that were reconstructed experimentally with 4 (μm) resolution or better. Downscaling to typical AFM dimensions would theoretically yield sub-nanometer resolution. Numerical and experimental data are shown to demonstrate the advantageous of the new approach.

1 Introduction

Atomic force microscopy (AFM) first presented by Binning et al. [1] is used for measuring nano-scale topography features in the semiconductors industry, and for atomic resolution measurments in physics and biology research. The non-contact configuration of the sensor (NC-AFM) is based on a micro-cantilever with a sharp stylus at its tip, vibrating at or near resonance. The tip interacts with the surface, while the distance dependent Van der Waals (VDW) force between the tip and the surface, acts on the cantilever, altering its resonance frequency. A common technique used today is the frequency modulated AFM (FM-AFM), presented by Albrecht et al. [2]. The latter showed that the measured frequency shift correlate to the surface topology.

The present work enhances the one-dimensional measurement capability of current AFM devices to multidimensional measurement capability by allowing the measuring tip to vibrate at two or more directions simultaneously. Here, each direction of excitation is associates to a distinct mode of vibration, hence has a distinct resonance frequency. By measuring the frequency shift in each mode separately, one can evaluate the tip to surface distance in each measured direction [3]. An Auto-resonance (AR) control scheme combined with modal filtering [4] is used for fast frequency tracking of the different modes, and a fast frequency estimation algorithm [5] is used for evaluating the frequency shift. Using this method for two (or more) spatial modes simultaneously results with a multi-dimensional measurement, that has the ability to measure and reconstruct complex geometries. In addition, Since the AR locks into the natural frequency instantaneously, the measuring process can be expedited. This concept was demonstrated in the lab, by a

large-scale sensor, where the VdW forces have been replaced by magnetic ones, and the sensing tip is a passive magnet. The sensing tip interacts with a ferromagnetic specimen having a complex topology.

The paper introduces the multidimensional vibrating sensor, demonstrated on a large-scale experimental system in section 2. Section 3 shows some experimental results that are analysed to assess the sensitivity and resolution of the current system. Finally, Section 4 draws some conclusions.

2 Experimental system layout, actuation and sensing

This chapter explains the multimode actuation and sensing concept on the laboratory large-scale experimental system, and the way to interpret the measured data to reconstruct the topology of the measured specimen.

2.1 Multimode actuation and sensing

The measuring magnetic tip is attached to a rectangular cantilever beam which has two orthogonal modes of vibration with distinct natural frequencies at 447 (Hz) and 549 (Hz). Finite element model of the cantilever showing the two direction of motion is illustrated in Figure 1.

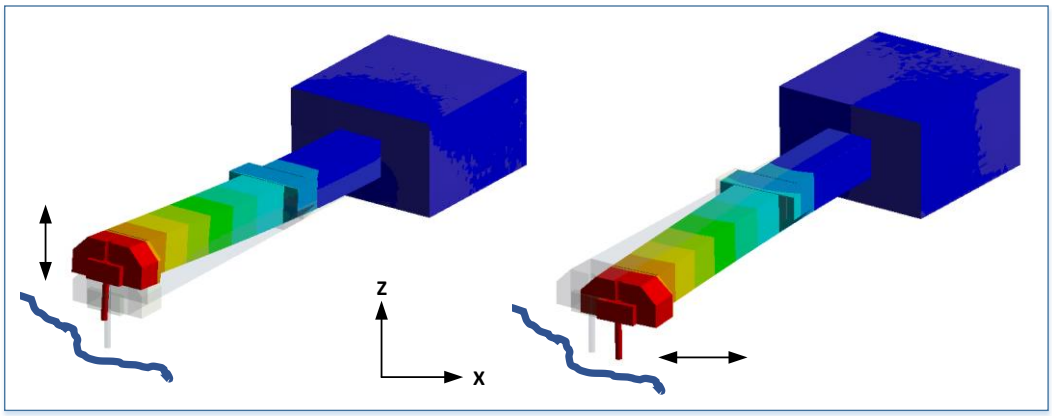


Figure 1: Multidirectional vibration- Left: vetical motion of the tip (z), Right: horizontal motion of the tip (x)

The interaction between the magnetic tip and the surface, due to the magnetic field, contributes to the total potential energy of the cantilever undergoing bending. Hence, one can write:

$$V_{Total} = V_{cantilevel} + V_{ts} = \frac{1}{2}\left(k_x x^2(t) + k_z z^2(t)\right) + V_{ts}(x(t), z(t)) \qquad (1)$$

where V_{ts} is the tip to surface interaction potential, k_x, k_z are the modal stiffness and $x(t), z(t)$ are the time dependent modal coordinates of the tip associated with the direction of vibration in x, z respectively.

The magnetic potential depends on the distance between the magnetic tip and the ferromagnetic surface. Hence, the interaction contribution can be added to the constant stiffness matrix of the cantilever:

$$\mathbf{K} = \begin{bmatrix} k_x & 0 \\ 0 & k_z \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 V_{ts}}{\partial x^2} & \frac{\partial^2 V_{ts}}{\partial x \partial z} \\ \frac{\partial^2 V_{ts}}{\partial x \partial z} & \frac{\partial^2 V_{ts}}{\partial z^2} \end{bmatrix} \triangleq \mathbf{K}_0 + \Delta\mathbf{K} \qquad (2)$$

Non-zero of-diagonal terms in $\Delta\mathbf{K}$ indicates that the specimen is inclined. In the present research, the information in x and z appears at 2 distinct frequencies and we obtain only the diagonal terms of $\Delta\mathbf{K}$, one at a time. The distance of the tip to the specimen is deduced from these terms only.

In order to excite each direction at a different frequency simultaneously we employ modal control according to the configuration shown in Figure 2. The cantilever is excited with two voice-coil actuators placed symmetrically at $45°$ on both sides of the cantilever. Two laser displacement sensors (Keyence LK-H008) are also positioned at $45°$ on both sides of the cantilever measuring the displacement of the cantilever

close to the tip. Modal filtering [6] is performed digitally in order to project the measured data and excitation forces on the exact modes directions.

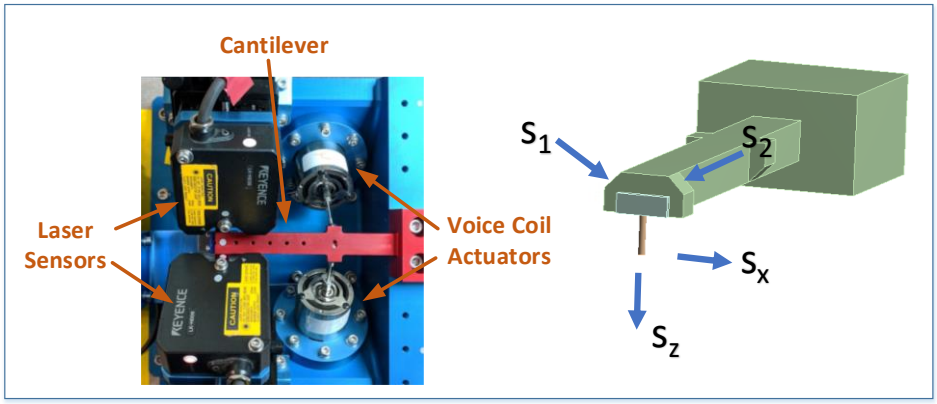


Figure 2: Left: laboratory realization using 2 voice-coil actuators acting at 45 degrees, and two optical (Keyence™) sensors measure in the same directions. Right: Illustration of the measurement of two sensors from inclined surfaces on the cantilever close to the magnetic tip.

The modal filtering (MF) uses the bi-orthogonality between the vibration modes, by taking a linear combination of the two sensors signals so that all modes are canceled but one. An illustration for the modal filtering of the displacement signals, measured by the laser sensors, is presented in

Figure 3 , where S_1, S_2 are real measured signals and S_z, S_x are the modal filtered displacements from the following equation:

$$\begin{aligned} S_x(t) &= S_1(t) - S_2(t) \\ S_z(t) &= S_1(t) + S_2(t) \end{aligned} \tag{3}$$

The separation of modes can be clearly seen on the frequency domain after applying a fast Fourier transform (FFT) to the measured and modal-filtered signals. The latter is demonstrated on the measured signals, and a similar approach is performed on the excitation forces operating on the cantilever by two voice-coil actuators.

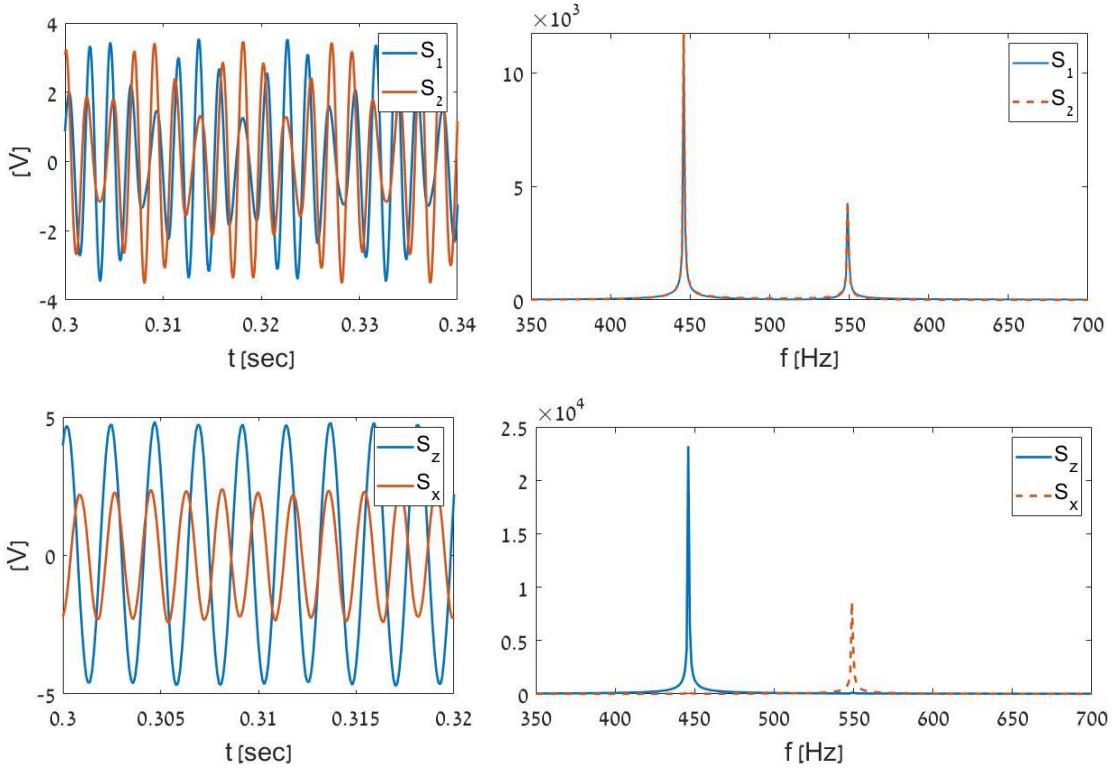


Figure 3: Top-left: measured signals S_1, S_2 from the experimental system. Top-right: S_1, S_2 frequency content showing both resonance frequencies are present in both measured signals. Bottom-left: S_x, S_z after applying modal filtering (3). Bottom-right: S_x, S_z frequency content showing the separation of frequencies in the modal-filtered signals.

Modulation of the natural frequency is detected by auto-resonance (AR) resonance tracking control scheme, combined with a new frequency estimation algorithm [5] , that uses Linear Least Squares (LLS) to fit an instantaneous phase to the noisy signal for several periods, and then fits a time dependent line whose slope is an estimate of the signal's frequency. Autoresonance, also known as self-excitation, is a well-known nonlinear feedback method used for automatically exciting a system at its natural frequency [7], [8]. It locks into resonance from the first cycle, and has the potential to increase the imaging speed compared to the common scheme employed in AFM - Phase Locked Loop (PLL), whose frequency locking speed is greatly reliant on the settling time of the cantilever. Both resonance tracking and frequency estimation are performed simultaneously for the two modes of vibration using modal filtering.

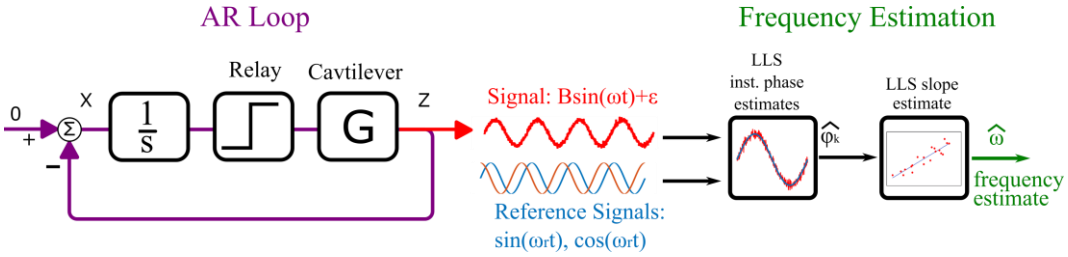


Figure 4: Schematic digram of the resonance tracking and frequency estimation. The AR feedback loop described here consist of a phase shifting element P which shifts the phase of the input signal by 90 degrees, and a digital 'relay' or sign function that forces the amplitude of the input signal to constant values [4].

Assuming white noise, the variance of the frequency estimated using the frequency estimation algorithm [5] depends significantly on the total estimation time t_N, and also on the sampling rate and on the SNR - signal to noise ratio between the variance of the single tone signal without the noise, to the variance of the noise. This variance bounds the minimal frequency shift (Δf_{min}) that can be measured by the sensor. Measurement accuracy and speed should be balanced for a desired working point.

$$\sigma_f^2 \approx \frac{12}{F_s t_N^3 SNR} \tag{4}$$

2.2 From frequency modulation to surface topology identification

Two main scanning methods can be used to scan a specimen [9]: Constant height or constant gap (frequency). In the former the probe is located in a constant height above the surface, and the change is the resonant frequency is measured and then converted into the change in surface topography. In the latter, the gap between the probe and the sample stays constant using a standard PI control scheme that keeps a constant natural frequency for single and dual mode excitation. In the large-scale demonstrator presented in this paper, an XYZ stage instantaneous location (with a $0.2\,\mu m$ resolution) was used for the surface topology estimation.

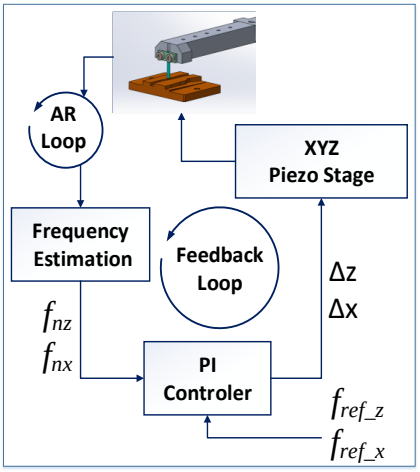


Figure 5: Control scheme for position control. The feedback loop keeps a constant resonance frequency by changing the position of the sample relative to the vibrating tip, keeping a constant gap between them. The XYZ stage displacements are then use to reconstruct the surface topology.

2.3 Laboratory experimental system implementation

The abovementioned measuring methodology is demonstrated using a large-scale experimental system. A schematic layout of the system is presented in Figure 6. The main subsystems consist actuation of the cantilever beam using two voice-coil actuators, measurement of the beam tip displacement using two optical sensors (Keyence™), and specimen displacement using XYZ accurate piezo stage (Nanomotion™). All the vibrating parts are mounted on an optical table through a rigid aluminum base in order to decrease the peripheral vibrations of the system, as detailed in Figure 7. The measured and actuation signals are generated in real time by a Xilinx FPGA connected to a dSpace™ signal processor, and are integrated to Matlab, together with the XYZ stage control.

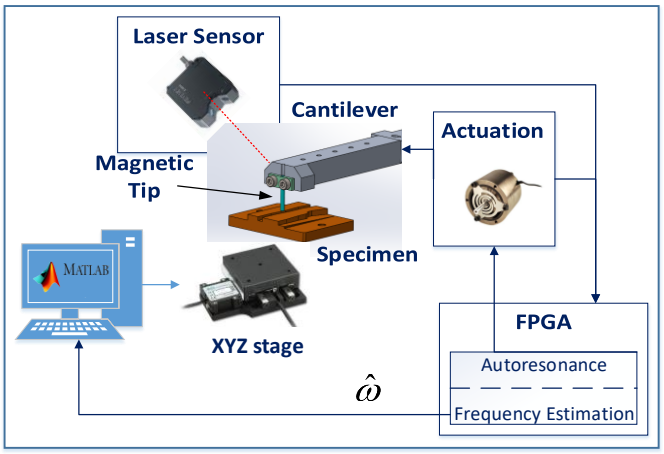


Figure 6: Large-scale experimental system layout

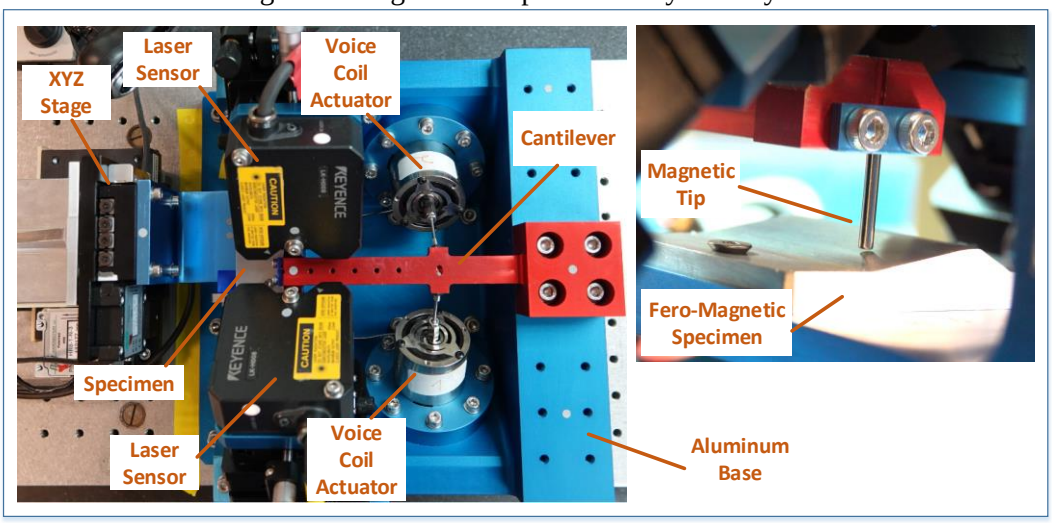


Figure 7: Detailed view of the large scale experimental system. Cantilever beam (in red) length- 130 mm and cross-section- 12mm X 15mm. Magnetic tip diameter- 2 mm and length- 20 mm

3 Experimental results

This chapter will present some of the main measured results of the large-scale experimental system, together with resolution and performance analysis and a relevant down-scale approximation to a real AFM.

3.1 Main results

First result in fig8 show the nonlinear dependency between the tip-surface distance (gap) and resonance frequency that effects the sensitivity of measurement for different gaps. Though, for small gaps and small oscillations (i.e. around 0.1 mm), linear relationship is obtained. Furthermore, once the control system (presented in chapter 2.2) moves the specimen to obtain the same reference frequency, the gap also remains the same, and the nonlinear dependency is effectively eliminated. Hence, the measurement is carried out with the same sensitivity.

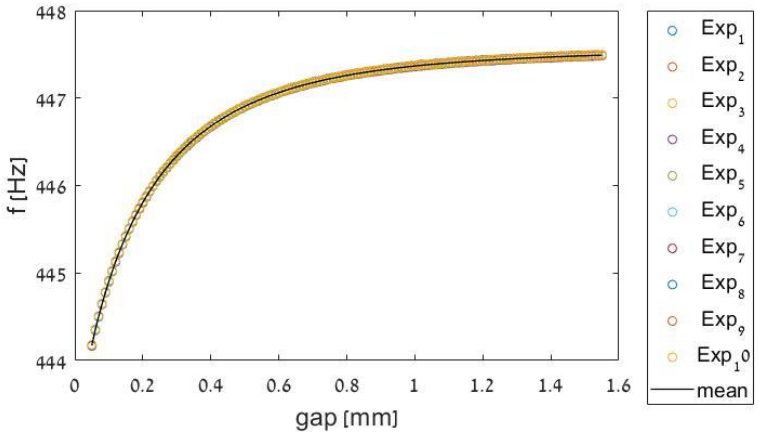


Figure 8: Dependency of resonance frequency on the distance between the measuring tip and the specimen. The experiment was repeated 10 times showing repeatability.

Steep and complex geometries, such as walls, grooves and trenches cannot be fully reconstructed by measuring a single spatial direction. Using the multidimensional measurement method presented in this paper, complex geometries were reconstructed experimentally as demonstrated in Figure 9: Left: measuring tip next to the steep walled and narrow trench specimen. Right: reconstructed geometry showing both the geometry of the wall and of the inclined bottom surface (3D).Figure 9 and Figure 10.

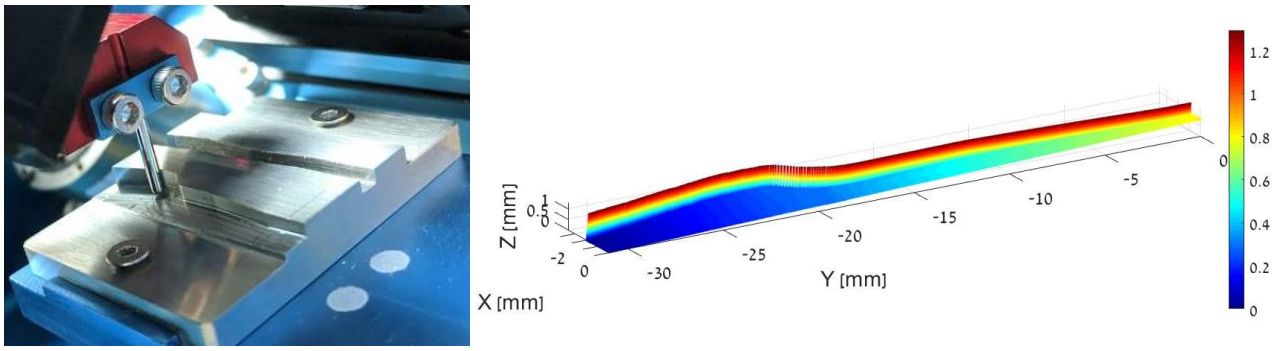


Figure 9: Left: measuring tip next to the steep walled and narrow trench specimen. Right: reconstructed geometry showing both the geometry of the wall and of the inclined bottom surface (3D).

The narrow trench geometry shown on the left in Figure 10 has been reconstructed, including the steep walls curvature and slight machined angle of the bottom surface. Still, since the magnetic field is not as localized as VdW forces in AFM, parasitic cross coupling effects distort the reconstructed geometry. These effects are not expected to occur on the nanoscale where tip-sample interaction, dominated by the local VdW forces, is more localized then the magnetic forces employed here.

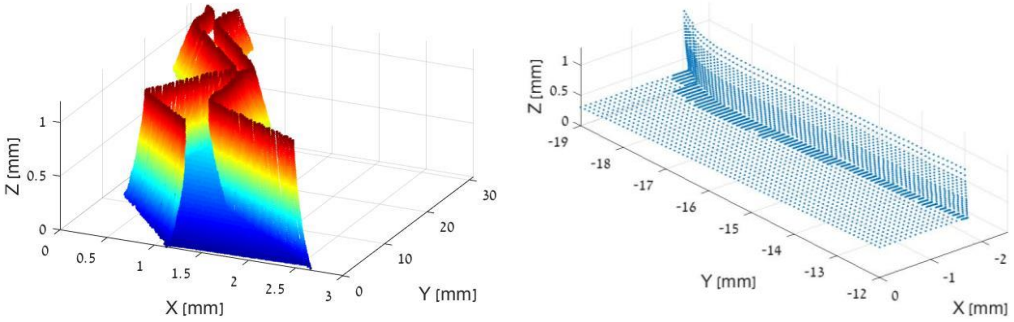


Figure 10: Left: reconstructed geometry of a narrow trench. Right: scanning step size manipulation using multiple mode information. Near the right angled corner, small steps were taken, which was made possible by sensing both vertical and horizontal gaps

The right side of **Error! Reference source not found.** illustrates the use of the additional information in each measuring point for improving the scanning algorithm, e.g. decreasing the scanning step size in

advance, while approaching a vertical wall by using the horizontal sensing mode to sense that the geometrical gradient become large.

3.2 Performance analysis

The experimental system is a large scaled AFM, hence the performance should be compared to a commercial FM-AFM [10] using relevant scaling factor. Magnet size, shape and orientation, spatial resolution of XYZ stage, frequency estimation resolution, sensors resolution, external disturbances and noise are some of the many parameters affect the measuring resolution of the large-scale experimental system. The resolution of the commercial AFM was measured for a specific image size, hence a similar method was exploit in the large-scale system. As consequence, a suitable non-dimensional comparison is the ratio between the resolution and the image size. The resolution of the large-scale experimental system was measured by scanning a flat rectangular surface with 10,201 data points, and calculating the standard deviation of the measurements. The RMS measured resolution is 1.3 (μm) for the vertical mode (z), and 4 (μm) for the horizontal mode (x). The reason for the difference between the modes is the spatial effects of the magnetic field created from the cylindrical shape of the magnet. The magnetic forces that affect the frequency shift are larger in the vertical direction, parallel to the axis of the cylindrical magnet, hence the resolution in the vertical direction is better.

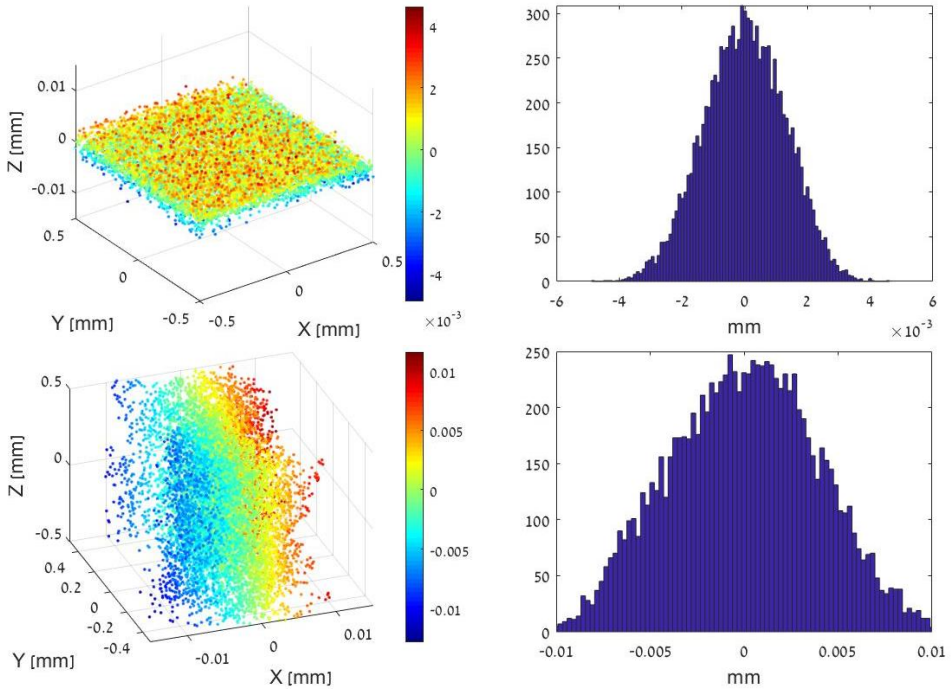


Figure 11: Total resolution measurement of the large-scale experimental system. Top left: vertical mode (z) measured data points. Top right: statistical analysis for the vertical mode showing a histogram representing the measured gap probability distribution. Bottom left: horizontal mode (x) measured data points. Bottom right: statistical analysis for the horizontal mode.

Table 1 compares the large-scale experimental system to a commercial FM-AFM system [10] showing similarity in non-dimensional parameters $\Delta f_{min} / f_0$ and resolution/image size, where f_0 is the basic resonance frequency of the cantilever, and Δf_{min} is the minimal measured frequency shift. The similarity approves the validity of the up-scaling done in the demonstration of the multidimensional measuring method.

Parameter	Commercial AFM [10]	Large Scale System Vertical Mode (z)	Large Scale System Horizontal Mode (x)
f_0 (Hz)	$330 \cdot 10^3$	447	549
Δf_{min} (Hz)	20	0.012	0.012
$\Delta f_{min} / f_0$	$6.06 \cdot 10^{-5}$	$2.68 \cdot 10^{-5}$	$2.18 \cdot 10^{-5}$
image size	0.2 (μm)	1 (mm)	1 (mm)

resolution	0.38 (nm)	1.3 (μm)	4 (μm)
resolution/image Size	0.0019	0.0013	0.004

Table 1: comparing parameters between a commercial AFM to the large-scale experimental system

4 Conclusions

A method able to measure complex geometries with multidirectional probe has been presented. The method combines modal filtering with Autoresonance, for fast resonance tracking, simultaneously in two or more spatial directions. The method was demonstrated on a large-scale experimental system that was able to reconstruct 3D relevant topographies, such as inclined surfaces, steep walls and trenches, with 4 (μm) resolution or better. A nano-scale system based on the same principles is currently being constructed, and is expected to improve measurement speed and ability to measure complex geometries with Angstroms resolution.

Acknowledgments

This research has been funded by The Israel Innovation Authority, within Multi-Dimensional Metrology (MDM) Consortium, 2018-2020.

References

[1] G. Binnig, C. Gerber, and C. F. Quate, “Atomic Force Microscopy,” *Phys. Rev. Lett.*, vol. 56, p. 930, 1986.

[2] T. R. Albrecht, P. Grütter, D. Horne, and D. Rugar, “Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity,” *J. Appl. Phys.*, vol. 69, no. 2, pp. 668–673, 1991.

[3] F. J. Giessibl, “A direct method to calculate tip-sample forces from frequency shifts in frequency-modulation atomic force microscopy,” *Appl. Phys. Lett.*, vol. 78, no. 1, pp. 123–125, 2001.

[4] S. Davis and I. Bucher, “Automatic vibration mode selection and excitation; combining modal filtering with autoresonance,” *Mech. Syst. Signal Process.*, vol. 101, pp. 140–155, 2018.

[5] S. Davis and I. Bucher, “A new approach to single-tone frequency estimation via linear least squares curve fitting,” Apr. 2019.

[6] L. Meirovitch and H. Baruh, “Control of Self-Adj oint Distributed-Parameter Systems,” *J. Guid. Control. Dyn.*, vol. 5, no. 1, pp. 60–66, 1982.

[7] W. E. Vander Velde, “Multiple-input describing functions and nonlinear system design.” McGraw-Hill, New York, 1968.

[8] K. J. ÅStröM, “Oscillations in Systems with Relay Feedback,” pp. 1–25, 2011.

[9] N. C. Santos and F. A. Carvalho, *Atomic Force Microscopy Methods and Protocols*. 2019.

[10] H. Seo, D. Goo, G. Jung, R. Stomp, and J. Wei, “Park NX-Hivac : Phase-lock Loop for Frequency Modulation Non-Contact AFM,” *Park AFM SYSTEMS*. [Online]. Available: https://www.parksystems.com/index.php/applications/manufacturing/semiconductor/197-high-throughput-and-non-destructive-sidewall-roughness-measurement-using-3-dimensional-afm. [Accessed: 04-Apr-2019].

Fault diagnosis and prognosis for rolling bearings

Early bearing defect detection in a noisy environment based on a method combining singular value decomposition and empirical mode decomposition

Mourad Kedadouche, Zhaoheng Liu and Marc Thomas
Department of Mechanical Engineering, École de technologie supérieure,
Montréal, QC H3C 1K3, Canada
mourad.kedadouche@hotmail.fr, zhaoheng.liu@etsmtl.ca, marc.thomas@etsmtl.ca

Abstract

This paper proposes a new method combining Empirical Mode Decomposition (EMD) and Singular Value Decomposition (SVD) for bearing fault diagnosis. The method includes three steps. First, the signal is decomposed using EMD. Secondly, the instantaneous amplitudes are computed for each component using the Hilbert Transform (HT). Lastly, the Singular Value Vector is applied to the matrix of Cross-Power Spectral Density (CPSD) of the instantaneous amplitude matrix and the SVD versus frequency is analysed. The proposed method is first validated by using various noisy simulated signals. The results show that the proposed method is robust versus the noise to detect the bearing frequencies that are representative of the defect even in a very noisy environment and that the amplitude of the first SVD at each bearing frequency is very sensitive to the defect severity. The proposed method is also applied to two different experimental cases on bearings with very low degradation. The results show that the proposed method is able to detect bearing defects at an early stage of degradation for both experimental cases.

Keywords: Bearing fault, Empirical Mode Decomposition (EMD), Hilbert transform (HT), Cross-Power Spectral Density (CPSD), Singular Value Decomposition (SVD).

1. Introduction

Bearing wear can be considered as a major cause of defects in rotating machinery. Unexpected failures in bearings may cause significant economic losses. Empirical Mode Decomposition (EMD) is an interesting technique for fault diagnosis of rotating machinery. EMD can decompose the signal into several components called Intrinsic Mode Functions (IMFs) [1]. With EMD, the principal “modes” representing the signal can be identified. This method has attracted much attention for signal processing and engineering applications over the past decade [2]. The fundamental idea when using the EMD method is to decompose the vibratory signal into multiple components and the suitable IMF allows for computing the envelope spectrum and analyse their statistical features. Hybrid methods based on EMD and other techniques, like the Wavelet Packet Transform (WPT), the Support Vector Machine (SVM), Spectral Kurtosis (SK) and the Teager-Kaiser Energy Operator (TKEO), have also been applied to bearing fault diagnosis [3–7].

After performing EMD on a signal, some IMFs are associated to bearing faults, others with information unusable for diagnosing such faults. The useful IMF (if it exists) can be selected to perform the Hilbert spectrum. A few studies focus on developing an indicator to select automatically this useful IMF. Wenliao *et al*.[8] used the Wigner-Ville distribution to select the optimum IMFs and the filter bandwidth . Ricci *et al* [9] proposed a new indicator, named Merit Index, to select the appropriate IMF. The Merit Index is a linear combination between the periodicity degree of the IMF and its absolute skewness value. Yi *et al*.[10]

proposed a new indicator called Confidence Index based on combination of correlation coefficient, skewness and kurtosis. Jacek *et al*.[11], Peng et al. [12], Wang *et al*. [13], and Guo and Tse [14] utilized the correlation coefficient as an indicator analysis to select the candidate IMFs.

However, it is well known that the repetitive impacts due to the defect excite all bearing natural frequencies. If only the best IMF is selected, information included in other IMFs excited by the fault is then lost. Selecting all IMFs excited by the fault appears thus more suitable. In [15-16], the authors propose to select all the IMFs excited by the fault. A hybrid method based on EMD and run-up excitation is proposed to select the useful IMFs. By using a swept excitation when running up a rotating machine, the resonance frequency bands of the mechanical system is obtained from the spectrogram of the signal. In [17-18], the authors proposed to select all IMFs selected by the fault for early detection of the defect. The selection is made through an indicator-based kurtosis.

In this study, a new approach exploiting all IMFs of the signal to improve fault diagnosis is proposed. To compress all information extracted from each IMF, Singular Value Decomposition (SVD) is used in this paper. The SVD method has been widely used in fault feature extraction and identification for mechanical systems [19–21]. Before performing SVD, a matrix obtained from the original one-dimensional signal must be constructed. Various matrices exist, for example: the Toeplitz matrix, cycle matrix and Hankel matrix. To improve fault feature extraction, a new approach exploiting the instantaneous amplitude of each IMF obtained by EMD is proposed. The matrix is defined as the Cross-Power Spectral Density (CPSD) of all instantaneous amplitudes of the obtained IMF. Cross-spectral analysis is a powerful tool for investigating the relationship between signals in the frequency domain. Inspired by the frequency domain technique [17], the power spectral density matrix for each frequency is decomposed by applying SVD to the matrix. The singular value plot of the spectral density matrix concentrates information from all spectral density functions. The first singular value should approximately equal the sum of the terms on the diagonal of the PSD matrix. This means that the power of the signals at this frequency can be attributed to the vibratory signature [22]. The following sections give details of the proposed method used for rolling bearings fault diagnosis. The paper first presents the theoretical background of EMD and the proposed approach in Section 2. To validate the approach, the method is applied to a simulated signal and real data from damaged bearing in sections 3 and 4. Section 5 concludes the paper.

2. The proposed approach

2.1 A brief description of EMD

The EMD method can decompose signal in a multiple intrinsic mode functions (IMFs). The decomposed signal may be written as [1-2]:

$$x(t)=\sum_{i=1}^{N}C_i(t)+r_N(t) \qquad (1)$$

where $C_i(t)$ is the $ith\ IMF$ and $r_N(t)$ is the residual signal.

This method could suffer of a mixing mode problem and derived methods such as EEMD, CEEMD, CEEMDAN could also be used in this case, but it is not always necessary.

2.2 The proposed approach based on SVD and EMD

The approach proposed for diagnosing faults on rolling bearings is illustrated in Figure 1. In this paper, the signal matrix obtained from EMD is constructed as follows:

$$[C_1; C_2; C_3; ...C_n] \quad (2)$$

The first step is to compute the instantaneous amplitude of each intrinsic mode function $C_i(t)$. The instantaneous amplitude (*IA*) is computed by means of the Hilbert Transform. The analytical signal is given by the following expression:

$$z(t) = C(t) + j\tilde{C}(t) = a(t)e^{i\varphi(t)} \quad (3)$$

where $a(t)$ is the amplitude function given by the expression:

$$IA_i(t) = a(t) = \sqrt{(C_i)^2 + (\tilde{C}_i)^2} \quad (4)$$

The signal matrix of instantaneous amplitudes is constructed as follows:

$$M = [IA_1; IA_2; IA_3; ...IA_n] \quad (5)$$

The second step is to compute the cross-power spectral density of matrix M. The cross-power spectral density is defined by [23]:

$$P_{xy}(w) = \sum_{m=-\infty}^{\infty} R_{xy}(\tau)\, e^{-jw\tau} \quad (6)$$

where $R_{xy}(\tau) = E\{x(t)y(t+\tau)\}$ with $E[\cdot]$ denoting the expectation operator on t. (7)

The diagonal elements of the matrix represent the auto-power spectral density (the same *IA*). The off-diagonal elements are the complex cross-spectral densities between two different IA.

$$P(w_i) = \begin{bmatrix} P_{IA_1IA_1}(w_i) & P_{IA_1IA_2}(w_i) & ... & P_{IA_1IA_n}(w_i) \\ P_{IA_2IA_1}(w_i) & P_{IA_2IA_2}(w_i) & ... & P_{IA_2IA_n}(w_i) \\ ... & ... & ... & ... \\ P_{IA_nIA_1}(w_i) & P_{IA_nIA_2}(w_i) & ... & P_{IA_nIA_n}(w_i) \end{bmatrix} \quad (8)$$

The flow chart of the proposed method is shown in Fig. 1.

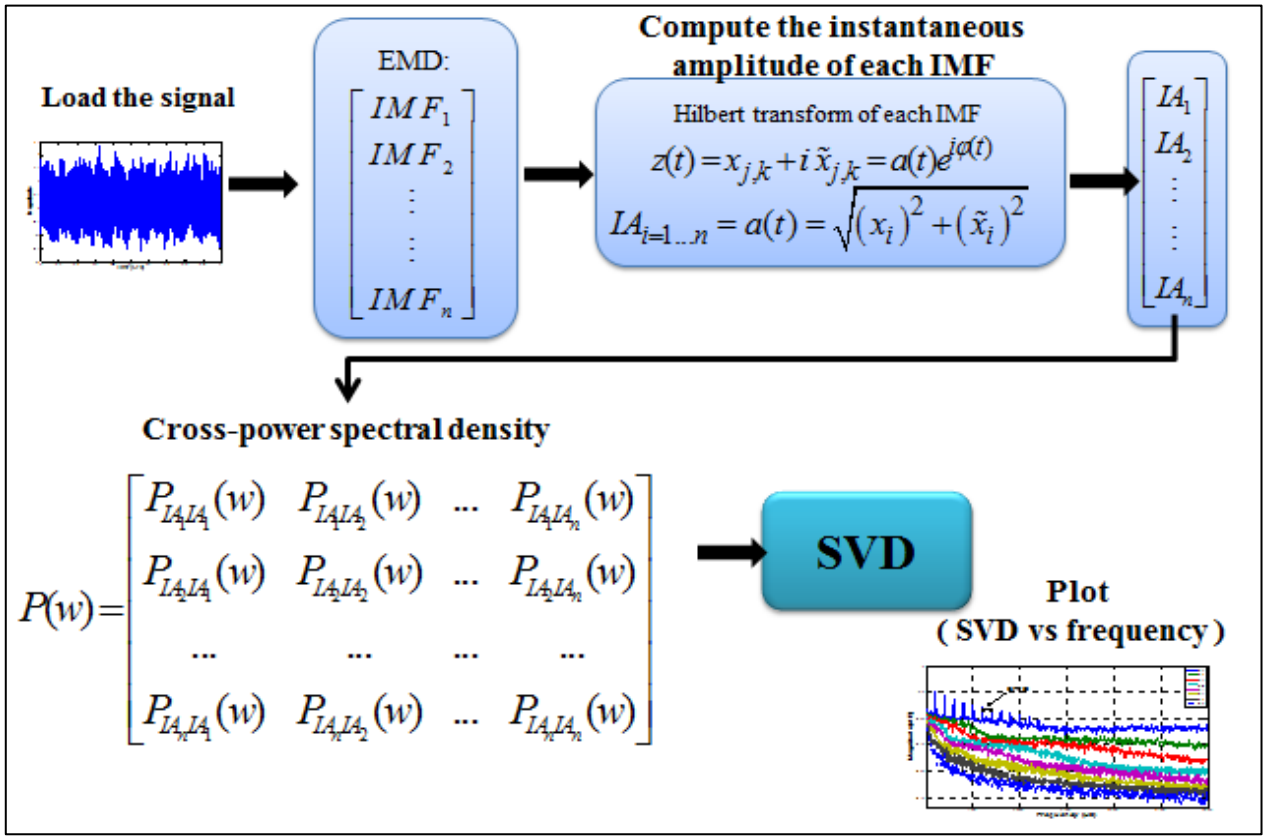


Figure 1: Flow chart of the proposed method

Assuming that $P(w_i)$ is an $n \times n$ matrix, the power spectral density matrix for each frequency (w_i) is decomposed by applying SVD to matrix $P(w_i)$.

$$P(w_i) = U \sum V^T \tag{9}$$

where U and V are orthogonal and Σ is a diagonal matrix of the singular values ($\sigma_{1,1} \geq \sigma_{2,2} \geq \sigma_{3,3} \ldots \geq 0$ and $\sigma_{k,j} = 0$ if $k \neq j$).

$$\Sigma(w_i) = \begin{bmatrix} \sigma(w_i)_{1,1} & \ldots & \ldots & 0 \\ 0 & \sigma(w_i)_{2,2} & \ldots & 0 \\ 0 & \ldots & \ldots & 0 \\ 0 & \ldots & \ldots & \sigma(w_i)_{n,n} \end{bmatrix} \tag{10}$$

As mentioned in the introduction, the first singular value should approximately equal the sum of the terms on the diagonal of the PSD matrix [22]. The plot of the first singular value versus frequency is thus used to identify the features extracted from the signal.

3. Validation with simulated data

3.1 Fault detection

In order to validate the proposed method and evaluate its effectiveness, a simulated numerical bearing signal is used. The simulated signal is similar to the signal used in [17-18]. The mathematical expression of the signal is given as:

$$x(t) = Ae^{-\alpha t'} \sin(2\pi f_n t) + n(t) \tag{11}$$

where

$$t' = mod\left(t, \frac{1}{F_m}\right) \tag{12}$$

Resonant frequency f_n is set to 1,800 Hz. The BPFO is set to 100 Hz. Amplitude A is set to 1. Sampling frequency F_s is set to 12,000 Hz. A random signal $n(t)$ with variance $\sigma^2 = 0.01$ is added to x(t) .

Figure 2 illustrates the simulated roller bearing signal. Figure 3 shows the 12-IMF obtained by the EMD method. It can be seen from this figure that the shocks related to defect are distributed across the first eight IMFs. As discussed in Section 2, all IMFs are considered in the proposed approach. The result is presented in Figure 4. Figure 4 exhibits the plot of the first singular value versus frequency. The fundamental of the BPFO (100 Hz) and its harmonics up to 1,200 Hz are clearly identified. The initial conclusion is thus that the proposed method can effectively detect the defect. No need to select the useful IMFs to accomplish the diagnosis.

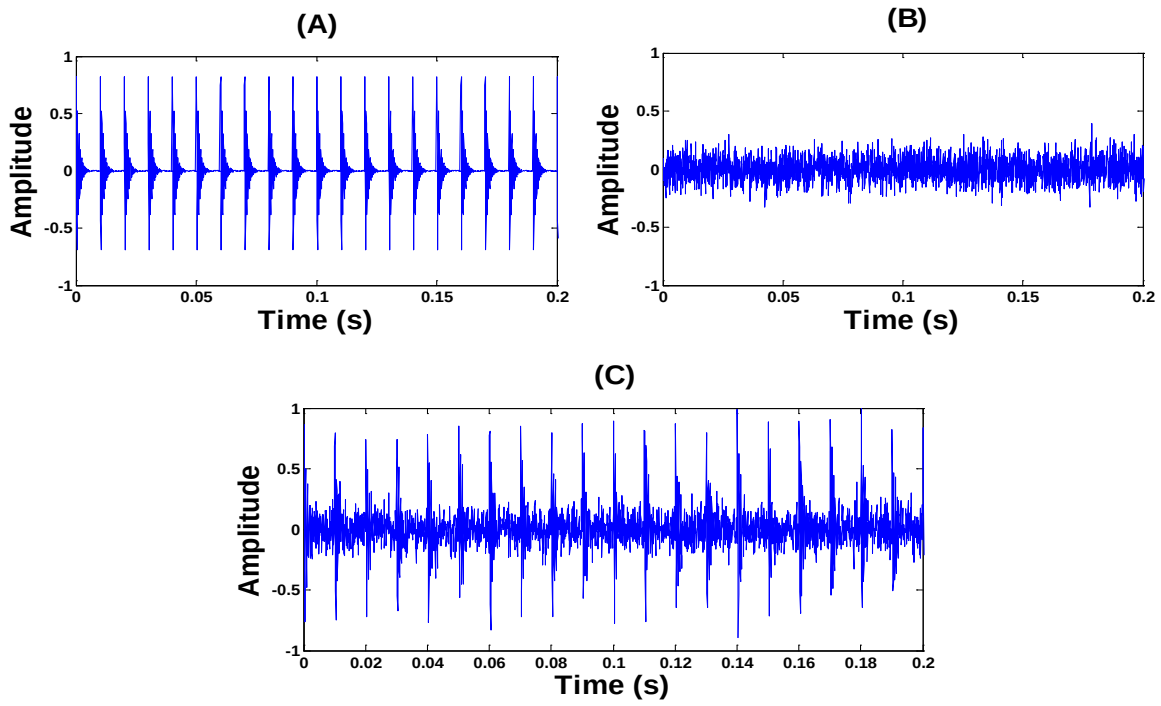


Figure 2: Simulated signal

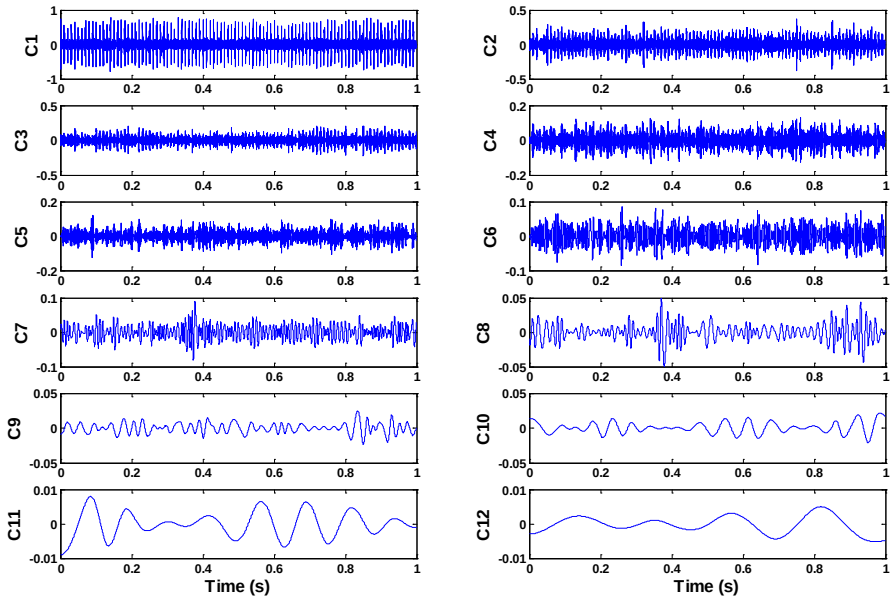


Figure 3: EMD of the simulated signal

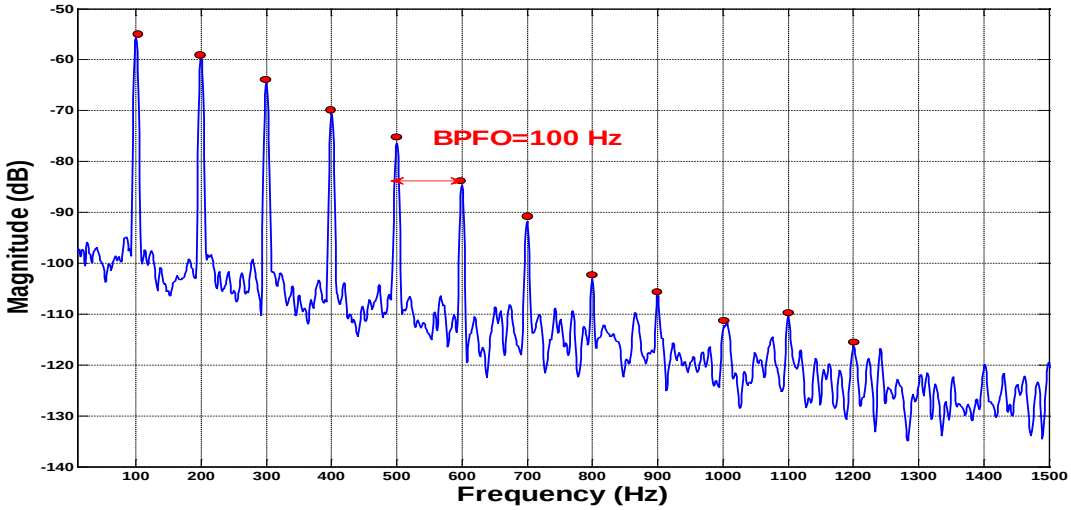


Figure 4: Result obtained for the simulated signal

It is well know that the amplitude of vibration due to bearing defects increases as the fault worsens and high peak levels may be observed. To confirm the efficiency of the method and its sensitivity to the severity of defect-induced vibrations, the simulated signal given by Equation (11) is simulated with A set to 1, 1.3, 1.6 and 1.9. The resulting waveforms are given in Figure 5.

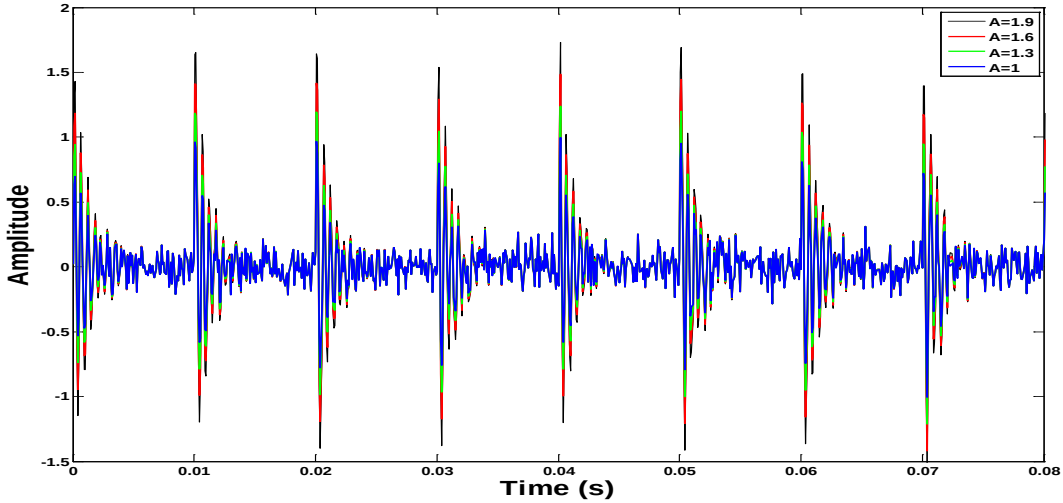


Figure 5: Simulated signal for different levels of A

Figure 6 shows the results obtained by the new approach for different levels of A. The magnitude of the BPFO and number of bearing frequency harmonics increase as the value of A increases. As seen in Figure 7, when $A=1.3$, the magnitude of the BPFO increases by 6 dB. The same conclusions may be drawn from the cases shown in Figure 7A. Therefore, the method is sensitive to the severity of the shocks. BPFO magnitude and the mean of all harmonics can be used as an indicator to track the severity of the defect.

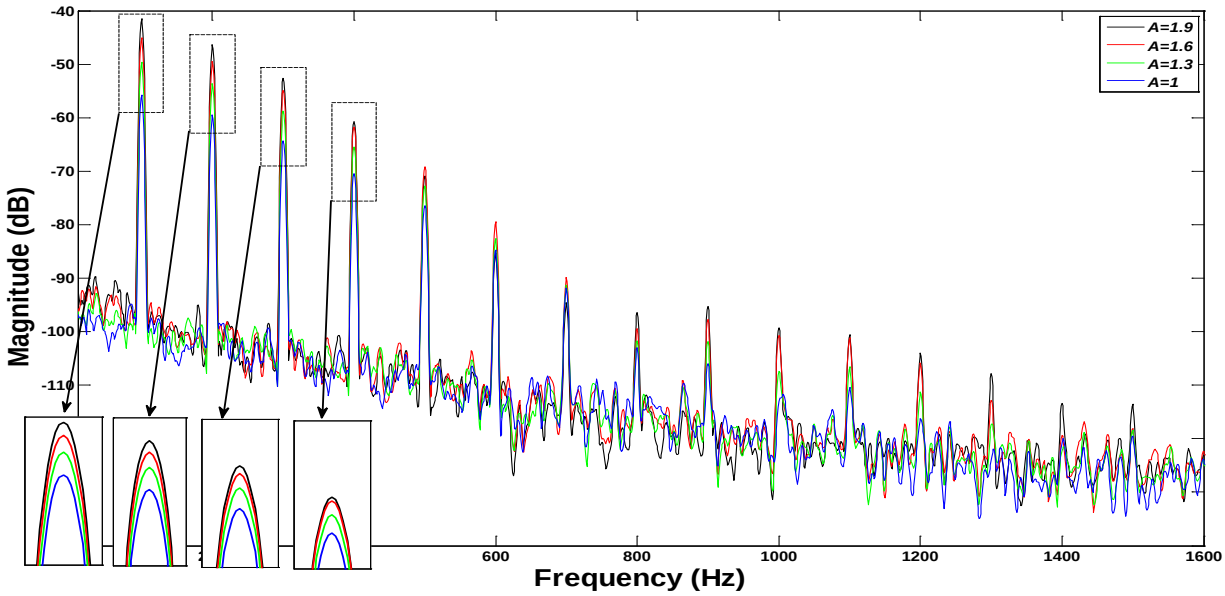


Figure 6: Results obtained for different levels of A

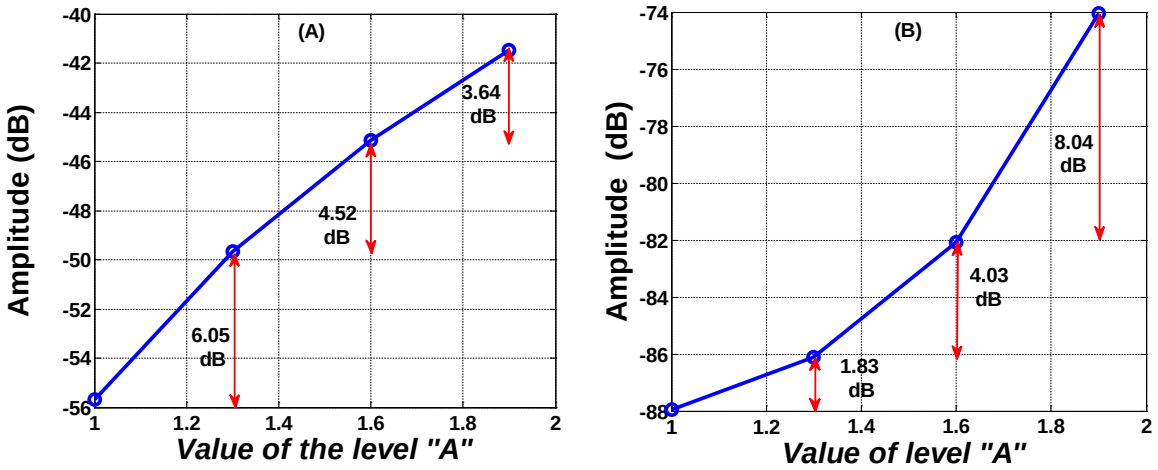


Figure 7: (A) Change in BPFO (B) Change in mean amplitude of all harmonics in Figure 6

The noise level in the simulated signal above is fairly low, though any industrial application would probably involve additive noises, potentially masking the signature of the defect, especially in the case of early bearing degradation. Another test was thus conducted to prove that the proposed method is able to detect the defect even if the noise level is higher.

3.2 Sensitivity to noise

The effectiveness of this method is evaluated based on its ability to extract bearing fault-related information. When a bearing is at an early stage of degradation, the signature of the defect may be masked by noise and difficult to extract. Gaussian white noise signals with variance σ^2 of 0.05, 0.1, 0.2 and 0.4 were thus added to the original simulated signal, $x(t)$, in order to evaluate the ability of the method to extract the defect-related information or signature when it is completely hidden by noise.

Figure 8 shows the simulated signals with the four values of variance σ^2 of added noise. It can be seen that defect-related shocks are masked by noise for $\sigma^2 = 0.1,\ 0.2$ and 0.4. The four noisy signals were processed using the proposed method and the results obtained from all signals are exhibited in the Figure 9. Note that even when the noise is very high, the method is able to identify the BPFO, while other tested state of the art methods were not available at this high noise level.

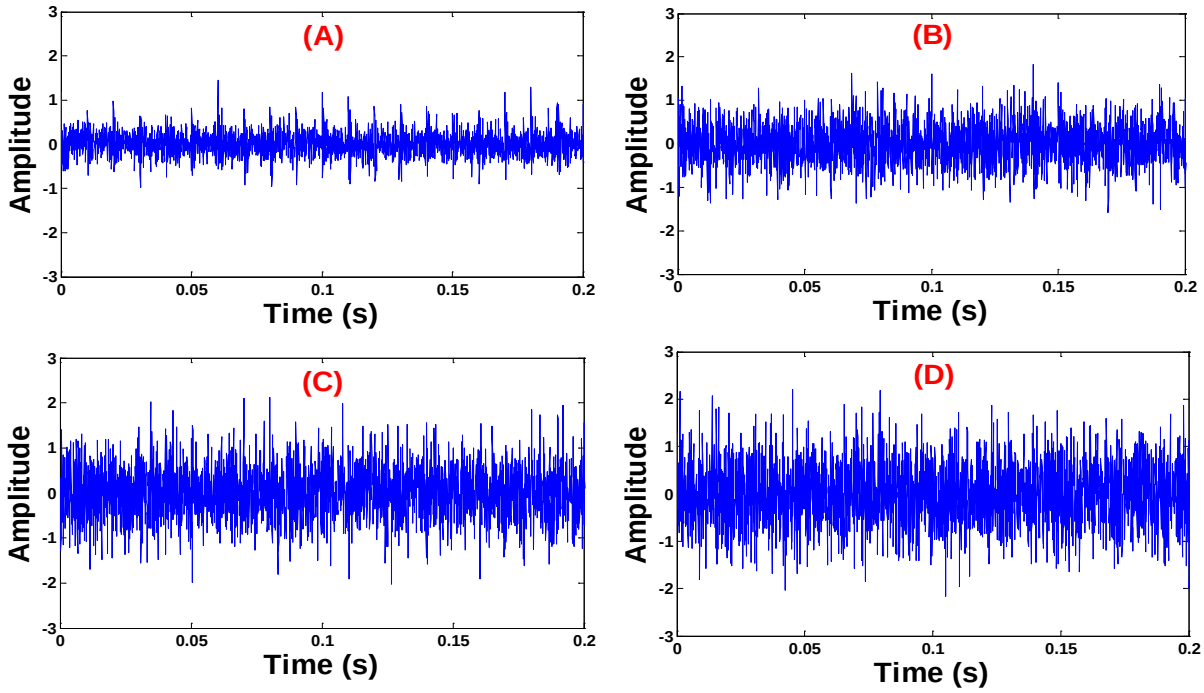


Figure 8: Simulated noisy signals: (A) $\sigma^2 = 0.05$; (B) $\sigma^2 = 0.1$; (C) $\sigma^2 = 0.2$; (D) $\sigma^2 = 0.4$

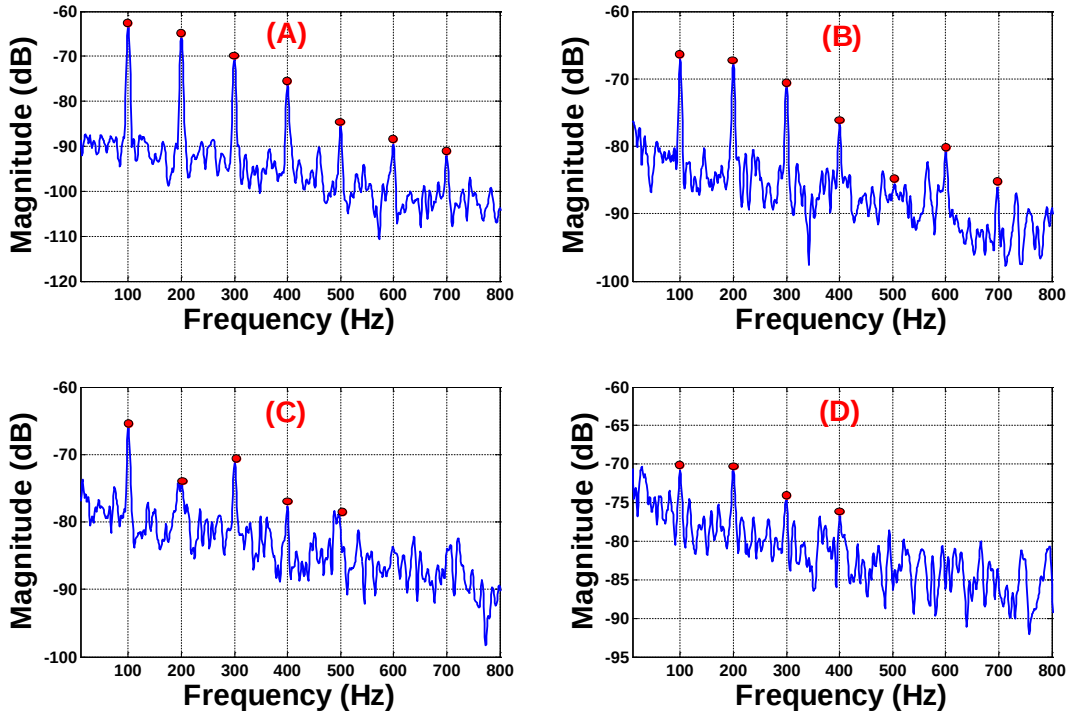


Figure 9: First singular value in frequency domain: (A) $\sigma^2 = 0.05$; (B) $\sigma^2 = 0.1$; (C) $\sigma^2 = 0.2$; (D) $\sigma^2 = 0.4$

4. Experimental data

The proposed method was investigated on two different history cases with different low levels of severity.

4.1 First case study

In the first case study, two defected bearings (SKF 1210 EKTN9) with very low severity levels are investigated. The test bench is shown in (Figure 10A). The first defected bearing noted D1, has an artificial defects about 200 µm deep with a groove width (W) of 50 µm and the second (D2) with W = 100 µm. The data were recorded when rotating at 600 rpm, using an accelerometer with a sensitivity of 100 mV/g. The frequency of the BPFO is about 76.46 Hz. The sample frequency is set to 12500 and 64,000 samples are used (acquisition time of 5 seconds).

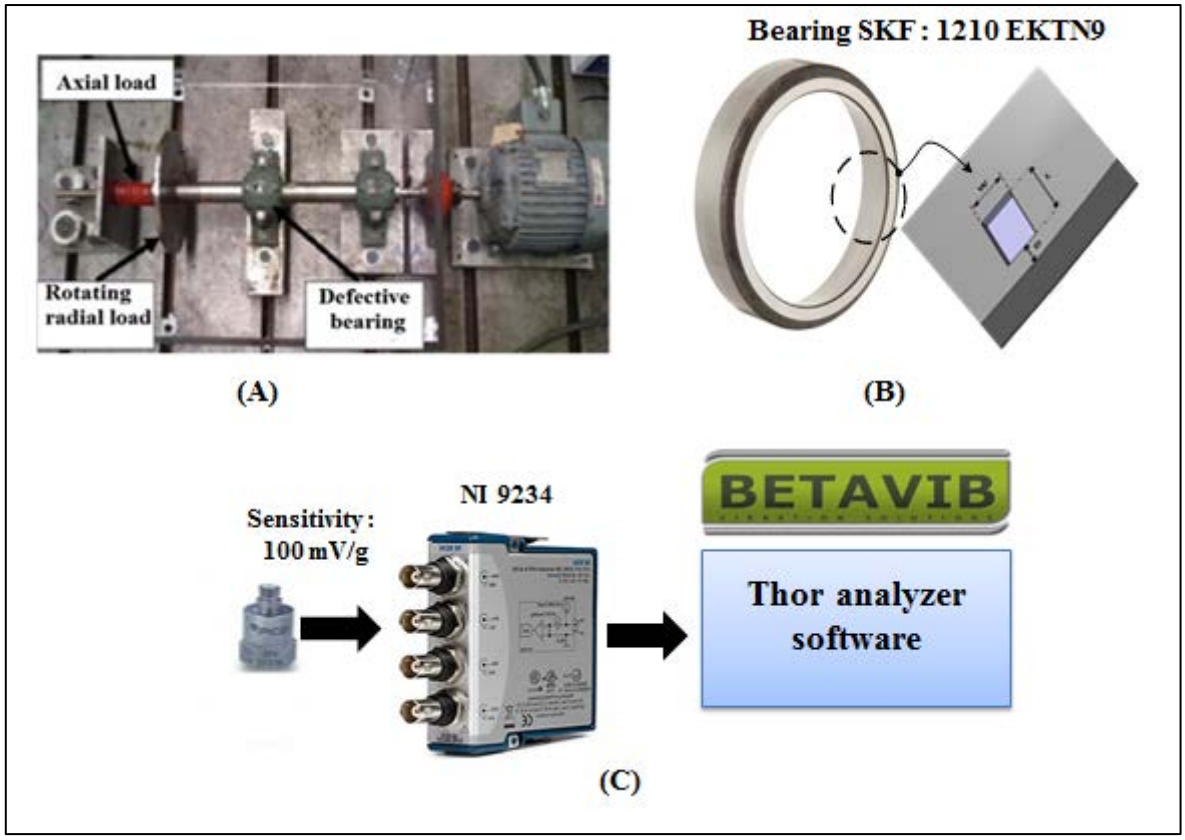


Figure 10: Test bench

Figure 11 shows the time signal of the acquired data. For the healthy bearing (Figure 11-A), quasi-random shock signals are observed. Defect D1 (50 µm) is at early stage of degradation. The time signal of defect D2 (Figure 11C) shows a series of impulse responses at BPFO and the amplitude is modulated periodically at the shaft speed. The shocks caused by defect D2 are more perceptible in the D2 signal than those due to defect D1, which are masked by noise.

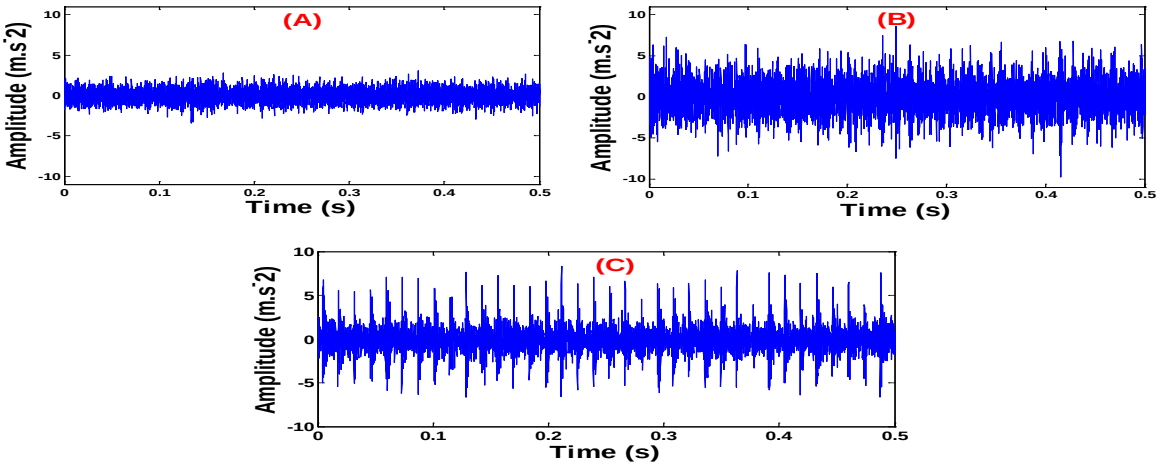


Figure 11: (A) Healthy bearing, (B) Defective bearing D1, (C) Defective bearing D2

The signals were processed using the proposed method. The defective bearing (50 µm) very clearly shows the BPFO (72.46 Hz) and its harmonics (Fig. 12). For defect D2 (100 µm), the BPFO harmonics are clearly identified and an increase in amplitude is observed (Fig. 13). The amplitude of the BPFO increases by 12.43 dB. This reveals that defects are more clearly identified and well-defined using the proposed approach.

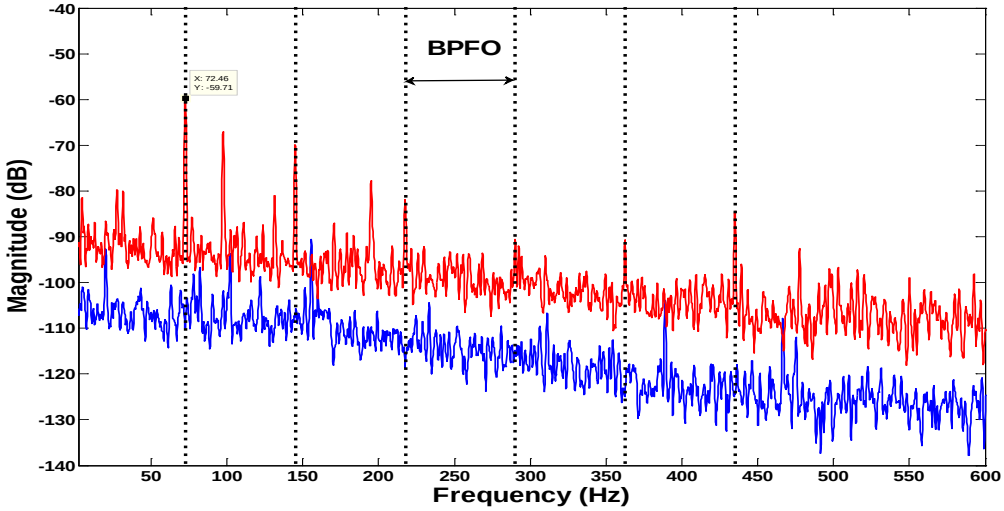


Figure 12: First singular value in frequency domain: Defect D1 (red), healthy bearing (blue)

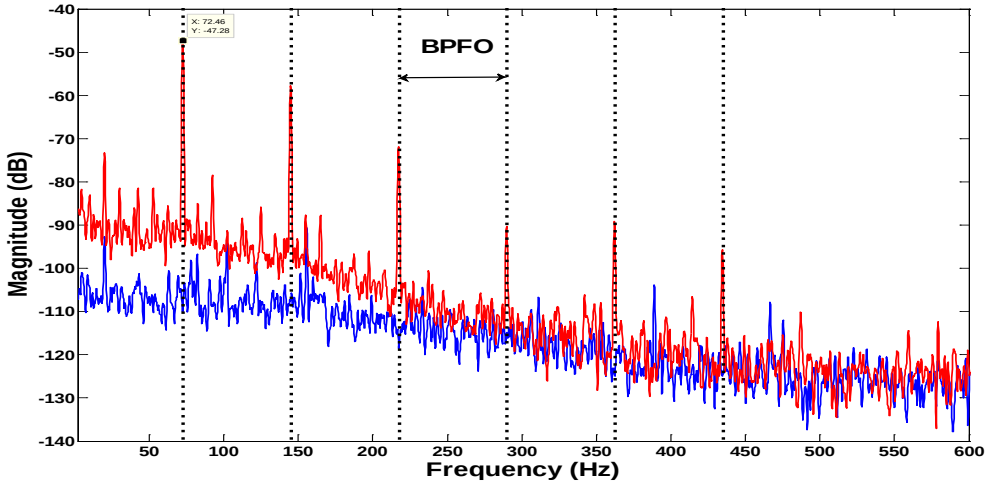


Figure 13: First singular value in frequency domain: Defect D2 (red), healthy bearing (blue)

4.2 Second case study

In this case, bearing (SKF 6205-2RS JEM) with defects in different locations are investigated. The vibration data are available in [24]. The test bench is shown in Figure 14. The first defect has a simulated single-point fault on inner race and the second defect has a simulated single-point fault on outer race. The fault size is 0.007" in diameter and 0.011" in depth.. The vibration data was collected when rotating at 1796 rpm. The frequency of the BPFO is about 107.6 Hz and the BPFI is 161.4 Hz. The sampling frequency (*Fs*) is 12,000 Hz and 60,000 data samples are used.

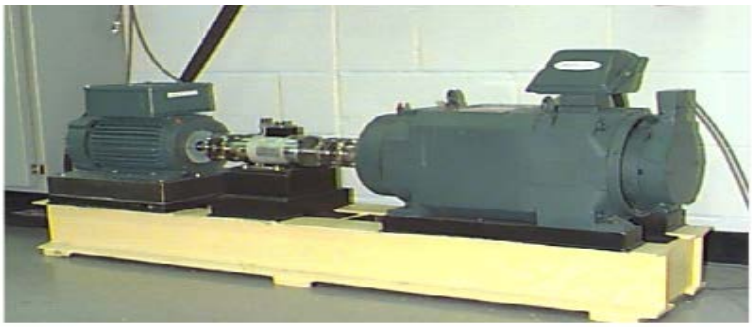


Figure 14: Test bench

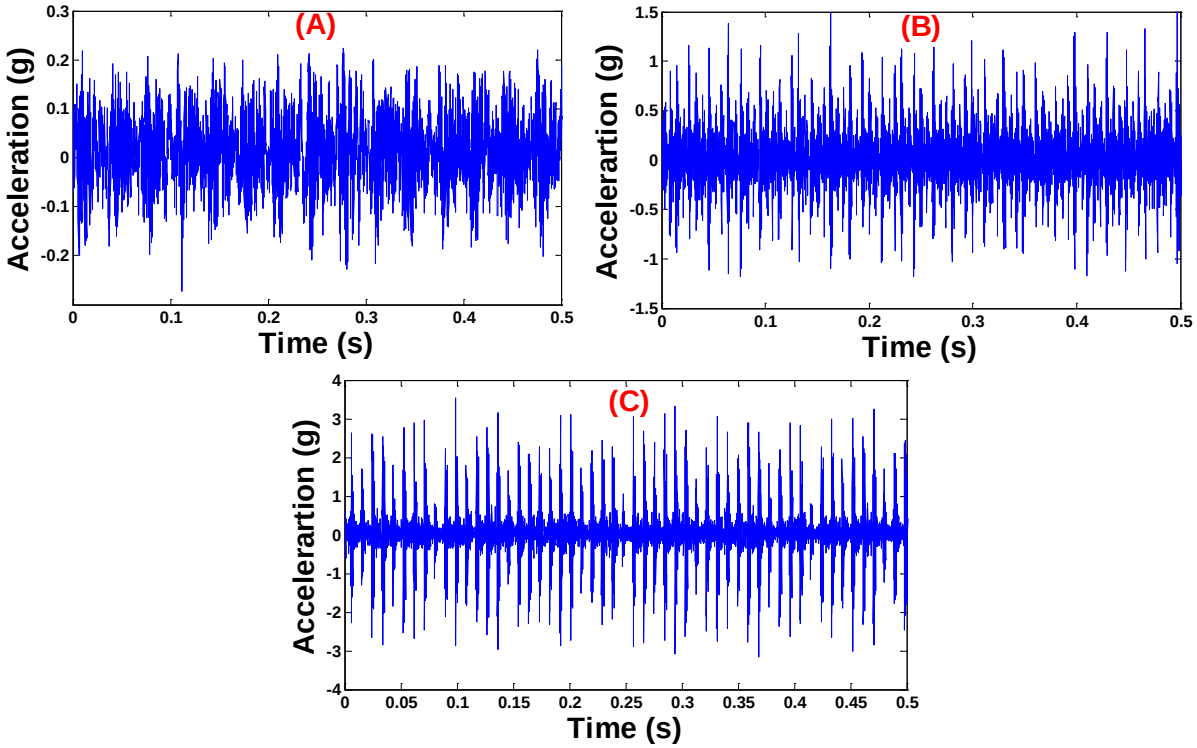


Figure 15: (A) Healthy bearing, (B) Inner ring fault (C) Outer ring fault

The time signals of the healthy bearing, inner ring fault and outer ring fault are shown in Figure 15. The time signal of the inner ring fault in Figure 15B shows that the amplitude is modulated periodically at the shaft speed. This is due to the rate at which the fault passes through the load zone. The signal is processed with the proposed method. Figure 16 illustrates the result. The result shows a series of harmonics of BPFI at 161.4 Hz, with sidebands spaced at the shaft speed to either side of each harmonic, as well as a number of harmonics of the shaft speed. This phenomena is usually related to a bearing looseness [25]. The

time signal of outer ring fault is shown in Figure 15C. Unlike that of the inner ring defect, the time signal of the outer ring fault should reveal a series of uniform impulse, but the signal is modulated at the shaft speed. The result obtained by the proposed method is shown in Figure 17. A series of harmonics of the BPFO are detected. As explained in [25], this modulation suggests a rotating load caused probably by mechanical looseness.

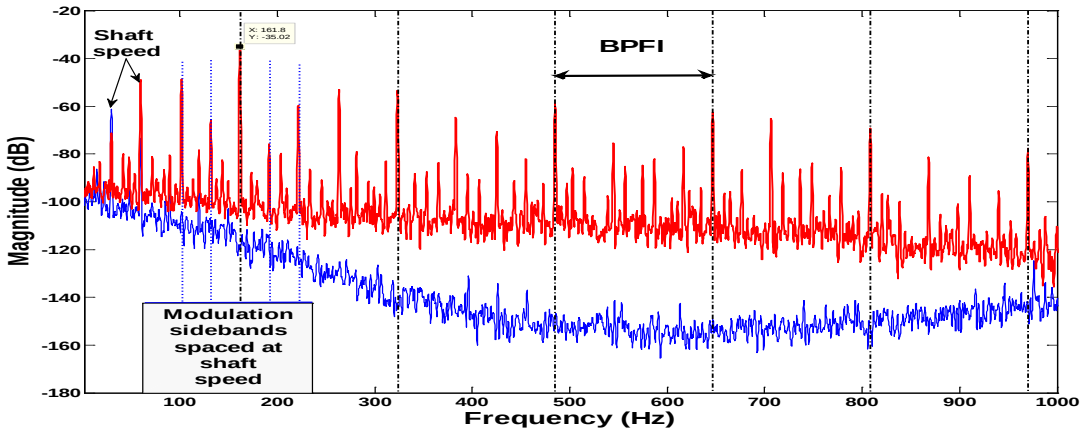


Figure 16: First singular value in frequency domain: Inner race fault (red), healthy bearing (blue)

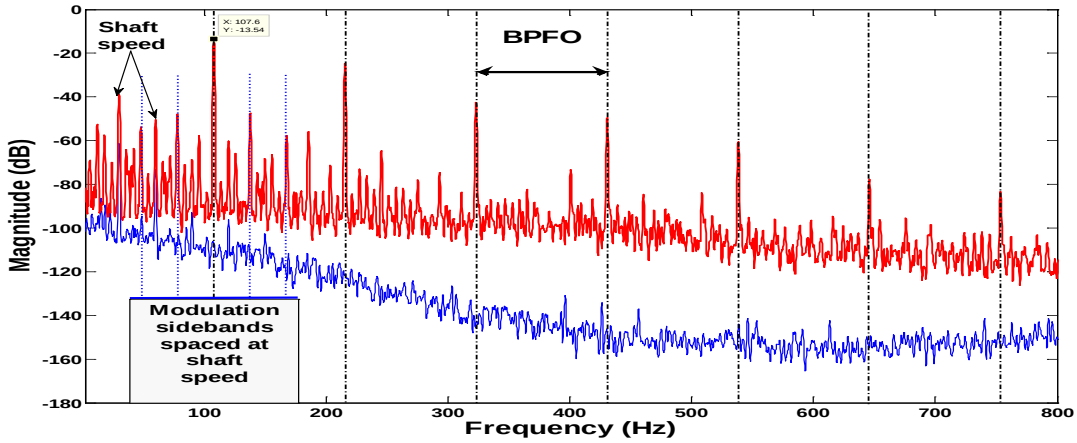


Figure 17: First singular value in frequency domain: Outer race fault (red), healthy bearing (blue)

5 Conclusion

A new feature extraction method for bearing fault diagnosis is presented in this paper. In the proposed method, there is no need to select the useful IMF to accomplish the diagnosis. The matrix of cross-power spectral density of all IMFs is decomposed in the frequency domain using SVD to extract defect-related information. The method was first validated by means of a simulated signal. The results have shown that this method may be used even in a very noisy environment. The proposed method can effectively detect defects even if the induced shocks are completely masked by noise and that the features extracted are sensitive to defect shock amplitude, making them useful indicators to track defect severity. Two test cases are presented to verify the efficiency of the method. Bearings in an early stage of degradation with two levels of defect severity (50 µm and 100 µm) and defects in different locations are investigated. The results show that the method can effectively extract all information related to the defect. In this study, we validated the proposed method using rolling bearings. Future work will extend and generalize the method for fault diagnosis of other rotating machinery, such as gears.

Acknowledgments

The financial support of the Natural Sciences and Engineering Research Council of Canada (NSERC) is gratefully acknowledged. The assistance of the Bearing Data Center of Case Western Reserve University, which provided some experimental results, is also gratefully acknowledged.

References

[1] Huang, N. E., Shen, Z. and Long, S.R. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. Proceedings of the Royal Society of London, vol. 454, no. 1971 (1998): 903-995.

[2] Lei, Y. G., Lin, J., He, Z. J. and Zuo, M. J. *A review on empirical mode decomposition in fault diagnosis of rotating machinery*. Mechanical Systems and Signal Processing, 35 (1-2), 2012: 108-126.

[3] Peng, Z.K., Tse, P.W. and Chu, F.L. *A Comparison study of improved Hilbert–Hung transform and wavelet transform: application to fault diagnosis for rolling bearing*. Mechanical Systems & Signal Processing, vol. 19 no. 2, (2005): 974-988.

[4] Zhang, Y., Zuo, H. and Bai, F. *Classification of fault location and performance degradation of a roller bearing*. Measurement, vol. 46, no. 3 (2013): 1178-1189.

[5] Guo, W., Tse, P. W. and Djordjevich, A. *Faulty bearing signal recovery from large noise using a hybrid method based on spectral kurtosis and ensemble empirical mode decomposition*. Measurement. vol. 45, no. 5 (2012): 1308-1322.

[6] Yang Y., Yu, D.J. and Cheng, J.S. *A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM*, Measurement, vol. 9, no. 40 (2007): 943–950.

[7] Kedadouche, M., Thomas, M. and Tahan, A., *Monitoring machines by using a hybrid method combining MED, EMD, and TKEO*, Advances in Acoustics and Vibration, vol. 2014, 10 pages, 2014.

[8] Wenliao Du, Zhiyang Wang, Xiaoyun Gong, Liangwen Wang, and Guofu Luo, *Optimum IMFs Selection Based Envelope Analysis of Bearing Fault Diagnosis in Plunger Pump*, Shock and Vibration, vol. 2016, Article ID 1248626, 8 pages, 2016. doi:10.1155/2016/1248626

[9] Ricci R., and Pennacchi P. *Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions*. Mechanical Systems and Signal Processing 25, no. 3 (2011): 821-838

[10] Yi C, Lin J, Zhang W, Ding J. Faults Diagnostics of Railway Axle Bearings Based on IMF's Confidence Index Algorithm for Ensemble EMD. *Passaro VMN, ed. Sensors (Basel, Switzerland)*. 2015;15(5):10991-11011. doi:10.3390/s150510991.

[11] Jacek D., and Zimroz R. *Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal*. Applied Acoustics 77 (2014): 195-203.

[12] Z. K. Peng, P. W. Tse, and F. L. Chu, *An improved Hilbert–Huang transform and its application in vibration signal analysis*, Journal of Sound and Vibration, vol. 286, no. 1-2, pp. 187–205, 2005.

[13] C. Wang, M. Gan, and C. Zhu, *Non-negative EMD manifold for feature extraction in machinery fault diagnosis*, Measurement, vol. 70, pp. 188–202, 2015.

[14] W. Guo and P. W. Tse, *A novel signal compression method based on optimal ensemble empirical mode decomposition for bearing vibration signals*, Journal of Sound and Vibration, 332 (2), 423–441, 2013.

[15] Tsao, W.-C., Li, Y.-F., Le, D.D. and Pan, M.-C.. "An insight concept to select appropriate IMFs for envelope analysis of bearing fault diagnosis." *Measurement*, vol. 45, no. 6 (2012): 1489-1498.

[16] Pan, M.-C. and Tsao, W.-C. *Using appropriate IMFs for envelope analysis in multiple fault diagnosis of ball bearings*. International Journal of Mechanical Sciences, vol. 69, (2013): 114-124.

[17] Kedadouche, M., Thomas, M. and Tahan, A. *A comparative study between empirical wavelet transforms and empirical mode decomposition methods: application to bearing defect diagnosis*. Mechanical Systems and Signal Processing, vol. 81 (2016): 88-107.

[18] Kedadouche, M., Liu, Z. and Vu, V. H. *A new approach based on OMA-empirical wavelet transforms for bearing fault diagnosis*. Measurement, vol. 90, (2016): 292-308.

[19] Cong, F., Chen, J., Dong, G. and Zhao, F. *Short-time matrix series based singular value decomposition for rolling bearing fault diagnosis*. Mechanical Systems and Signal Processing, 34 (1), 2013: 218-230.
[20] Cong, F., Zhong, W., Tong, S., Tang, N. and Chen, J. *Research of singular value decomposition based on slip matrix for rolling bearing fault diagnosis*. Journal of Sound and Vibration, 344 (2015): 447-463.
[21] Kedadouche, M. and Liu, Z. *Fault feature extraction and classification based on WPT and SVD: Application to element bearings with artificially created faults under variable conditions*. Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science (2016): 0954406216663782.
[22] Brincker, R., Andersen, P., and Zhang, L. *Modal identification from ambient responses using frequency domain decomposition*. 18th International Modal Analysis Conference (IMAC), San Antonio, Texas, (2000): 625-630.
[23] Bendat, J.S. and Piersol, A.G. Random Data, *Analysis and Measurement Procedures*, 2nd ed., New York: Wiley-Interscience (1986): 120–121.
[24] http://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.
[25] Smith, W. A. and Randall, R. B. *Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study*". Mechanical Systems and Signal Processing, 64 (2015): 100-131.

Prognostics of Rolling Element Bearings based on Entropy indicators and Particle Filtering

Junyu QI[1, 2], Alexandre MAURICIO[1, 2], Konstantinos GRYLLIAS[1, 2]

[1]Department of Mechanical Engineering, KU Leuven,
[2]Dynamics of Mechanical and Mechatronic Systems, Flanders Make
Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium
Junyu.qi@kuleuven.be, alex.ricardomauricio@kuleuven.be, konstantinos.gryllias@kuleuven.be

Abstract

Rolling element bearings' damage is the main cause for unexpected breakdown in rotating machinery. Therefore there is a continuous industrial interest on condition monitoring of bearings, targeting towards the development and proposal of robust diagnostic techniques which can detect accurately, robustly and early the generation of the fault. On the other hand, industry is not interested only in the proper damage detection and identification of faults but is mainly targeting towards the robust estimation of the Remaining Useful Life (RUL) of machine elements. The proper estimation of the RUL could be linked directly with the maintenance planning and warehouse organization, providing immediately profits in terms of employees health and safety, environmental protection and continuous production. A plethora of diagnostics and prognostics indicators have been proposed during the last decade focusing towards the accurate representation and tracking of the health state of bearings and other machine elements. However, in certain cases (e.g. nonstationary operating conditions), the classic techniques for bearings prognostics (e.g. statistical analysis, frequency analysis and time-frequency analysis) underperform due to the high noise influence or the high machines' complexity. Therefore the classical diagnostics indicators may identify the fault quite late and fail to identify properly the RUL.

In this paper, prognostics indicators based on the measurement of disorder (e.g. entropy) are used in order to track the degradation process of the machinery. An advanced health indicator, termed Info-entropy, is built based on the Spectral Entropy, the Envelope Spectral Entropy and the Spectral Negentropy. It is compared with the Spectral Kurtosis, the Kurtosis and the RMS based on the criterion of trendability. The indicators are estimated on the well-known bearing dataset Prognostia and Particle filtering is used in order to estimate the RUL of bearings.

1 Introduction

In the era of Industry 4.0, Condition Monitoring (CM) is currently attracting unprecedented attention. Rotating machinery systems and components, such as Rolling Element Bearings (REBs), are utilized frequently in almost each mechanical equipment, including electric motors, wind turbines and compressors, etc. The breakdown of large amount of machineries happens due to the failure of bearings, which are mainly caused by misalignment, corrosion, insufficient lubrication and oil contamination, resulting further in accelerated wear in the surface. As the damage accumulates in the continuous operation, the functionality of REBs reduces through the stages of mild (stage 1), moderate (stage 2) and profound fault (stage 3), as presented in Figure 1.

Usually, stage 1 endures for long time, and as the fault reaches to a certain level, the degradation rate starts accelerating and as a result the stage 2 lasts for shorter time compared to stage 1. In stage 3, the functionality reduces to the End of Life (EoL) incredibly fast. Concerning the change of functionality, a number of maintenance concepts have been proposed in the last decades, i.e., the proactive maintenance (to stage 1), preventative maintenance (to stage 2) and reactive maintenance (to stage 3). The reactive

maintenance (fault detection and diagnosis) have been explored widely and a number of systematic, automatic and accurate methods are successfully established. However, the period of the reactive maintenance usually sustains shortly till the EoL, as stage 3 shown in Figure 1, and it requires expensive cost. On one hand, the hazardous performance of the bearings influences the normal operating condition of the equipment, which is related with the continuous production quality. On the other hand, it may lead to accidental breakdown of machinery. Preventative maintenance is localized in stage 2, while the defect is still in the impending or the incipient stage and the functionality can be corrected by proper intervention, e.g. cleanness, lubrication and alignment. The proactive maintenance in stage 1 costs less expensive, where the defect incubates slowly and the REB exhibits an acceptable performance. Therefore, if the life of in-service components is able to be predicted in the early time (stage 1), the maintenance strategy and actions can be optimally planned. Compared with traditional CM techniques, the predictive maintenance is recently catching particular attention. The state of the art of methodologies [1] includes mainly three directions, i.e. the Physics model based Prognostics (PbP), the Data-driven based Prognostics (DbP) and the Hybrid-approach based Prognostics (HbP).

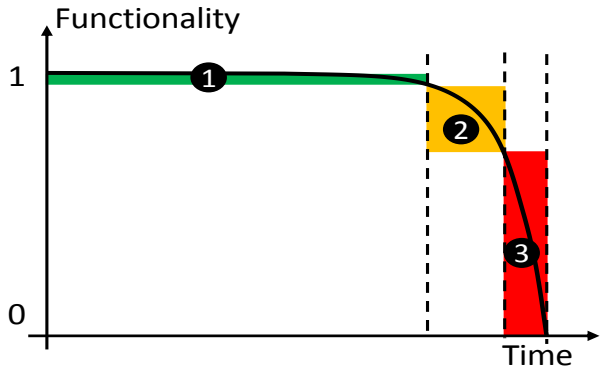


Figure 1: Degradation process and maintenance strategy

The PbP usually describes the degradation phenomenon in an explicit mathematical formulation. The establishment of a physical Finite Element Model (FEM) depends on the available sufficient information, e.g. the geometry, the load and the material, etc. In the past, some physics models have been developed, e.g. spall progression models, crack-growth model, gas path model, mixed lubrication model, etc. [2]. Theoretically, the PbP is built systematically and well adaptive to the degradation evolution process. However, the modeling in the system level requires complete domain knowledge and is expensive to be developed, therefore, the PbP is usually applied to a specific component.

With the advantage of flexibility, applicability and economy, the DbP is preferred by the industrial domain. Unlike the difficulty of building the physics model, an analytic model is directly derived from massive data. In general, the DbP is classified as Machine Learning (ML) and Statistical Model Based (SMB) techniques [3]. ML learns the hidden information from big data. Instead of an explicit model, ML works like a black box and till now often the output is hardly interpretable. In order to reduce the computation cost and enhance the knowledge understanding, the direction of SMB is considered in this paper, e.g. Kalman filter (KF), Particle Filter (PF), linear regression, Hidden Markov Model (HMM), etc. An analytic model, e.g. the exponential model [4], is learned from training data, and further is updated with testing data.

Medjaher et al. [5] presented a method of bearing prognostics based on the regression method. Yet, it is not an optimal solution for the stochastic degradation progress. Alternatively, the Bayesian Recursive Estimation (BRE) is able to represent the model, considering the uncertainty. For example, HMM is used for the RUL prediction, considering different failure modes [6]. However, the probability transition matrix between the model's states need to be presupposed. Qi et al. [7] employed the Unscented Kalman filter (UKF) for the bearing RUL prediction. Additionally, in order to cope with the limitation of the single model, the Switching Kalman Filter was investigated for the prognostics of a helicopter bearing. [8]. PF achieves advantageous performance than KF versions without the restrictions of Gaussian noise, system linearity or existence of low nonlinearities. Li et al. [9] predicted bearing RUL combining the PF with an improved exponential model. If Health Indicator (HI) exhibits a degradation pattern with multiple degradation rates,

the single statistical model is probably not adequate anymore. Therefore, a multi-model PF is introduced to track the linear and nonlinear stage, and the corresponding mode is recognized automatically by the calculation of the likelihood [10].

The result of SMB is usually dependent on several factors, e.g., the statistical model, the HI, the failure threshold, etc. To extract high quality HI, different signal processing techniques have been developed in the domain of time, frequency and time-frequency. However, many classic indicators are frequently deficient in the real applications. Ahmad et al. [11] proposed an improved HI removing the fluctuation based on the Linear Rectification Technique (LRT). In order to track the degradation in the best way, some researchers focused on the change in the specific frequency band. Singleton et al. [12] extracted a HI over the specific frequency band of the Choi-Williams Transform. Soualhi et al. [13] built a HI based on the Hilbert Huang Transform, decomposing the signal into several Intrinsic Mode Functions (IMF) in different frequency bands. However, an optimal HI with a remarkable tendency is usually challengeable to attain in reality, especially the tendency in the full lifetime. As a result, classic HIs, such as RMS and Kurtosis, may fail to track the trend in case of high noise, low impulsiveness and varying operating condition. The problem of low trendability triggers the proposal of start prediction point [9], [11] and the RUL prediction by the start prediction point is usually starting in the very late stage, e.g. in stage 3 of Figure 1. Moreover, it also influences the statistical model selection in PF.

In order to enhance the prediction performance, a new HI Info-entropy is proposed for the bearing prognostics in this paper. Following the logic of [12]-[13], Info-entropy assesses the chaos level for a specific frequency band and is composed of different entropy indicators in the frequency domain, i.e. the Spectral Entropy (SpecE), the Spectral Flatness (SpecF) and the Spectral Negentropy (SpecNegE). Based on the criterion of trendability, the best entropy indicators from a certain frequency band is selected. Then, the Info-entropy is collaborated with a classic exponential model and PF. The main contributions of the proposed methodology for SMB prognostics are summarized as follows.

1) Through the chaos measurement in multiple frequency bands, Info-entropy recognizes the hidden signature of the vibration signal of the REB. It provides a new perspective for the HI extraction compared to the traditional impulsiveness based features. Even in cases of ignoring the characteristic bearing frequencies, the low frequency resolution and the repetitive transients, the Info-entropy could detect the inherent change.

2) The in-service machinery degrades with time and its corresponding functionality reduces also, as can be seen in Figure 1. The Info-entropy is sensitive to frequency distributions, demonstrating the characteristic of high trendabillity. With the full lifetime degradation tracking, the aforementioned concept of start prediction point is no longer the practical bottleneck. Consequently, the use of Info-entropy allows for the start of prediction from the initial moment and potentially is useful for long-term RUL prediction.

3) Likewise, as Info-entropy present a high quality of trendability, the corresponding statistical model selection could be simplified. Compared with the complex model, e.g. for the non-monotonic trend, the actual calculation rested on the simple exponential model, becomes faster and efficient.

The rest of paper is organized as follows. The mathematical theory of PF and the construction of Info-entropy is explained in Section 2. The proposed methodology is detailed step by step in Section 3. In Section 4 and 5, the experimental dataset and the RUL prediction results are described. Finally, the paper is concluded in Section 6.

2 Mathematical theory

The theory of Particle filter and the proposed HI based on the Info-entropy are explained in this Section.

2.1 Particle filter

As a kind of BRE method, PF uses a number of particles to estimate the posterior distribution of a stochastic process. It is particularly useful for the estimation of a nonlinear dynamic system with non-Gaussian noise. The principles of PF is based on the Bayesian theory and sampling strategy. Bayesian theory includes steps of the prediction and update. In the prediction step, the priori Probability Density Function (PDF) of the k_{th} predicted state x_k is gained with the posteriori knowledge $p(x_k|z_{1:k-1})$. Then, the posteriori PDF $p(x_k|z_{1:k})$ is updated when the k_{th} measurement $z_k$is available. According to Monte Carlo Sampling (MCS), the state x_k is approximated by a set of particles $x_k^i, i = 1 \dots N$, which are sampled from a certain probability distribution. The posteriori probability of Bayesian estimation is able to be approximated

with MCS, $p(x_k|z_{1:k}) \approx \frac{1}{N}\sum_{i=1}^{N}\delta(x_k - x_k^i)$, where δ denotes the Dirac delta function. Each particle has the same weight $\frac{1}{N}$. However, particles degeneracy arises in the sequential importance sampling. The amount of effective particles decreases after some certain recursive steps. Hence, the resampling duplicates the high weight particles and replaces the lower ones. The importance distribution $q(x_k|x_{k-1}, z_k)$ is generally chosen as $p(x_k^i|x_{k-1}^i)$, then $w_k^i \propto w_{k-1}^i p(z_k|x_k^i)$.

PF mainly consists of three steps, the particles generation, the weight calculation and the resampling.

1) Particles generation

As the initial step, the amount (N) of particles are firstly generated from a priori distribution and then are transmitted through the propagation model, $x_k^i \sim p(x_k|x_{k-1}^i), i = 1 \ldots N$.

2) Weight calculation

The likelihood of each particle x_k^i is calculated with the measurement z_k, $w_k^i = p(z_k|x_k^i)$, and is normalized as $\breve{w}_k^i = \frac{w_k^i}{\sum_{i=1}^{N} w_k^i}$.

3) Resampling

Being the most frequently used resampling method, the Multinomial Resampling (MR) compares the weight of particles $\breve{w}_k^i$ with a random threshold between 0 and 1, and only the particles, which are above the threshold, are kept. It is worthy to mention, that the Systematic Resampling (SR) is utilized to replace the classic MR and to enhance the PF performance. Unlike the MR method, SR divides the whole interval into M subspaces $\mathrm{U^i} = ((\mathrm{i}-1)/\mathrm{M}, \mathrm{i}/\mathrm{M}), \mathrm{i} = 1, \ldots, \mathrm{M}$, and the particles are taken with a corresponding random threshold in the U^i, higher particle diversity is thus guaranteed.

2.2 Info-entropy extraction

Many classic HIs are extracted based on the impulsiveness of the signal, e.g., the statistical indicators, the Spectral Kurtosis, frequency indicators extracted by the envelope spectrum, after the application of Hilbert transform, etc. Nevertheless, the fault information is easily affected by the noise, non-stationarities and repetitive transients. Moreover, the behavior of rotating machinery performs physically nonlinear, due to the instantaneous variation of the friction, damping and loading conditions. Thus, the chaos assessment is taken into account. In this aspect, entropy is the versatile measurement of disorder, unevenness of distribution and complexity. The concept of entropy, different entropy indicators and the info-entropy will be explained in this Section.

2.2.1 Entropy

As property of thermodynamics, the entropy of an isolated system never decreases. It always evolves toward the thermodynamic equilibrium, which means each of the microstates S_i inside the system has an equal probability p_i.

2.2.2 Entropy indicators

A number of variants of entropy have been proposed in the time series, e.g. the Approximate Entropy, the Sample Entropy, the Kolmogorov Entropy and the Permutation Entropy. Yet, the assessment in the time domain is often insufficient. The frequency information is explored here, as the compensation to capture the latent variation. Thus, the Spectral Entropy, the Spectral Flatness and the Spectral NegEntropy, i.e., the NegEntropy in the Squared Envelope (NegE-SE) and the Squared Envelope Spectrum (NegE-SES) are introduced.

SpecE has been proposed to measure the amplitude distribution of the Fourier spectrum of a signal $x(t)$. It is defined as the same form of entropy:

$$SpecE = -\sum_{i=1}^{L} \mathrm{p_i} log_2\ (\mathrm{p_i}) \tag{1}$$

with,

$$p_i = \frac{X(i)}{\sum_{i=1}^{L} X(i)} \tag{2}$$

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \tag{3}$$

When the spectral amplitudes distribute evenly, SpecE yields a large value. On the contrary, it outputs a small value if the amplitudes are concentrated around a specific region. To avoid the influence of the data length, the SpecE is usually normalized as in Equation 4, and ranges between 0 and 1.

$$SpecE = \frac{-\sum_{i=1}^{L} \mathrm{p_i} log_2\,(\mathrm{p_i})}{log_2(L)} \tag{4}$$

SpecF, also called Wiener entropy, has been initially used for the whiteness characterization of speech analysis [14]. It is calculated by the ratio between the geometric mean and the arithmetic mean of the power spectrum, as presented in equation 5.

$$SpecF = \frac{\exp\left(\frac{1}{L}\sum_{i=1}^{L} \ln\left(|X(i)|^2\right)\right)}{\frac{1}{L}\sum_{i=1}^{L} \ln\left(|X(i)|^2\right)} \tag{5}$$

Equivalent to the SpecE, SpecF approaches to 1 if the spectrum uniformly scatters and to 0 if the spectrum is spiky.

As the opposite of entropy, NegEntropy is used as a measure of distance from normality. The Spectral NegEntropy in the frequency band $(f - 0.5 \cdot \Delta f, f + 0.5 \cdot \Delta f)$ is defined in [15] as:

$$\Delta \mathrm{I_\varepsilon(f,\Delta f)} = \left\langle \frac{\varepsilon_\mathrm{x}(\mathrm{t,f,\Delta f})^2}{\langle \varepsilon_\mathrm{x}(\mathrm{t,f,\Delta f})^2 \rangle} \ln\left(\frac{\varepsilon_\mathrm{x}(\mathrm{t,f,\Delta f})^2}{\langle \varepsilon_\mathrm{x}(\mathrm{t,f,\Delta f})^2 \rangle}\right) \right\rangle \tag{6}$$

where, ε_x is the SE of signal $x(t)$, $\langle * \rangle$ indicates the average operator and ΔI_ε represents the NegEntropy in the SE (NegE-SE). Similarly, NegE-SES is written as:

$$\Delta \mathrm{I_E(f,\Delta f)} = \left\langle \frac{\mathrm{E_x(t,f,\Delta f)^2}}{\langle \mathrm{E_x(t,f,\Delta f)^2} \rangle} \ln\left(\frac{\mathrm{E_x(t,f,\Delta f)^2}}{\langle \mathrm{E_x(t,f,\Delta f)^2} \rangle}\right) \right\rangle \tag{7}$$

where, ΔI_E is the NegEntropy in the SES, and E_x denotes SES of $x(t)$. In addition, to overcome the shortcomings of NegE-SE and NegE-SES, the weighted average Spectral NegEntropy (NegE-Ave) has been considered:

$$\Delta \mathrm{I_{1/2}(f,\Delta f)} = (\Delta \mathrm{I_\varepsilon(f,\Delta f)} + \Delta \mathrm{I_E(f,\Delta f)})/2 \tag{8}$$

SpecNegE has reverse behavior with Spectral Entropy. If vibration signal shows more impulsiveness or repetitive transient, it returns a higher value.

2.2.3 Info-entropy

In order to capture the informative event of machinery, spectral NegEntropy has been adopted in Infogram [15]. The signal is firstly decomposed into different frequency bands using a 1/3-binary tree. Then, the demodulation frequency band is localized by the maximal spectral NegEntropy, which is caused from the faulty component.

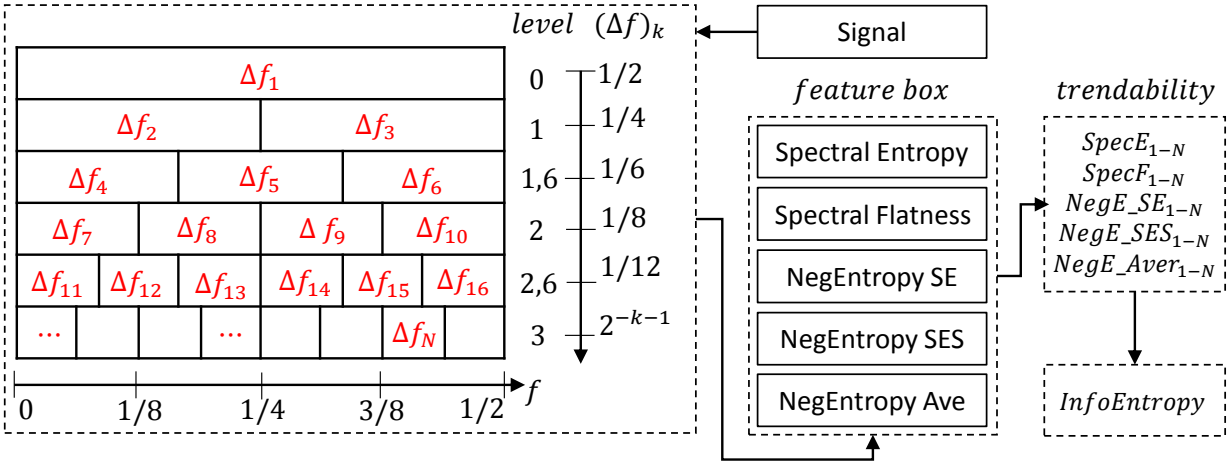


Figure 2: The illustration of Info-entropy proposition

The proposed Info-entropy is structured as presented in Figure 2. At first, the decomposed signals in different frequency bands Δf_1, Δf_2, ⋯, Δf_N are calculated by five kinds of Spectral Entropy and NegEntropy. Then, each entropy indicator has an amount of N values and is collected in a feature box, i.e., $SpecE_{1-N}$,

$SpecF_{1-N}$, $NegE\_SE_{1-N}$, $NegE\_SES_{1-N}$ and $NegE\_Ave_{1-N}$. In order to catch the degradation evolution for the specific operating condition, the anticipant feature is selected based on the trendability, as shown in Equation 9. In this way, the most sensitive and informative frequency band is localized. Finally, the selected Info-entropy can be any of the feature box with an optimal score, based on Equation 10, depicting well the whole degradation process.

$$R = \frac{n(\sum xy)-(\sum x)(\sum y)}{\sqrt{(n\sum x^2-(\sum x)^2)(n\sum y^2-(\sum y)^2)}} \quad (9)$$

where x and y represent, respectively, the time index and the feature. n indicates the length of the feature. The trendability measures the correlation between the time and the feature. The value R varies between -1 and 1. The final score of the HI is expressed in the form of equation (10). The perfect score is equal to 0 and signifies high trendability for the full lifetime.

$$Score = 1 - abs(R) \quad (10)$$

3 Proposed methodology

The proposed methodology is detailed in Figure 3. Firstly, the data is recorded in a certain operating condition. However, the vibration signals are easily polluted during the experiment, by e.g. the deterministic composition of gear transmission, anomalies, noise etc. and different preprocessing techniques are widely derived to enhance the bearing signal. In order to solve the problem of outliers, thresholding using the Standard Deviation (SD) is often used. But SD is problematic to the extreme values, e.g.to infinite valueσ. Compared to the criterion of SD, the application of three Median Absolute Deviations (MAD) is more robust to detect and remove anomalies. Therefore, the Moving MAD (MMAD) is proposed as α pre-processing technique. The MAD is calculated w.r.t the D sequence points. The samples beyond the threshold Thr_1 (e.g. $Thr_1 = 3 \cdot MAD$) are recognized as outliers and αρε replaced with Thr_1. Moreover, α prewhitening technique is suggested before calculating the Spectral NegEntropy [15].

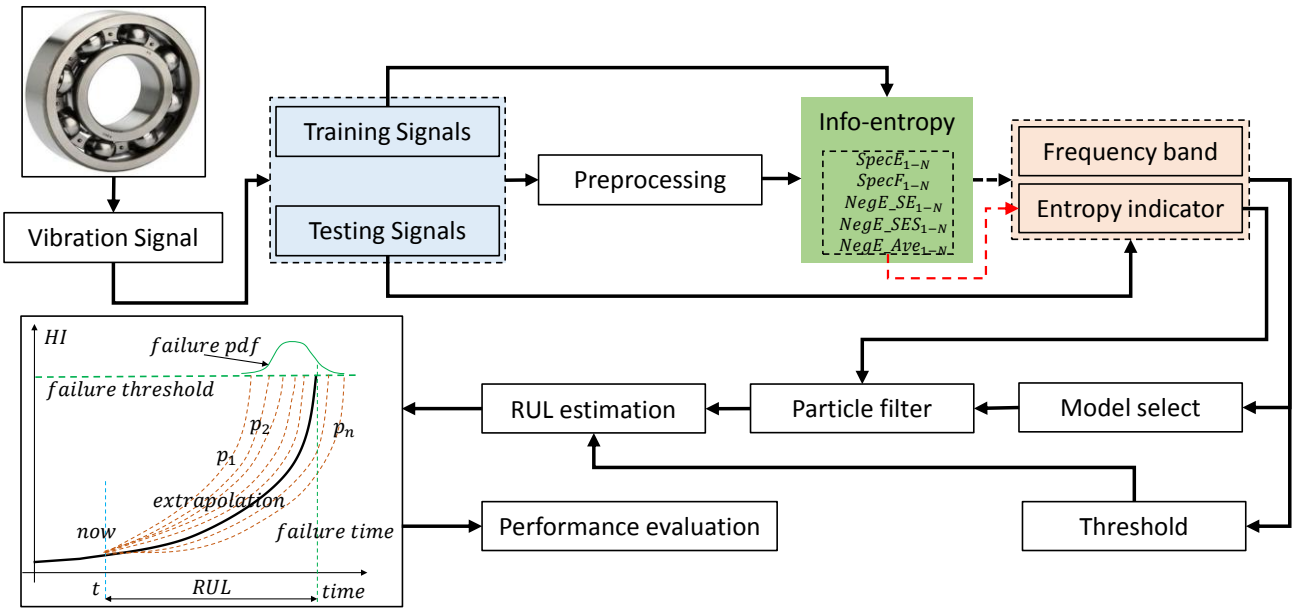


Figure 3: The flowchart of proposed methodology

Then, following the explanation of Info-entropy in Figure 2, the training dataset is utilized to localize the most informative frequency band and the corresponding entropy indicator in the feature box with the best score is selected. Based on the extracted HI, the statistical model is easily determined to represent the degradation process of the bearing. If the model parameter b_i indicates the degradation path of the i_{th} bearing B_i, the similarity level between HIs is measured by the ratio between the slowest degradation b_{min} w.r.t the closest path b_k.

$$\text{ratio} = \frac{b_{min}}{b_k} \quad (9)$$

In case, the Info-entropy of all the members of training sets $U(B_i)$ present high similarity (e.g. $ratio > \epsilon$, ϵ is the threshold of similarity level), the failure threshold thr is defined as the average value of the maximum or minimum of the His and the initial model parameters pr_0 are calculated by the average of the fitted values from $U(B_i)$. Otherwise, when the HIs show low similarity in the degradation path (e.g. $ratio < \epsilon$), the thr and pr_0 are set based on the closest HIs. They depend on the two bearings B_{n1} and B_{n2} of $U(B_i)$ if the starting point HI_0 of testing data is localized in the middle of their HI_0. Yet, when HI_0 falls outside the ranges of HI_0 of $U(B_i)$, thr and pr_0 are chosen based on the HI with the closest HI_0.

Finally, after the selection of the statistical model, the initial parameters and the failure threshold, PF starts to update the model parameters with the availability of new measurements. The RUL is calculated recursively by the extrapolation to the thr. As a set of particles is used, the RUL is then expressed as the probability of distribution. The RUL result is evaluated with three classic metrics, i.e., the Cumulative Relative Accuracy (CRA), the Convergence (C_m) and the Error (Err) [16].

The Relative Accuracy (RA) evaluates the error of the RUL prediction in relation to the actual RUL at the specific time. In order to measure the error at multiple time instances, the CRA aggregates the RA of a given time span.

$$\mathrm{CRA}_\lambda = \frac{1}{|l_\lambda|}\sum_{l_\lambda} \mathrm{w_k RA_k} \tag{11}$$

where, l_λ is the set of all time indexes, w_k and RA_k are, respectively, the weight factor and the RA at the time step k. The CRA approaches to 1 in the case of the ideal RUL prediction.

By the measurement of the Euclidean distance, convergence expresses how fast the prediction goes close to the actual RUL.

$$\mathrm{C_m} = \sqrt{(\mathrm{x_c} - \mathrm{t_p})^2 + \mathrm{y_c}^2} \tag{12}$$

where x_c and y_c are the centroid coordinates of the area under the prediction error, such as the accuracy and t_p is the first inspection time. A smaller distance shows that the predicted RUL converges faster to the actual RUL.

In addition, the percentage error is also frequently used, and is defined as follows:

$$\mathrm{Err} = \frac{\mathrm{RUL_{re} - RUL_{pr}}}{\mathrm{RUL_{re}}} \times 100\% \tag{13}$$

where, RUL_{re} and RUL_{pr} are the actual and the predicted RUL, respectively. The percentage is equal to 0 when the predicted RUL coincides with the actual RUL.

4 Experimental degradation data

The proposed methodology is evaluated on the experimental dataset PRONOSTIA [17]. The setup consists of a motor, a custom gearbox, two support bearings and the test bearing, which bears load from a pneumatic actuator. The experiment is implemented under three constant operating conditions (radial load and rotating speed). The data is measured with a sampling frequency of 25.6 kHz for a duration of 0.1 seconds every 10 seconds. Acceleration signals are recorded simultaneously in the horizontal and vertical direction. The EoL in the experiment is defined based on the vibration amplitude. The measurement stops when the vibration at one of the two measurement directions exceeds 20g in a sequence of samples. In the end, seven measurements are implemented for the first two conditions and three measurements for the condition 3.

5 Results and discussion

Based on the experimental dataset and the established procedure of the proposed methodology, the results of the RUL prediction are discussed. In this section, the experimental dataset is firstly preprocessed before calculating the Info-entropy. In the view of trendability, the extracted HI is compared with some

existing papers. Then, the statistical model, the initial parameters and the failure threshold are decided. Finally, the RUL prediction results are presented.

5.1 Data preprocessing

Following the explanation given in section 3, the preprocessing is important to remove the environmental influences. After visual inspection, the time wave of the vibration signals contain outliers and unsymmetric points. The window length D in MMAD is empirically chosen equal to 50. After cleanness, the samples against the degradation behavior are efficiently separated from the normality. In addition, a prewhitening technique based on the linear prediction is utilized before the Spectral NegEntropy. The order is set equal to 50.

5.2 Info-entropy extraction

Based on the description of Section 4, seven bearings have been measured in the first operating condition. In order to demonstrate the priority of the proposed HI in comparison with other state of the art papers [9], [11], only the same 4 bearings (bearing 1-1, bearing 1-3, bearing 1-4 and bearing 1-7) in the first operating conditions are predicted in this paper.

After conditioning of the vibration signals, the Info-entropy is firstly calculated on the six training sets. The attained informative frequency band and the entropy indicator are then employed into the rest testing data. In the process of Info-entropy calculation, five entropy indicators, i.e. $SpecE_{1-N}$, $SpecF_{1-N}$, $NegE\_SE_{1-N}$, $NegE\_SES_{1-N}$ and $NegE\_Ave_{1-N}$ are simultaneously achieved. In this paper, only the example of NegE-SE of the bearing 1-1 is presented in Figure 4. The NegE-SE in each frequency band performs an increasing tendency and it coincides with the degradation evolution and the accumulation of damage. The Info-entropy is able to find the highest trendability for all training datasets in certain frequency band, which represents the characteristic information of the same operating condition.

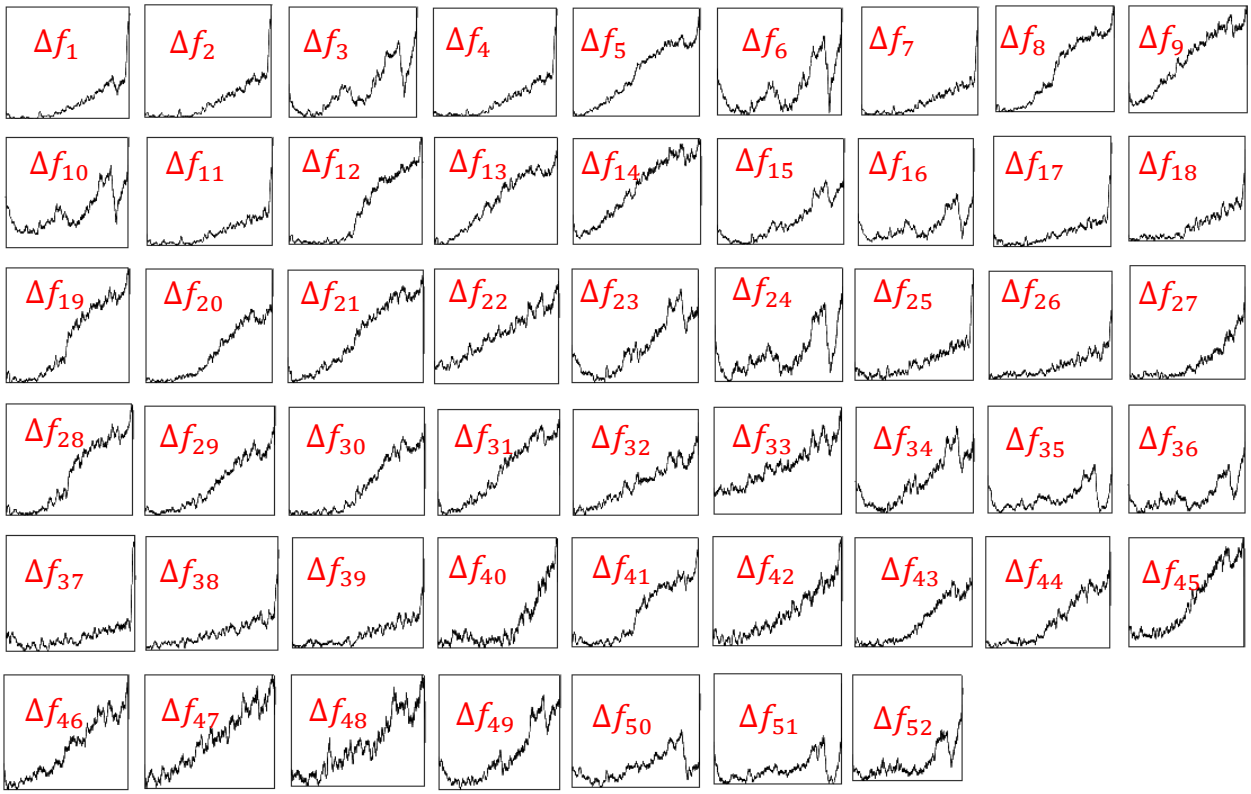


Figure 4: Spectral NegE-SE of decomposed exemplary signal (bearing 1-1)

	Bearing 1-1	Bearing 1-3	Bearing 1-4	Bearing 1-7
RMS	0.3387	0.4676	0.2707	0.6127
Kurtosis	0.3576	0.1863	0.3633	0.2786

SK	0.1840	0.1692	0.5237	0.5131
Info-entropy	0.0215	0.0467	0.1778	0.0428

Table 1: Trendability comparison of different His

The extracted feature in the frequency bands $\Delta f_1 - \Delta f_{52}$ in Figure 4 is generated by a 1/3 binary-tree, where the decomposition level L_v is empirically set equal to 4. Theoretically, the larger level L_v is able to capture more information with more decomposed signals. However, it requires also more computation cost. Based on the score of Equation 15, NegE-SE is selected as the Info-entropy among five different entropy indicators over the frequency band Δf_5, i.e., 4267 Hz-8533 Hz. The NegE-SE of four to be tested bearings are shown in Figure 5d. To highlight the superiority of the proposed HI, two statistical features Kurtosis and RMS for the same bearings in [9] and the Spectral Kurtosis (SK) are introduced as a comparison, and their trends are shown together in the Figure 5.

Analysing Figure 5a, RMS indicates the stable trend in the early stage. Especially for bearing 1-4 and 1-7, it suddenly reaches to failure after the long term 'healthy' running. Compared with RMS, Kurtosis in Figure 5b is sensitive to the impulsiveness and the slope of degradation is more distinct, e.g. in bearing 1-7. However, its trend drops in the late stage, where the fault size may accumulate and the vibration signal becomes less impulsive than the early stage fault. SK in Figure 5c has similar results as Kurtosis, but with slight higher slop. It is obvious that Info-entropy tracks well the full lifetime degradation for the four bearings. The slope of each bearing is high and cluster together. In addition, their corresponding scores based on the Equation 10 are listed in Table 1. The values of Kurtosis of bearings 1-3 and 1-7 approach closer to 0 than RMS. However, the SK of the bearing 1-1 and 1-3 triumphs over Kurtosis. It should be mentioned that the Info-entropy reveals the best score compared to the other indicators for each bearing. Therefore, with the advantage of trendability, the Info-entropy will be applied in the next section.

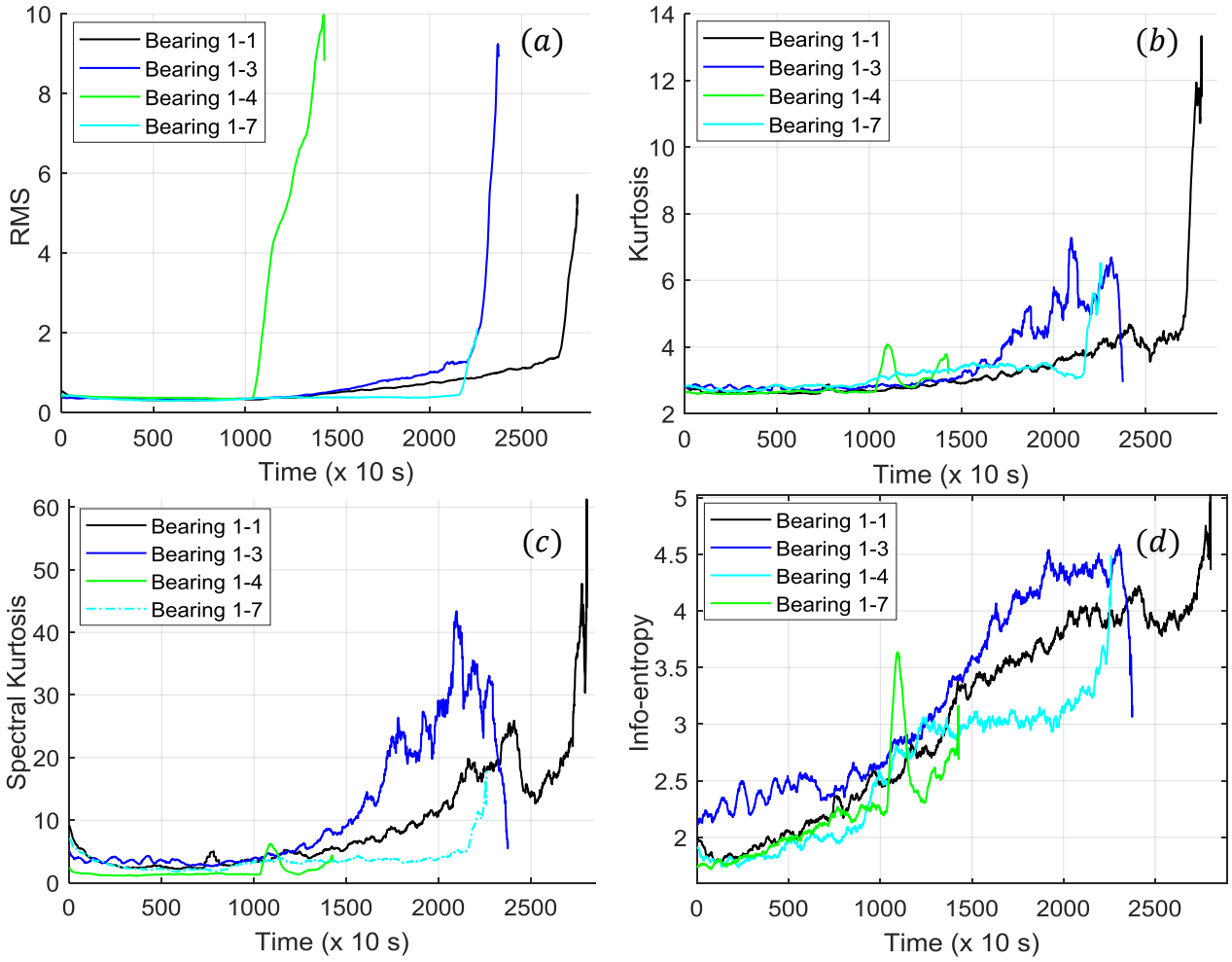


Figure 5: Comparison of different HIs

5.3 RUL results

5.3.1 Parameters definition

From the extracted Info-entropy, the degradation trend of each bearing generally follows an exponential path. Hereby, a classic exponential model is thus chosen for PF:

$$f(t) = ae^{b \cdot t} \tag{14}$$

where, a and b are the model parameters.

The failure threshold and the initial parameters are realized based on the strategy in Section 3. The threshold of similarity level ϵ is set equal to 0.6. The ratio is lower than 0.6 for the bearings in condition 1. In addition, as known, Spectral Entropy and Spectral Flatness range between 0 and 1. To facilitate the implementation, Spectral NegEntropy is normalized by the maximal value of HIs from the training dataset.

5.3.2 Results

After the selection of the HI, the statistical model and the parameters definition, the RUL of the aforementioned four bearings (length $Len = 1400, 1802, 1139, 1502$, duration $d = Len \times 10\ s$) is predicted based on the proposed methodology. In order to evaluate the effectiveness of different algorithms, the PF based methods, e.g., the Classic Particle Filter (CPF) with MR strategy and the Enhanced Particle Filter (EPF) with SR, are compared with other SMB methods, i.e. the CKF, the EKF, the UKF and the simple regression (REG). Finally, their performance are evaluated using the performance metrics, e.g. the CRA, the C_m and the Err.

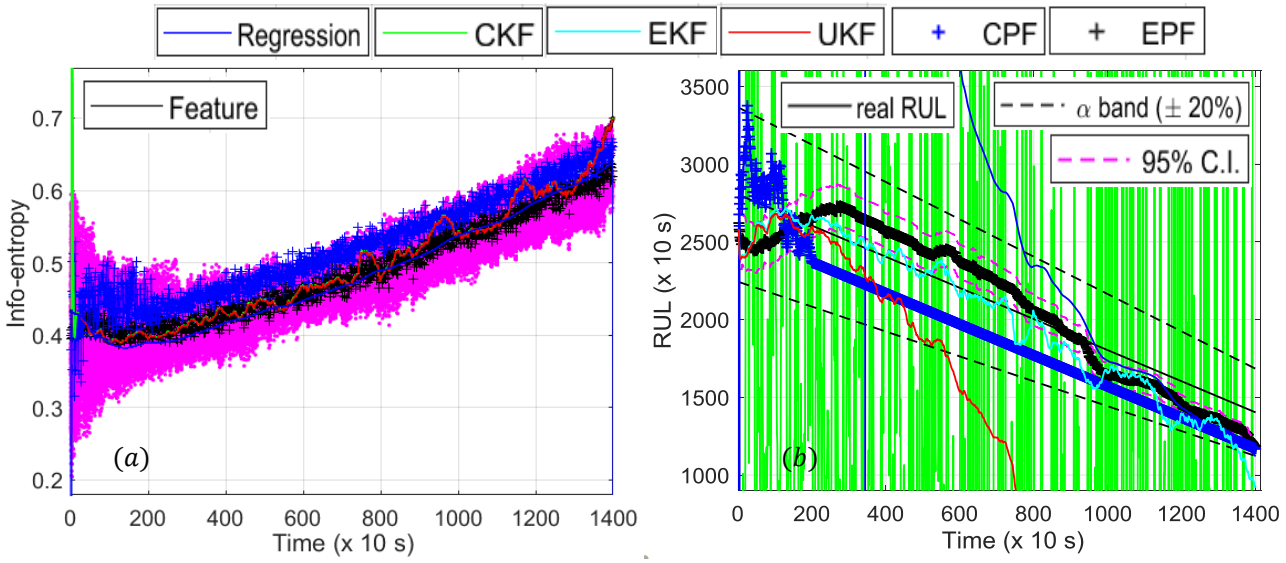


Figure 6: Feature estimation (links) and RUL estimation (right) by different methods for bearing 1-1

The results of bearing 1-1 are presented in Figure 6. In the Figure 6a, the feature is estimated perfectly using the KF versions (i.e. CKF, EKF and UKF). The particle filter generates a number of particles to approximate the posterior distribution of the state in each step. The mean value of the CPF and the EPF is plotted in the Plus (+) sign. From the Figure 6a, the PF tracks the general trend of the feature till the end, and the EPF approximates closer to the feature than CPF. By updating the model parameters in each step, the RUL result is able to be calculated by the extrapolation to the predefined threshold. The corresponding RUL estimation is presented in Figure 6b. The CKF estimates the RUL with severe fluctuation without convergence, while EKF and UKF predict in a similar manner before 200 (x 10 s). Afterwards, UKF diverges from the actual RUL, probably because of the random generated sigma points. However, the EKF follows almost the actual RUL before 900 (x 10 s) and then deviates to the outside of the (+/-) 20% error band. As comparison, the EPF performs the comparable prediction as EKF before 900 (x 10 s) and much closer to the actual RUL than the simple regression. Between 900 (x 10 s) and the end, the EPF goes slightly far from the actual RUL and has a better performance than EKF. It should be noticed that the full lifetime prediction by EPF falls within the error band. Moreover, the uncertainty of the RUL prediction is evaluated with the 95% confidence band, starting in wide uncertainty and decreasing with the time passing by. In

contrast, the performance of CPF seems reliable with the stable distance to the actual RUL line, yet, it has the severe problem of particle degeneracy after 200 (x 10s).

The result of the second example bearing 1-4 is presented in Figure 7. Similarly with the results in Figure 6, the feature estimation of KFs demonstrates better than PFs in Figure 7a, especially for the part after 1000 (x 10 s). In Figure 7b, the performance of different algorithms is shown. The worst results are achieved by the CKF. UKF demonstrates good results before 400 (x 10 s) but then diverges to the outside of the error band. The RUL prediction by EKF seems more reliable than REG, CKF and UKF, but with large biased error before 1000 (x 10 s). Then, it converges fast to the actual RUL and diverges again in the very late stage. Analogously to the Figure 6b, CPF predicts in the form of a straight line after 180 (x 10 s). Although the observed prediction error is small, yet, it has a serious problem of particle degeneracy. As comparison, the EPF performs better results than EKF and REG before 1000 (x 10s), and converges with a biased error w.r.t the actual RUL. Looking back to the Info-entropy in Figure 5d, the failure threshold is set based on the bearing 1-7, which is much higher than the real life of bearing 1-4. This is the reason that the prediction of EPF and EKF has a biased error in this case. Therefore, it can be concluded that obviously the failure threshold set has a critical influence to the final RUL estimation.

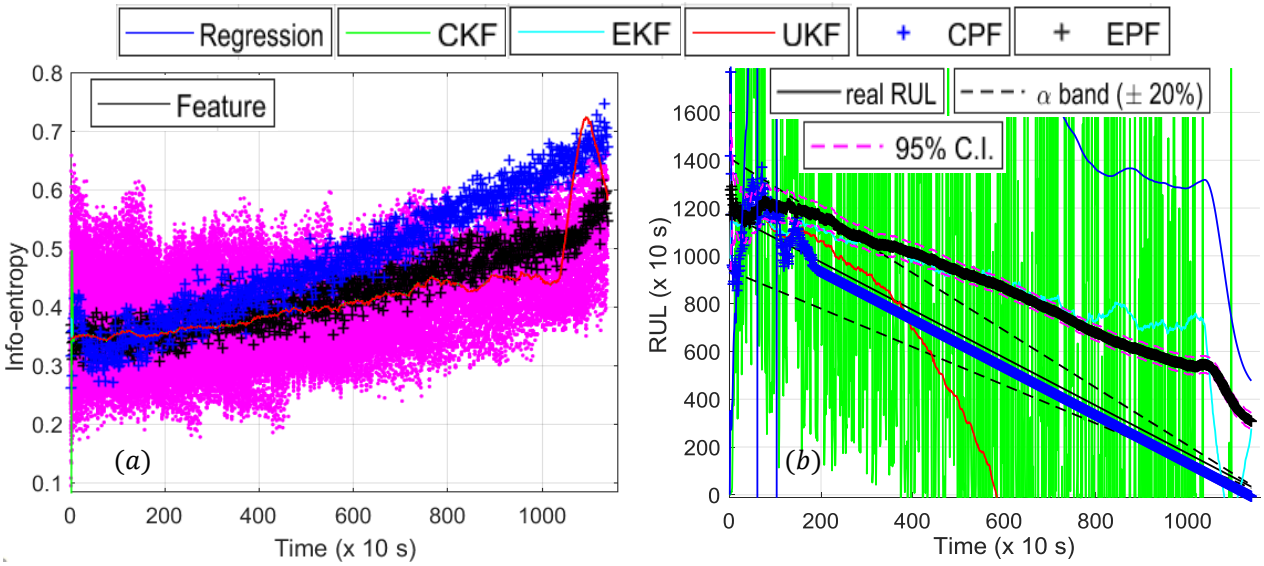


Figure 7: Feature estimation (links) and RUL estimation (right) by different methods for bearing 1-4

In addition to the visual depiction, the performance of these four bearings are presented in Table 2. For all bearings, the CRA of the EPF is the best ,closest to the 1, while the EKF performs better than the rest. With the huge fluctuation, CKF gains a negative CRA value for all bearings here. In the view of the Err, EPF performs the best score almost in total, with the exception of bearing 1-4. REG shows advantage than CKF, EKF and UKF in some bearings as it is calculated however based on the all history data in each step. Moreover, the convergence of EPF is not always best at CRA and Err and it may be surpassed by the CPF and CKF. However, the CPF is easily caught with the problem of degeneracy and the CKF usually has the larger prediction error.

Unlike the RUL prediction using the concept of start prediction point in [9], [11]-[13], with its help the good RUL results are possible to be achieved at the very late stage, i.e. the stage 3 of Figure 1. Using the proposed HI, the RUL in this paper is predicted from the beginning. The prediction of some bearings performs even better than [9] and [11], for example, the CRA of bearing 1-1 has better score than an exponential model in [9] and the bearing 1-3 and 1-7 has smaller error than the LRT technique in [11]. Therefore the proposed Info-entropy is able to track the degradation very early and can achieve good results in the long terms. The prognostics in the proposed methodology becomes more meaningful than state of the art papers, because the maintenance operators have plenty of time to schedule the repair plan, further avoiding a hazardous machinery breakdown.

6 Conclusion

In this paper, a statistical model based prognostics methodology for the RUL estimation of rolling element bearings is proposed. The methods of HI construction, statistical model selection and RUL

estimation are explained in detail. Traditionally, a HI with a high trendability is expected, being capable to reflect the full lifetime degradation. However, the degradation status is commonly not easily tracked, especially in case of less impulsiveness and low frequency resolution. Many classic signal processing techniques may fail to extract high quality HI, which is critical to the selection of the failure threshold, of the model and even of the concept of start prediction point. In order to overcome these problems, entropy, as a chaos measurement, is considered. An HI Info-entropy is constructed based on the entropy indicators in the frequency domain, i.e., the Spectral Entropy, the Spectral Flatness and the Spectral NegEntropy. After the signal decomposition, the most informative frequency band is localized and the entropy indicator that follows the optimal degradation trend is selected.

Through the validation on the experimental dataset, Info-entropy is effective to capture the bearing degradation trend in the full lifetime. It provides a new perspective to construct the high trendability HI in a complex operation and further simplify the prognostics process, e.g. the model selection, the elimination of the concept of the start prediction point. In the end, the Info-entropy is combined with a classic exponential model, and is incorporated with six different methods, i.e. mathematical regression, classic Kalman filter, Extended Kalman filter, Unscented Kalman filter and Particle filter. Through the analysis, systematic resampling based Particle filter is efficient to better enhance the particle diversity than the classic resampling method and it achieves more accurate prediction results than others.

		Bearing 1-1	Bearing 1-3	Bearing 1-4	Bearing 1-7
REG	CRA	-0,07	0,02	-3,69	-5,41
	Err	20,20	-10,19	-1308,02	5,84
	C_m	912,92	855,53	356,43	1110,79
CKF	CRA	-1,59	-7,26	-8,12	-1,77
	Err	617,46	727,73	411,52	649,89
	C_m	167,54	357,92	3125,91	176,38
EKF	CRA	0,91	0,30	-0,30	0,35
	Err	34,38	-36,49	-717,90	-52,45
	C_m	681,80	892,43	410,18	804,69
UKF	CRA	-0,08	-3,46	-10,70	-0,86
	Err	496,72	546,77	283,19	546,74
	C_m	224,40	767,07	7740,37	391,44
CPF	CRA	0,87	0,64	0,81	0,34
	Err	16,64	64,57	119,43	101,15
	C_m	687,63	848,54	541,94	671,61
EPF	CRA	0,93	0,73	-0,50	0,44
	Err	15,52	-0,34	-808,25	-4,37
	C_m	696,84	949,01	407,38	792,10
PF [9]	CRA	0,87	0,76	0,87	0,93
	Err	/	/	/	/
	C_m	8,99	4,28	275,40	290,20
LRT [11]	CRA	0,94	/	/	/
	Err	/	7,00	0	19,00
	C_m	/	/	/	/

Table 2: Comparison of RUL estimation by different methods

References

[1] K. Javed, R. Gouriveau and N. Zerhouni, “State of the art and taxonomy of prognostics approaches, trends of prognostics applications and open issues towards maturity at different technology readiness levels,” *Mech. Syst. Signal Process.*, vol. 94, pp. 214–236, 2017.

[2] A. Cubillo, S. Perinpanayagam and M. Esperon-Miguez, “A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery,” *ADV. MECH. ENG.*, vol. 8, no. 8, pp. 1-21, 2016.

[3] B. Sun, S. Zeng, R. Kang and M. G. Pecht, “Benefits and Challenges of System Prognostics,” *IEEE Trans. Rel.*, vol. 61, no. 2, pp. 323-335, Jun. 2012.
[4] N. Gebraeel, M. Lawley, R. Li and J. Ryan, “Residual Life Distribution from Component Degradation Signals: A Bayesian Approach,” *IIE Trans.*, vol. 37, no. 6, pp. 543-557, 2005.
[5] K. Medjaher, N. Zerhouni, and J. Baklouti, “Data-driven prognostics based on health indicator construction: Application to PRONOSTIA’s data,” in *Proc. Eur. Control Conf.*, 2013, pp. 1451–1456.
[6] D. Tobon-Mejia, K. Medjaher, N. Zerhouni, G. Tripot, “Hidden Markov models for failure diagnostic and prognostic.,” in *Proc. Prognostics Syst. Health Manage. Conf.*, Shenzhen, China, 2011 May, pp. 1–8.
[7] J. Qi, A. Mauricio, M. Sarrazin, K. Janssens and K. Gryllias, “Remaining Useful Life Prediction of Rolling Element Bearings Based on Unscented Kalman Filter,” in *Proc. Int. Conf. on Condition Monitoring of Machinery in Non-stationary operations*, Santander, Spain, 2018 June 20-22, pp. 123–138.
[8] C. K. R. Lim and D. Mba, “Switching Kalman filter for failure prognostic,” *Mech. Syst. Signal Process.*, vol. 52-53, pp. 426-435, Feb. 2015.
[9] N. Li, Y. Lei, J. Lin, and S. X. Ding, “An improved exponential model for predicting remaining useful life of rolling element bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762–7773, Dec. 2015.
[10] P. Wang, R. Qiang and R. X. Gao, “Multi-Mode Particle Filter for Bearing Remaining Life Prediction.,” in *Proc. Int. Manufacturing Science and Engineering Conf.*, Texas, USA, 2018 June.
[11] W. Ahmad, S. Khan and J. Kim, “A Hybrid Prognostics Technique for Rolling Element Bearings Using Adaptive Predictive Models.,” *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, Feb. 2018.
[12] R. K. Singleton, E. G. Strangas, and S. Aviyente, “Extended Kalman filtering for remaining-useful-life estimation of bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.
[13] A. Soualhi, K. Medjaher and N. Zerhouni, “Bearing Health monitoring based on Hilbert-Huang Transform, Support Vector Machine and Regression.,” *IEEE Trans. Instrum. Meas.*, vol. 64, no. 1, pp. 52–62, Jan. 2015.
[14] A. Gray and J. Markel, “A Spectral-Flatness Measure for Studying the Autocorrelation Method of Linear Prediction of Speech Analysis.,” *IEEE Trans Acoust.*, vol. 22, no. 3, pp. 207-217, Jun. 1974.
[15] J. Antoni, “The infogram: Entropic evidence of the signature of repetitive transients.,” *Mech. Syst. Signal Process.*, vol. 74, Nov. 2018.
[16] A. Saxena, J. Celaya, B. Saha, S. Saha and K. Goebel, “On Applying the Prognostic Performance Metrics.,” in *Proc. Prognostics Health Manage. Conf.*, San Diego, USA, 2009 September 27- October 1, pp. 1-16.
[17] P. Nectoux, R. Couriveau, K. Medjaher and E. Ramasso, “PRONOSTIA: An experimental platform for bearings accelerated degradation tests.,” in *Proc. IEEE Int. Conf. on Prognostics Health Manage.*, Denver, USA, 2012 June, pp. 1-8.

Spall Evolution in a Rolling Element Bearing

Dmitri Gazizulin[1], Roni Schneck[2], Renata Klein[3], Jacob Bortman[1]

[1]PHM Lab, Department of Mechanical Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 8410501, Israel
dima.gzizo@gmail.com
jacbort@bgu.ac.il

[2]Department of Materials Engineering, Ben-Gurion University of the Negev, P.O. Box 653, Beer Sheva 8410501, Israel
roni.shneck@gmail.com

[3] R.K. Diagnostics, Gilon, P.O.B. 101, D.N. Misgav 20103, Israel
renata.klein@rkdiagnostics.co.il

Abstract

Rolling element bearing (REB) is one of the basic mechanical components in rotating machinery. It is common to divide the REB wear into two stages, damage initiation and damage propagation. There has been a growing awareness of the need to understand the damage mechanism during the propagation phase. The current work includes a discussion on the ongoing research and the methodology for the development of the prognostic method for damage propagation. The methodology integrates experiments, diagnostic methods, and physics-based models. Endurance tests were conducted in order to learn about the damage propagation process and model validation. Furthermore, Finite Element model of the spalled bearing was developed and validated. The FE model aims to investigate and simulate the damage propagation process. The simulation results are in good agreement with the experimental observation.

1 Introduction

Failure prognosis of rolling element bearings (REBs) is crucial in rotating machinery PHM. The damage evolution in REBs consists of two main phases: damage initiation and propagation. The conventional REB life models address the lifetime of the bearing to the damage initiation, i.e. first spall formation [1]. However, after the first spall formation, the bearing might be fully operational for millions of cycles. After the first spall formation, it propagates in the circumferential direction of the raceway, until the bearing becomes non-operational [2].

Many diagnostic tools have been developed in order to monitor the spall propagation. However, there is lack of reliable prognostic of the remaining useful life (RUL). Physic-based prediction of the damage propagation in the REB, after the first spall generation. The difficulties in prognosis of the propagation stage necessitate deep understanding of the damage mechanisms, the stochastic nature of the spall propagation process, and its modeling [3, 4]. The main goal of the current work is to develop a physics-based prognostic method for the spall propagation in REBs. Herein, we present a methodology for the development of the prognostic method, which combines experiments, diagnostic methods, and physics-based models.

2 Methodology

The proposed methodology for prognostic of the RUL is based on a combination of physics-based models, diagnostic methods and experiments [7]. The concept is presented in Figure 1. This section includes a description of the objectives that need to be reached in order to achieve the main goal, RUL prognostics.

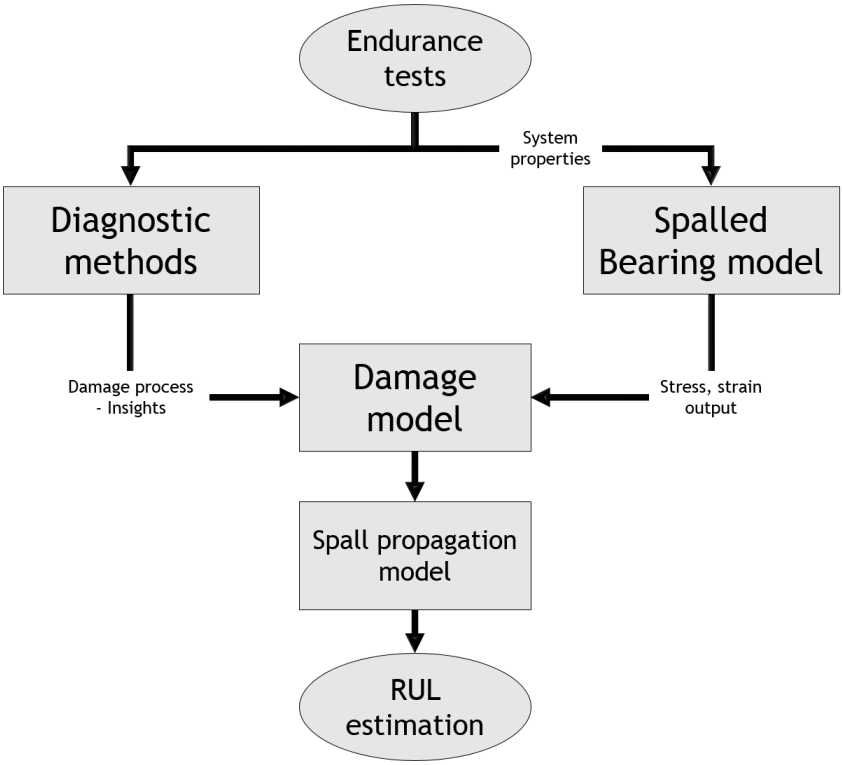


Figure 1: Research flow chart describing the steps toward the development of the prognostic method.

The first objective is the quantitative and qualitative understanding of the damage-driven mechanisms, e.g. plastic strains, residual stresses, etc., of the spall propagation process. In addition, it is important to learn, based on the existing literature and experiments, about the effects of the bearing's features (e.g. hardness, ball mass), and operational conditions (e.g. speed, load) on the propagation process and the trend of the spall growth [8, 9]. This objective can be attained by conducting endurance tests. Figure 2 shows an example of a diagnostic indicator evolution, BPFO Z-score distance vs. time. The tests can add insight regarding the spall propagation process and can be used for the physical model validation.

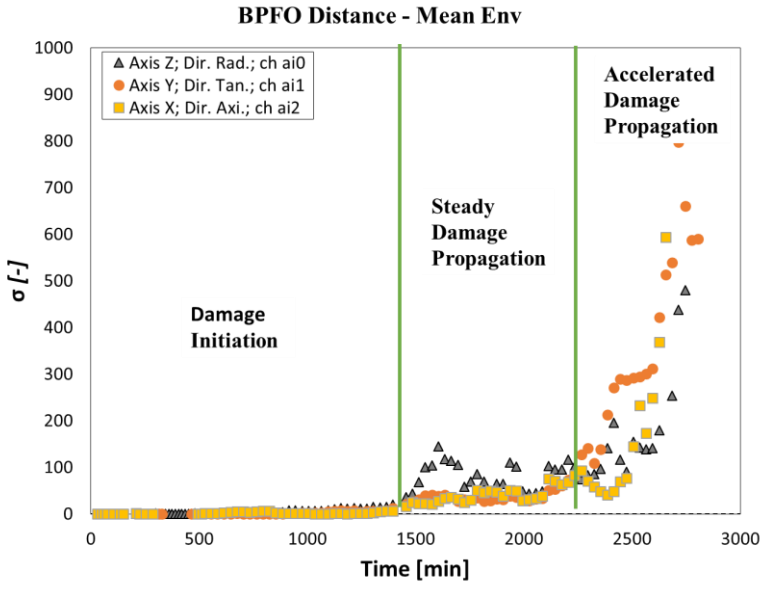


Figure 2: Damage evolution trajectory during the endurance test. Each data point correspond to the vibration measurement during the test.

The second objective is the development of a model for the damage propagation process. First, the material response in the presence of a defect must be analyzed. This analysis, coupled with the endurance test results, will shed light on the mechanism governing the damage propagation process. For this purpose a physics-based model of a spalled bearing has been developed. The model is used for analysis of the material response in the presence of a defect. Moreover, the model was used for damage evolution simulation. Figure 3 shows examples of a simulation result and the metallurgical analysis of a bearing from an endurance test. The accomplishment of this objective will provide a mean estimated damage trajectory. However, the damage propagation is a stochastic process. Hence, dispersed results are expected.

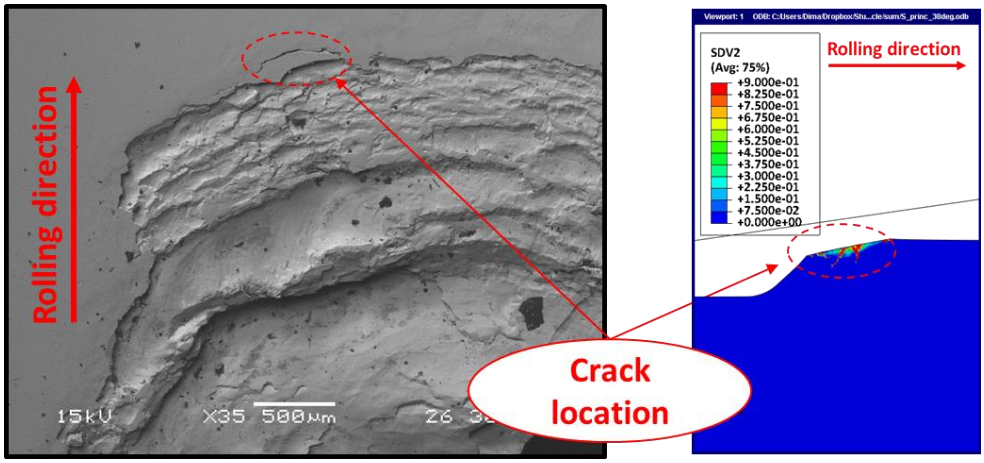
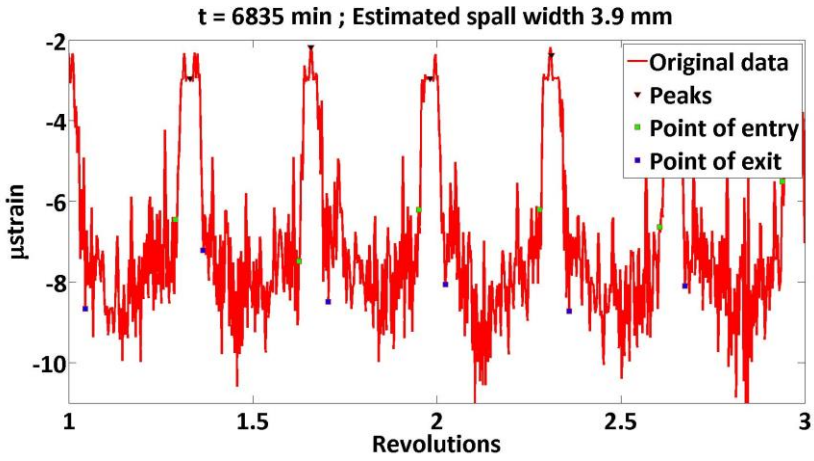
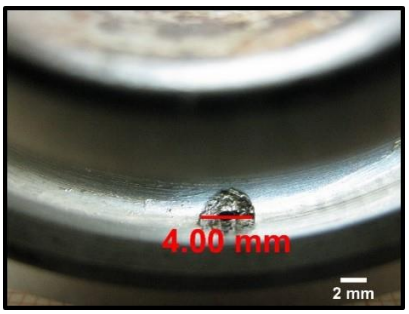


Figure 3: SEM Image of the spall trailing edge and simulated cracks.

The third objective of the research, and probably the most challenging, is to model the stochastic nature of the damage propagation. The propagation of the spall in the REB is a stochastic process. Even under well controlled experimental conditions, using allegedly identical bearings, the results of the endurance tests vary [2]. The prognostic method must consider the uncertainties and the progress of the probability distribution. One of the common methods is to use diagnostic condition indicators in the early stages of the damage in order to monitor its propagation, e.g. oil debris, vibration level, etc. For this purpose, methods for the spall width estimation via time domain analysis were developed [10]. The spall width is estimated based on the acceleration and strain signatures, Figure 4. The spall width and vibration based CIs will be used for the estimation of the damage model parameters by the trend identification of the spall propagation process. The accomplishment of this goal will complete the development of the physics-based prognostic method.



(a)



(b)

Figure 4: Strain data vs. cycle, based on which the spall size was estimated. The (a) estimated size is 3.9 *mm*, and the (d) measured size is 4.0 *mm*.

3 Summary and Conclusions

The remaining useful life (RUL) estimation using physics-based prognostic method is schematically illustrated in Figure 5. The calibration of the damage model can be implemented by comparing the simulations results with the data extracted from endurance tests. For example, a diagnostic method for the defect severity estimation (fault size, vibration level, etc.) can be implemented during the first stages of the tests. The results obtained by the diagnostic method can be used for the estimation of the damage model parameters and their uncertainties. The integration of the prognostic and the diagnostic methods, has the potential for reliable online estimation of the RUL including probability distribution of the result.

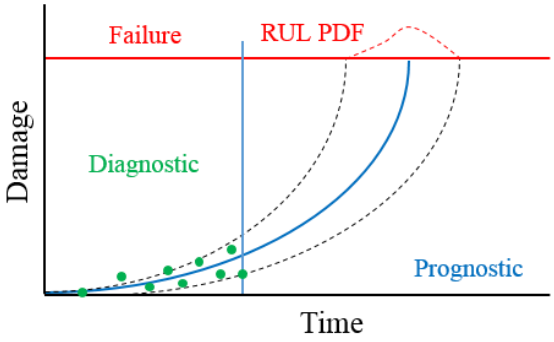


Figure 5: Bearing prognosis - first the damage propagation process is monitored; next, the model parameters are estimated; and the RUL is calculated [7].

Acknowledgments

The authors would like to express their deepest appreciations to Pearlstone Center for their support and funds of this work.

References

[1] F. Sadeghi, B. Jalalahmadi, T. S. Slack, N. Raje, and N. K. Arakere, "A review of rolling contact fatigue," *Journal of tribology*, vol. 131, no. 4, p. 041403, 2009.

[2] L. Rosado, N. H. Forster, K. L. Thompson, and J. W. Cooke, "Rolling contact fatigue life and spall propagation of aisi m50, m50nil, and aisi 52100, part i: experimental results," *Tribology Transactions*, vol. 53, no. 1, pp. 29–41, 2009.

[3] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, and S. Liang, "Adaptive prognostics for rolling element bearing condition," *Mechanical systems and signal processing*, vol. 13, no. 1, pp. 103–113, 1999.

[4] J. Qiu, B. B. Seth, S. Y. Liang, and C. Zhang, "Damage mechanics approach for bearing lifetime prognostics," *Mechanical systems and signal processing*, vol. 16, no. 5, pp. 817–829, 2002.

[5] M. Hoeprich, "Rolling element bearing fatigue damage propagation," *ASME, Transactions, Journal of Tribology*, vol. 114, no. 2, pp. 328–333, 1992.

[6] M. L. French and W. M. Hannon, "Angular contact ball bearing experimental spall propagation observations," *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, vol. 229, no. 8, pp. 902–916, 2015.

[7] D. Gazizulin, R. Klein, and J. Bortman, "Physics based methodology for the estimation of bearings' remaining useful life: Physics-based models, diagnostic methods and experiments," in *Proceedings of the European Conference of the PHM Society*. PHM Society, 2018.

[8] N. K. Arakere, N. Branch, G. Levesque, V. Svendsen, and N. H. Forster, "Rolling contact fatigue life and spall propagation of aisi m50, m50nil, and aisi 52100, part ii: stress modeling," *Tribology Transactions*, vol. 53, no. 1, pp. 42–51, 2009.

[9] N. A. Branch, N. K. Arakere, V. Svendsen, and N. H. Forster, "Stress field evolution in a ball bearing raceway fatigue spall," *Journal of ASTM international*, vol. 7, no. 2, pp. 1–18, 2009.

[10] G. Kogan, J. Bortman, and R. Klein, "A new model for spall-rolling-element interaction," *Nonlinear Dynamics*, vol. 87, no. 1, pp. 219–236, 2017.

[11] J.-L. Chaboche, "Continuum damage mechanics: Part i - general concepts," *Journal of Applied Mechanics*, vol. 55, no. 1, pp. 59–64, 1988. [Online]. Available: http://dx.doi.org/10.1115/1.3173661

[12] S. Marble and B. P. Morton, "Predicting the remaining life of propulsion system bearings," in *Aerospace Conference, 2006 IEEE*, 2006, pp. 8 pp.–.

Data Mining Classification & Machine Learning methods

Multi-label fault diagnosis based on Convolutional Neural Network and Cyclic Spectral Coherence

Zhuyun Chen[1], Alexandre Mauricio[2,3], Weihua Li[1], Konstantinos Gryllias[2,3]

[1]School of Mechanical and Automotive Engineering, South China University of Technology, 510640, GuangZhou, China
(mezychen@gmail.com; whlee@scut.edu.cn)
[2]Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300, Leuven, 3001, Belgium
[3]Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Leuven, Belgium
(alex.ricardomauricio@kuleuven.be; konstantinos.gryllias@kuleuven.be)

Abstract

Rotating machines are widely used in manufacturing industry, where sudden failures of key components such as bearings may lead to unexpected breakdown of machines and cause economic loss and human casualties. In addition, machines usually are operating under different working conditions leading to the dynamic changes of fault characteristic, thus presenting big challenges of reliable and accurate fault diagnosis. Data-driven based Deep Learning (DL) fault diagnosis methods are powerful tools to capture hierarchical features from raw input to classify fault patterns by stacking multiple non-linear transformation layers. It constructs and trains deep models relying on huge historical data and requiring less expert knowledge to obtain decision-making. These techniques present effectiveness and advantages in many intelligent fault diagnosis tasks. However, many DL methods are developed for the diagnosis of single fault type without considering the correlations of fault modes. In this paper, we develop a novel fault diagnosis method based on cyclostationary tool and Convolutional Neural Networks (CNN) to tackle these problems. The proposed method presents three characteristics: 1) Cyclic Spectral Coherence (CSCoh) is adopted to provide bearing discriminative patterns for specific type of faults. 2) A fault occurred on the same component (fault pattern), but having different fault severity levels can be regarded a multi-label classification problem, where the fault pattern and the fault severity level are considered to be two specific faults. 3) A novel CNN is constructed by introducing a sigmoid activation output and binary-cross entropy loss function to conduct the multi-label classification task. Specifically, CSCoh is adopted to capture correlation features of periodic phenomenon in the frequency domain. CSCoh is a bi-variable map of two frequency values, which could be used to enhance signatures masked by strong noise, characterizing the fault vibration signals obtained from the rotating machinery under different operating conditions. Then a CNN is developed for multi-label fault classification, which includes fault patterns and fault severity levels identification. The proposed method is evaluated in the experimental study of rolling element bearing fault diagnosis, in which data are collected under different working conditions. The experiment results demonstrate that the proposed method could presents good classification performance and superiority compared with other approaches.

1 Introduction

Rotating machines play important role in manufacturing industry. Rolling element bearings, as the key components of machines easily suffer from the sudden failures due to the long operating under the harsh conditions. This unexpected breakdown of machines may lead to economic loss and even human casualties. Thus, it is essential to develop the condition monitoring techniques for the early and accurate defect detection of such components.

Recently, data driven based-DL intelligent fault diagnosis methods have achieved increasing attentions, due to the powerful feature leaning capability from raw input. DL algorithms refer to deep neural networks,

where multiply non-linear transformation layers are stacked to construct the hierarchical architectures. Each layer can be regarded as a data pre-processing unit, where the input is converted into abstract features. With the increase of the layers, the high-level layer can learn more discriminative representations which are helpful for the diagnosis tasks [1]-[2]. The typical DL algorithms, such as Auto Encoder (SAE), Deep Belief Network (DBN), Convolutional Neural Network (CNN) and Long Short-Term Memory network (LSTM), have been applied for the fault diagnosis and detection [3]-[5].

Wang [6] took advantages of the CNN to learn features automatically from the raw vibration signals. Then, Hidden Markov Models (HMM) were employed as strong stability tools to classify rolling element bearing faults. Chen [7] proposed a SAE-DBN method for fault diagnosis by utilizing the multiply sensor information. In the first step, eighteen statistic indexes were extracted from the raw vibration signals, which were fed into sparse autoencoders for feature fusion. Finally, the fused features were input into DBN for fault diagnosis. Janssens [8] extracted the frequency spectra from two vibration accelerometers, and then a 2D CNN with one convolutional layer was designed to learn useful features for bearing fault detection. The proposed method demonstrated its advantages compared to that with hand-crafted features. Sun [9] presented a sparse Deep Stacking Network (DSN) to improve motor diagnosis performance, where the output label of DSN was coded as binary 0 and 1, which leads to more accurate and robust classification results. Chopra [10] adopted SAE for unsupervised features extraction from the engine data, and the majority voting based criteria was used to determine the engine fault type. Althobiani [11] utilized both the Teager Kaiser energy operator and statistical measures to reveal the fault patterns contained in collected signals, and then further adopted DBN for diagnosis of reciprocating compressor. The proposed method provided highly reliable and applicable. Tamilselvan [12] applied DBN for aircraft engine health diagnosis and electric power transformer health diagnosis, which obtained high classification accuracy and presented good generalization performance. In addition, Ince [13] developed a 1D CNN to conduct end-to-end motor fault diagnosis from raw signal input. Jia [14] proposed a normalized CNN for improving the bearing diagnosis performance under imbalanced data by embedding normalized layers and weighted Softmax loss.

From those works mentioned above, different measurements, such as raw time-series signals, frequency spectra, time domain and frequency domain statistical indexes, were adopted as the input of the DL algorithms, which obtained high diagnosis results. However, most of the studies are focusing on distinguishing different fault patterns while ignoring the diagnosis of fault severity levels. The fault severity identification is meaningful to detect the early fault occurrences and guide the decision-making.

Gan [15] adopted the Wavelet Packet Transform (WPT) to extract representative features and then designed a two-layer Hierarchical Diagnosis Network (HDN) for rolling bearing faults. In this study, different DBNs are stacked together, respectively, for classifying fault patterns and fault severity levels. However, multiply DBNs should be designed and trained for meeting the classification tasks. In addition, Wen [16] proposed a Hierarchical Convolutional Neural Network (HCNN), which can be used to recognize the fault patterns and fault severity levels at the same time. In this work, two fully-connected branches are designed at the end of pooling layer to conduct diagnosis. The first branch is used for the diagnosis of fault patterns, and the second branch is used for the recognition of fault severity levels. However, the drawback is that each branch should be constructed with an independently loss function, and a parameter was introduced to balance the trade-off of two losses of HCNN, which requires much expertise knowledge and computational cost for training.

Inspired by previous works, a novel approach integrating CNN and CSCoh is proposed for the multi-label fault classification of rolling element bearings. Firstly, CSCoh is adopted to capture correlation features of periodic phenomenon in the frequency domain, which provides a good discriminative input for CNN. Then a novel CNN is constructed for implementing multi-label fault classification by introducing a novel activation function and loss function. Compared to the other methods, the proposed method only requires a CNN to obtain the fault patterns and fault severity levels at the same time. In addition, compared to other methods, the proposed only replaces the activation function and loss function, without introducing extra parameters and computational cost, which is more suitable for the real industrial applications.

The remaining part of the paper is organized as follows. In Section 2, the theory of the cyclic spectral analysis and the CNN are provided. The diagnosis procedure using the proposed method is introduced in

Section 3. In Section 4, a comprehensive experimental description and the analytical are introduced. Section 5 describes the conclusion of this paper.

2 Introduction to cyclic spectral analysis and Convolutional Neural Networks

2.1 cyclic spectral analysis

In rotating machines, the bearing defects usually generate modulated signals by the characteristic frequencies of the bearings. Such signal, though not periodic, usually can be described as cyclostationary, whose statistical properties vary periodically with time [17]-[18]. The common spectral analysis technique is. Fast Fourier Transform (FFT). It is based on the assumption that the analysed signals are stationary, which can not accurately describe the cyclostationary. To deal with the problems, the cyclic spectral analysis techniques are developed to detecting and identify the hidden periodic behaviour of signals [19]-[20].

For a cyclostationary signal $x(t)$, the second-order moment of cyclostationarity can be defined as an instantaneous AutoCorrelation Function (ACF) with a cyclic T, which is defined as:

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) = E\{x(t+\tau/2)x(t-\tau/2)^*\} \tag{1}$$

Then, the second-order statistical descriptor of cyclostationarity, called the Cyclic Spectral Correlation (CSC), can be estimated by implementing the double Fourier transform on the ACF, which is given by:

$$\mathrm{CSC}(\alpha, f) = \int\int R(t,\tau) e^{-j2\pi(\alpha t + f\tau)} dt d\tau \tag{2}$$

The CSC is a bi-variable map of two frequency values. The parameters f and α are called the spectral frequency and cycle frequency, respectively. Contrary to the classic spectral analysis, it provides an additional frequency dimension, revealing both the carriers and their modulations. Spectral frequency f is linked to the carrier component, and the cyclic frequency α is linked to its modulation. It can be observed that for α is equal to zero, it is the classical power spectrum. Furthermore, for α is not equal to zero, it is the power spectrum for that specific cyclic component. Then the Cyclic Spectral Coherence (CSCoh) can be used to measure the degree of correlation between two spectral components given estimated by:

$$\mathrm{CSCoh}(\alpha, f) = \frac{\mathrm{CSC}(\alpha, f)}{\mathrm{CSC}(0, f)\,\mathrm{CSC}(0, f-\alpha)} \tag{3}$$

The CSCoh can be interpreted as the CSC of a whitened signal, which tends to equalize regions with very different energy levels, magnifying weak cyclostationary signals [20].

2.2 Convolutional Neural Network

CNN as a category of multi-layer neural network has achieved great success in areas such as image recognition, image classification, object detections, recognition faces [21]. A typical CNN usually is constructed by the four main operations, convolutional layer, activation layer, pooling or sub sampling layer, and fully-connected layers. Different kinds of layers play different roles. By stacking multiply convolutional, pooling and fully-connected layers, CNN can learn from low-level features to high-order or more abstract features. The layer types considered in this work are introduced.

2.2.1 Convolutional layer

Convolution is the first layer of CNN. The primary purpose of convolution is to extract the features by implementing the convolution operation on raw input data with learned convolution kernel/weights [22]. For each input $\mathbf{x}_i$ and convolution kernel $\mathbf{k}_j$, the output feature map can be calculated as follows,

$$\begin{aligned} \mathbf{y}_{i,j} &= f(\mathbf{b}_j + \textstyle\sum_i \mathbf{k}_j * \mathbf{x}_i) \\ f(\mathbf{x}) &= \max(0,\mathbf{x}),\ \mathbf{x}>0 \end{aligned} \tag{4}$$

where, * denotes the convolution operation, $\mathbf{k}$ and $\mathbf{b}$ are the value of the kernel and the bias. $f(\,)$ is the activation function, which is usually selected as the Rectified Linear Unit (ReLU) to enable better training of CNN.

2.2.2 Pooling layer

In the second step, the pooling layer is followed, which is used to reduce the spatial dimension and gain computation performance and some translation invariance. This is achieved by summarizing the feature responses in a region of neurons in the previous layer. For an input feature map $\mathbf{x}_i$, the output feature map is obtained,

$$\mathbf{y}_i = \max_{r\times r}(\mathbf{x}_i) \tag{5}$$

where r is the pooling size, and the common pooling operation adopted is known as max-pooling, which slides a window, and gets the maximum on the window as the output.

2.2.3 Fully-connected layer

In the fully-connected layer, the neurons are fully connection to all activations in the previous layer, and a Softmax classifier is usually attached to compute the class score. For the input vector $\mathbf{z}_i$ (i=1, 2, …, N), where N is the number of samples. the Softmax computes the exponential of the given input vector, and the sum of exponential values of all the values in the inputs. Then the ratio of the exponential of the input value and the sum of exponential values is the output of the Softmax function, which can be defined as,

$$Softmax(\mathbf{z}_i) = \frac{\exp(\mathbf{z}_i)}{\sum_j \exp(\mathbf{z}_i)} \tag{6}$$

The output corresponds to the probabilities of each class, and the target class will have the high probability. Softmax will enforce that the total sum of all the probabilities equals to one. That means, in order to increase the probability of a particular class, the module will correspondingly decrease the probability of at least one of the other classes. Thus, the final output will only have one true label. In order to effectively update the neural network, the Cross-Entropy (CE) loss can be adopted by minimizing the loss function between the probability output and the true target class, which is defined as,

$$Loss_{CE} = \sum_{i=1}^{C} \mathbf{y}_i(\log(\hat{\mathbf{y}}_i)) \tag{7}$$

where $\mathbf{y}$ is the true label of the data set, $\hat{\mathbf{y}}$ is the Softmax output, and C is the number of class.

3 The proposed CSCoh-CNN fault diagnosis framework

3.1 The architecture of the proposed CNN

In this section, the architecture of the proposed CNN is designed. Compared to the traditional CNN architecture, the proposed architecture introduces a new activation function in the output layer and a new loss function of CNN.

3.1.1 Sigmoid activation function

In the traditional CNN, the Softmax is usually regarded as the final fully-connected layer to predict the classes. While in the proposed CNN, it is replaced with Sigmoid activation function, which can be defined as,

$$Sigmoid(\mathbf{z}_i) = \frac{1}{1+\exp(-\mathbf{z}_i)} \tag{8}$$

For each value of the Sigmoid input, the Sigmoid function returns an independently real-valued output, which can be used to estimate the true output. For the Softmax output, the high value will have the higher probability than other values. That means for a classification problem, there is only one right class output, the outputs are mutually exclusive. While for the Sigmoid output, since the output are independently, it allows to have high probability for all of the classes, and the high value will have the high probability but not the higher probability. That means, for a multi-label classification, Sigmoid can output multiply correct classes, once a probability of one of the output nodes is above the threshold which is usually set to 0.5.

In order to better explain the differences of Softmax and Sigmoid, a fault diagnosis case is taken for example, presented in figure 1. When a Ball Fault with defect diameters of 14 mil (BF14) of rolling element bearing occurs, it can be observed that the traditional Softmax can correctly predict the BF fault with a probability of 85%. But it can only provide a true class output, which fails to diagnosis the severity level at the

same time. On the contrary, for the Sigmoid activation output, it not only can obtain the fault pattern: BF with a probability of 92%, but also the fault severity level which is estimated with a probability value of 84%.

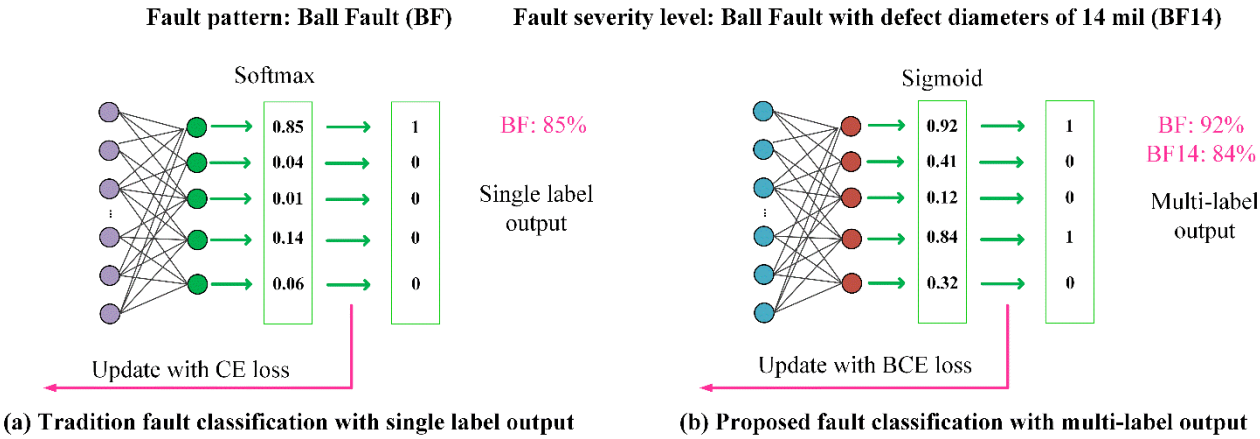


Figure 1: The differences between Softmax and Sigmoid activation function

3.1.2 Binary Cross-Entropy Loss

In order to optimize the CNN with multi-label classification tasks, the Binary Cross-Entropy (BCE) loss is adopted by splitting a multi-label classification problem in *C* binary classification problems. Unlike the CE loss, BCE is independently for each class, meaning that the loss computed for every output component is not affected by other output class. The loss function can be defined as,

$$Loss_{BCE} = \sum_{i=1}^{C} \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1-\mathbf{y}_i)\log(1-\hat{\mathbf{y}}_i) \tag{9}$$

where the update of the weight can be easily implemented by Back-Propagation (BP) algorithm, which is the same as that of the traditional CNN.

3.2 Fault diagnosis based on the proposed CSCoh-CNN

In this section, a multi-label fault diagnosis framework combining CSCoh and CNN is constructed as presented in figure 2. Inspired by the typical LeNet-5 [22], the proposed CNN architecture is designed by stacking two convolutional layers, two pooling layers, one fully-connected layer, and one Sigmoid classification layer.

In the convolutional layer configurations, a small kernel size (3 x 3) is applied in each convolution layer to capture the detail information and reduce the number of parameters. The convolution stride is fixed to 1 pixel. The number of filters is set equal to 6 in the first convolutional layer and the second one is doubled (12) to increase the feature learning capability. In the pooling layers, the max pooling is carried on the input over a (2 x 2) pixel window with stride 2. Therefore, the feature map size is halved to reduce the time complexity. Then the fully-connected architectures is set to 100 neurons. The number of Sigmoid output nodes correspond to the number of predicted classes.

The specifically training procedure can be summarized as follows,

Step 1: The raw vibration data are collected from the test rig, which are pre-processed by cyclic spectral analysis to obtain 2D CSCoh maps. The maps are further downsized to size 112×112 by balancing the computational cost and accuracy.

Step 2: The CNN is constructed by stacking multiply convolutional and pooling layers. Especially, Sigmoid activation function is adopted to predict the independently probability of each class. Accordingly, the BCE is adopted for measuring the distribution between the multi-label output and the target output.

Step 3: CNN network is updated by minimizing the BCE loss to improve the performance of the model in each epoch. The training procedure is the same as that of the traditional CNN.

Step 4: At the testing phase, the testing samples are fed into the trained CNN model to obtain the final diagnosis result.

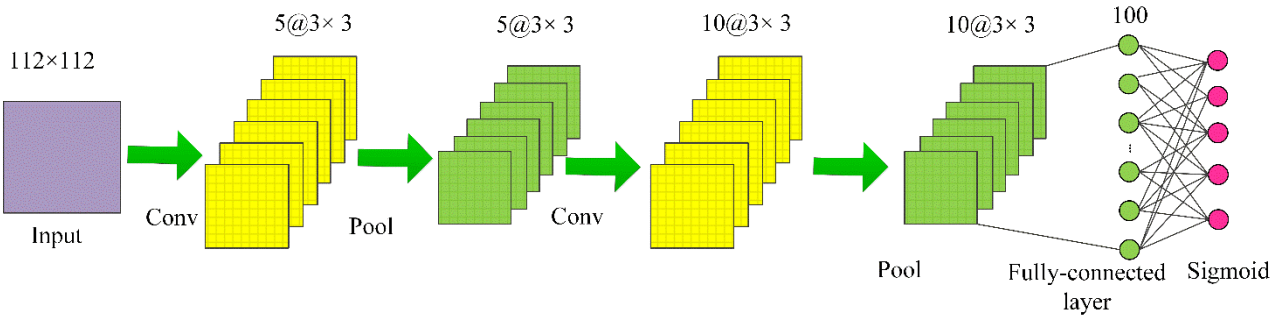


Figure 2: The fault diagnosis framework of the proposed method

4 Experiment verification

4.1 Experiment Setup

The experimental data of rolling element bearings have been acquired from the public bearing data center provided by Case Western Reserve University (CWRU), which is regarded as a benchmark dataset. The test rig is mainly composed of an induction motor, a transducer and a dynamometer. The vibration data are collected near the driving end of motor with a sampling frequency of 48 kHz. The motor bearings were seeded with faults using electro-discharge machining (EDM). In addition to the Normal Condition (NC), bearings with defect diameters of 7 mil, 14 mil and 21 mil have been introduced separately at the inner raceway, ball and out raceway. Each bearing is tested under four different loads (0, 1, 2 and 3 hp). Thus, the faults with two different diagnosis levels ('Level 1' and 'Level 2') can be detected. 'Level 1' means to discriminate the bearing with different fault patterns. While 'Level 2' denotes to further diagnosis the bearing with different severity levels, such as BF7 and BF17 cases, which is more challenge. The detailed description of data is listed in Table 1.

Level 1: Fault pattern	Level 2: Fault severity levels (mil)	Class encoding
Normal Condition (NC)	0	[1,0,0,0,0,0,0,0,0,0,0,0]
Ball Fault (BF)	7	[0,1,0,0,1,0,0,0,0,0,0,0]
	14	[0,1,0,0,0,1,0,0,0,0,0,0]
	21	[0,1,0,0,0,0,1,0,0,0,0,0]
Inner race Fault (IF)	7	[0,0,1,0,0,0,0,1,0,0,0,0]
	14	[0,0,1,0,0,0,0,0,1,0,0,0]
	21	[0,0,1,0,0,0,0,0,0,1,0,0]
Outer race Fault (OF)	7	[0,0,0,1,0,0,0,0,0,0,1,0]
	14	[0,0,0,1,0,0,0,0,0,0,0,1]
	21	[0,0,0,1,0,0,0,0,0,0,0,1]

Table 1: Description of the bearing health conditions

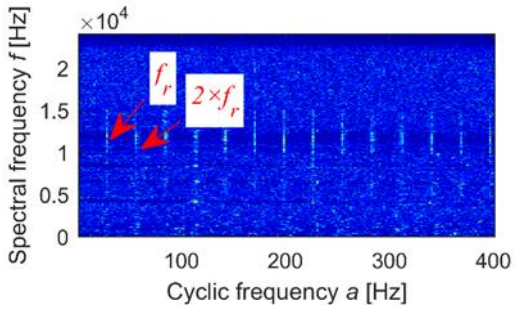
4.2 Analysis of the CSCoh 2D maps

In order to obtain the 2D CSCoh maps from the vibration measurements, 24000 data points (time duration of 0.5 seconds) are considered from the time-series signals to form one sample. There are 20 samples obtained from each health condition under every working load. In addition, it should be noted that, due to the limited sampling time, there are only 14 samples obtained for IF14 under the load 1.

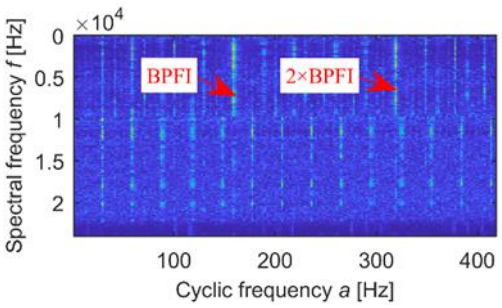
When defects of bearings occur, the bearing fundamental fault frequencies can be detected to analyse their dynamic behaviors. In order to validate the effectiveness of CSCoh in revealing the discriminative information of different fault types, four health conditions including the NC, IF with defect diameter of 7 (IF7), OF with defect diameter of 7 (OF7) and BF with defect diameter of 14 (BF14) are presented in figure 3.

It can be seen that the CSCoh maps provide unique representations for given fault types. In figure 3 (a), the fundamental shaft frequency (f_r) and its harmonic presents in the lower frequency are clearly observed, which is consistent with the dynamic behavior of the normal condition. In figure 3 (b) and figure 3 (c), Ball Pass Frequency of Inner-race (BPFI) and the Ball Pass Frequency of Outer-race (BPFO) and its harmonic can

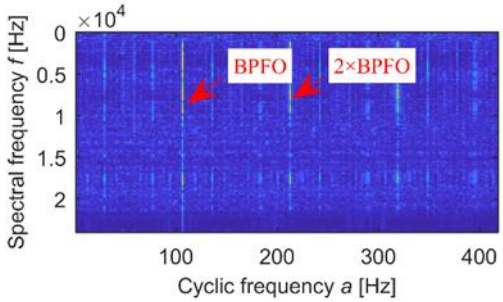
be clearly captured respectively, corresponding to the occurrence of the specific faults. It should be noted that in the case of BFs, presented in figure 4 (d), the weak amplitude of the Fundamental Train Frequency (FTF) and the Ball Spin Frequency (BSF) can be detected only in a few of samples of BF14, which reveal the existence of the ball fault. This demonstrates that the proposed is able to provide a good discriminative features when defects of bearings occurred.



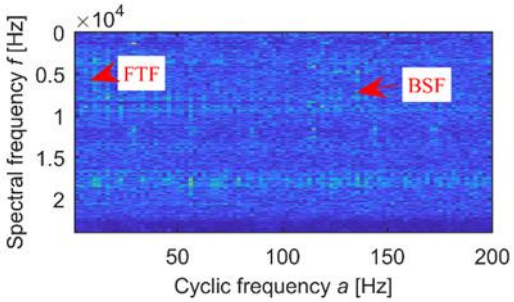
(a) Normal Condition (NC)



(b) Inner-race Fault with diameter of 7 mil (IF7)



(c) Outer-race Fault with diameter of 7 mil (OF7)



(d) Ball Fault with diameter of 7 mil (BF14)

Figure 3: The 2D CSCoh maps of different health conditions

4.3 A Fault diagnosis using the proposed method

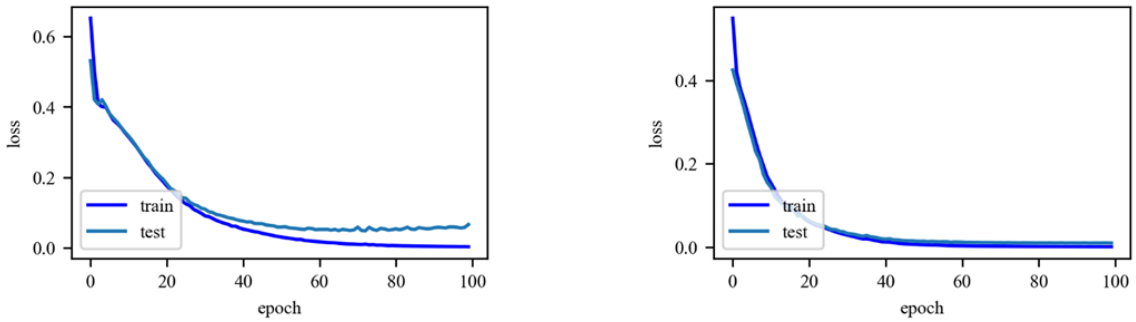
4.3.1 Effect of training sample ratio on classification performance

In order to study the influence of the training sample size on the classification rate, two datasets (dataset A and dataset B) are constructed to evaluate the performance. In dataset A, 20% of the samples are taken as the training data, while the rest for testing. In dataset B, 50% of samples are taken as training, and the rest for testing. Dataset A is constructed to simulate the insufficient training for the network. While dataset B is designed to sufficient training of network.

For CNN training, Adam algorithm is utilized to adjust the network weights with a batch size of 50. The epoch is set to 100, Ten trials are implemented to reduce the random. And the loss curves are shown in figure 4. From the figure 4, it can be seen that the training losses in both datasets are smooth, and remain stable, when they reach a certain number of iterations. It reveals that the models are well trained under the two training data. In the test stage, the loss curve in dataset A decreases slowly than that of the train stage, and it is close to a fixed value, and keep stable. It is possible that CNN with a large number of parameters trained on the insufficient training data suffers from the overfitting problem. By adding more training samples, the loss curve as shown in figure 4(b), are obviously decreasing.

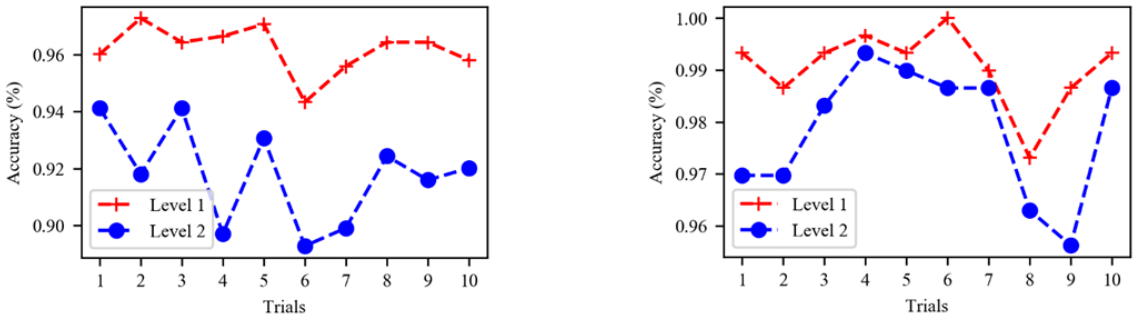
In addition, the diagnosis results of ten trials are presented in figure 5. “Level 1” denotes the classification accuracy of the fault patterns, where NC, BF, IF and BF are correctly recognized. “Level 2” reflects the total recognition accuracy, where the fault patterns and fault severity levels are all correctly classified.

From the results, it can be observed that, the results of all ten trials present relative high accuracy in both datasets. In addition, ‘Level 1’ is much higher than ‘Level 2’, since the former only needs to diagnosis the specific fault patterns, while the later requires to discriminative the fault severity levels of each fault patterns. Moreover, CNN with dataset A is much lower than that of dataset B, especially in ‘Level 2’. This is because that ‘Level 2’ contains more discriminative classes, which is more difficult for diagnosis. Therefore, the results of dataset B is able to obtain better classification performance compared to that of dataset A, since more of the training samples are contained.



(a) Loss curve of CNN using dataset A (b) Loss curve of CNN using dataset B

Figure 4: The loss curve of CNN with different datasets



(a) Results of 10 trials using dataset A (b) Results of 10 trials using dataset B

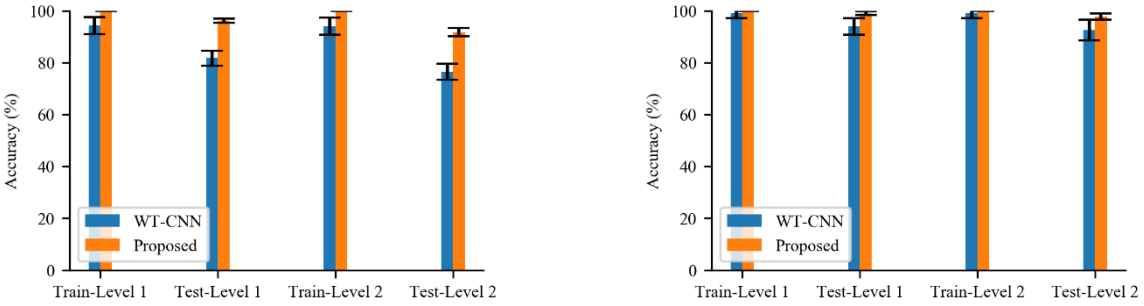
Figure 5: The results of ten trials of different datasets

4.3.2 Comparison with other methods

Wavelet Transform (WT), which decomposes the signals into wavelets, is usually considered as an effective tool for pre-processing nonstationary and transient signals [23]. The two-dimensional time frequency representation of WT provides a high resolution in both the time-domain and frequency-domain, which provides good information about the health conditions of rotating machinery.

In this section, a comparison of the WT scalograms and the CSCoh maps is carried out. Morelet wavelet basis is adopted to implement the time-frequency transformation. In order to provide a relative fair comparison, all the pre-processing data are fed into CNN for training, and the results are conducted with ten trials. Final results are averaged. The training and testing accuracy of different methods are shown in figure 6.

It can be seen that, the training accuracies of the proposed method in both datasets are 100.0%. While the training accuracies of WT are relative lower, and present larger standard deviations. In addition, the testing accuracy, especially, in ‘Level 2’, the proposed method also obviously performs better than that of WT-CNN.



(a) Result comparison of different methods using dataset A (b) Result comparison of different methods using dataset B

Figure 6: Result comparison of different methods

Furthermore, the testing accuracy of each types under different severity levels are further listed in Table 2 and Table 3, respectively. It can be observed that WT-CNN achieves average testing accuracy of 76.7% and 92.8%, in dataset A and dataset B, respectively. On the contrast, the proposed method achieves average testing accuracy of 92.2% and 97.9%, respectively, which is better that that of the WT-CNN. In addition, the diagnosis accuracies of the BF7 and the BF14 are obviously lower that of the other fault types. This is because that the characteristic frequencies in those cases are not obvious in the 2D CSCoh maps, which makes CNN difficulty to obtain good classification performance. From the analysis, it can be concluded that the proposed method is effective in extracting discriminative features and conducting the multi-label classification tasks.

Methods	Accuracy (%) of each fault severity level using dataset A										
	NC	BF7	BF14	BF21	IF7	IF14	IF21	IF7	IF14	IF21	AVG
WT-CNN	100	52.7	71.7	64.2	100.0	59.8	99.5	93.4	39.5	86.5	76.7
Proposed	100	100.0	79.0	58.4	100	96.1	95.9	100.0	96.4	96.5	92.2

Table 2: The average testing accuracy of each fault severity level using dataset A

Methods	Accuracy (%) of each fault severity level using dataset B										
	NC	BF7	BF14	BF21	IF7	IF14	IF21	IF7	IF14	IF21	AVG
WT-CNN	100.0	95.7	85.0	82.3	100.0	92.0	100	95.4	79.5	98.0	92.8
Proposed	100.0	100.0	95.7	89.6	100.0	96.5	98.3	100.0	99.0	99.4	97.9

Table 3: The average testing accuracy of each fault severity level using dataset B

Conclusion

In this work, a new DL-based fault diagnosis framework, combining CSCoh and CNN is proposed for multi-label fault classification. Firstly, CSCoh is considered, as a pre-processing step, to reveal the fault nature of each fault types. Then, a novel CNN is constructed for conducting fault classification with multiply labels by introducing the Sigmoid activation function and BCE loss function. The proposed method is verified on the data collected from the CWRU motor bearing test rig. Two different datasets including the insufficient training and sufficient training data are designed to evaluate the effectiveness of the methodology. It has been demonstrated that the proposed method not only achieves high classification performance, but also presents better generalization performance compared to WT-CNN fault diagnosis method.

References

[1] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, *Deep learning and its applications to machine health monitoring*, Mech. Syst. Signal Process. 115 (2019) 213-237.

[2] S. Khan, T. Yairi, *A review on the application of deep learning in system health management*, Mech. Syst. Signal Process. 107 (2018) 241-265.

[3] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, *Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data*, Mech. Syst. Signal Process. 72-73 (2016) 303-315.

[4] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, M. Modarres, *Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings*. Shock and Vibration, vol. 2017, Oct. 2017.

[5] B. Zhang, S. Zhang, W. Li, *Bearing performance degradation assessment using long short-term memory recurrent network*, Computers in Industry. 106 (2019) 14-29.

[6] S. Wang, J. Xiang, Y. Zhong, Y. Zhou, *Convolutional neural network-based hidden Markov models for rolling element bearing fault identification*, Knowledge-Based Systems, 144 (2018) 65-76.
[7] Z. Chen, W. Li, *Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network*, Ieee T Instrum Meas. 66 (2017) 1693-1702
[8] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, *Convolutional Neural Network Based Fault Detection for Rotating Machinery*, J. Sound Vibrat. 377 (2016) 331-345.
[9] C. Sun, M. Ma, Z. Zhao, X. Chen, *Sparse Deep Stacking Network for Fault Diagnosis of Motor*, IEEE Transactions on Industrial Informatics. 14 (2018) 3261-3270.
[10] P. Chopra and S. K. Yadav, *Fault detection and classification by unsupervised feature extraction and dimensionality reduction*, Complex Intell. Syst., vol. 1, no. 1, pp. 25–33, Jan. 2015.
[11] F. Althobiani and A. Ball, *An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks*, Expert Syst. Appl., vol. 41, no. 9, pp. 4113–4122, Dec. 2013.
[12] [P. Tamilselvan and P.Wang, *Failure diagnosis using deep belief learning based health state classification*, Rel. Eng. Syst. Safety, vol. 115, pp. 124–135, Jul. 2013.
[13] Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, *Real-time 751 motor fault detection by 1-d convolutional neural networks*, IEEE Trans. 752 Ind. Electron., vol. 63, no. 11, pp. 7067–7075, Nov. 2016
[14] F. Jia, Y. Lei, N. Lu, S. Xing, *Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization*, Mech. Syst. Signal Process. 110 (2018) 349-367.
[15] M. Gan, C. Wang, C.a. Zhu, *Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings*, Mech. Syst. Signal Process. 72-73 (2016) 92-104.
[16] L. Wen, X. Li, L. Gao, Y. Zhang, *A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method*, IEEE Trans. Industr. Electron. 65 (2018) 5990-5998.
[17] P. Borghesani, J. Antoni, *A faster algorithm for the calculation of the fast spectral correlation*, Mech. Syst. Signal Process. 111 (2018) 113-118.
[18] A. Mauricio, J., Qi, K. Gryllias, 2019. *Vibration-Based Condition Monitoring of Wind Turbine Gearboxes Based on Cyclostationary Analysis*. Journal of Engineering for Gas Turbines and Power, 141(3), p.031026.
[19] K. Gryllias, A. Mauricio, J. Qi, 2018, September. *Advanced cyclostationary-based analysis for condition monitoring of complex systems*. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 385-389). IEEE.
[20] A. Mauricio, J. Qi, W. Smith, R. Randall, K. Gryllias, *Vibration Based Condition Monitoring of Planetary Gearboxes Operating Under Speed Varying Operating Conditions Based on Cyclo-non-stationary Analysis*, 61 (2019) 265-279.
[21] Y. LeCun, G. Bengio, Hinton, *Deep learning*, Nature, 521 (2015) 436-444.
[22] Y. LeCun, L. Bottou, Y. Bengio, Y, P. Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278-2324.
[23] Yan, R.X. Gao, X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications*, Signal Processing, 96 (2014) 1-15.

A Semi-supervised Support Vector Data Description- based Fault Detection Method for Rolling Element Bearings based on Cyclic Spectral Coherence

Chenyu LIU [1,2], Konstantinos GRYLLIAS [1,2]

[1]Department of Mechanical Engineering,
Faculty of Engineering Science, KU Leuven
[2]Dynamics of Mechanical and Mechatronic Systems, Flanders Make
Celestijnenlaan 300, BOX 2420, 3001 Leuven, Belgium
chenyu.liu@kuleuven.be, konstantinos.gryllias@kuleuven.be

Abstract

Modern Internet-of-Things (IoT)-driven condition monitoring exploits data from various surveillance tools to reflect the health status of machinery. However, many diagnostic tasks are still hardly achievable, considering the complex operating environment of mechanical components. Detecting rolling element bearing faults, for timely repair and maintenance before a total breakdown, is one of the challenges which has a significant impact on rotating machinery. Nowadays advanced signal processing techniques are combined with high level machine learning approaches, focusing towards automatic fault diagnosis. A plethora of diagnostic indicators have been proposed to track the system degradation. Cyclic Spectral Correlation (CSC) and Cyclic Spectral Coherence (CSCoh) have been proved as powerful tools in signal processing community. Due to the periodic mechanism of the bearing faults' impacts, the diagnostic indicators extracted from CSCoh domain can better detect bearing faults. On the other hand, supervised machine learning approaches with labeled training and testing datasets cannot be realistically obtained under industrial conditions. In order to overcome this limitation, a novel fault detection approach based on semi-supervised learning and Support Vector Data Description (SVDD) is proposed in this paper. The proposed SVDD model utilizes CSCoh domain indicators to build the feature space and fits a hyper-sphere to calculate the Euclidean distances in order to isolate the healthy and faulty data. Meanwhile a systematic fault detection decision strategy is proposed to estimate the bearing status simultaneously with the detection of fault initiation. Two run-to-failure bearing datasets are used to implement the proposed fault detection framework and the results show that the SVDD model with CSCoh indicators can precisely detect the bearing faults. Moreover, the CSCoh based indicators show superior characteristics in the detection process comparing to time and Fast Kurtogram (FK) based Squared Envelope Spectrum (SES) domain indicators.

1 Introduction

Rolling element bearings are important but fragile components and are highly sensitive to external operation environment. Premature bearing failure caused by mounting errors, poor lubrication, debris contamination etc. may trigger a breakdown of machinery or even a fatal accident [1]. Therefore, the early bearing fault detection arises as a critical mission in the frame of predictive maintenance, receiving extensive attention in recent years.

Data-driven based fault detection methods have been widely applied to monitor rolling element bearings in rotation machinery. Its framework consists of four major steps: 1) data acquisition, 2) construction of Health Indicators (HI), 3) determination of decision threshold, 4) detection of anomaly [2]. Vibration signals are widely used in the data acquisition step to detect machine faults. HIs, derived from different domains, are used to track the bearing fault progression. Time domain HIs, such as root mean square (RMS), kurtosis, skewness etc., are widely employed [3]. They represent the statistical features of the time sequence but cannot characterize nonlinearities of bearing vibrations. On the other hand, a plethora of signal processing methods have been developed to enhance the embedded defect information, including Envelope Analysis, Spectral Kurtosis,

Short-Time Fourier Transform, Cyclic Spectral Analysis etc. [4]. These approaches allow the extraction of sophisticated HIs which correlate with bearing fault characteristic frequencies with more expressiveness and robustness in practical applications.

Among the above mentioned signal processing methods, cyclic spectral analysis recently gained much attention due to its capability of revealing the second-order cyclostationary periodicities of rolling element bearing signals. It has been proved that the vibration signals of faulty bearings, especially with localized fault, will exhibit cyclostationary behaviour [6]. Cyclic Spectral Correlation (CSC) and Cyclic Spectral Coherence (CSCoh) have been developed as efficient tools in cyclostationary analysis [7]. They represent the potential fault modulation information into frequency-frequency domain bivariable maps. The integration over the spectral frequency, leads to the estimation of the Enhanced Envelope Spectrum (EES), which demonstrates the modulation frequencies and their harmonics. Due to the periodic mechanism of the bearing faults' impacts, the EES can provide a clearer representation of bearing faults compared to other methods such as the Squared Envelope Spectrum (SES) [8]. Some attempts have been proposed to construct HIs from EES. In [9], a self-running bearing diagnosis framework with scalar indicators derived from EES has been set up . Moreover the summation of the amplitudes of three harmonics, which appear at the characteristic fault frequencies of EES, has been used initially as diagnostic indicators [10], and further as bearing prognostic indicators.

Another key problem is the definition of a threshold, which can trigger the detection based on the HIs, estimated on signals emitted by faulty bearings. On one hand, traditional statistic decision threshold strongly depends on the expert knowledge of the indicators' statistical distribution as well as on the empirical understanding of bearing faults. On the other hand, the detection threshold should be able to precisely isolate the faults meanwhile tolerating outliers to maintain a robust performance of the false alarm and misdetection rate under different external environments. Nowadays machine learning approaches show great potential to achieve this comprehensive target. In the research of automatic fault diagnosis, machine learning algorithms are considered as black-box models which can exhibit system behaviours from data. Therefore, it can provide continuous information for the rapid deterioration of bearings and detect anomalies. Support vector machine (SVM) is one of the initial machine learning attempts in real-world fault detection application [11]. The major principle of SVM is to seek an optimal hyper-plane as the decision surface that can maximize the margin of separation between two classes. In the case of bearing fault detection, the hyper-plane of SVM has been used as the detection threshold to separate indicators extracted from bearing signals into healthy and faulty class [12, 13]. Artificial neural network (ANN) is also widely applied to fulfill the task of fault detection. A supervised non-linear autoregressive neural network with exogenous input, was trained in [14], to model the healthy condition of bearings meanwhile setting up the detection threshold using the Mahalanobis distance to detect the bearing faults.

In real industrial conditions, the bearings are operating for most of the time in healthy conditions and the faulty datasets are rather limited and usually not labelled. Therefore standard supervised machine learning approaches, which are based on known labels of the classes, cannot be used for fault detection. In order to overcome this limitation, semi-supervised fault detection techniques have been proposed, based on the training of the methodology exclusively on healthy data. Support Vector Data Description (SVDD) has been proposed by Tax and Duin [15] and has been successfully used in one-class classification problems for various applications. SVDD is inspired by SVM but uses a hyper-sphere, instead of a hyper-plane, to solve the non-linear separable problem. SVDD provides a geometric description of the observed data by calculating the Euclidean distances between the data and the center of the hyper-sphere. It is trained by minimizing the volume of the sphere with most training data. When testing with the trained SVDD, any data found to have larger distance than the boundary of the sphere, are considered to be anomalies. In the bearing fault detection, the hyper-sphere of the SVDD can be used as the detection threshold.

In this paper, the CSCoh based diagnostic indicators and the SVDD classifier are combined in order to detect bearing faults. A comprehensive bearing status decision strategy is proposed within this framework. The methodology is tested and evaluated on experimental data from run-to-failure bearing datasets. The results demonstrate the efficacy of the method, presenting high detection rate with low false alarm and misdetection rate. The rest of the paper is organised as following: the background of Cyclic Spectral Coherence is discussed in Section 2 and the theory of SVDD is introduced in Section 3. The proposed methodology and the experimental setup are illustrated in Section 4 and 5, respectively. The results are presented and analysed in Section 6. Finally, the paper closes with some conclusions.

2 Cyclic Spectral Coherence

The concept of cyclostationarity was first proposed in the field of telecommunications and later was introduced to the mechanical engineering community. As a subcategory of non-stationary processes, cyclostationarity is a stochastic process that exhibits hidden periodicities embedded in systems. When a fault is generated in rolling element bearings in different locations (inner race, outer race or rolling element), a series of shocks in repetition form will be generated simultaneously, modulated by other frequencies like the shaft speed. This phenomenon can be described as cyclostationary and can be exploited to detect bearing damage. Different orders of cyclostationarity are defined based on the order of moments. The first-order cyclostationarity (CS1) is described as:

$$R_{1x}(t) = R_{1x}(t+T) = \mathbb{E}\{x(t)\} \tag{1}$$

where the first-order moment, called the statistical mean, $R_{1x}(t)$ is periodic with period T. $x(t)$ represents the signal function at time t. $\mathbb{E}\{x(t)\}$ is the ensemble average, which represents the average of the same stochastic process of repeated experiments. CS1 exhibits as periodic waveforms in vibration signals of rotating machinery and can be generated by imbalances, misalignments or flexible coupling. Moreover, second-order cyclostationary (CS2) depicts the periodicity of second-order moments with the autocorrelation function:

$$R_{2x}(t,\tau) = R_{2x}(t+T,\tau) = \mathbb{E}\{x(t)x(t-\tau)^*\} \tag{2}$$

where τ is the time-lag. CS2 provides a distinction of the stochastic process with amplitude or frequency modulation. CS2 has been proved extremely effective to achieve diagnosis on rotating components that are not completely phase-locked with shaft speeds, such as rolling element bearings.

The Cyclic Spectral Correlation (CSC) is designed as a tool to describe CS1 and CS2 signals in frequency-frequency domain. It is defined as the autocorrelation function of two frequency variables using the two-dimensional Fourier transform, as shown in Eq. 3:

$$CSC_x(\alpha, f) = \lim_{W\to\infty} \frac{1}{W}\mathbb{E}\{\mathscr{F}_W[x(t)]\mathscr{F}_W[x(t+\tau)]^*\} \tag{3}$$

where α is the cyclic frequency related with the modulation and f is the spectral frequency representing the carrier. $\mathscr{F}$ stands for the Fourier transform and W is the time duration. Furthermore, a normalization procedure can be added to the CSC, to minimize uneven distributions, which is known as Cyclic Spectral Coherence (CSCoh). The CSCoh is estimated as:

$$CSCoh_x(\alpha, f) = \frac{CSC_x(\alpha, f)}{\sqrt{CSC_x(0,f)CSC_x(0,f-\alpha)}} \tag{4}$$

The integration of the CSCoh over the spectral frequency f from zero to the Nyquist frequency can lead to the Enhanced Envelope Spectrum (EES) which is an improved version of the Squared Envelope Spectrum, as described in Eq. 5:

$$IES(\alpha) = \frac{1}{F_2 - F_1}\int_{F_1}^{F_2} |CSCoh_x(\alpha, f)|df \tag{5}$$

3 Semi-supervised SVDD

3.1 Theory

Semi-supervised learning approaches focus on partially labelled datasets. The objective of a semi-supervised model is to classify some of the unlabelled data, leveraging information from the labelled part. SVDD has been developed as one-class classification method [15] in the way of semi-supervised learning. It can be seen as an extension of SVM, which uses a hyper-sphere instead of a hyper-plane as the classification decision surface. The hyper-sphere is characterized by the radius R and the center a, which are the decision variables. The labelled data "targets" are regarded as the training set of SVDD. The data, being exactly on the boundary, are the support vectors and the data outside are "outliers". The primary goal of SVDD is to construct a hyper-sphere with a minimum radius, which simultaneously contains the maximum number of targets. The objective function can be described by Eq. 6:

$$F(R,a) = \min R^2 + C\sum_{i=1}^{n}\xi_i \tag{6}$$

This function subjects to two constraint conditions: $\|x_i - a\|^2 \leq R^2 + \xi_i, \forall i = 1,\ldots,n$ and $\xi_i \geq 0, \forall i = 1,\ldots,n$. Here $x_i \in \mathbb{R}$ is the training dataset with n samples. ξ_i is the slack variable which represents the tolerance of targets being outside the boundary. The slack variable is used to avoid an extremely large radius hyper-sphere that reduces the descriptive ability. C is the penalty factor that controls the trade-off between the radius and the rejected number of data points. It can also be described as $C = \frac{1}{nf}$, where f is the outlier fraction. Eq. 6 is a convex quadratic programming problem which cannot be solved for unknown R. However, it can be transformed to an equivalent dual problem with Lagrange duality. By applying Lagrange multipliers, the constraint conditions can be fused into the objective function, which turns it to a dual form as Eq. 7:

$$L = \max \sum_{i=1}^{n}\alpha_i(x_i \cdot x_i) - \sum_{i,j}\alpha_i\alpha_j(x_i \cdot x_j) \tag{7}$$

where $(x_j \cdot x_i)$ stands for the inner product of x_i and x_j, $\alpha_i \in \mathbb{R}$ are the Lagrange multipliers. New conditions are set for the dual form: $\sum_{i=1}^{n}\alpha_i = 1$ and $0 \leq \alpha_i \leq C, \forall i = 1,\ldots,n$. The dual form is composed of the data itself which makes the problem solvable. The training data x_i and its corresponding α_i are related with the radius R and the center a of the hyper-sphere according to Eq. 8:

$$\begin{cases} x_i - a^2 < R^2 \iff \alpha_i = 0 \\ x_i - a^2 = R^2 \iff 0 \leq \alpha_i \leq C \\ x_i - a^2 > R^2 \iff \alpha_i = C \end{cases} \tag{8}$$

3.2 Flexible data description

The inner product $(x_i \cdot x_j)$ in Eq. 7 can be replaced with kernel functions $K(x_i, x_j)$ to make a more flexible description for non-linear datasets. The kernel function can map the data to a higher dimension feature space which makes the non-linear data separable. The kernel functions have various choices such as Gaussian, linear, polynomial etc. Among all the functions, Gaussian kernel is the most common choice and is also adopted in this paper as described in Eq. 9:

$$K(x_i, x_j) = \exp\frac{-\|x_i - x_j\|^2}{2s^2}; \tag{9}$$

where s represent the kernel width parameter. By using the kernel function, the distance of any observation z to the center of the sphere can be described as follows:

$$\text{dist}^2(z) = K(z,z) - 2\sum_i \alpha_i K(x_i, z) + \sum_{i,j}\alpha_i\alpha_j K(x_i, x_j) \tag{10}$$

Considering the previous spherical data boundary conditions in Eq. 8, $\text{dist}^2(z) < R^2$ represents the position inside the sphere and $\text{dist}^2(z) = R^2$ is on the boundary of the sphere. When $\text{dist}^2(z) > R^2$, then z is recognized as an outlier. It should be noticed that the width of the kernel function K will influence on the calculation of the distance and is optimized in the proposed methodology, as explained in the following section.

4 Proposed methodology

4.1 Health Indicator construction

In order to realise the fault detection, HIs need to be constructed to evaluate the current condition as well as the degradation level of the bearing. The goodness of HIs can be characterized by monotonicity, trendability and prognosability. In real applications, the HIs should be able to detect bearing faults in early stage meanwhile minimizing the false alarm and misdetection rate. Time domain statistic features, such as Root Mean Square (RMS), kurtosis, skewness etc. are the most used HIs since they are effective to capture instant faults, such as using kurtosis in the detection of bearing spall initiation. However, the effectiveness of time domain HIs is restricted when the faults are distributed across the bearing surface. During the developing of damage, the

vibration signals exhibit more cyclostationarity with masked fault signatures. To overcome this limitation, the HIs, in this paper, are constructed using EES features from the CSCoh domain, as shown in Tab. 1. The CSCoh domain HIs are defined as the amplitude of the harmonics of the bearing characteristic frequencies extracted from the EES, i.e. the Shaft Rotation speed (SR), the Ball Pass Frequency Outer race (BPFO), the Ball Pass Frequency Inner race (BPFI) with sidebands and the Ball Spin Frequency (BSF) with sidebands. Four (4) harmonics and their sum are chosen to keep the robustness of the HIs and to facilitate the early start of detection.

S1R	EES peak at SR	S2I	EES peak at 2nd harmonic BPFI and sidebands
S2R	EES peak at 2nd harmonic of SR	S3I	EES peak at 3rd harmonic BPFI and sidebands
S3R	EES peak at 3rd harmonic of SR	S4I	EES peak at 4th harmonic BPFI and sidebands
S4R	EES peak at 4th harmonic of SR	SIR	Sum of 4 EES peaks of BPFI harmonics and sidebands
SSR	Sum of 4 EES peaks of SR harmonics	SIOR	Sum of 4 EES peaks of BPFO and BPFI harmonics
S1O	EES peak at BPFO	S2B	EES peak at two times BSF and sidebands at FTF
S2O	EES peak at 2nd harmonic of BPFO	S4B	EES peak at four times BSF and sidebands at FTF
S3O	EES peak at 3rd harmonic of BPFO	S6B	EES peak at six times BSF and sidebands at FTF
S4O	EES peak at 4th harmonic of BPFO	S8B	EES peak at eight times BSF and sidebands at FTF
SOR	Sum of 4 EES peaks of BPFO harmonics	SBF	Sum of 4 EES peaks of BSF and sidebands
S1I	EES peak at BPFI and sidebands		

Table 1: Health indicators from CSCoh domain

To compare the classification performance by using different health indicators, 9 statistic features from the time domain are also extracted during the test, as shown in Tab.2. Meanwhile, the same 21 frequency domain features from the Fast Kurtogram (FK) based SES are also extracted as another comparison group. The results from these indicators are illustrated in the following section.

RMS	$RMS = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2}{n}}$	Variance	$VR = \frac{\sum_{i=1}^{n}(x_i-m)^2}{(n-1)\sigma^2}$	Shape Factor	$SF = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$
Kurtosis	$KU = \frac{\sum_{i=1}^{n}(x_i-m)^4}{(n-1)\sigma^4}$	Mean	$m = \frac{\sum_{i=1}^{n} x_i}{n}$	Impulse Factor	$IF = \frac{\max\lvert x_i\rvert}{\frac{1}{n}\sum_{i=1}^{n}\lvert x_i\rvert}$
Skewness	$SK = \frac{\sum_{i=1}^{n}(x_i-m)^3}{(n-1)\sigma^3}$	Crest Factor	$CF = \frac{\max\lvert x_i\rvert}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}}$	Clearance Factor	$CLF = \frac{\max\lvert x_i\rvert}{(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\lvert x_i\rvert})^2}$

Table 2: Health indicators from time domain

4.2 Kernel width optimization using cross validation with grid search

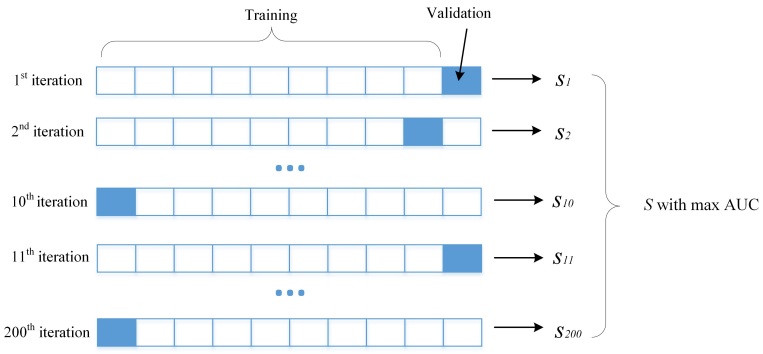


Figure 1: Optimization of kernel width using 10-fold Cross validation with grid search

The boundary of the hyper-sphere can be described with a kernel function for the SVDD classifier. In this paper, the Gaussian kernel is used since it has been proved effective in former research [15]. On the other hand, the kernel width acts as a hyperparameter in the SVDD model and influences the classification performance. The Receiver Operating Characteristic (ROC) curve is commonly used to evaluate the performance of a binary

classifier. The corresponding area under the ROC curve (AUC) is the key metric. The classifier with a larger AUC value performs better than the others. In order to tune the kernel width meanwhile proceeding cross validation during the training process, cross validation with grid search is adopted to search for the kernel width, that can maximize the AUC of the classifier. With 10-fold cross validation method, 200 iterations are proceeded to search for the kernel width s, which gives the maximum AUC values as depicted in Fig. 1. In this way, the model could achieve an automatic parameter optimization to avoid a stochastic selection of the hyperparameter. The results from the ROC curve will be discussed in Section 6.

4.3 Bearing fault detection procedure

The practical need for the bearing fault detection approach is a systematic decision strategy as a complement of the SVDD classifier based on CSCoh indicators. The flowchart of the proposed bearing fault detection process is shown in Fig. 2. The entire procedure consists of three major parts: feature extraction, classification and fault detection. The CSCoh based HIs are extracted from the EES to construct the feature space in the feature extraction stage. The features are then normalized by subtracting the mean value and dividing by the standard deviation to avoid the calculated Euclidean distances be governed by particular features in the SVDD classifier. Semi-supervised methods are able to learn from fewer labelled data points with the help of a large number of unlabelled data points, therefore the training data are selected from the health conditions which normally can be gathered at the beginning period of the bearing operation. In the classification stage, the entire dataset is split into training and testing sets with 25% and 75% of the whole dataset, respectively. The training set includes the validation part to realise the parameter optimization and meanwhile is labelled as healthy. A banana dataset is generated as artificial outliers around these real data to train the hyper-sphere as the classifier.

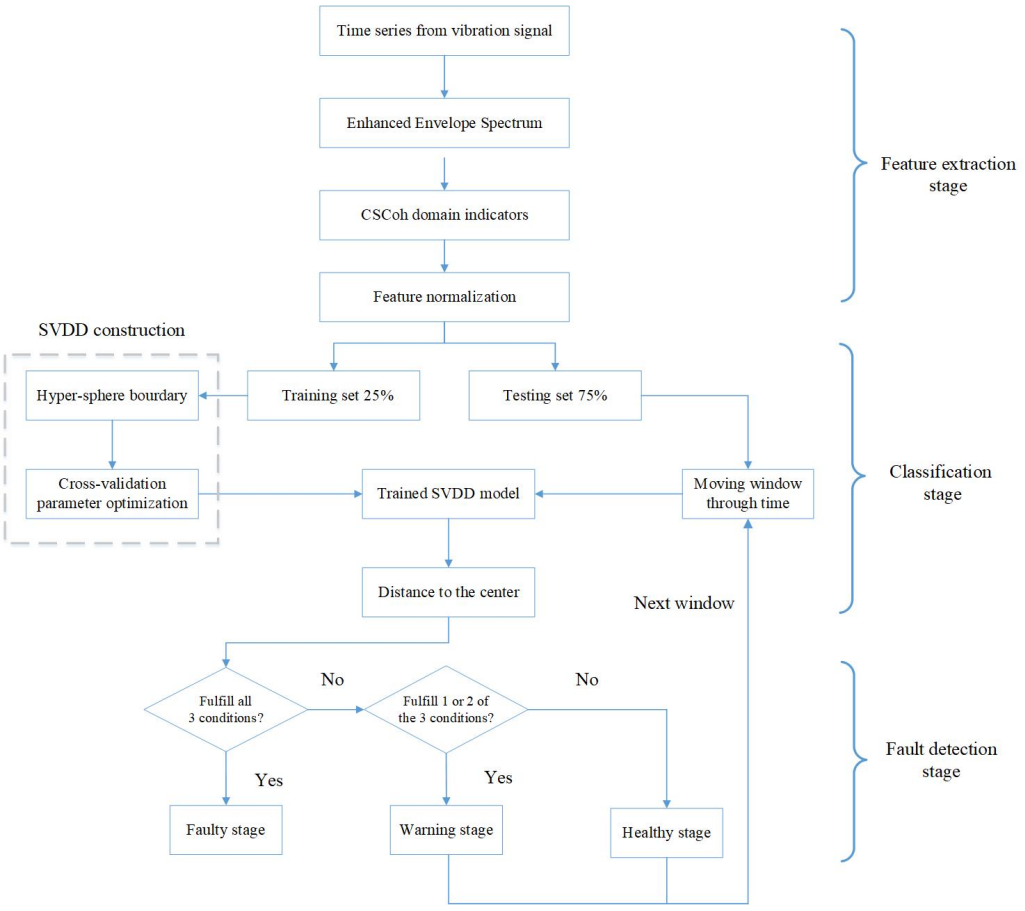


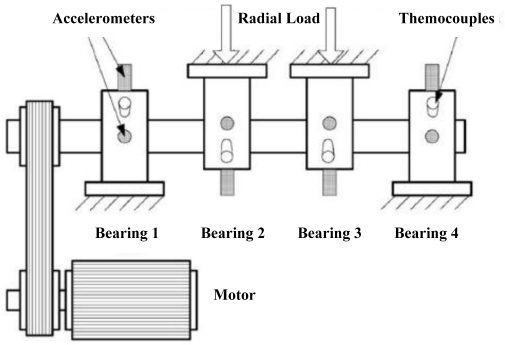
Figure 2: Procedure for the SVDD-based fault detection

Once the SVDD classifier is trained, a moving window is applied on the testing set. The data in the window are sent to the SVDD model to calculate the Euclidean distances to the center of the hyper-sphere. Considering the distances and the threshold, i.e. the boundary of the hyper-sphere, a fault detection decision strategy is built following 3 conditions: 1) 50% or more of the distances in the window are above the threshold. 2) 50% or

more of the distances in the window continuously pass the threshold. 3) The average distance of the data in the window is equal or greater than the threshold. These conditions can keep the detection threshold robust from the influence of random dominant outliers meanwhile reducing the misdetection rate. Three bearing statuses are then defined, according to the conditions. When all the three conditions are fulfilled, the status of the bearing is considered as faulty. In contrast, if less than 3 conditions are satisfied, then the status is considered as warning. If the distances from the bearing data obey none of the three conditions, it is considered as in a healthy status. Additionally, the starting point of the warning status represents the existence of a premature anomaly and the starting point of faulty status is the detection point of a mature bearing defect.

5 Experimental setup and dataset

The proposed SVDD based semi-supervised methodology is tested and validated using the NASA Intelligent Maintenance Systems (IMS) dataset [16]. The IMS dataset was collected during the rolling element bearing endurance experiment by using a dedicated test rig in the University of Cincinnati. The layout of the test rig is shown in Fig. 3, which consists of an electric motor coupled with a rub belt with 2000 rpm stationary speed, 4 double row Rexnord ZA-2115 bearings mounted on a common shaft lubricated by a circulation system, a 6000 lbs radial load applied on the bearing 2 and 3, a shaft and PCB 353B33 High Sensitivity Quart ICP accelerometers. The collecting of data was stopped when the accumulated debris exceeded a certain level on a magnetic plug and was considered as the end of bearing life. The physical parameters and the characteristic frequencies can be found in Fig. 3.



Physical Parameters	
Pitch diameter	71.5 mm
Rolling element diameter	8.4 mm
Number of rolling element per row	16
Contact angle	15.7
Static load	26690 N
Characteristic Frequencies	
Shaft rotation speed	33 Hz
Ball pass frequency outer race (BPFO)	236 Hz
Ball pass frequency inner race (BPFI)	297 Hz
Ball spin frequency (BSF)	139 Hz
Fundamental train frequency (FTF)	15 Hz

Figure 3: IMS test rig layout (left) and characteristics of bearings (right)

The run-to-failure experimental results were published in three datasets with numerous of files. Some details should be noticed when processing these datasets. According to the document describing the experiment, the sampling frequency should be 20 kHz but the file for one second includes 20,480 points so the sampling frequency probably was 20.48 kHz and this one is used in this paper. Moreover, the measurement for Dataset 1 is not continuous but with several interruptions during recording. These interruptions influences the behaviour of some HIs and this issue will be discussed in the following section. The signals from the dataset will be marked with time (Day) and number of signal (# signal number).

6 Results and discussion

6.1 Case 1: Dataset 1 Bearing 3

The 9 time domain HIs, the 21 FK based SES domain HIs and the 21 CSCoh domain HIs estimated processing the Dataset 1 Bearing 3, after normalization, are presented in Fig. 4. It is noticeable that the time axis is not in equal intervals for this map and other figures involving Dataset 1 in this section, due to the recording interruptions. The time axis represents the real time points in days from beginning (Day 0) to the end of the experiment (Day 34.5) but not the recording time. Analysing the time and the FK based SES features in the map, it can be visually observed that there is a significant changing step before Day 7.1 (#157), caused by a long period interruption. Therefore the data before Day 7.1 are abandoned to avoid contamination of the

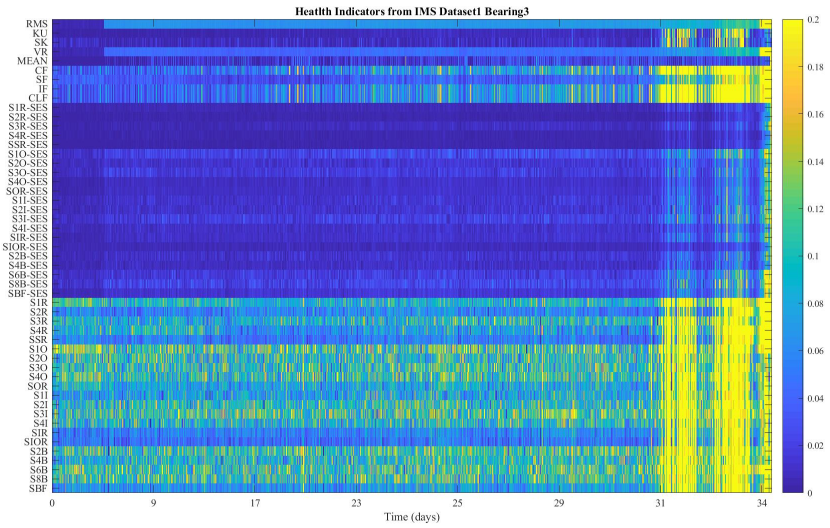


Figure 4: Health indicator map for Dataset 1 Bearing 3

healthy training set, which is selected from Day 7.4 (#200) to Day 17 (#600). The remaining data are used for testing.

The distances of the CSCoh indicators to the SVDD center are shown on the left of Fig. 5. Based on the proposed decision framework, the fault detection result is shown on the right of Fig. 5. The warning stage starts from Day 29.2 (#1570) till Day 30.9 (#1795) and the fault is detected also at Day 30.9 (#1796).

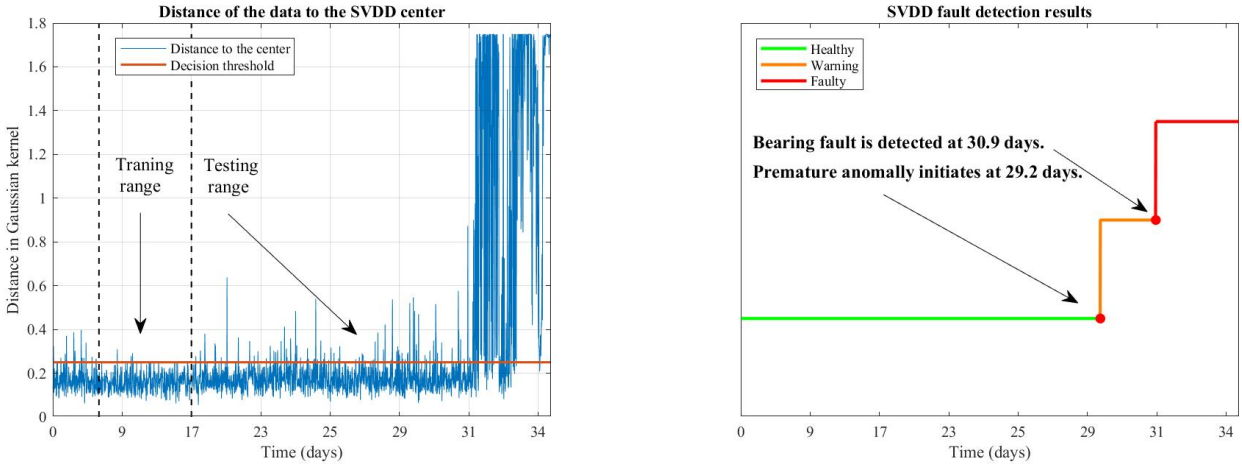


Figure 5: Fault detection restuls (left) and testing ROC curve (right)
of CSCoh indicators of Dataset 1 Bearing 3

The evaluation of the detection performance uses results from the study [17] as reference. The testing ROC curve of the SVDD classifier is plotted in Fig. 6. As discussed in Section 4, the ROC curve is employed to evaluate the classification ability of a binary classifier. The horizontal and vertical axis of a ROC curve represent the False Positive rate (FPr) and the True Positive rate (TPr). The classifier behaves better when the curve is closer to the left up point where FPr=0, TPr=1. Additionally, the area under ROC curve (AUC) can be used as a quantitative index in the evaluation. A larger AUC represents a better classification performance. Furthermore, more indexes are used to evaluate the classifier in the later part.

The HIs from the time and the FK based SES domain are also used to train the SVDD model as shown in Fig. 7 and 8. Compared to the results of CSCoh, it is obvious that the distances of the time and the FK based SES domain indicators exhibit wide fluctuations around the threshold. The outliers above the threshold spread in the whole testing range for both the two HI groups. Hence, the fault detection results show longer warning stages for the two groups with 7.9 and 8 days respectively. The time indicators detected the fault at Day 31 (#1820) and the FK based SES indicators give the detection at the same time as the CSCoh indicators at Day

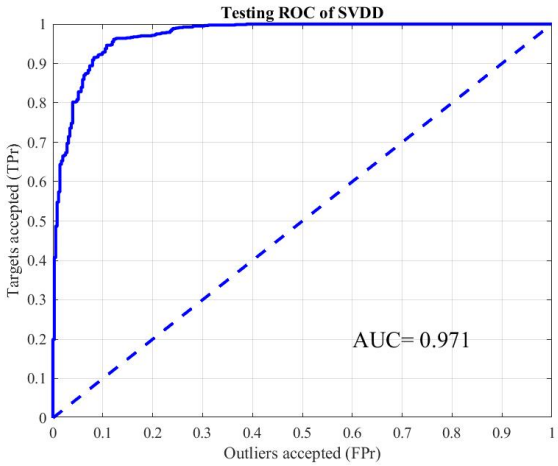


Figure 6: ROC curve for SVDD with CSCoh indicators of Dataset 1 Bearing 3

30.9 (#1796). The detection results show that time and FK based SES indicators are sensitive to the changing of external environment due to the recording interruptions. On the contrary, the distances of SVDD results from CSCoh indicators keep steady during most of the testing period which indicates that the CSCoh HIs are robust in detecting the bearing faults.

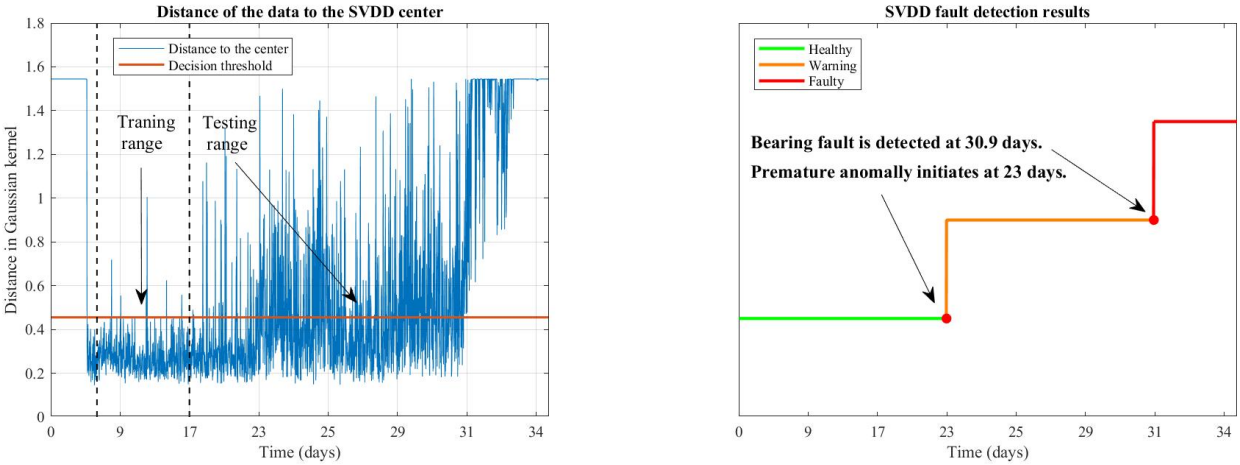


Figure 7: Distances to SVDD center (left) and fault detection results (right) for time domain indicators of Dataset 1 Bearing 3

The metrics derived from the classification confusion matrix are adopted to compare the performance of SVDD with different HIs, including the False Positive Fraction (FPF), the False Negative Fraction (FNF), the False Positive Rate (FPR) and the False Negative Rate (FNR). The calculation of the four values is shown in Fig. 9. The metrics represent the classification performance, measured from the estimated labels and the true labels, based on [17]. More specifically, FPR can be seen as the evaluation of false alarms and FNR represents the misdetection of the classifier. Lower values of these metrics indicate a better classification performance. The metrics together with the AUC value of all three group of HIs are listed in Tab. 3. The SVDD classifier apparently performs better with indicators from CSCoh than the other two groups, presenting superiority in the accurately detection of bearing faults.

6.2 Case 2: Dataset 2 Bearing 1

The HI map of Dataset 2 Bearing 1 is shown in Fig. 10. Since the measurement of this dataset is continuous without any interruption, all the indicators behave more monotonic than in Dataset 1. A clear separation can be seen at Day 4.9 for all the indicators which represent a dramatic increasing of the HIs. On the other hand, the HIs from the CSCoh show a forehead increase at Day 3.8 for the indicators constructed with the harmonics and

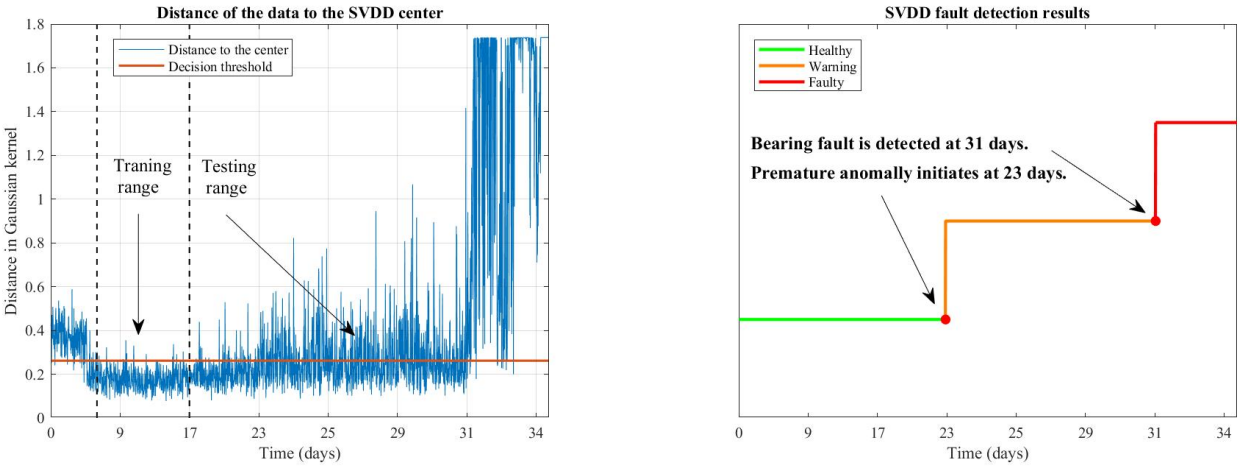


Figure 8: Distances to SVDD center (left) and fault detection results (right) for FK based SES indicators of Dataset 1 Bearing 3

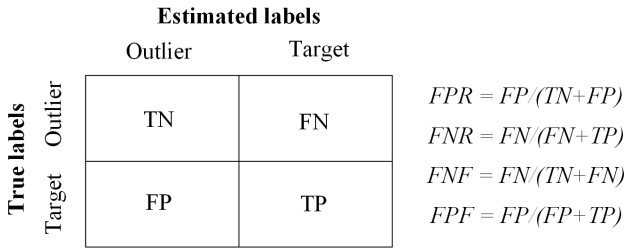


Figure 9: Calculation of the evaluation metrics.

	FPF	FNF	FPR	FNR	AUC
Time	0.474	0.195	0.370	0.085	0.905
FK based SES	0.248	0.040	0.525	0.023	0.963
CSCoh	0.076	0.023	0.177	0.019	0.971

Table 3: Performance of SVDD classifiers with different indicators for Dataset 1 Bearing 3

the sum of BPFO, which is the fault type in this case.

The distances from the SVDD and the fault detection results of the CSCoh indicators are shown in Fig. 11. The training set is selected from Day 0.3 (#50) to Day 2.1 (#300) and the remaining data are used for testing. Less distances pass the threshold which represent the monotonicity of the indicators. The warning stage is 0.1 days from Day 3.6 (#521) to Day 3.7 (#533) which is relatively short due to the indicators with more tendency. Then the bearing fault is detected at Day 3.7 (#534).

The indicators from the time and the FK based SES are also sent to the SVDD model to realise fault detection and the results are shown in Fig. 12 and 13. The CSCoh indicators provide the earliest detection compared to the time indicators (Day 3.8, #550) and the FK based SES indicators (Day 3.7, #541). The amplitudes of the characteristic frequencies and their harmonics from EES present more monotonicity and trendability which provide evidence in the construction of feature space during the training of SVDD model. As a result, the detection performs better using the CSCoh indicators.

The metrics of the classifiers' performance among different groups of HIs show less variance, compared to the first case. The CSCoh indicators gain lower FPF, FPR and FNR. The time domain indicators get lower FNF but the AUC is still lower than for the CSCoh.

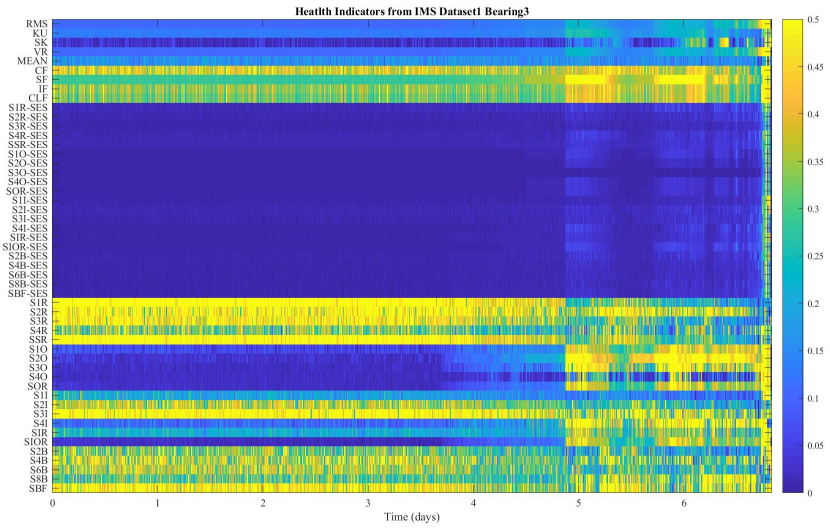


Figure 10: Health indicator map for Dataset 2 Bearing 1

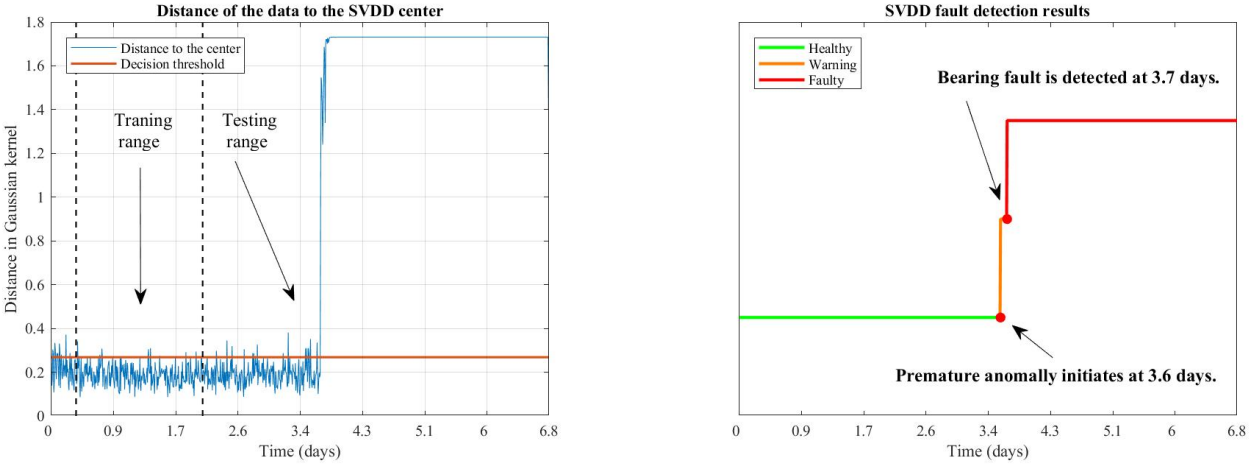


Figure 11: Distances to SVDD center (left) and fault detection results (right) for CSCoh indicators of Dataset 2 Bearing 1

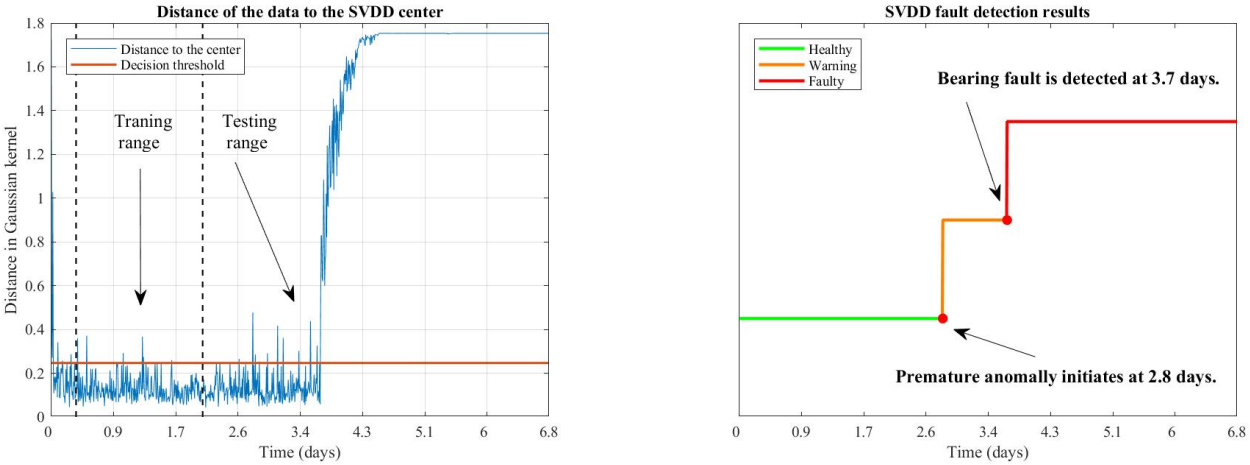


Figure 12: Distances to SVDD center (left) and fault detection results (right) for time domain indicators of Dataset 2 Bearing 1

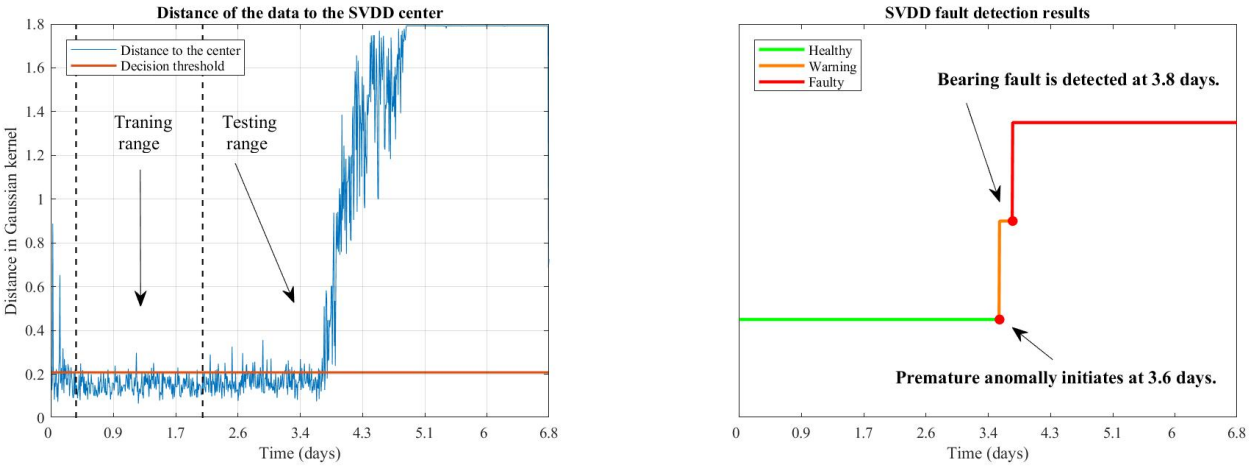


Figure 13: Distances to SVDD center (left) and fault detection results (right) for FK based SES indicators of Dataset 2 Bearing 1

	FPF	FNF	FPR	FNR	AUC
Time	0.078	0.018	0.028	0.041	0.991
FK based SES	0.090	0.020	0.038	0.005	0.992
CSCoh	0.056	0.020	0.028	0.005	0.992

Table 4: Performance of SVDD classifiers with different indicators for Dataset 2 Bearing 1

7 Conclusion

In this paper, a novel rolling element bearing automatic fault detection approach is proposed by combining CSCoh domain indicators and a semi-supervised SVDD technique. Experimental results from run-to-failure bearing datasets prove that the bearing faults can be accurately detected with the proposed methodology. Compared to the indicators from time domain and the FK based SES domain, CSCoh diagnostic indicators are more robust to the changing of external environment and improve the performance of the SVDD classifier. The proposed semi-supervised methodology therefore has a strong practical significance in industrial applications since it can be effective in both bearing failure warning and damage detection.

Acknowledgements

Chenyu Liu would like to acknowledge the support from China Scholarship Council.

References

[1] R. Randall , J. Antoni, *Rolling element bearing diagnostics-A tutorial*, Mechanical Systems and Signal Processing, 2011 February, 1, vol. 25, no. 2, pp.485-520.

[2] B. Zhang, G. Georgoulas, M. Orchard, A. Saxena, D. Brown, G. Vachtsevanos, S. Liang. *Rolling element bearing feature extraction and anomaly detection based on vibration monitoring*, 16th Mediterranean Conference on Control and Automation, 2008 June 25, pp.1792-1797.

[3] K. Gryllias, S. Moschini, J. Antoni, *Application of cyclo-nonstationary indicators for bearing monitoring under varying operating conditions*, Journal of Engineering for Gas Turbines and Power, 2018 Jan 1, vol 140, no. 1.

[4] C. Yi, D. Wang, W. Fan, KL. Tsui, J. Lin, *EEMD-based steady-state indexes and their applications to condition monitoring and fault diagnosis of railway axle bearings*, Sensors, 2018 Feb 27, vol. 18, no. 3, pp.704.

[5] Y. Lei, F. Jia, J. Lin, S. Xing, SX. Ding, *An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data*, IEEE Transactions on Industrial Electronics, 2016 May, vol. 63, no. 5, pp.3137-3147.

[6] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, *Cyclostationary modelling of rotating machine vibration signals*, Mechanical Systems and Signal Processing, 2004 Nov 1, vol. 18, no. 6, pp.1285-1314.

[7] A. Mauricio, A. Miguel, W. Smith, R. Randall, J. Antoni, K. Gryllias, *Cyclostationary-based tools for bearing diagnostics*, Online Proceedings of ISMA 2018 including USD 2018, 2013, pp.905-918.

[8] A. Mauricio, J. Qi, K. Gryllias. *Vibration-Based Condition Monitoring of Wind Turbine Gearboxes Based on Cyclostationary Analysis*, Journal of Engineering for Gas Turbines and Power, 2019 Mar 1, vol. 141, no. 3.

[9] S. Kass, A. Raad, J. Antoni, *Self-running bearing diagnosis based on scalar indicator using fast order frequency spectral coherence*, Measurement, 2019 May 1, vol. 138, pp.467-484.

[10] J. Qi, A. Mauricio, K. Gryllias, *Enhanced Particle Filter and Cyclic Spectral Coherence based Prognostics of Rolling Element Bearings*, In PHM Society European Conference, 2018 Jun 28, vol. 4, no. 1.

[11] K. Gryllias, I. Antoniadis, *A Support Vector Machine approach based on physical model training for rolling element bearing fault detection in industrial environments*, Engineering Applications of Artificial Intelligence, 2012 Mar 1, vol. 25, no. 2, pp.326-44.

[12] T. Shon, Y. Kim, C. Lee, J. Moon, *A machine learning framework for network anomaly detection using SVM and GA*, In Proceedings from the sixth annual IEEE SMC information assurance workshop, 2005 Jun 15, pp.176-183.

[13] R. Perdisci, G. Gu, W. Lee, *Using an Ensemble of One-Class SVM Classifiers to Harden Payload-based Anomaly Detection Systems*, In ICDM, 2006 Dec 18, vol. 6, pp.488-498.

[14] P. Bangalore, LB Tjernberg, *An artificial neural network approach for early fault detection of gearbox bearings*, IEEE Transactions on Smart Grid, 2015 Mar, vol. 6, no. 2, pp.980-987.

[15] D. Tax, R. Duin, *Support vector data description*, Machine learning, 2004 Jan 1, vol. 54, no. 1, pp: 45-66.

[16] H. Qiu, J. Lee, J. Lin, G. Yu. *Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics. Journal of sound and vibration*, 2006 Feb 7, vol. 289(4-5), pp.1066-1090.

[17] K. Gryllias, I. Antoniadis, C. Yiakopoulos, *A novel semi-supervised mathematical morphology-based fault detection & classification method for rolling element bearings*, In Proceedings of the 22nd International Congress on Sound and Vibration, 2015.

Machine teaching to optimize algorithms performances on restricted dataset.

Alexandre CARBONELLI[1], Maurin WOUSSEN [1]

[1]Vibratec, 28 Chemin du Petit Bois, 69130 Écully, France
alexandre.carbonelli@vibratec.fr

Abstract

Machine Learning techniques open important ways of development of physical models in almost every field. Performances reached rely on two major pillars: the (physical) model and the data. When a huge amount of data is available, the intrinsic characteristics of the chosen algorithm become less critical. On the other hand, with limited amount of data, all the human knowledge on the system to be modelled becomes critical to exploit.

One of the multiple applications of Machine Learning (ML) technics concerns the meta-models. Indeed, in this paper, we show how we can bypass a computation scheme by using clever regression models. The approach is performed on a system from which we want to know instantaneously the first natural frequencies without performing each time finite elements (FE) computation. We study the performance reached by studying the number of training cases required to teach the algorithm how to link inputs to outputs within a satisfying accuracy. Different algorithms are tested with very encouraging results as going into higher dimensional problem. The final aim of this study is to provide global guidelines for the most efficient Machine Teaching.

1 Introduction

Reducing the computation time is an objective is many cases. Whether it is for having instantaneously a result or to integrate the meta-model in a bigger loop, the idea remains the same: to link inputs to outputs without going through a whole computation process. The best way to achieve that is for instance to have an analytical formula based on physics consideration which can represent exactly the phenomenon studied. According to the task, it is unfortunately not always possible to use an analytical formula. The typical representative case of such a task would be for instance a classification problem: to determine if a given picture represents a cat or a dog, it seems obvious that we will not find an analytical formula able to do that. To illustrate our idea we can consider Figure 1.

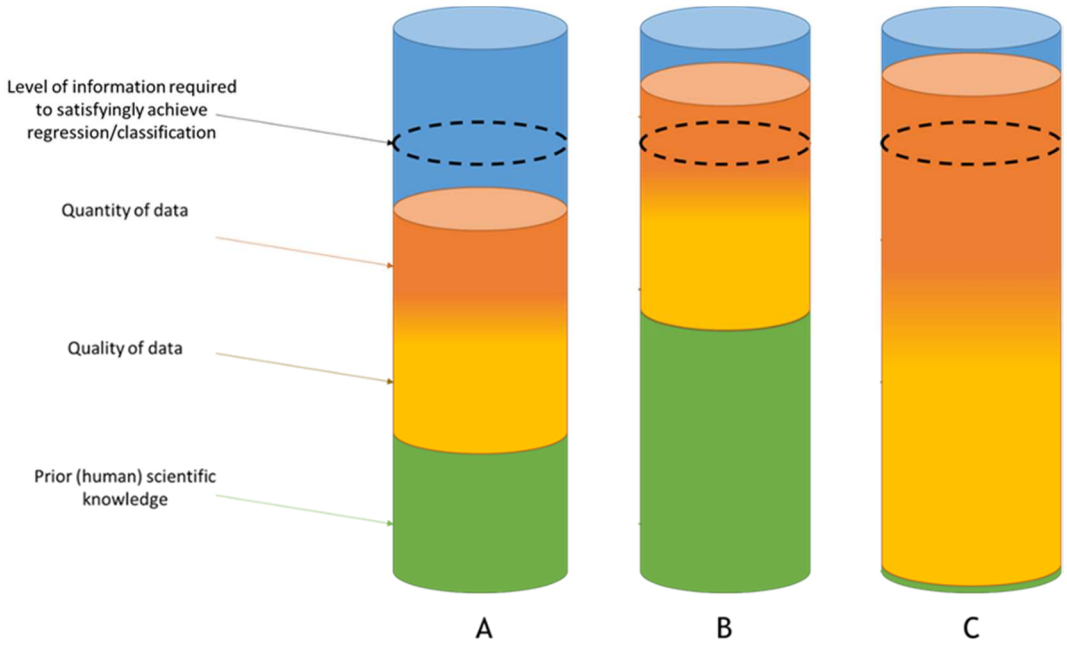


Figure 1: General approach of a regression/classification problem

When we lack information (case A), we can either use scientific knowledge (case B) or generate more data (case C) or of course both in order the fill the gap.
The general parameters suggested to be considered are:

- The level of information required to ***satisfyingly*** achieve the regression/classification implies to set a threshold beyond which we estimate that the task is a success,
- The quantity of data is directly linked to their quality (we consider that in the next part), meaning that many data of poor quality could not be that helpful according to the accuracy aimed for,
- The prior (human) scientific knowledge establishes the base of the whole stack (for instance, an adapted analytical formula fills the gap to the level of information required by its own).

This approach will be later included into a classification task consisting in detecting in a block made of carbon is cracked or not by studying its natural frequencies. We focus in this paper on the determination (through regression) of the natural frequencies of the block without any additional FE computation than those used for the model building. The accuracy of the reference natural frequencies (those obtained for a healthy block) must be high in order to be the more sensitive possible in the incoming classification task.

2 The concept of Machine Teaching

The Machine Teaching concept as we thought it relies on a direct analogy on human learning. Let's consider a student who has a set of examinations.

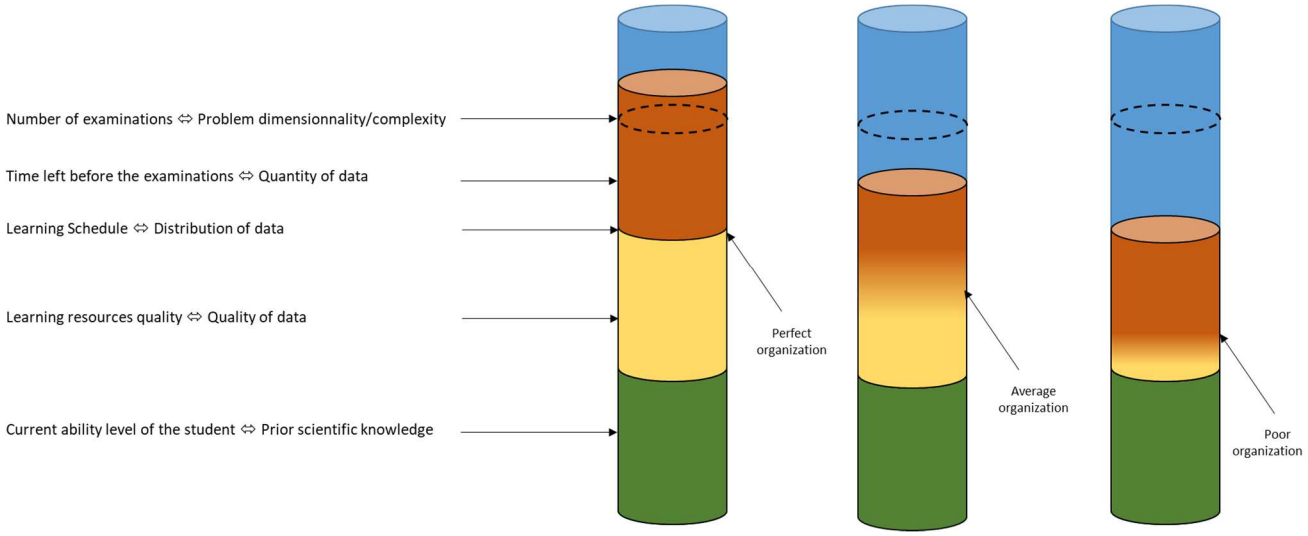


Figure 2: Illustration of the Machine Teaching concept

We suggest on Figure 2 the following symmetries. The number of examinations can be related to the complexity of the task or its dimensionality. The time before the examinations can be related to the quantity of data that one can generate in a given time. The learning resources (writing quality, the teacher implication and so on) can be related to the quality of data (does one need to correct and to complete manually the dataset, taking risk to insert errors…). The current ability of the student can be related on the priori scientific knowledge: what does he knew before any learning? His knowledge basis. Finally, the learning schedule is directly related for us to the concept of Machine Teaching. Indeed, having well distributed data (not necessarily uniform) permits to optimize the supply of each new data, avoid generalization problem, eliminate redundancies and perform better. To continue the analogy, for a student with a limited amount of time, a good schedule learning during which he will be able to the see different cases. The most representative cases will definitely improve his learning compared to spend most of the time on one chapter and ignoring the six left. Therefore a good teaching can be related to one ability to understand the knowledge compulsory to integrate and to provide consistent organization of the data.

3 Presentation of the study case

3.1 The free-free beam

The study of the free vibration of a beam is not that challenging. Nevertheless, classical approach using Euler-Bernoulli's and Timoshenko's beam theories [1-2] provides satisfying as long as some assumptions are satisfied, such as the ratio between the length of the beam and its cross-sectional dimensions (in which case the Euler-Bernouilli's beams present an important error). In our case, we don't want to make any assumptions about the beam behaviour and we use therefore FE model with solid elements to feed the meta-model. The Figure 3 shows two beam samples, which can be rectangular or circular. The Table 1 lists the different parameters considered and their variation ranges and Table 2 sums up the meshing details used.

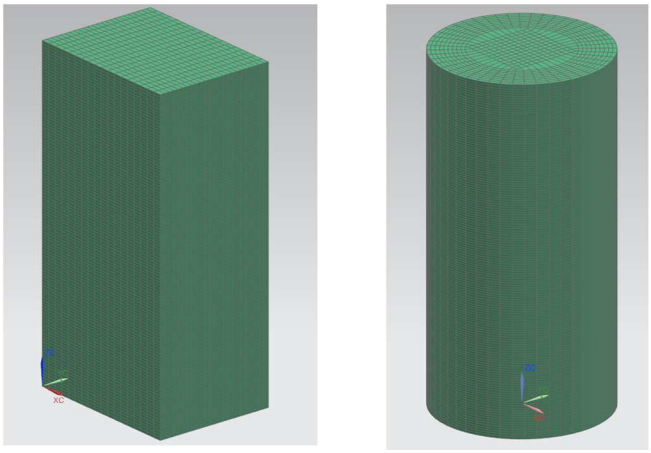


Figure 3: Examples of considered rectangular and circular beams.

Parameters	Minimum	Maximum
Length (m)	2	9
Section[1] (m)	0.3	1
Young's modulus (MPa)	7.92	11.88
Poisson's ratio	0.1	0.4

Table 1: Parameters studied and their variation ranges

	Rectangular beams	Circular beams
Element's type	Cube	Rectangular cuboid
Type of grid	Structured	Structured
Number of elements	50 112	93 600
Number of nodes	55 575	97 999

Table 2: Meshing details

3.2 Automatic generation of samples

The study was realized in two steps, a first naïve approach to the problem, and then a second step where we developed and improved a method. The different cases firstly used to feed the meta-model are generated automatically and are divided in N_1 circular and N_2 rectangular. We separate two different types of set, the training data used for the learning procedure, generated in a grid way, we will explain later the reason of this choice. And the test data, used for the meta-model validation, which are randomly generated, to emphasize the generalization on the entire study domain. We recall the different subsets in Table 3 and Table 4*: Subsets labels of the second step***Erreur ! Source du renvoi introuvable.**.
L, b, h, E, ν respectively stand for the length, the height, the thickness, the Young's Modulus and the Poisson's coefficient of the beam.

[1] Section corresponds to the larger length of the section's beam, i.e. diameter for circular beams or maximum between thickness and height for rectangular beams.

	First Step			
	Rectangular beams		Circular beams	
	Train	Test	Train	Test
Variables involved	b, h, L	b, h, L	b, h, L	b, h, L
Number of samples	512	141	400	75
Generation	Grid	Random	Grid	Random

Table 3: Subsets labels of the first step

	Second Step			
	Rectangular beams		Circular beams	
	Train	Test	Train	Test
Variables involved	b, h, L, E, ν	b, h, L, E, ν	b, h, L, E, ν	b, h, L, E, ν
Number of samples	100	50	100	50
Generation	Grid	Random	Grid	Random

Table 4: Subsets labels of the second step

We begin with 512 samples of rectangular beams for the training process, such as we have eight different values of length, thickness and height to make the grid. The initial distribution is naively uniform.
Its test set counts 141 random samples, with a criteria on the ratio, length over the working dimension of the section, where it can't overstep 3.33 for each configuration. The first training data for the circular beams contains 400 samples, such as we have twenty values of length and twenty values of diameter, to make a grid. We have 75 samples for its test set.

For the second step, we include the material properties in the study and we try to reduce the number of samples needed for the learning. For both of the circular and rectangular cases, we generate a grid such as we have 5 values of length, 5 values of thickness, 2 values of Young's modulus and 2 values of Poisson's ratio. We'll explain later the reasons of that configuration choice. As we add two new features, we also have to create new test sets of 50 samples each, generated randomly.

Finally, a total of 1428 samples have been automatically generated. To give an idea of the length of the process, one computation takes about three minutes. Then, to generate the complete set of input data, about 72 hours have been required.

3.3 Automatic extraction of data

The outputs required are simply the three first traction/compression (Z axis) eigenfrequencies and the three first flexural eigenfrequencies in both directions (X and Y axis). Each computation output file are post-process and for each case the 9 outputs are written related to the parameters defined in part 3.1.

4 Methodology for building the meta-model

Facing a lack of information, we firstly choose the random forest as our starting strategy to predict the eigenfrequencies. Indeed, the ensemble methods can have good results on small subsets, where they start from a high variance and thanks to bagging algorithm, decreases it during the learning process to improve the generalization [4]. However, many parts of the response values are underrepresented, which distorts the training phase, leading to a bad accuracy. We had to change our point of view on the procedure and try to develop a way to guide the algorithm during the training so that we can control how it learns along the process and "help" it. That's why we choose to build a custom learning method, closer to the machine teaching area. This method is based on nonlinear least squares [5] and the use of a specific grid as training data.
We can see in the Table 5: Comparison of relative error between random forest and a custom method a comparison on the worst mode's results, between random forest and our custom method. To do so, we calculate the relative error e such as:

$$e = \sqrt{\frac{(y-\tilde{y})^2}{y^2}} \times 100 \quad (1)$$

with y, the observation value and $\tilde{y}$, the prediction from the modelling function. This error is determined for each point of the test subset.

	Random Forest			Custom Method		
	F1X	F1Y	F1Z	F1X	F1Y	F1Z
Mean (%)	9.57 ; 29.1	13.7 ; 29.3	0.210 ; 5.38 10^{-2}	9.35 10^{-2} ; 0.122	0.100 ; 0.122	0.212 ; 0.265
Max (%)	46.54 ; 154	39.98 ; 156	2.42 ; 0.346	0.264 ; 0.554	0.255 ; 0.554	0.414 ; 0.843

Table 5: Comparison of relative error between random forest and a custom method
Rectangular beams / Circular beams

This table shows the limit of classical machine learning algorithms in this context. We had to modify existing methods to have a better fitting and more complex regression curve.

The main idea behind the method used is to deconstruct the learning procedure, so that we can control the tendency of our response depending on each features separately, during each step. We want to approximate the output with a function such as:

$$f: \quad I_1 \times ... \times I_N \rightarrow \mathbb{R} \qquad (2)$$
$$(v_1 \times ... \times v_N) \rightarrow f((v_1 \times ... \times v_N), \beta)$$

Where N is the number of features, $I_k \subset \mathbb{R}, k = \{1, ..., N\}$ the set where the feature is taken from, and β a set of parameters used by the function f. In our situation, we can choose the configuration $(v_1 \times ... \times v_N)$ we want to append in our samples set. Then, we choose to build a grid, as a combination of values from chosen subsets. The strategy is given in Appendix.

5 First step

5.1 Strategy

To start the study, several comments have been done to make easier the resolution. We choose to consider each eigenfrequency separately, instead of using multioutput regression algorithms which tends to be less accurate. To approximate the transverse modes along the Y axis, we used the symmetric properties of these modes and take the regression function on the transverse modes along X axis, where we inverse the thickness and height values. We now had to build 6 modelling functions. We also consider that the height does not influence the transverse modes along X axis, so the approximation will involve only two dimensions. For the longitudinal modes, we noticed that the eigenfrequencies are highly correlated with the length, so the regression function will only depend on this feature.
Our custom method were mainly used to build the transverse modes interpolation. So we have two layers, according to the features involved. The thickness layer is approximated with polynomials of second order and the parameters to estimate as related to a rational function such as:

$$f(x) = \frac{\beta_1 x^2 + \beta_2 x + \beta_3}{\beta_4 x^2 + \beta_5 x + 1} \quad (3)$$

Where x is the feature and $\beta_i, i = 1, ..., 5$ the parameters. The final function looks like:

$$f(b,L) = \frac{\alpha_1 L^2 + \alpha_2 L + \alpha_3}{\alpha_4 L^2 + \alpha L + 1}\, b^2 + \frac{\beta_1 L^2 + \beta_2 L + \beta_3}{\beta_4 L^2 + \beta_5 L + 1}\, b + \frac{\gamma_1 L^2 + \gamma L + \gamma_3}{\gamma_4 L^2 + \gamma L + 1} \quad (4)$$

With $\alpha_i, \beta_i, \gamma_i, i = 1, \dots, 5$ the final parameters.
For the longitudinal modes, we did a log-log transformation then a linear regression on the result, such as $f(L) = L^{\beta_1} e^{\beta_2}$.

5.2 Results

5.2.1 Longitudinal modes

The relative error mentioned (1) is used to explore the function's accuracy and validate the model. We deduct some statistic information of its tendency, such as the mean above the values, the standard deviation around it and maximum and minimum value of the error data. This procedure is kept for all the results in the study.
The mentioned data are presented in **Erreur ! Source du renvoi introuvable.**.

	Rectangular beams			**Circular beams**		
	F1Z	F2Z	F3Z	F1Z	F2Z	F3Z
Count	141	141	141	75	75	75
Mean	0.212	0.202	0.198	0.265	0.263	0.296
Standard Deviation	$9.31\ 10^{-2}$	$8.99\ 10^{-2}$	0.138	0.215	0.241	0.295
Min	$7.63\ \ 10^{-2}$	$4.86\ 10^{-2}$	$9.36\ 10^{-3}$	$1.12\ \ 10^{-2}$	$2.01\ 10^{-3}$	$1.53\ 10^{-3}$
Max	0.414	0.432	0.608	0.843	0.973	1.42

Table 6: Error (%) for longitudinal modes with circular beams during the study's first step

We can already be satisfied by the error which is less than 1%, except for the third longitudinal modes with the circular beams, where we notice a peak at 1.42%. All the means are around 0.2% of error which is promising for the efficiency of the method used.

We also represent the error depending on the ratio, to enhance the influence of ratio as major error vector. The results are represented in the Figure 4.

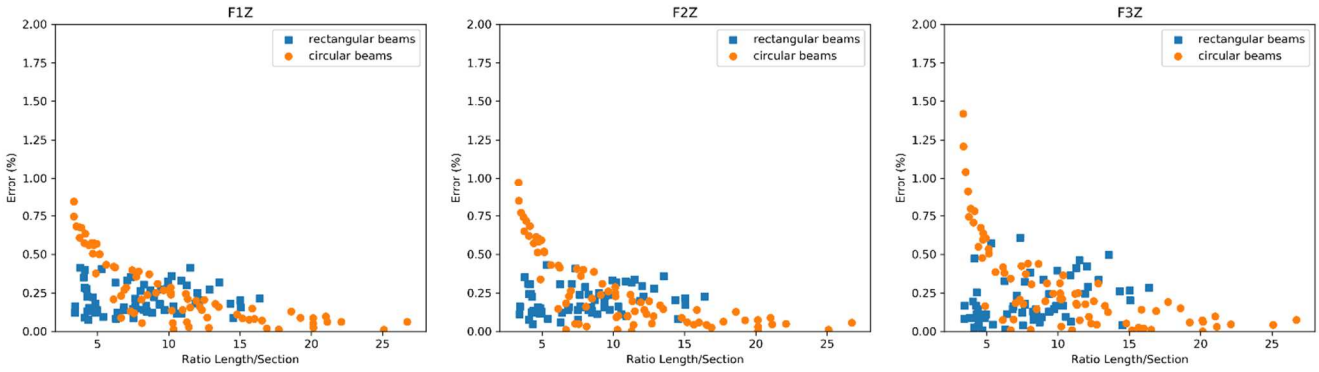


Figure 4: Errors in function the ratio for longitudinal modes during the study's first step

The error has a dependency with the ratio which increases when this ratio decreases. The difference of behavior between circular and rectangular beams is due to the difference of the grid's construction used.

5.2.2 Transverse modes

Table 7: Error (%) for transverse modes with rectangular beams during the study's first step and Table 8*: Error (%) for transverse modes with circular beams during the study's first step* contain the statistic of the error on the transverse modes. We still have an error less than 1% except a peak for the third modes which results of its higher complexity. The means tends to evolve around 0.15% which is still promising.

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	141	141	141	141	141	141
Mean	$9.35\ 10^{-2}$	0.128	0.126	0.100	0.141	0.139
Std. Dev.	$6.76\ 10^{-2}$	$8.58\ 10^{-2}$	$8.98\ 10^{-2}$	$6.55\ 10^{-2}$	$8.53\ 10^{-2}$	$8.31\ 10^{-2}$
Min	$1.04\ \ 10^{-3}$	$2.44\ 10^{-3}$	$1.45\ 10^{-3}$	$5.45\ 10^{-3}$	$6.17\ 10^{-3}$	$5.08\ 10^{-4}$
Max	0.264	0.323	0.305	0.255	0.324	0.344

Table 7: Error (%) for transverse modes with rectangular beams during the study's first step

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	75	75	75	75	75	75
Mean	0.122	0.160	0.185	0.122	0.153	0.187
Std. Dev.	0.130	0.153	0.187	0.130	0.153	0.187
Min	$9.52\ 10^{-3}$	$1.21\ 10^{-3}$	$2.92\ 10^{-3}$	$9.52\ 10^{-3}$	$1.21\ 10^{-3}$	$2.92\ 10^{-3}$
Max	0.554	0.877	1.23	0.554	0.877	1.23

Table 8: Error (%) for transverse modes with circular beams during the study's first step

We have globally the same accuracy as for the longitudinal modes, the distribution is wider which can be explained by the more complex behavior of these quantities. The use of 2 layers here can also add some deviation for some values.

We have a good accuracy, even if we have some peaks for small ratio which confirming the changing behavior of this area as shown Figure 5.

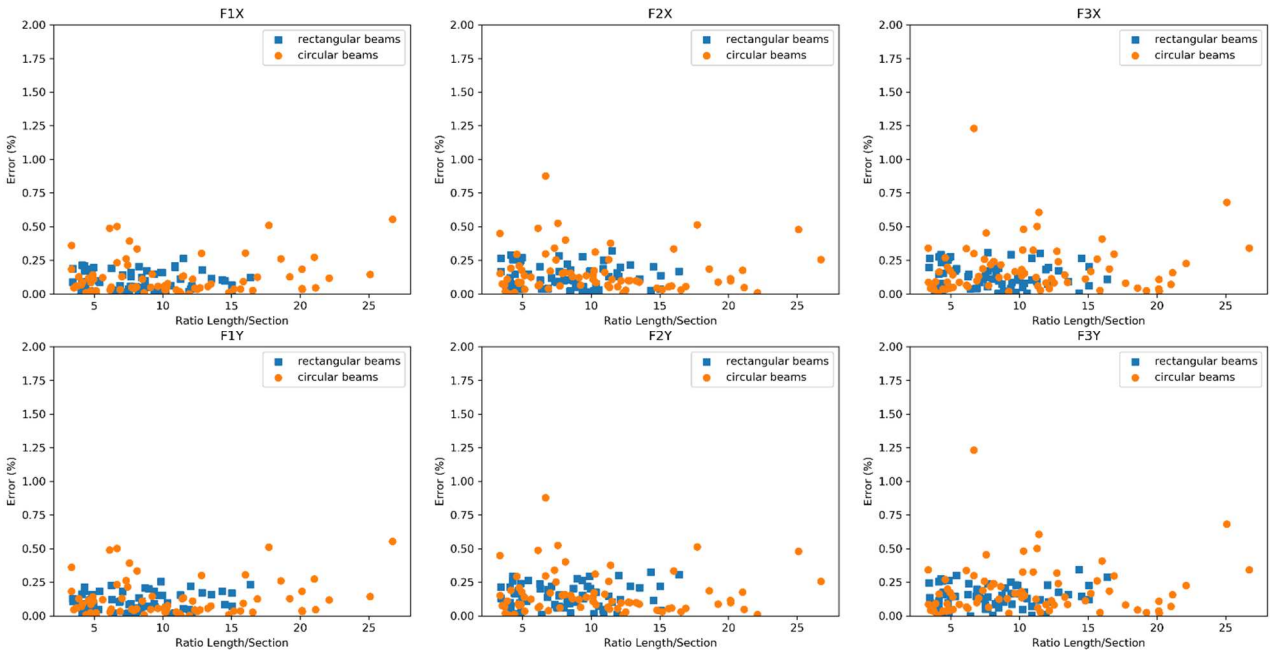


Figure 5: Error in function of the ratio for transverse modes during the study's first step

5.3 Critics and way of improvement

Even if the accuracy could be higher, we have satisfying results for this first meta-model, which gives us 1% error on our generalization data. This results is satisfying for the model. As we are in a problem with maximum two dimensions, having only one feature to fit the longitudinal works well in this specific case.

Moreover, we want to include in the second step the material properties of the beam, so we will have to be more flexible on the complexity of our regression function, even for the longitudinal modes. A more specific way of determining the training samples has to be built, so that we minimize its size. The next step will be to set a guideline to choose an initial subset and optimize the number of samples needed to have a great accuracy.

6 Second step

6.1 Strategy

As we have two new features involved, another grid needs to be made. We want to minimize the number of samples, so the first idea is to keep as much variance as possible in the domain of study, while maintaining the properties of a grid. Nevertheless, during the study, we set a constrain on the ratio, where for a beam too close of a cube, we are not sure to find a third longitudinal mode, due to its geometry. To respect this constrain, instead of taken the length between 2 and 9 meters, we set its boundaries between 3.33 and 9 meters.

The optimization is mainly done by the number of values we choose for each feature. Thanks to theoretical properties, we know that the eigenfrequencies are linearly dependent with Young's modulus and Poisson ratio [6]. So for the layers associated to material features, we needed two values each to make a linear regression. Then for the other features, we want to approximate data with curves which have a certain complexity, so we decided to choose five values as our beginning training set. The five values will allow us to choose function including five parameters and this gives us a first flexibility on function complexity.

We took the same comments as the previous step, where only six modes are approximated and the height is not included as a relevant feature for both the transverse and longitudinal modes. Moreover, for the rectangular beams, we generate randomly the height always strictly lower than the thickness, so that we avoid redundancy in the training data.

We also improve the custom method, now for each layer and for each parameters, the function to interpolate can be chosen. This increases a lot the complexity of the final function and allows us to fit more precisely our data and have a visualization of the fitting at each step of the process. We chose among the following functions to adjust the data:

- Linear : $f(x) = \beta_1 x + \beta_2$
- Polynomial of degree 2 : $f(x) = \beta_1 x^2 + \beta_2 x + \beta_3$
- Polynomial of degree 3 : $f(x) = \beta_1 x^3 + \beta_2 x^2 + \beta_3 x + \beta_4$
- Rational function of first order : $f(x) = \frac{\beta_1 x + \beta_2}{\beta_3 x + 1}$
- Rational function of second order : $f(x) = \frac{\beta_1 x^2 + \beta_2 x + \beta_3}{\beta_4 x^2 + \beta_5 x + 1}$
- Exponential : $f(x) = \beta_1 e^{\beta_2 x}$

With $\beta_i, i = 1, \dots, 5$ the parameters to adjust.

6.2 Results

6.2.1 Longitudinal modes

In the Table 9*: Error (%) for longitudinal modes with rectangular beams during the study's second step*, we present the results of longitudinal modes for the second step of the study.

	Rectangular beams			Circular beams		
	F1Z	F2Z	F3Z	F1Z	F2Z	F3Z
Count	50	50	50	50	50	50
Mean	0.378	0.412	0.582	0.215	0.278	0.351
Standard Deviation	0.138	0.223	0.442	0.221	0.263	0.291
Min	$1.62\ 10^{-2}$	$6.31\ 10^{-3}$	$4.20\ 10^{-4}$	$5.03\ 10^{-4}$	$6.00\ 10^{-5}$	$8.71\ 10^{-4}$
Max	0.570	0.954	2.30	0.565	0.763	0.886

Table 9: Error (%) for longitudinal modes with rectangular beams during the study's second step

The error is slightly higher than the first step and increases with the order of the mode, but the means are under 0.6% of error, which is a promising accuracy for this initial grid.

The Figure 6: Error in function of the ratio for longitudinal modes during the study's second presents

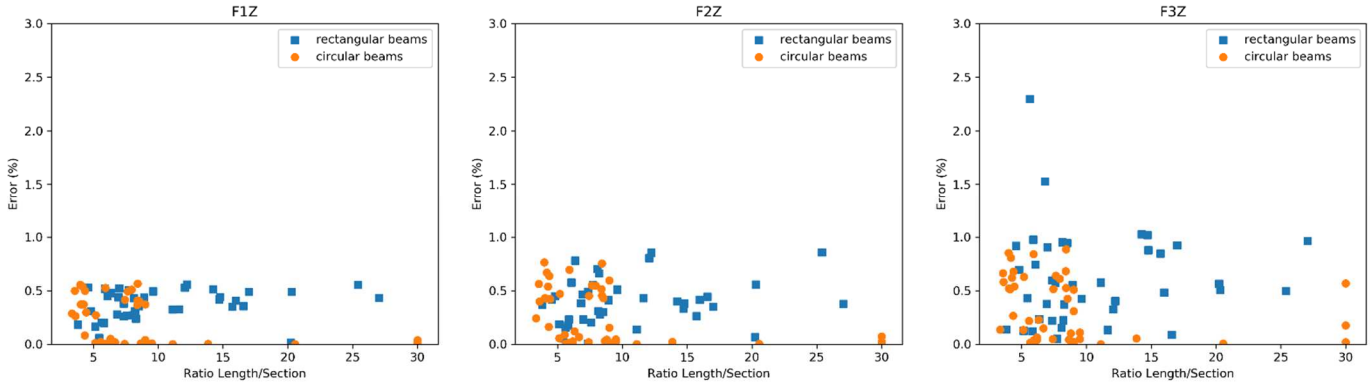


Figure 6: Error in function of the ratio for longitudinal modes during the study's second

the error depending on the ratio of the beams.
The rise of the error depending on the mode's order is more prominent here, with still some perturbations for small ratio.

6.2.2 Transverse modes

In the Table 10*: Error (%) for transverse modes with rectangular beams during the study's second step* and

Table 11: Error (%) for transverse modes with circular beams during the study's second step, we present the results of transverse modes for the second step of the study.

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	50	50	50	50	50	50
Mean	0.430	0.487	0.512	0.399	0.678	0.829
Standard Deviation	0.203	0.250	0.379	0.166	0.494	0.623
Min	$1.93\ 10^{-2}$	$6.66\ 10^{-2}$	$2.76\ 10^{-2}$	$3.79\ 10^{-2}$	$1.45\ 10^{-2}$	$6.23\ 10^{-1}$
Max	1.05	1.24	1.92	0.837	2.48	3.03

Table 10: Error (%) for transverse modes with rectangular beams during the study's second step

	F1X	F2X	F3X	F1Y	F2Y	F3Y
Count	50	50	50	50	50	50
Mean	0.186	0.309	1.36	0.189	0.309	1.36
Standard Deviation	0.173	0.472	1.00	0.172	0.470	1.00
Min	$4.02\ 10^{-3}$	$1.26\ 10^{-4}$	$2.73\ 10^{-2}$	$4.02\ 10^{-3}$	$1.26\ 10^{-4}$	$3.73\ 10^{-2}$
Max	0.475	3.25	3.64	0.475	3.24	3.63

Table 11: Error (%) for transverse modes with circular beams during the study's second step

The same remark as the previous section can be made, the error is higher than the first step but still acceptable, considering the number of training data.

In the Figure 7*:* Error in function of the ratio *for transverse modes during the study's second*, the graphs show the limits of the method and the grid, where we have generalization issues, especially for small ratios and third modes.

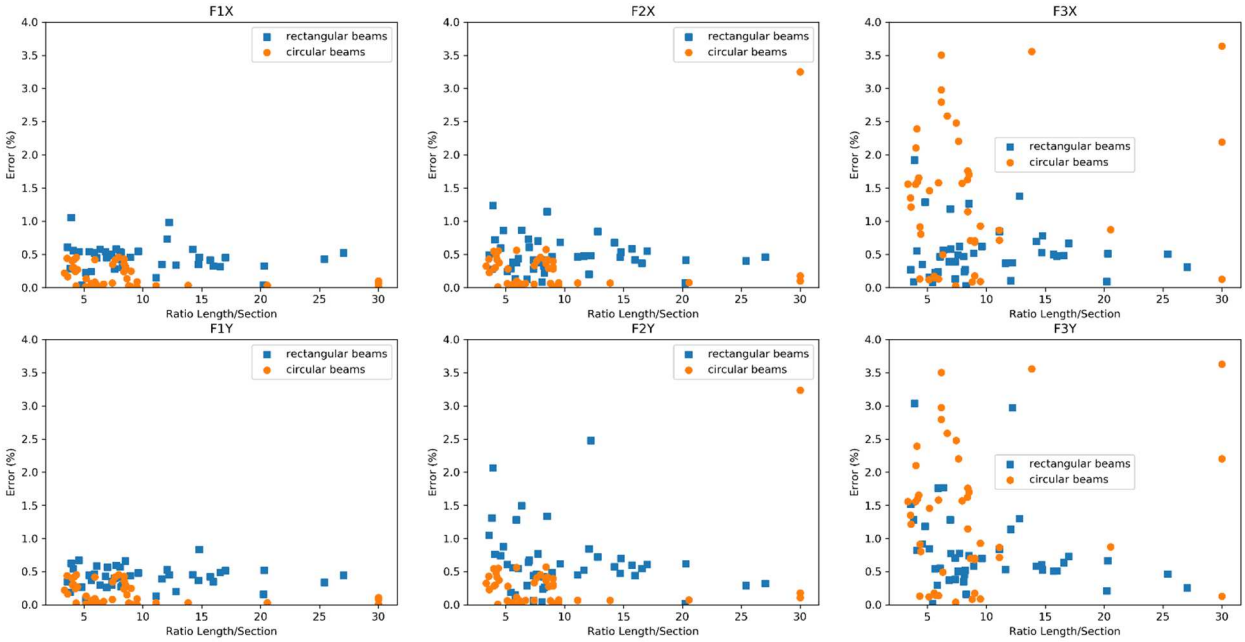


Figure 7: Error in function of the ratio for transverse modes during the study's second

6.3 Critics and ways of improvement

The error increases but is still acceptable. Moreover, we have to consider the context of the learning which is more complex than the previous step. The error is twice higher than the first step, under 3% error of accuracy. It is worth remembering that we added two features and decreases the training data from 512 to 100 samples.

A first improvement would be to stabilize generalization's error. To do so, we will propose a guideline to select leading samples to add in the training data which will increase the accuracy. The criteria to build the initial grid can also be improved. Indeed, we didn't train the model with length under 3.33

meters and it impacts the generalization as we see in the results. A new methodology could spare the lack information in this ratio's area.

6.4 Guideline for sample selection

We propose a first guideline to increase the precision by adding some relevant samples. The initialization starts with the construction of a first grid, with a minimum number of samples. Here it will depends on the nature of the features and the domain of study. If some follows a specific tendency, then we know the minimum number of requisite values to approach them.
In order to get a satisfying learning, we need to have a uniform representation of the response on its boundaries. Then, with the modelling function learnt with the initial grid, the purpose is to find the missing configuration with that function, to restore a uniformly represented response.
We can either use directly the reciprocal function of our final regression function, or generate a set of values with that function and determine the closest ones to our wanted configuration. Once we have all the inputs needed, we have to adjust those values to a grid which will be added to the initial grid and used in the next iteration.

7 Discussion and perspectives

The first aim of this study was to build a model to obtained directly the 9 first eigenfrequencies of interest of a circular or rectangular beam in order to lately integrate this model in a larger loop which purpose is to classify carbon blocks whether there are susceptible to contained cracks or not, and this the more precisely possible. This has been achieved with success.
The second aim was to investigate the concept of Machine Teaching in the sense that we were searching for guidelines to minimize the number of input data required conserving a given accuracy goal. To illustrate the idea, we can think about Figure 1, and more particularly at the link between the quality and the quantity of data. Indeed, by structuring cleaver the distribution of the input data taking into account the produced output data, we strongly believe that we can reach higher levels of regression or classification ability while minimizing the number of initial computations required to feed the model. That is of first importance when the computation time last hours or days.
This study based on a simple model permitted to not be hampered by too long computational times in order to find the optimal strategy. The further work will consist in applying those guidelines on a more complex problem and to implement it for optimization purposes.

References

[1] Méndez-Sánchez, R. A., Morales, A., & Flores, J. (2005). Experimental check on the accuracy of Timoshenko's beam theory. *Journal of Sound Vibration*, *279*, 508-512.
[2] Mokhtari, A., Sarvestan, V., Ghayour, M., & Mirdamadi, H. R. Free vibration of undamped Euler-Bernoulli and Timoshenko beams by using spectral finite element method for different boundary conditions.
[3]Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of machine learning research*, *12*(Oct), 2825-2830.
[4] Qi, Y. (2012). Random forest for bioinformatics. In *Ensemble machine learning* (pp. 307-323). Springer, Boston, MA.
[5] Kenneth Levenberg (1944). A Method for the Solution of Certain Non-Linear Problems in Least Squares.
[6] Michel Lalanne, Patrick Berthier, Johan Der Hagopian. (1984). Mechanical vibrations for engineers. *J. Wiley.*

A Appendix

The grid used is built with the following pattern:

$$
\begin{aligned}
&choose\ m_1\ points\ in\ I_1, such\ as\ , \left\{v_1^{(1)}, \dots, v_{m_1}^{(1)}\right\} \\
&\qquad\qquad\qquad\vdots \\
&choose\ m_N\ points\ in\ I_N, such\ as\ , \left\{v_1^{(N)}, \dots, v_{m_N}^{(N)}\right\} \\
&for\ i_1 = 1, \dots, m_1 : \\
&\quad \ddots \\
&\qquad for\ i_N = 1, \dots, m_N: \\
&\qquad\quad Add\ \ (v_{i_1}^{(1)}, \dots, v_{i_{N_1}}^{(N)})
\end{aligned}
$$

This kind of combination leads us to $\prod_{i=1}^{N} m_i$ samples. In this way of data generation, we can explore a domain study with a chosen discretization, which can be useful in our case. A second benefit of the grid, is the possibility to observe the response depending on only one feature, where all the others are fixed. We will use this ability for the visualization.

We start by unrolling the grid's construction operations, by fixing each feature, then we have a 2D graph with the response associated to the set $\left\{v_1^{(N)}, \dots, v_{m_N}^{(N)}\right\}$. Then, we interpolate the graph by a chosen function, with the nonlinear least squares algorithm, which gives us the best parameters to fit the interested values. After repeating this curve fitting for each configuration on the last loop, we interpolate the parameters of the chosen function with the set $\left\{v_1^{(N-1)}, \dots, v_{m_{N-1}}^{(N-1)}\right\}$. The process is done when all the features were involved in the learning process.

We can illustrate that decomposition with an example. Let's consider a problem with three real input features, *A*, *B* and *C,* and *F* as a real output feature. In order to make a grid, we choose arbitrary two points for each input, and by applying the previous building method, we get the following combinations (Table).

A	**B**	**C**	**F**
$\boldsymbol{A_1}$	B_1	C_1	F_1
$\boldsymbol{A_1}$	B_1	C_2	F_2
$\boldsymbol{A_1}$	B_2	C_1	F_3
$\boldsymbol{A_1}$	B_2	C_2	F_4
$\boldsymbol{A_2}$	B_1	C_1	F_5
$\boldsymbol{A_2}$	B_1	C_2	F_6
$\boldsymbol{A_2}$	B_2	C_1	F_7
$\boldsymbol{A_2}$	B_2	C_2	F_8

Table 12: One grid example

Then, we can apply the decomposition, which can be illustrated with a graph, like the one in Figure 8. The set of nodes corresponding to one feature is a step of the algorithm, which can also be described as a feature layer. Only the last layer will proceed the response, corresponding to the first step of the method. Then all the layers will deal with the parameters of the function fitted in the layer before.

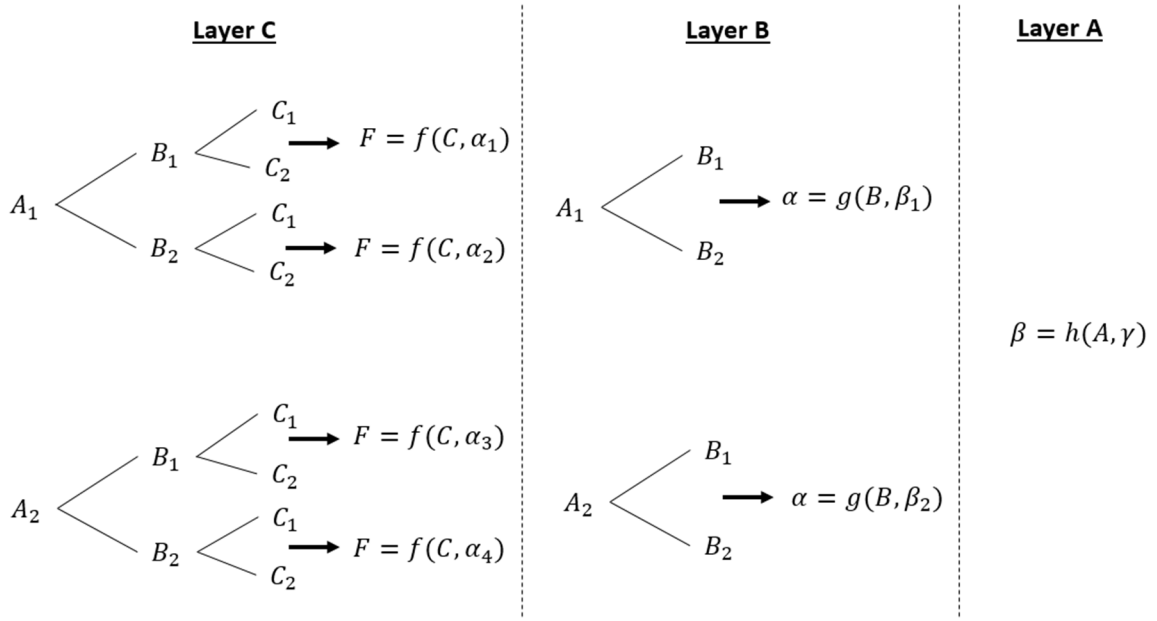


Figure 8: Graph of each layer of the example

The purpose is to manually select the best fitted function among a lot of possibilities, in terms of intrinsic properties such as convexity, asymptotic behavior or the level of complexity. The construction of the final modelling function is done by a composition of all the functions involved during each feature's layer. In our example, the final regression function would be:

$$\tilde{F}\ (A, B, C) = f(C, g\big(B, h(A, \gamma)\big)) \tag{5}$$

The interest of this method, is firstly the high complexity of the final function, often with several input features, where we adjust several functions on a 2D graph. It also decomposed the training phase in several steps, where a human can control the level of accuracy. The possibility of choosing the modelling function at each layer for each parameter, depending of the tendency followed by the values is a determinant way to obtain maximum precision with a highly restricted data set.

Effects of Particle Swarm Optimization Algorithm Parameters for Structural Dynamic Monitoring of Cantilever Beam

Xiao-Lin LI[1,2], Roger SERRA[1,2] and Julien OLIVIER[1,3]

[1] Institut National des Sciences Appliquées Centre Val de Loire, Blois France,
[2] Laboratoire de Mécanique Gabriel Lamé - EA 7494, Blois, France
[3] Laboratoire d'Informatique Fondamentale et Appliquée de Tours, EA 6300, Blois, France

Abstract

Nowadays, particle swarm optimization (PSO) algorithm has become a widespread optimization method. However, it is well known that its main parameters (inertia weight, two learning factors, velocity constraint and population size) have a critical effect on its performance. Currently the effects of PSO parameters on structural health monitoring have not been comprehensively studied. Therefore, in this paper, the PSO algorithm is used for damage detection assessment of a cantilever beam, and the simulation results are used to analyze the effects of PSO parameters. There are five levels for each parameter in our experiment, mean fitness value and success rate for each level are used as criteria to measure the convergence and stability of the PSO algorithm. Considering the effect of population size on CPU time, a trade-off strategy is presented to further determine the selection of population size.

1 Introduction

Over the last few years, there have been increasing demands to develop structural dynamic monitoring system over different kinds of aerospace, mechanical and civil engineering structures because of the huge economic and life-safety benefits. Vibration testing is the widest used method for structural damage detection[1, 2]. The main idea behind damage detection techniques based on structural dynamic changes is the fact that the modal parameters of a structure are functions of the physical parameters (such as mass, stiffness and damping) thus the existence of damage leads to changes in the modal properties of the structure. The inverse method for damage detection using vibration data and solving by optimization algorithms have received extensive attention in recent years. The usual approach is to minimize an objective function, which is defined in terms of discrepancies between the predicted model parameters and the initial model parameters. Using classic optimization methods to solve it often meet some difficulties. However, the particle swarm optimization (PSO) algorithm can be used on complicated optimization problems that cannot be expressed explicitly. PSO algorithm is one of the newest intelligent method, this parallel evolutionary computation technique was developed by Kennedy and Eberhart in 1995[3].The basic idea comes from the study of group behaviors such as predation of birds. PSO algorithm was first used for function optimization and neural network training[3]. Since the algorithm has many advantages such as comparative simplicity, easy to implement and few parameters to be adjusted, PSO has found its application in many complex engineering optimization problems, including structural damage detection of beam structure[4, 5].

It is well known that in various optimization methods, parameters is one of the key factors which have a great effect on the performance. For different kinds of optimization problems, the matching and cooperation modes between parameters are different. Even for the same type of optimization problem, if problem scales are different, parameter selections are not completely the same. Although PSO algorithm has few parameters to adjust, how to determine them is also an important problem. However, currently the effects of PSO parameters on structural damage detection have not been comprehensively studied. In this paper, the damage detection of a cantilever beam by PSO algorithm for two damage patterns are simulated, and the experimental results are analyzed to study the effects of PSO parameters. There are five levels for each parameter in our experiment, the mean fitness value and success rate for each level are used as criteria to measure the convergence and stability

of the PSO algorithm. Considering the effect of population size on CPU time, a *Ratio* compromise strategy is proposed to further determine the selection of population size.

2 Problem formulation

2.1 Structural dynamic finite element formulation

The governing equation for an Euler-Bernoulli beam with negligible damping is given by:

$$\frac{\partial^2}{\partial x^2}\left[EI(x)\frac{\partial^2 \omega(x,t)}{\partial x^2}\right] + m(x)\frac{\partial^2 \omega(x,t)}{\partial t^2} = f(x,t) \tag{1}$$

where $EI(x)$ denotes the flexural stiffness, $m(x)$ denotes the mass per unit length of the beam, $w(x,t)$ represents the transverse displacement of the beam. The beam is discretized into a number of elements, with displacement and slope as nodal degrees of freedom and cubic interpolation function.For an n-degree of freedom system of uniform beam, the stiffness matrix and the consistence mass matrix are given in [6], respectively.

The characteristics of the beam are given in the table below. The total mass is $3.237\ kg$.

Young modulus E (Pa)	Poisson ratio ν	Density ρ $(Kg.m^{-3})$	Length L (m)	Width b (m)	Thickness h (m)
2e+11	0.33	7850	1	5.3e-3	2.5e-2

Table 1: Beam properties

For a properly modeled structure, the structural dynamic eigenvalue equation is given by:

$$(K - \lambda M)\Phi = 0 \tag{2}$$

where K and M are the global stiffness and mass matrices respectively, and λ and Φ represent the natural frequency and vibration mode shape vectors. It can be assumed that when a structural modification occurs, the local stiffness of the structure changes whereas the change in mass may be neglected. Hence, the equation (2) could be rewritten for a modified system as:

$$(K_d - \lambda_d M)\Phi_d = 0 \tag{3}$$

where K_d and M are the global modified stiffness and mass matrices respectively, and λ_d and Φ_d represent the new natural frequency and vibration mode shape vectors for the modified structure respectively. In many studies on structural health monitoring, the structural modification has been simulated by decreasing one of the local element 's stiffness parameters like a inertia moment I, cross sectional area S or elasticity (Young) modulus E. In this work, the structural modification of each element was simulated using Young modulus reduction factor (also called stiffness reduction factor) x as a scalar variable between $[0,1]$ where zero value corresponds to no modification and a value near to one corresponds to failure condition introduced as follows[5]:

$$x_i = \frac{E - E_i}{E} \tag{4}$$

where E is the initial Young modulus and E_i is the reduced Young modulus of the i^{th} element. In this case, the stiffness matrix K will be modified as:

$$K_d = \sum_i (1 - x_i) K_i \tag{5}$$

2.2 Optimization problem

Equation (3) forms the basis of the structural modification detection method through an inverse procedure giving the new vibration natural frequencies and the mode shapes. As the structural modification causes change in vibration natural frequencies and which are easier to measure than mode shapes (limited number of

accelerometers) and the error associated is comparatively less. Hence, they are used as structural modification indicators in this study. PSO algorithm is used to search a particular stiffness reduction factor x so that the predicted numerical natural frequencies exactly match with the initial natural frequencies. The initial numerical model of the structure is generally considered for the optimization. When the exact match between modified and initial natural frequencies is observed, the value of stiffness reduction factor represents the actual modification location and amount. The usual approach to solve the inverse problem of structural dynamic monitoring involves minimization of the fitness function (or objective function) which is defined in terms of discrepancies between the predicted natural frequencies and the initial natural frequencies. In this study, the fitness function can be presented just like in literature [7]:

$$F = \sum_{s=1}^{n} \left(\frac{(f_s^r)^2 - (f_s^p)^2}{(f_s^r)^2} \right)^2 \tag{6}$$

where f_s^r and f_s^p are the initial and predicted natural frequencies respectively. n is the number of input response parameters chosen (natural frequencies) and for this study is taken as five.

2.3 Particle swarm optimization algorithm

The particle swarm optimization technique is a population based stochastic technique in nature (bio-inspired) so-called evolutionary computational model which is based on swarm intelligence. PSO is developed by Kennedy and Elberhart [3] and primarily used to tackle continuous optimization problems. The system is initialized firstly in a set of randomly generated potential solutions, and then performs the search for the optimum one iteratively by swarms following the best particle. Compared to others evolutionary algorithms, PSO has much more profound intelligence background and could be performed more easily. Based on its advantages, the PSO is suitable for engineering applications, in the fields of evolutionary computing, optimization and many others.

As suggested in literature [8], a fully connected topology is elected as PSO algorithm topology. Set for the D-dimensional search space, mth particles compose a population $\{X_1, X_2, \ldots, X_m\} \subset R^n$, and the ith particle position is $X_i = (x_{i1}, x_{i2}, \ldots, x_{iD})^T$, the velocity of the particle X_i can be represented by another D-dimensional vector $V_i = (v_{i1}, v_{i2}, \ldots, v_{iD})^T$. The best position previously visited of the particle X_i is denoted as $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})^T$, and the best position among all particles in the population is $P_g = (p_{i1}, p_{i2}, \ldots, p_{iD})^T$. Each particle adjusts its position dynamically according to the principle of following the current optimal particle, the particle X_i updates its speed and position according to (7) and (8).

$$v_{id}^{t+1} = \omega\, v_{id}^t + c_1\, r_1 (p_{id}^t - x_{id}^t) + c_2\, r_2 (p_{gd}^t - x_{id}^t) \tag{7}$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \tag{8}$$

where t is iteration time, d represents the dimension of the particle, $d = 1, 2, \ldots, D$, i represents the number of the particle, $i = 1, 2, \ldots, m$. r_1, r_2 are random between 0 and 1, ω is the inertia weight, c_1 and c_2 are the learning factors to adjust each iteration step length.

many literatures has emphasized the importance of ω, the linear decreasing inertia weight ω^t has been widely used[9], which is defined as follow:

$$\omega^t = \omega_{max} - \frac{\omega_{max} - \omega_{min}}{t_{max}} \times t \tag{9}$$

where t_{max} is the maximum number of iteration. In this strategy, ω^t changes with iteration. At beginning, the value of ω^t is ω^{max} , and ω^t decrease during the execution of the algorithm. At the end the value of ω^t is ω_{min}. ω_{max} and ω_{min} are set to 0.9 and 0.4 respectively.

3 Analysis of parameters on the algorithm performance

In the experiment, a steel cantilever beam is considered for structural damage detection. Figure 1 shows the sketch of the beam with element number using in the finite element simulations, 30 equal Euler-Bernoulli beam elements are chosen for finite element modeling.

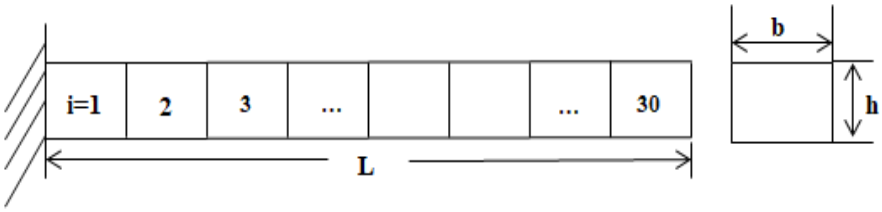


Figure 1: The Euler-Bernoulli cantilever beam model

PSO algorithm has some important parameters, such as population size (m), inertia weight (ω), two learning factors (c_1, c_2) and maximum velocity $(Vmax)$. The effects of PSO parameters are analyzed by simulating structural damage detection of a cantilever beam. 200 iterations and 1000 runs are set for two damage patterns in Table 2.

Damage Pattern I		Damage Pattern II	
Element	Damage(%)	Element	Damage(%)
5	10	5	10
		7	10

Table 2: Simulated damage patterns in cantilever beam

Table 3 shows the parameter setting. The value of $Vmax$ is set to $Vmax = \gamma Xmax$ and $\gamma \in (0,1]$, $Xmax$ denotes the dynamic range of the variable on each dimension. In each test, there are five levels for each parameter, mean fitness value is equal to the average value of fitness function F, success rate is equal to the ratio of the number of successful runs to total number of runs, which are all considered as criteria for parameter performance measurement. When fitness function provides smaller mean fitness value it shows better convergence performance of the algorithm. The higher success rate provides the stronger stability of the algorithm. The convergence and stability for each parameter at five levels are compared.

Table 3: parameter setting for five levels

level	population size m	inertia weight ω	learning factors (c_1, c_2)	maximum velocity $Vmax$
1	20	0.25	$(0,4)$	$0.2Xmax$
2	40	0.5	$(1,3)$	$0.4Xmax$
3	60	0.75	$(2,2)$	$0.6Xmax$
4	80	1.0	$(3,1)$	$0.8Xmax$
5	100	ω^t	$(4,0)$	$Xmax$

3.1 Effect of population size m

The choice of population size is related to the complexity of the problem. As the complexity of the problem increases, the population size also grows. The five levels of population size are given in Table 3, the other parameters are chosen as $\omega = \omega^t, c_1 = c_2 = 2, Vmax = Xmax$ [9]. Mean fitness value and success rate for two damage patterns are shown in Figure 2. It is clearly that along with the increase of population size m, the convergence and stability of the PSO algorithm is becoming stronger and stronger.

However, for a given problem, the parameters that affect CPU time are mainly finite element number, population size, maximum iterations and number of PSO runs, they are given except for population size. Therefore, when the effect of population size are analyzed, CPU time needs to be taken into consideration. The larger the population size represents the longer CPU time. Then, a trade-off strategy (10) is proposed as a criterion to further determine the population size m. Obviously, a larger $Ratio$ means a better performance with selected population size. Thus, it can be seen in Figure 3 that for two damage patterns,the optimal choices are $m = 20$

and $m = 60$,respectively.

$$Ratio = \frac{\text{success rate}}{\text{population size}} \tag{10}$$

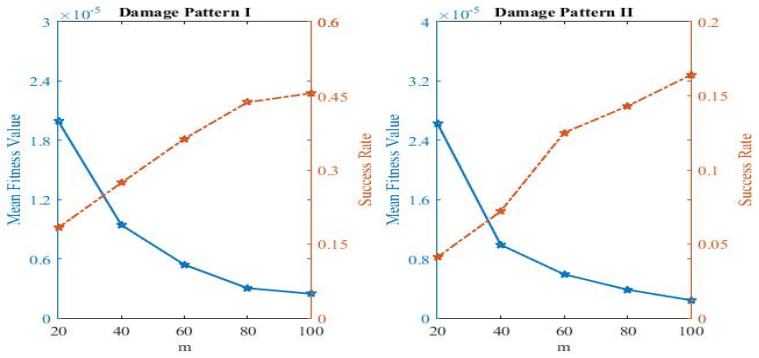


Figure 2: Effect of population size on PSO algorithm for two damage patterns

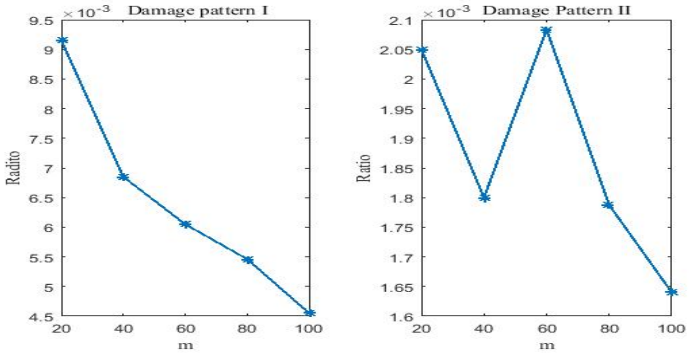


Figure 3: Selection of population size for two damage patterns

3.2 Effect of inertia weight ω

The inertia weight ω affects the particle's global and local search ability. The stochastic process theory in [10] shows that the range of ω is $[0,1]$. From the above,the best setting for population size are: $m = 20$ for damage pattern I, $m = 60$ for damage pattern II. And $c_1 = c_2 = 2, Vmax = Xmax$. In order to examine the balance between global and local exploration, the five levels of inertia weight are compared, and the simulation results are shown in Figure 4.

When ω is small, PSO algorithm hardly converges and the success rate is low. Along with the increase of ω, PSO algorithm has a better convergence and stability. Although a better convergence and stability can also been obtained when ω takes linear-decreasing strategy, the optimal choice is $\omega = 1$ which is different from the general linear-decreasing strategy.

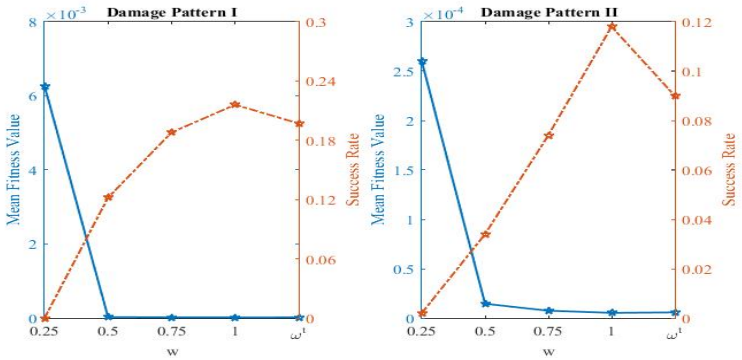


Figure 4: Effect of inertia weight on PSO algorithm for two damage patterns

3.3 Effect of learning factors c_1, c_2

A general rule for setting the two learning factors is $c_1 + c_2 <= 4$, there is a close relationship between c_1 and c_2. Therefore, the values of the two learning factors are considered simultaneously. Considering the simplicity of the experiment, the following relation between c_1 and c_2 will be used:$c_1 + c_2 = 4$, the five levels of the learning factors are compared with $\omega = 1, Vmax = Xmax$, $m = 20$ for damage pattern I, $m = 60$ for Damage Pattern II.

From figure 5, it can be seen that $(c_1, c_2) = (3, 1)$ is the optimal choice for the convergence and stability of PSO algorithm. That means, for structural damage detection of the cantilever beam, the algorithm shows a good performance when the population put much more attracted to the best location found by itself.

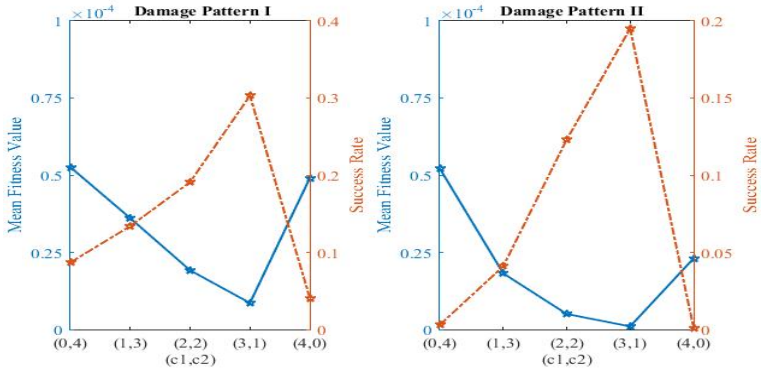


Figure 5: Effect of leaning factors on PSO algorithm for two damage patterns

3.4 Effect of maximum velocity $Vmax$

The velocity of the particles can be limited to $[-Vmax, Vmax]$ by a maximum velocity, which acts as a constraint to control the global exploration capability of the population, It is clearly that $\omega = 1, c_1 = 3, c_2 = 1$,$m = 20$ for damage pattern I,$m = 60$ for damage pattern II are optimal choices for structural damage detection. Then, the convergence and stability of $Vmax$ for the five levels are compared.

The simulation results for mean fitness value and success rate are shown in Figure 6. Along with the increase of the maximum velocity, mean fitness value is decreasing and success rate is increasing, which means the convergence and stability of the algorithm is becoming stronger and stronger. Usually set $Vmax$ as a constant, $Vmax = Xmax$ is the best choice for structural damage detection of the cantilever beam.

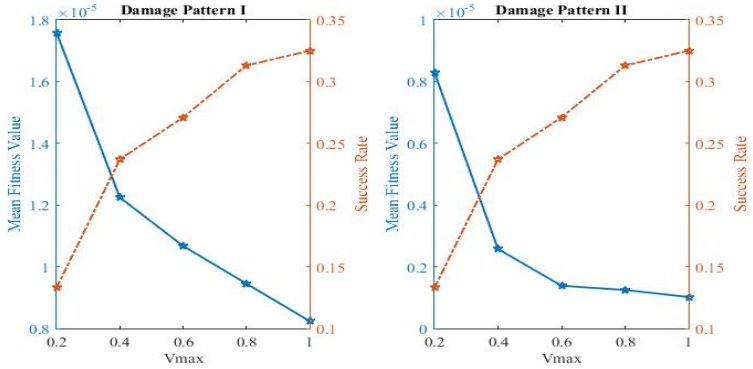


Figure 6: Effect of maximum velocity on PSO algorithm for two damage patterns

4 Validation

From the above paper, the optimal parameter configuration for damage pattern I is $(m, \omega, c_1, c_2, Vmax) = (20, 1, 3, 1, Xmax)$, at this time, mean fitness value $= 8.60e-06$, success rate $= 0.36$. For damage pattern two is $(m, \omega, c_1, c_2, Vmax) = (60, 1, 3, 1, Xmax)$, the corresponding mean fitness value and success rate are $9.78e-07$

and 0.18, respectively. Under the optimal configuration, PSO shows a better convergence and stability than other configurations used in our experiment by using relatively low time costs. Therefore, the effectiveness is verified by simulating structural damage detection of cantilever beam.

5 Conclusion

In this paper, first of all, the five levels for each parameter are designed to perform damage detection of cantilever beam. Then mean fitness value and success rate obtained from simulation results are used as criteria to evaluate the convergence and stability of the algorithm. Considering CPU time, the *Ratio* strategy is proposed to further determine the selection of population size. A parameter guideline are given for structural damage detection of cantilever beam.

References

[1] S. W. Doebling, C. R. Farrar and M. B. Prime, *A Summary Review of Vibration-Based Damage Identification Methods*, The Shock and Vibration Digest, 1998.

[2] Y.J. Yan, L. Cheng, Z.Y. Wu and L.H. Yam, *Development in Vibration-based Structural Damage Detection Technique*, Mechanical Systems and Signal Processing, 2007.

[3] J. Kennedy and R.C. Eberhart, *Particle swarm optimization*, Proceedings of IEEE international conference on neural networks, 1995.

[4] B. Nanda, D. Maity and D. K. Maiti,*Vibration Based Structural Damage Detection Technique using Particle Swarm Optimization with Incremental Swarm Size*, International Journal of Aeronautical and Space Sciences, 2012.

[5] G. Ghodrati Amiri, A. Zare Hosseinzadeh and S.A. Seyed Razzaghi, *Generalized Flexibility-based Model Updating Approach via Democratic Swarm Optimization Algorithm for Structural Damage Prognosis*, International Journal of Optimization in Civil Engineering, 2015.

[6] J. Lee, *Identification of Multiple Cracks in a Beam Using Natural Frequencies*, Journal of Sound and Vibration, 2009.

[7] L. Rubio, J. Fernández-Sáez and A. Morassi. *Identification of Two Cracks with Different Severity in Beams and Rods from Minimal Frequency Data*, Journal of Vibration and Control, 2014.

[8] A. Carlisle and G. Dozier, *An Off-the-shelf PSO Proceedings of the Workshop on Particle Swarm Optimization*,Proceeding of the 2001 Workshop on Particle Swarm Optimization, Indianapolis IN, 2001.

[9] R. Perera, S-E Fang and A Ruiz, *Application of Particle Swarm Optimization and Genetic Algorithms to Multiobjective Damage Identification Inverse Problems with Modelling Errors*, Information Processing Letters, 2007.

[10] M. Jiang, Y. P. Luo and S. Y. Yang, *Stochastic Convergence Analysis and Parameter Selection of the Standard Particle Swarm Optimization Algorithm*, Information Processing Letters, 2007.

List of participants

- Abboud Dany
- Abdenour Soualhi
- Addou Yacine
- Aghazadeh Fatemeh
- Alcorta Roberto
- Andre Hugo
- Antoni Jérôme
- Arruda Jose Roberto
- Assoumane Amadou
- Astolfi Davide
- Audebert Sylvie
- Bachar Lior
- Baguet Sébastien
- Bajric Rusmir
- Barcet Sylvain
- Bareille Olivier
- Baudin Sophie
- Benaicha Youness
- Berger Sébastien
- Bertoni Renaud
- Billon Kevin
- Blanchard Corentin
- Bonnardot Frédéric
- Borghesani Pietro
- Bortman Jacob

- Bouhaddi Noureddine
- Bruand Guillaume
- Camerini Valerio
- Carbonelli Alexandre
- Castellani Francesco
- Cédric Peeters
- Chiementin Xavier
- Cocconcelli Marco
- Combet Francois
- Daga Alessandro Paolo
- Dallali Manel
- Darraz Abdelhakim
- Demore Félix
- Dimitrijevic Zoran
- Dion Jean-Luc
- Dollon Quentin
- Dubey Anurag
- Ege Kerem
- El Badaoui Mohamed
- Eltabach Mario
- Gagnon Martin
- Gao Hanwei
- Garibaldi Luigi
- Garnell Emil
- Gazizulin Dmitri
- Griffaton Julien
- Gryllias Konstantinos
- Guo Yu
- Heyns Stephan
- Hurel Gabriel
- Jaboviste Kévin
- Kehr-Candille Veronique

- Kieu Thinh
- Koechlin Samuel
- Lacaze Florian
- Le Deunf Meryem
- Leaman Felix
- Lerouge Matthieu
- Li Xiaowen
- Li Xiao-Lin
- Liu Chenyu
- Liu Hongcheng
- Lizoul Khalid
- Maamar Asia
- Macchi Lucas
- Manin Lionel
- Marcel Lucie
- Martin Guillaume
- Matania Omri
- Matten Gaël
- Mauricio Alexandre
- Melot Adrien
- Molina Vicuna Cristián
- Naccarato Gianni
- Ompusunggu Agusmian
- Ouisse Morvan
- Paillot Guillaume
- Passos Sébastien
- Perez Matthias
- Pham Quang Hung
- Pichard Géraldine
- Proteau Antoine
- Qi Junyu
- Qin Yi

- Quoc-Cuong Nguyen
- Rubin Eyal
- Sadoulet-Reboul Emeline
- Sanches Leonardo
- Sauvage Olivier
- Schmidt Stephan
- Serra Roger
- Sghaier Emna
- Sierra-Alonso Edgar F.
- Snoun Cherif
- Soave Elia
- Sun Yulan
- Telbani Afaf
- Teloli Rafael
- Thomas Marc
- Thomas Jean-Hugh
- Thomas Xavier
- Topenot Margaux
- Troncossi Marco
- Vouagner Pascal
- Vyas Nalinaksh
- Wang Yuzhu
- Wang Yi
- Weisser Thomas
- Woussen Maurin
- Xiang Sheng
- Xiong Yiwei
- Yoann Hebrard
- Zhang Xin

Author Index