



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

CONFERENCE VIBRATION SURVEILLANCE AND NOISE

Table of contents

Monday	1
Angular approaches	1
Cyclo-non-stationary analysis for bearing fault identification based on instantaneous angular speed estimation, Sierra-Alonso Edgar F. [et al.]	1
Investigation of the Influence of the Operating Parameters on theMagnetic Encoder Geometric Error Compensation, Cakar Halil Ibrahim [et al.]	19
Interpolation of periodic hidden signal measured at steady-operating conditions on hydroelectric turbine runners, Pham Quang Hung [et al.]	28
Adaptive tacho information estimation for surveillance of rotatory machine under nonstationary conditions, Wang Yi [et al.]	40
Diagnostics and Dynamic models	48
Characterization of a Bouc-Wen model-based damper model for automobile comfort simulation, Gao Hanwei [et al.]	48
Modal identification of machining robots, Maamar Asia [et al.]	62
Demodulating of the 3-D tip clearance of turbine blades using BP neural network optimized by genetic algorithm, Liu Hongcheng [et al.]	72
Research on the Variation Mechanism of the 3-D Tip Clearance of a Cracked Blade under Multi-parameters in the Aero-engine Acceleration Process, Xiong Yiwei [et al.]	79
Modeling and identification of mechanical systems	86

The Dynamics of Helicopters with Nonlinearities on the Fuselage, Sanches Leonardo [et al.]	86
Comparison of pseudo-static and modal spectral seismic analyses of motor-driven pump units: is 1.5 security coefficient of pseudo-static method relevant?, Audebert Sylvie [et al.]	97
Characterization of the damping added by a foam on a plate by an inverse vibration problem, Le Deunf Meryem [et al.]	107
Coupled bending torsional vibrations of non-ideal energy source rotors going through critical speeds, Sghaier Emna [et al.]	113
NAFID - A Grid Tool for output only modal analysis, Vu Viet-Hung	119
Study of the static and dynamic behaviour of PU foam: from the material sample to the automotive seat., Blanchard Corentin [et al.]	128
Condition Monitoring	137
Edge computing for advanced vibration signal processing, Helsen Jan [et al.] . . .	137
Remote diagnosis by analyzing heterogeneous data, Guerry Joris [et al.]	145
TOOL CONDITION MONITORING METHOD IN MILLING PROCESS USING WAVELET TRANSFORM AND LONG SHORT-TERM MEMORY, Ag-hazadeh Fatemeh [et al.]	153
Signal processing	162
Autonomous Embedded Vibroacoustic Measurements: an efficient tool for railway monitoring, Clerc Christian [et al.]	162
Blind vibration filtering using envelope spectrum indicators for bearing and gear fault detection without knowledge of machine kinematics, Cédric Peeters [et al.] .	173
Vibration representation in time and phase domains, applications to aircraft engines, Griffaton Julien [et al.]	180
Fault diagnosis and prognosis for rolling bearings	190
Fourier-Bessel series expansion based blind deconvolution method for bearing fault detection, Soave Elia [et al.]	190

Combination of vibration analysis and Acoustic Emission measurements to better characterize damage and mechanical behaviour of aerospace high speed rolling bearing, Yoann Hebrard	203
Neuroevolution for bearing diagnosis, Sleiman Rita [et al.]	213
Electrical motors	226
Long-term vibration monitoring of induction motors in the food industry with low-cost MEMS accelerometers, Ompusunggu Agusmian [et al.]	226
Structural dynamics of electric machines subjected to PWM excitations, Topenot Margaux [et al.]	235
Experimental evidence of MCSA for the diagnosis of ball-bearings, Immovilli Fabio [et al.]	241
Uncertainties, stochastic, robustness	252
A model reduction method to analyze the dynamic behavior of vibrating structures with uncertain parameters, Kieu Duc-Think [et al.]	252
On the monitoring of noisy data as a multidimensional shell, Gagnon Martin [et al.]	267
Confounding factors analysis and compensation for high-speed bearing diagnostics, Daga Alessandro Paolo [et al.]	277
Smart Structures	291
EMBEDDED SENSING MICRO-COMPONENTS FOR FIBRE REINFORCED COMPOSITE MATERIAL SYNTHESIS AND MONITORING, Bareille Olivier [et al.]	291
Vibration Control of Cable-Driven Parallel Robot for 3D Printing, Lacaze Florian [et al.]	297
Tuesday	306
Angular approaches	307

NUMERICAL AND EXPERIMENTAL LOADS ANALYSIS ON A HORIZONTAL- AXIS WIND TURBINE IN YAW, Castellani Francesco [et al.]	307
Gears and Bearings faults Detection: from Instrumentation to Classification, Bertoni Renaud [et al.]	322
Measurement and use of transmission error for diagnostics of gears, Randall Robert [et al.]	323
Rolling bearing diagnosis based on H_infinity filter order tracking, Assoumane Amadou [et al.]	331
Dynamic Characterization Of Hydroelectric Turbine In Transient Using OBMA And Phase-Shift Analysis, Dollon Quentin [et al.]	343
A new method for identifying diagnostic rich frequency bands under varying op- erating conditions, Schmidt Stephan [et al.]	356
Diagnostics and Dynamic models	368
Challenging the multiplicative model used for gear vibration, Hubert Elisa [et al.]	368
Detection sensitivity study of local faults in spur gears based on realistic simula- tions, Bachar Lior [et al.]	379
Towards a better understanding of helical gears vibrations – dynamic model val- idated experimentally, Silverman Nadav [et al.]	382
Modeling and identification of mechanical systems	385
Localization and quantification of damage by frequency based methods : Numeri- cal and Experimental applications on bending vibration beams, Dubey Anurag [et al.]	385
ARX model for experimental vibration analysis of grinding process by flexible manipulator, Nguyen Quoc-Cuong [et al.]	392
Use of virtual sensors for the analysis of forces exerted by the load inside a tum- bling mill, Molina Vicuna Cristián [et al.]	406
Signal processing	428
Comparison and Improvement of Techniques for Transmission-Path Restoring, Mata- nia Omri [et al.]	428

Influence of Gaussian Signal Distribution Error on Random Vibration Fatigue Calculations, Wang Yuzhu [et al.]	432
Fault diagnosis and prognosis for rolling bearings	442
A non-parametric generalization of the synchronous average in the cyclo-non-stationary framework, Abboud Dany [et al.]	442
High Frequency Demodulation Technique for Instantaneous Angular Speed Estimation, Bonnardot Frédéric [et al.]	452
Development of a vibration monitoring strategy based on cyclostationary analysis for the predictive maintenance of helicopter gearbox bearings, Camerini Valerio [et al.]	458
A new indicator designed from the spectral coherence, proposition and application to bearing diagnosis, Kass Souhayb [et al.]	477
Multi band integration on the cyclostationary bivariable methods for bearing diagnostics., Mauricio Alexandre [et al.]	489
Data Mining Classification & Machine Learning methods	502
A Deep Learning Protocol for Condition Monitoring and Fault Identification in a Rotor-Bearing System from raw Time-Domain data, Sonkul Nikhil [et al.] . . .	502
WIND TURBINE GEARBOXES FAULT DETECTION THROUGH ON-SITE MEASUREMENTS AND VIBRATION SIGNAL PROCESSING, Castellani Francesco [et al.]	510
Gears and bearings defaults: from classification to diagnosis using machine learning, Barcet Sylvain [et al.]	520
Vibration Feature for Detecting Eccentric Workpiece/Runout Faults During Continuous Gear Grinding Processes, Ompusunggu Agusmian Partogi [et al.]	532
Toward the quality prognostic of an aircraft engine workpiece in Inconel Alloy 625: case study and proposed system architecture, Proteau Antoine [et al.] . . .	546
Fault prognosis of planetary gearbox using acoustic emission and genetic algorithm: a case study, Leaman Felix [et al.]	561
Rotating machine diagnosis using acoustic imaging and artificial intelligence, Darraz Abdelhakim [et al.]	571

Macroscopic-Microscopic Attention in LSTM Networks based on fusion Features for prediction of bearing remaining life, Qin Yi [et al.]	579
Milling diagnosis using machine learning approaches, Knittel Dominique [et al.] .	587
Passive control of vibrations	597
Experimental identification of the corrective effect of a non-circular pulley : application to timing belt drive dynamics, Passos Sébastien [et al.]	597
Robust optimization of nonlinear energy sinks used for dynamic instabilities mitigation of an uncertain friction system, Snoun Cherif [et al.]	608
Energy exchange between a nonlinear absorber and a pendulum under parametric excitation, Hurel Gabriel [et al.]	618
Smart Structures	625
Hybrid crankshaft control for the reduction of torsional vibrations and rotational irregularities, Paillot Guillaume [et al.]	625
SEMI-ACTIVE TORSIONAL VIBRATIONS CONTROL OF A ROTOR USING A SMART ELECTRO-RHEOLOGICAL DYNAMIC ABSORBER, Sun Yulan [et al.]	636
Shunted piezoelectrical flexensionnal suspension for vibration insulation, Billon Kevin [et al.]	648
Exploring periodicity and dispersion diagrams in muffler design, Lima Vinicius D. [et al.]	652
Wednesday	660
Angular approaches	661
Angular vibration on-site measurements and application to torsional analysis on industrial cases, Combet Francois [et al.]	661
Towards the use of hybrid models for diagnosis and prognosis in turbomachinery health management, Heyns Stephan [et al.]	671
Condition Monitoring	681

CMBBase, a universal gateway to condition monitoring datasets, Capdessus Cécile [et al.]	681
Experimental investigation of sensor mounting positions for localized faults detection of epicyclic gear sets, Guo Yu [et al.]	690
Towards 3D AFM Using Multiple Vibration Modes, Rubin Eyal [et al.]	697
Fault diagnosis and prognosis for rolling bearings	706
Early bearing defect detection in a noisy environment based on a method combining singular value decomposition and empirical mode decomposition, Kedadouche Mourad [et al.]	706
Prognostics of rolling element bearings based on Entropy indicators and Particle Filtering, Qi Junyu [et al.]	722
Spall Evolution in a Rolling Element Bearing, Gazizulin Dmitri [et al.]	735
Data Mining Classification & Machine Learning methods	740
Multi-label fault diagnosis based on Convolutional Neural Network and Cyclic Spectral Coherence, Chen Zhuyun [et al.]	740
A semi-supervised Support Vector Data Description- based fault detection method for rolling element bearings based on Cyclic Spectral Coherence, Liu Chenyu [et al.]	751
Machine teaching to optimize algorithms performances on restricted dataset., Carbonelli Alexandre [et al.]	764
Effects of the Particle Swarm Optimization parameters for structural dynamic monitoring of cantilever beam, Li Xiao-Lin [et al.]	778
List of participants	785
Author Index	788

Monday

Angular approaches

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

Edgar F. Sierra-Alonso^{1,2} Jerome Antoni¹ German Castellanos-Dominguez²

¹Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon,
LVA EA677, F-69621, Villeurbanne, France
edgar-felipe.sierra-alonso@insa-lyon.fr

²Signal 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 constant 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 nonstationarity 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 the low Signal to Noise Ratio (SNR); consequently, a short-time approach methodology robust to noise for IAS estimation 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 and 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 below 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 interest¹, 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:

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

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 Displacement² $\phi_{i,k}(t) \in \mathbb{R}^+$ is defined as:

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

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

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

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

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:

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

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 range³, 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 window⁴. 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 $\mathbf{x}[n] = [x[n], x[n+1], \dots, x[n+L-1]]^T$, 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:

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

$$\mathbf{Z}[n] = [\mathbf{e}^{j\phi_1[n]}, \mathbf{e}^{j\phi_1[n+1]}, \dots, \mathbf{e}^{j\phi_1[n+L-1]}]^T \quad (6)$$

$$\mathbf{e}^{j\phi_1[n]} = [e^{j1\phi_1[n]}, e^{j2\phi_1[n]}, \dots, e^{jI\phi_1[n]}] \quad (7)$$

$$\mathbf{a}[n] = [a_1, a_2, \dots, a_I]^T \quad (8)$$

$$\mathbf{a}[n] = [A_1 e^{j\psi_1}, A_2 e^{j\psi_2}, \dots, A_I e^{j\psi_I}]^T \quad (9)$$

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:

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

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

⁴the 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 $\mathbf{x}[n]$:

$$\phi_1[n] = \beta_1 n^2 / 2 + \beta_2 n \quad (11)$$

$$= \beta_1[n] + \beta_2[n] \quad (12)$$

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:

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

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:

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

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:

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

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:

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

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 moments 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:

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

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

$$\hat{S}_{2x}[f_n] := \langle |\text{STFT}_x[m, f_n]|^2 \rangle_m \quad (18)$$

$$\hat{S}_{4x}[f_n] := \langle |\text{STFT}_x[m, f_n]|^4 \rangle_m \quad (19)$$

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:

$$\text{SK}[f_n] = \frac{\hat{S}_{4x}[f_n]}{\hat{S}_{2x}^2[f_n]} - 2 \quad (20)$$

Please note that $\text{SK}[f_n]$ 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 $\text{SK}_n[f_m]$. 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

$F_{t \rightarrow f} \{ \mathbb{E} \{ |z(t)|^2 \} \}$, where $z(t)$ is the analytic signal⁵ 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:

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

then the spectral correlation is defined as:

$$S_z(\alpha, f) = F_{\tau \rightarrow f} \left\{ F_{t \rightarrow \alpha} \{ R_z(t, \tau) \} \right\} \quad (22)$$

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:

$$M_z(\alpha) = \int_{\mathbb{R}} S_z(\alpha, f) df \quad (23)$$

$$= F_{t \rightarrow \alpha} \{ R_z(t, 0) \} \quad (24)$$

$$= F_{t \rightarrow \alpha} \{ \mathbb{E} \{ |z(t)|^2 \} \} \quad (25)$$

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 $\mathbf{x}[n] = [x[n], x[n+1], \dots, x[n+L-1]]^T$, let us consider a filtered version using the SK as:

$$\mathbf{x}_{\text{SK}}[n] = \text{DFT}_{f_m \rightarrow n}^{-1} \{ \mathbf{X}_n[f_m] \text{SK}_n[f_m] \} \quad (26)$$

where $\mathbf{X}_n[f_m] := \text{DFT}_{n \rightarrow f_m} \{ \mathbf{x}[n] \}$ being $\text{DFT}_{n \rightarrow f_m} \{ \cdot \}$ the Discrete Fourier transform of a given signal, and $\text{DFT}_{f_m \rightarrow n}^{-1} \{ \cdot \}$ 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:

$$\hat{S}_{|\mathbf{z}_{\text{SK}}|^2}[f_m] = \frac{1}{NU} \sum_{n=1}^N \left| \text{DFT}_{n \rightarrow f_m} \{ |\mathbf{z}_{\text{SK}}[n]|^2 \} \right|^2 \quad (27)$$

where $\mathbf{z}_{\text{SK}}[n]$ is the analytical signal, a signal whose modulus is approximately the envelope⁶ 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 $\hat{S}_{|\mathbf{z}_{\text{SK}}|^2}[f_m]$. 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.

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

⁶the 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:

$$\hat{x}[n] = h[n] \otimes \sum_{i=1}^I a_i[n] e^{j\phi_i[n]} + \eta[n] \quad (28)$$

where $\otimes(\cdot, \cdot)$ 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 $n_1 \approx 2.5fs$ and $n_2 = n_1 + 0.1fs$. 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.

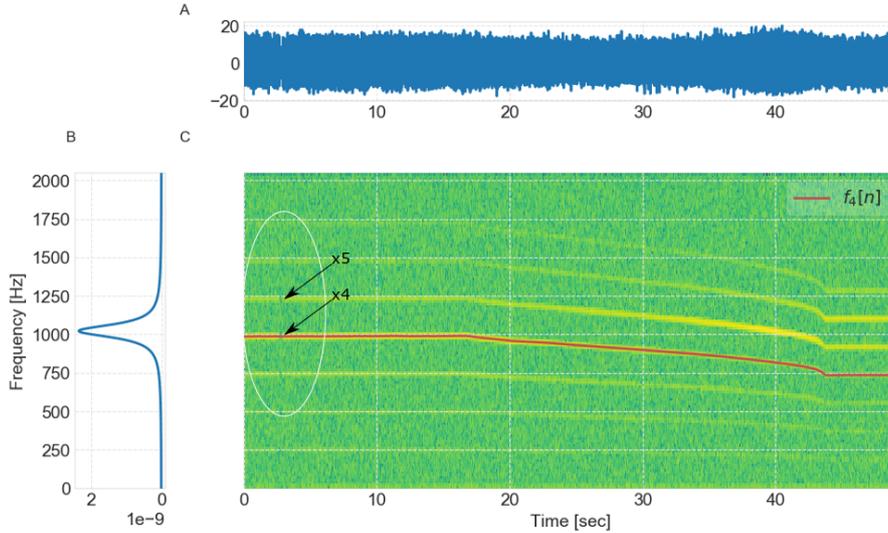


Figure 1: Numerical signal with AWGN of 3dB, a) in time, b) transfer function $H(f)$, c) spectrogram showing a random discontinuity of 0.1 sec reflected on all orders.

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 noise⁷ 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:

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

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 $N > L$, 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.

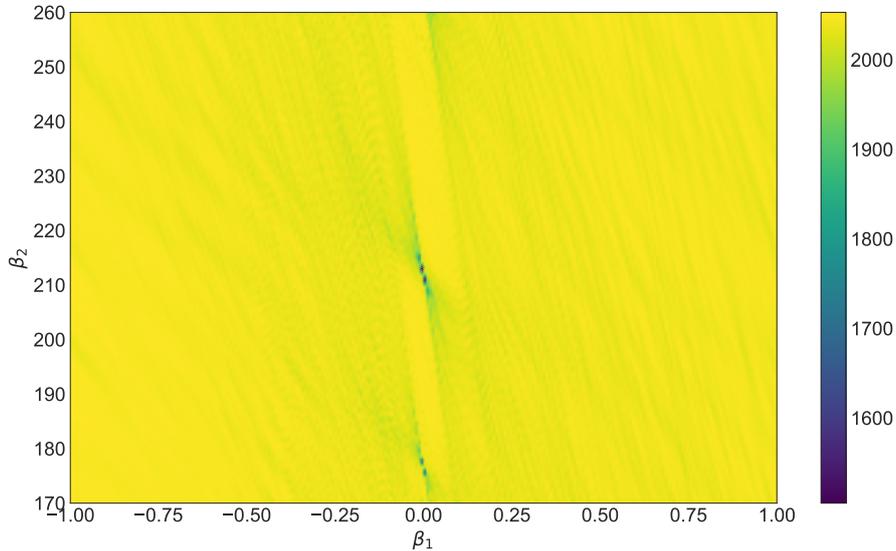


Figure 2: Cost function in the Eq. (13) for a segment of the the exemplary signal in Fig. 1 at 35sec.

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 $\beta_2n = \beta_2$ per each n -the segment. where the n in β_2n 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 $\eta_{\text{PN}}[f_m] \approx 1/f_m$.

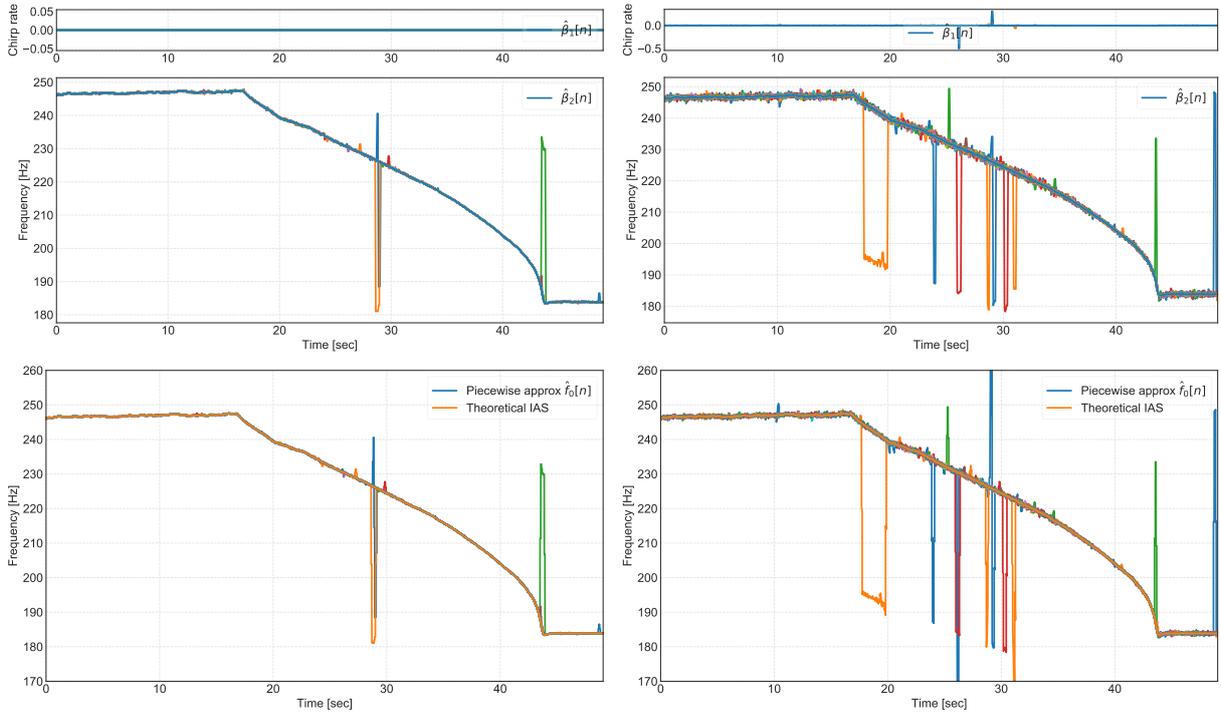


Figure 3: AWGN test, first-column: $\beta_1[n]$, $\beta_2[n]$, and IAS estimated no linear approximation, second-column: $\hat{\beta}_1[n]$, $\hat{\beta}_2[n]$ and IAS estimation after linear approximation, for all the considered signal polluted with noise in to achieve a for SNR in $[-10, 10]$ dB.

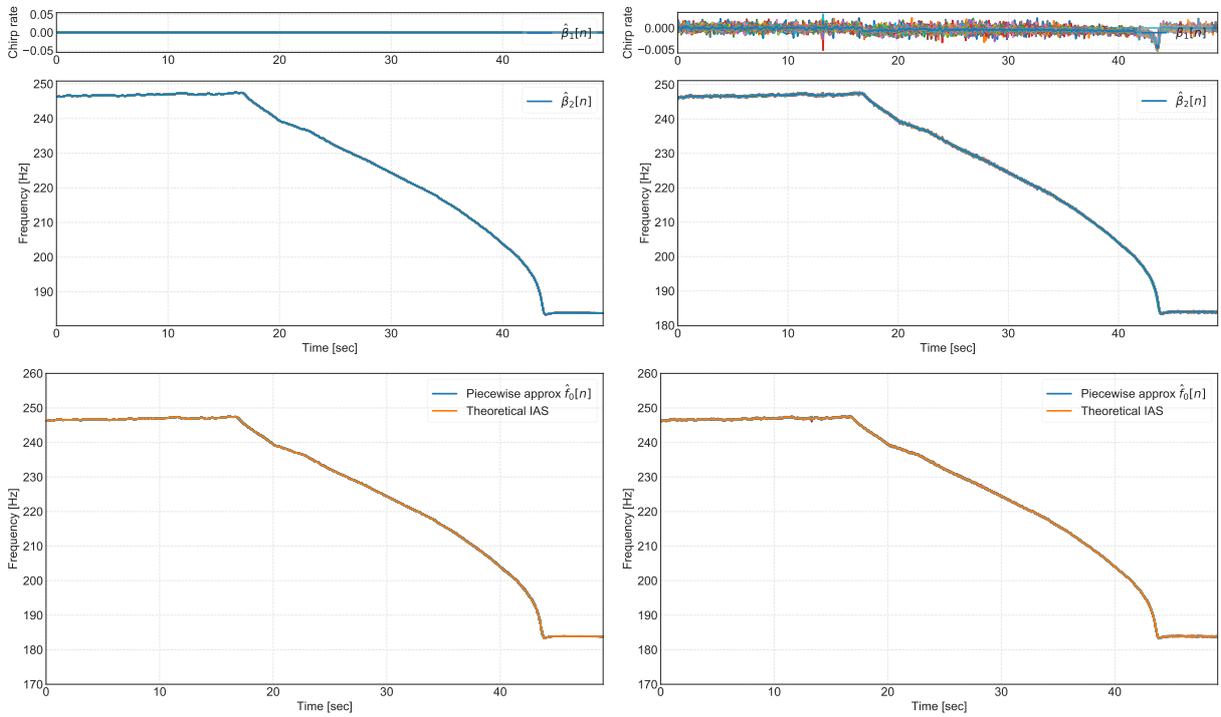


Figure 4: Pink noise test, first-column: $\beta_1[n]$, $\beta_2[n]$, and IAS estimated no linear approximation, second-column: $\hat{\beta}_1[n]$, $\hat{\beta}_2[n]$ and IAS estimation after linear approximation, for all the considered signal polluted with noise in to achieve a for SNR in $[-10, 10]$ 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.

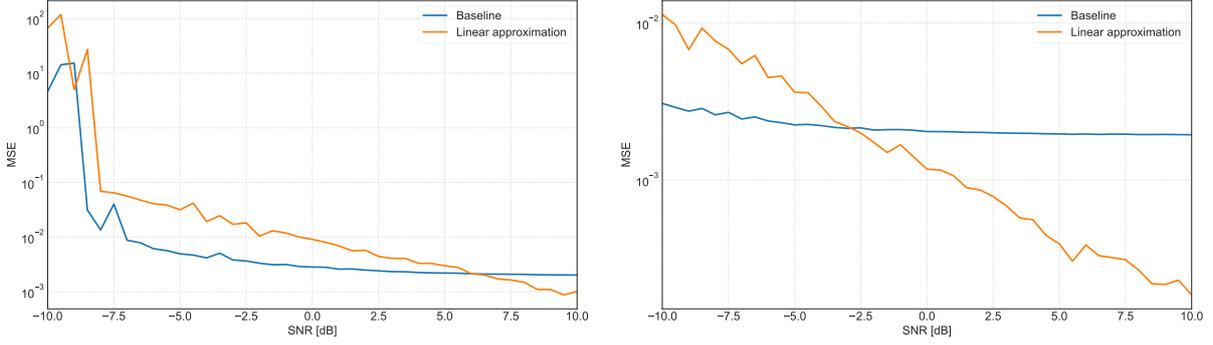


Figure 5: SNR vs MSE tests for an SNR between $[-10, 10]$ dB, in the left the AWN scenario, and in the right the Pink-noise scenario, for the baseline and linear IF reconstruction of the IAS $\hat{f}_1[n]$ using Eqs. (11) and (12)

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 -3 dB 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 x_1 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 x_2 . 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 x_1 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

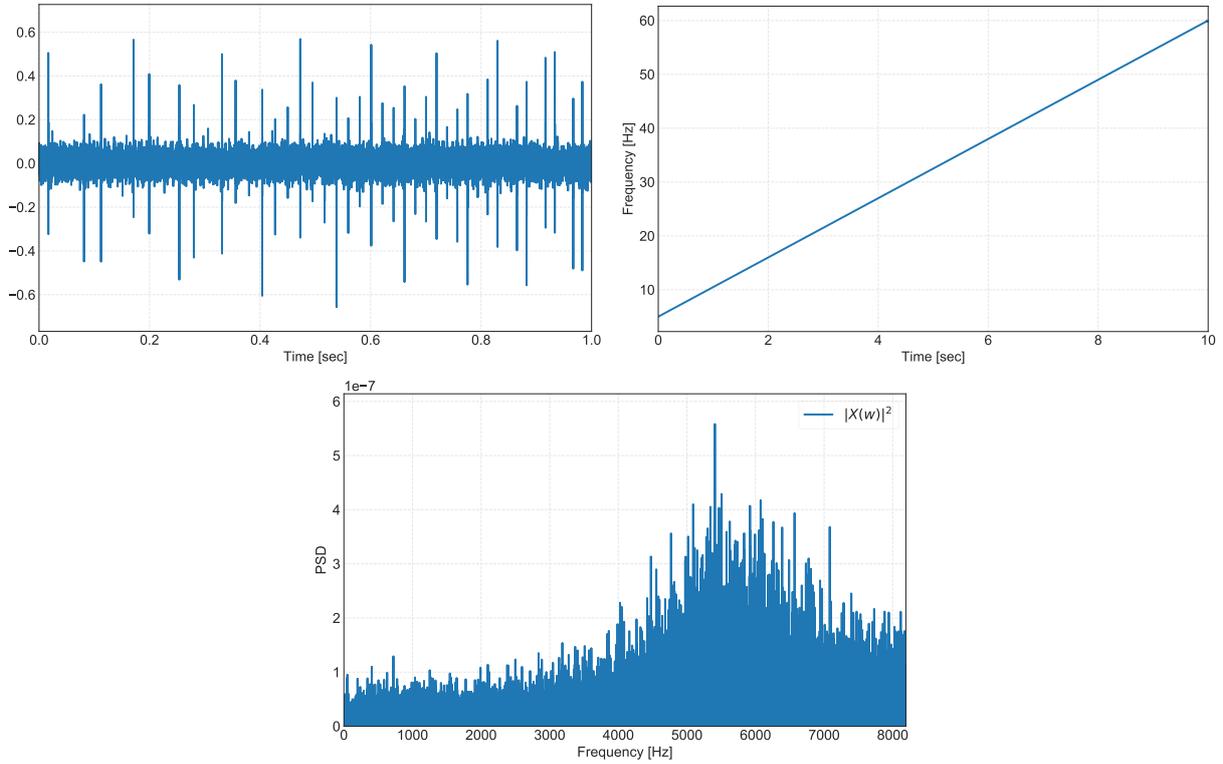


Figure 6: Top-row: left) numerical signal in time with a BPFI at 5.875Hz, right) IAS profile, bottom-row: spectrum of the simulated signal.

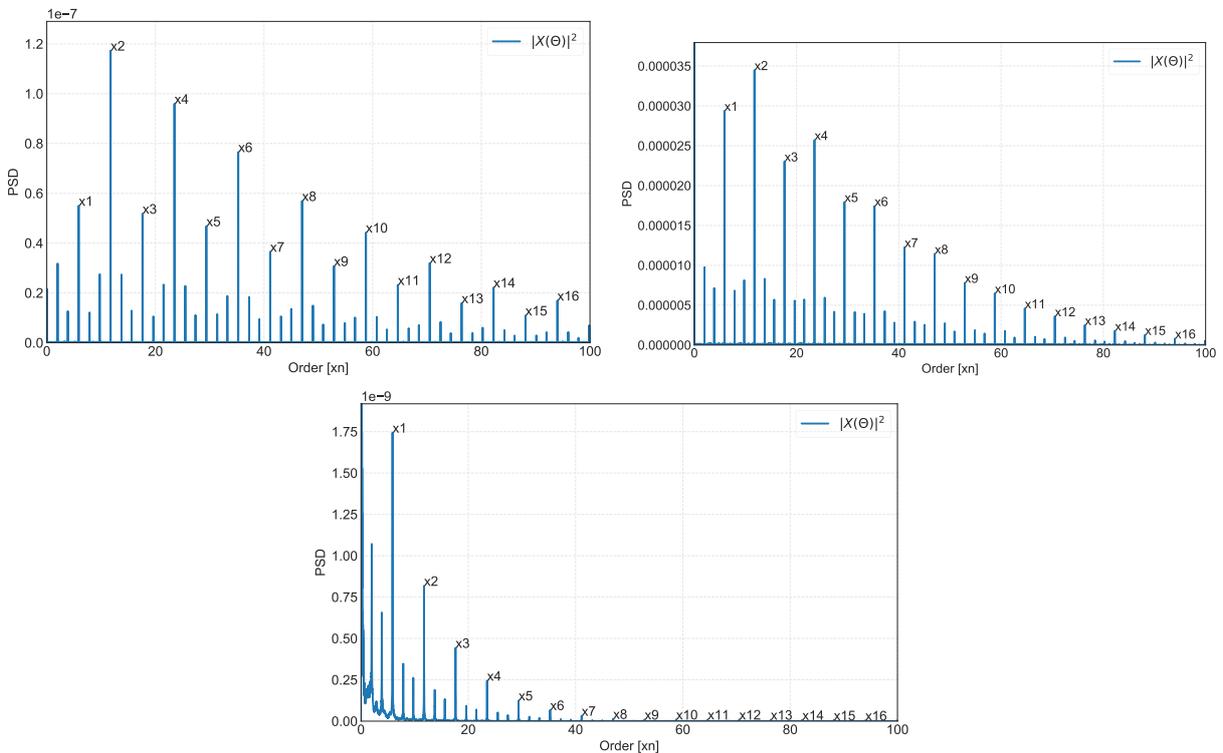


Figure 7: Top-row: baseline methodologies, left) angular domain transform, SK filtering, then envelope spectrum, right) SK filtering, angular domain transform, then envelope spectrum, bottom-row: proposal) angular domain transform, then, envelope spectrum convolved with a time-frequency dependent SK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

envelope spectrum.

3.2 Case study: an aircraft engine

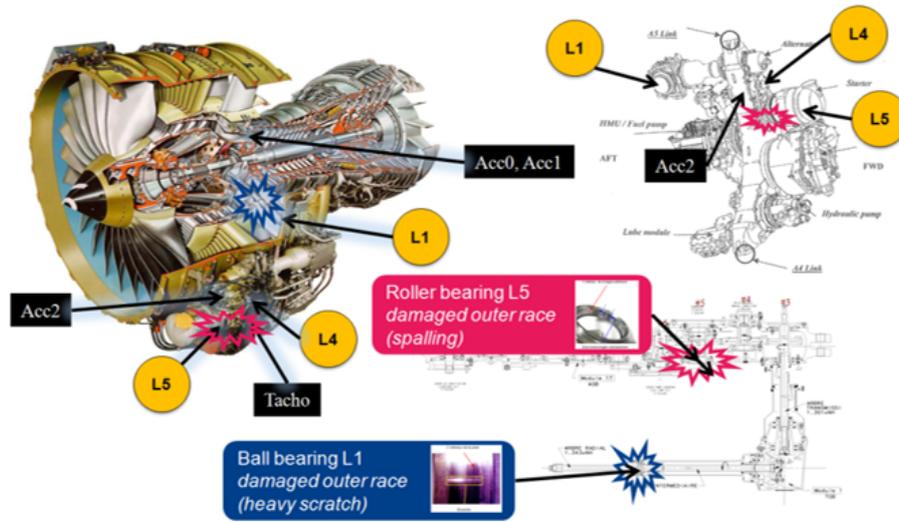


Figure 8: General overview of the engine and the accessory gearbox. Shafts (L1-L5) are identified by labels in amber colour (image taken from [4]).

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.

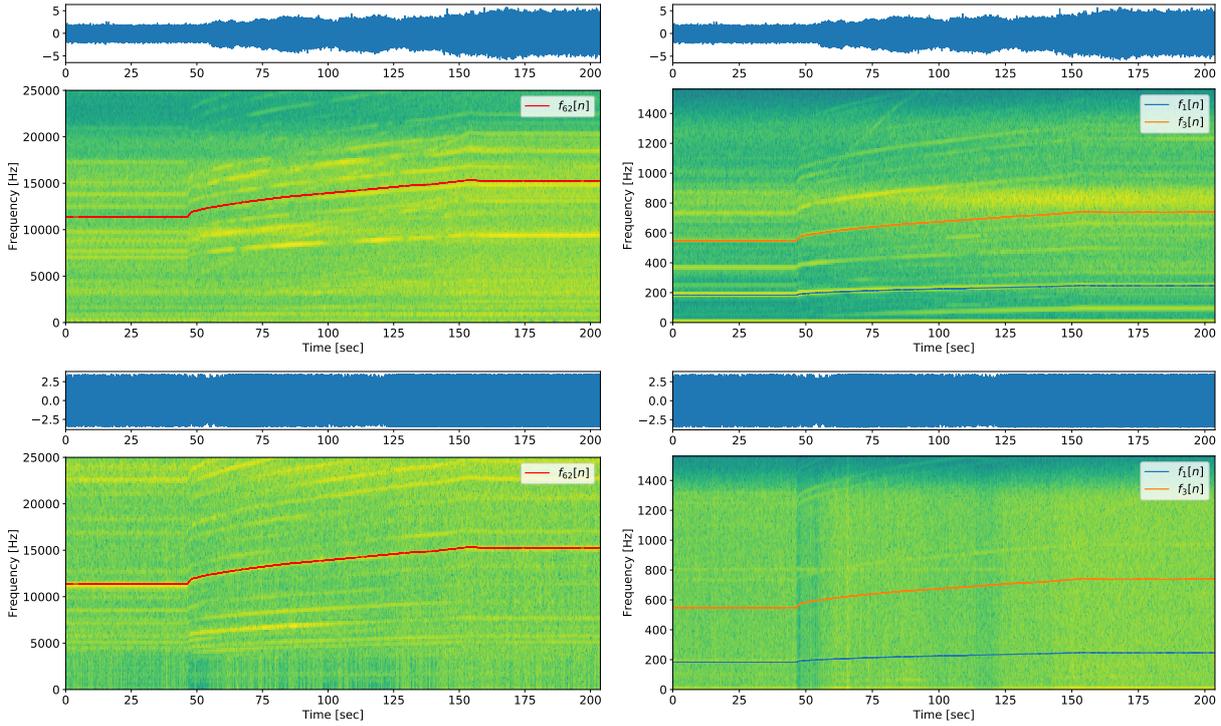


Figure 9: Top-row: spectrograms ACC1, left, full frequency range $[0, 25]$ kHz, right, down-sampled to $[0, 1.56]$ kHz, bottom-row: spectrograms for ACC2 same frequency ranges than in top figures.

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,

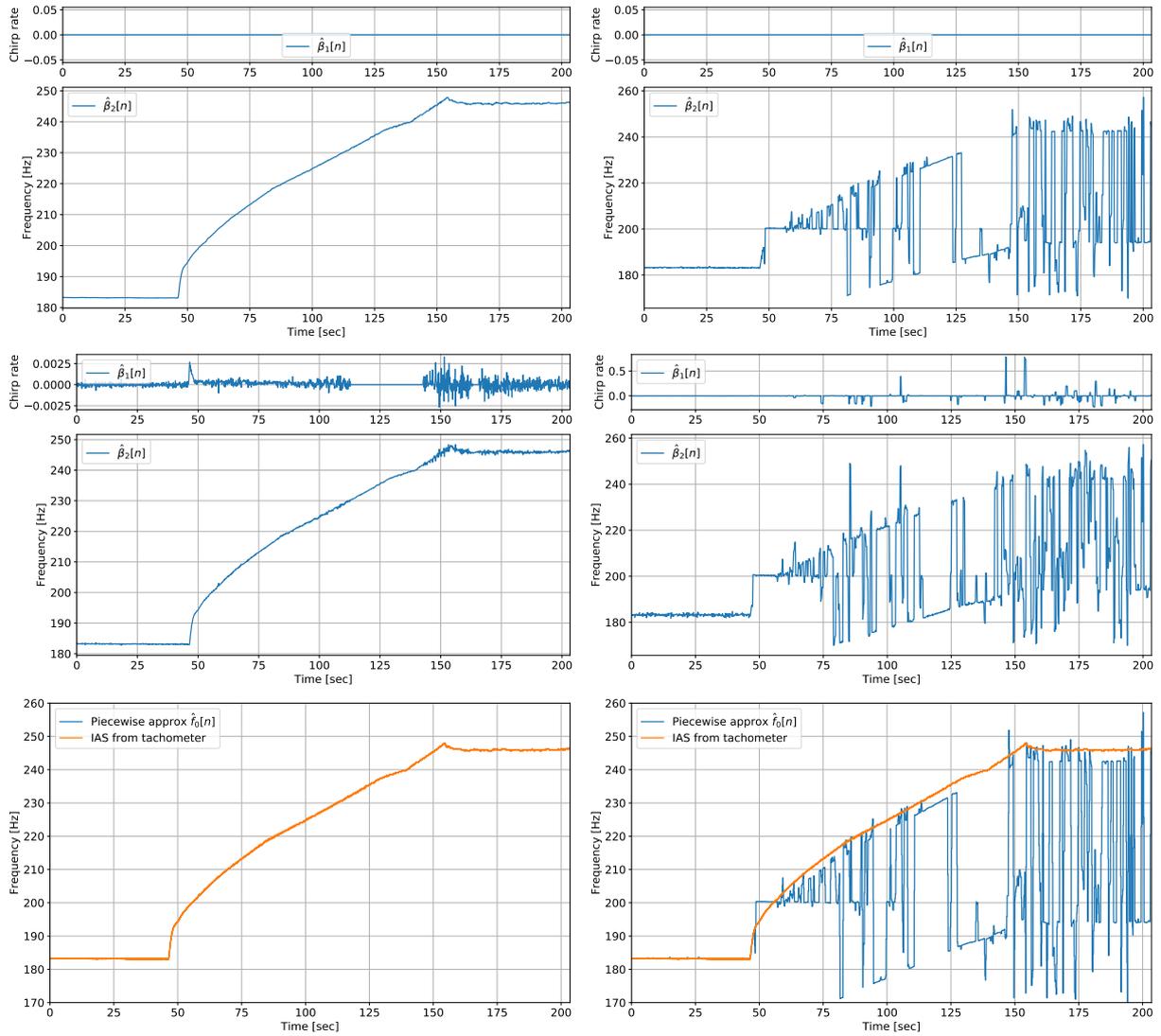


Figure 10: Left-column ACC1, right-column ACC2, rows: i) phase estimation without linear approximation, ii) phase with linear approximation $\{\beta_i[n]\}_{i=1}^2$, iii) IAS estimation after linear approximation.

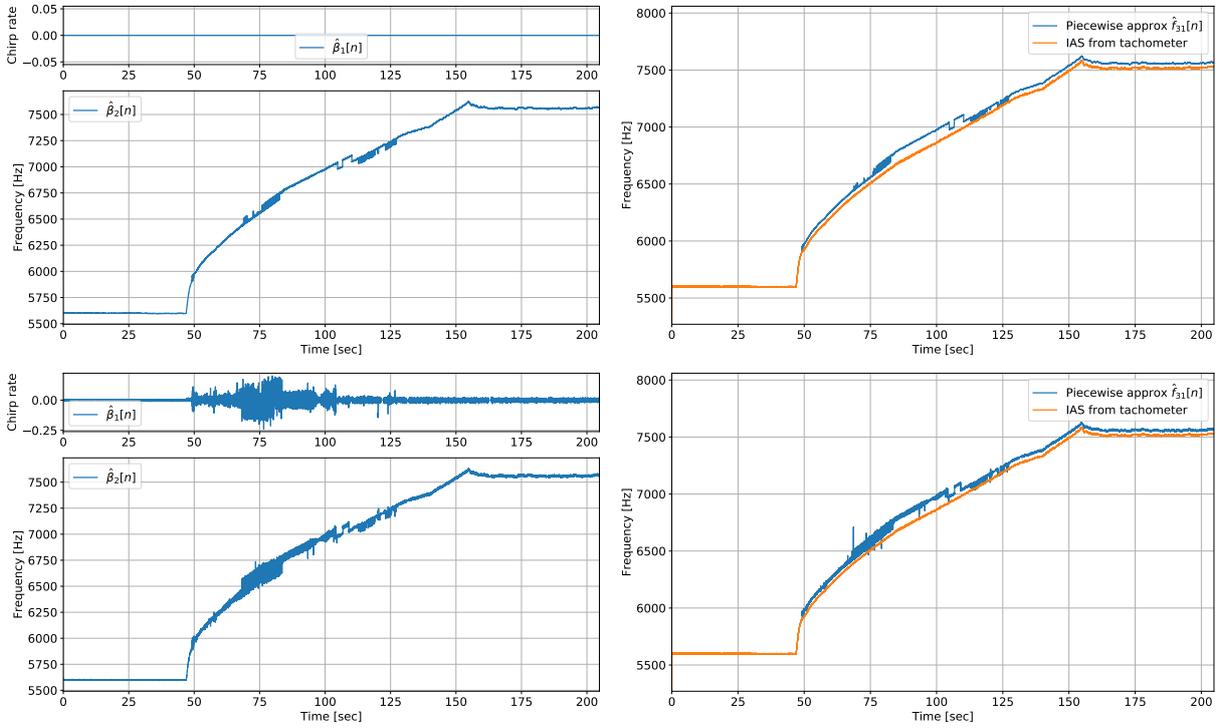


Figure 11: ACC2: IAS estimation approximation taking as fundamental frequency $f_{311}[n]$, top-row: without linear approximation, bottom-row: with linear approximation.

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 necessities 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 sampling 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

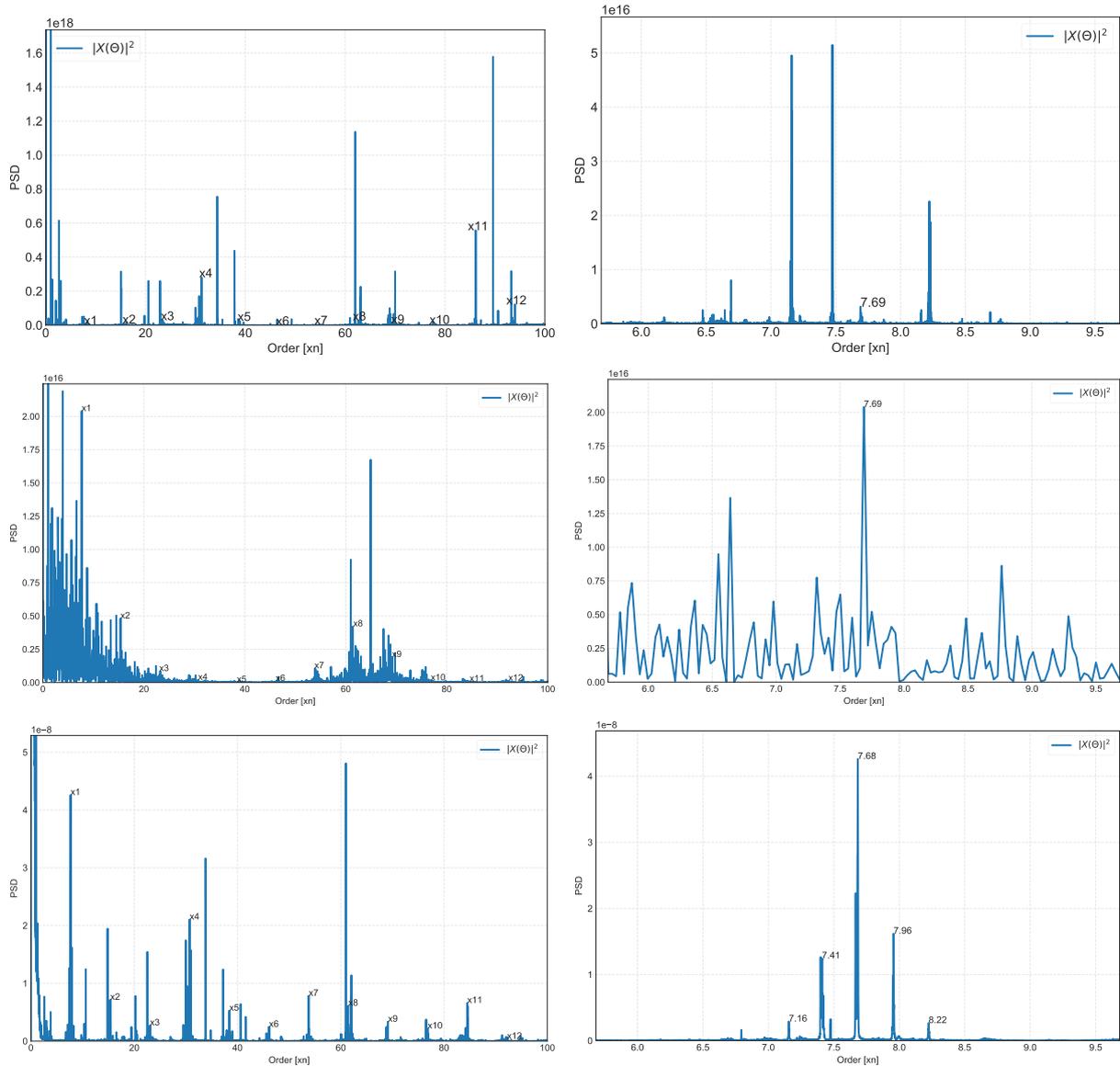


Figure 12: Top-row: baseline one, angular domain transform, SK filtering, then envelope spectrum, Middle-row: SK filtering, angular domain transform, then envelope spectrum, Bottom-row: **proposal** angular domain transform, then, envelope spectrum convolved with a time-frequency dependent SK based filter.

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 improperly 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 GE 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:

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

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:

$$\overline{\tau}[j] = \frac{\Delta\theta}{\overline{\omega}_{LF}[j]} \rightarrow \Delta\theta[j] = \overline{\omega}_{LF}[j] \cdot \overline{\tau}[j] \quad (2)$$

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

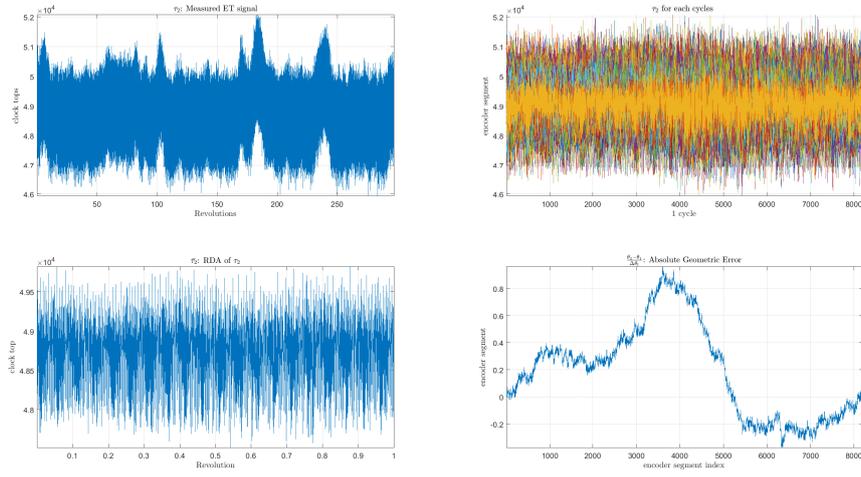


Figure 1 – GE Estimation of one of the measurements

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

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

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.

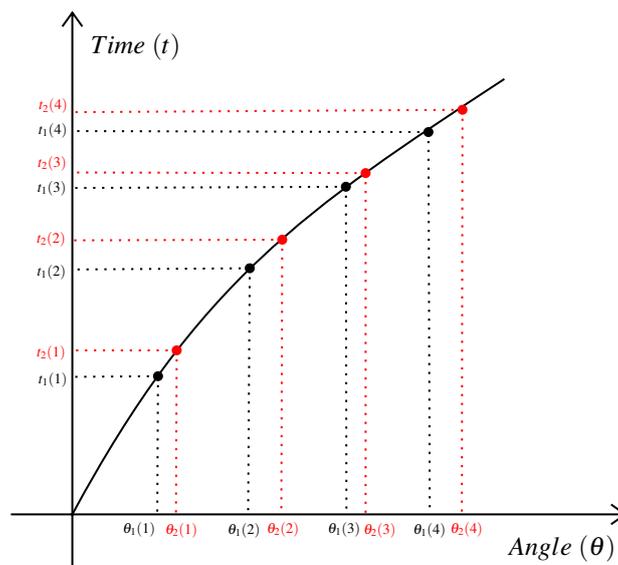


Figure 2 – Angle versus Time

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 cyclostationary behaviour, the energy contained by the integer orders appears as the ideal indicator.

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

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.

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

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.

$$I = \begin{pmatrix} 1 & I_{12} & I_{13} & \dots & I_{1m} \\ I_{21} & 1 & I_{2m} & \dots & I_{22} \\ \dots & \dots & \dots & \dots & \dots \\ I_{m1} & I_{m2} & I_{m3} & \dots & 1 \end{pmatrix} \quad (6)$$

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 epicyclic 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.

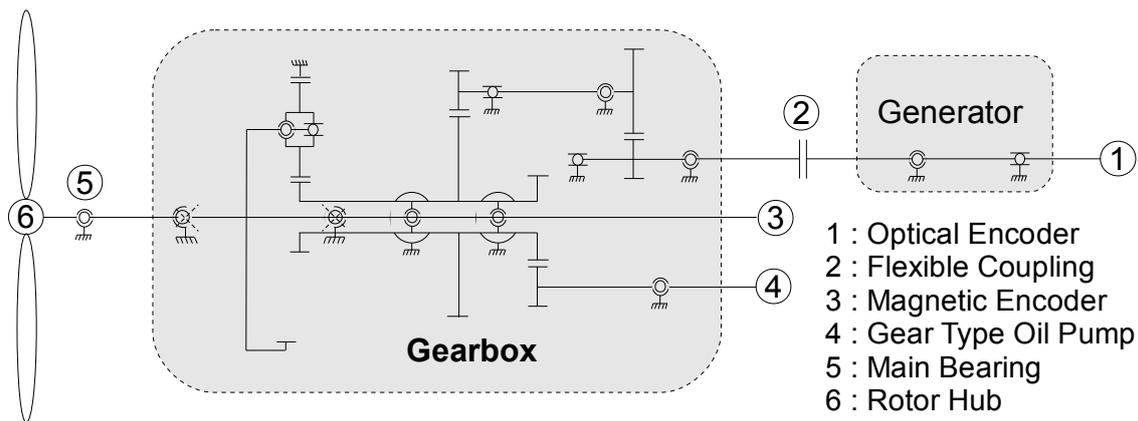


Figure 3 – Kinematic Scheme of the wind turbine set-up.

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.

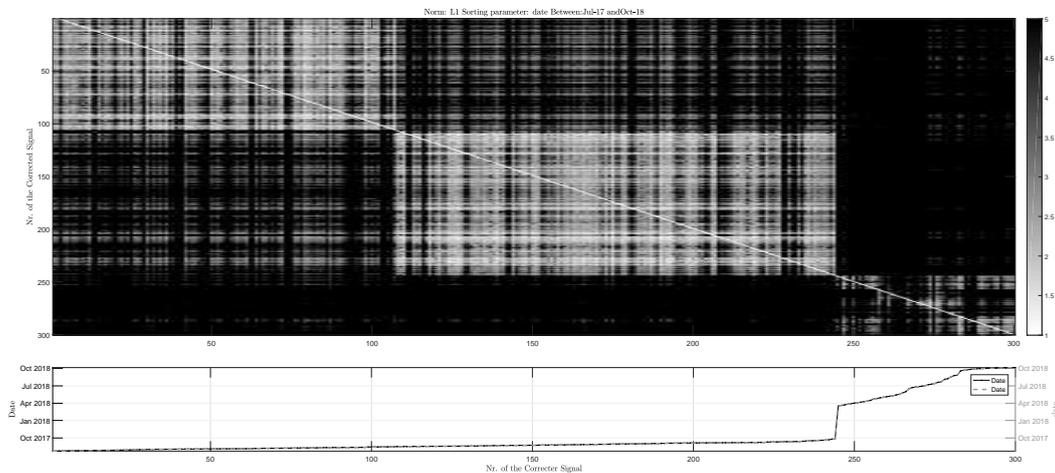


Figure 4 – Cross correction map sorted by date. The bottom plot reads the date of each measurement.

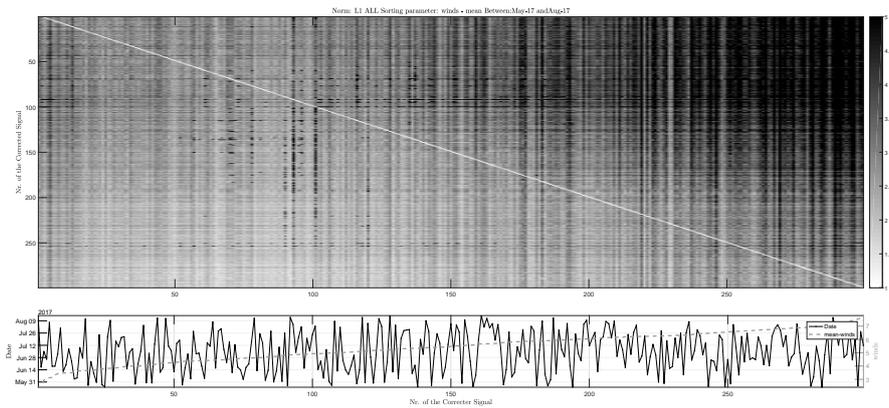
Unlike the maps obtained using other parameters, this map clearly exhibits vast light and dark regions. Lighter regions means every signal within it correct and are corrected in a better way between themselves. The top left region is limited by the measurement 105th, which has been acquired just before encoder removal, on 25th 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 244th 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.

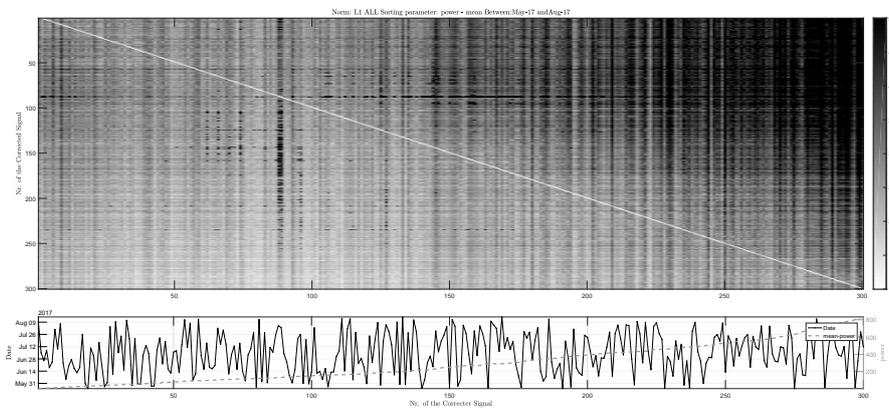
3.2.2 Influence of the Power and Wind Speed

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.

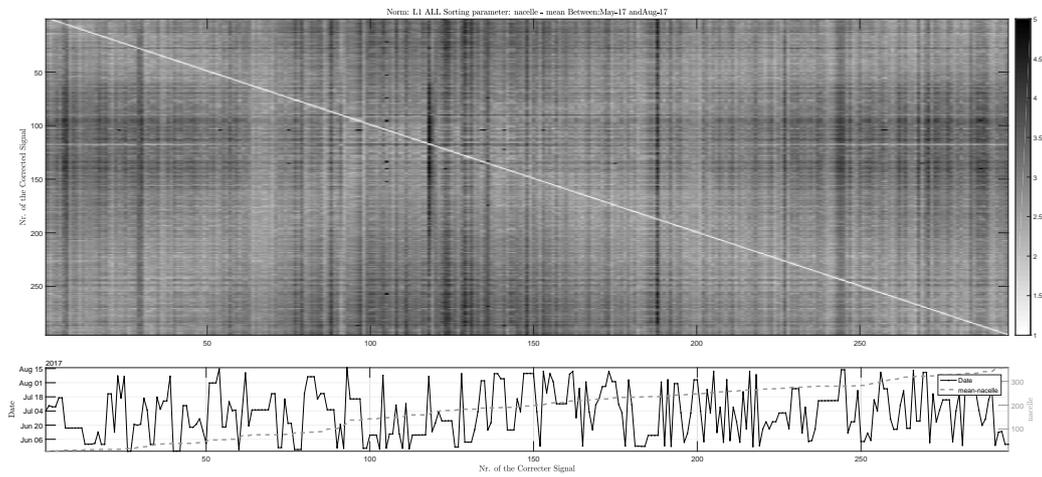


Figure 6 – Map demonstrates the influence of the nacelle values on GE correction efficiency. *The map is sorted by mean nacelle values.*

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] J.L. 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 Pham¹, M Gagnon², J Antoni³, S A Tahan⁴, C Monette⁵

¹Phd Student at ÉTS Montréal and INSA de Lyon

²Institut de recherche d'Hydro-Québec, Varenne, Québec, Canada

³Univ Lyon, INSA-Lyon, Laboratoire Vibrations Acoustique, Villeurbanne, France

⁴École de technologie supérieure (ÉTS), Montréal, Québec, Canada

⁵Andritz 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.

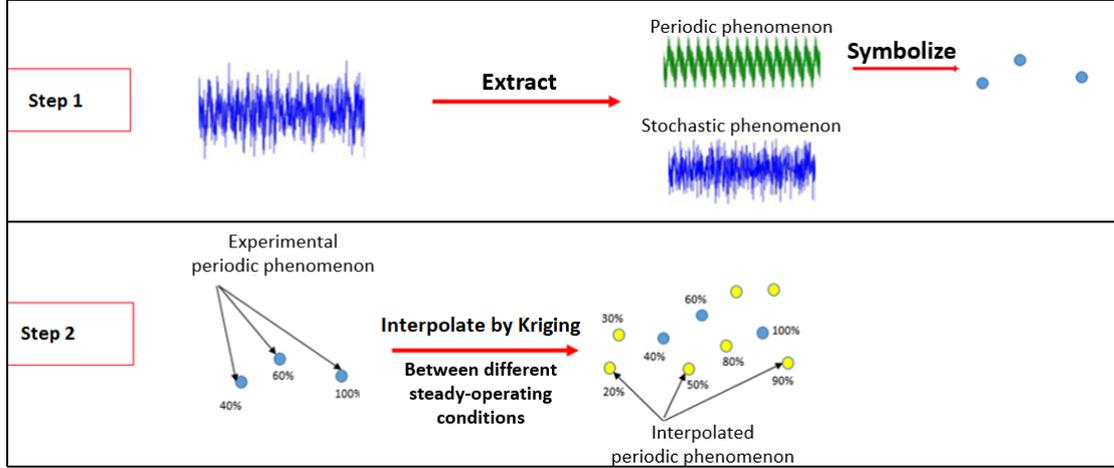


Figure 1 – Proposed interpolation tool

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 cyclostationarity 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].

$$S[n] = CS1[n] + CS2[n] + R[n] \quad (1)$$

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).

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

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

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.

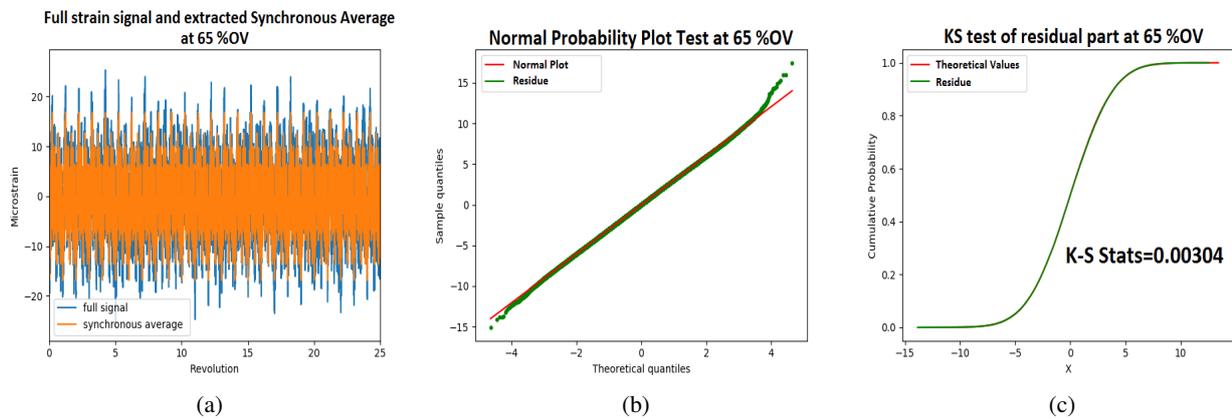


Figure 2 – Results of SA extraction process at 65 %OV. (a) Extracted Synchronous Average ; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.

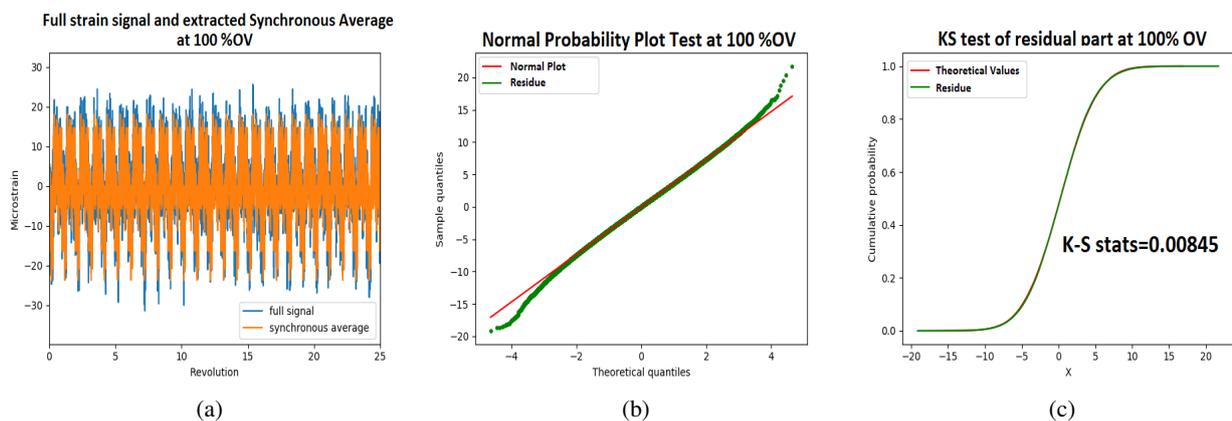


Figure 3 – Results of SA extraction process at 100 %OV. (a) Extracted Synchronous Average; (b) Normal probability test of the residual part; (c) Kolmogorov-Smirnov test of the residual part.

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.

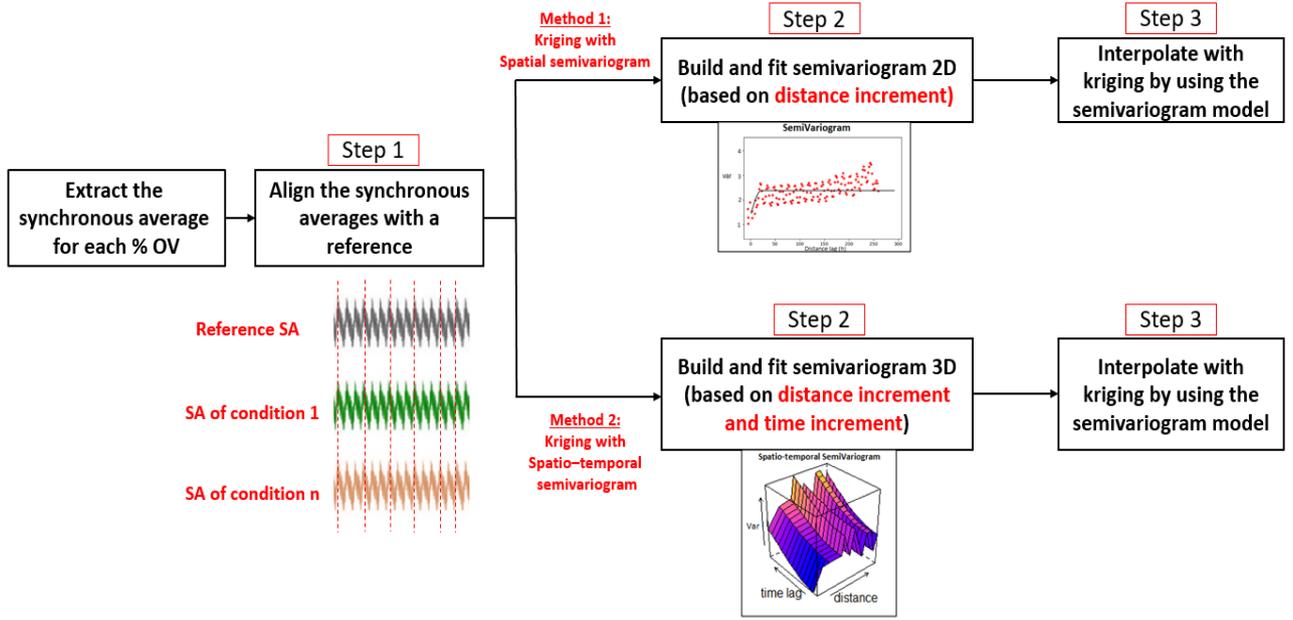


Figure 4 – Proposed interpolation process

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.

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

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).

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

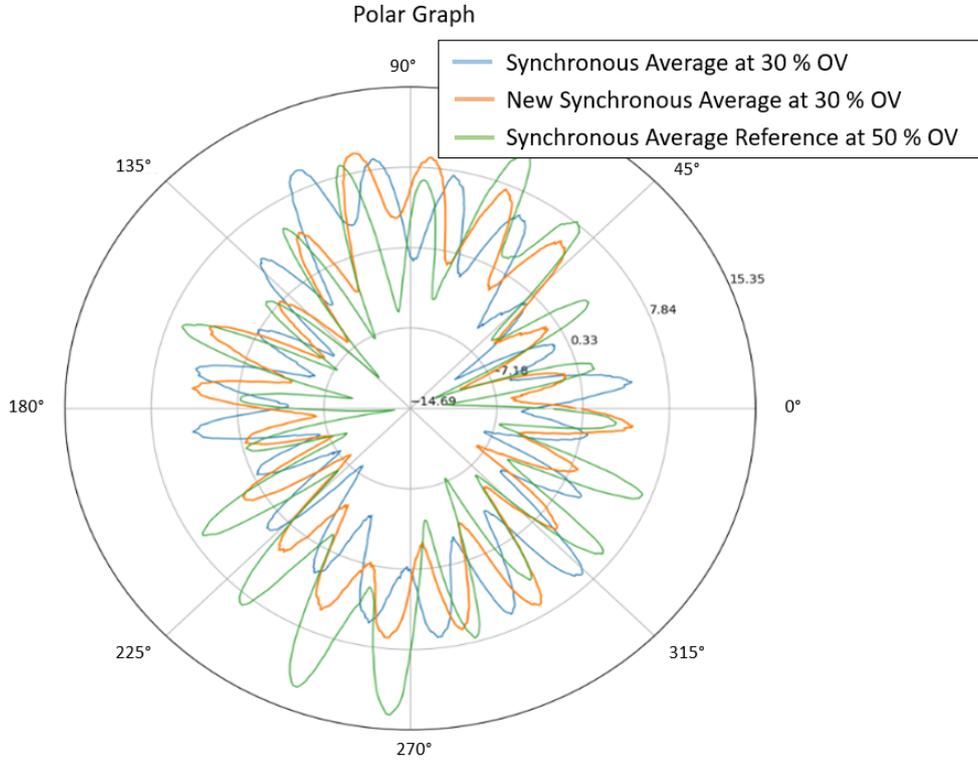


Figure 5 – Synchronous Average in polar field

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

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:

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

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.

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

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:

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

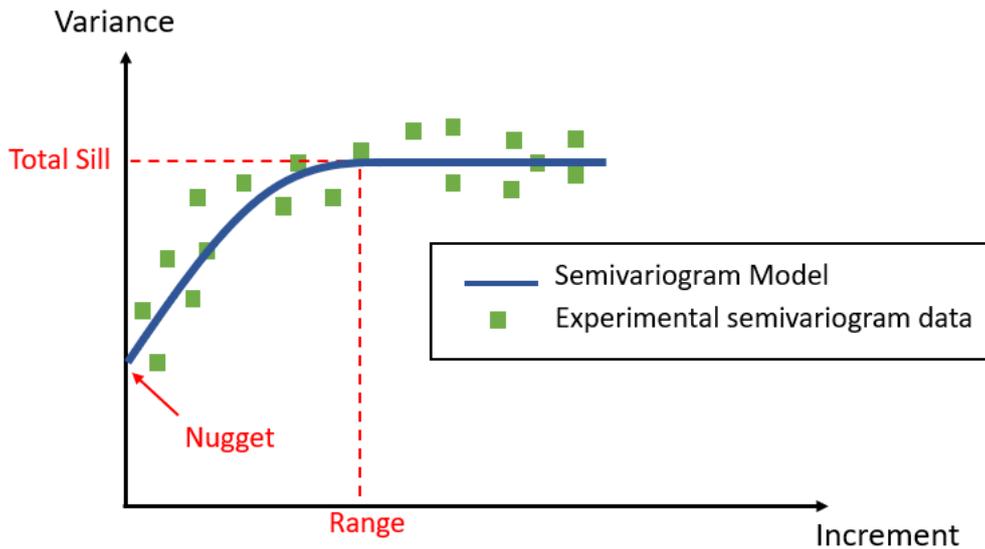


Figure 6 – Semivariogram model

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.

$$\hat{\gamma}(h, a) = \frac{1}{2N(h, a)} \sum_{(i, j) \in N(h, a)} [S(\%OV_i, \theta_i) - S(\%OV_j, \theta_j)]^2 \quad (9)$$

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.

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

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	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
	Second Approach: Nugget effect + long-range gaussian model + short-range hole effect model		

Table 1 – Information of input, output 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).

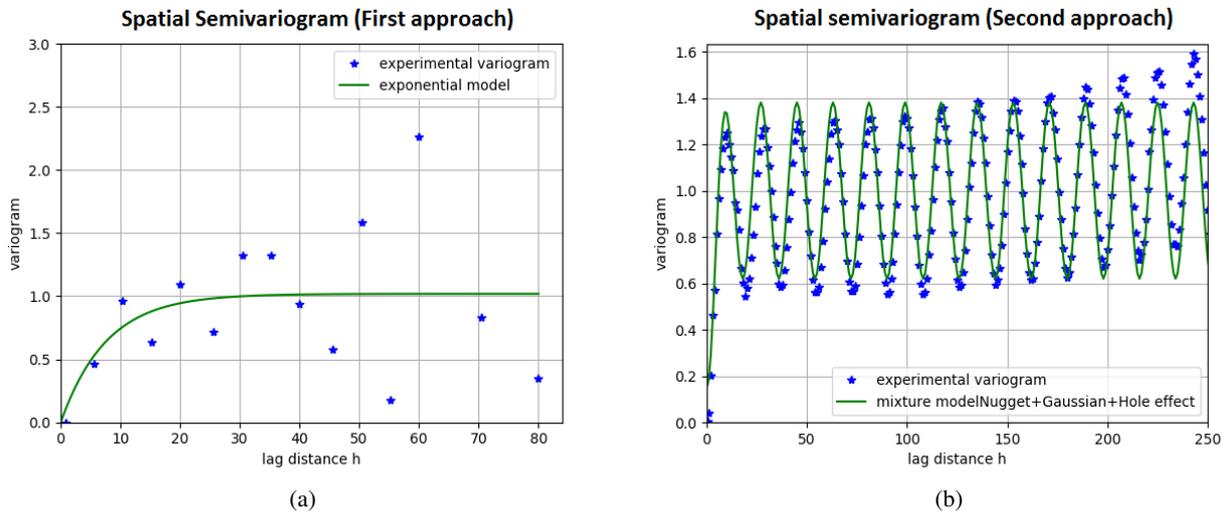


Figure 7 – Experimental semivariogram and fitted model in the case of spatial method: (a) First approach; (b) Second approach

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

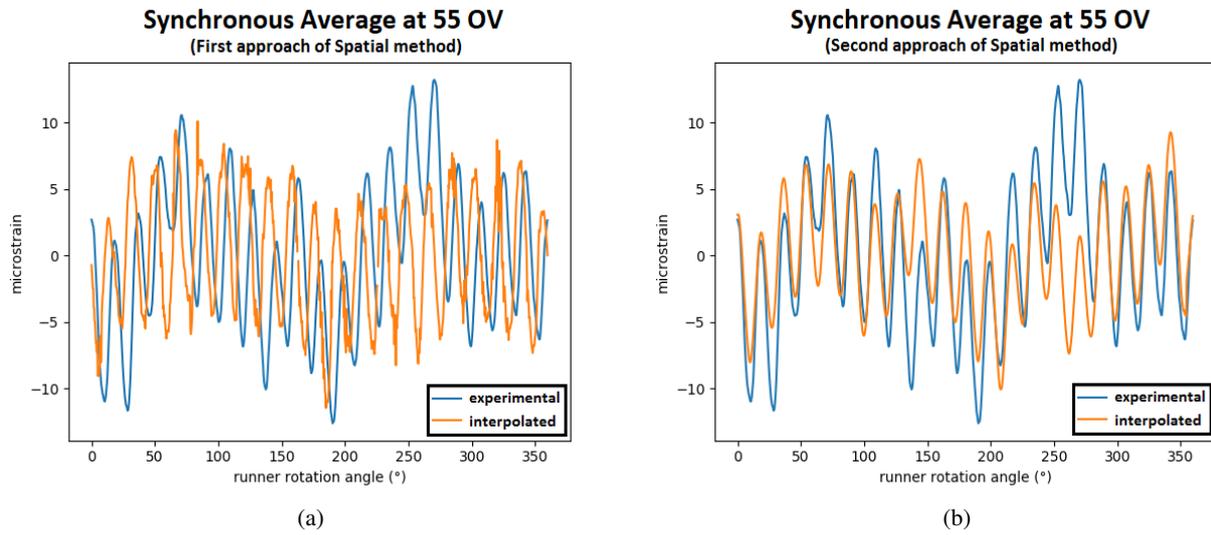


Figure 8 – Experimental and interpolated data at 55 %OV in the case of spatial method: (a) First approach; (b) Second approach

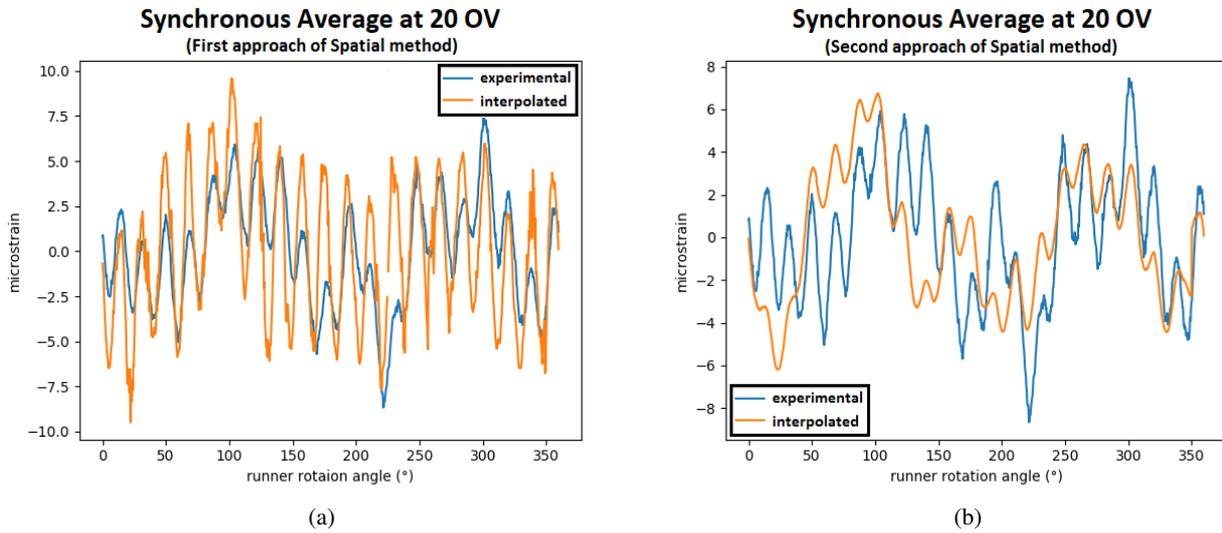


Figure 9 – Experimental and interpolated data at 20 %OV in the case of spatial method: (a) First approach; (b) Second approach

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.

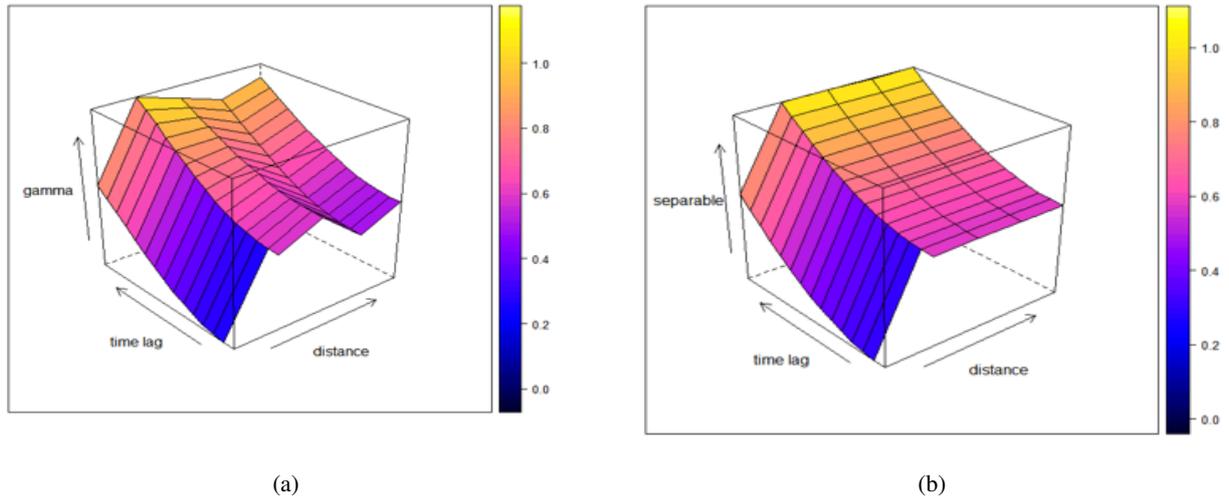


Figure 10 – (a) Experimental spatio-temporal semivariogram; (b) Separable model used to fit the experimental semivariogram

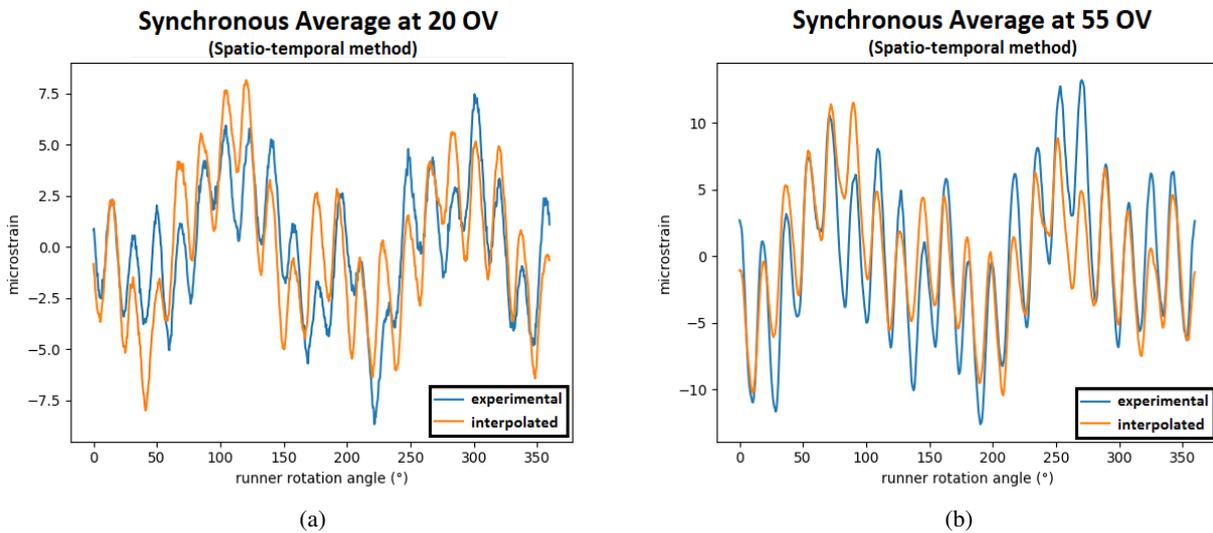


Figure 11 – Experimental and interpolated data in the case spatio-temporal method: (a) 20 %OV; (b) 55 %OV

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.

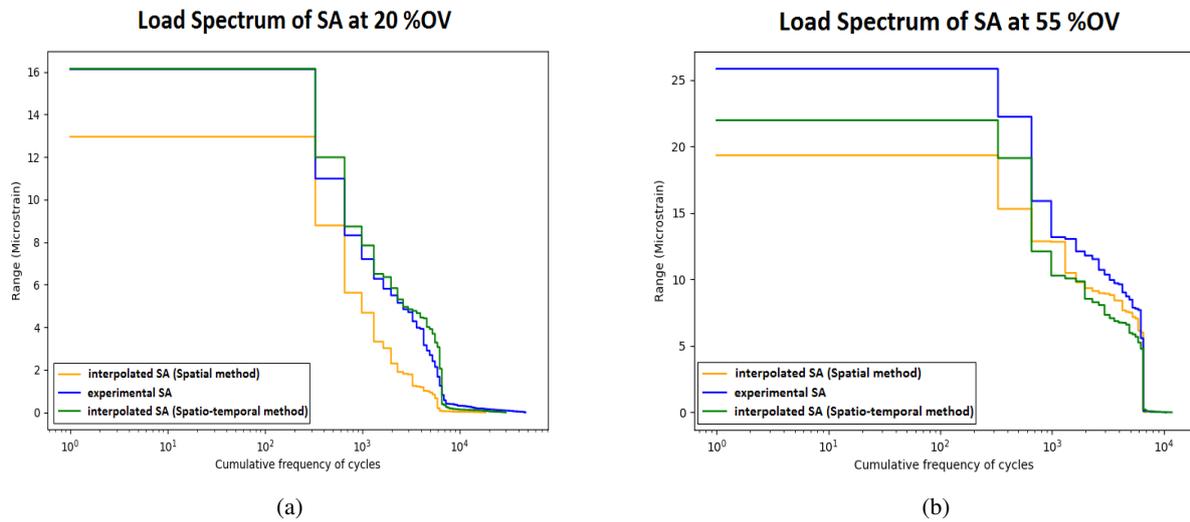


Figure 12 – Load spectrum comparison between original SA and interpolated SA of two kriging methods: (a) 20 %OV; (b) 55 %OV

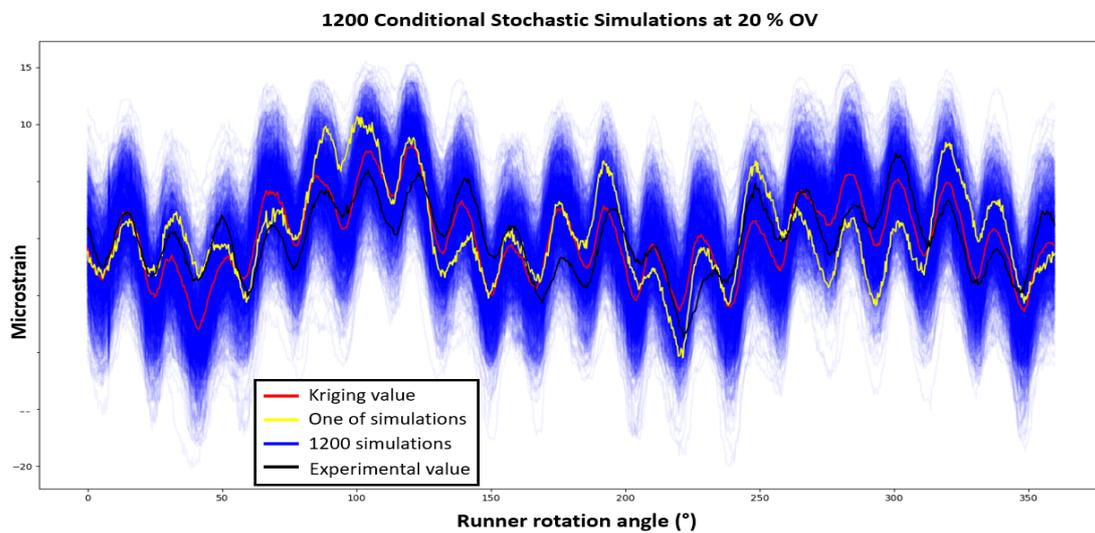


Figure 13 – Confidence interval by 1200 conditional stochastic simulations at 20 %OV

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 Hou³, 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 made a lot of 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 mono-component via numerical approximation. However, because of 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 circumstances. 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 needed 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 be 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 tachometer-less OT method for REB fault detection

The meliorated NMD method are shown in last subsection, and the realization process of NMD based tachometer-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 tachometer 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].

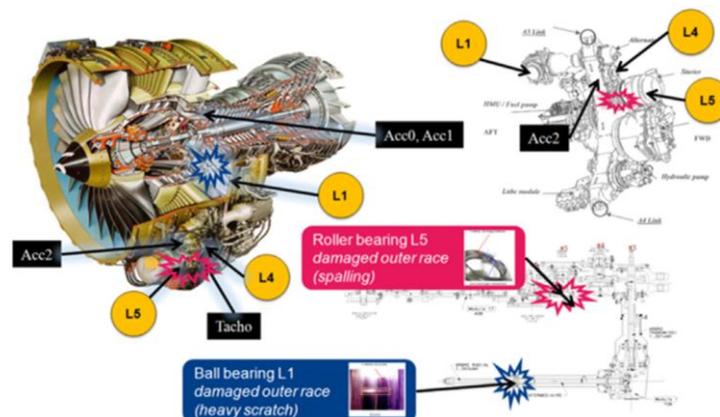


Figure 1: General overview of the engine and the accessory gearbox. [1]

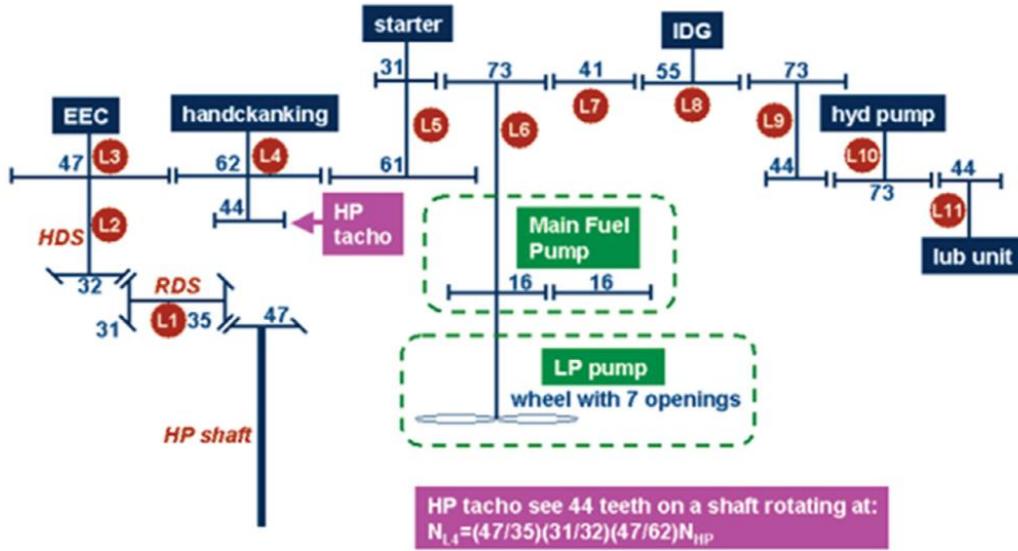


Figure 2: kinematics of the aircraft engine gearbox. [1]

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 hypothesis 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 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 hypothesis of noise. As a

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

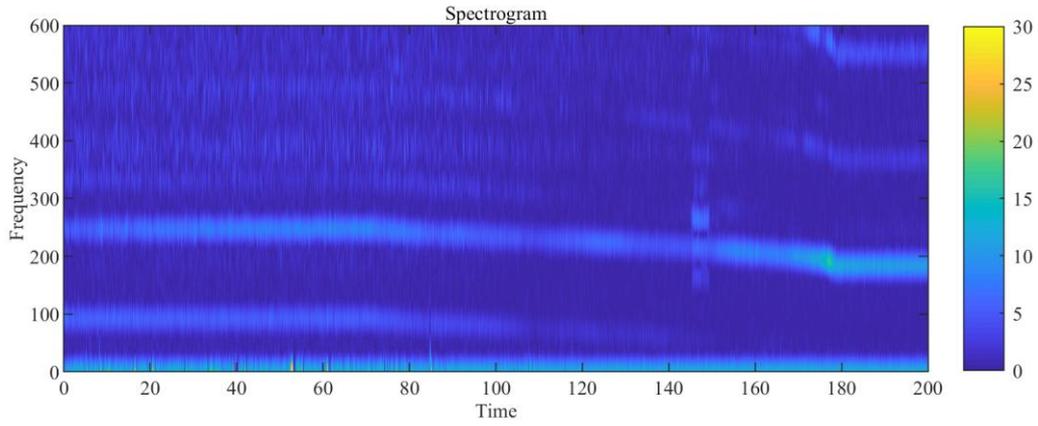


Figure 3: The partially enlarged drawing of the TFD of the original signal.

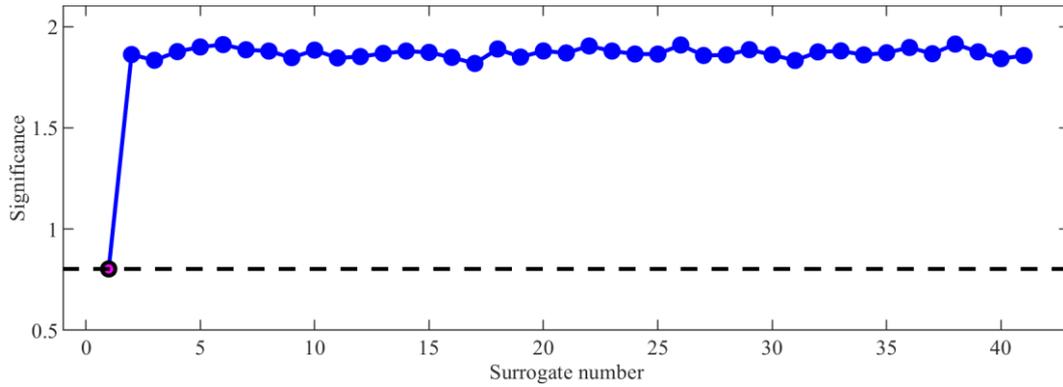


Figure 4: The Fourier transform surrogate test.

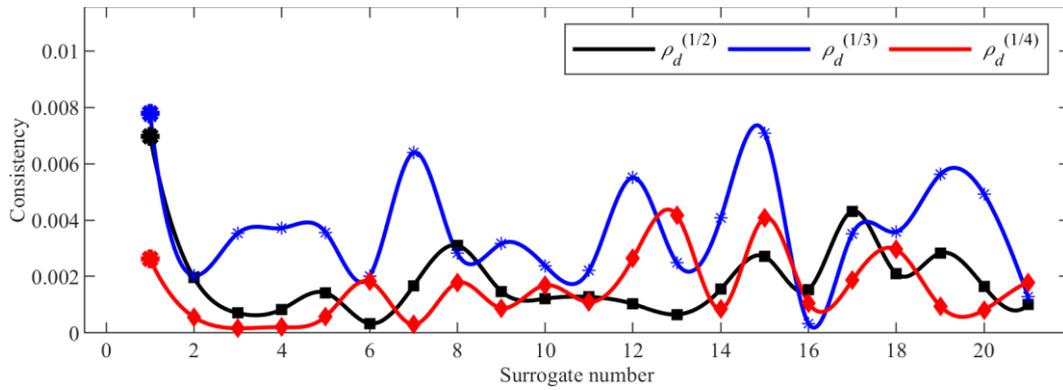


Figure 5: The consistency of the candidate sub-harmonics obtained by time-shifted surrogate test.

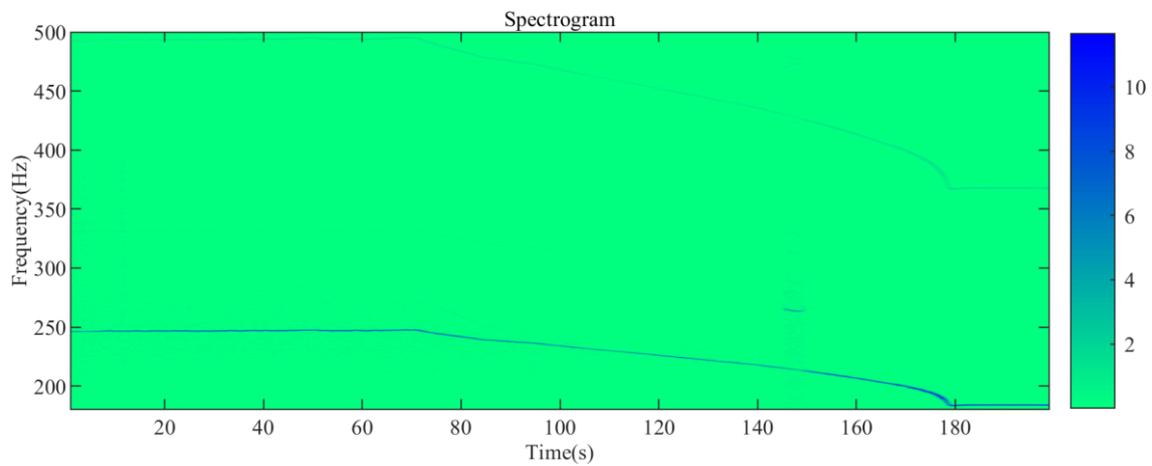


Figure 6: The extracted component in TFD domain.

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.

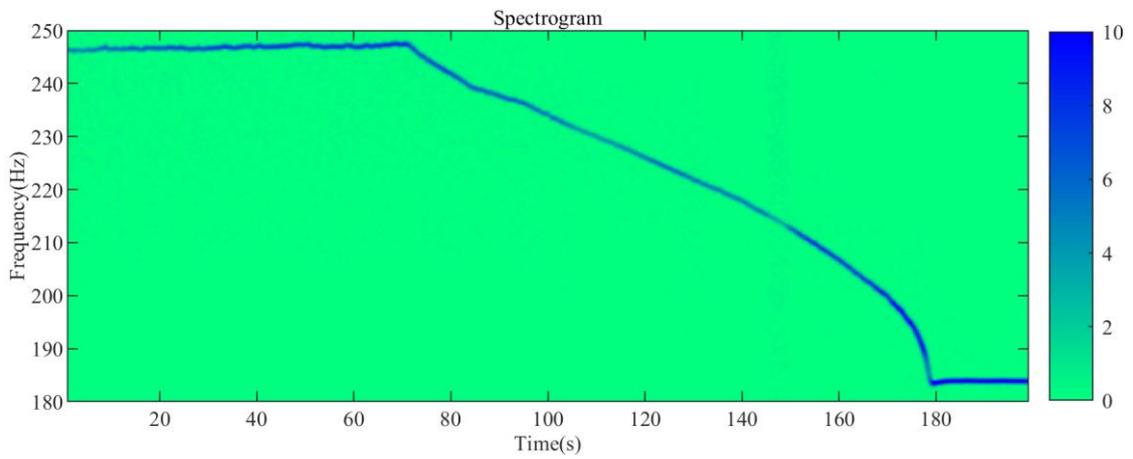


Figure 7: The extracted fundamental component of the rotating speed signal.

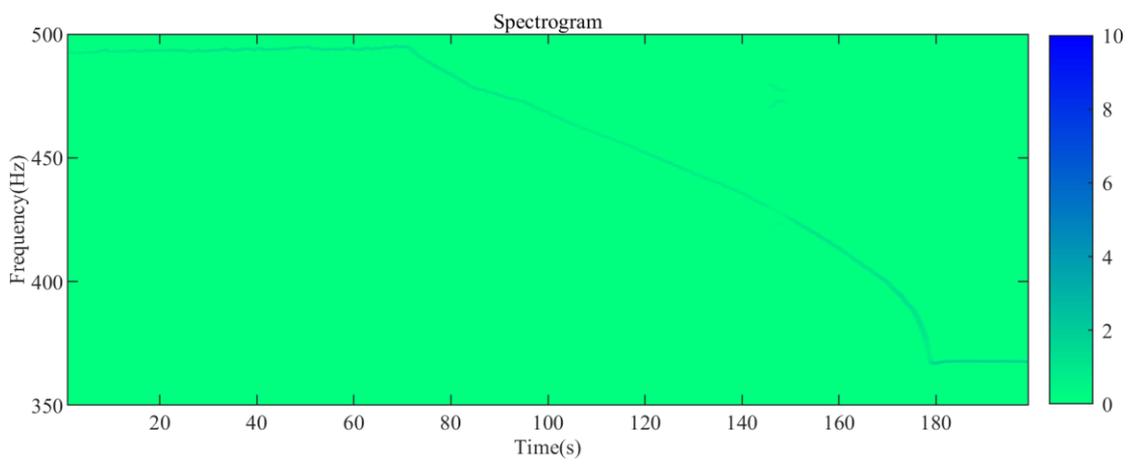


Figure 8: The extracted second harmonic of the rotating speed signal.

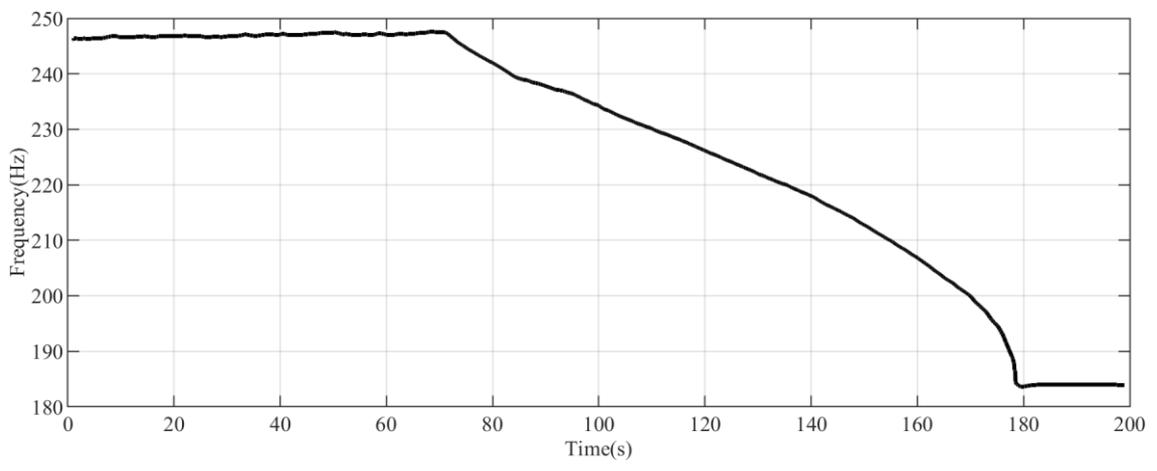


Figure 9: The estimated instantaneous frequency of the rotating speed signal.

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, *Tacholeless 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 tacholeless 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 tacholeless order tracking method for wind turbine planetary gearbox fault detection*, Measurement, 2019 May. vol. 138, pp.266-277.

- [13]Yue Hu, Xiaotong Tu, Fucui Li, Hongguang Li, Guang Meng, *An adaptive and tacholeless 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 JÉZÉQUEL¹, Eric CABROL², Bernard VITRY²

¹LTDS, Ecole Centrale de Lyon, UMR5513, 69134, Ecully Cedex, France
hanwei.gao@ec-lyon.fr

²Technocentre 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

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

where the signum function $\text{sign}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$, 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

$$\begin{aligned} f &= \sigma_0 u \\ \dot{u} &= v \cdot \text{sign}\left(1 - \text{sign}(v) \frac{\sigma_0 u}{F_c}\right) \left|1 - \text{sign}(v) \frac{\sigma_0 u}{F_c}\right|^i \end{aligned} \quad (2)$$

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 $\sigma_0 \rightarrow \infty$ 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.

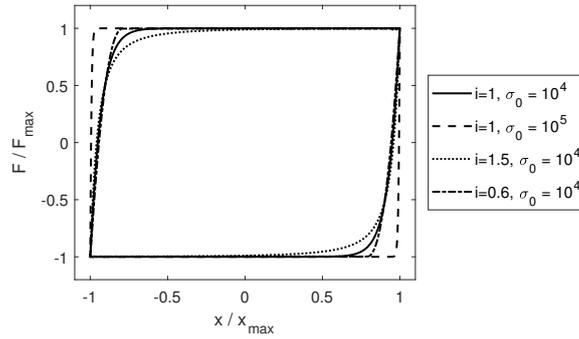


Figure 1 – Comparison of different parameters settings of i and σ_0 of Dahl model

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

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

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.

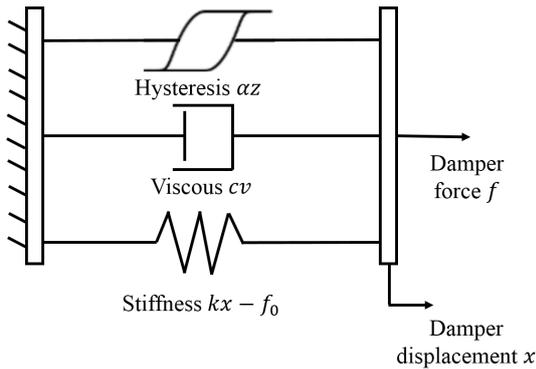


Figure 2 – Construction of the Original Bouc Wen model

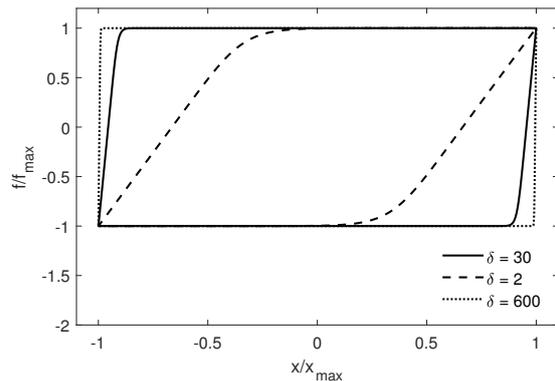


Figure 3 – Influence of δ on Damper force-displacement curve

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

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

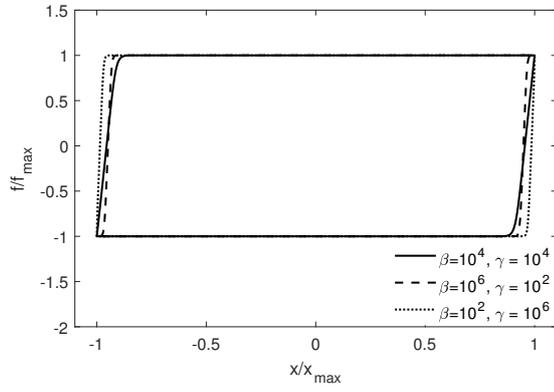


Figure 4 – Influence of β and γ on Damper force-displacement curve

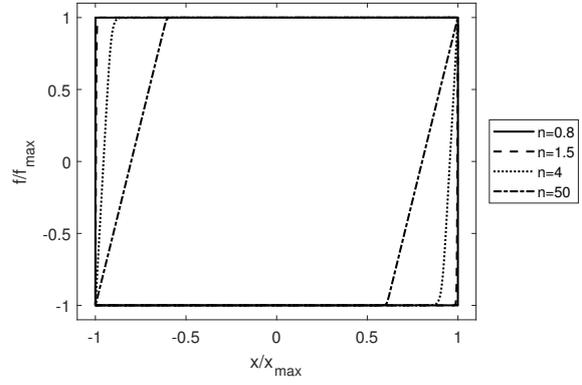


Figure 5 – Influence of n on Damper force-displacement curve

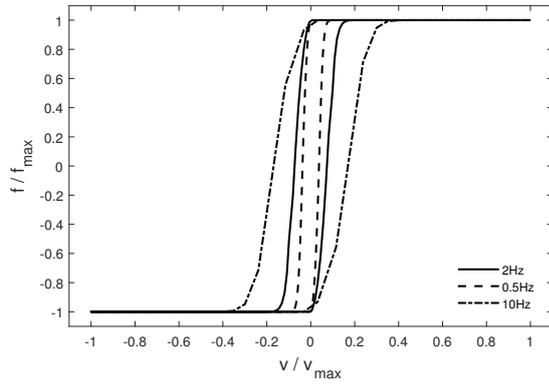


Figure 6 – Hysteresis curve under different excitation frequencies

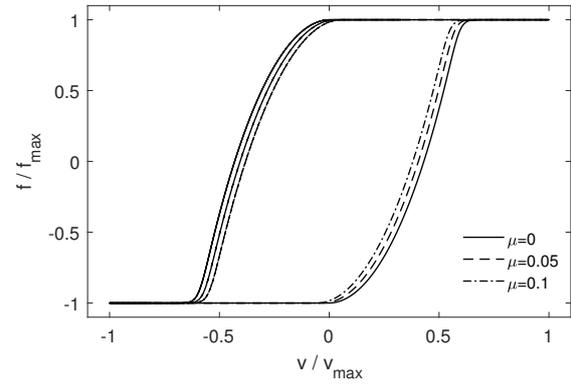


Figure 7 – Influence of μ on Damper force-displacement curve

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

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

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

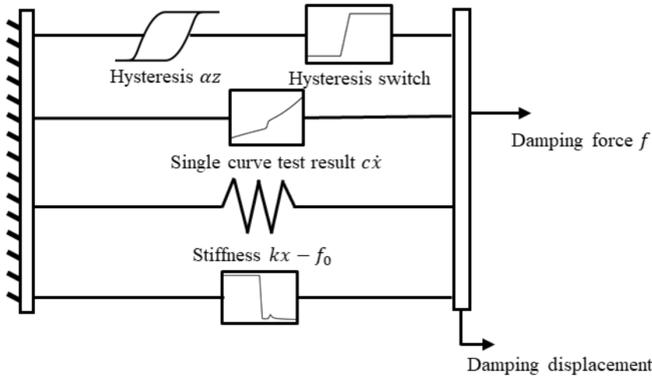


Figure 8 – Construction of a modified Bouc-Wen model

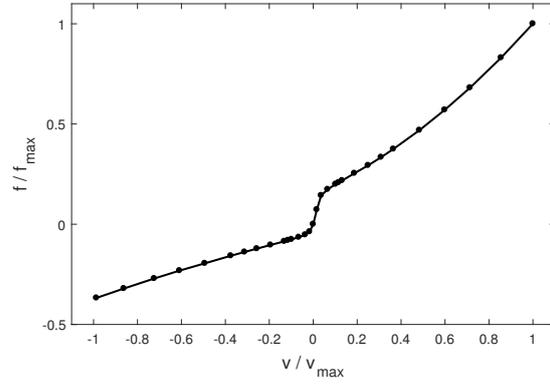


Figure 9 – Example of a single damping curve

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

$$F_{gas} = Ax^2 + Bx + C \quad (6)$$

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.

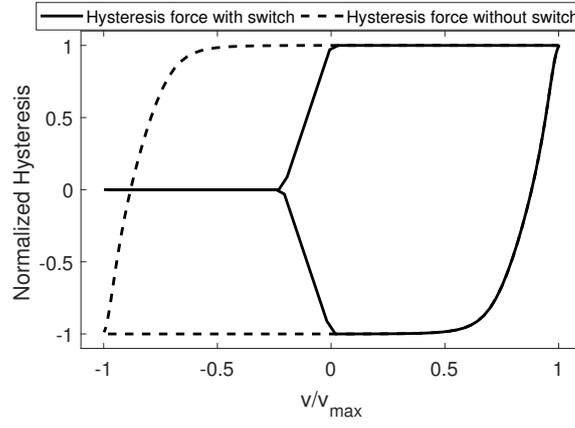


Figure 10 – The hysteresis force-velocity curve after passing the switch

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

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

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, \dots, 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 $e(\mathbf{x})$. 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.

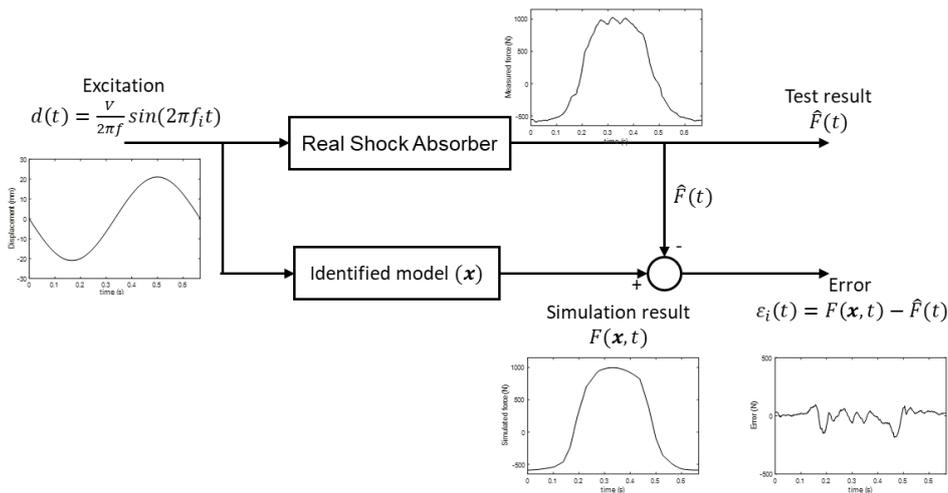


Figure 11 – The hysteresis force-velocity curve after passing the switch

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

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

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].

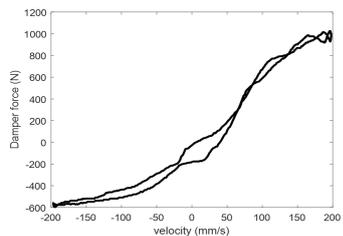
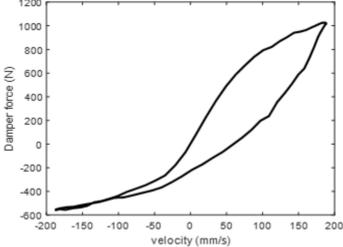
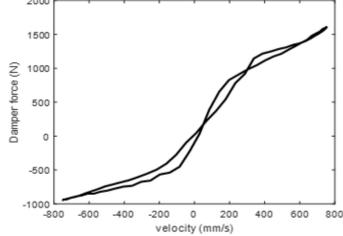
Group	LVLF	LVHF	HVHF
Velocity	0.1-0.4m/s	0.1-0.3m/s	0.5-1.2m/s
Frequency	1.5Hz	12Hz	12Hz
Nb. Signal	6	3	4
Curve example			

Table 1 – Repartition 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.

$$\begin{aligned} e_{LVLF}(\mathbf{x}) &= \frac{1}{6} \sum_{i=1}^6 e_i(\mathbf{x})/e_i(\mathbf{x}_0) \\ e_{LVHF}(\mathbf{x}) &= \frac{1}{3} \sum_{i=7}^9 e_i(\mathbf{x})/e_i(\mathbf{x}_0) \\ e_{HVHF}(\mathbf{x}) &= \frac{1}{4} \sum_{i=10}^{13} e_i(\mathbf{x})/e_i(\mathbf{x}_0) \end{aligned} \quad (9)$$

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

$$\begin{aligned} \forall x \in \mathcal{E}, \forall j \in \{1, \dots, m\}, f_i(x) < f_j(x^*) \\ \Rightarrow \exists i \neq j, f_i(x) > f_j(x^*) \end{aligned} \quad (10)$$

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 \mathbf{x}

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

The initial, minimum and maximum values of these parameters are defined in Table 2. In this example, the gas pressure 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:

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

Under the constraints:

$$x \in DesignSpace \quad (13)$$

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

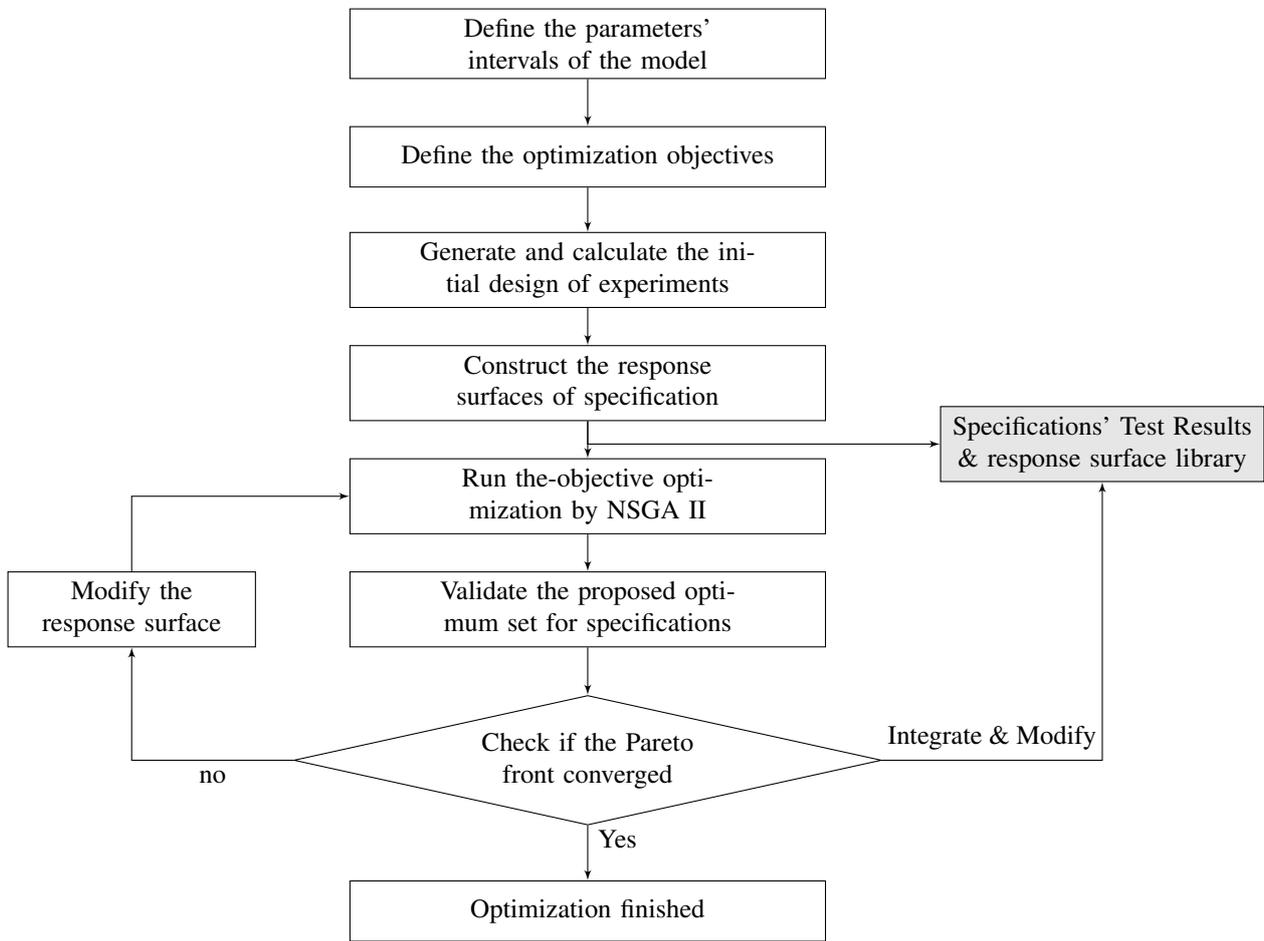


Figure 12 – Optimization process without robustness notions

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.

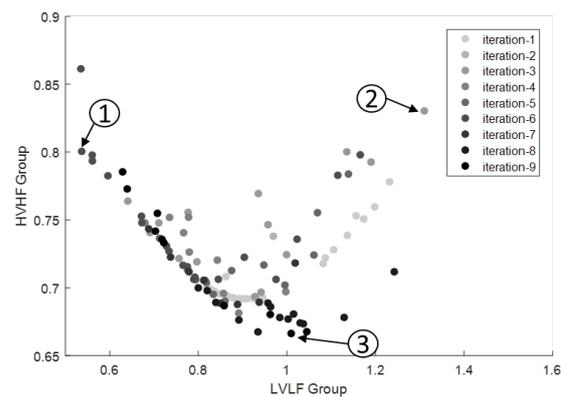
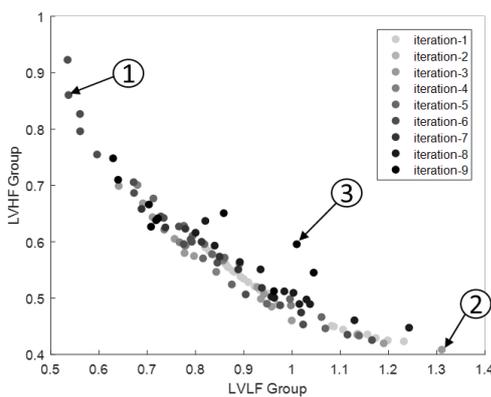


Figure 13 – Non-dominated solutions in LVLG-LVHF plane Figure 14 – Non-dominated solutions in LVLG-HVHF plane

Figure 15 shows the projection of the 3D Pareto surface on the LVHF-HVHF plan. The darker colour means better correction to LVLG 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

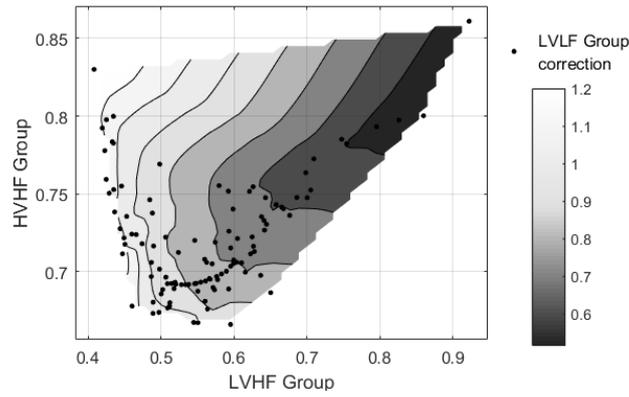


Figure 15 – Projection of Pareto surface on in LVHF-HVHF plot

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

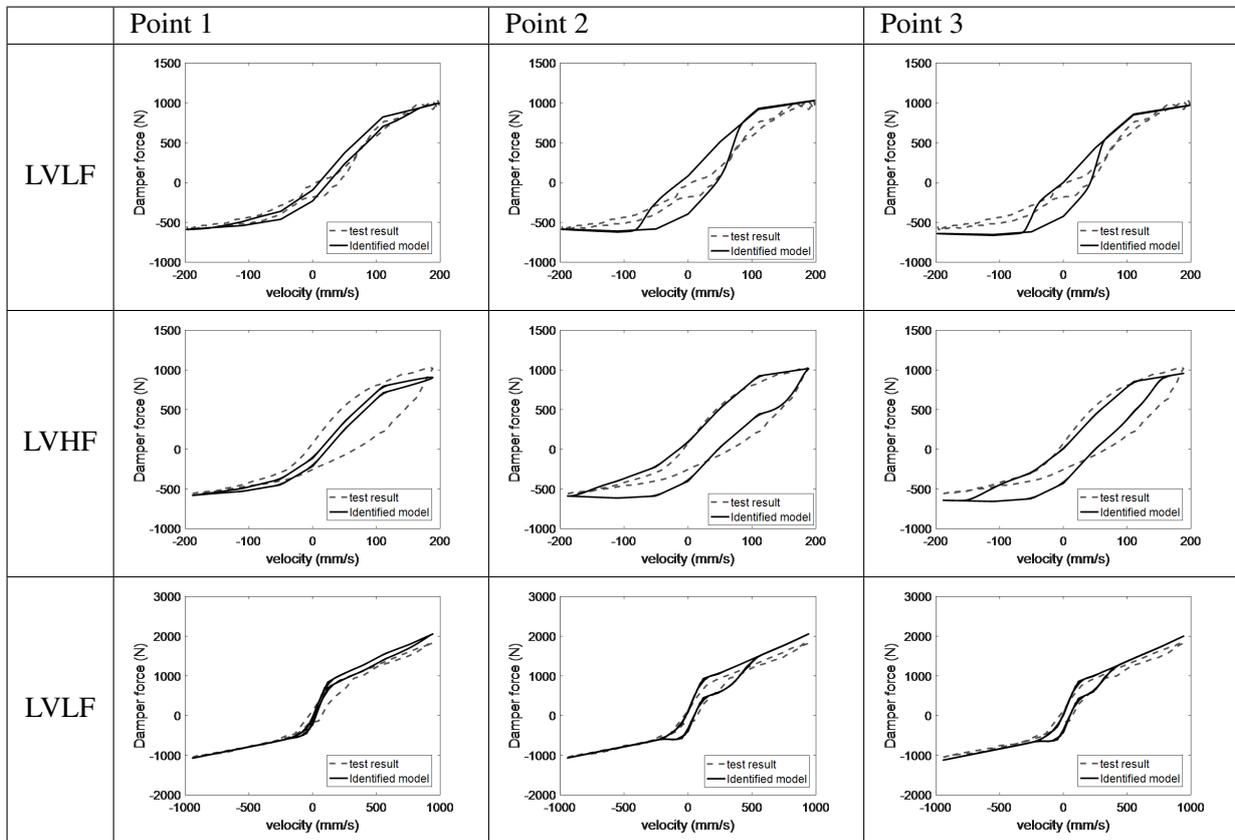


Table 4 – Correction of test results from choice of different orientations (Point 1: Best correction for low frequency. Point 2 and 3: Best corrections for high frequency)

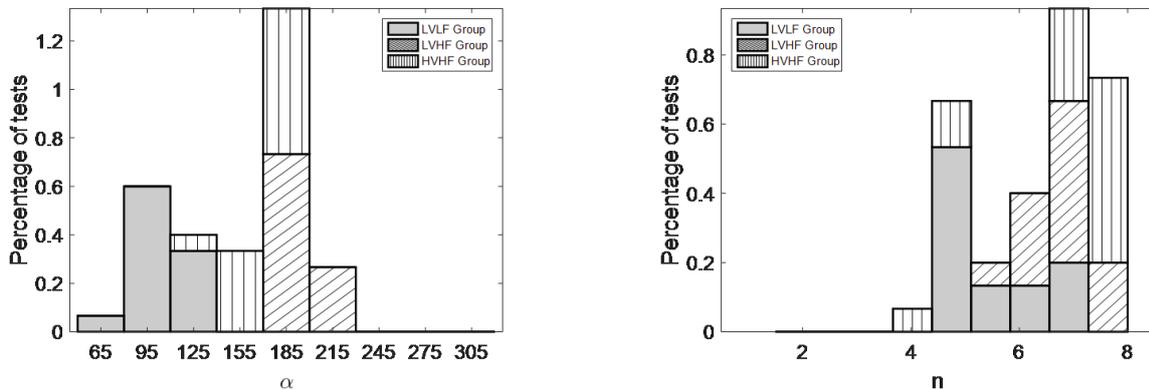


Figure 16 – Resumes of α (left) and n (right) of 15 best corrections in each group

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.

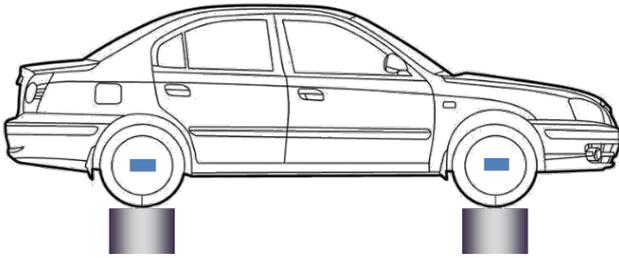


Figure 17 – The full vehicle comfort test bench

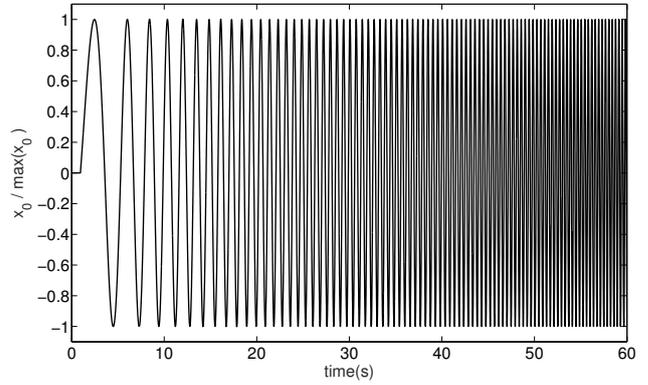


Figure 18 – One part of excitation signal

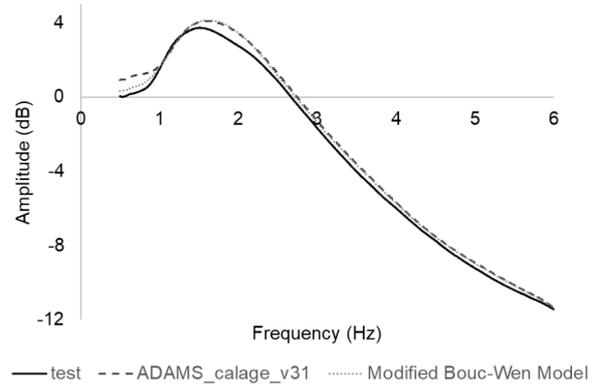
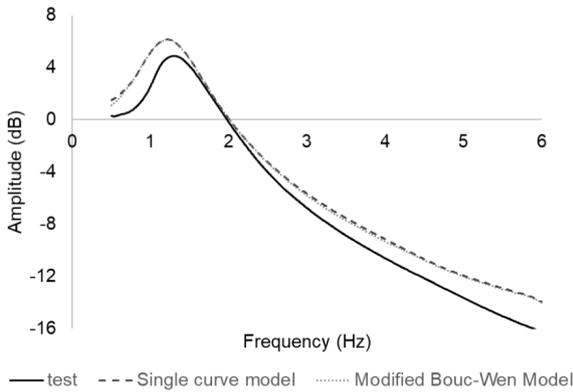


Figure 19 – Comparison of transfer function for front (left) and rear (right) suspensions

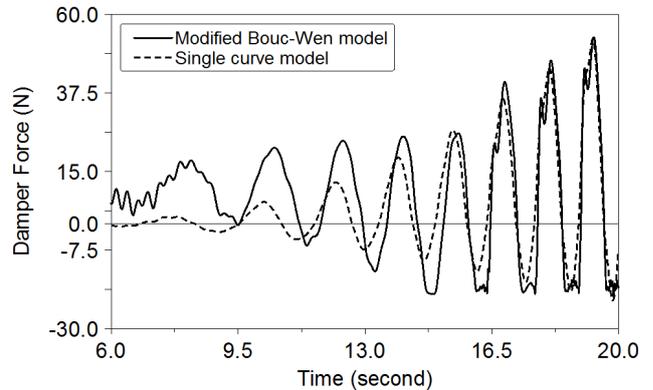
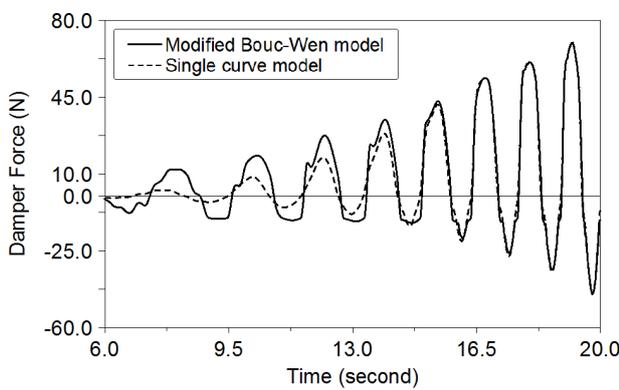


Figure 20 – Comparison of transfer function for front (left) and rear (right) suspensions

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 d'Expertise, 2018.

Modal identification of machining robots in service

Asia MAAMAR¹, Vincent GAGNOL¹, Thien-Phu LE², Laurent SABOURIN¹

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.

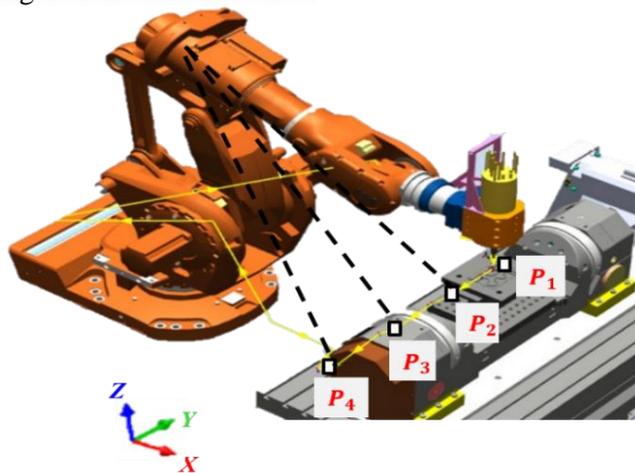


Figure 1: Robot Tool Center Point (TCP) positions along the Y direction in the robot's workspace

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 .

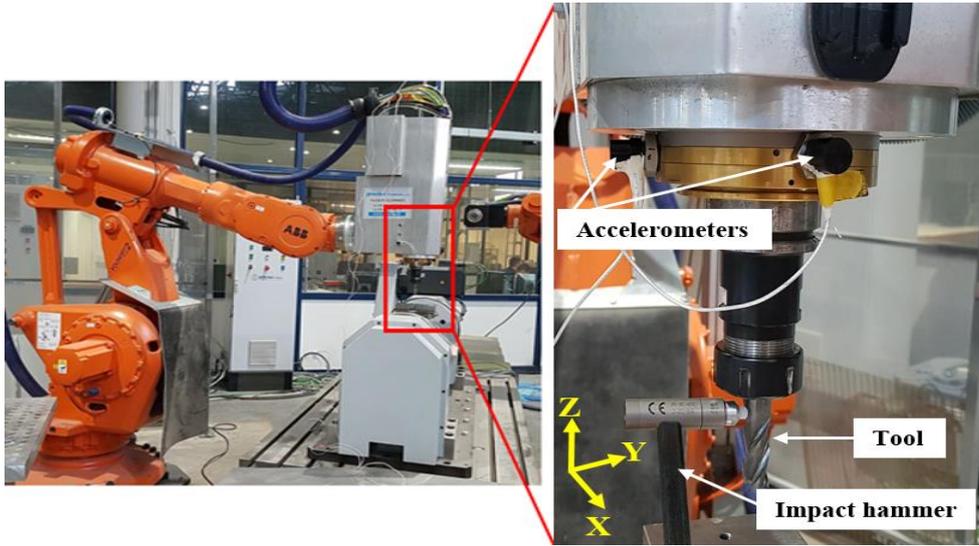


Figure 2: Experimental protocol for an impact testing on an ABB IRB 6660 machining robot

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

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

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_{yy} . 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	1 st position (P_1)		2 nd position (P_2)		3 rd position (P_3)		4 th position (P_4)	
	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:

$$\zeta = \frac{\sigma * 100}{\mu} \quad (2)$$

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

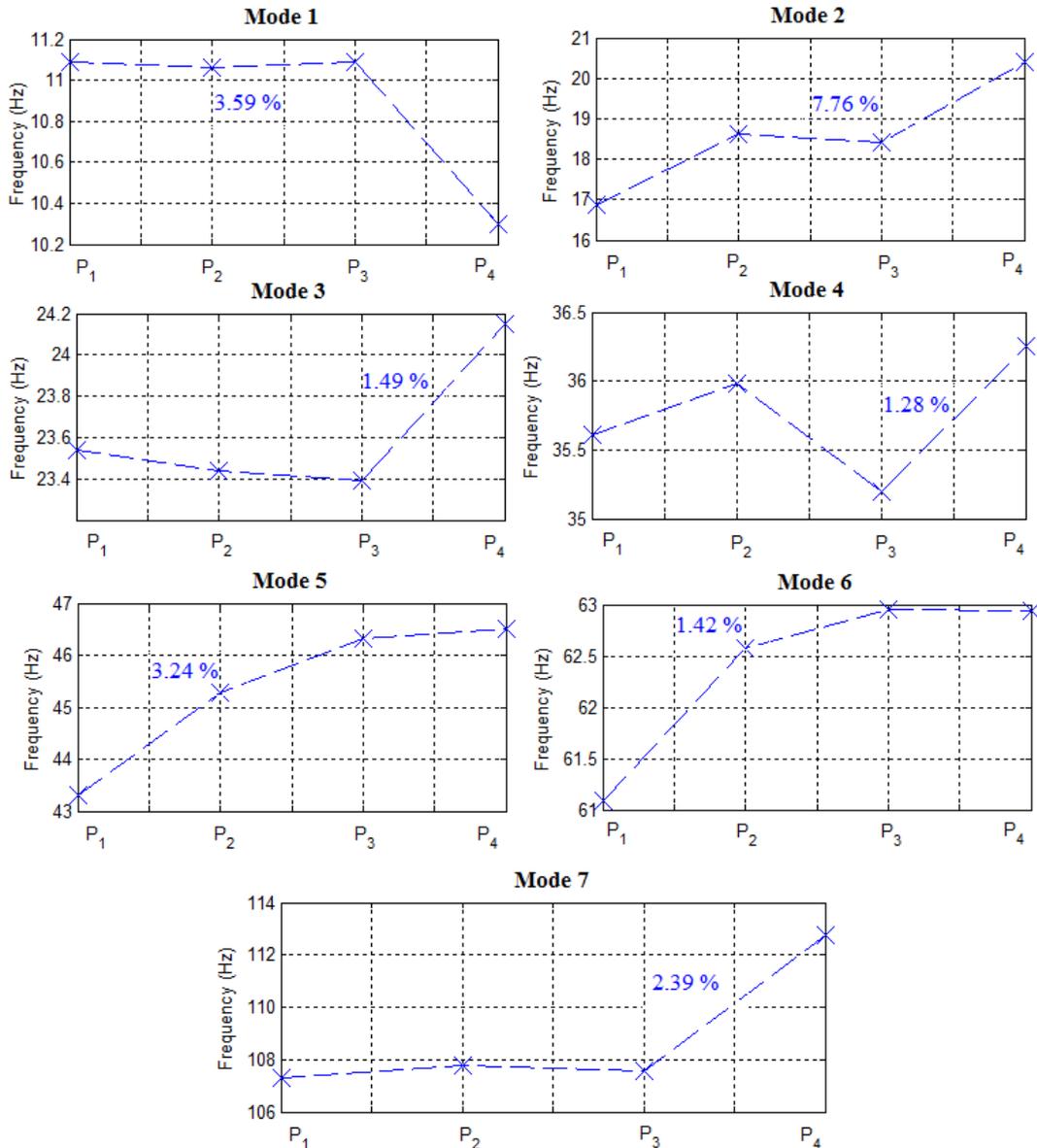


Figure 3: Evolution of the natural frequencies of the robot ABB IRB 6660 in function of its position

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.

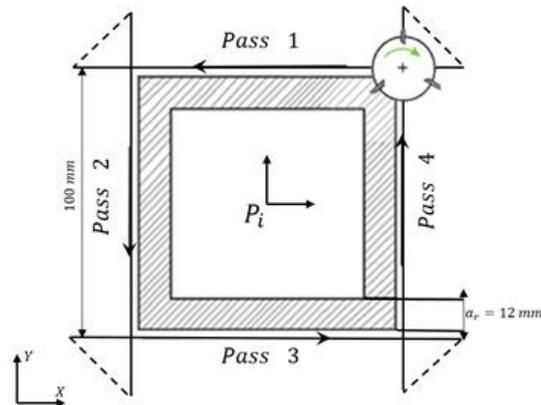


Figure 4: Machining operation when robot is in position 1 and configuration 1

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.

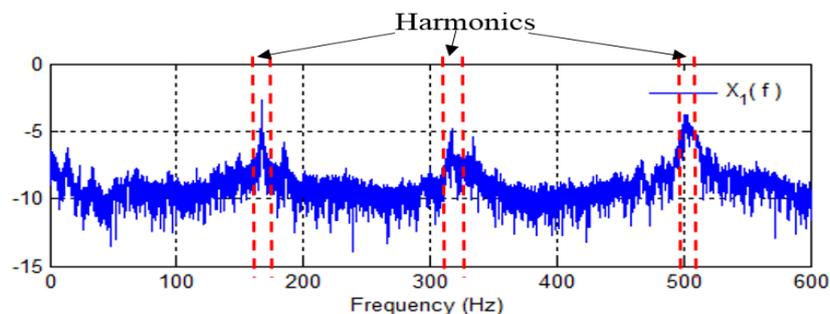


Figure 5: FFT of the measured acceleration in the x direction in pass 1 and presence of harmonics associated to the machining process

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 .

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

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

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

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

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

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:

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

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.

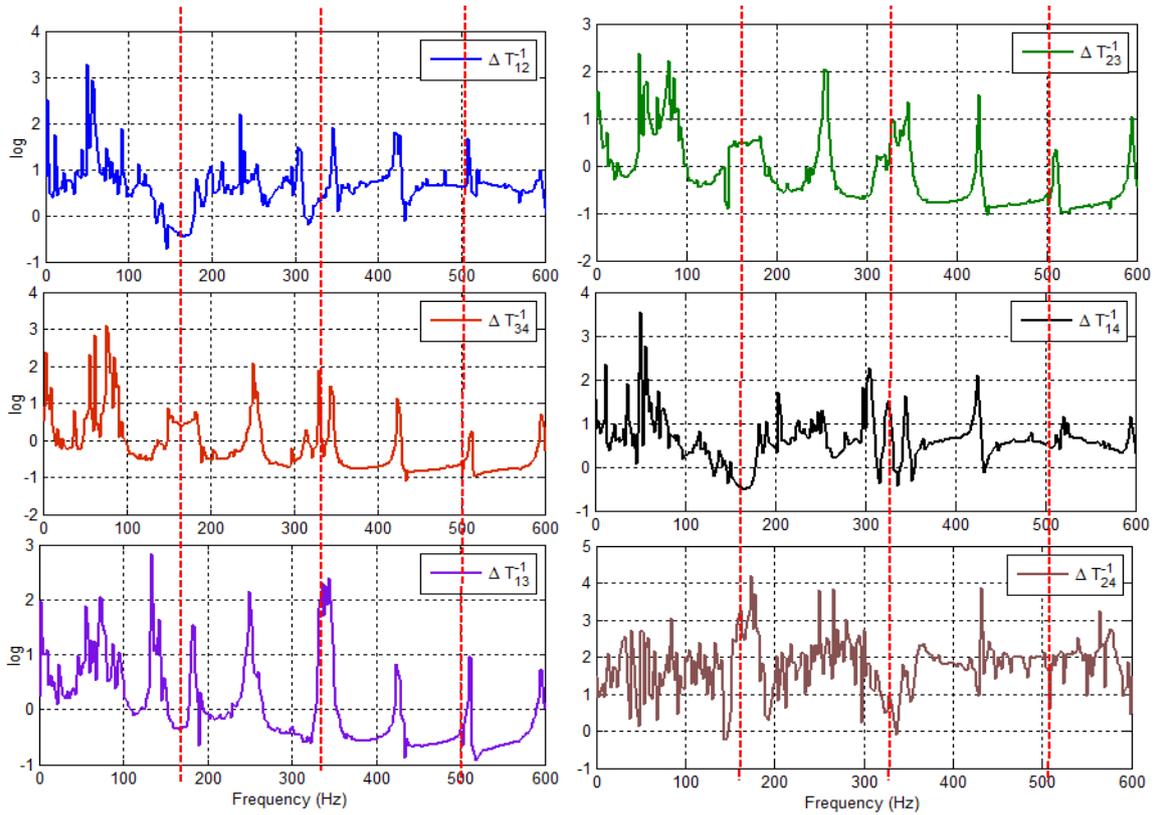


Figure 6: Rational transmissibility function ΔT_{ij}^{-1} and presence of harmonic components

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.

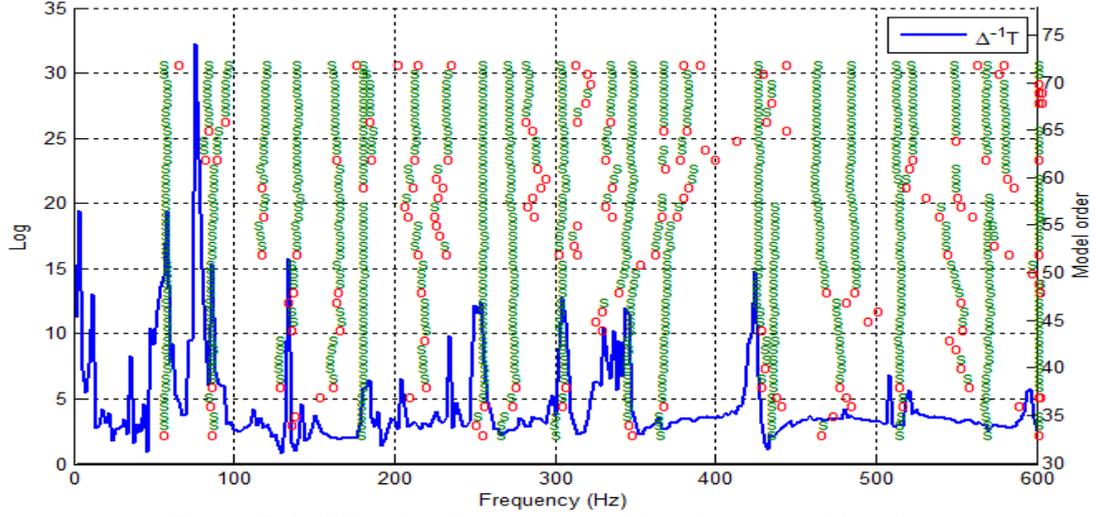


Figure 7: Stabilization diagram (s : stable pole, o : unstable pole)

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 \mathbf{T} is considered, and a singular value decomposition is performed.

$$\mathbf{T} = \begin{bmatrix} T_{xy}^1 & T_{xy}^2 \\ T_{xy}^3 & T_{xy}^4 \end{bmatrix} \quad (8)$$

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.

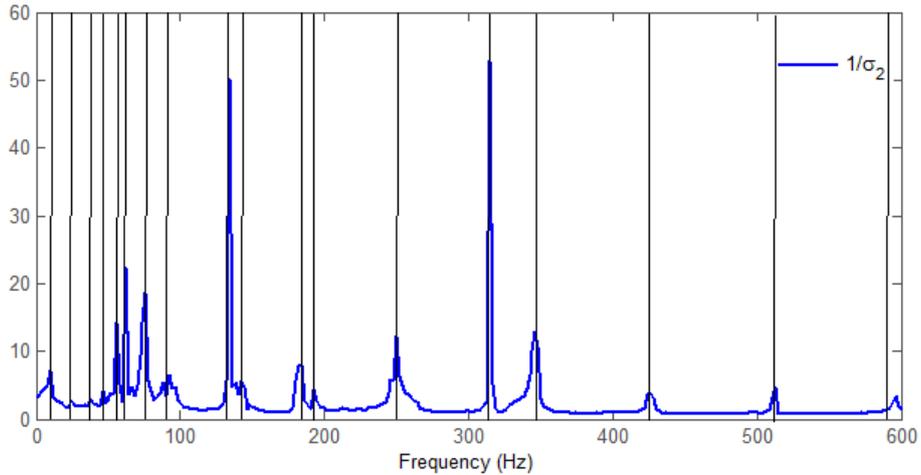


Figure 8: Selection of the system's poles by means of singular value decomposition

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.

Mode	AME		AMO	
	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*, 21^{ème} 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 Identification 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 Liu¹, Xiaodong Zhang^{1,2}, Yiwei Xiong¹, Siying Xie¹,

¹School of Mechanical Engineering - Xi'an Jiaotong University

²Key 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

- (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 relies 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.

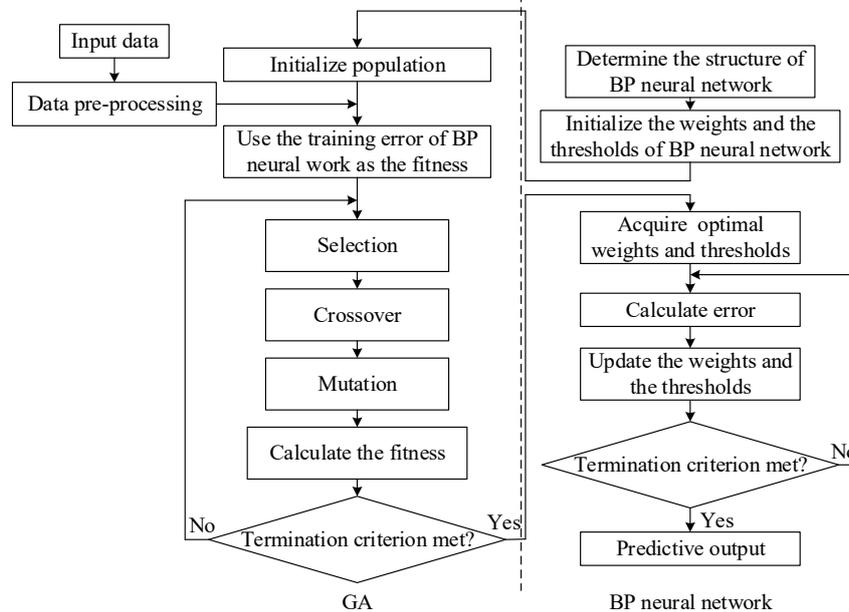


Figure 2 Flow chart of BP neural network optimized by GA

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 $+15^\circ$ 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.

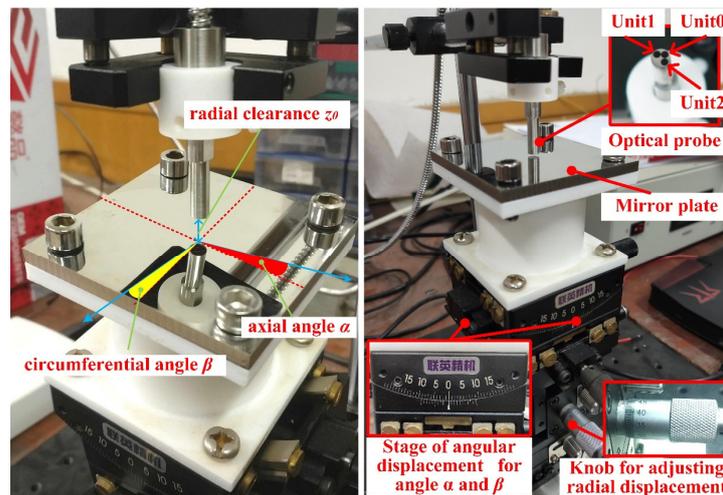


Figure 3 Calibration table for the 3-D tip clearance

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, unit1 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.

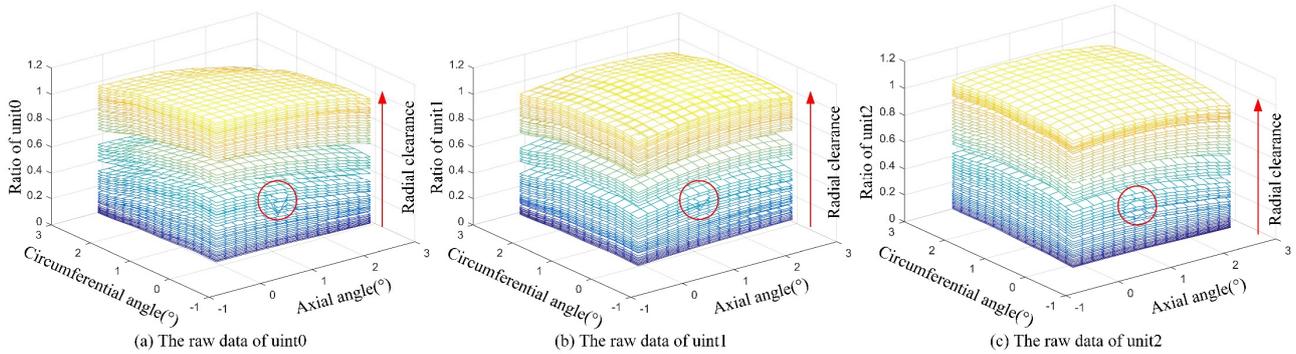


Figure 4 The raw data collected by the calibration experiments

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.

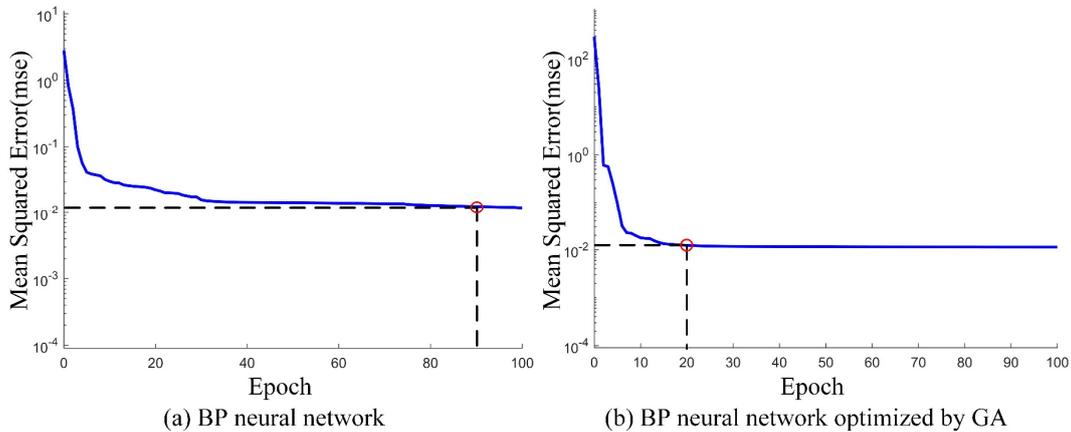


Figure 5 Neural network training error curve

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.

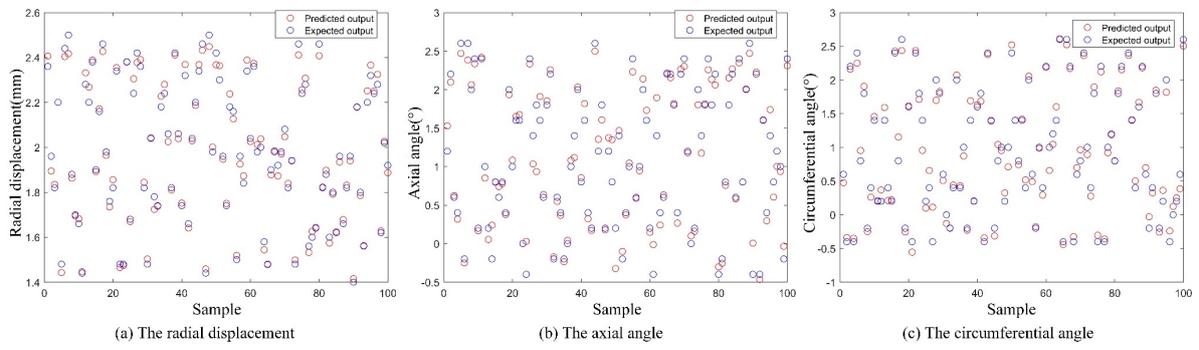


Figure 6 Comparison of the predicted output and the expected output of the 3-D tip clearance

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.

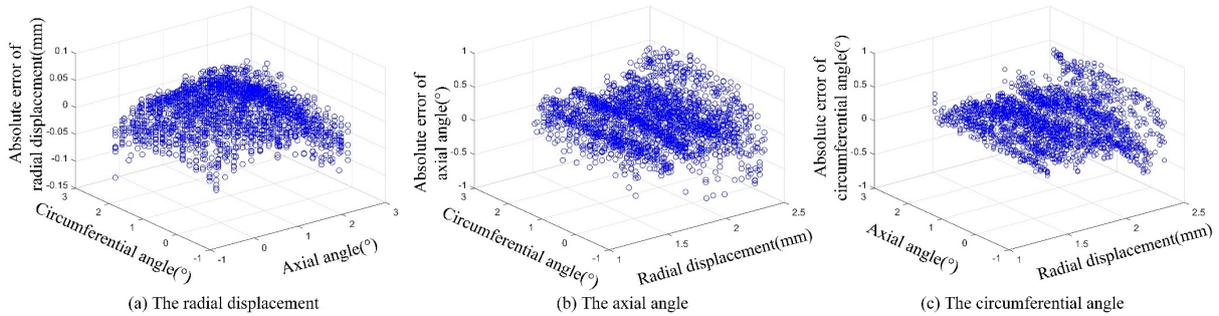


Figure 7 Demodulation error distribution of the 3-D tip clearance

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 XIONG¹, Xiaodong ZHANG^{1,2}, Siying XIE¹, Hongcheng LIU¹

¹School of Mechanical Engineering, Xi'an Jiaotong University, No.28 Xianning West Road, 710049, Xi'an, Shaanxi, P.R. China

²Key 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.

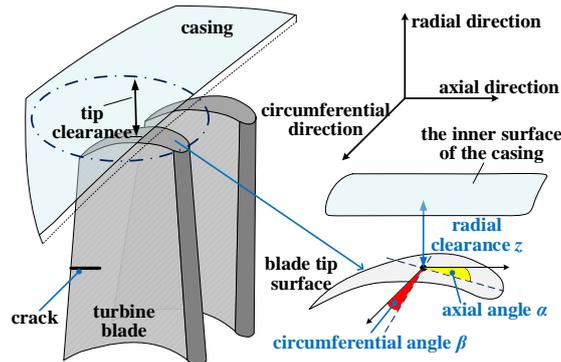


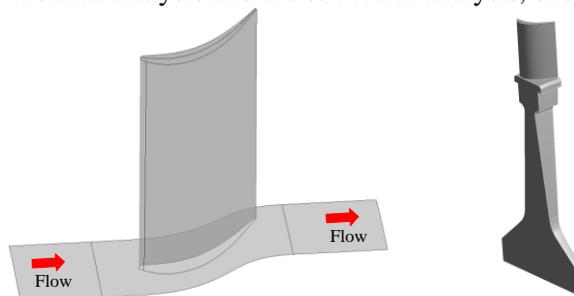
Figure 1: The 3-D tip clearance of a cracked turbine blade

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).

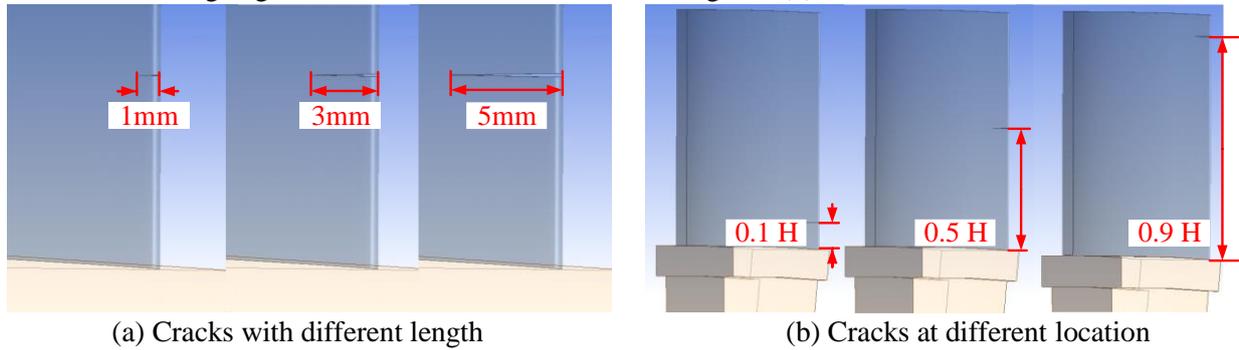


Figure 3: Different cracks at the trailing edge of the turbine blade

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.

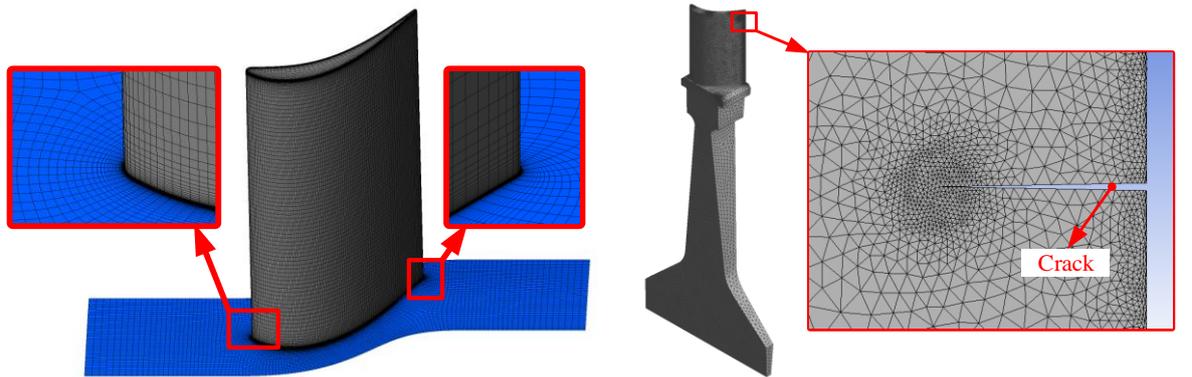


Figure 4: The computational grids of the fluid domain and the turbine blisk

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 $h = q(T_w - T_{aw})$, 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.

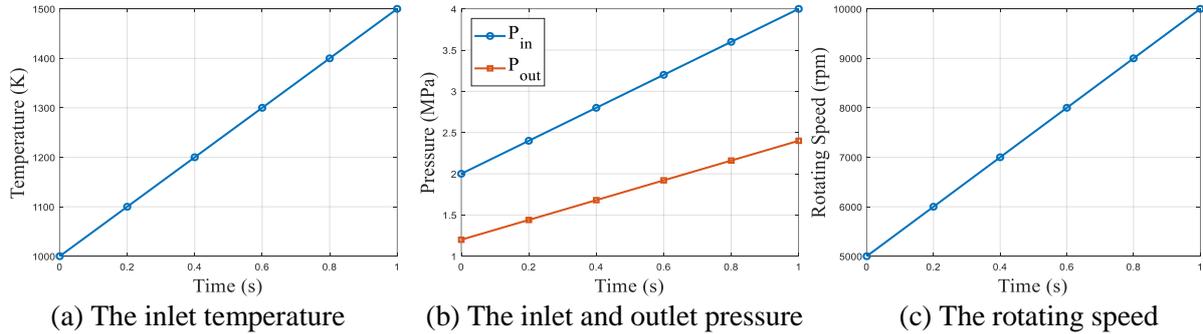


Figure 5: Time-varying loads in the aero-engine acceleration process

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.

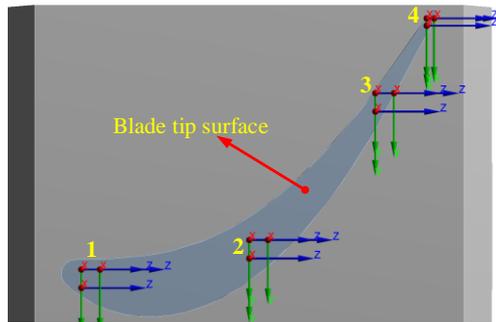


Figure 6: Monitoring points on the blade tip surface

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.

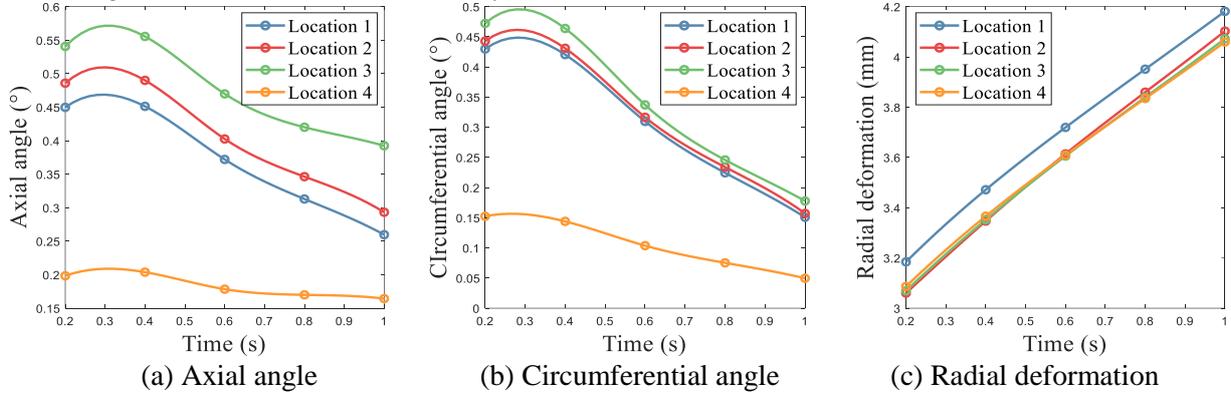


Figure 7: The 3-D tip clearance of a cracked blade in the acceleration process

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.

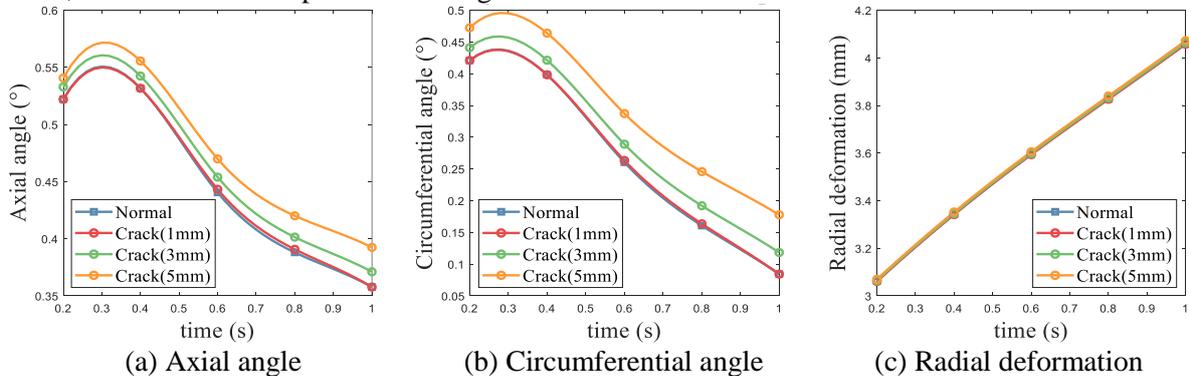


Figure 8: The 3-D tip clearance of the fault blades with different crack length at the monitoring location 3

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.

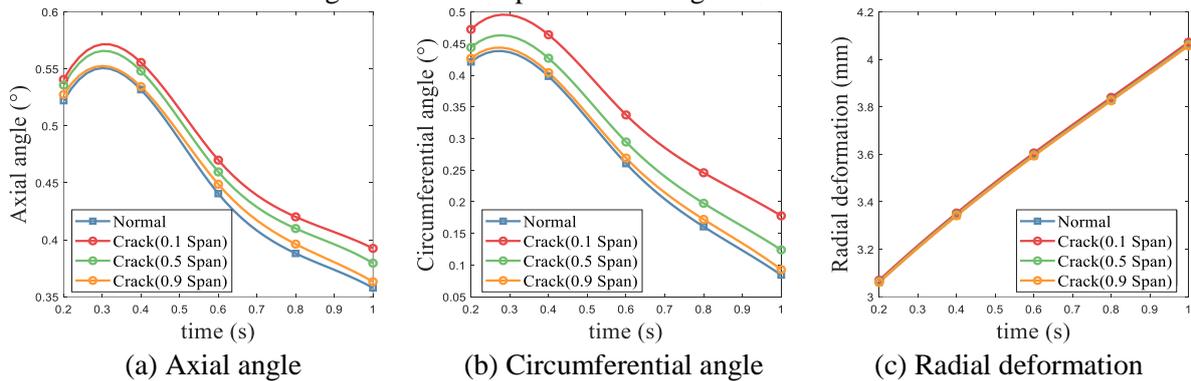


Figure 9: The 3-D tip clearance of the fault blades with different crack location at the monitoring location 3

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 angle 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 angle 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 SILVA¹

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

Leonardo SANCHES^{2*}, Guilhem MICHON²⁺

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

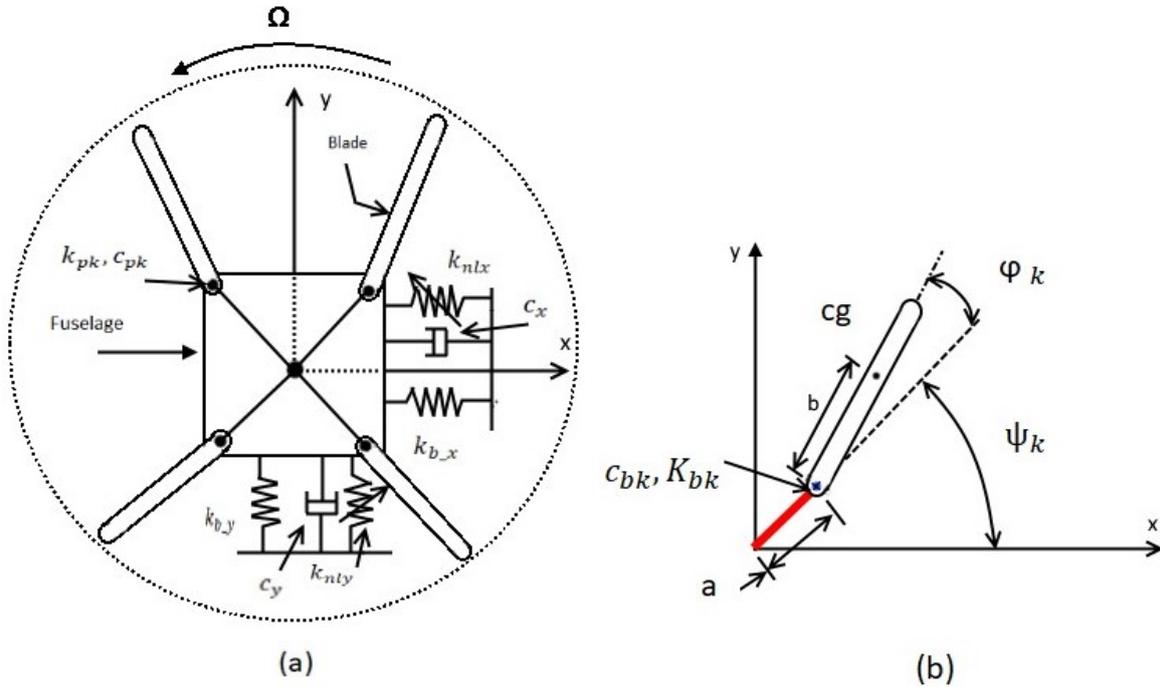


Figure 1 – Sketch of Helicopter Mechanical Model

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:

$$\mathbf{M}(t)\ddot{\mathbf{u}} + \mathbf{G}(t)\dot{\mathbf{u}} + \mathbf{K}(t)\mathbf{u} = \mathbf{F}_{nl} \quad (1)$$

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:

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

with,

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

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

$$\mathbf{M}_C \ddot{\mathbf{q}} + \mathbf{G}_C \dot{\mathbf{q}} + \mathbf{K}_C \mathbf{q} = \mathbf{F}_{nl} \quad (4)$$

where,

$$\mathbf{M}_C = \begin{bmatrix} m_f + N_b m_b & 0 & 0 & 0 & -2 b m_b & 0 \\ 0 & m_f + N_b m_b & 0 & 0 & 0 & 2 b m_b \\ 0 & 0 & N_b (m_b b^2 + I_{zb}) & 0 & 0 & 0 \\ 0 & 0 & 0 & N_b (m_b b^2 + I_{zb}) & 0 & 0 \\ -2 b m_b & 0 & 0 & 0 & N_b/2 (m_b b^2 + I_{zb}) & 0 \\ 0 & 2 b m_b & 0 & 0 & 0 & N_b/2 (m_b b^2 + I_{zb}) \end{bmatrix} \quad (5)$$

$$\mathbf{G}_C = \begin{bmatrix} c_x & 0 & 0 & 0 & 0 & 0 \\ 0 & c_y & 0 & 0 & 0 & 0 \\ 0 & 0 & N_b c_b & 0 & 0 & 0 \\ 0 & 0 & 0 & N_b c_b & 0 & 0 \\ 0 & 0 & 0 & 0 & N_b/2 c_b & -N_b (m_b b^2 + I_{zb}) \Omega \\ 0 & 0 & 0 & 0 & N_b (m_b b^2 + I_{zb}) \Omega & N_b/2 c_b \end{bmatrix} \quad (6)$$

$$\mathbf{K}_C = \begin{bmatrix} k_{fx} & 0 & 0 & 0 & 0 & 0 \\ 0 & k_{fy} & 0 & 0 & 0 & 0 \\ 0 & 0 & N_b (k_b + \Omega^2 a b m_b) & 0 & 0 & 0 \\ 0 & 0 & 0 & N_b (k_b + \Omega^2 a b m_b) & 0 & 0 \\ 0 & 0 & 0 & 0 & k_\tau & -c_b \Omega N_b/2 \\ 0 & 0 & 0 & 0 & c_b \Omega N_b/2 & k_\tau \end{bmatrix} \quad (7)$$

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,

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

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

Property	Value	Units
m_f	2902	kg
m_b	31.9	kg
I_{z_b}	259	kg m ²
a	0.2	m
b	2.5	m
ω_x	6π	rad/s
$\omega_{bk} (k = 1 \dots 4)$	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.

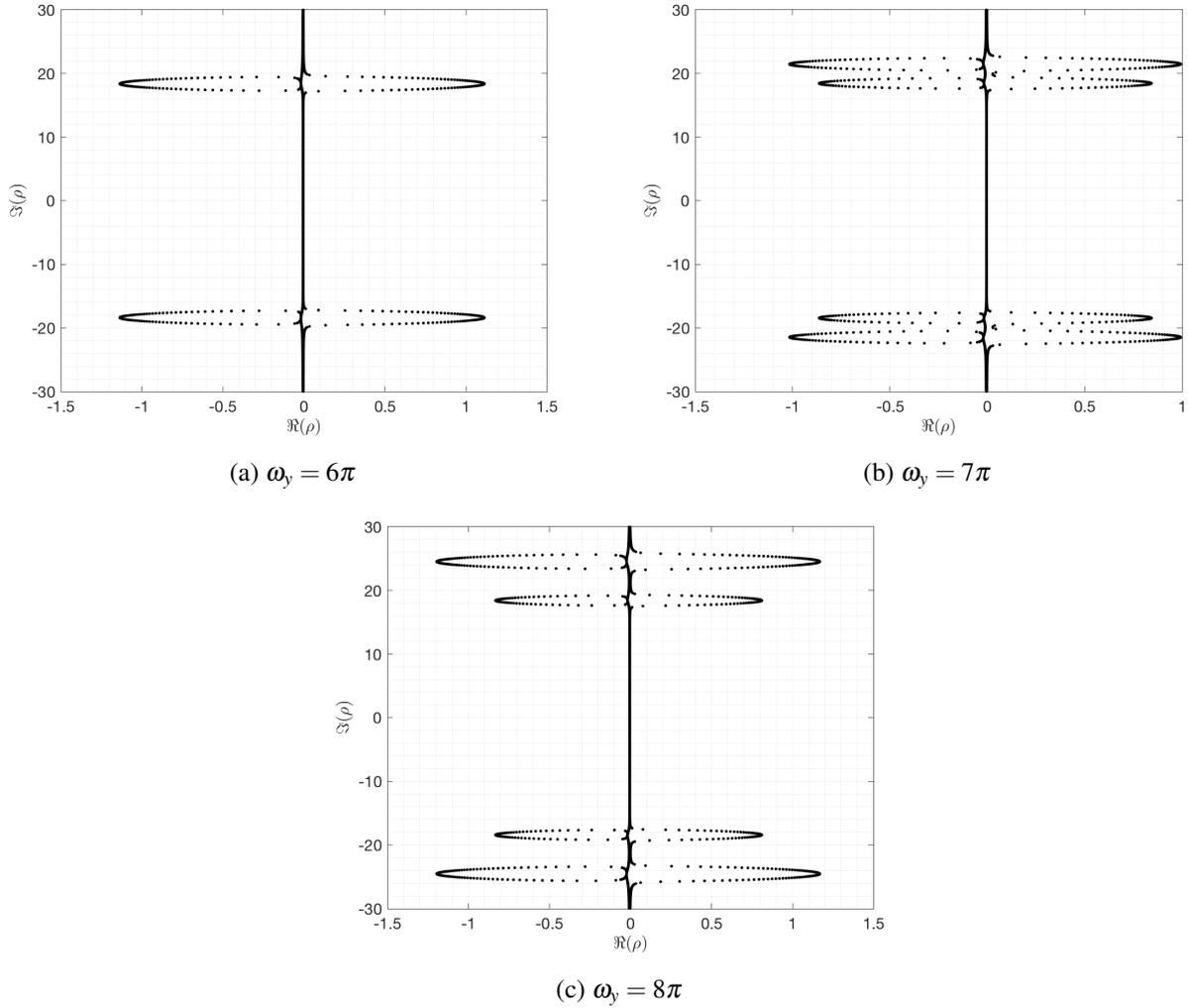


Figure 2 – Eigenvalue Evolution with respect to rotor speed and fuselage configurations.

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., $\Re(\rho) > 0$) 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

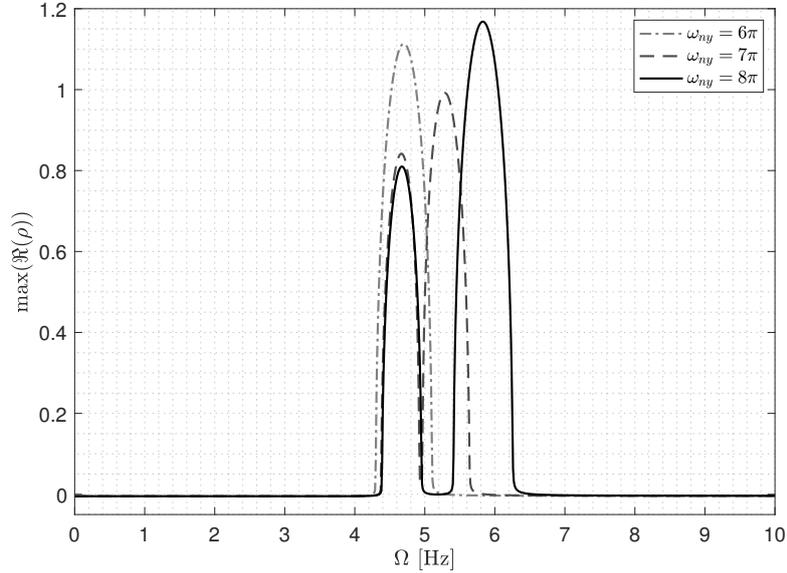


Figure 3 – Instability boundaries for different fuselage configurations

$\omega_y = 6\pi$ 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 $k_{nlx} = k_{nly} = 1 \times 10^5 N/m^3$.

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.

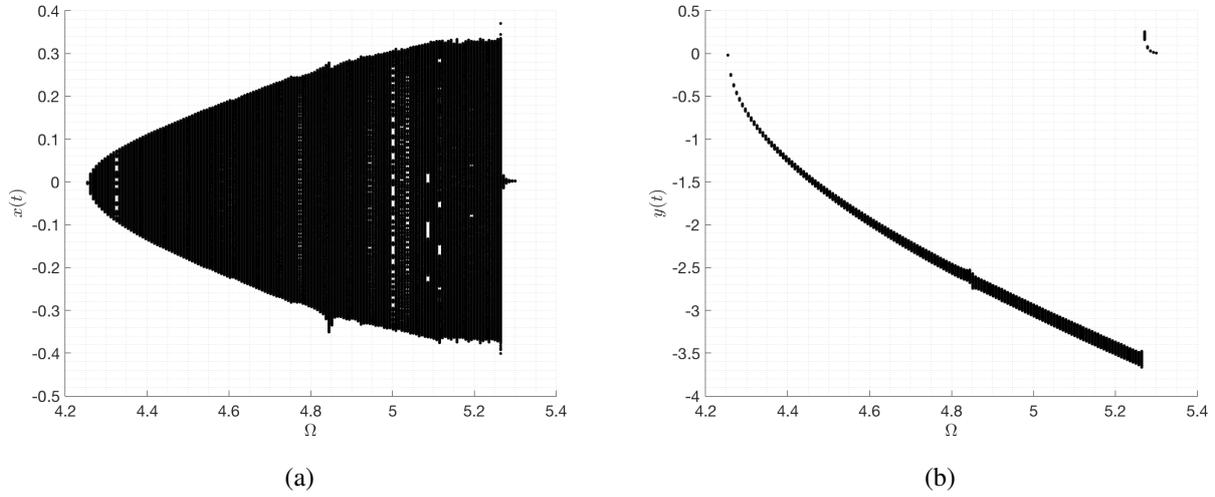


Figure 4 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 6\pi$: a) $x(t)$ and b) $y(t)$

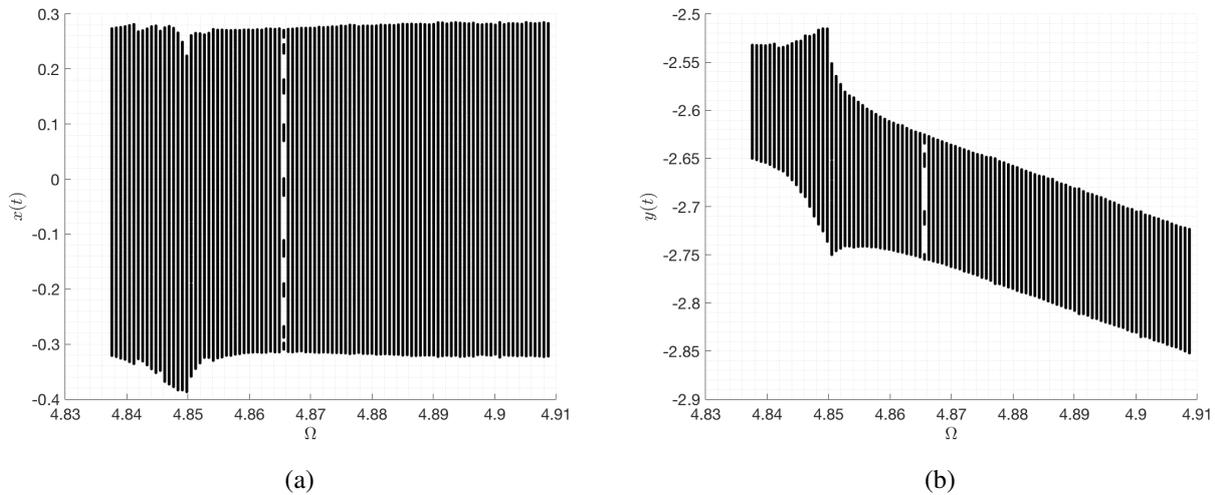
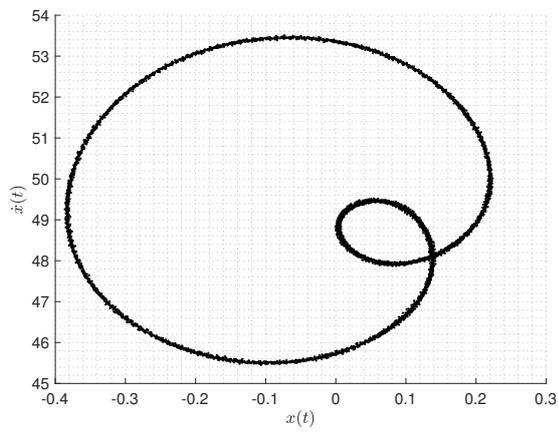


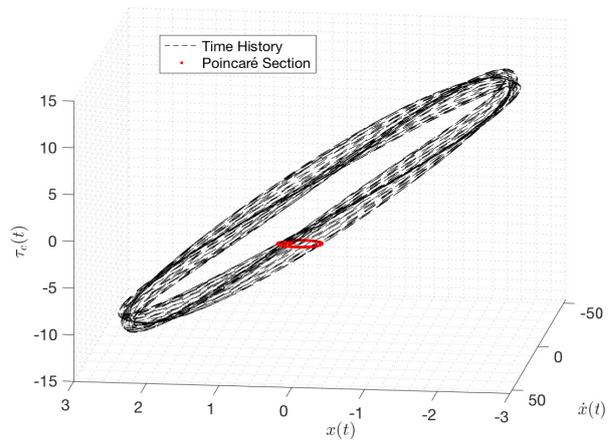
Figure 5 – Detailed Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 6\pi$: a) $x(t)$ and b) $y(t)$

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., $\omega_x = 6\pi$

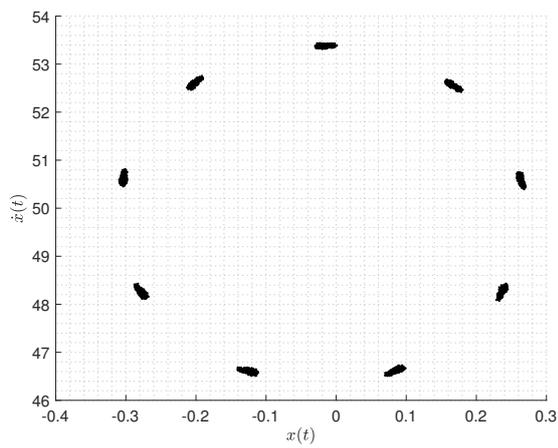


(a) Poincaré Section

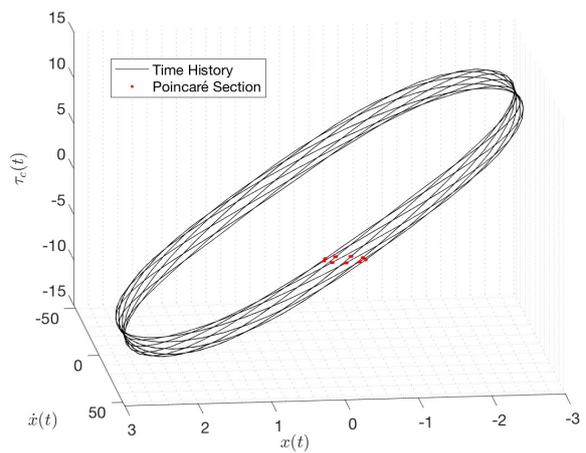


(b) Phase Subspace

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 $\omega_y = 6\pi$

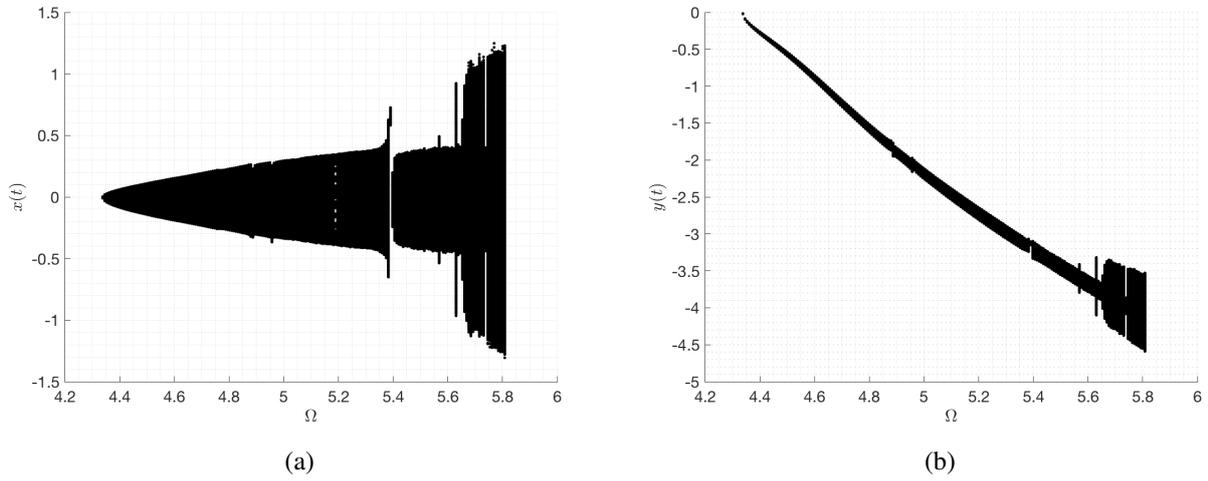


Figure 8 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 7\pi$: a) $x(t)$ and b) $y(t)$

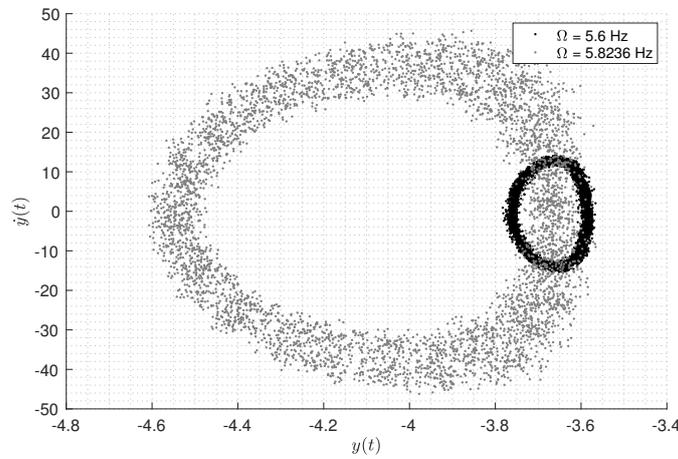


Figure 9 – Comparison of Poincaré Sections for Helicopter with $\omega_x = 6\pi$ and $\omega_y = 7\pi$

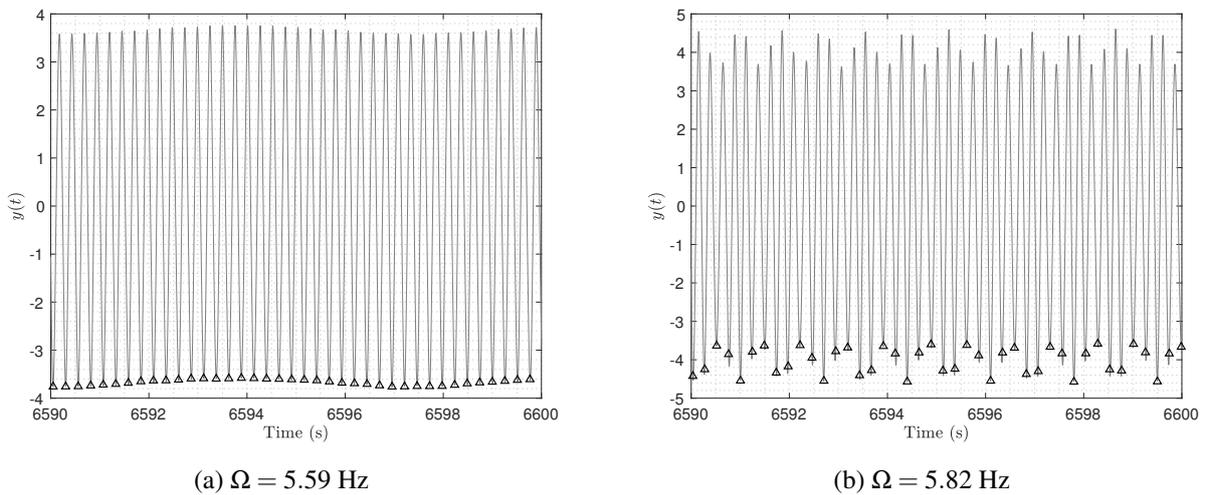


Figure 10 – Temporal Responses of the helicopter with $\omega_x = 6\pi$ and $\omega_y = 7\pi$ - Δ Poincaré section points

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.

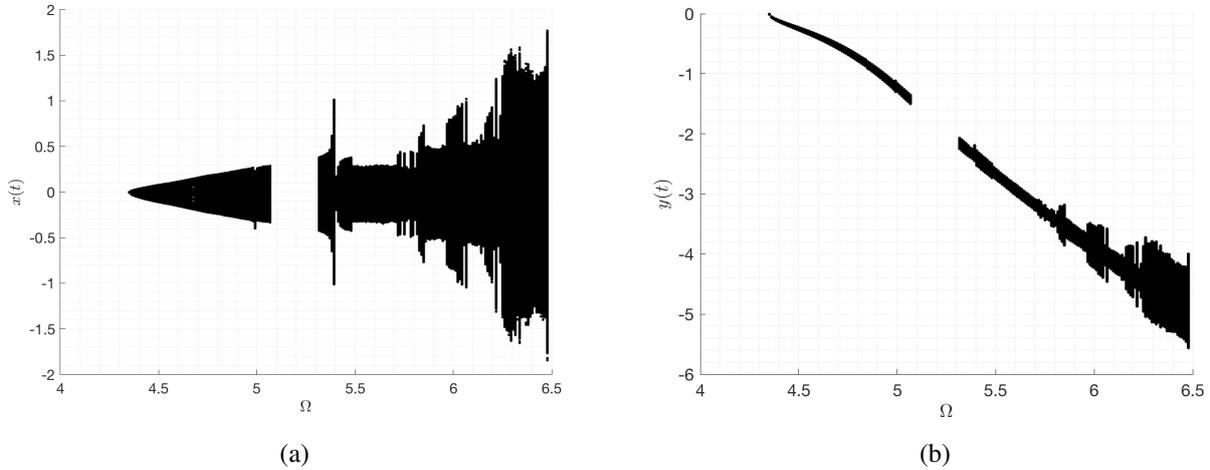


Figure 11 – Bifurcation Diagram for $\omega_x = 6\pi$ and $\omega_y = 8\pi$: a) $x(t)$ and b) $y(t)$

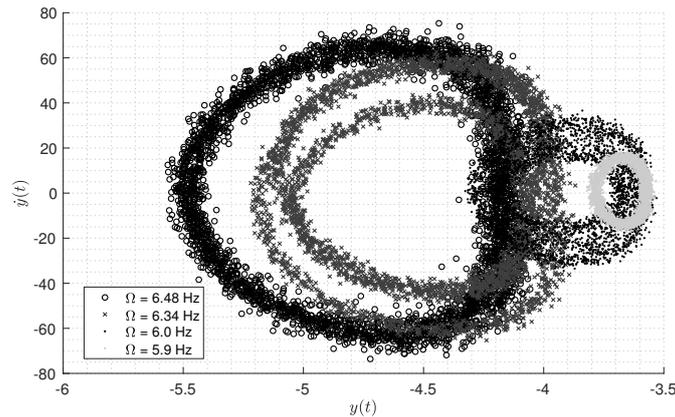


Figure 12 – Comparison of Poincaré Sections of a Helicopter with $\omega_x = 6\pi$ and $\omega_y = 8\pi$

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 rotor-stator 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 ROUSSEU³

¹EDF R&D, EDF Lab Paris-Saclay, ERMES Department, F-91140 Palaiseau, France

²IMSIA, UMR EDF-CEA-ENSTA-CNRS 9219, F-91140 Palaiseau, France

sylvie.audebert@edf.fr

EDF DT, F-69363 Lyon, France

damien.rousseau@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 equipments

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:

$$F_{Htotal} = \sqrt{F_x^2 + F_y^2} \quad (2)$$

Total tearing load simply is:

$$F_{Ztotal} = F_z \quad (3)$$

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%.

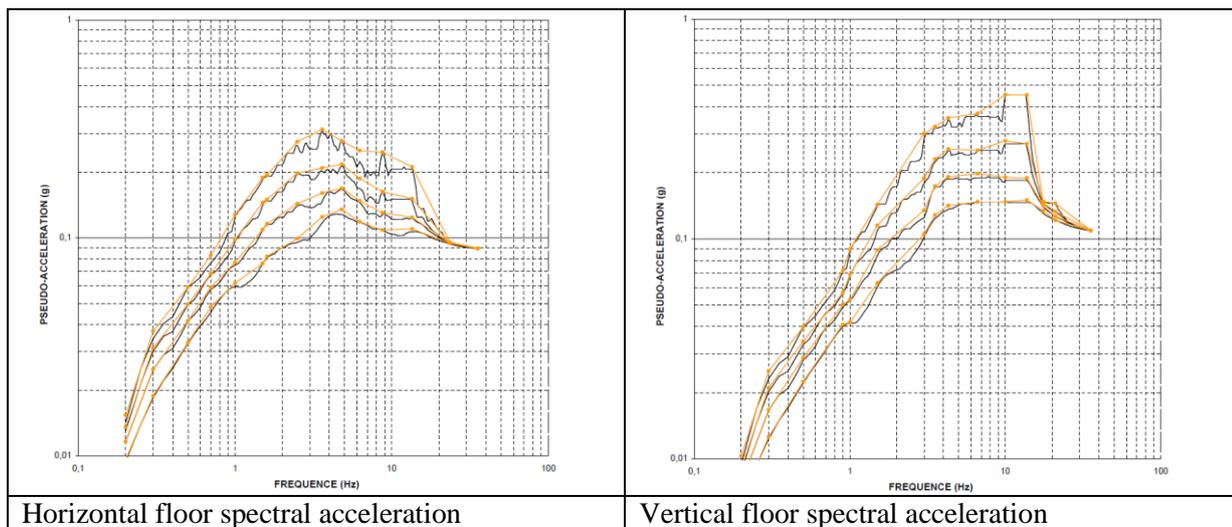


Figure 1: Horizontal and vertical floor spectral accelerations

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 stucked 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).

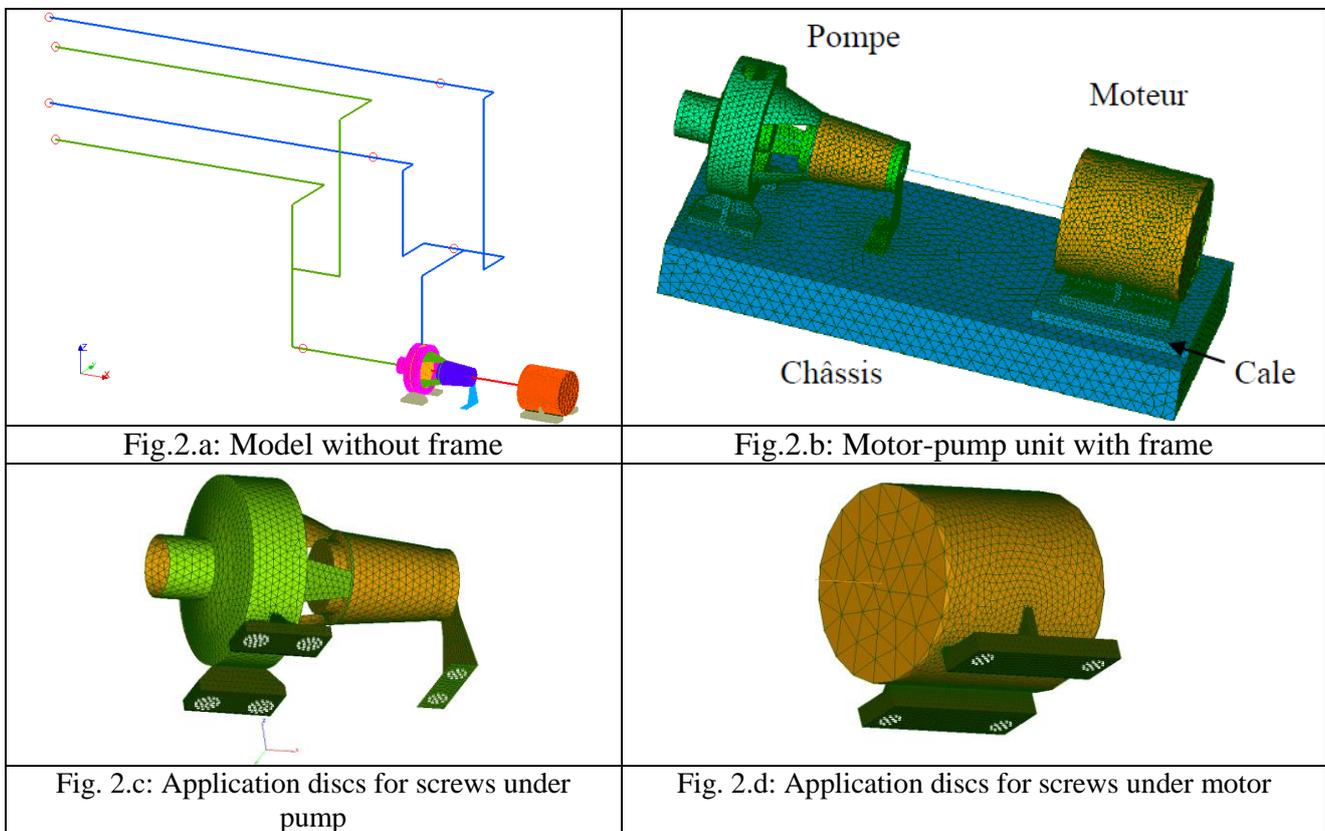


Figure 2: The horizontal pump unit: model with its pipes (without frame); model with frame (without pipes) ; pump model; motor model

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.

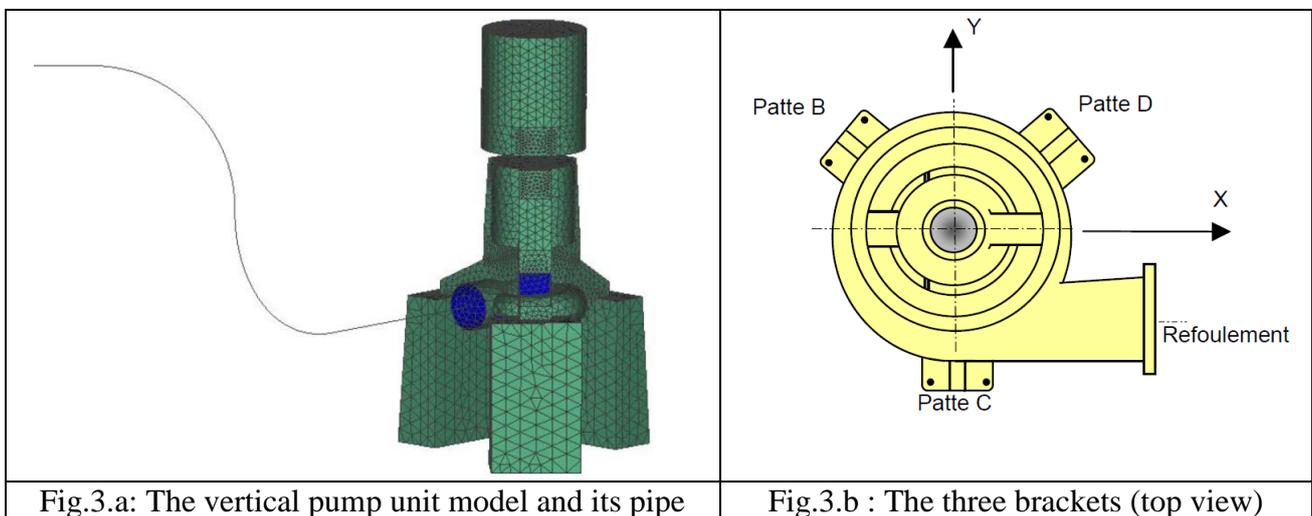


Figure 3: The vertical pump unit

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, conformly 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.

Mode number	Horizontal stiff pump unit		Vertical flexible pump unit			
	Num. eigenfrequency (Hz)	Characterisation	Num. eigenfrequency (Hz)	Exp. eigenfrequency (Hz)	MAC	Characterisation
1	4.2	Pipe	14.5	14.3	0.91	Pipe
2	15.9	Pipe	18.4	18.3		0.86
3	16.3	Pipe	21.5	21.6		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	Resultant		Resultant	
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 PÉZERAT¹, Frédéric ABLITZER¹ and Serge PUVILLAND²

¹Laboratoire 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 these 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

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

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

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

introduces the Poisson coefficient ν 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

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

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 $\frac{\partial^4 w}{\partial x^2 \partial y^2}$, 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

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

presents schemas with finite differences

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

and corrective coefficients

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

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

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

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

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

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

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

where

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

with $L = L_x - 4\Delta_x$ and

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

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

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

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

$$k_c = a \cdot k_f, \quad (9)$$

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 (56x56 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.

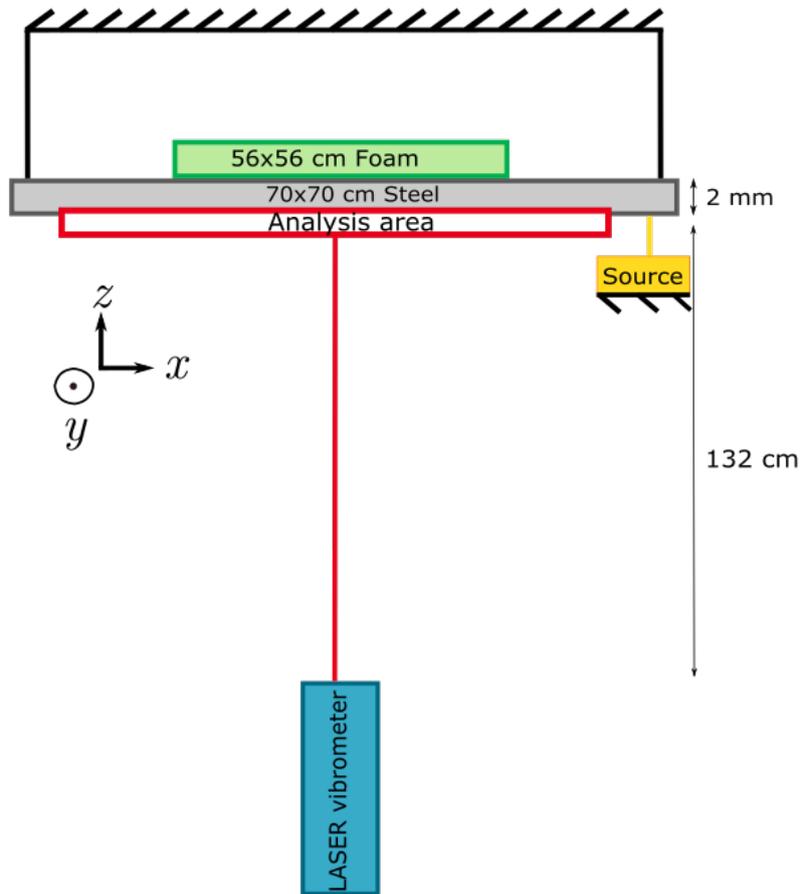


Figure 1: Experimental set-up

The measurement mesh has $33 \times 33 = 1089$ 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.

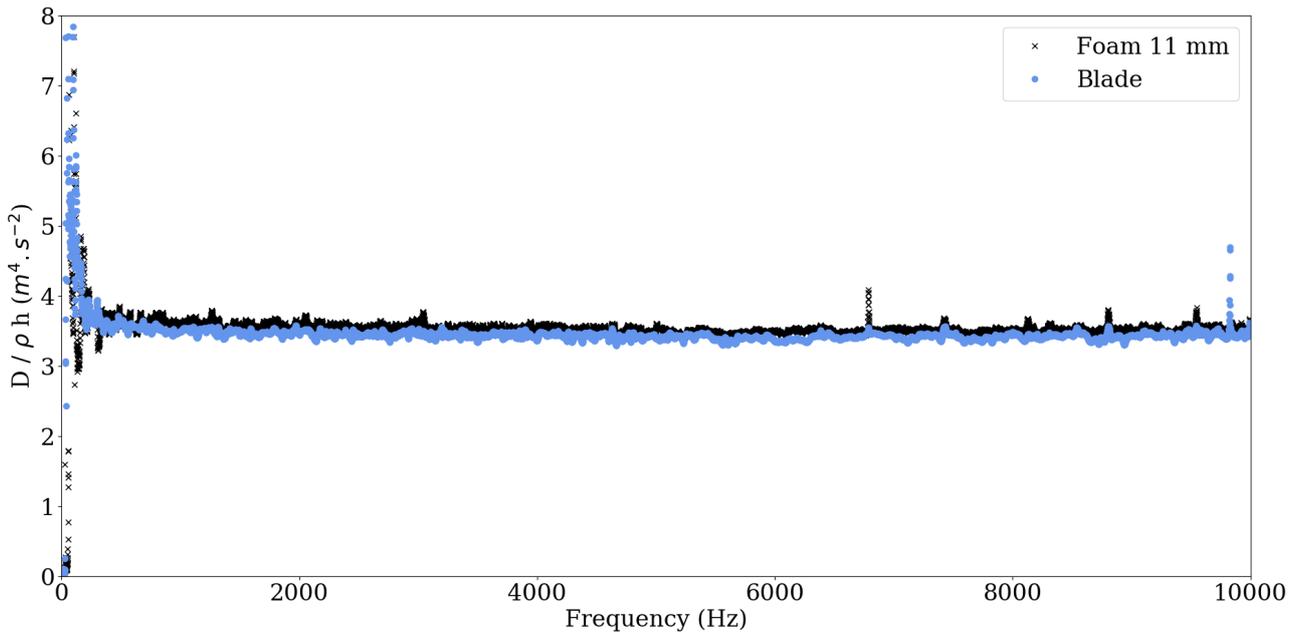


Figure 2: Stiffness $\frac{D}{\rho h}$ of the structure with only the plate and different foam thicknesses

The stiffness for the two configurations varies. The average value of the stiffness is $3,528 \text{ m}^4 \cdot \text{s}^{-2}$ for the configuration with the 11 mm foam and $3,454 \text{ m}^4 \cdot \text{s}^{-2}$ 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%.

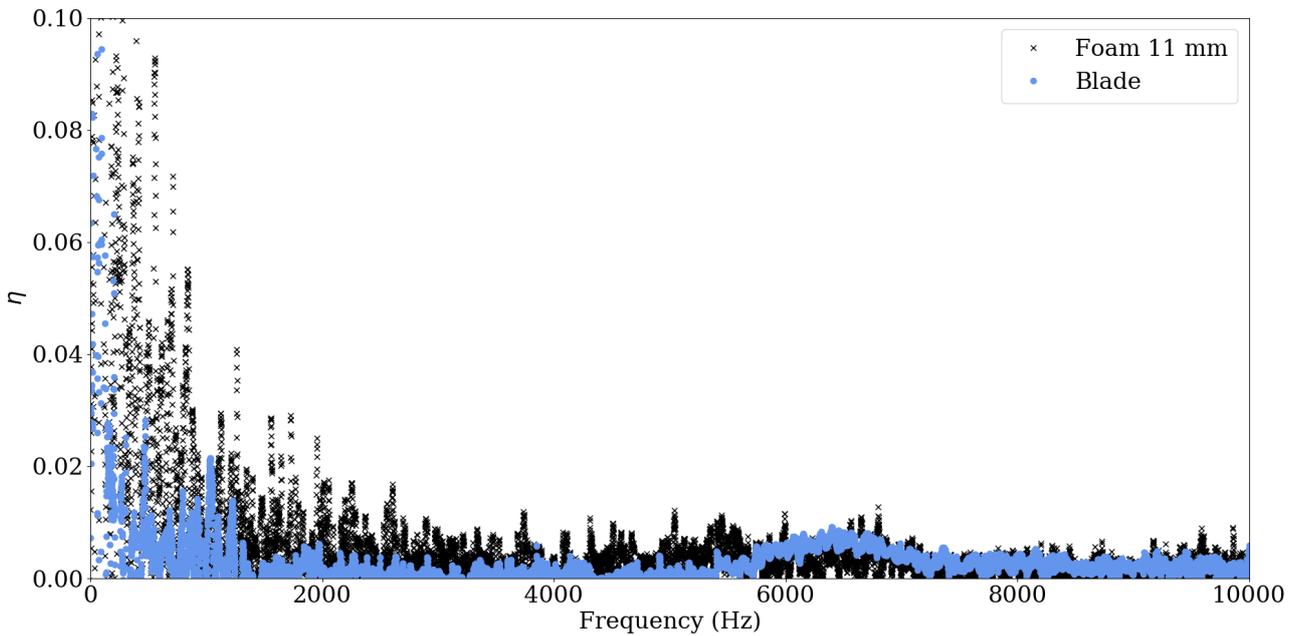


Figure 3: Loss factor η of the structure with only the plate and different foam thicknesses

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 SGHAIER¹, Jean-Luc DION¹, Nicolas PEYRET¹, Adeline BOURDON², Didier REMOND²

¹Laboratoire 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

²University 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 excitation 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 separate ways [1, 2, 3]. Fewer studies have been performed for the exploration of coupled bending-torsional behavior. However, those studies for coupled behavior are usually made under some simplifying assumptions. 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 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 vibrations 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 stationary regime for a given speed of rotation Ω . Al-Bedour [7] studied the particular case of Jeffcott 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 stationary and non-stationary 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-stationary 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:

$$\theta_z = \Omega + \theta_t \quad (1)$$

This way of introducing the degree of freedom θ_z gives more freedom for the simulation of the rotor behavior under non-stationary 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 analytical 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.

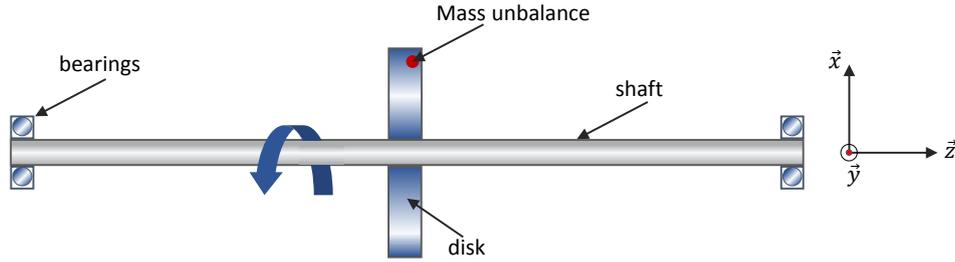


Figure 1: Illustration of the rotor

The finite element method is used to write the equation of motion over a shaft element under the considered 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:

$$\{\delta^e\} = (u_1; v_1; w_1; \theta_{x_1}; \theta_{y_1}; \theta_{z_1}; u_2; v_2; w_2; \theta_{x_2}; \theta_{y_2}; \theta_{z_2})_{\{1,12\}} \quad (2)$$

The energetic 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:

$$\left(\sum_e [M_s^e] + [M_D] + [M_u] \right) \{\ddot{\delta}\} + \left(\sum_e [C_s^e] + [G_D] \right) \{\dot{\delta}\} + \left(\sum_e [K_s^e] + [K_D] \right) \{\delta\} = \{F_{ext}\} + \sum_e [S_s^e] + \sum_e \{F_{se}^{coup}\} + \{F_D^{coup}\} + \{F_{Nlu}\} \quad (3)$$

$$[S_s^e] = -\ddot{\theta}_{z_2} [A_1] \{\delta^e\} - \dot{\theta}_{z_2} [Gyr] \{\dot{\delta}^e\} + (\ddot{\theta}_{z_2} - \ddot{\theta}_{z_1}) [A_2] \{\delta^e\} + (\dot{\theta}_{z_2} - \dot{\theta}_{z_1}) [Gyr^*] \{\dot{\delta}^e\} \quad (4)$$

$$\{F_{se}^{coup}\} = -\{N_3^g(l)\} \left(\{\delta^e\}^t [A_1^g] \{\delta^e\} \right) - \{N_3^g(l)\} \left(\{\delta^e\}^t [A_1^g] \{\dot{\delta}^e\} \right) + \left\{ \frac{\partial N_3^g}{\partial z} \right\} \left(\{\delta^e\}^t [A_2^g] \{\delta^e\} \right) + \left\{ \frac{\partial N_3^g}{\partial z} \right\} \left(\{\delta^e\}^t [A_2^g] \{\dot{\delta}^e\} \right) \quad (5)$$

$$[Gyr] = [A_1^g] - [A_1^g]^t \quad ; \quad [Gyr^*] = [A_2^g] - [A_2^g]^t \quad (6)$$

$$[A_1^g] = -2 \frac{\rho I_p}{l} \int_0^l \left\{ \frac{\partial N_2^g}{\partial z} \right\} \left\{ \frac{\partial N_1^g}{\partial z} \right\}^t dz \quad ; \quad [A_2^g] = -2 \frac{\rho I_p}{l} \int_0^l \int_0^z \left\{ \frac{\partial N_2^g}{\partial z} \right\} \left\{ \frac{\partial N_1^g}{\partial z} \right\}^t dz \quad (7)$$

$$\begin{cases} u(z,t) = \{N_1(z)\}^t \{\delta_u^e(t)\} & ; & w(z,t) = \{N_3(z)\}^t \{\delta_w^e(t)\} \\ v(z,t) = \{N_2(z)\}^t \{\delta_v^e(t)\} & ; & \theta_z(z,t) = \{N_3(z)\}^t \{\delta_{\theta_z}^e(t)\} \end{cases} \quad (8)$$

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-stationary 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-stationary regime and would vanish if the torsional deformation is neglected in the study. Finally, the vector $\{F_{se}^{coup}\}$ 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 gyroscopic effect terms are taking this form due to the considered assumptions of non-ideal energy source, non-stationary 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.

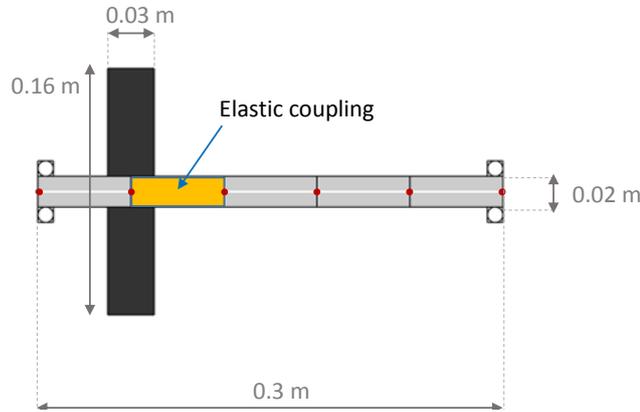


Figure 2: studied rotor

The campbell diagram of the rotor shows two critical speeds in the speed range between $0rpm$ and $45000rpm$. The first critical speed $\Omega_{cr1} = 2038rpm$ and the second one is at $\Omega_{cr2} = 20388rpm$ (see fig.3).

A linear torque is induced to the rotor. As we can see on the results for the evolution of the instantaneous 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 undesirable because it causes high lateral vibrations magnitude if not enough damping is present in the structure. It is a manifestation of the interaction 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 perform time-frequency analysis 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 $f_u - f_{b1}$ and $f_u + f_{b1}$ are contained in the torsional displacement signal, such as f_{b1} is the transverse natural frequency at Ω_{cr2} which is given by $f_{b1} = 154Hz$. Finally,

only when crossing the second critical speed, a frequency of $4f_u$ 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 $4f_u$ contributes to the composition of the torsional signal.

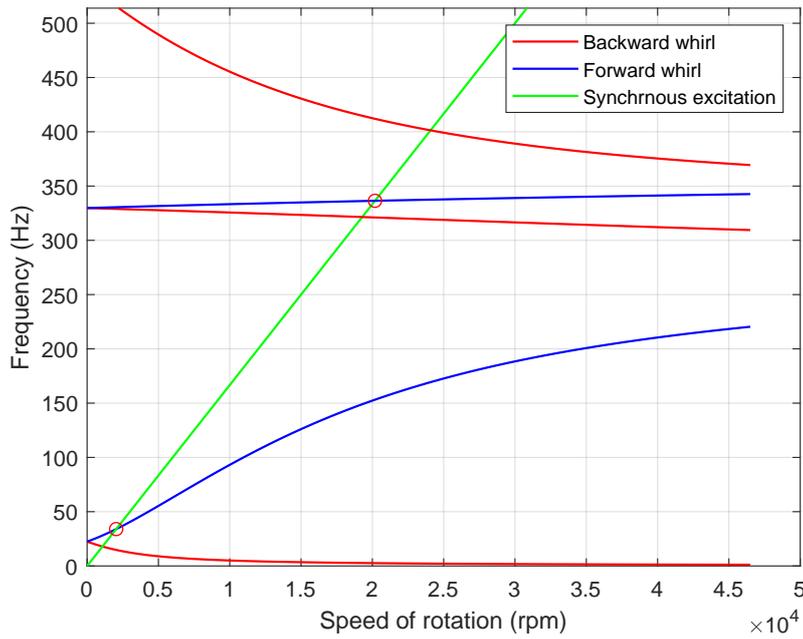


Figure 3: Campbell diagram

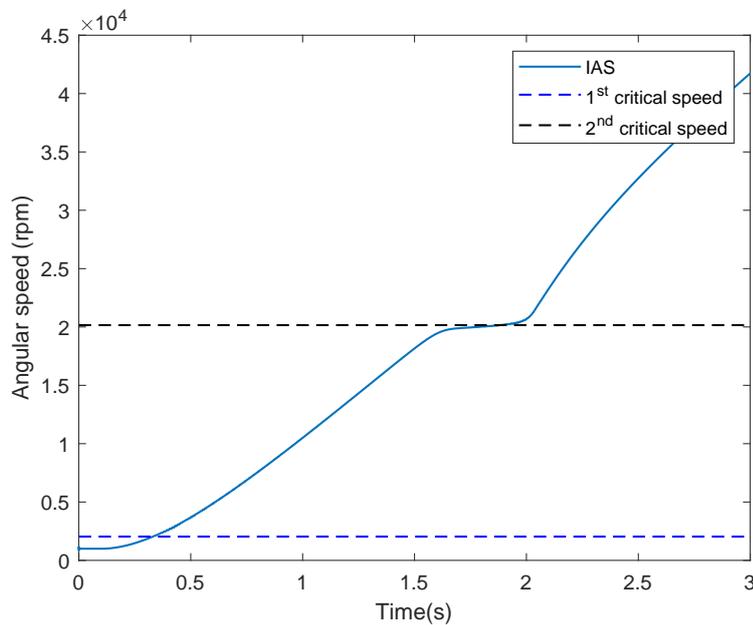


Figure 4: Angular velocity as a function of time

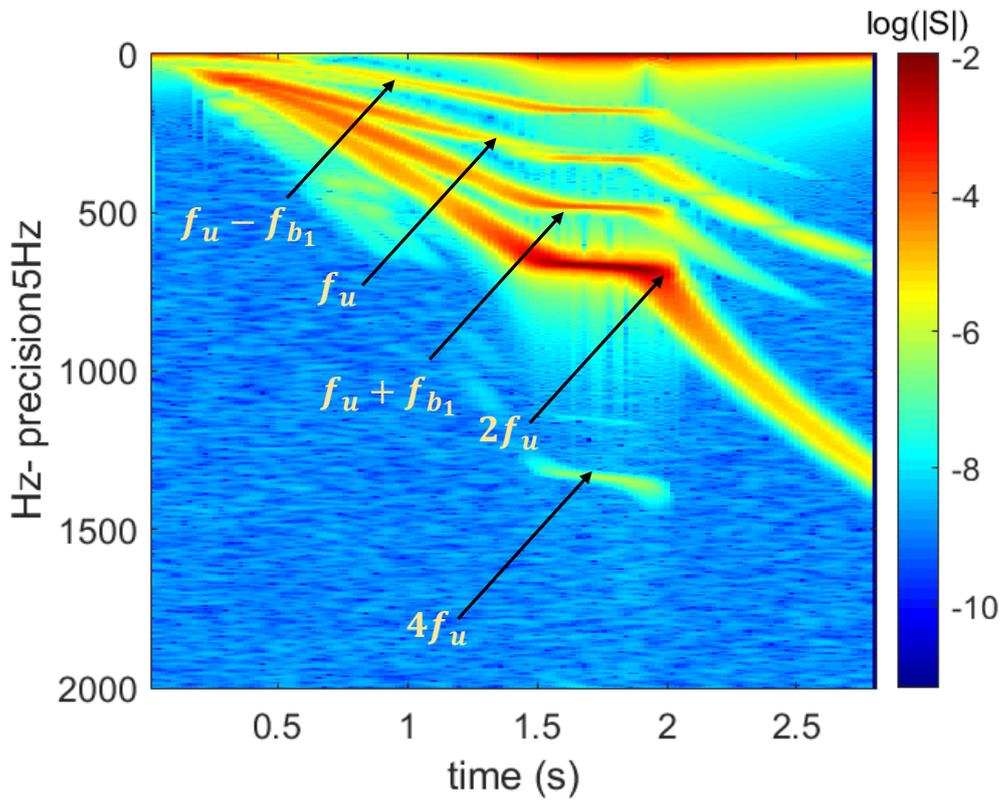


Figure 5: Time-frequency analysis of torsional vibrations on the second element

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-stationary 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. Shiao, 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

$$Y = \begin{bmatrix} y_{1,1} & y_{1,2} & y_{1,3} & \cdots & y_{1,n} \\ y_{2,1} & y_{2,2} & y_{2,3} & \cdots & y_{2,n} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ y_{m,1} & y_{m,2} & y_{m,3} & \cdots & y_{m,n} \end{bmatrix} = [y_{[1]}, y_{[2]}, y_{[3]}, \dots, y_{[n]}] \quad \text{where } y_{[i]} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \vdots \\ y_{m,i} \end{bmatrix} \quad (1)$$

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 $i = 1, \dots, n$.

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

$$y_{[t+k]} = A_1 y_{[t+k-1]} + A_2 y_{[t+k-2]} + \cdots + A_p y_{[t+k-p]} + e_k \quad (2)$$

or equivalence to

$$y_{[t+k]} = [A_1, A_2, \dots, A_p] \begin{bmatrix} y_{[t+k-1]} \\ y_{[t+k-2]} \\ \vdots \\ y_{[t+k-p]} \end{bmatrix} + e_k \quad (3)$$

where $A_j \in \mathbb{R}^{m \times m}$ for $j = 1, \dots, p$ 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 $k = 0, \dots, N$. Therefore, the relationship between the current response and the previous (past) response is written in compact form as

$$B = \Phi_A R + E \quad (4)$$

where $B \in \mathbb{R}^{m \times N}$ is the matrix of N responses, $\Phi_A \in \mathbb{R}^{m \times pm}$ 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

$$B = [y_{[t]}, y_{[t+1]}, \dots, y_{[t+N]}] \quad (5)$$

$$\Phi_A = [A_1, A_2, \dots, A_p] \quad (6)$$

$$E = [e_0, e_1, \dots, e_N] \quad (7)$$

$$R = \begin{bmatrix} y_{[t-1]} & y_{[t]} & \cdots & y_{[t+N-1]} \\ y_{[t-2]} & y_{[t-1]} & \cdots & y_{[t+N-2]} \\ \vdots & \vdots & \vdots & \vdots \\ y_{[t-p]} & y_{[t+1-p]} & \cdots & y_{[t+N-p]} \end{bmatrix} \quad (8)$$

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]

$$\Phi_A = BR^T (RR^T)^{-1} \quad (9)$$

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]

$$\Phi = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I & 0 & \dots & 0 & 0 \\ 0 & I & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I & 0 \end{bmatrix} \quad (10)$$

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

$$\lambda_r = \frac{\ln(U_{r,r})}{\Delta t} \quad (11)$$

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

$$\omega_r = \sqrt{(\text{real}(\lambda_r))^2 + (\text{imag}(\lambda_r))^2} \text{ (rad/s)} \Rightarrow f_r = \frac{\omega_r}{2\pi} \text{ (Hz)} \quad (12)$$

$$\zeta_r = -\frac{\text{real}(\lambda_r)}{\omega_r} \quad (13)$$

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.

$$f^{(p)} = [f_1^{(p)}, f_2^{(p)}, \dots, f_{n_p}^{(p)}] \quad (14)$$

$$\zeta^{(p)} = [\zeta_1^{(p)}, \zeta_2^{(p)}, \dots, \zeta_{n_p}^{(p)}] \quad (15)$$

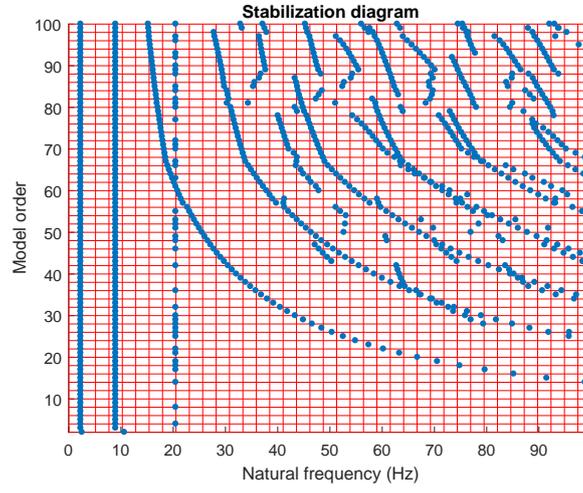


Figure 1 – Illustration of the grid technique.

When the model order varies in the interval $[p_1, p_2, \dots, p_m]$ the number of frequencies/damping ratios is $[n_{p_1}, n_{p_2}, \dots, n_{p_m}]$ in which $n_{p_1} < n_{p_2} < \dots < n_{p_m}$. 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

$$F = \begin{bmatrix} f^{(p_1)} \\ f^{(p_2)} \\ \vdots \\ f^{(p_m)} \end{bmatrix}, \quad Z = \begin{bmatrix} \zeta^{(p_1)} \\ \zeta^{(p_2)} \\ \vdots \\ \zeta^{(p_m)} \end{bmatrix} \quad (16)$$

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 $[p_1, p_m]$, 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 $p = [p_1, p_m]$ in the ranges $[f_{\min}, f_{\max}]$ and $[\zeta_{\min}, \zeta_{\max}]$

$$f_{\min} \leq F \leq f_{\max} \quad \text{and} \quad \zeta_{\min} \leq Z \leq \zeta_{\max} \quad (17)$$

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

$$N_f = \frac{f_{\max} - f_{\min}}{\Delta f} \quad (18)$$

Step 4 Define the number of repeating frequency in the interval $[\Delta f, 2\Delta f]$, denoted by N_{rf} , where:

$$1 \leq N_{rf} \leq p_m \quad (19)$$

Set $k = 1$

Step 5 Classify frequencies and damping ratios in the interval as follows

$$[I, \bar{F}] = \text{FIND_STABLE_FREQUENCIES}(f_{\min} + (k-1)\Delta f < F \leq f_{\min} + k\Delta f) \quad (20)$$

$$\bar{Z} = \text{GET_DAMPING_RATIOS}(Z, I) \quad (21)$$

$$\bar{N} = \text{COUNT_STABLE_FREQUENCIES}(\bar{F}) \quad (22)$$

Step 6 If $\bar{N} \geq N_{rf}$ then \bar{F} , \bar{Z} and the index matrix I (index of \bar{F} in F) is saved to a file.

Step 7 If $k \leq N_f$ 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

$$f_{\min} = 0, f_{\max} = \frac{1}{2\Delta t}, \zeta_{\min} = 0, \zeta_{\max} = 100. \quad (23)$$

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).

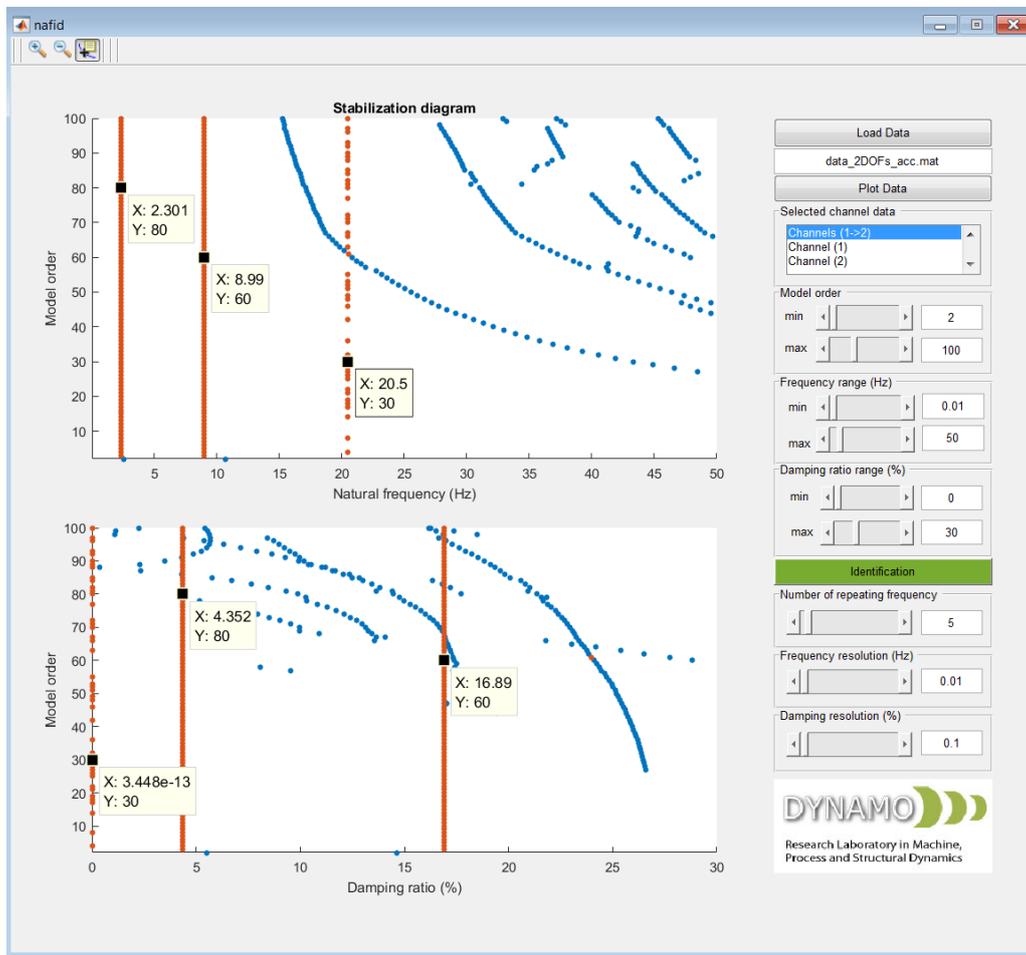


Figure 2 – Interface of the NAFID-tool implemented in Matlab.

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).

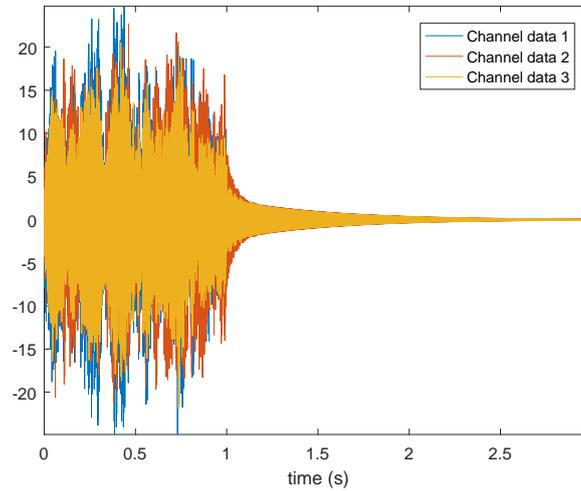


Figure 3 – Time response of 3 channels.

In order to estimate natural frequencies and damping ratios of this system, the following parameters are used in the NAFID-tool: $f_{\min} = 0$ 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 $q = [q_1, q_2, q_3, q_4, q_5, q_6]^T$, equations of motion of the system is given by

$$M\ddot{q} + C\dot{q} + Kq = f \quad (24)$$

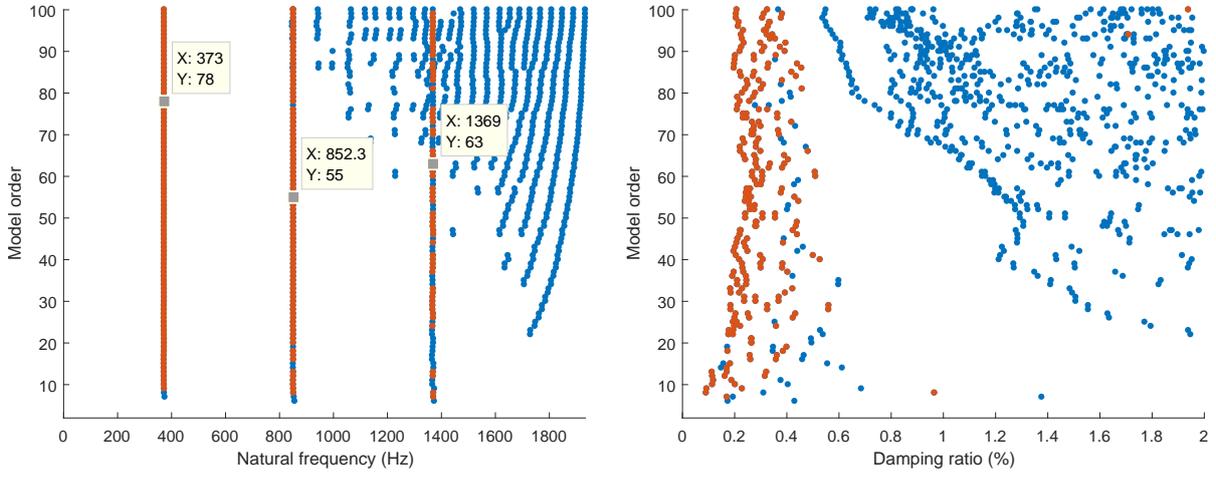


Figure 4 – Modal parameters of the 3DOF system using the NAFID-tool

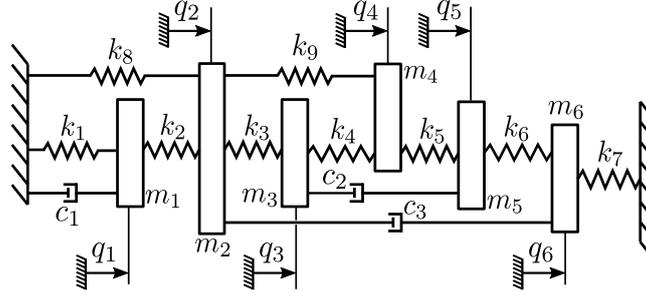


Figure 5 – The mechanical system with 6 DOFs.

where

$$M = \begin{bmatrix} m_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & m_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & m_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & m_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & m_6 \end{bmatrix}, \quad C = \begin{bmatrix} c_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & c_3 & 0 & 0 & 0 & -c_3 \\ 0 & 0 & c_2 & 0 & -c_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -c_2 & 0 & c_2 & 0 \\ 0 & -c_3 & 0 & 0 & 0 & c_3 \end{bmatrix} \quad (25)$$

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 & 0 & 0 \\ -k_2 & k_2 + k_3 + k_8 + k_9 & -k_3 & -k_9 & 0 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 & 0 & 0 \\ 0 & -k_9 & -k_4 & k_4 + k_5 + k_9 & -k_5 & 0 \\ 0 & 0 & 0 & -k_5 & k_5 + k_6 & -k_6 \\ 0 & 0 & 0 & 0 & -k_6 & k_6 + k_7 \end{bmatrix}, \quad (26)$$

For numerical simulation, physical parameters of the system are given by: $m_1 = m_2 = m_5 = 2$ (kg), $m_3 = m_4 = m_6 = 1$ (kg), $k_5 = k_8 = k_9 = 2.0 \cdot 10^6$ (N/m), $k_1 = k_2 = k_3 = k_4 = k_6 = k_7 = 10^6$ (N/m) and $c_1 = c_2 = c_3 = 10^3$ (Ns/m). The system is excited by initial condition with $q = 0$ and $\dot{q} = [0, 10, 0, 0, 0, 0]^T$ 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: $f_{\min} = 0$ and $f_{\max} = 500$ (Hz); $\zeta_{\min} = 0.0$ and $\zeta_{\max} = 70(\%)$; $N_{\text{tf}} = 5$ 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.

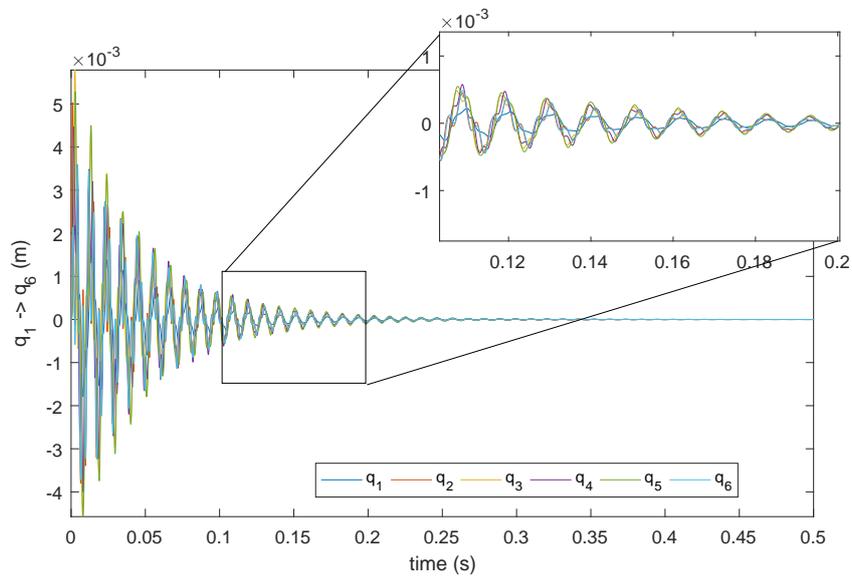


Figure 6 – Displacement vector is calculated using ODE45.

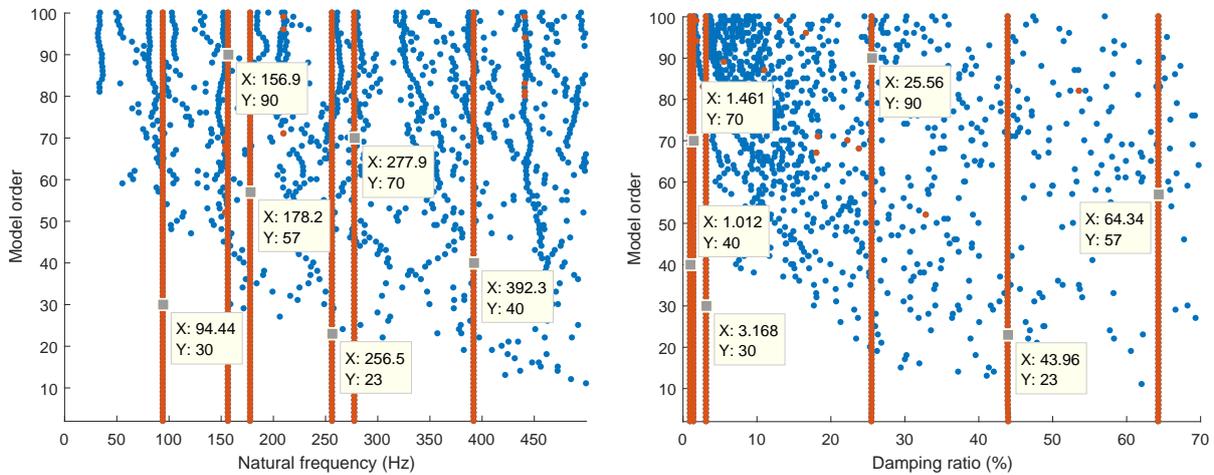


Figure 7 – Modal parameters of the 6DOF-mechanical system using the NAFID-tool

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. Laffleur, 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 Weisser², Romain Barbeau¹,
Anne-Isabelle Mallet-Da Costa¹, Evelyne Aubry²

¹Faurecia Automotive Seating, CS 10001, 91152 Etampes, France

²IRIMAS, 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 $95\text{mm} \times 95\text{mm} \times 40\text{mm}$ 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 $100\text{mm}/\text{min}$ up to 80% of deformation. The results are presented in figure 1.

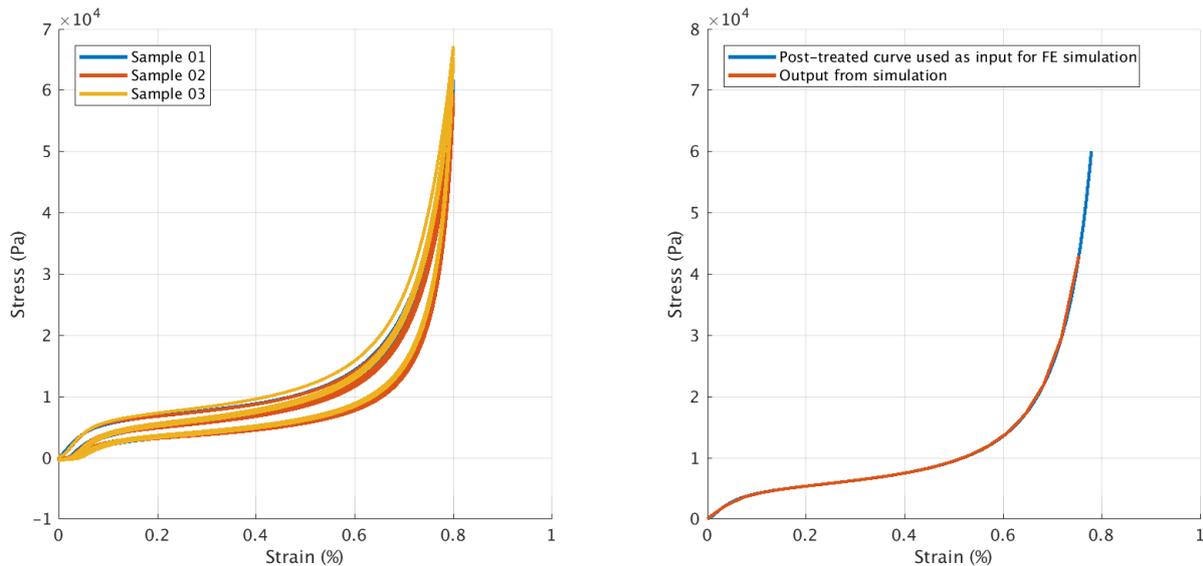


Figure 1: Left: Raw stress-strain curves measured on three foam samples. Right: Post-processed curve used as input for the finite element simulation (blue) and simulation result (red).

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-strain 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 additional 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

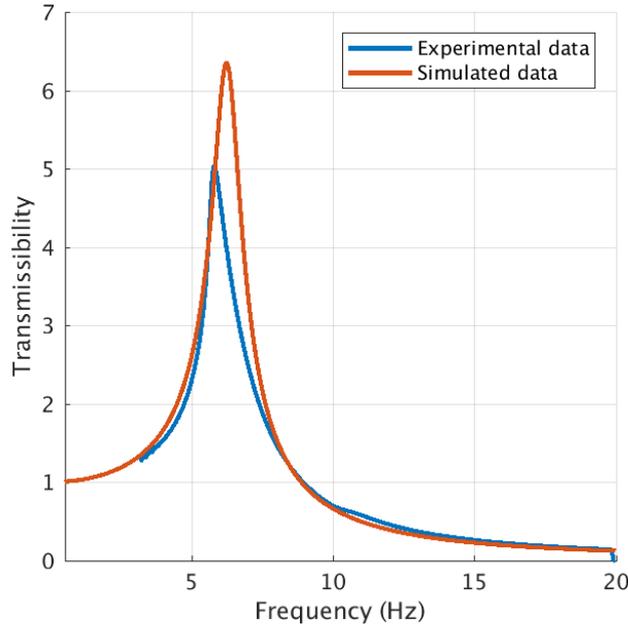


Figure 2: Experimental transmissibility curve (blue) and simulated curve with an equivalent elasticity modulus (red).

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:

$$T(\omega) = \frac{k + jc\omega}{-m\omega^2 + jc\omega + k}, \quad (1)$$

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 assumption 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 study 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 representing the bottom of the thighs, and three additional surfaces associated to the contact between the lead buttock and the seat bolsters.

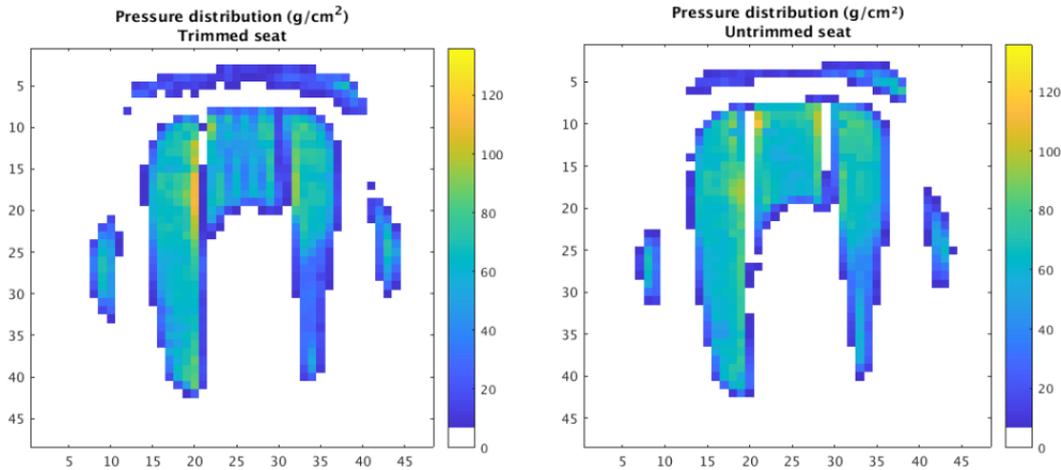


Figure 4: Pressure maps obtained depending on the seat setup.

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 interesting to see that the maximum pressure is $114.7g.cm^{-2}$. When this value is multiplied by the gravity acceleration ($g = 9.81ms^{-2}$), it is possible to obtain the maximum applied stress. Here, $\sigma_{max,exp} = 11.25kPa$. This value is lower than the maximum stress obtained in the simulation ($\sigma_{max,sim} = 43kPa$, see figure 1) and shows therefore that the tabulated 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 $30Hz$ and with an RMS value of $0.5ms^{-2}$. The seat is first excited during $2min$ until the steady state is considered as reached, and then the measurement is performed during $6min$. 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.

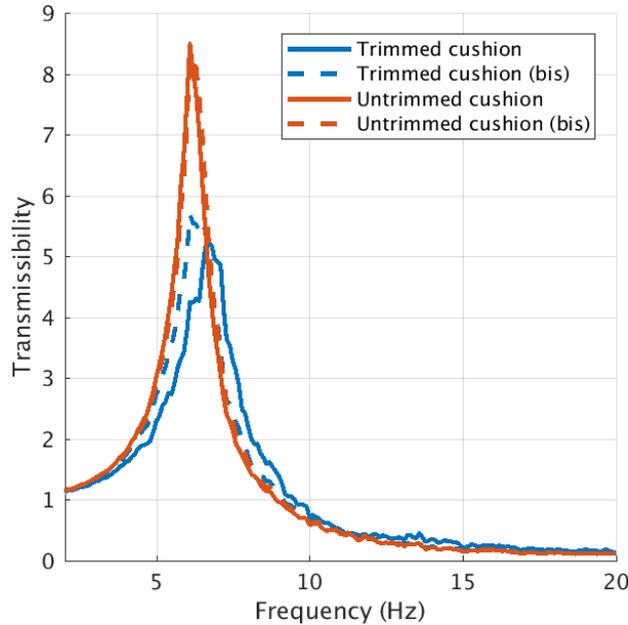


Figure 5: Transmissibilities obtained for the trimmed and untrimmed seat cushion with a lead buttock.

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, transmissibilities 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 assumption 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 PEETERS²,
Pieter-Jan DAEMS², Ann NOWÉ¹, Jan HELSEN²

¹Artificial Intelligence Lab Brussels

²Acoustics & 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.

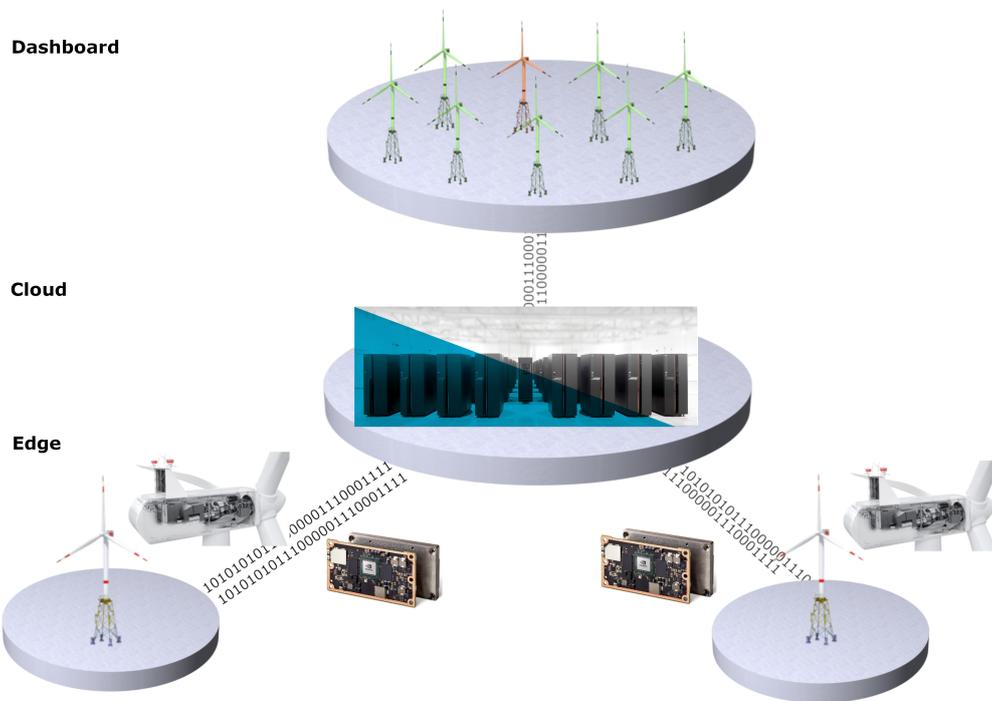


Figure 1 – Streaming data in the context of an edge-cloud balanced architecture for condition monitoring.

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. Learning 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 at 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: $PEI = \sqrt{\frac{1}{N_p} \sum_{n=1}^{N_p} x_p^2(n)}$, 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 harmonics. 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 assess 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 AVRCloud 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.

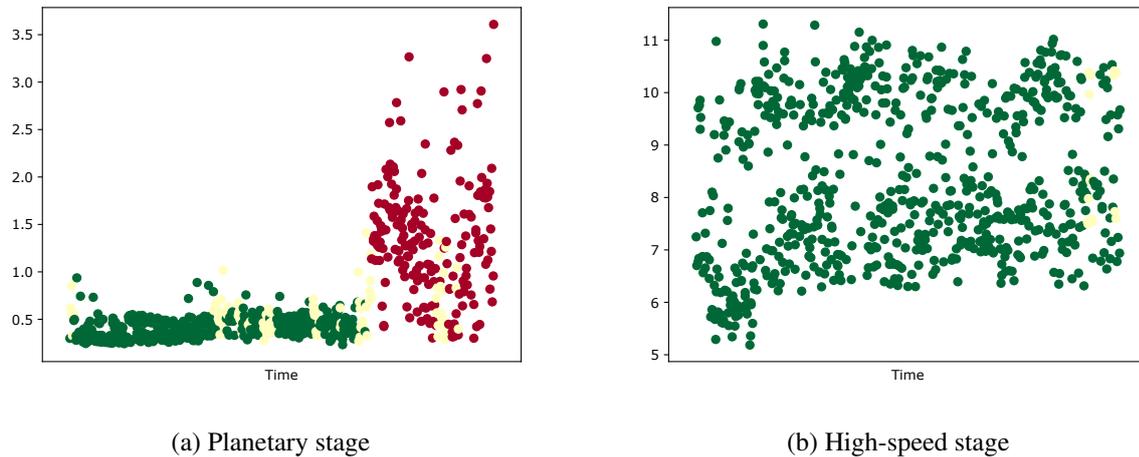


Figure 2 – Comparison of peak to peak based health indicator between two affected and unaffected channels.

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 Bayesian 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 Wetenschappelijk 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.

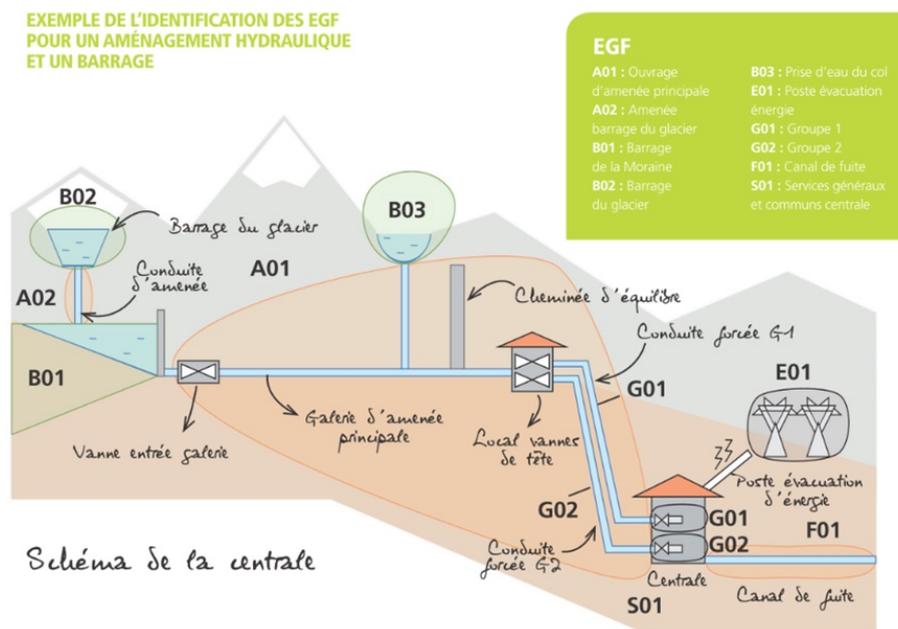


Figure 1: Schematic diagram of a hydraulic installation and name of the main elements (in French)

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.

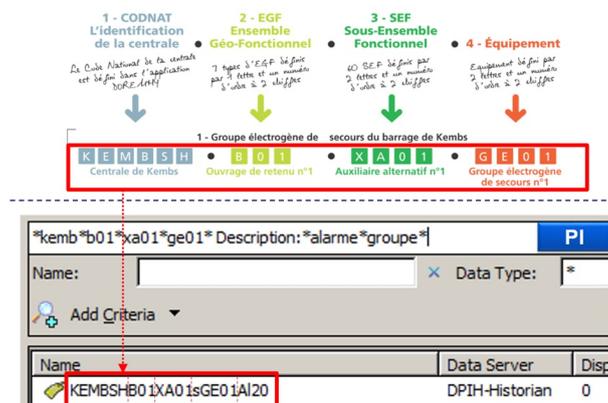


Figure 2: Parsing of the name of a PI tag linked to the equipment tree in which the monitored equipment is located

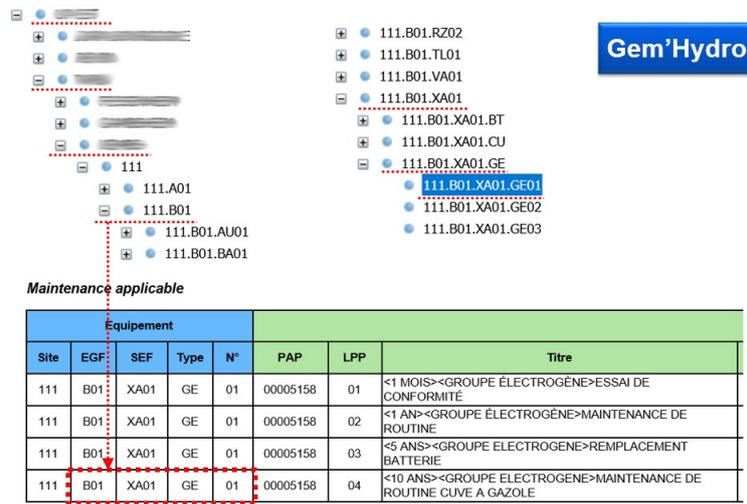


Figure 3: Parsing of the name of a maintenance event in Gem'Hydro related to the equipment tree in which the maintained equipment is located

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 OSIssoft data historian, a robust and widely proven industrial solution. The PI OSIssoft 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		
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.	REQUALIF. G04 T°C + JOINT A FUIITE_01062018 H-41554603-2018-000143
31/05/2018	A98.G04.AL04.AT04		00365448-01	MP	<R><ALT>VISITE X-MA-0814 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 basse 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 FUIITE DU G4	Remplacement du joint fuite effectué	H-41554404-2018-000003 RT GMH - 2018 - G4 - REMPLACEMENT DU JOINT À FUIITE
						Entretien léger à 500h + nettoyage des têtes de bobines pour essayer de faire remonter 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

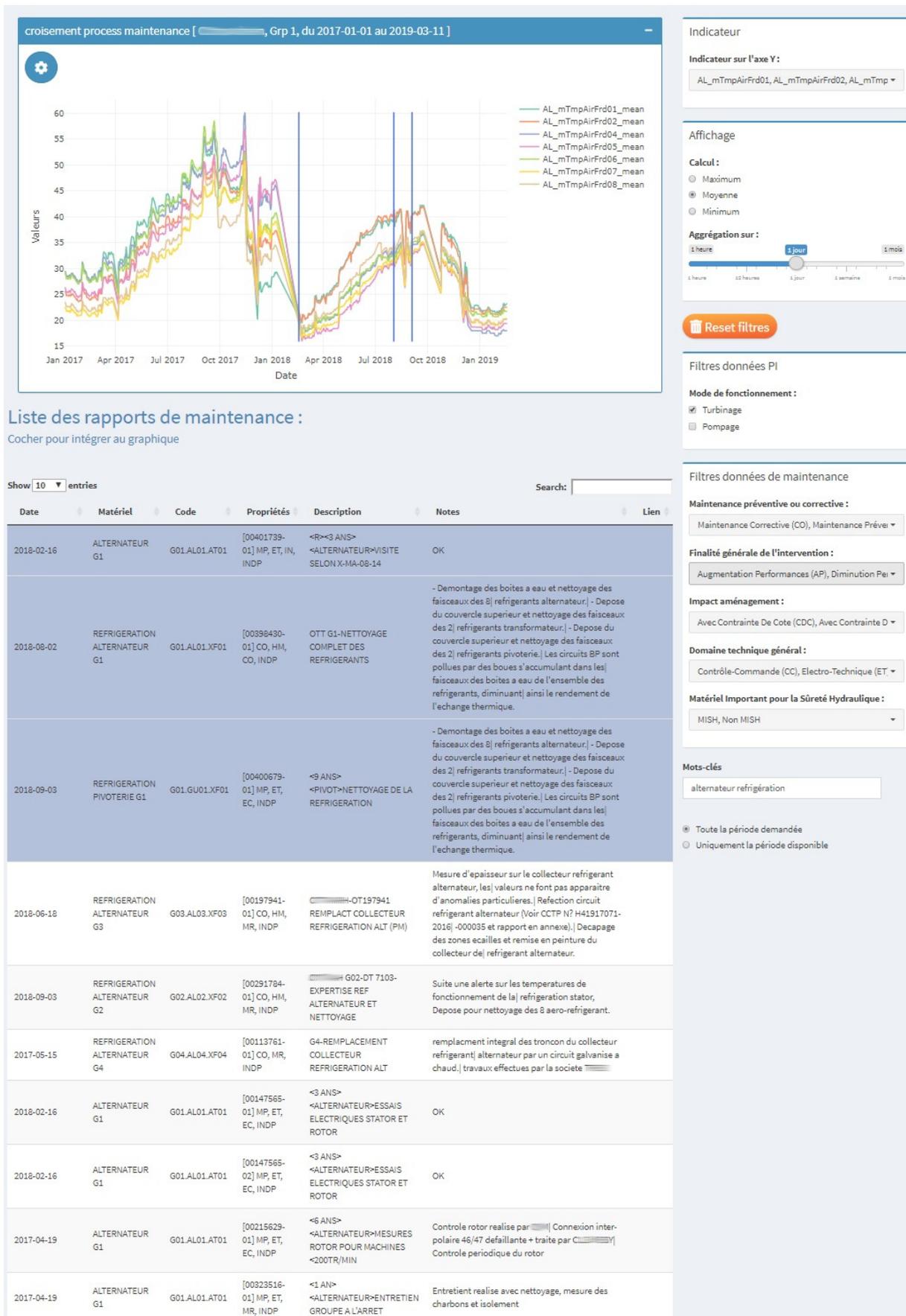


Figure 6: SuperViz'Orte cross-reference module screenshot (in French). The context restricts time series to alternator cold air temperatures and maintenance operations related to the alternator refrigeration system.

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).

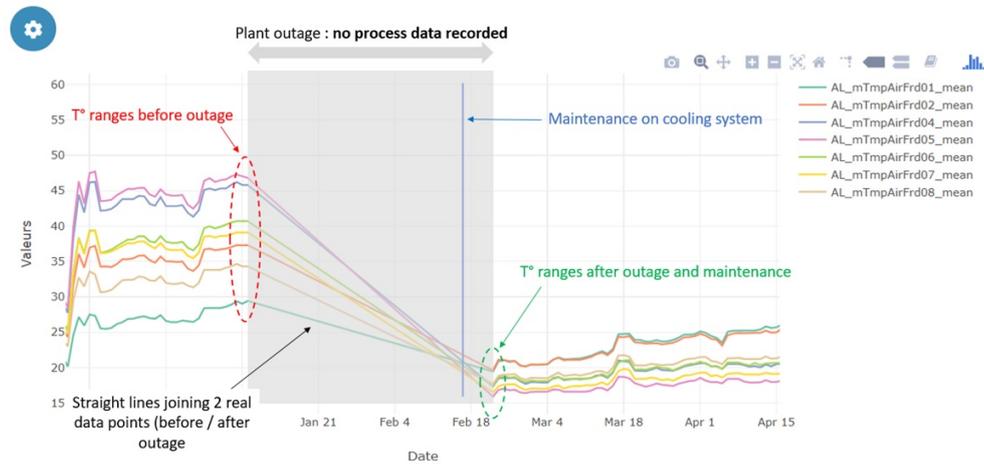


Figure 7: Focus on the first maintenance intervention: the plant outage and the maintenance performed are indeed at the origin of the temperature drops

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 Aghazadeh¹, Antoine Tahan² and Marc Thomas³

¹ Département de Génie mécanique, École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec), Canada

fatemeh.aghazadehkouzekonani.1@ens.etsmtl.ca

² Département de Génie mécanique, École de Technologie Supérieure, 1100, rue Notre-Dame Ouest, Montréal (Québec), Canada

antoine.tahan@etsmtl.ca

³ Département 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.

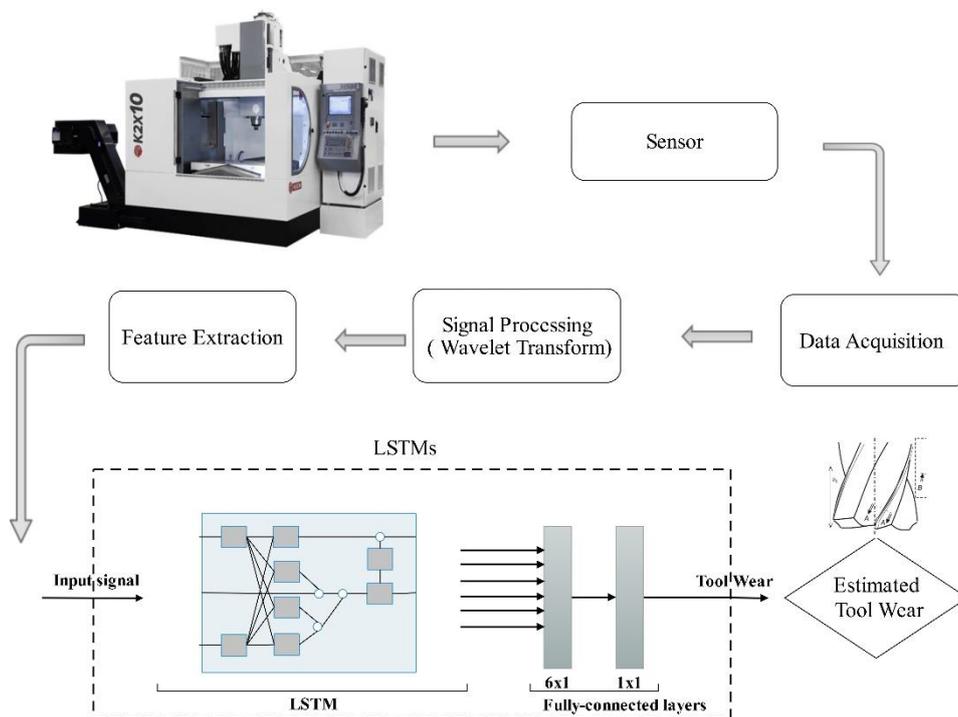


Figure 1. The monitoring system framework

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].

$$WT_x(t, a) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(u) \psi \left(\frac{u-t}{a} \right) du$$

where wavelet $\psi \left(\frac{u-t}{a} \right)$ 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]:

$$\begin{aligned} g_l^{(t)} &= \theta(W_l^{gx} h_{l-1}^{(t)} + W_l^{gh} h_{l-1}^{(t-1)} + b_l^g) \\ i_l^{(t)} &= \sigma(W_l^{ix} h_{l-1}^{(t)} + W_l^{ih} h_{l-1}^{(t-1)} + b_l^i) \\ f_l^{(t)} &= \sigma(W_l^{fx} h_{l-1}^{(t)} + W_l^{fh} h_{l-1}^{(t-1)} + b_l^f) \\ o_l^{(t)} &= \sigma(W_l^{ox} h_{l-1}^{(t)} + W_l^{oh} h_{l-1}^{(t-1)} + b_l^o) \\ s_l^{(t)} &= g_l^{(t)} \odot i_l^{(t)} + s_l^{(t-1)} \odot f_l^{(t)} \\ h_l^{(t)} &= \theta(s_l^{(t)}) \odot o_l^{(t)} \end{aligned}$$

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 $200 \times 54 \times 4$. 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

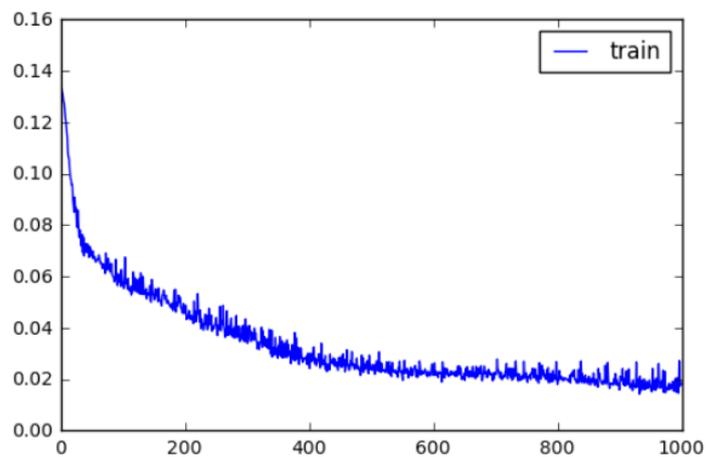


Figure 3. Loss function during training process

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.

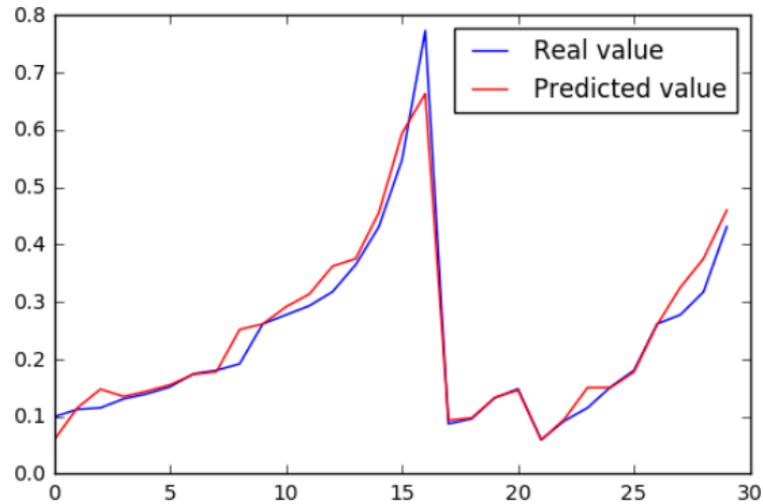


Figure 4. Estimated and real tool wear values using vibration signals

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 s-transform. 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 diagnosis 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 Laszló 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 Wetzell. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.
- [22] Francois Chollet et al. *Keras*, 2015.
- [23] Mart'in 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 AUGEZ¹, Aurélien CLOIX¹, Christian CLERC¹

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,

$$F(f) = Z(f) \cdot [A_w(f) + A_r(f) + A_c(f)]^{-1} \quad (1)$$

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

$$Y^2(f) = F(f) \cdot A_{cb}(f) \quad (2)$$

With:

$Y(f)$ 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 ($1/\lambda$) measured by the corrugation measurement trolley (e.g. Figure 1) on a portion of the track, and
- the vertical acceleration recorded on the bogie $Y(f)$, in running conditions at the speed V of the train on the same portion of track.



Figure 1: rail corrugation analysis trolley

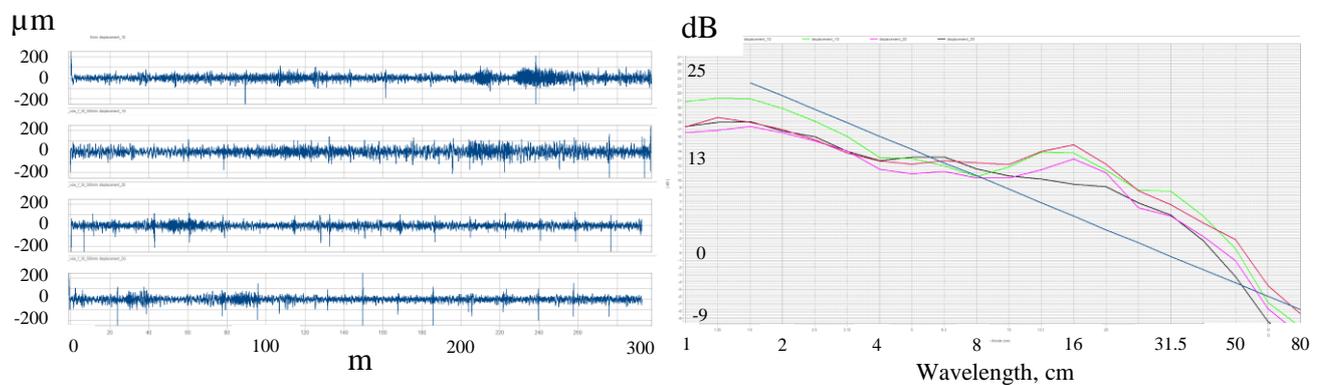


Figure 2: Left side: roughness in μm measured on 4 track lines.
Right side: Associated roughness spectra ($1/\text{m}$) of a portion of track

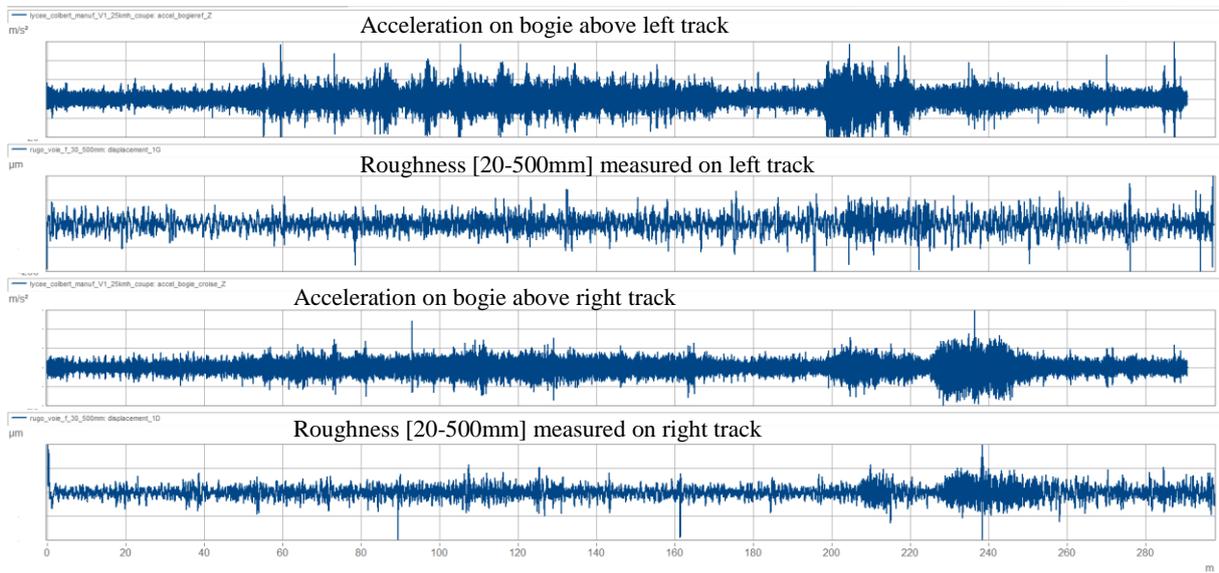


Figure 3: Accelerations measured on a bogie and roughness on the same portion of track at 25 km/h

A post-processing of these two measurements makes it possible to compute the transfer function $Y^2(f)/Z^2(f)$, by a conversion of the roughness abscissa ' 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 $[Y^2(f)/Z^2(f)]$ 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: $[Y_0^2(f)/Z_0^2(f)]^{-1}$. The roughness PSD is then directly computed using the product of the acceleration with the transition function.

$$Z^2(f) = [Y_0^2(f)/Z_0^2(f)]^{-1} \cdot Y^2(f) \quad (3)$$

With:

$Y^2(f)$ 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.

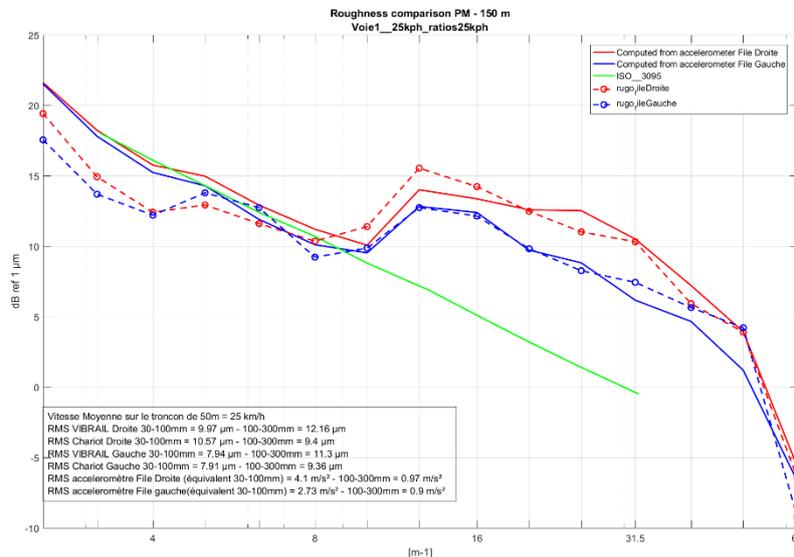


Figure 4 : 1 Third octave spectra (in 1/m) measured (with trolley in dashed lines) and computed (with inverse method in solid lines). In green, the ISO 3095 standard

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.

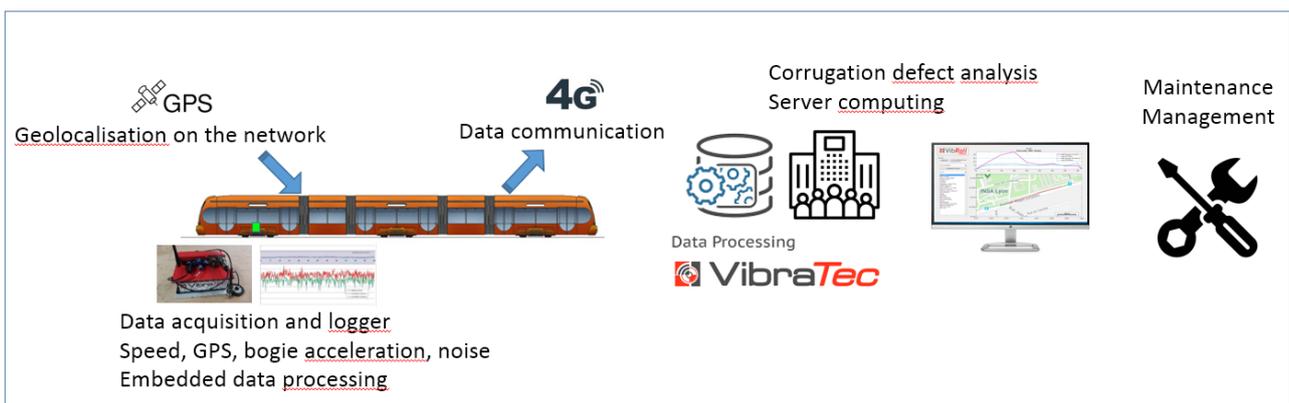


Figure 5: Global approach – from raw data to corrugation analysis

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

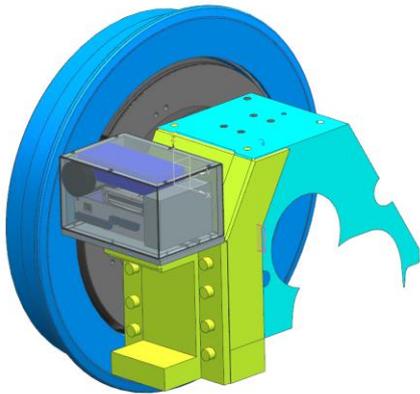


Figure 7: localisation of VibRail Concept n°1 on a tramway bogie

Examples of raw data are presented in Figure 8 and Figure 9.

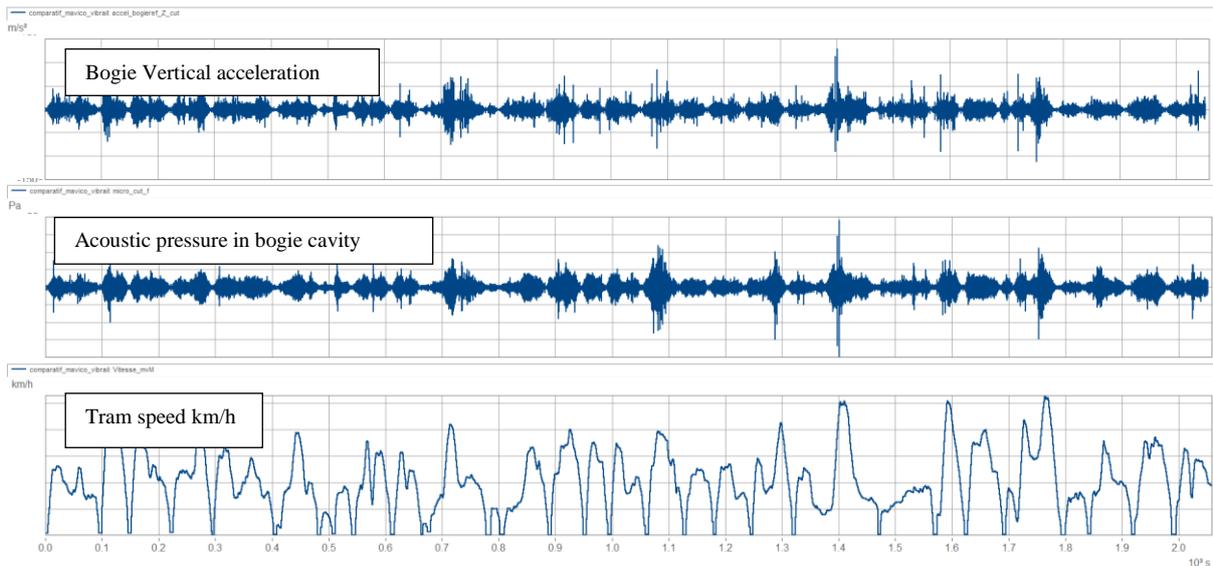


Figure 8: raw data measured by the device on the Lyon tramway network T1 Line



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.

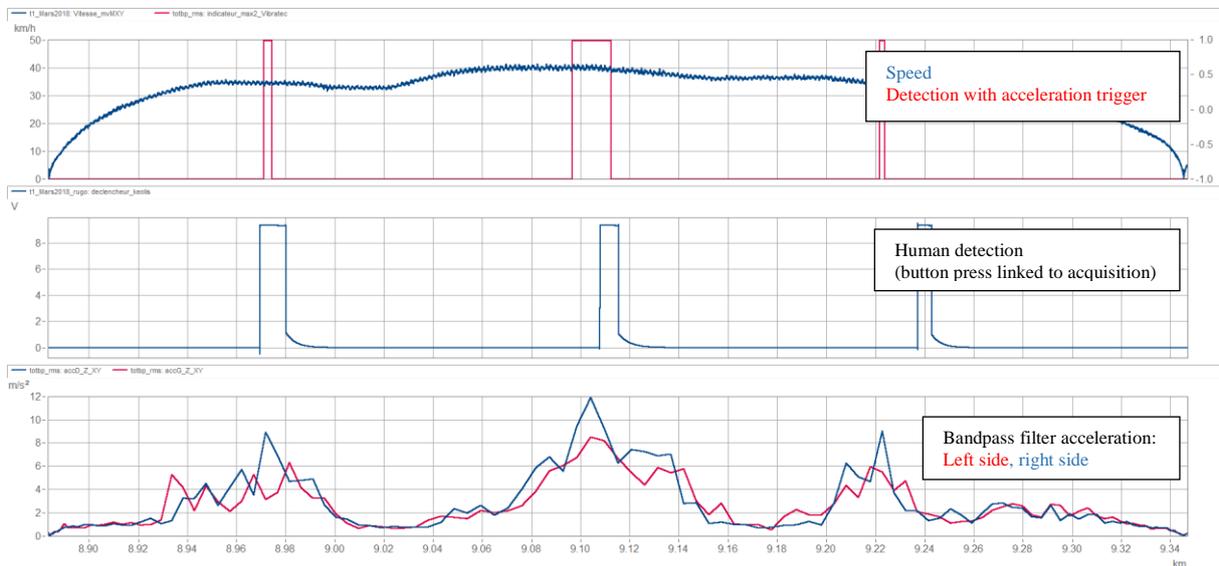


Figure 10: correlation between bandpass filtered acceleration and human detection on an interstation

In this example, a trigger was activated when the bandpass filtered RMS acceleration reached 8m/s^2 . 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 $[30-100]\text{mm}$ 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.

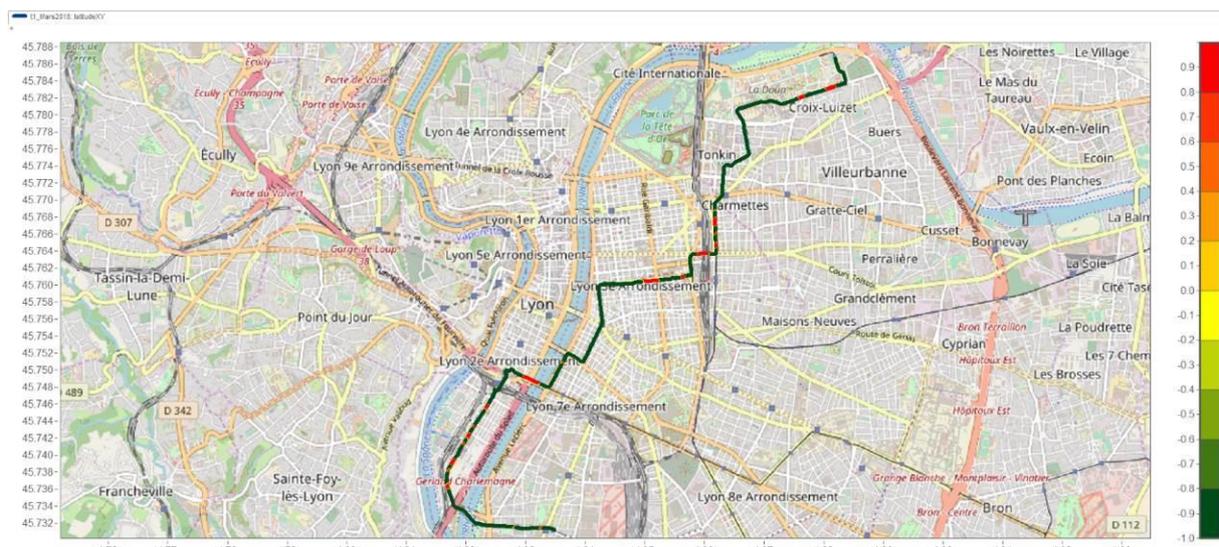


Figure 11: Example of a GPS map of the roughness level (Lyon tramway network T1 Line).

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).

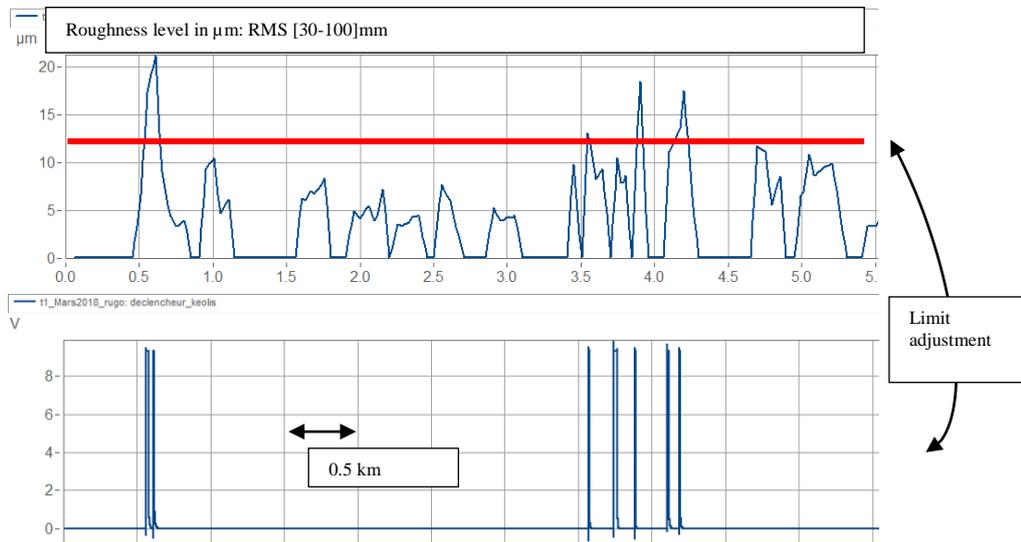


Figure 12: Top: Rail roughness - RMS level in the [30-150]mm wavelength. Bottom: Correlation with subjective detection

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).

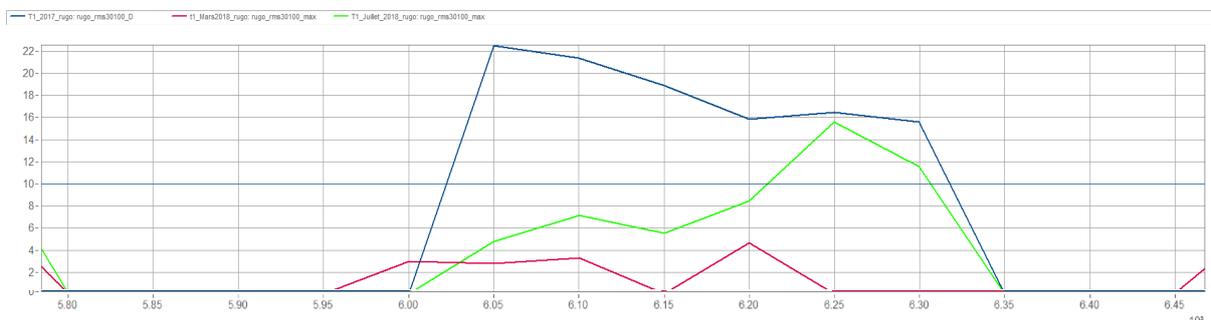


Figure 13: rail roughness - RMS level in the [30-150]mm wavelength. Status July 2017, Status March 2018, Status July 2018

In this example, the grinding campaign reduced the roughness from 22 μm to 4 μm 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 Peeters¹, Jérôme Antoni², Jan Helsen¹

¹Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
cedric.peeters@vub.be

²LVA, 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} :

$$\mathbf{s} = \mathbf{x} * \mathbf{h} \quad (1)$$

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:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \quad (2)$$

$$\begin{bmatrix} s_{N-1} \\ \vdots \\ x_{L-1} \end{bmatrix} = \begin{bmatrix} x_{N-1} & \dots & x_0 \\ \vdots & \ddots & \vdots \\ x_{L-1} & \dots & x_{L-N-2} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{N-1} \end{bmatrix}$$

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:

$$\varepsilon_{\mathbf{x}} = |\mathbf{s}|^2 = |\mathbf{X}\mathbf{h}|^2 \quad (3)$$

It can also be written as:

$$\varepsilon_{\mathbf{x}} = \begin{bmatrix} s_0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_{L-N+1} \end{bmatrix}^H \mathbf{X}\mathbf{h} = \mathbf{diag}(\mathbf{s}^H)\mathbf{X}\mathbf{h} \quad (4)$$

with \mathbf{s}^H being the Hermitian transpose of \mathbf{s} , and $\mathbf{diag}(\mathbf{s}^H)$ being a diagonal matrix with the values of the vector \mathbf{s}^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:

$$RPE = \frac{\sigma_e^2}{\sigma_{SE}^2} \quad (5)$$

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:

$$x(n) = \sum_{i=1}^N a_i x(n-i) + e(n) \quad (6)$$

with a_i the autoregressive coefficients, and where the error $e(n)$ can be obtained by:

$$e(n) = \sum_{i=0}^N a_i x(n-i) \text{ with } a_0 = 1. \quad (7)$$

Figure 1 illustrates the straightforward process. The autoregressive filtering necessary to obtain the prediction error \mathbf{e} of the squared envelope $\mathbf{\varepsilon}_x$ can be written in matrix notation as follows:

$$\mathbf{e} = \mathbf{A} \text{diag}(\mathbf{s}^H) \mathbf{s} \quad (8)$$

with \mathbf{A} a band matrix containing the autoregressive coefficients as follows:

$$\mathbf{A} = \begin{bmatrix} a_0 & a_1 & \dots & a_N & 0 & \dots & 0 \\ 0 & a_0 & a_1 & \dots & a_N & \ddots & \vdots \\ & & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & a_0 & a_1 & \dots & a_N \\ & & & & & \ddots & \ddots & \vdots \\ & & & & & & a_0 & a_1 \\ 0 & & \dots & & & & 0 & a_0 \end{bmatrix} \quad (9)$$

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:

$$\frac{\sigma_{SE}^2}{\sigma_e^2} = \frac{\mathbf{\varepsilon}_x^H \mathbf{\varepsilon}_x}{\mathbf{e}^H \mathbf{e}} = \frac{\mathbf{h}^H \mathbf{X}^H \text{diag}(\mathbf{s}) \text{diag}(\mathbf{s}^H) \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \text{diag}(\mathbf{s}) \mathbf{A}^H \mathbf{A} \text{diag}(\mathbf{s}^H) \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XW_1X} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XW_2X} \mathbf{h}} \quad (10)$$

The generalized Rayleigh quotient [13] can be recognized in Eq. 10 and can be maximized using an iterative maximization of the eigenvalues:

$$\lambda = \frac{\mathbf{h}^H \mathbf{R}_{XW_1X} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XW_2X} \mathbf{h}} \quad (11)$$

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:

$$\mathbf{R}_{XW_1X}\mathbf{h} = \mathbf{R}_{XW_2X}\mathbf{h}\lambda \quad (12)$$

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, $f_{BPF0} = 0.31Hz$
- 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.

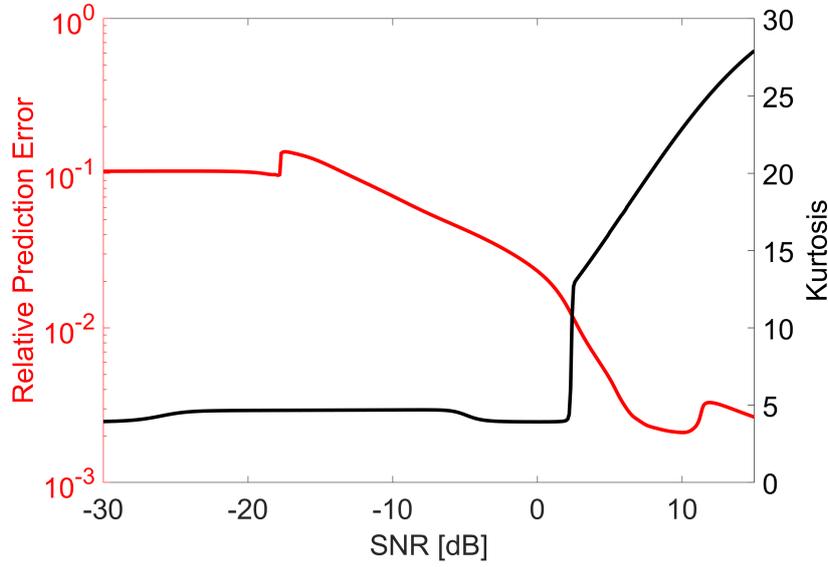


Figure 2: Trending of Relative Prediction Error (RPE) versus kurtosis after blind filtering using the proposed LPE approach and MED respectively.

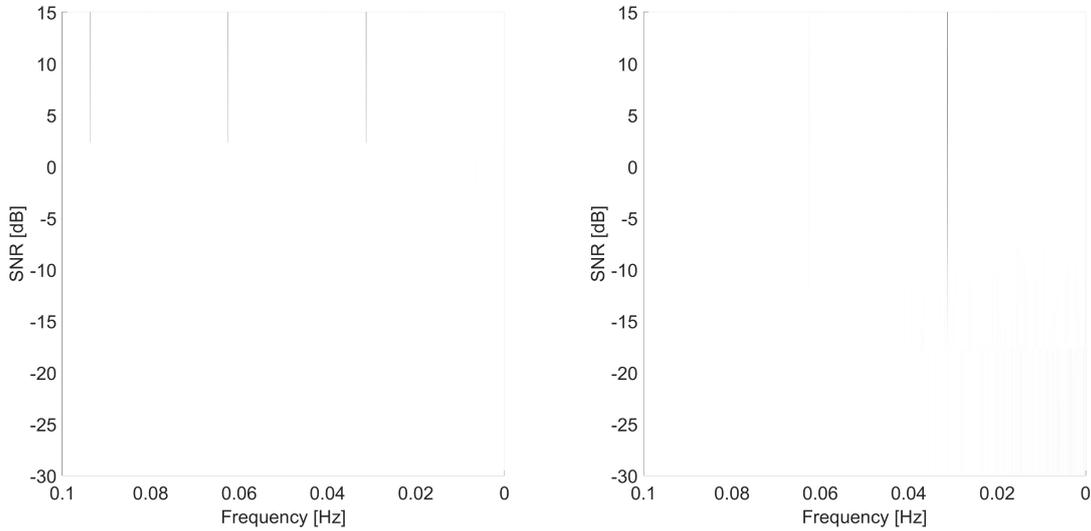


Figure 3: Color maps (in grayscale) of the squared envelope spectra after blind filtering using the MED (left) and proposed approach (right) respectively. The black lines appearing at approximately -17 dB and 2 dB SNR for respectively the proposed and MED method correspond to the 0.31 Hz fault frequency.

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.

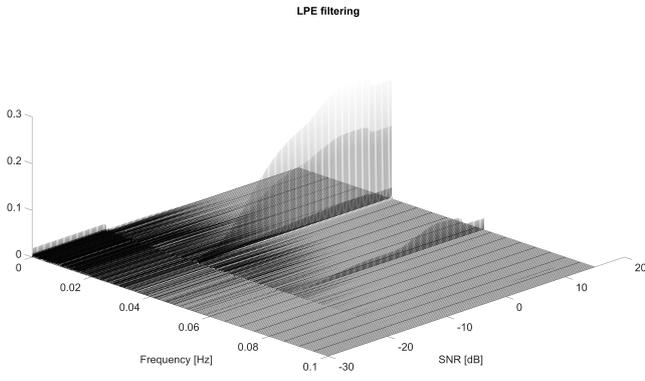


Figure 4: Squared envelope spectrum waterfall plot after filtering with the proposed method based on the relative prediction error.

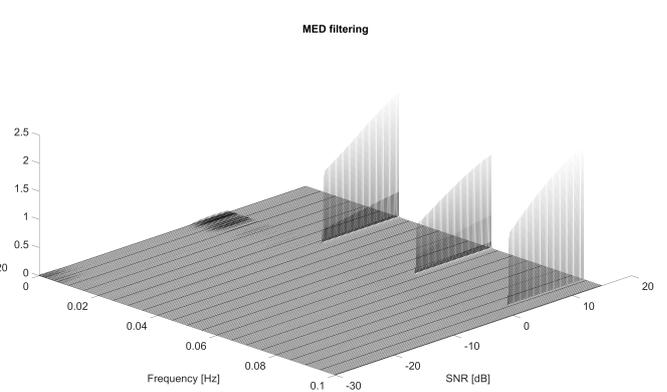


Figure 5: Squared envelope spectrum waterfall plot after filtering with the MED method based on kurtosis.

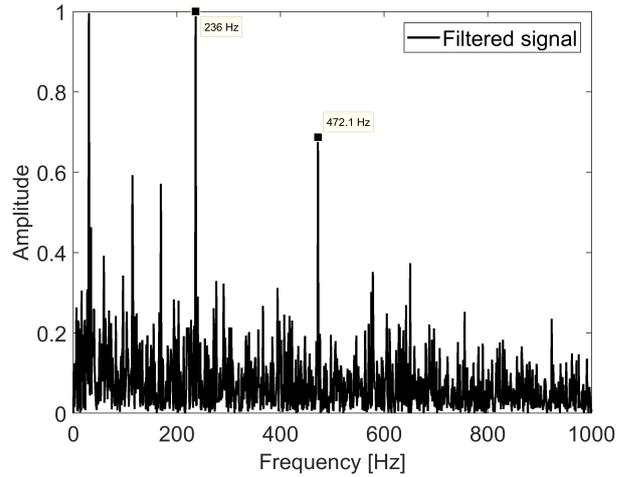
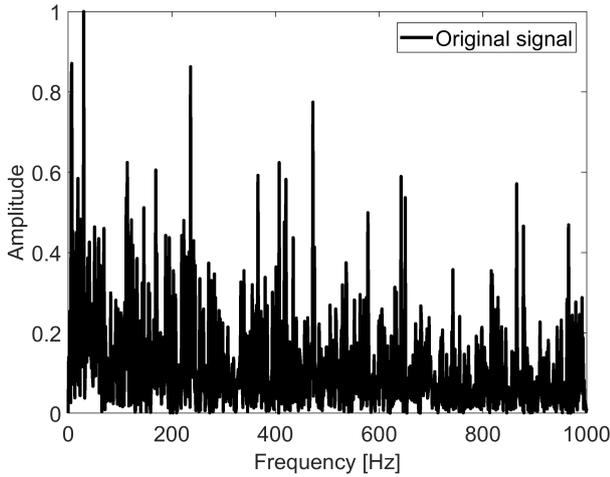


Figure 6: Experimental case: (Left) The envelope spectrum of the raw input signal. (Right) The envelope spectrum after filtering with the proposed LPE method.

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 GRIFFATON¹, Jose L. GOMEZ¹

¹Safran 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’.

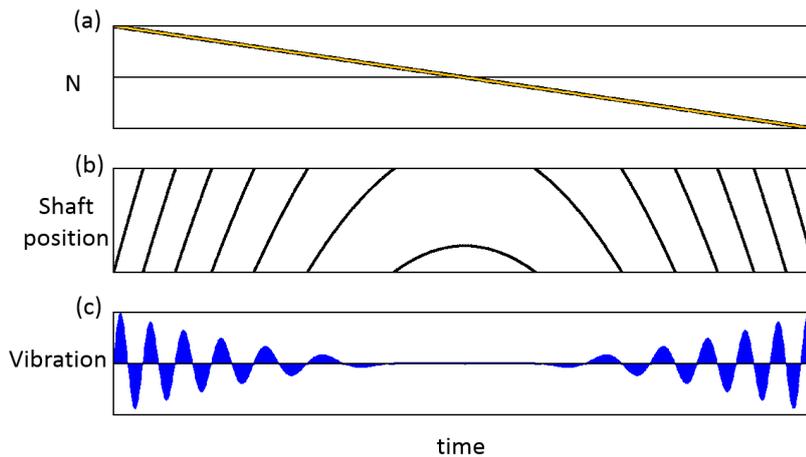


Figure 1: Standard time history representation with a simple simulation. Time history of the rotating speed N of a shaft (a), time history of the shaft angular position (b), and time history of a simulated vibration signal (c).

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).

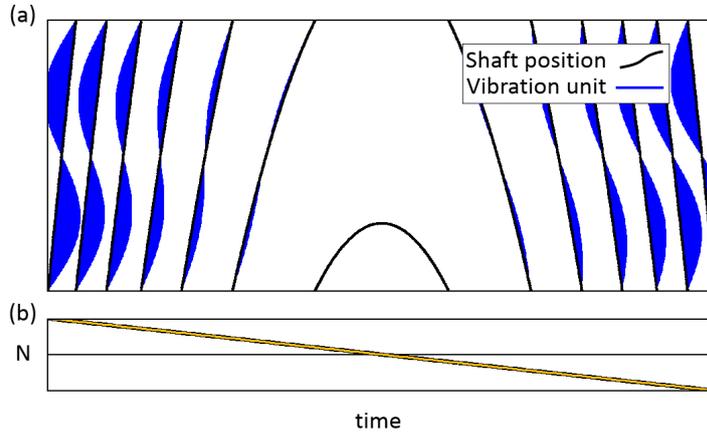


Figure 2: Representation in the time/phase domain (a) and time history of the shaft rotating speed (b).

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.

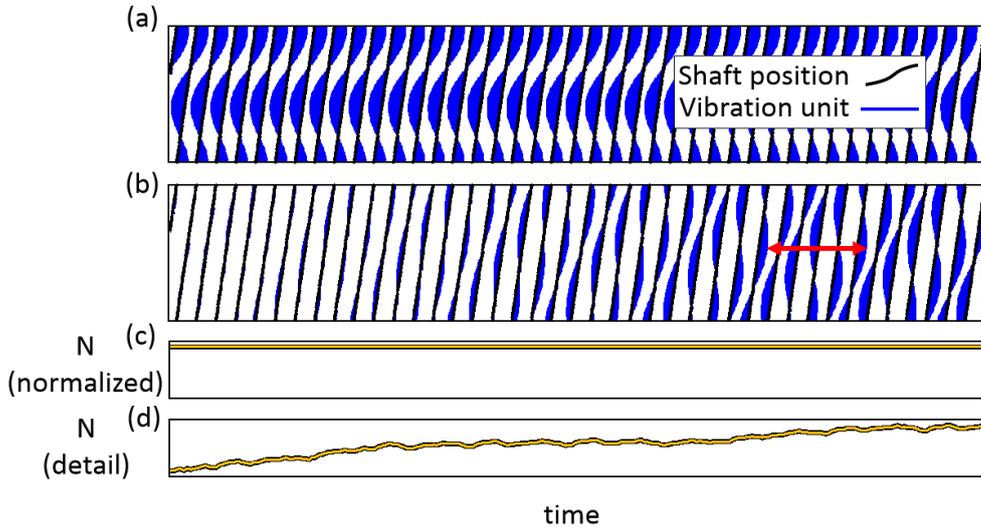


Figure 3: Application to a constant rigid unbalance and the progressive increase of an oil unbalance. Representation in the time/phase domain of the vibration amplitude tracked on the first order 1 N of the rotating speed (a) and of the vibration amplitude tracked around 0.8 N (b). Time history of the 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.

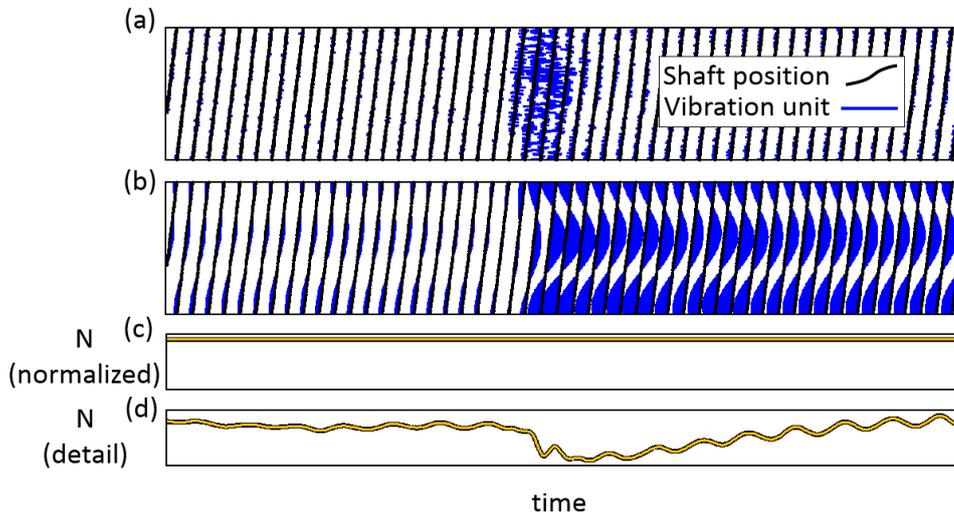


Figure 4: Application to ice shedding, fleeting event and sustained unbalance. Representation in the time/phase domain of the vibration amplitude (a) and of the vibration amplitude tracked around 1 N (b). Time history of the normalized rotating speed (c) and of the detail of the rotating speed (d). The unbalance increases in amplitude during the shedding while preserving its phase location.

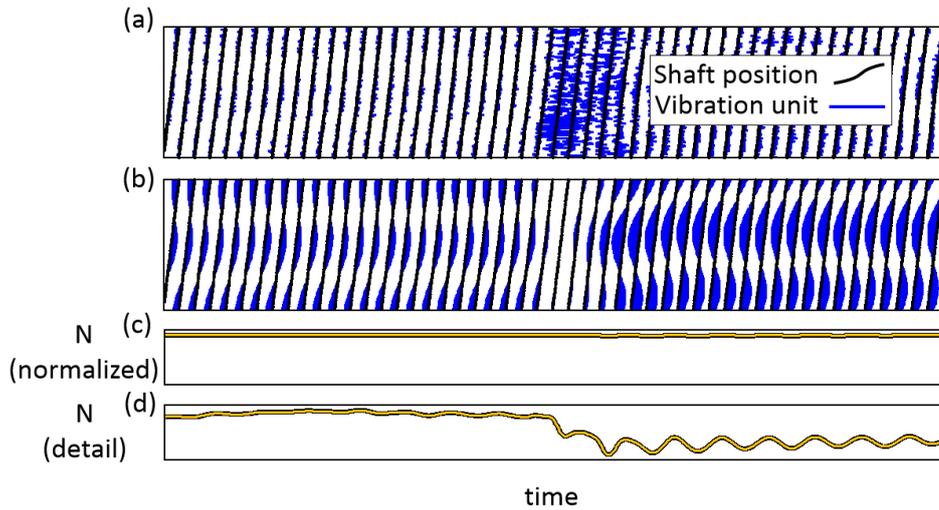


Figure 5: Application to ice shedding, fleeting event and change in unbalance amplitude and position. Representation in the time/phase domain of the vibration amplitude (a) and of the vibration amplitude tracked around 1 N (b). Time history of the normalized rotating speed (c) and of the detail of the 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.

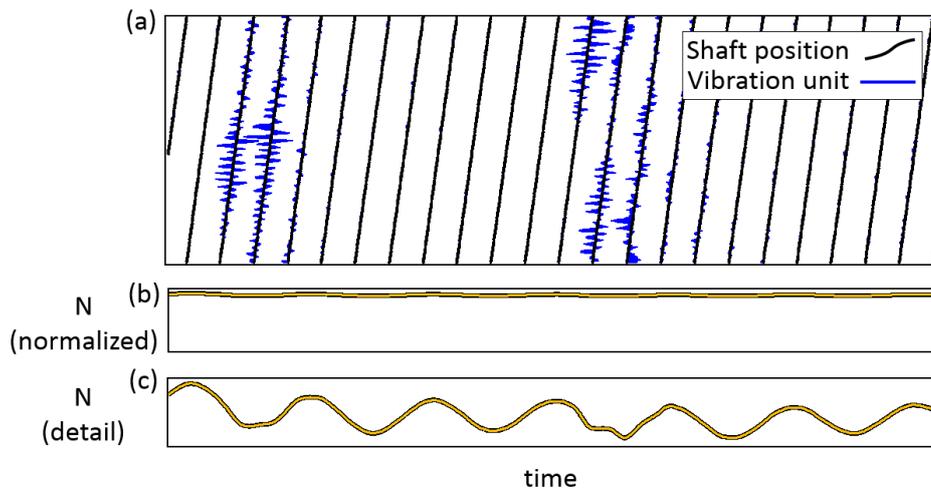


Figure 6: Application to the ingestions of several birds. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

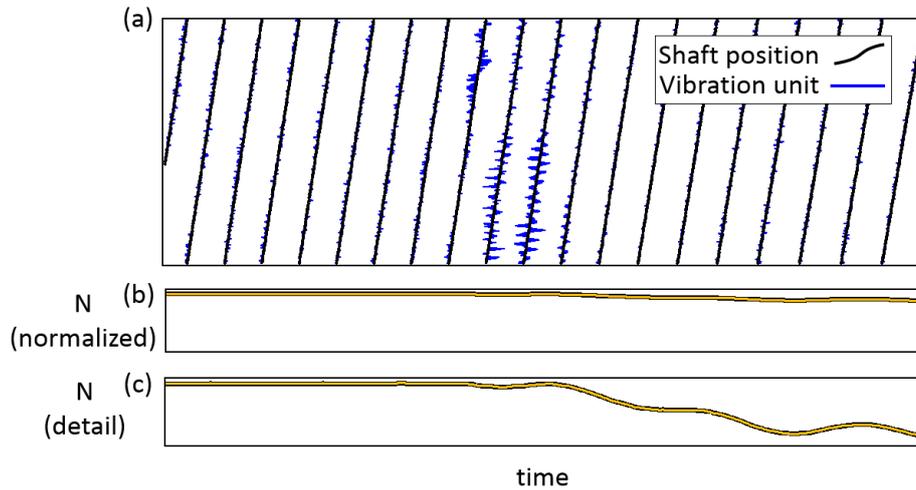


Figure 7: Application to a compressor stall. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

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.

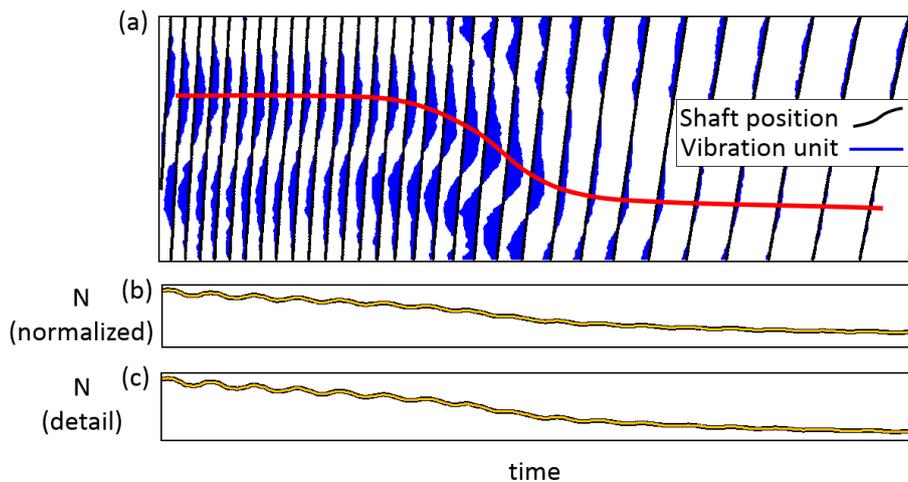


Figure 8: Application to the crossing of the rotating speed and a mode of the shaft. Representation in the time/phase domain of the vibration amplitude of a strain gauge mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (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.

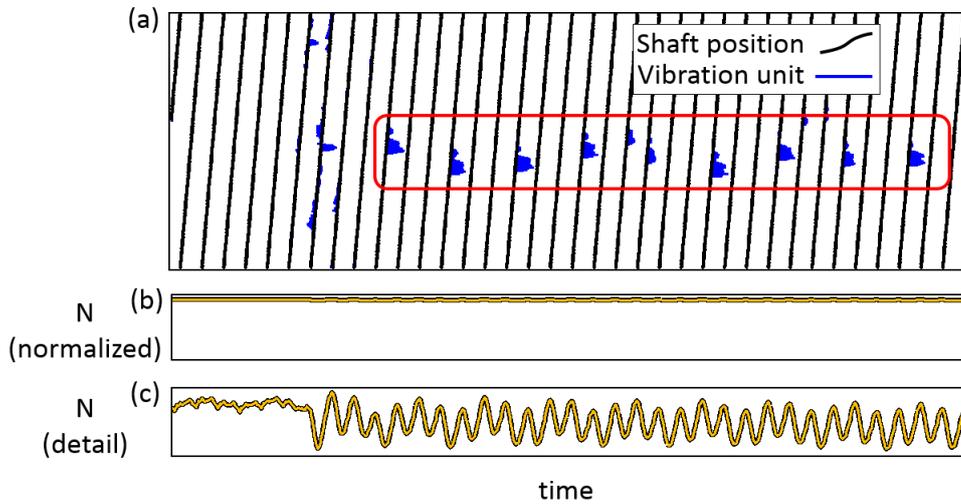


Figure 9: Repetitive shocks on a static case when a rotor interacts with a stator. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a case of the engine (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

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.

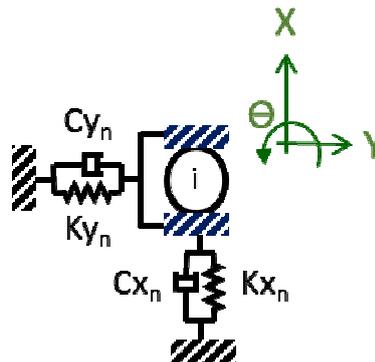


Figure 10: Lumped mass model Schema.

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.

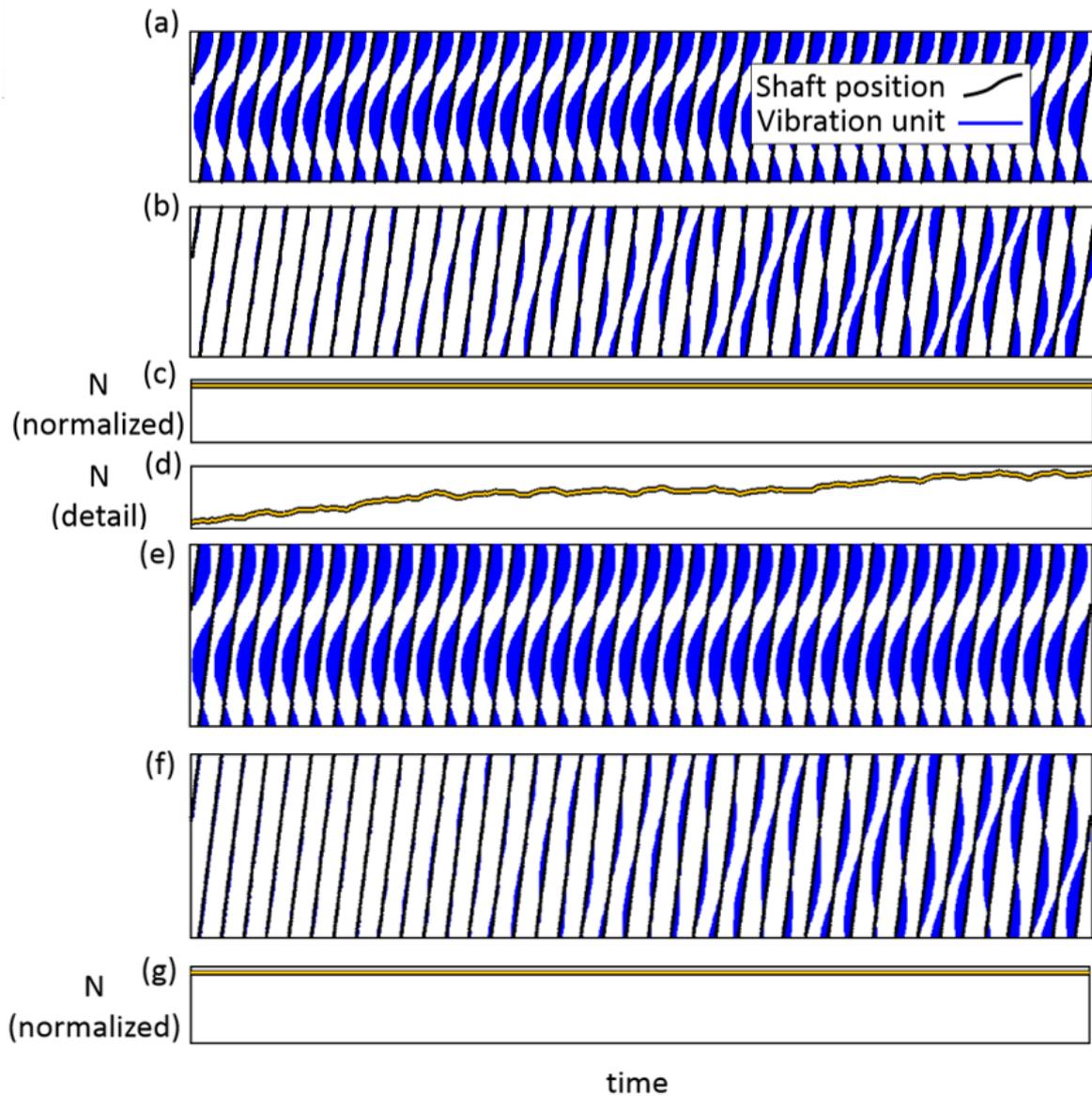


Figure 11: Application to a constant rigid unbalance and the progressive increase of an oil unbalance. Representation in the time/phase domain of the vibration amplitude tracked on the first order 1 N of the rotating speed (measurement (a), simulation (e)) and of the vibration amplitude tracked around 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.

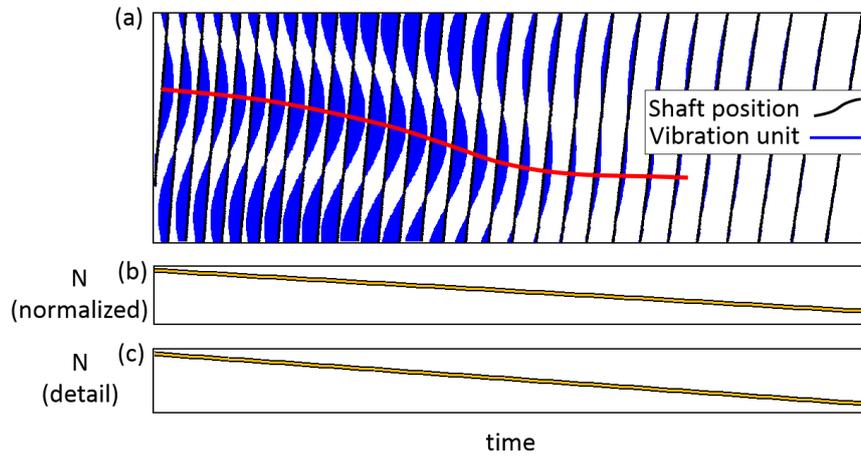


Figure 12: Simulation of the response to a crossing of the rotating speed and a mode of the shaft. Representation in the time/phase domain of the vibration amplitude of a strain gauge mounted on a stator close to the shaft (a). Time history of the normalized rotating speed (b) and of the detail of the 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.

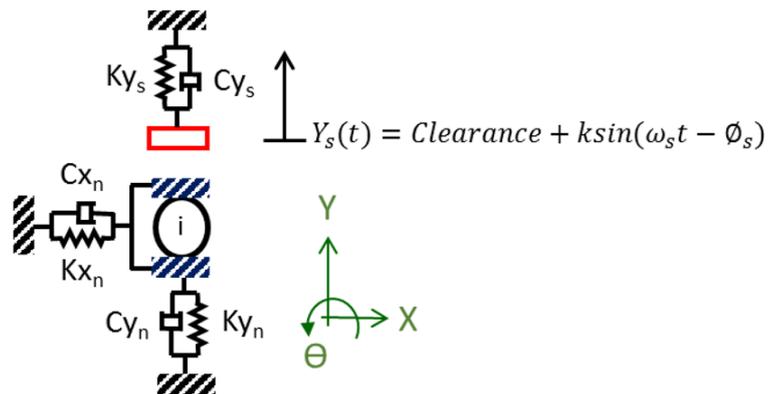


Figure 13: Rotor-stator contact model schema.

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.

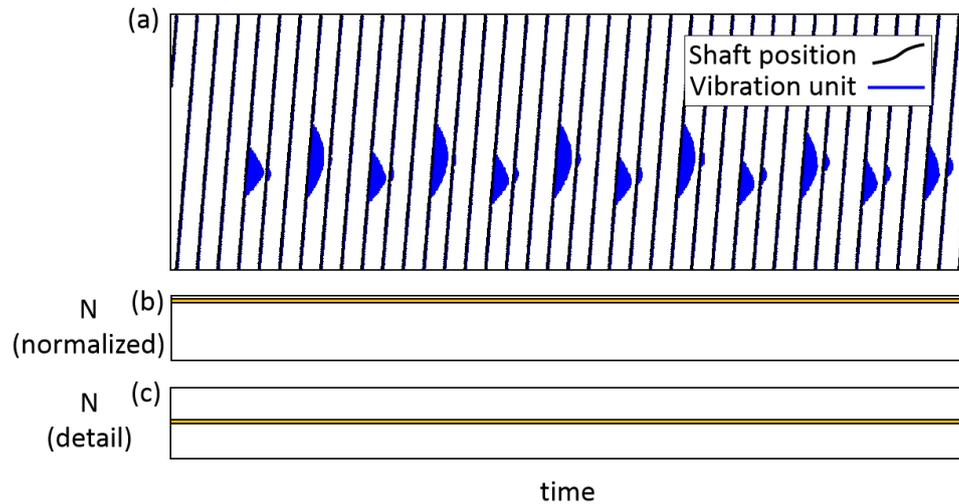


Figure 14: Simulation of the response to repetitive shocks. Representation in the time/phase domain of the vibration amplitude of an accelerometer mounted on a case of the engine (a). Time history of the normalized rotating speed (b) and of the detail of the rotating speed (c).

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. Soave¹, G. D'Elia¹, G. Dalpiaz¹

¹University 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 exploits 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 terms 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 directly 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 useful indicator. In fact, this statistic 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 et 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 \mathbf{t} , defined as a train of equispaced impulses. The main issue of MOMEDA is related to the definition of \mathbf{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 its 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 Cyclostationarity (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 a 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 cyclostationary 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 cyclostationarity maximization aims to point out the improvement given by the proposed algorithm, in particular in terms 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 \mathbf{x} , can be considered as composed by three different parts: an impulsive pattern \mathbf{s}_0 related to a local fault, a periodic component \mathbf{p} and a Gaussian background noise \mathbf{n} , all convolved with their respective Impulse Response Function (IRF), viz:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n \quad (1)$$

where \mathbf{g}_s , \mathbf{g}_p and \mathbf{g}_n are the IRFs related to \mathbf{s}_0 , \mathbf{p} and \mathbf{n} , respectively and $*$ refers to the convolution operation.

BD aims to estimate an inverse filter \mathbf{h} (assumed to be a FIR filter) in order to deconvolve the excitation \mathbf{s}_0 from \mathbf{x} minimizing the other contributions, such that:

$$\mathbf{s} = \mathbf{x} * \mathbf{h} = (\mathbf{s}_0 * \mathbf{g}_s + \mathbf{p} * \mathbf{g}_p + \mathbf{n} * \mathbf{g}_n) * \mathbf{h} \approx \mathbf{s}_0 \quad (2)$$

In the first part of eq. 2 the convolution can be expressed for discrete signals in matrix form, as follow:

$$\mathbf{s} = \mathbf{X}\mathbf{h} \quad (3a)$$

$$\begin{bmatrix} s[0] \\ \vdots \\ s[L-1] \end{bmatrix} = \begin{bmatrix} x[N-1] & \dots & x[0] \\ \vdots & \ddots & \vdots \\ x[L-1] & \dots & x[L-N-2] \end{bmatrix} \begin{bmatrix} h[0] \\ \vdots \\ h[N-1] \end{bmatrix} \quad (3b)$$

where L and N are the number of samples of \mathbf{s} and \mathbf{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 CY-CBD 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:

$$\alpha = \frac{k}{T} \quad (4)$$

where k is the sample index and T the cyclic period, e.g. the characteristic period of a bearing fault. The BD algorithm described 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:

$$ICS_2 = \frac{\sum_{k>0} |c_s^k|^2}{|c_s^0|^2} \quad (5)$$

with

$$c_s^k = \frac{1}{L-N+1} \sum_{n=N-1}^{L-1} |s[n]|^2 e^{-j2\pi \frac{k}{T} n} \quad (6a)$$

$$c_s^0 = \frac{\|\mathbf{s}\|^2}{L-N+1} \quad (6b)$$

or, in matrix form:

$$c_s^k = \frac{\mathbf{E}^H |\mathbf{s}|^2}{L-N+1} \quad (7a)$$

$$c_s^0 = \frac{\mathbf{s}^H \mathbf{s}}{L-N+1} \quad (7b)$$

where

$$|\mathbf{s}|^2 = [|s[N-1]|^2, \dots, |s[L-1]|^2]^T \quad (8a)$$

$$\mathbf{E} = \begin{bmatrix} e^{-j2\pi \frac{1}{T}(N-1)} & \dots & e^{-j2\pi \frac{K}{T}(N-1)} \\ \vdots & \ddots & \vdots \\ e^{-j2\pi \frac{1}{T}(L-1)} & \dots & e^{-j2\pi \frac{K}{T}(L-1)} \end{bmatrix} \quad (8b)$$

Now it is possible to rewrite eq. 5 by substituting eq. 7a and 7b as:

$$ICS_2 = \frac{|\mathbf{s}|^{2H} \mathbf{E} \mathbf{E}^H |\mathbf{s}|^2}{|\mathbf{s}^H \mathbf{s}|^2} \quad (9)$$

Finally, ICS_2 can be rewritten by substituting eq. 3a into eq. 9, viz:

$$ICS_2 = \frac{\mathbf{h}^H \mathbf{X}^H \mathbf{W} \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XWX} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XX} \mathbf{h}} \quad (10)$$

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:

$$\mathbf{W} = \text{diag} \left(\frac{\mathbf{E} \mathbf{E}^H |\mathbf{s}|^2}{(L-N+1) \mathbf{s}^H \mathbf{s}} \right) (L-N+1) = \begin{bmatrix} \ddots & & 0 \\ & \frac{\mathbf{E} \mathbf{E}^H |\mathbf{s}|^2}{(L-N+1)} & \\ 0 & & \ddots \end{bmatrix} \begin{matrix} (L-N+1) \\ \sum_{n=N-1}^{L-1} s[n]^2 \end{matrix} \quad (11)$$

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:

$$\mathbf{R}_{XWX} \mathbf{h} = \mathbf{R}_{XX} \mathbf{h} \lambda \quad (12)$$

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:

$$x(n) = \sum_{i=1}^L C_i J_0 \left(\frac{\beta_i n}{L} \right), n = 0, 1, \dots, L-1 \quad (13)$$

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:

$$C_i = \frac{2}{L^2 [J_1(\beta_i)]^2} \sum_{n=0}^{L-1} n x(n) J_0 \left(\frac{\beta_i n}{L} \right) \quad (14)$$

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:

$$\beta_i \approx \frac{2\pi f_i L}{f_s} \quad (15)$$

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 cyclostationarity 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:

$$c_F^k = \frac{2}{(L-N+1)^2 [J_1(\beta_k)]^2} \sum_{n=N-1}^{L-1} n |s(n)|^2 J_0\left(\frac{\beta_k n}{L-N+1}\right) \quad (16)$$

where the roots of the zero-order Bessel function can be expressed as:

$$\beta_k = \frac{2\pi \frac{k}{T_s} (L-N+1)}{f_s} \quad (17)$$

where the term k/T_s represents the cyclic frequency, according to eq. 4.

Eq. 16 can be expressed in matrix form as follows:

$$c_F^k = \frac{2}{(L-N+1)^2} \frac{\mathbf{J}_{0_n}^H |\mathbf{s}|^2}{\mathbf{J}_1 \mathbf{J}_1^H} \quad (18)$$

where

$$|\mathbf{s}|^2 = [|s[N-1]|^2, \dots, |s[L-1]|^2]^T \quad (19a)$$

$$\mathbf{J}_{0_n} = \begin{bmatrix} J_0\left(\frac{\beta_1(N-1)}{L-N+1}\right)(N-1) & \dots & J_0\left(\frac{\beta_K(N-1)}{L-N+1}\right)(N-1) \\ \vdots & \ddots & \vdots \\ J_0\left(\frac{\beta_1(L-1)}{L-N+1}\right)(L-1) & \dots & J_0\left(\frac{\beta_K(L-1)}{L-N+1}\right)(L-1) \end{bmatrix} \quad (19b)$$

$$\mathbf{J}_1 = [J_1(\beta_1) \dots J_1(\beta_k) \dots J_1(\beta_K)] \quad (19c)$$

Starting from eq. 7b and eq. 18, eq. 5 can be rewritten as:

$$ICS_{2F} = \frac{4}{(L-N+1)^2 |\mathbf{J}_1 \mathbf{J}_1^H|^2} \frac{|\mathbf{s}|^{2H} \mathbf{J}_{0_n} \mathbf{J}_{0_n}^H |\mathbf{s}|^2}{|\mathbf{s}^H \mathbf{s}|^2} \quad (20)$$

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:

$$P[\mathbf{s}] = \frac{4}{(L-N+1)^2} \sum_{k>0} \frac{J_{0_n}^k (J_{0_n}^k |s|^2)}{[J_1(\beta_k)]^2} = \frac{4}{(L-N+1)^2} \frac{\mathbf{J}_{0_n} \mathbf{J}_{0_n}^H |\mathbf{s}|^2}{|\mathbf{J}_1 \mathbf{J}_1^H|^2} \quad (21)$$

Substituting eq. 3a and eq. 21 into eq. 20 it is possible to write the ICS_2 as a generalized Rayleigh quotient:

$$ICS_{2F} = \frac{\mathbf{h}^H \mathbf{X}^H \mathbf{W} \mathbf{X} \mathbf{h}}{\mathbf{h}^H \mathbf{X}^H \mathbf{X} \mathbf{h}} = \frac{\mathbf{h}^H \mathbf{R}_{XW} \mathbf{h}}{\mathbf{h}^H \mathbf{R}_{XX} \mathbf{h}} \quad (22)$$

where the weighting matrix can be expressed as:

$$\mathbf{W} = \text{diag}\left(\frac{P[|\mathbf{s}|^2]}{\mathbf{s}^H \mathbf{s}}\right) (L-N+1) = \begin{bmatrix} \ddots & & 0 \\ & P[|\mathbf{s}|^2] & \\ 0 & & \ddots \end{bmatrix} \frac{1}{\sum_{n=N-1}^{L-1} s[n]^2} \quad (23)$$

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 \mathbf{p} as follows:

$$\mathbf{x} = \mathbf{s}_0 * \mathbf{g}_s + \mathbf{n} * \mathbf{g}_n \quad (24)$$

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:

$$\mathbf{x}_{SDOF} = Ae^{-\zeta\omega_n t} \sin(\omega_d t) \quad (25)$$

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:

$$\omega_d = \omega_n \sqrt{1 - \zeta^2} \quad (26)$$

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
$\omega_n(\text{rad/s})$	18.84	62.83

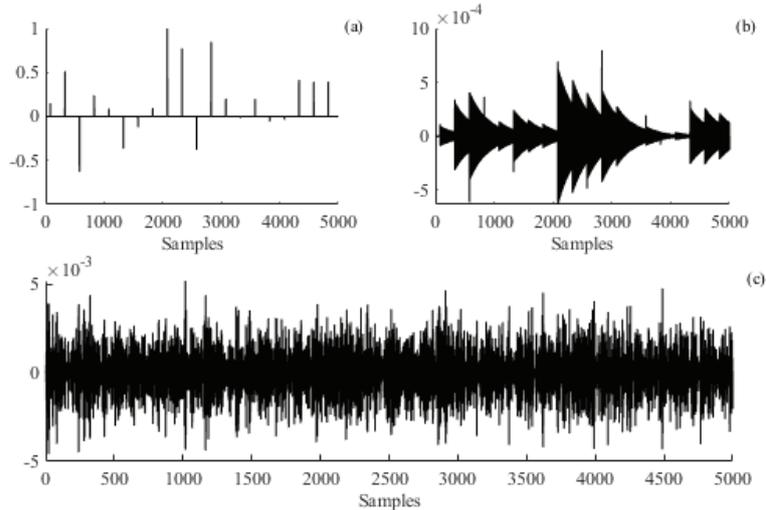


Figure 1: Simulated signal used for the validation: impulsive pattern with Gaussian distributed amplitudes (a), impulsive pattern convolved with its IRF (b) and overall signal (c).

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.

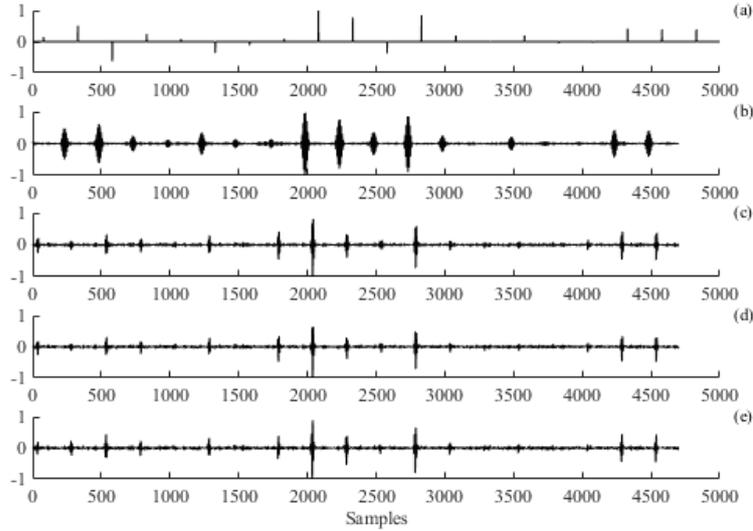


Figure 2: Comparison between the target source (a) and the patterns estimated through the FBBD method at increasing number of cyclic frequencies (from b to e).

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

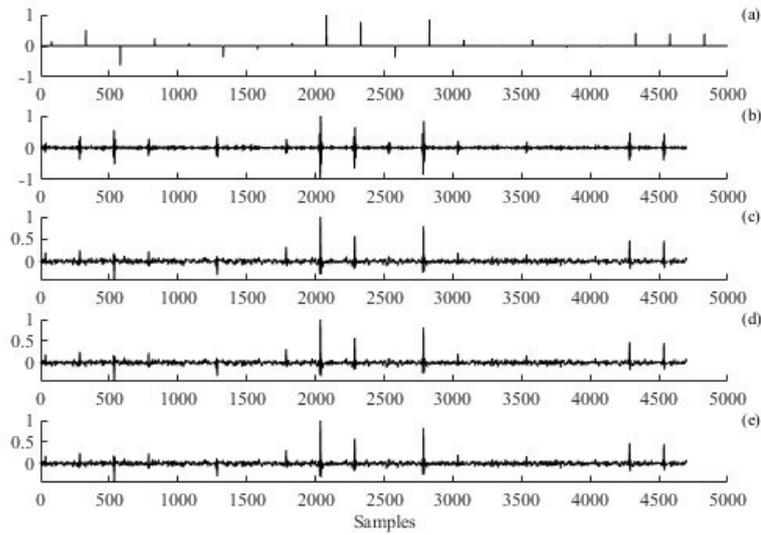


Figure 3: Comparison between the target source (a) and the patterns estimated through the CYCBD method at increasing number of cyclic frequencies (from b to e).

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.

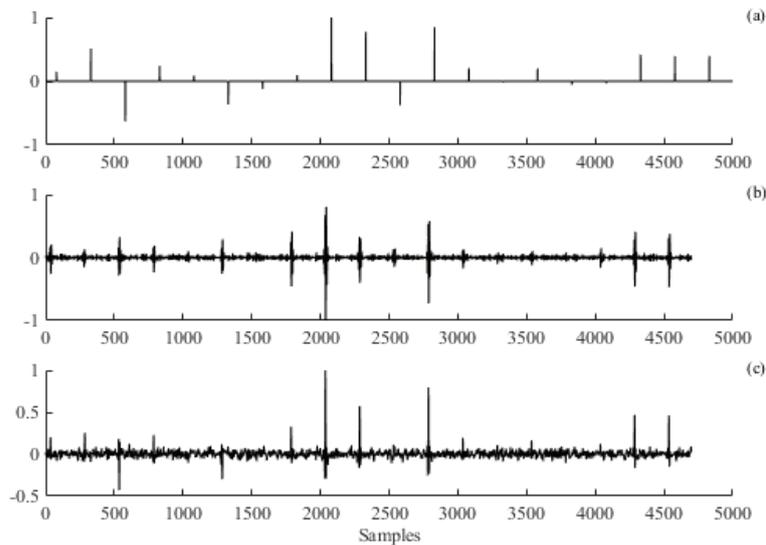


Figure 4: Comparison between the target source (a) and the sources estimated through the FBBD method (b) and the CYCBD method (c).

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.

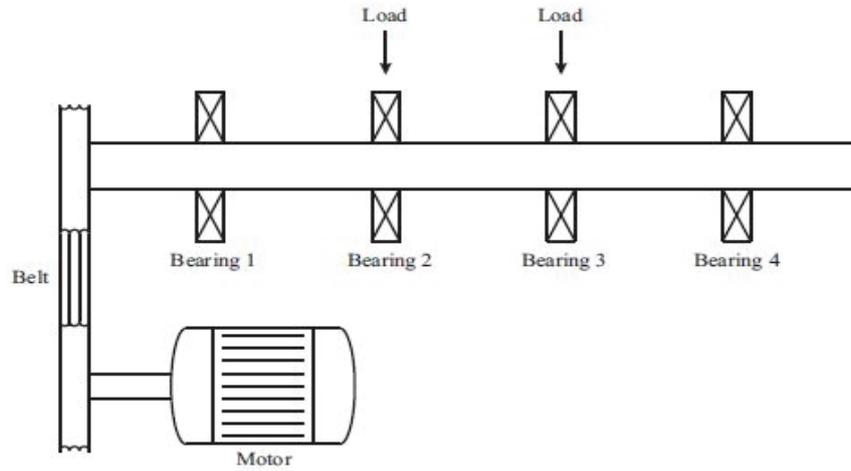


Figure 5: Scheme of the experimental test rig.

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 FBBDD 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 FBBDD seems to permit the identification of the lower amplitude impulses better than 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 FBBDD, 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 FBBDD method for the extraction of hidden cyclostationary 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 27.7kN 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).

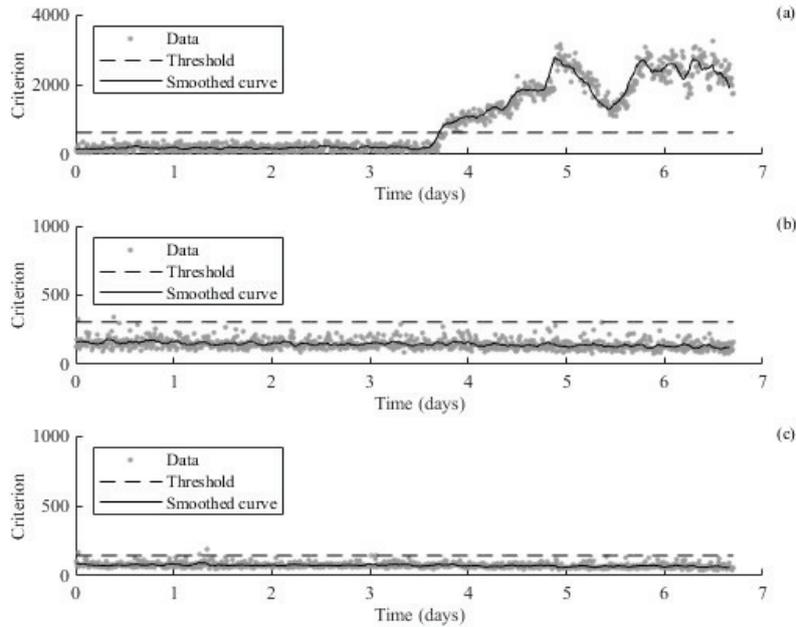


Figure 6: Final values of maximized criterion during the endurance test taking account of BPFO (a), BPFi (b) and BSF (c).

The FBB 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 FBB 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 Cyclostationarity and define a new BD criterion. The choice of the Fourier-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 FBBBD, 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 FBBBD 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 FBBBD 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*. *Geophysical Prospecting* 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 HEBRARD¹

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.

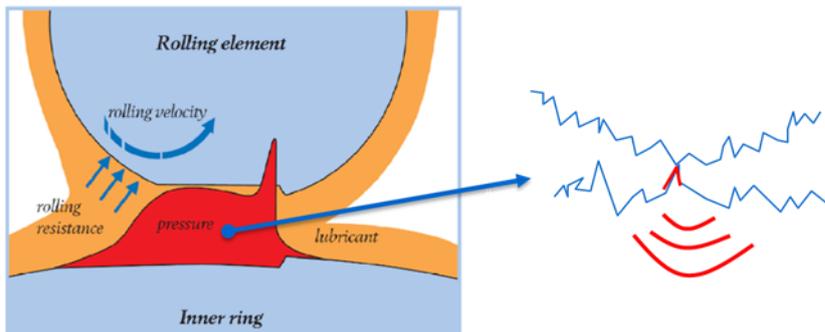


Figure 1. AE signal from asperity in rolling contact

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 200 μ m. 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.

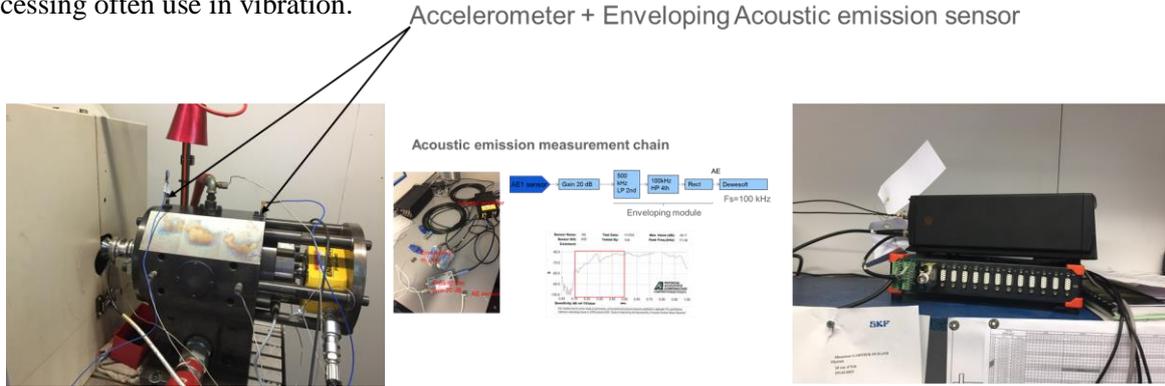


Figure 2 Experimental set-up

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	200 kHz	5-40 kHz
Mistras WD AE sensor	2 gHz	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.

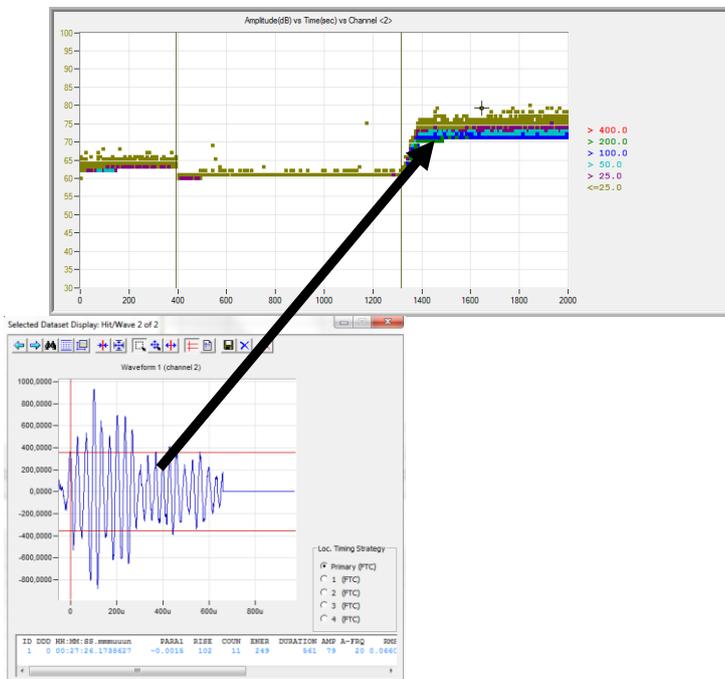


Figure 3 acquisition with smart threshold (above graph) and waveform and hit correspondence (graph below).

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.

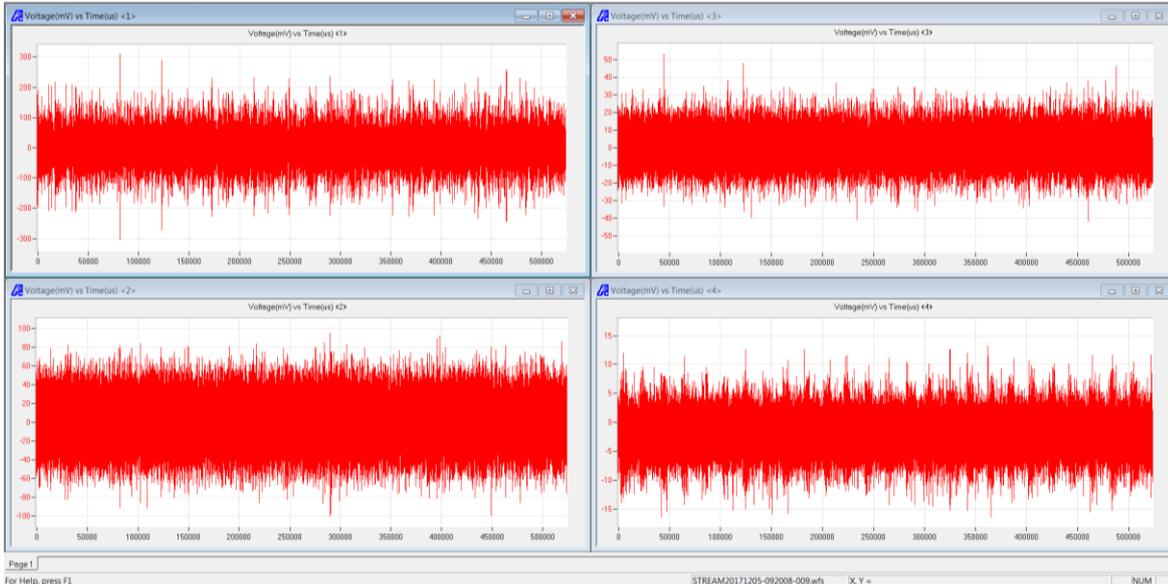
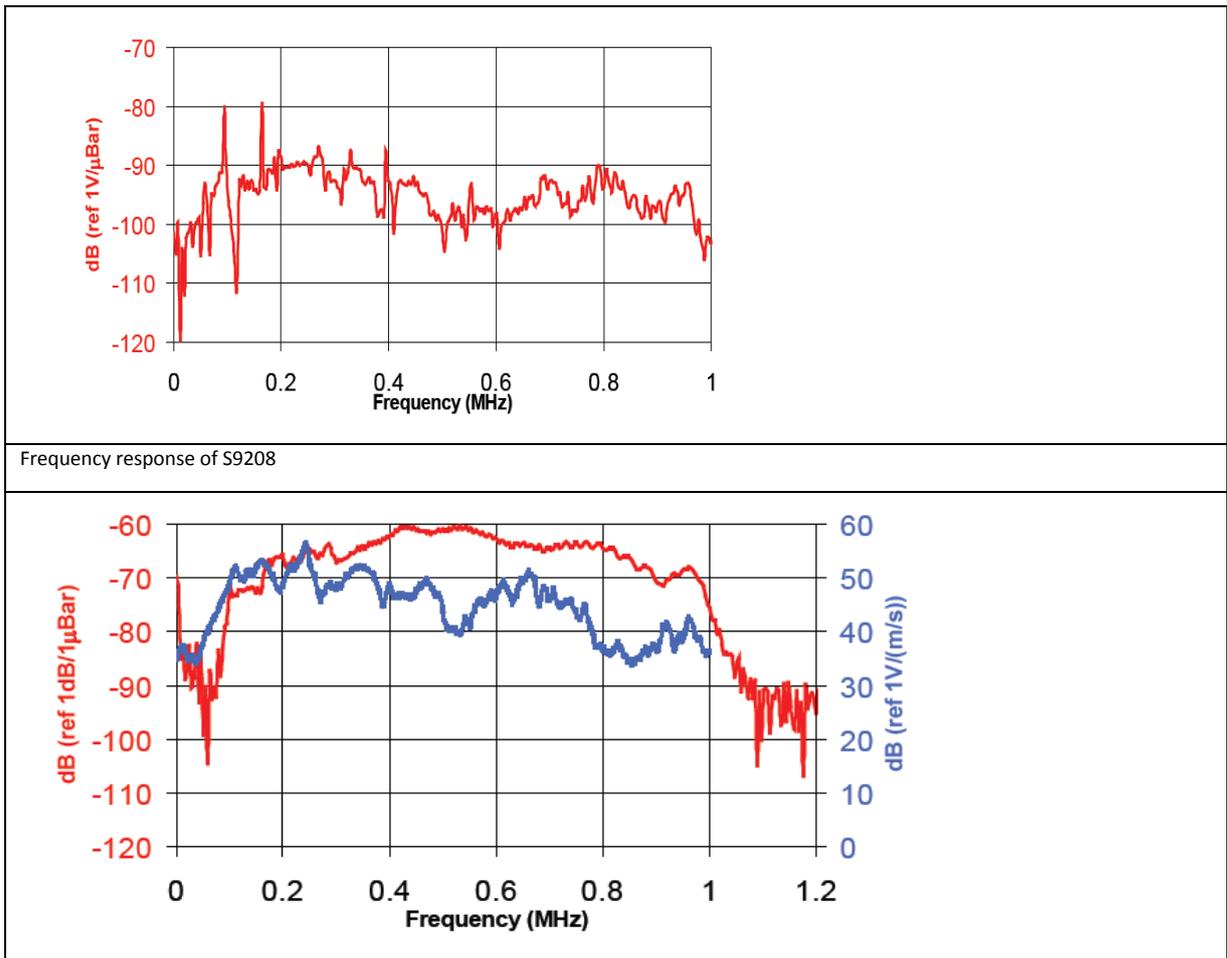


Figure 4 streaming acquisition during 5 sec at 2 Mega sample per second on the four channels.

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.



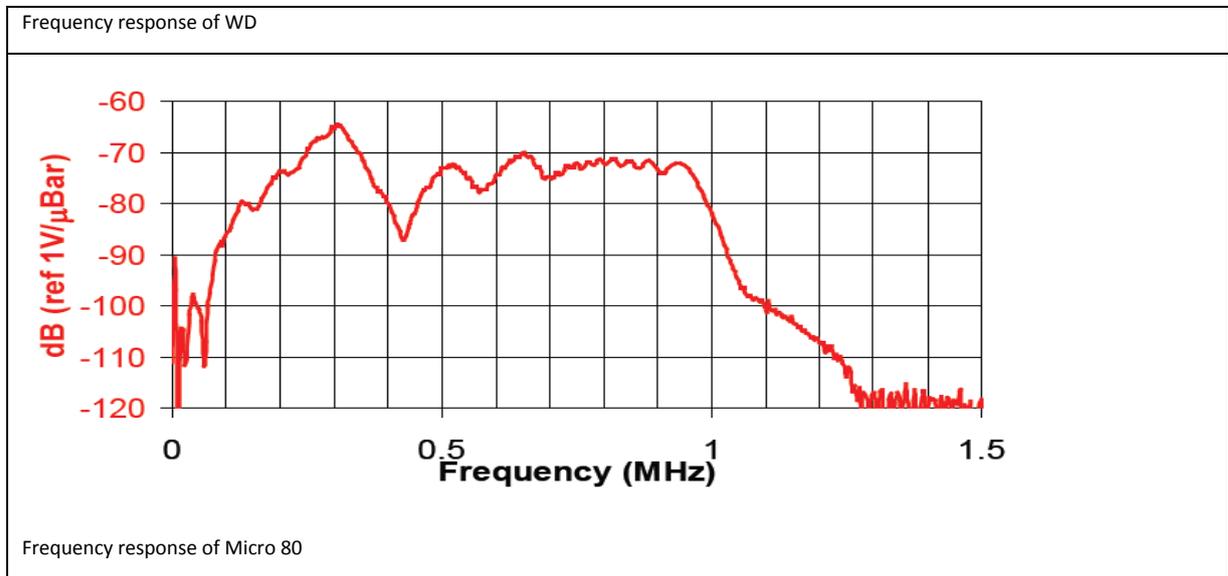


Figure 5 calibration curve of sensors S9203, WD and micro 80.

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).

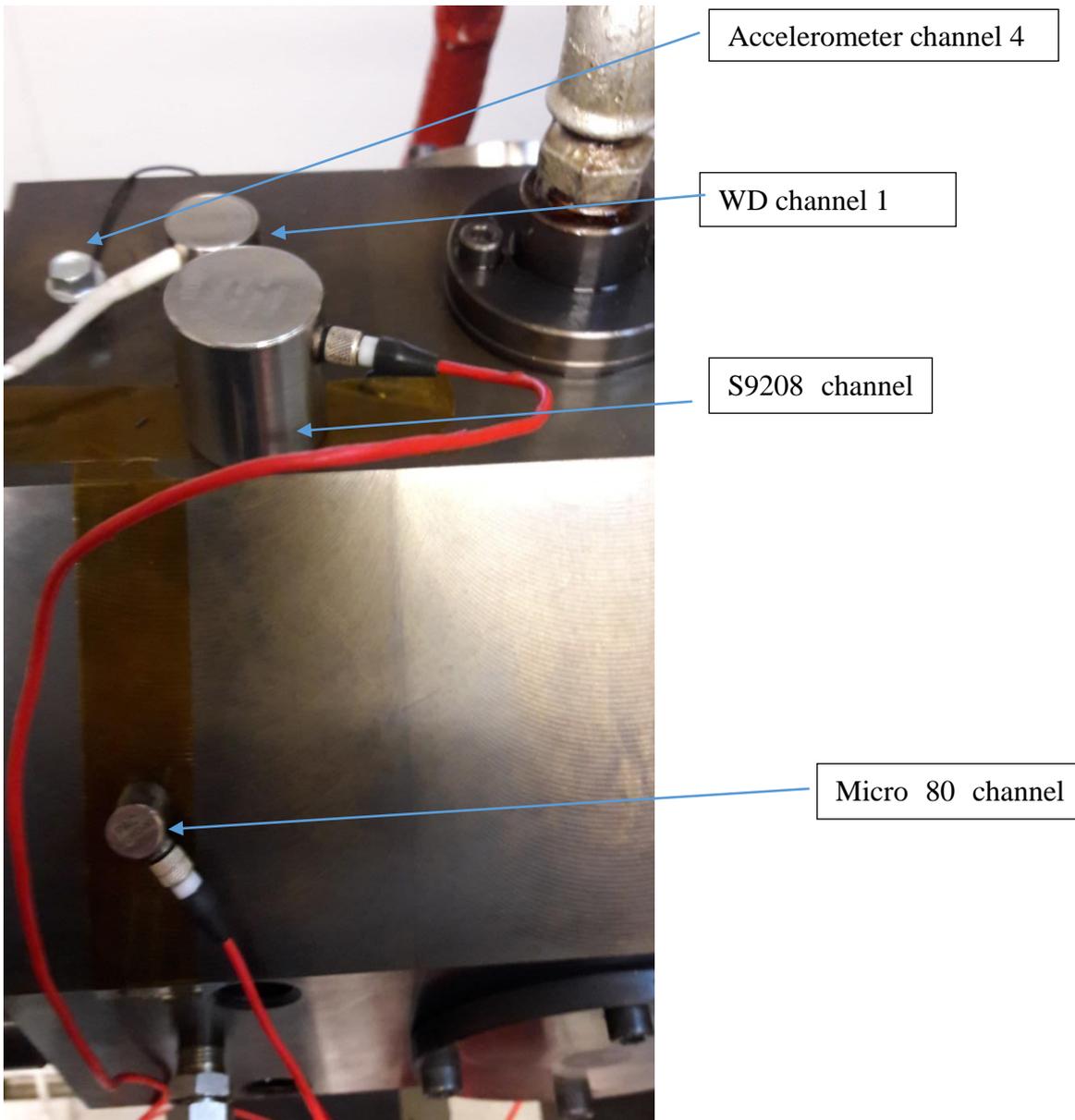


Figure 6 picture of the different sensors used for the monitoring and coupling mode.

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 compared and the WD (channel 1) and micro 80 (channel 2) are giving better results than accelerometer and S9209.

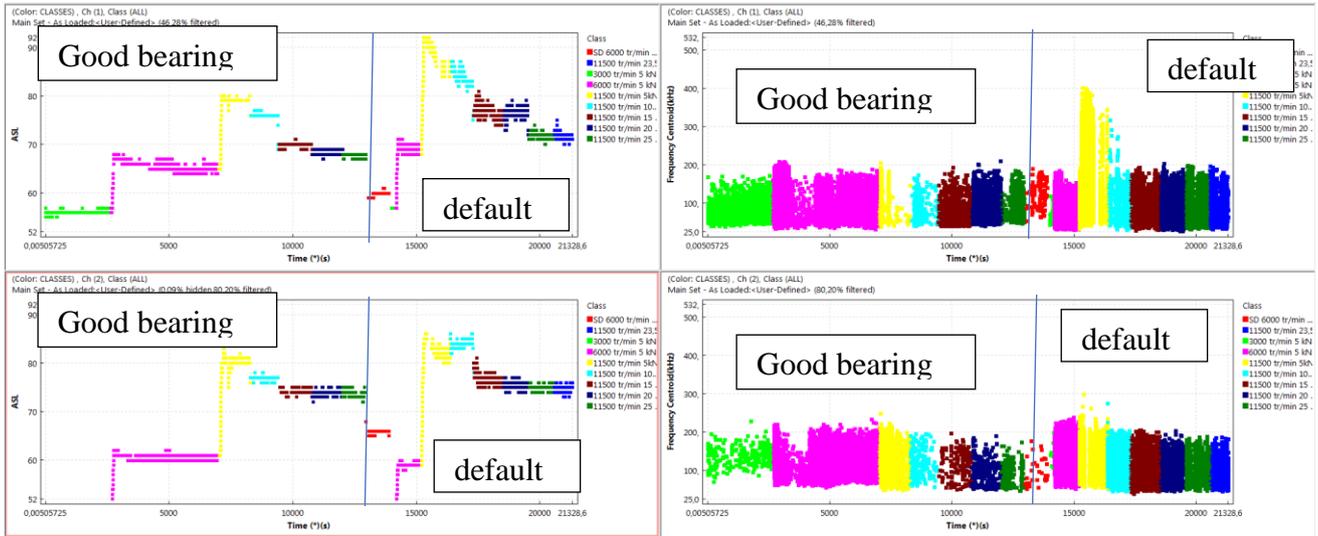


Figure 7 evolution of ASL and frequency centroid versus time with and without default. On the top graph channel 1 (WD) at the bottom channel 2 (Micro 80).

A clear difference of behavior can be seen above 6000 rpm at 5 kN and it enhances at 11 kN on the energy of the AE signal on transients. Also, the center mass of the frequency spectrum (called frequency centroid) increases a lot for the bearing with default it exhibits the best power of discrimination using our pattern 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).

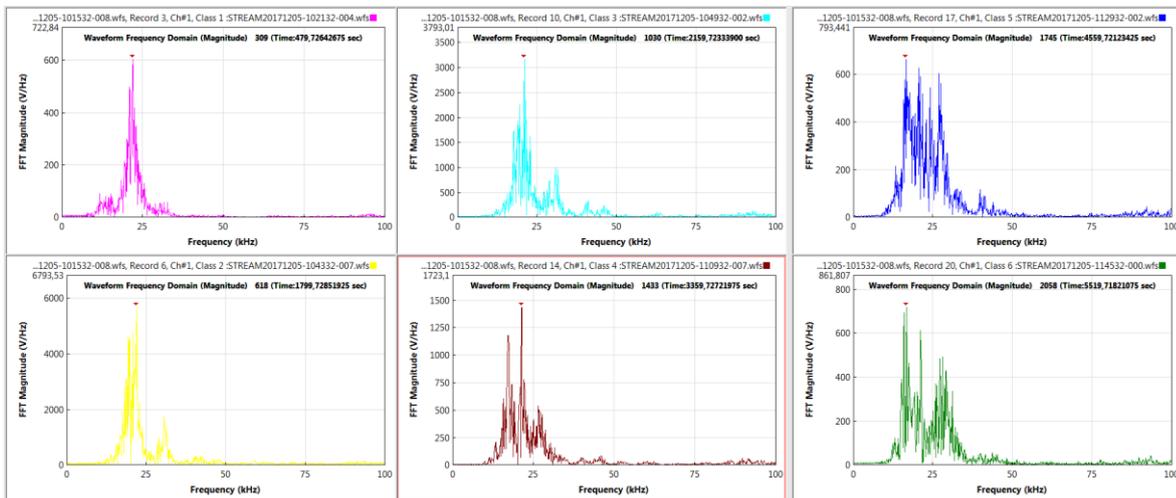


Figure 8 Fourier transform on 5s files on channel 1 for different case of load levels and rotating speed.

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.

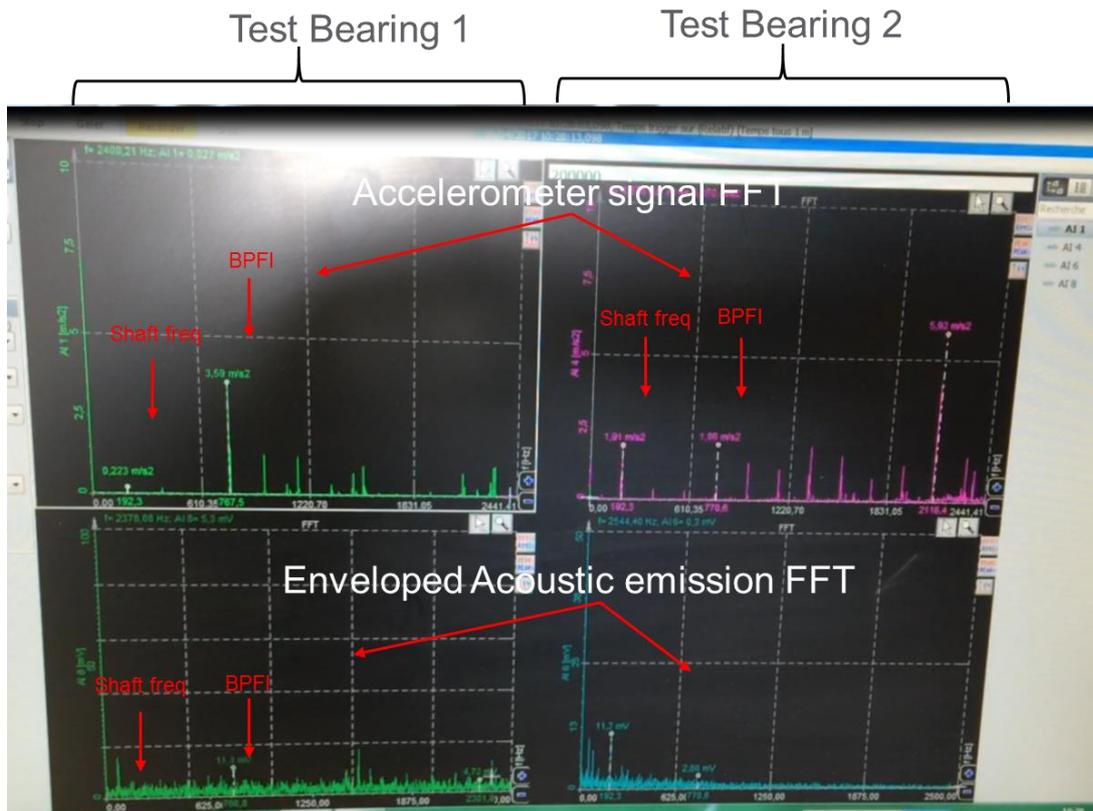


Figure 9 Enveloped spectrum with bearing frequencies identified

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.m.s, 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] Levallois, 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 SLEIMAN¹, Amani RAAD¹, Souhayb KASS², Jerome ANTONI²

¹Univ Libanaise, Faculté de Génie, Tripoli, LIBAN

²Univ 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 k th element of the target vector t_n and y_{kn} is the k th element of the prediction vector y_n for x_n , given a training data set $\{x_n, t_n\}$, $n=1 \dots N$.

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.

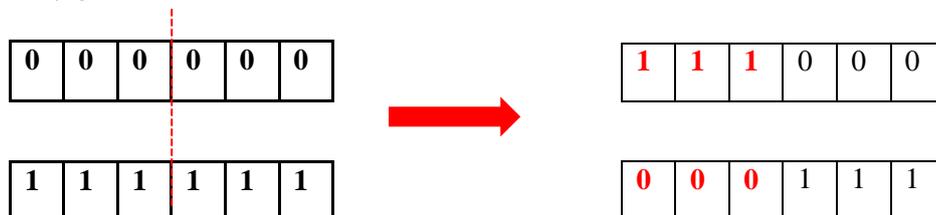
Once the fitness function is properly defined, the genetic algorithm generally starts by randomly generating the initial population. It should be large enough so that any solution in the search space can be later engendered [27]. Then, the algorithm loops over an iteration process to make the population 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 P_c , 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 P_m , 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.

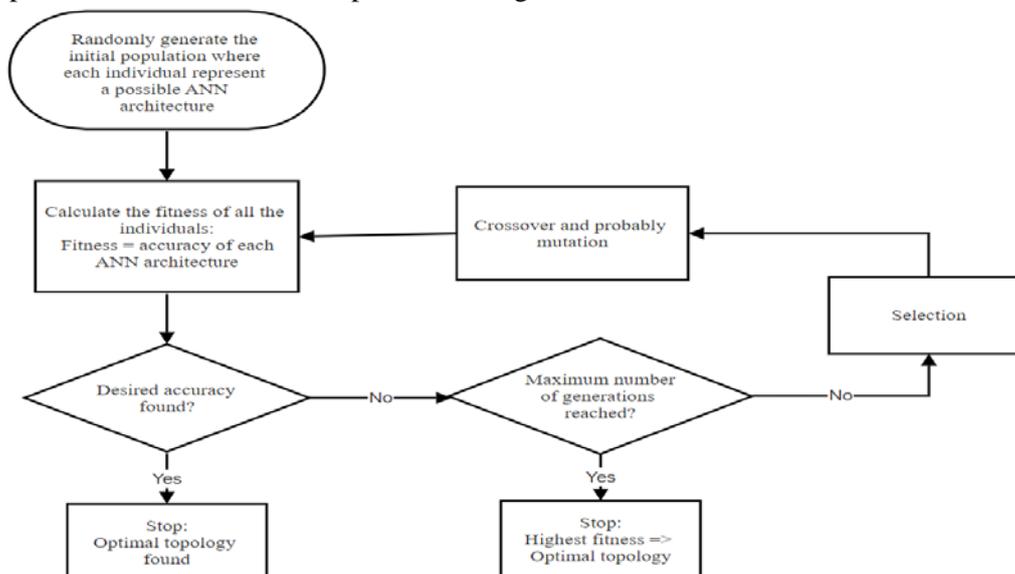
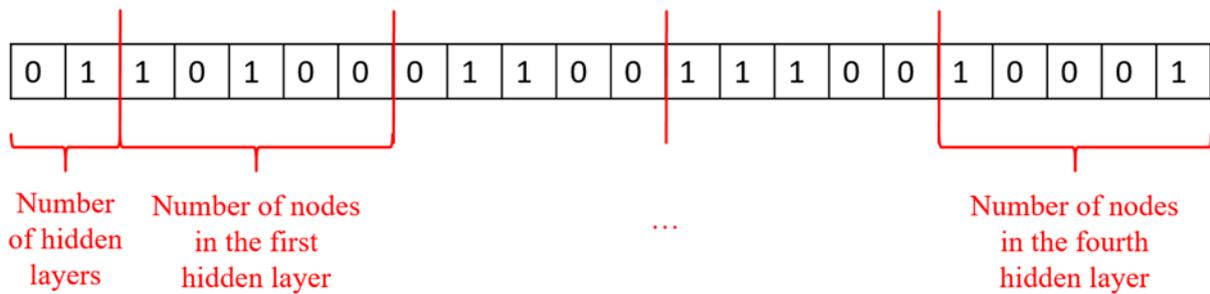


Fig.1 Flowchart of the ANN architecture optimization algorithm

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 divided 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:



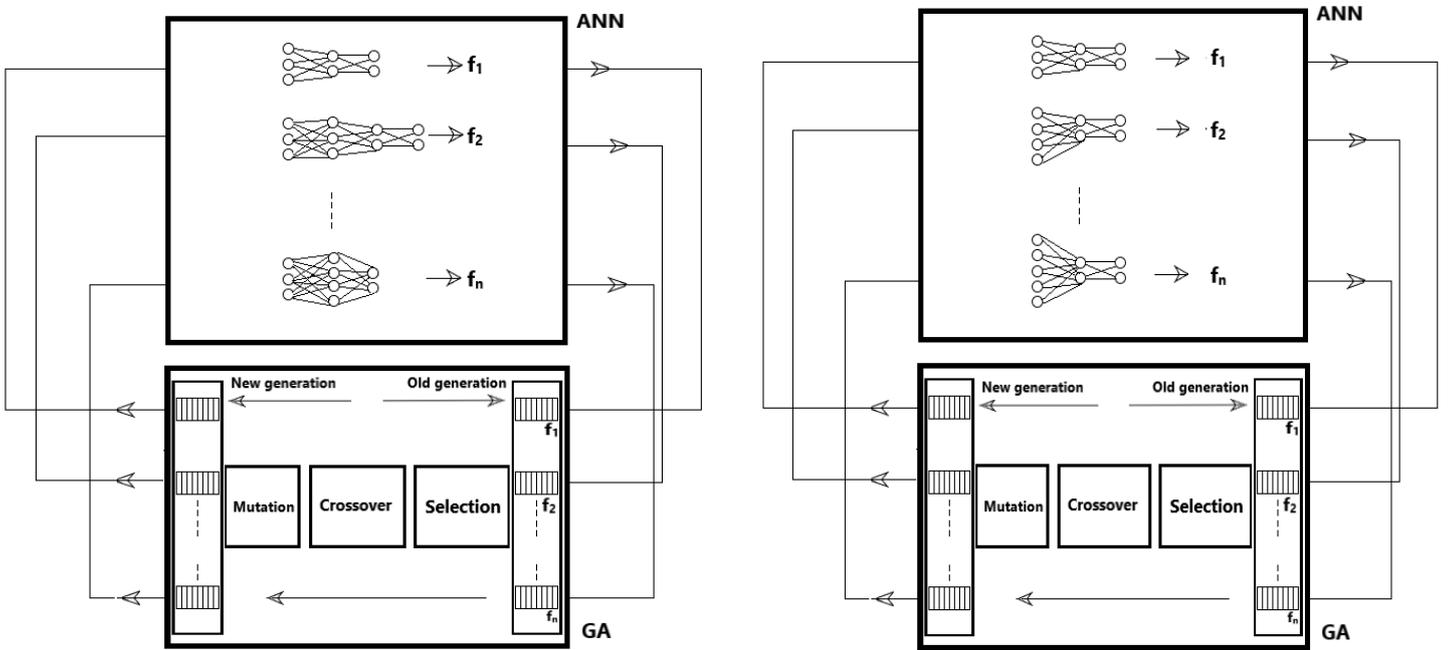


Fig.2 Combination between ANN and GA in two cases of ANN architecture and Feature optimization.

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	$\frac{\max x_i }{RMS}$
Shape factor	$\frac{RMS}{\frac{1}{N}\sum_{i=1}^N x_i }$
Impulse factor	$\frac{\max x_i }{\frac{1}{N}\sum_{i=1}^N x_i }$
Margin factor	$\frac{\max x_i }{\left(\frac{1}{N}\sum_{i=1}^N x_i ^2\right)^{\frac{1}{2}}}$

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:

$$\begin{cases} \alpha_1^L = \alpha_{c1} - X_{\alpha_{c1}} \\ \alpha_1^U = \alpha_{c1} + X_{\alpha_{c1}} \end{cases} \quad (3)$$

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:

$$PPF_1 = \begin{cases} \frac{m_1 - \lambda_{1-p}}{m_1} & m_1 \geq \lambda_{1-p} \\ 0 & m_1 < \lambda_{1-p} \end{cases} \quad (4)$$

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 α_{c1}^{corr} such that $\alpha_{c1}^{corr} = \alpha$ where the highest peak m_1 was found. Thus, B_2 is centered on $\alpha_{c2} = 2 * \alpha_{c1}^{corr}$ 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:

$$\alpha_{c_i} = \frac{i}{i-1} \sum_{n=1}^{i-1} \frac{\alpha_{c_n}^{corr}}{n}, \quad i > 1 \quad (5)$$

Once the PPF_i s 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:

BPM	BPM	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_BPM	PPF_BPM	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:

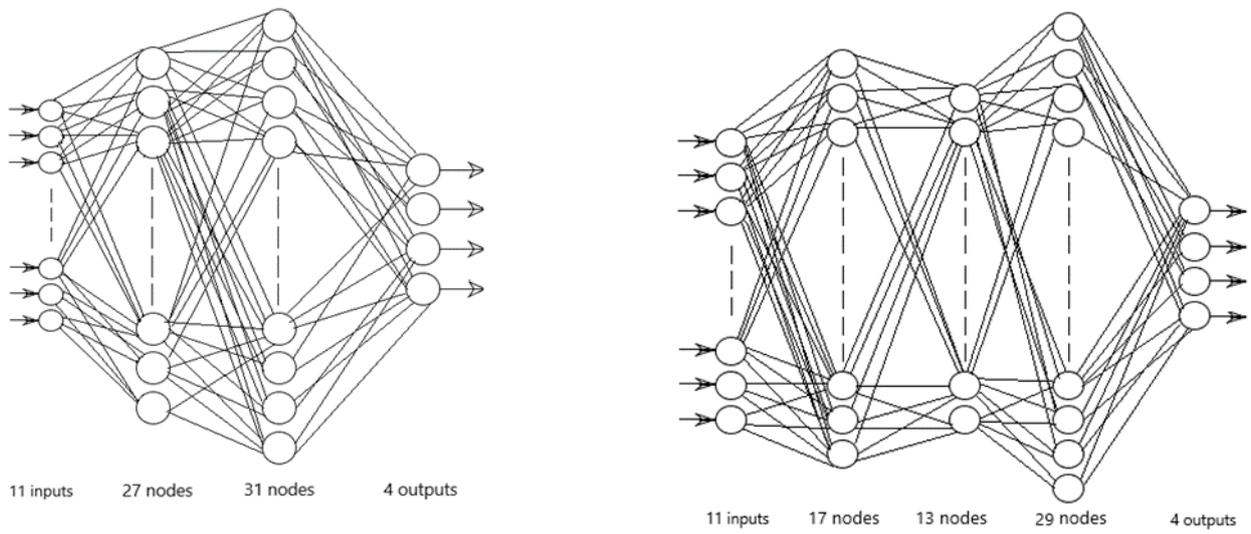


Fig.3 Two possible architectures obtained by the GA

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.

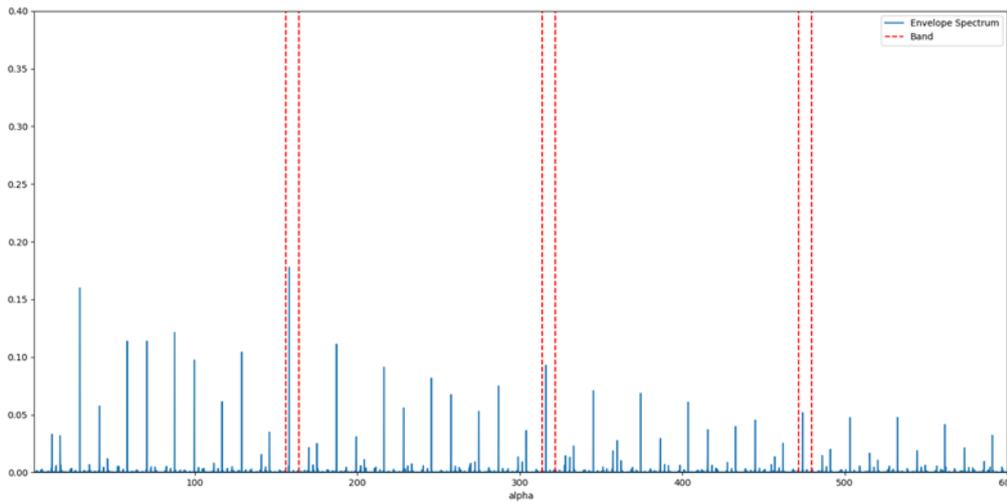


Fig.4 Envelope Spectrum for an inner race faulty signal

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.

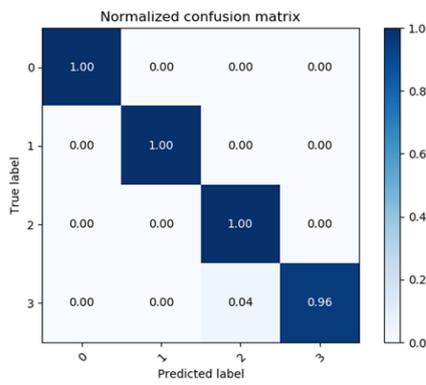


Fig.5 Normalized confusion matrix

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:

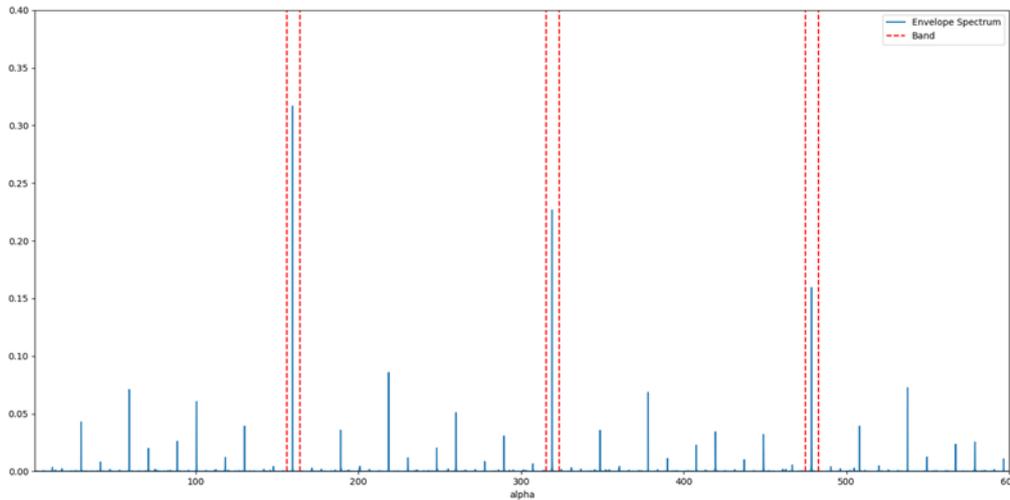


Fig.6 Envelope Spectrum for an inner race faulty signal

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:

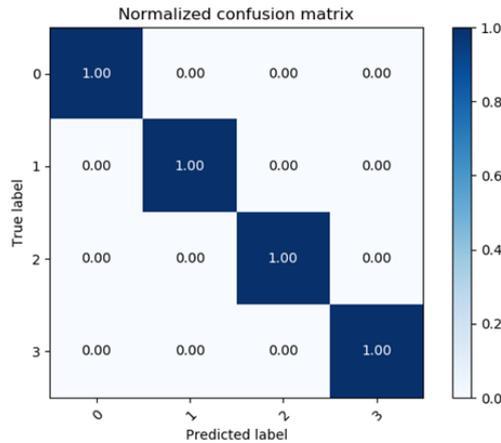


Fig.7 Normalized confusion matrix

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. Basavanahally, 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 evolutionary origins 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-Nimes. 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.ymsp.2015.04.021

Electrical motors

Long-term vibration monitoring of induction motors in the food industry with low-cost MEMS accelerometers

Agusmian Partogi OMPUSUNGGU¹, Karel JANSSEN¹, Vladimir DZYUBA¹

¹Flanders 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 $\mu\text{g}/\sqrt{\text{Hz}}$** .

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 $\mu\text{g}/\sqrt{\text{Hz}}$** . 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.

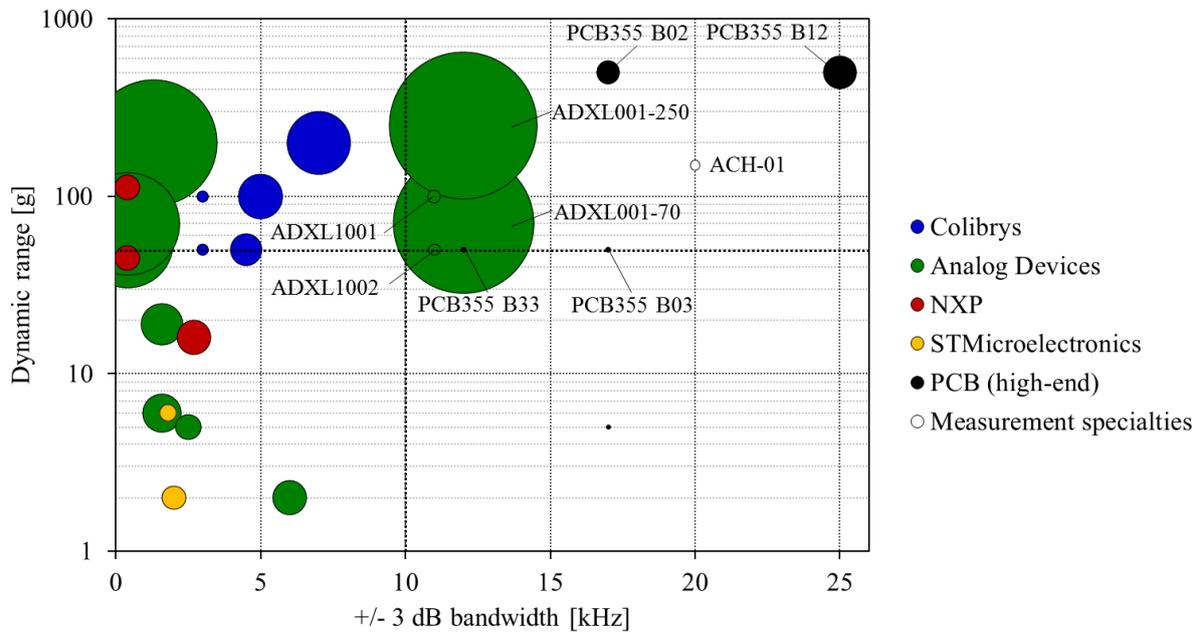


Figure 1: The market overview of analogue MEMS accelerometers updated version of [5]. The dashed lines indicate the minimum requirements set for condition monitoring applications. The diameter of the circles indicates the noise density as specified in the datasheets.

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.

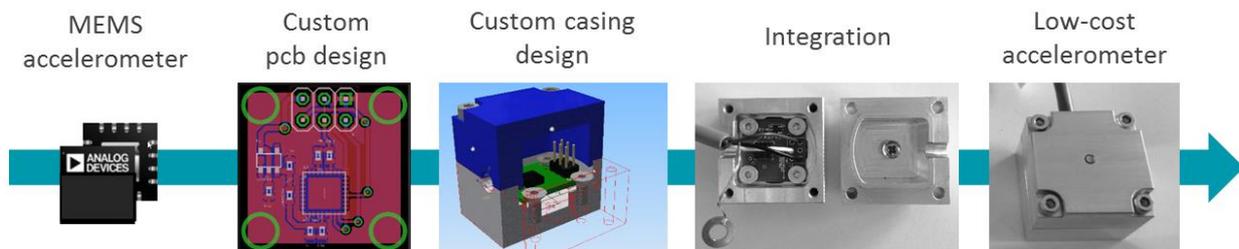


Figure 2: Sensor integration and packaging process.

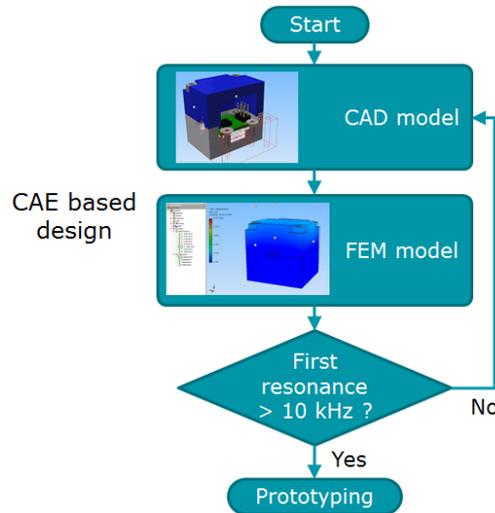


Figure 3: Packaging design flowchart.

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 is 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.

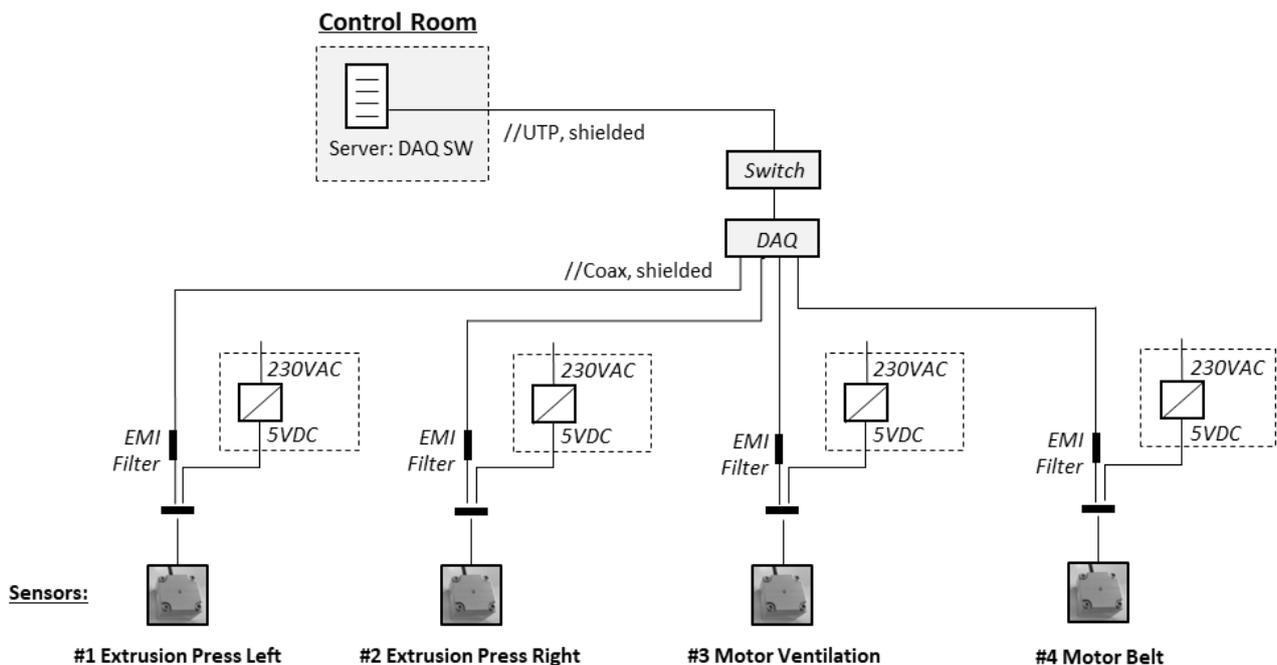


Figure 4: Architecture of the monitoring system installed in a production line.

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.

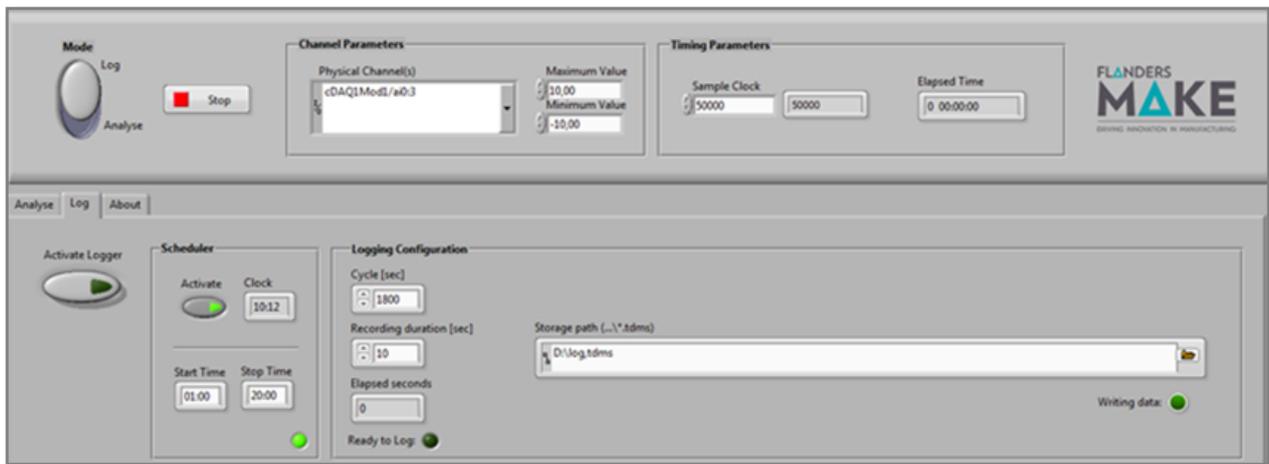


Figure 5: Custom data recording program.

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.

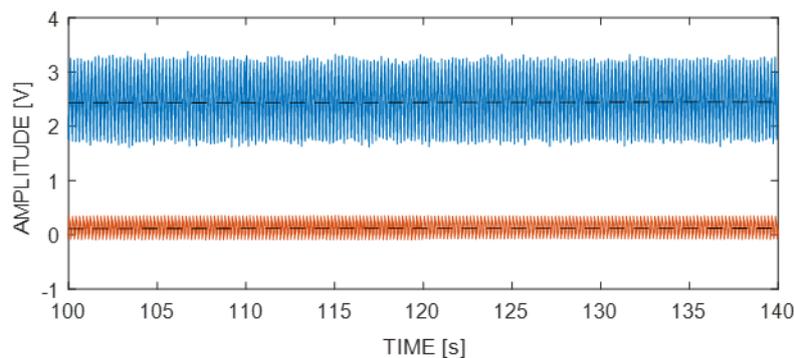


Figure 6: Raw sensor data of a functional (top) and failing (bottom) sensor, sampling at 50 kHz.

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.

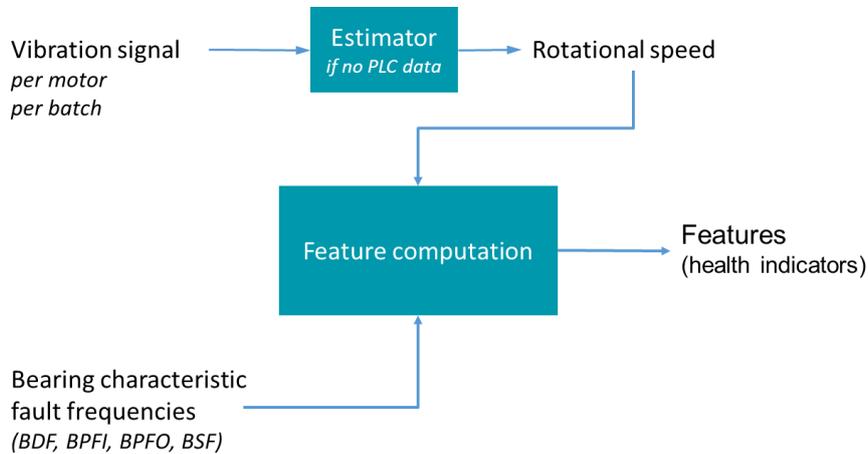


Figure 7: Inputs and steps of the feature computation algorithm.

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].

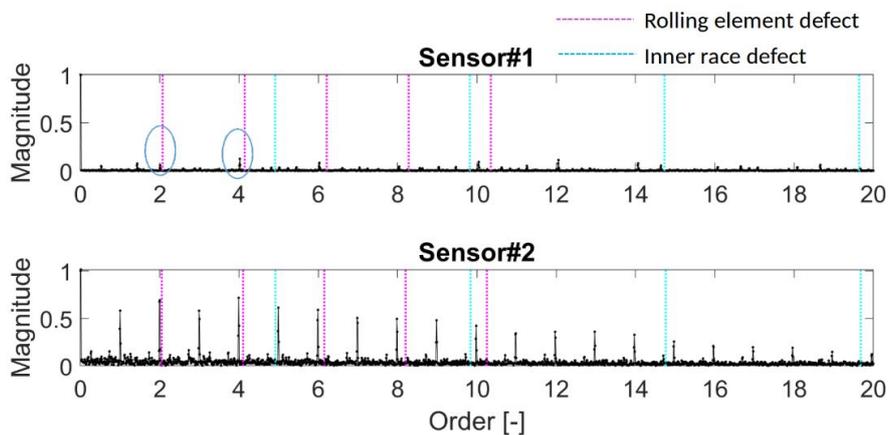


Figure 8: Final step of feature computation: matching *envelope spectral* peaks with bearing fault frequencies. Close matches for *Sensor#2* likely indicate bearing fault.

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.

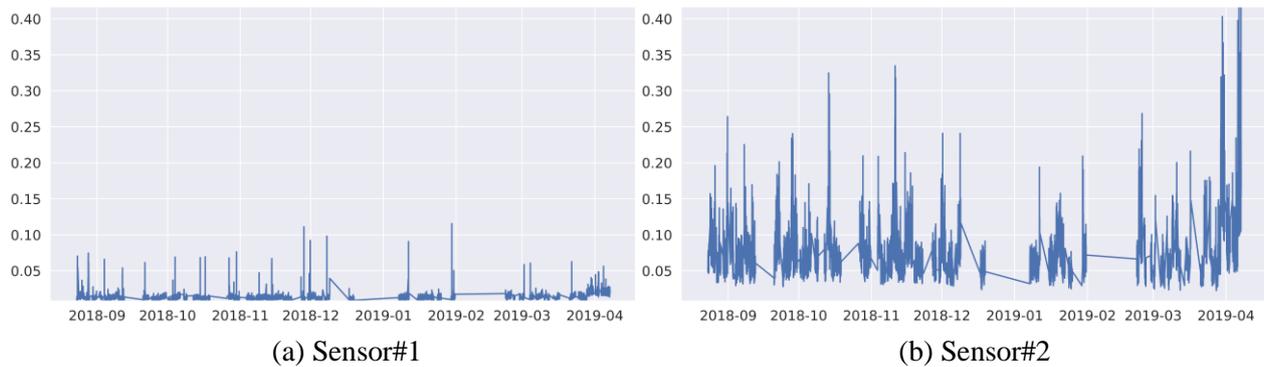


Figure 9: Evolution of the global bearing health indicator for two motors over eight months (missing values correspond to non-production periods, e.g., maintenance or holidays).

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 OUISSE¹, Gael CHEVALLIER¹, Damien VAILLANT²

¹Institut FEMTO-ST, Université de Bourgogne Franche-Comté,
24 Rue de l'Épitaphe, 25000 Besançon, France

²Alstom 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).

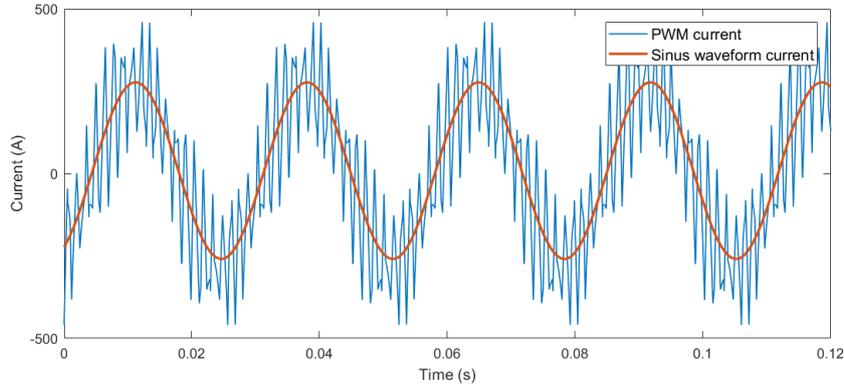


Figure 1: Current waveform with PWM or sinus

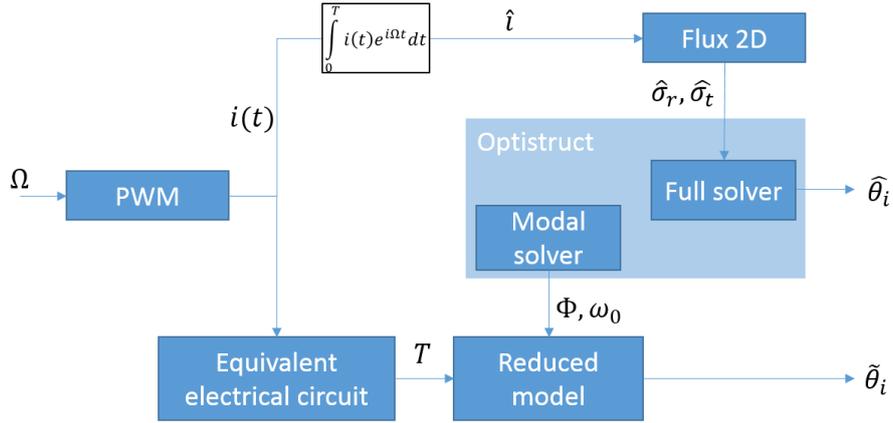


Figure 2: Workflow of methodologies for weakly coupled model and reduced model

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.

$$\begin{aligned}
 \overrightarrow{rot}(\vec{E}) &= -\frac{\partial \vec{B}}{\partial t} \\
 \overrightarrow{div}(\vec{B}) &= 0 \\
 \overrightarrow{rot}(\vec{H}) &= \vec{J} \\
 \vec{J} &= \sigma \vec{E} \\
 \vec{B} &= \mu \vec{H}
 \end{aligned} \tag{1}$$

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.

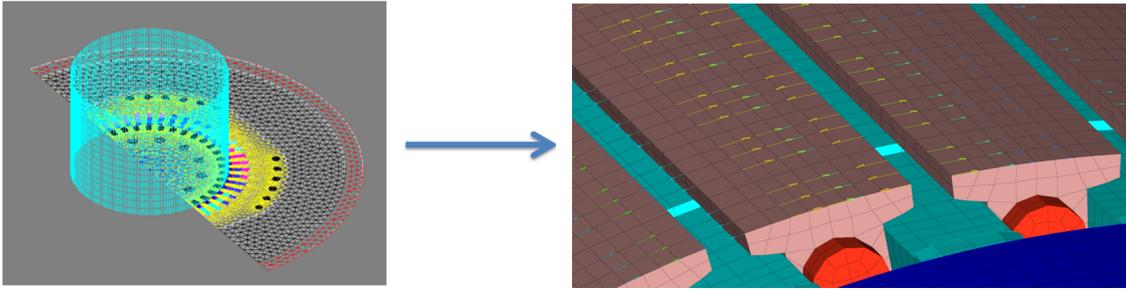


Figure 3: Maxwell pressure projection

$$\overrightarrow{rot} \left(v_0 [v_r] \overrightarrow{rot} (\vec{A}) - \vec{H}_c \right) + [\sigma] \left(\frac{\partial \vec{A}}{\partial t} + \overrightarrow{grad} (V) \right) = 0 \quad (2)$$

where $[v_r]$ is the magnetic reluctivity, v_0 is the vacuum reluctivity ($v_0 = 1/\mu_0 = 1/(4\pi \cdot 10^{-7})$ m/H), \vec{A} is the magnetic potential vector (Wb/m), \vec{H}_c is the coercitive 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).

$$\begin{aligned} \sigma_r &= \frac{1}{\mu_0} (B_r^2 - B_t^2) \\ \sigma_t &= \frac{1}{\mu_0} B_r B_t \end{aligned} \quad (3)$$

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.

$$\begin{aligned} -\omega^2 \rho \hat{u} &= \text{div}(\hat{\sigma}) \\ \hat{\sigma} &= H \hat{\varepsilon} \\ \hat{\varepsilon} &= \frac{1}{2} (\text{grad}(\hat{u}) + \text{grad}^T(\hat{u})) \end{aligned} \quad (4)$$

Flux 2D generates loadings for the mechanical model:

$$\hat{\sigma} \cdot n = \begin{cases} \sigma_r \\ \sigma_{t_1} \\ \sigma_{t_2} \end{cases} \text{ on } \Gamma_l \quad (5)$$

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 :

$$\theta = Cu \quad (6)$$

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 electro-technic 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 electro-technical 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).

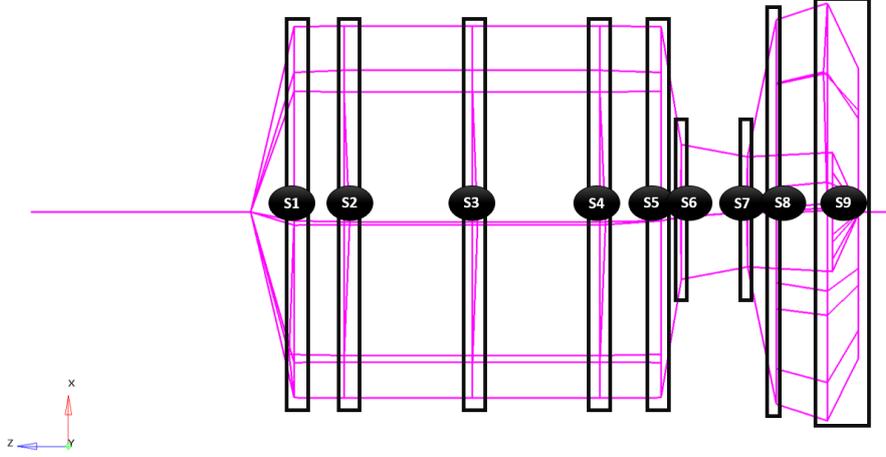


Figure 4: Rotor divided in 9 sections

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.

$$\begin{aligned} u &= \Phi q \\ u &= A\theta \\ \Phi_\theta &= \begin{cases} A^{-1}\Phi \\ C\Phi \end{cases} \end{aligned} \quad (7)$$

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:

$$q = (A\Phi_\theta)^{-1}A\theta = P\theta \quad (8)$$

The classic dynamic equation in translation is given by the following equation.

$$[M]\ddot{u} + [D]\dot{u} + [K]u = F \quad (9)$$

Writted on the modal basis, the dynamic equation in generalized coordinates is :

$$\ddot{q} + [\Delta]\dot{q} + [\Lambda]q = f \quad (10)$$

with $[\Delta] = 2\xi\omega_0[I]$.

By using the transformation (equation 8), the equation in the rotation coordinate frame becomes :

$$[J]\ddot{\theta} + [D_\theta]\dot{\theta} + [K_\theta]\theta = T, \quad (11)$$

with

$$\begin{cases} [J] = P^T P, \\ [K_\theta] = P^T [\Lambda] P, \\ [D_\theta] = P^T (2\xi\omega_0[I]) P. \end{cases} \quad (12)$$

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.

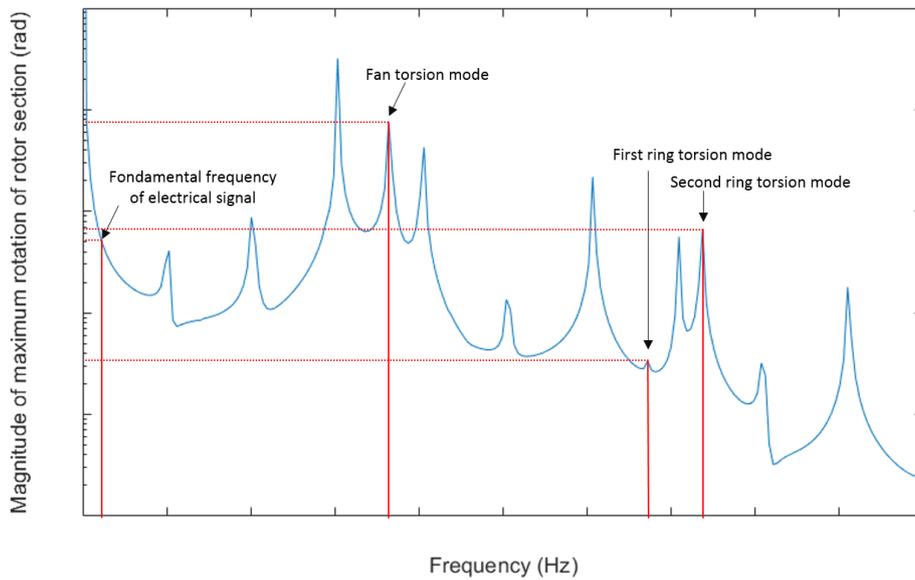


Figure 5: Frequency Response Function

	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 fundamental 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, Tassarolo, 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 IMMOVILLI¹, Marco LIPPI², Marco COCCONCELLI³

DISMI, University of Modena and Reggio Emilia, 42123 Reggio Emilia, Italy
{¹fabio.immovilli,²marco.lippi,³marco.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 detect faults at the early stage, to allow for scheduled maintenance, minimizing system downtime. Under these 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 result in a torque ripple that generates in the current a chain of components at frequencies F_{be} :

$$F_{be} = | f \pm k f_{car} | \quad (1)$$

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.

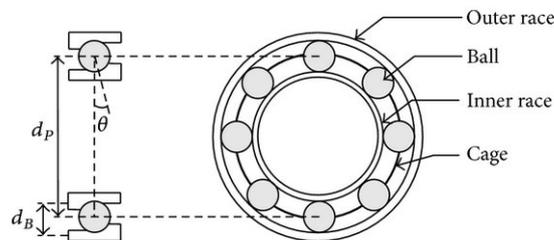


Figure 1: Bearing structure and characteristic dimensions.

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:

$$F_{cage} = \frac{1}{2} F_r \left(1 - \frac{D_b \cos \theta}{D_p} \right) \quad (2)$$

$$F_{outer} = \frac{Z}{2} F_r \left(1 - \frac{D_b \cos \theta}{D_p} \right) \quad (3)$$

$$F_{inner} = \frac{Z}{2} F_r \left(1 + \frac{D_b \cos \theta}{D_p} \right) \quad (4)$$

$$F_{ball} = \frac{D_p}{D_b} F_r \left[1 - \left(\frac{D_b \cos \theta}{D_p} \right)^2 \right] \quad (5)$$

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	[Ω]

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.

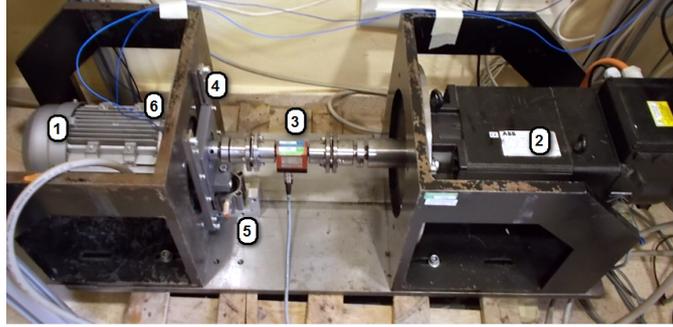


Figure 2: Test setup overview: (1) MUT; (2) brake/dynamometer; (3) torque meter; (4) crosshead; (5) pneumatic cylinder; (6) accelerometer position. The compressed air hose connecting the cylinder to the manifold with pressure regulator and transducer is not shown.

The test bearing is a SKF 6205 deep groove 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: Micrograph of the chemically etched outer race defect (left). Photo of the hydraulic press employed to impart bearing brinelling 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%	Brinelling

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 brinelling. 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 analysis. 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.

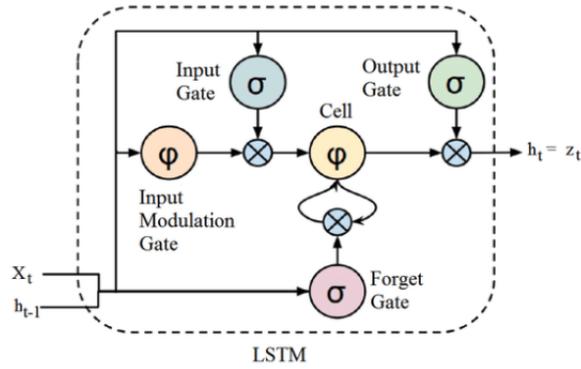


Figure 5: Depiction of LSTM cell. σ (input, output, forget gates) are typically the sigmoid function. Recurrent connections h_t and X_t propagate information through time.

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 recognition 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.

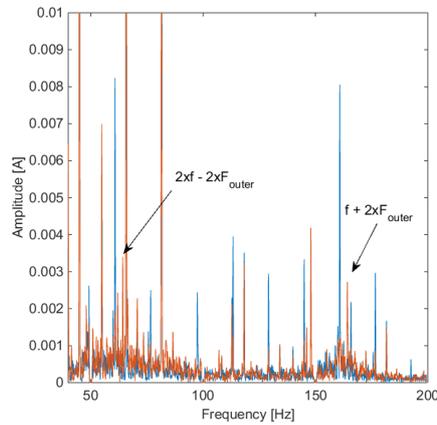


Figure 6: Comparison of FFT spectra of the notch-filtered current signal of the MUT operated at rated output torque and 6 bar pressure on the radial load cylinder: healthy bearing (blue), faulty bearing with localized outer race defect (red).

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.

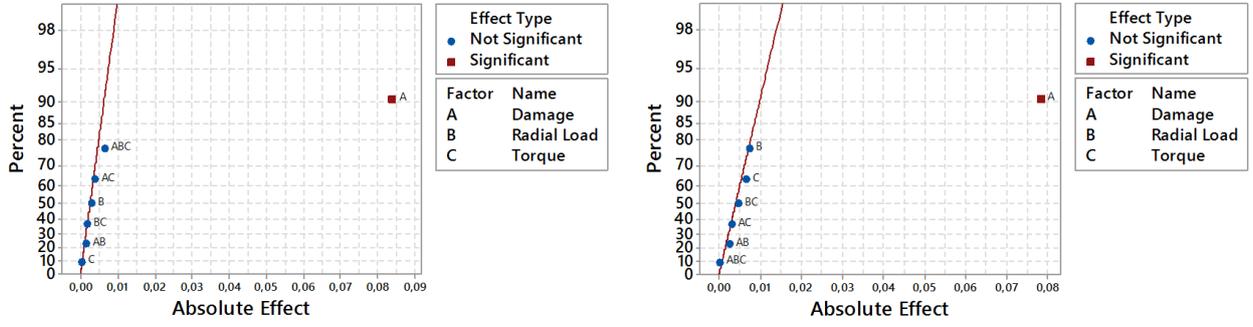


Figure 7: Test set 1 (single point defect on outer race): Half normal plot of the scalar current fault indicator (left). Test set 2 (brinelling defect): Half normal plot of the scalar current fault indicator (right).

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 healthy 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 related to the brinelling 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:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision P and recall R are related to the ratio of false positives and false negatives:

$$P = \frac{TP}{TP + FP} \quad (7)$$

and

$$R = \frac{TP}{TP + FN} \quad (8)$$

Finally F_1 is the harmonic mean between P and R ,

$$F_1 = \frac{2PR}{P + R} \quad (9)$$

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 were 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	F	%	H	F
H	47.68	2.33	H	47.80	2.21	H	42.74	7.26
F	21.79	28.20	F	20.31	29.68	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

$$[M]\{\ddot{q}\} + [K]\{q\} = \{F\}, \quad (1)$$

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:

$$-\omega^2 \begin{bmatrix} M_{BB} & M_{BI} \\ M_{IB} & M_{II} \end{bmatrix} \begin{Bmatrix} q_B \\ q_I \end{Bmatrix} + \begin{bmatrix} K_{BB} & K_{BI} \\ K_{IB} & K_{II} \end{bmatrix} \begin{Bmatrix} q_B \\ q_I \end{Bmatrix} = \begin{Bmatrix} F_B \\ F_I \end{Bmatrix}. \quad (2)$$

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:

$$\begin{Bmatrix} q_B \\ q_I \end{Bmatrix} = \begin{bmatrix} I & 0 \\ X_{st} & X_{el} \end{bmatrix} \begin{Bmatrix} q_B \\ \alpha \end{Bmatrix} \quad (3)$$

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

$$\{\tilde{q}_I\} \approx [X_{st}]\{q_B\} + [\tilde{X}_{el}]\{\tilde{\alpha}\} \quad (4)$$

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 $(\xi_1, \dots, \xi_r) = \xi$ as follows:

$$X(\xi) = \sum_{\alpha \in \mathbb{N}^r} \bar{x}_\alpha \phi_\alpha(\xi), \quad (5)$$

where $\phi_\alpha(\xi)$ are orthogonal polynomials which represent the stochastic components of the process, and \bar{x}_α 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

$$X(\xi) \approx \sum_{\alpha \in \mathcal{A}^{r,p}} \bar{x}_\alpha \phi_\alpha(\xi), \quad (6)$$

where p is the order of the PC expansion and $\alpha = \{\alpha_1, \dots, \alpha_r\} \in \mathbb{N}^r$. The index set used in the truncated expansion (6) is then defined as

$$\mathcal{A}^{r,p} = \{\alpha \in \mathbb{N}^r : \|\alpha\|_1 \leq p\}, \quad (7)$$

with

$$\|\alpha\|_1 = \sum_{i=1}^r \alpha_i. \quad (8)$$

Computing the QoI X comes down to finding the coefficients \bar{x}_α 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]

$$N_p = \text{card}(\mathcal{A}^{r,p}) = \frac{(p+r)!}{p!r!}. \quad (9)$$

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 $Q = k N_p$ (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 experimental 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]

$$\varepsilon_{reg}^2 = \sum_{q=1}^Q \left[X(\xi^{(q)}) - \sum_{\alpha \in \mathcal{A}^{r,p}} \bar{x}_\alpha \phi_\alpha(\xi^{(q)}) \right]^2, \quad (10)$$

where $\xi^{(q)} = (\xi_1^{(q)}, \dots, \xi_r^{(q)})$ (with $q = 1, \dots, Q$) 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; $X(\xi^{(q)})$ denotes the vector of the corresponding FE model evaluations. The PC coefficients are finally calculated as

$$\bar{x} = \left(\phi^T(\xi^{(q)}) \phi(\xi^{(q)}) \right)^{-1} \phi^T(\xi^{(q)}) X(\xi^{(q)}), \quad (11)$$

with $\phi(\xi^{(q)})$ the matrix defined by

$$\phi(\xi^{(q)}) = \begin{pmatrix} \phi_0(\xi^{(1)}) & \dots & \phi_{N_p-1}(\xi^{(1)}) \\ \vdots & \ddots & \vdots \\ \phi_0(\xi^{(Q)}) & \dots & \phi_{N_p-1}(\xi^{(Q)}) \end{pmatrix}. \quad (12)$$

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

$$\|\alpha\|_{m,w} = \left(\sum_{i=1}^r |w_i \alpha_i|^m \right)^{1/m}, \quad w_i \geq 1. \quad (13)$$

The corresponding anisotropic index set is then chosen as

$$\mathcal{A}_{m,w}^{r,p} = \{\alpha \in \mathbb{N}^r : \|\alpha\|_{m,w} \leq p\}, \quad (14)$$

where w is a set of weights w_i defined by

$$w_i = 1 + \frac{\max_{1 \leq j \leq r} S_j^T - S_i^T}{\sum_{k=1}^r S_k^T}. \quad i = 1, \dots, r. \quad (15)$$

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

$$S_i^T = \frac{1}{D_{PC}} \sum_{\alpha \in I_i^+} \bar{x}_\alpha^2 E[\phi_\alpha^2(\xi)], \quad (16)$$

where I_i^+ denotes the set of indices having a non-zero i^{th} component

$$I_i^+ = \{\alpha \in \mathcal{A}_{m,w}^{r,p} : \alpha_i \neq 0\}, \quad (17)$$

and D_{PC} the variance of the QoI

$$D_{PC} = \sum_{\alpha \in \mathcal{A}_{m,w}^{r,p}} \bar{x}_\alpha^2 E[\phi_\alpha^2(\xi)]. \quad (18)$$

The anisotropic hyperbolic polynomial chaos expansions are finally defined with the index sets $\mathcal{A}_{m,w}^{r,p}$ as

$$X_{\mathcal{A}_{m,w}^{r,p}}(\xi) = \sum_{\alpha \in \mathcal{A}_{m,w}^{r,p}} \bar{x}_\alpha \phi_\alpha(\xi). \quad (19)$$

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:

$$Err = E[(X(\xi) - \widehat{X}_{\mathcal{A}}(\xi))^2], \quad (20)$$

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

$$\widehat{X}_{\mathcal{A}}(\xi) = \sum_{\alpha \in \mathcal{A}} \bar{x}_\alpha \phi_\alpha(\xi). \quad (21)$$

The generalization error is estimated in practice by the following empirical error:

$$Err_{emp} = \frac{1}{Q} \sum_{q=1}^Q \left[\left(X(\xi^{(q)}) - \widehat{X}_{\mathcal{A}}(\xi^{(q)}) \right)^2 \right] \quad (22)$$

in which the differences are computed specifically at the Q observations of a NED $\xi^{(q)} = (\xi_1^{(q)}, \dots, \xi_r^{(q)})$. The latter will be used in the following to compute a coefficient of determination R^2 defined as

$$R^2 = 1 - \frac{Err_{emp}}{\widehat{V}[X]}, \quad (23)$$

where $\widehat{V}[X]$ is the variance of $X(\xi^{(q)})$:

$$\widehat{V}[X] = \frac{1}{Q-1} \sum_{q=1}^Q (X(\xi^{(q)}) - \bar{X})^2 \quad \text{with} \quad \bar{X} = \frac{1}{Q} \sum_{q=1}^Q X(\xi^{(q)}).$$

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 $(\xi^{(1)}, \dots, \xi^{(Q)}) \setminus \xi^{(i)}$ (that is the original NED from which the observation $\xi^{(i)}$ has been discarded) [7]:

$$\Delta^{(i)} = X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}^{(-i)}(\xi^{(i)}). \quad (24)$$

The leave-one-out error is then defined as

$$Err_{LOO} = \frac{1}{Q} \sum_{i=1}^Q (\Delta^{(i)})^2. \quad (25)$$

In practice, the predicted residual $\Delta^{(i)}$ may be computed as [17]

$$\Delta^{(i)} = \frac{X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}(\xi^{(i)})}{1 - h_i}, \quad (26)$$

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

$$Err_{LOO} = \frac{1}{Q} \sum_{i=1}^Q \left(\frac{X(\xi^{(i)}) - \widehat{X}_{\mathcal{A}}(\xi^{(i)})}{1 - h_i} \right)^2. \quad (27)$$

A determination coefficient S^2 equivalent to that of the empirical error, R^2 , may be defined for the leave-one-out error:

$$S^2 = 1 - \frac{Err_{LOO}}{\widehat{V}[X]}. \quad (28)$$

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_{target}^2 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 $w = \{1, \dots, 1\}$. The corresponding initial values of the determination coefficients are denoted as R_0^2 and S_0^2 .

Step 3: Training step - Enrichment of the PC basis

Increment the order value: $p \rightarrow p + 1 \in [1, \dots, p_{max}]$.

⇒ **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.

⇒ **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 $2 \cdot \text{card}(\mathcal{A}_{m,w}^p)$ go to step 5. If the conditioning is poor, i.e. the size Q_k of the NED ($\xi^{(q)}$) is smaller than $2 \cdot \text{card}(\mathcal{A}_{m,w}^p)$, 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.

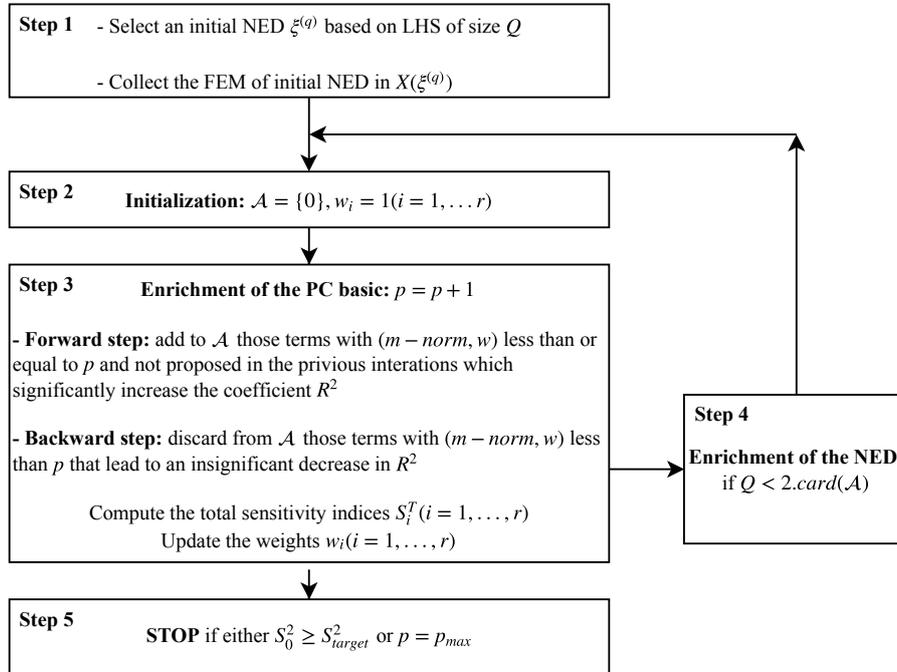


Figure 1: Algorithm applied to build a sparse polynomial chaos expansion with anisotropic hyperbolic index sets

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 $1 m \times 1 m$ 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 $0.2 m \times 1 m$, 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 $[100, 200]\text{ N/m}$. 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 $F = 40\text{ N}$ in the z -direction, located at the node of coordinates $(0.25\text{ m}, 0.25\text{ m})$ if the origin is chosen as the lower left corner of plate 1.

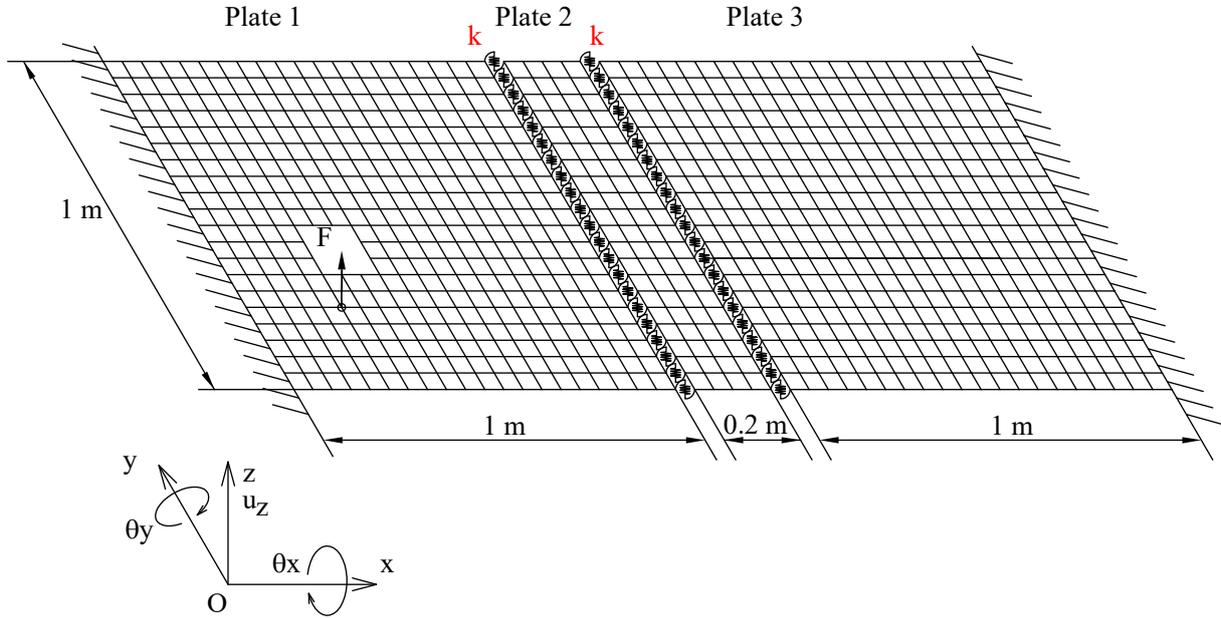


Figure 2: Model of three plates connected by springs

The frequency response function of the structure is studied within the frequency band $[0, 50]\text{ Hz}$ 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 .

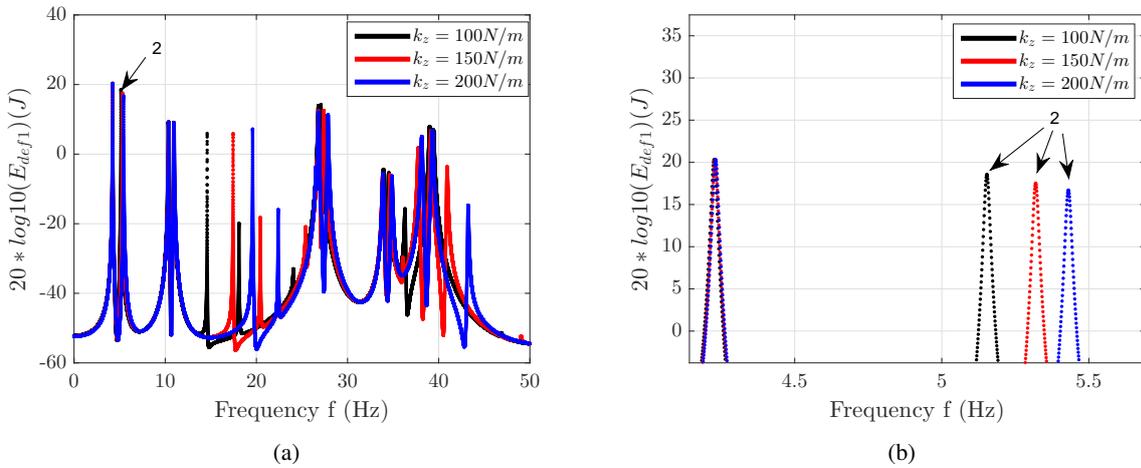


Figure 3: Deformation energy E_{def1} of plate 1: (a) whole frequency range (b) zoom on the first two peaks

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 $S_{target}^2 = 0.999$, and we use two identical thresholds $\epsilon_1 = \epsilon_2 = 0.001(1 - S_{target}^2)$.

As explained previously, the optimal order p of the sPC expansion of the QoI depends on the leave-one-out error S_0^2 that is computed at each step of the iterative procedure and compared to the target accuracy S_{target}^2 . For the first QoI (peak 2 resonance frequency), the values of S_0^2 obtained for the three sPC expansions (sPC-FEM, sPC-CB50130 and sPC-3080) are $S_0^2 = 0.9820$ for $p = 1$ and $S_0^2 = 0.9997$ 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)

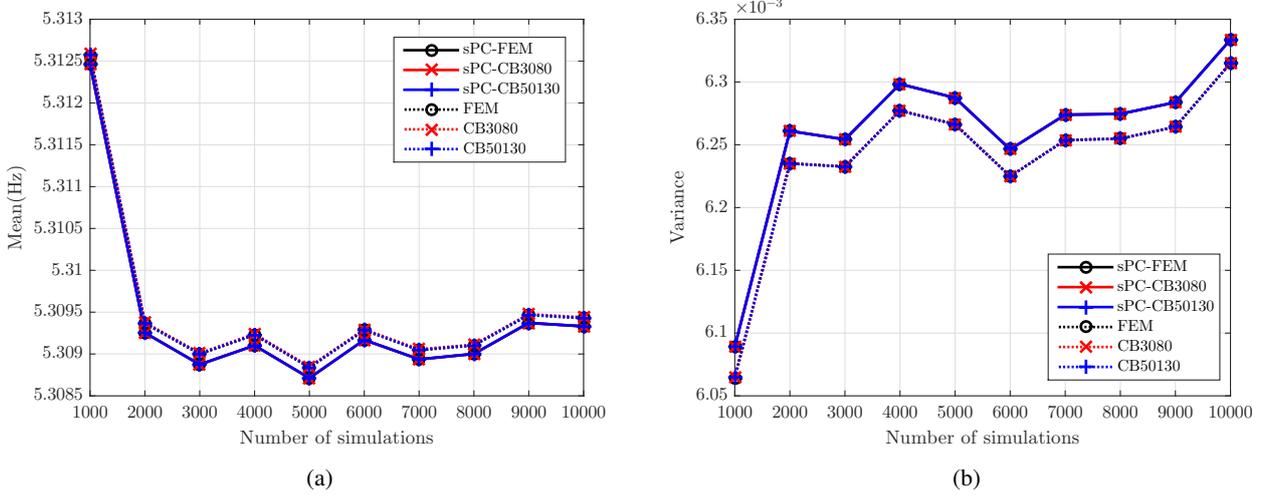


Figure 4: (a) Mean and (b) variance of the QoI according to the number of MC simulations

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 $\bar{f}_2 \pm 1.2 \cdot 10^{-4}$ 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 $2.5 \cdot 10^{-3} \%$ 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.

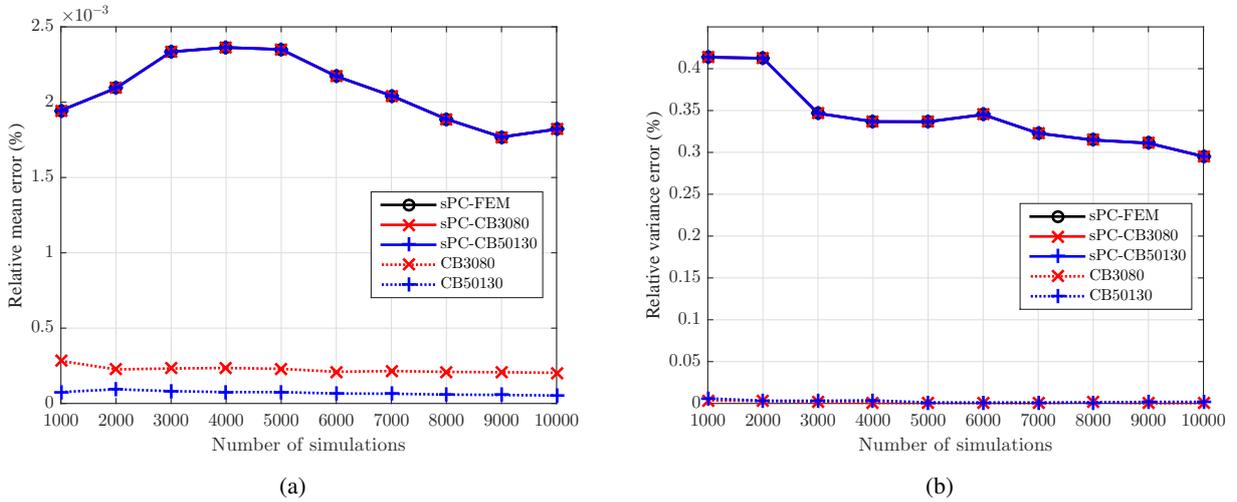


Figure 5: Relative errors of (a) the mean value of the QoI and (b) the variance of the QoI with respect to the reference solution according to the number of MC simulations

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 Nk_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.

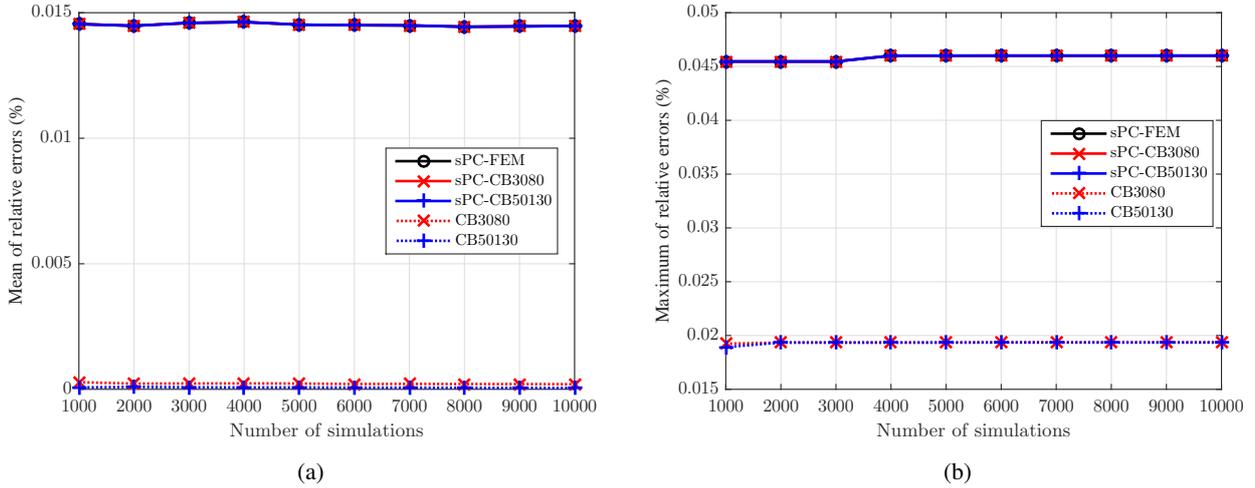


Figure 6: (a) Mean and (b) maximum values of the relative errors of the QoI with respect to the reference solution according to the number of simulations

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.

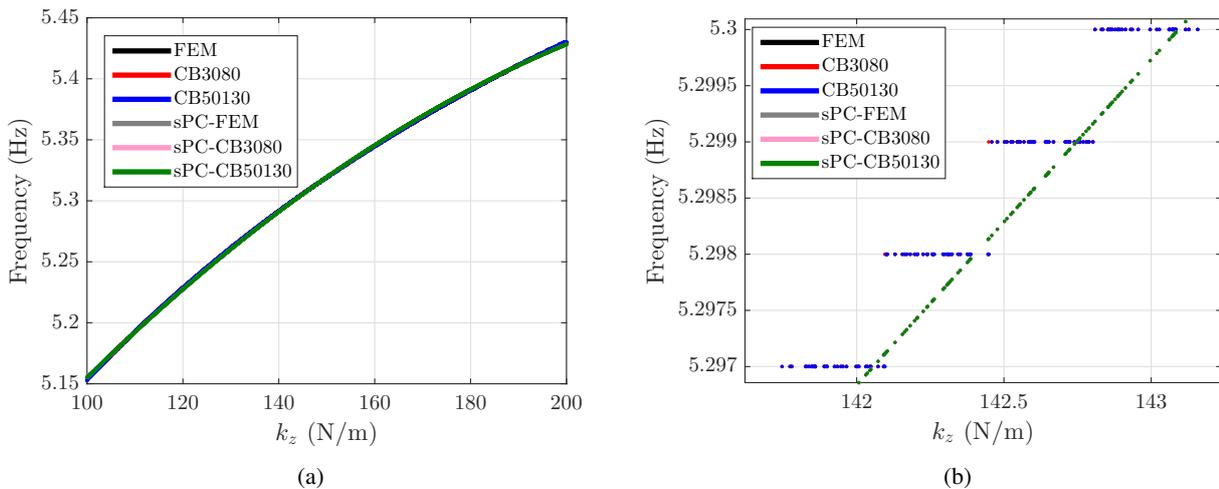


Figure 7: Peak 2 resonance frequency according to the spring stiffness k_z for $N = 10000$

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 used 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.

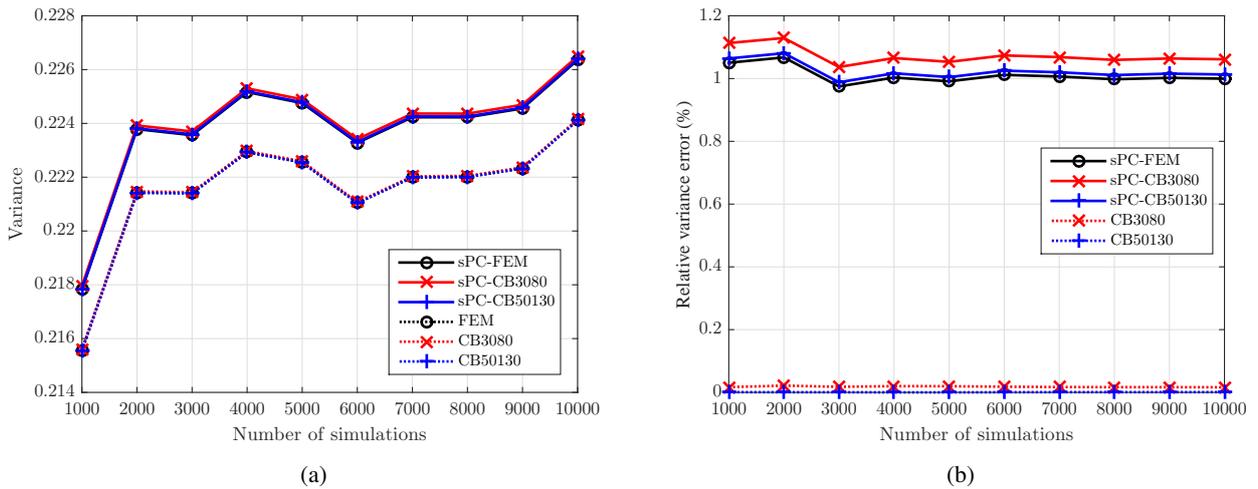


Figure 8: (a) Variance and (b) relative error of the variance of the second QoI according to the number of MC simulations

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.

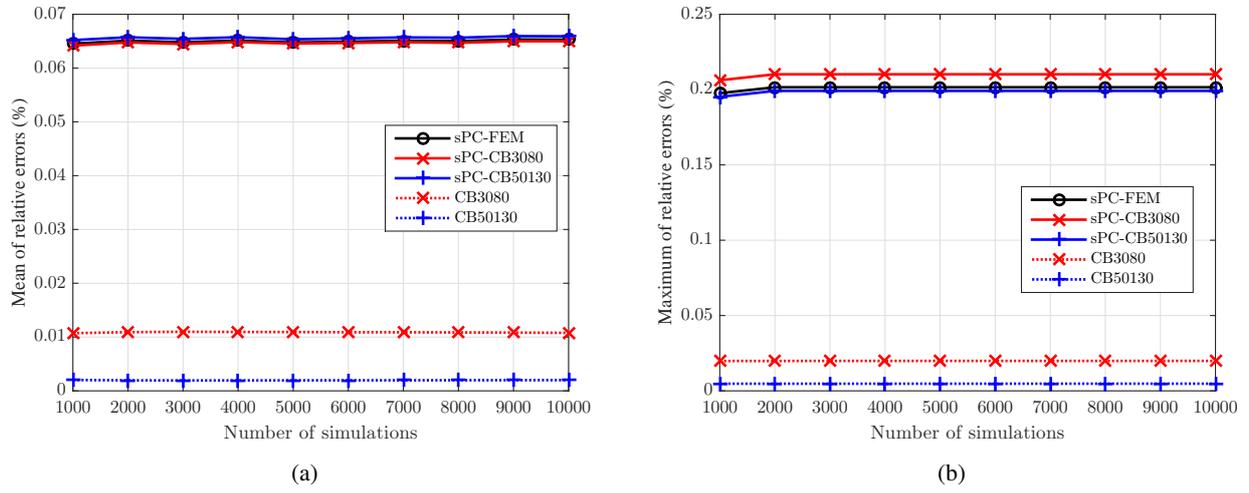


Figure 9: (a) Mean and (b) maximum values of the relative errors of the QoI with respect to the reference solution according to the number of simulations

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.

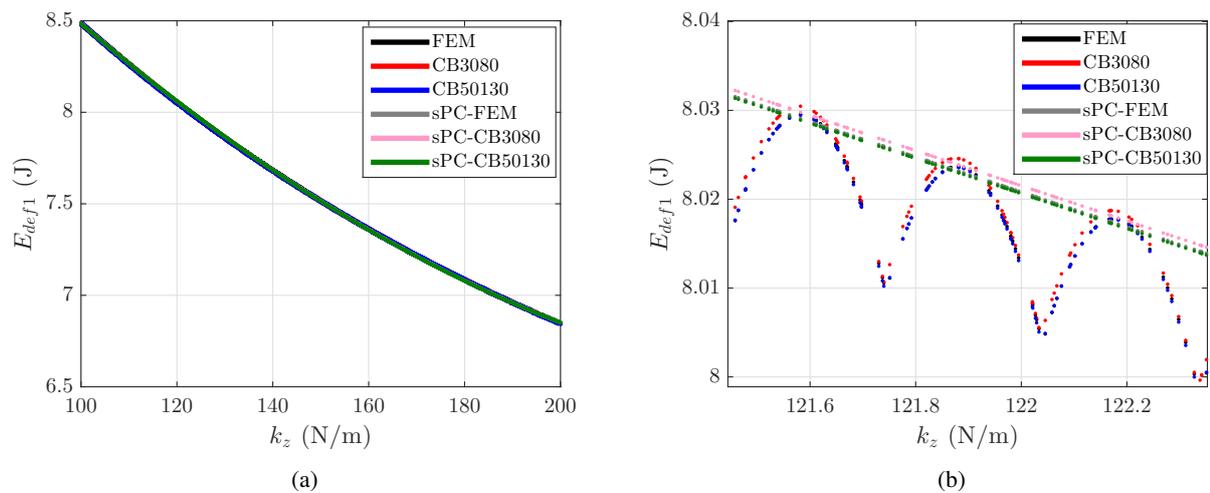


Figure 10: Peak 2 amplitude according to the spring stiffness k_z for $N = 10000$

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 parameters 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 Reliability*, 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 GAGNON¹, François LEONARD¹, James MERLEAU¹, Dominique TAPSOBA¹

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 \mathbf{R} repeated many times ($\mathbf{R}_i, i = 1, \dots, M$), a shell will be formed at a given distance $\mu_{\perp S}$ from the expected value \mathbf{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.

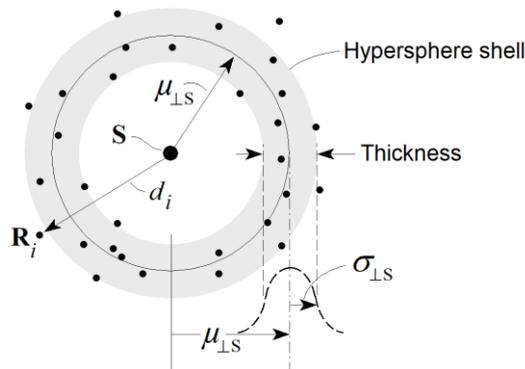


Figure 1: Multidimensional shell from noisy data

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 \mathbf{S} that we transmit over N dimensions contaminated by noise, the received signal is given by:

$$\mathbf{R} = [S_1 + \varepsilon_1, S_2 + \varepsilon_2, \dots, S_N + \varepsilon_N] = \mathbf{S} + \boldsymbol{\varepsilon} \quad (1)$$

where \mathbf{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:

$$\hat{\mathbf{S}} = \frac{1}{M} \sum_{i=1}^M \mathbf{R}_i \quad (2)$$

Furthermore, since the signature transmitted is constantly changing with the operating conditions of the equipment, the signature \mathbf{S} is a manifold rather than a single location as shown in Figure 2. This point is discussed further in the following sections.

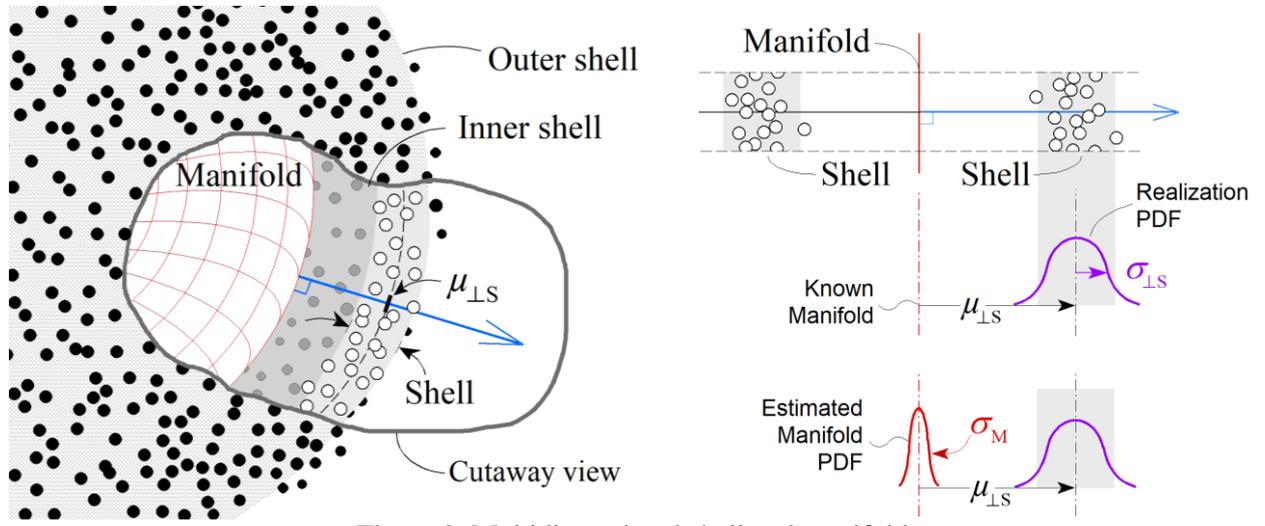


Figure 2: Multidimensional shell and manifold

In this multidimensional space, the distance $\mu_{\perp S}$ between the received signal \mathbf{R} and the transmitted signature \mathbf{S} at any given location on the manifold can be defined as the average of the realizations at this location as follows:

$$\hat{\mu}_{\perp S} = \frac{1}{M} \sum_{i=1}^M \|\mathbf{R}_i - \hat{\mathbf{S}}\| = \frac{1}{M} \sum_{i=1}^M d_i \quad (3)$$

where $\hat{\mathbf{S}}$ is the estimate of the transmitted signature of interest \mathbf{S} , a location on the manifold. In a similar manner, the dispersion can be obtained with:

$$\hat{\sigma}_{\perp S}^2 = \frac{1}{M} \sum_{i=1}^M (d_i - \hat{\mu}_{\perp S})^2 \quad (4)$$

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_M^2 and shell dispersion $\sigma_{\perp S}^2$ can be used to assess the likeliness of a given data point \mathbf{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.

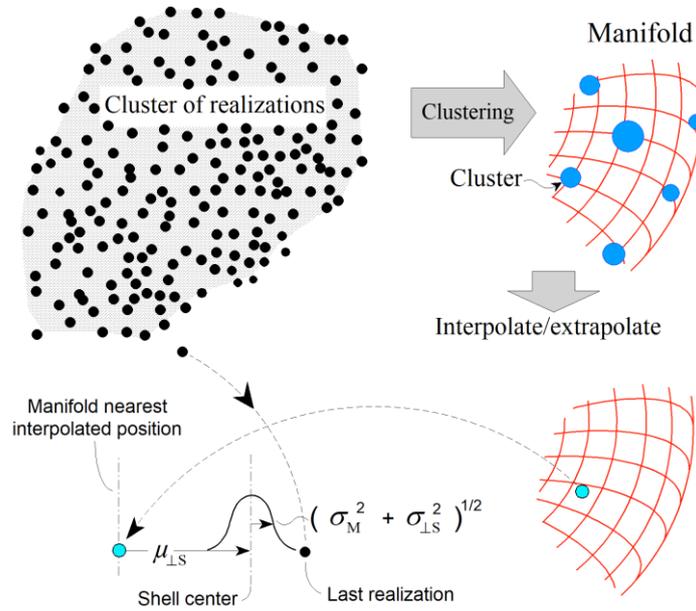


Figure 3: Monitoring methodology

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 $\mathbf{X}_i \equiv \mathbf{R}_i \cup \mathbf{O}_i$ where \mathbf{R}_i is a set of response indicators and \mathbf{O}_i is a set of operating condition indicators. These \mathbf{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 \mathbf{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 \mathbf{X}_i vectors, our goal is to generate clusters of similar \mathbf{O}_i to estimate the multidimensional manifold \mathcal{S} . Initially, for the creation of the clusters, it is important to have a reference dataset of validated history of \mathbf{X}_i that cover most operating conditions \mathbf{O}_i with corresponding responses \mathbf{R}_i . Afterwards, with each new \mathbf{X}_i , it is the dynamic clustering methodology that will determine if a new \mathbf{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 \mathbf{X}_i enters the data history. Figure 4 shows the typical process every new \mathbf{X}_i is subjected to.

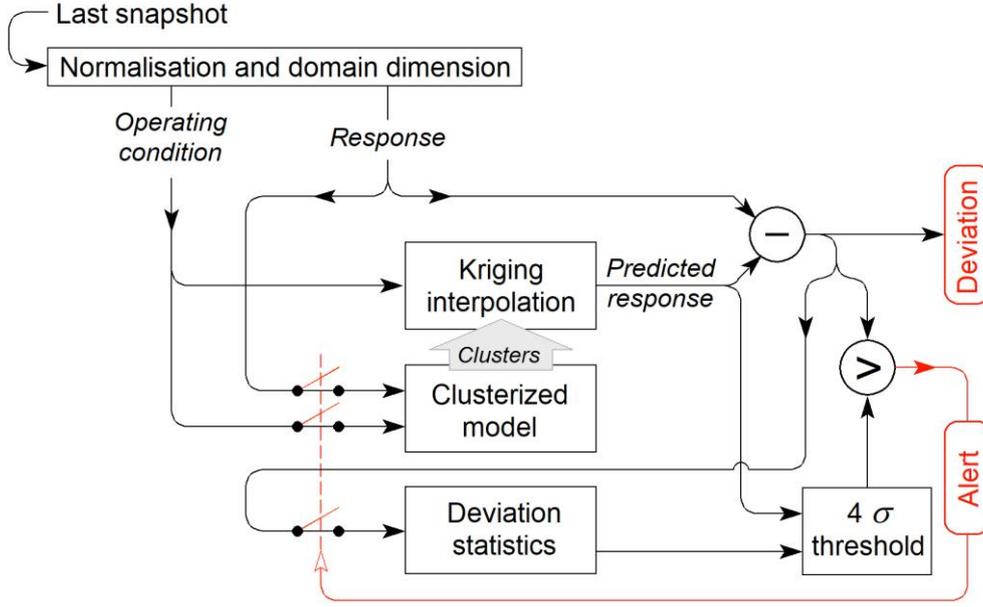


Figure 4: Data snapshots processing flowchart

The clustering process for a new snapshot \mathbf{X}_i looks like this:

- If the number of clusters $k < k_{max}$ then the centroid location is $\mathbf{Centroid}_k = \mathbf{O}_i$ and the population $p_k = 1$ which processes the new snapshot \mathbf{X}_i
- If the number of cluster $k \geq k_{max}$ then find the smallest distance to an existing cluster c_{min}

$$c_{min} = \min_k \|\mathbf{Centroid}_k - \mathbf{O}_i\| \quad (5)$$

- If $c_{min} > c_{max}$ then merge the two closest clusters together, $\mathbf{Centroid}_k = \mathbf{O}_i$, $p_k = 1$ and process the new snapshot \mathbf{X}_i .

$$c_{max} = c \frac{1}{K} \sum_{k=1}^K \|\mathbf{Centroid}_k - \mathbf{O}_i\| \quad (6)$$

- If $c_{min} \leq c_{max}$ then merge snapshot \mathbf{X}_i to the nearest cluster, update the population p_k and location $\mathbf{Centroid}_k$

Here, in equation 6, c is an arbitrary decision level usually set between 1 and 3. Furthermore, note that any abnormal \mathbf{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 \mathbf{O}_k with estimated expected response vectors $\mathbf{E}(\mathbf{O}_k)$ and standard deviation vectors $\mathbf{C}(\mathbf{O}_k)$ enables us to use a multidimensional interpolator to estimate the response vector for any new operating condition \mathbf{O}_i . The interpolated $\mathbf{E}(\mathbf{O}_i)$ with corresponding $\mathbf{C}(\mathbf{O}_i)$ 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 $\mathbf{E}(\mathbf{O}_i) = \widehat{\mathbf{S}}_i$ (in the notation of section 2) and standard deviation $\mathbf{C}(\mathbf{O}_i)$ at any new operating condition \mathbf{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 \mathbf{R}_i and the expected response at the operating condition $\mathbf{E}(\mathbf{O}_i)$:

$$d_i = \|\mathbf{R}_i - \mathbf{E}(\mathbf{O}_i)\| = \sqrt{\sum_{n=1}^N (r_{i,n} - e(\mathbf{O}_i)_n)^2} \quad (7)$$

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:

$$\hat{\mu}_{LS} = \bar{d} = \frac{1}{\sum_{j \in A} 1} \sum_{j \in A} d_j \text{ with } j \in A \text{ if } \mathbf{O}_j \approx \mathbf{O}_k \quad (8)$$

$$w_i = d_i - \bar{d} \quad (9)$$

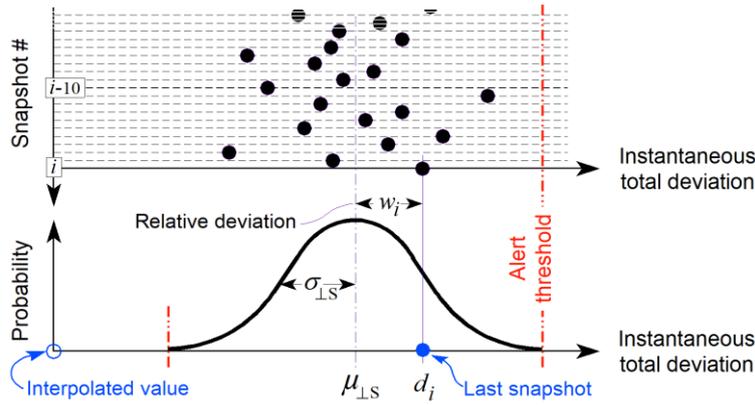


Figure 5: Illustration of the comparison metric

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 $\widehat{\sigma}_M = \|\mathbf{C}(\mathbf{O}_i)\|$ and standard deviation of the relative deviation w_i as follows:

$$\sigma' = \sqrt{\widehat{\sigma}_{LS}^2 + \widehat{\sigma}_M^2} \quad (10)$$

$$\widehat{\sigma}_{LS}^2 = \frac{1}{\sum_{j \in A} 1} \sum_{j \in A} w_j^2 \text{ with } j \in A \text{ if } \mathbf{O}_j \approx \mathbf{O}_k \quad (11)$$

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 ($\in R$)
- Water temperature
- Peak to peak thrust bearing axial acceleration ($\in R$)
- RMS thrust bearing axial acceleration ($\in R$)
- RMS generator guide bearing radial displacement X ($\in R$)
- RMS generator guide bearing radial displacement Y ($\in R$)
- Mean turbine guide bearing radial displacement X ($\in R$)
- RMS turbine guide bearing radial displacement X ($\in R$)
- Peak to peak turbine guide bearing radial displacement Y ($\in R$)
- Mean turbine guide bearing radial displacement Y ($\in R$)
- RMS turbine guide bearing radial displacement Y ($\in R$)
- Excitation tension
- Wicket gates opening ($\in O$)
- Mean power output ($\in R$)



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.

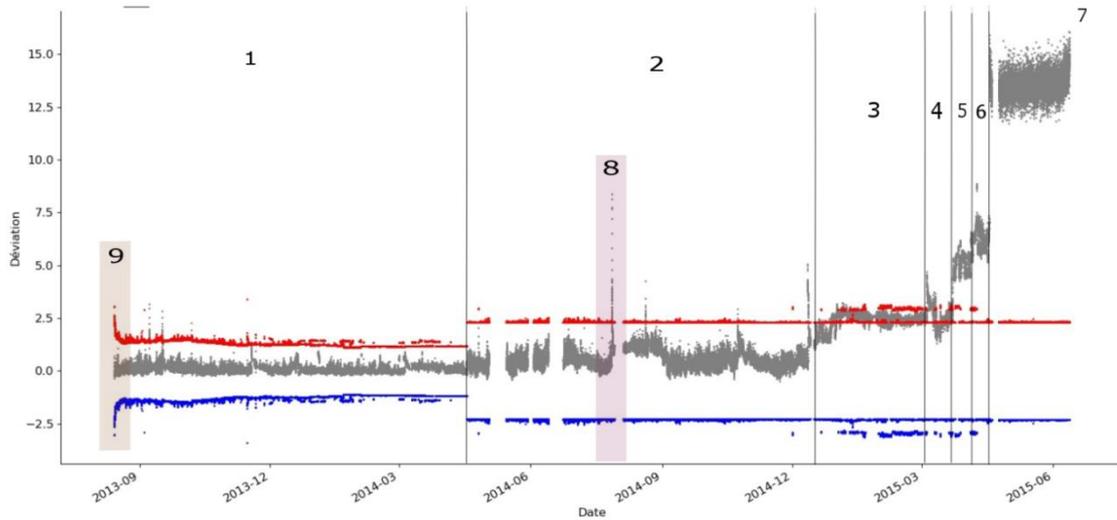


Figure 7: Multidimensional deviation

An advantage of the proposed methodology is that we have access to the contribution of each snapshot response 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.

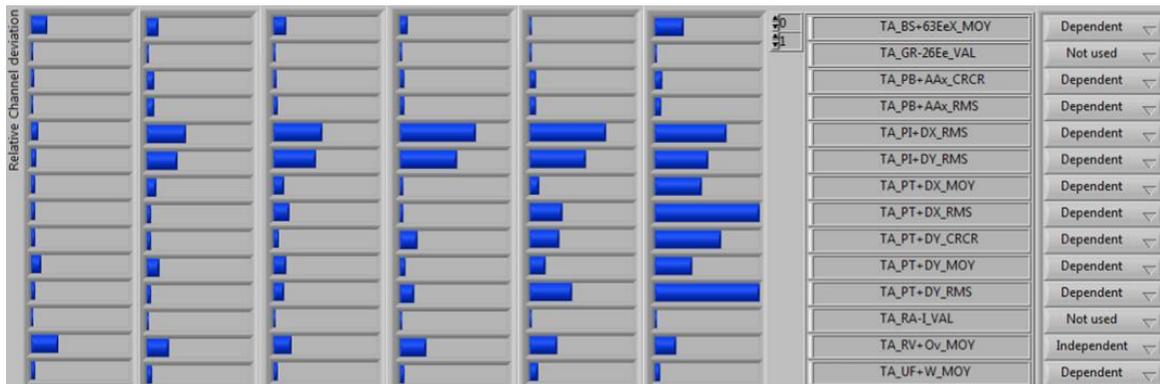


Figure 8: Screenshots for a given timestamp of the relative multidimensional deviation for each channel. From left to right: phase 2, 3, 4, 5, 6 and 7

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.

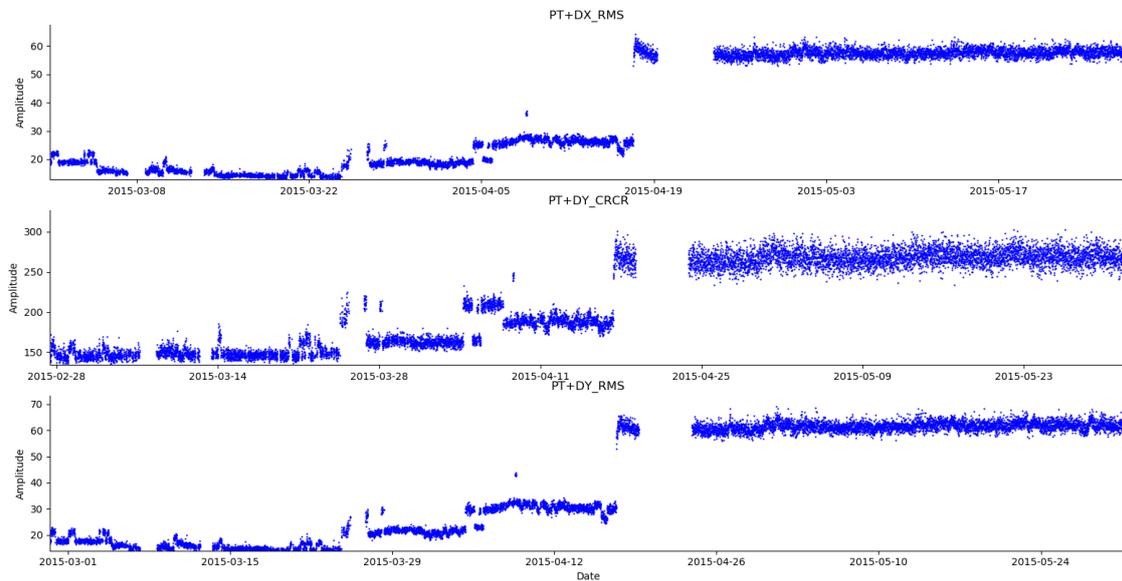


Figure 9: Selected individual responses prior to the propeller runner cone loss

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 Daga¹, Luigi Garibaldi¹, Alessandro Fasana¹, Stefano Marchesiello¹

¹Politecnico 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 $f_s = 102400$ 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

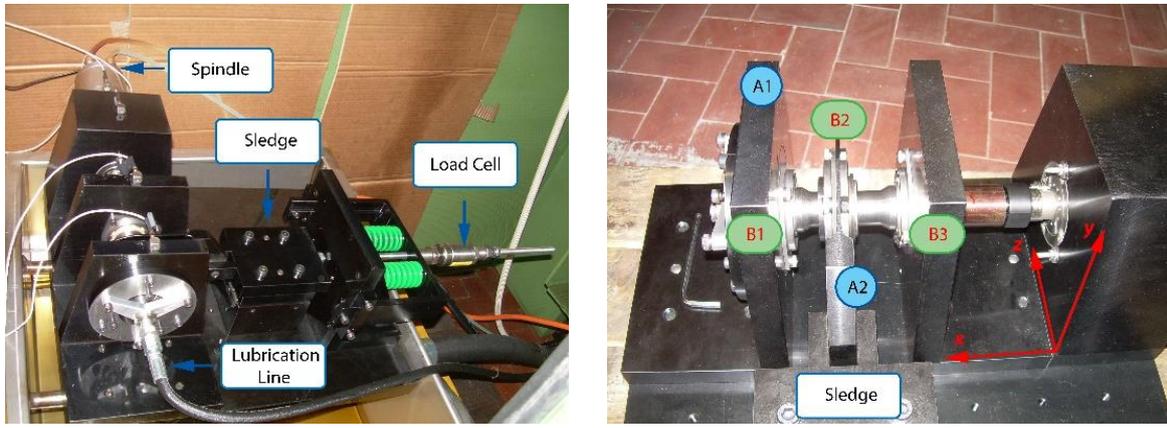


Figure 1. The experimental setup, the triaxial accelerometers location (A1 and A2) and orientation.

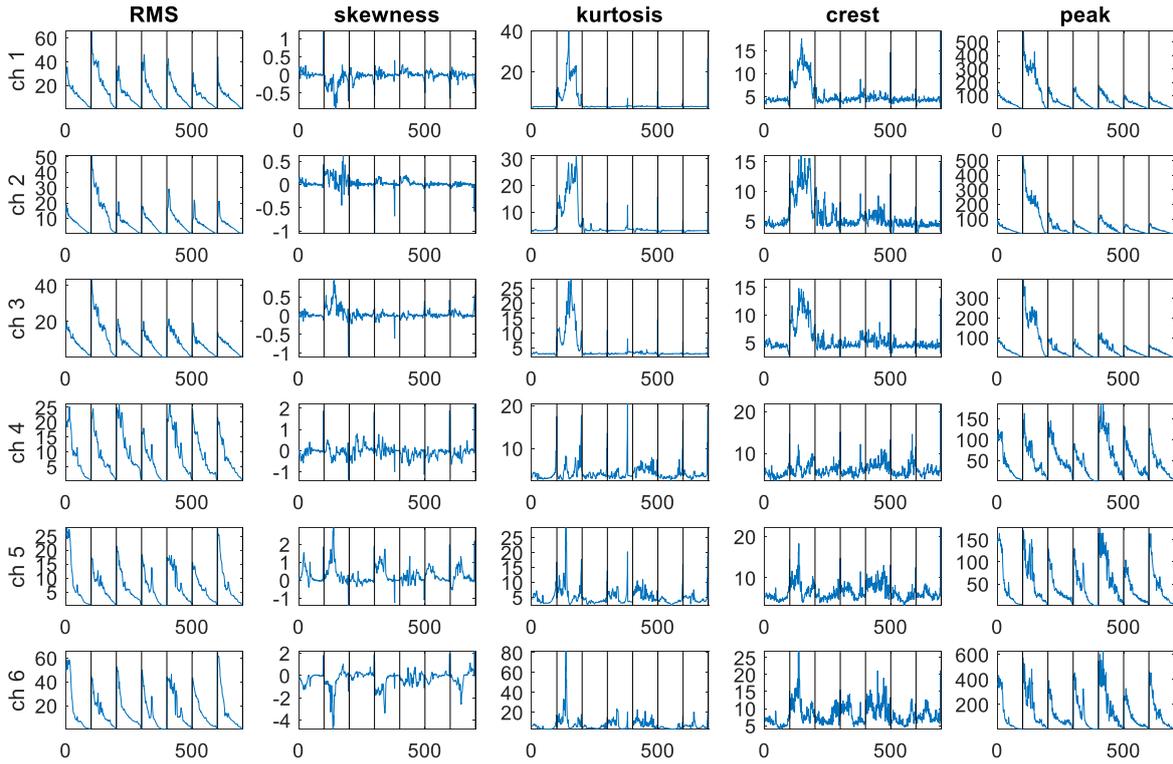


Figure 2. The considered dataset after features extraction for load condition 1 (0 N) while the speed is decreasing until a stop starting from 470 Hz. The black dotted lines divide the different damage conditions (0A to 6A). For each, 100 observations are plotted sequentially.

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).

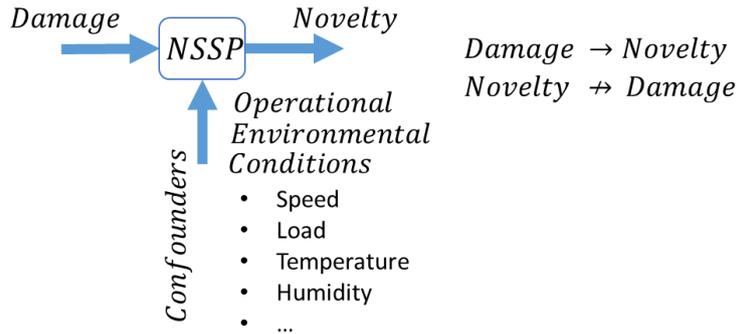


Figure 4. Non-stationary stochastic process and the effect of confounders

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.

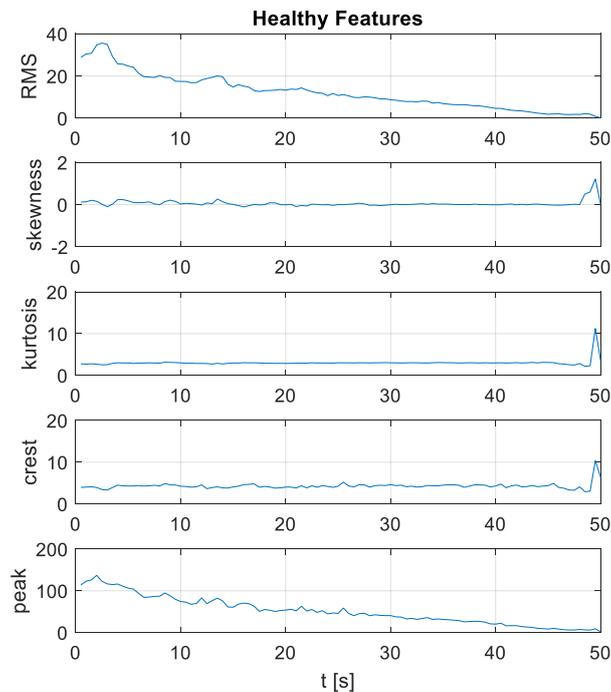


Figure 5. The five selected features from the first channel of the first accelerometer – 0N.

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.

$$y_t = \beta t + \varepsilon_t \quad \varepsilon_t \sim N(0, \sigma^2)$$

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.

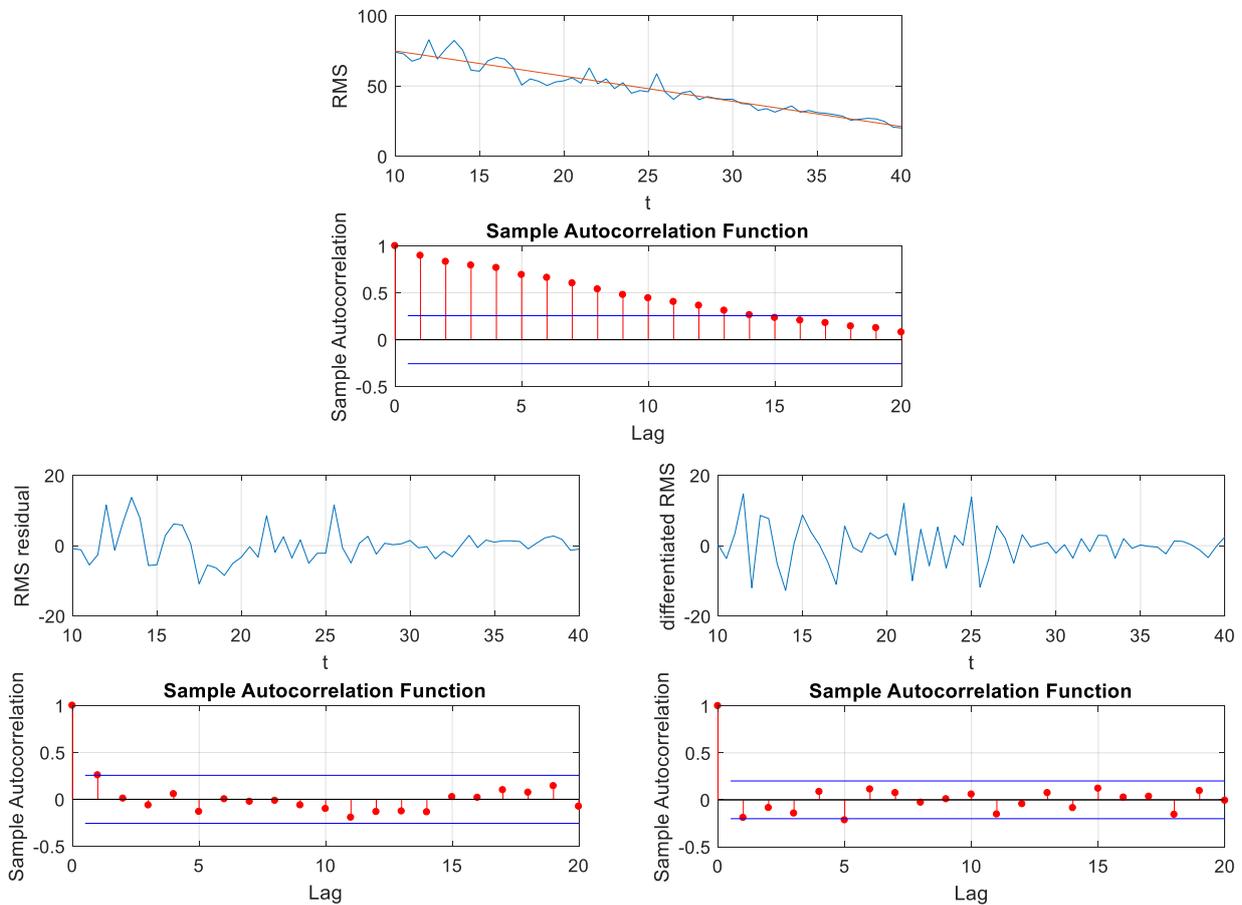


Figure 6. The raw RMS from 10 to 40s and its autocorrelation function (ACF). Below, on the left the residual from removing the linear regression and its ACF, on the right, the differencing (equivalent to the residual after an AR(1) fit) and its ACF.

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.

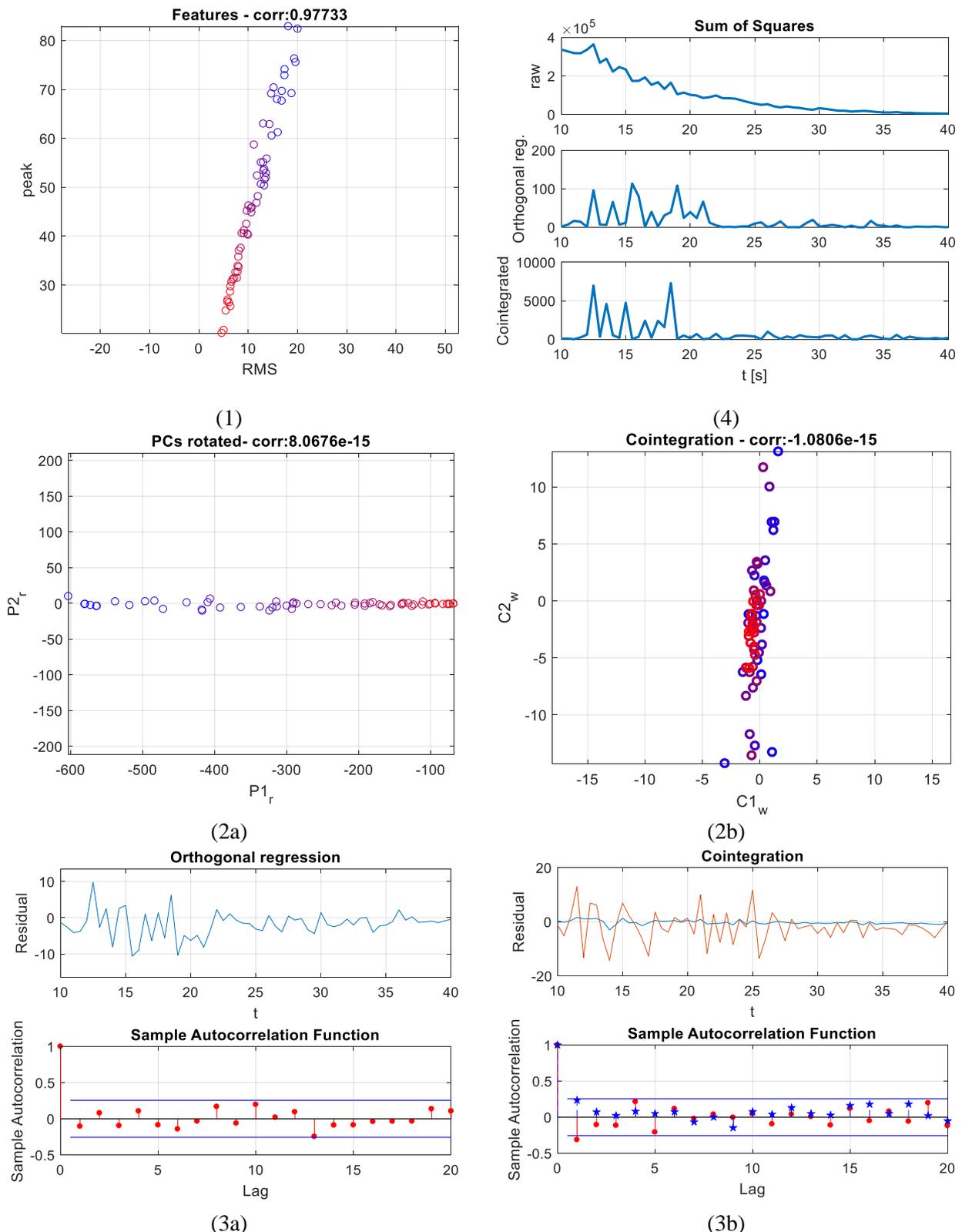


Figure 7. (1) The bivariate scatterplot with time evolution (from blue – 10s to red – 40s)
 (2a) PCA rotation (3a) Rotated PC2 corresponding to the residual and its ACF
 (2b) cointegration scatterplot (3b) stationary residual from cointegration and its ACF
 (4) sum of squares of 3a and 3b components 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).

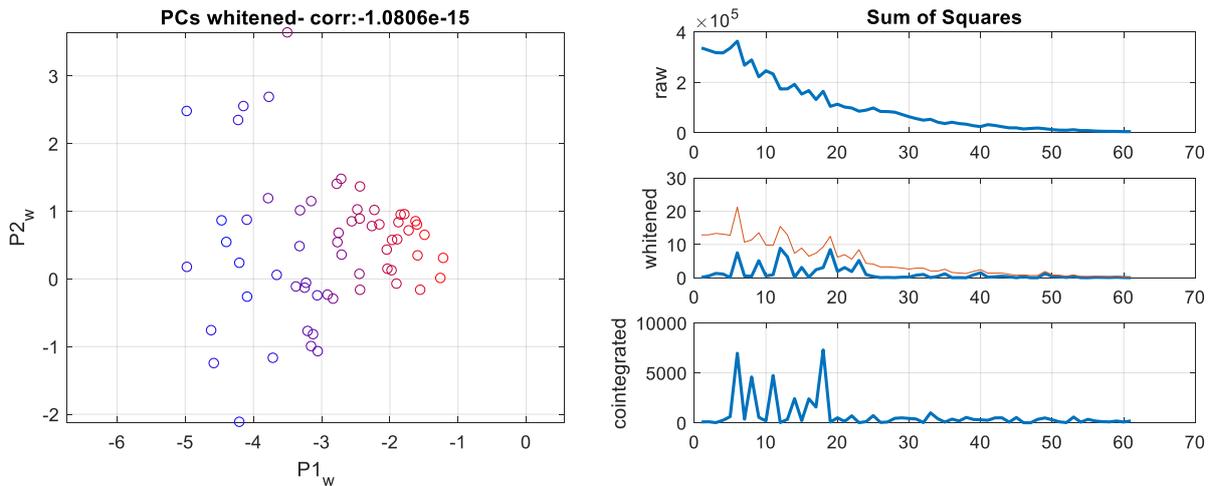


Figure 8. Standardized principal components (whitening) and sum of squares of the two PCs (in red) and of the second alone (blue, mid graph) compared to the squared Eulerian distance (raw) and to the sum of squares of the cointegrated residuals.

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 $\sigma_j^2 = \text{var}(z_j) = \lambda_j$. 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:

$$z_j = v_j'X = v_{j1}x_1 + v_{j2}x_2 + \dots + v_{jd}x_d = \sum_{k=1}^d v_{jk}x_k$$

$$Z_L = V_L'X$$

$$X_L = V_L Z_L = V_L V_L'X$$

The orthogonal regression is visualized in Figure 9.

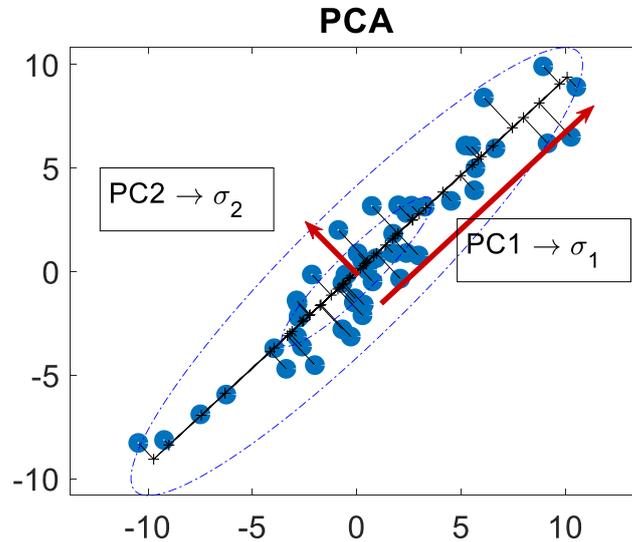


Figure 9. Visualization of the PCA orthogonal regression – the residuals corresponding to PC2 are highlighted.

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 $\text{var}(z_j) = 1$ 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:

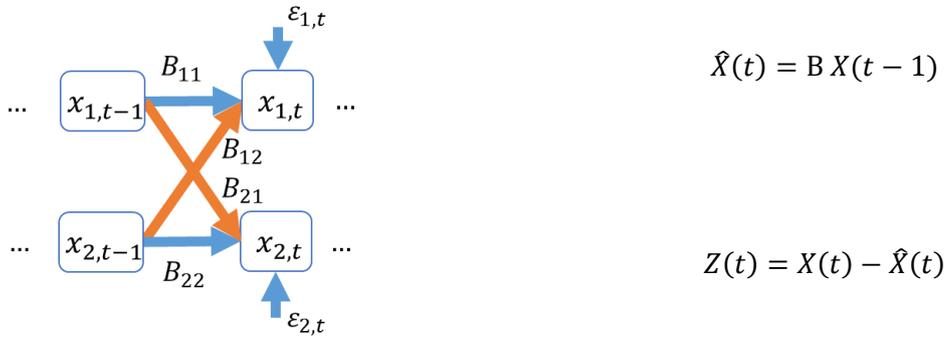


Figure 10. The VAR(1) model and residual.

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 two 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 $d = \frac{\mu_1 - \mu_2}{\sigma}$ (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.

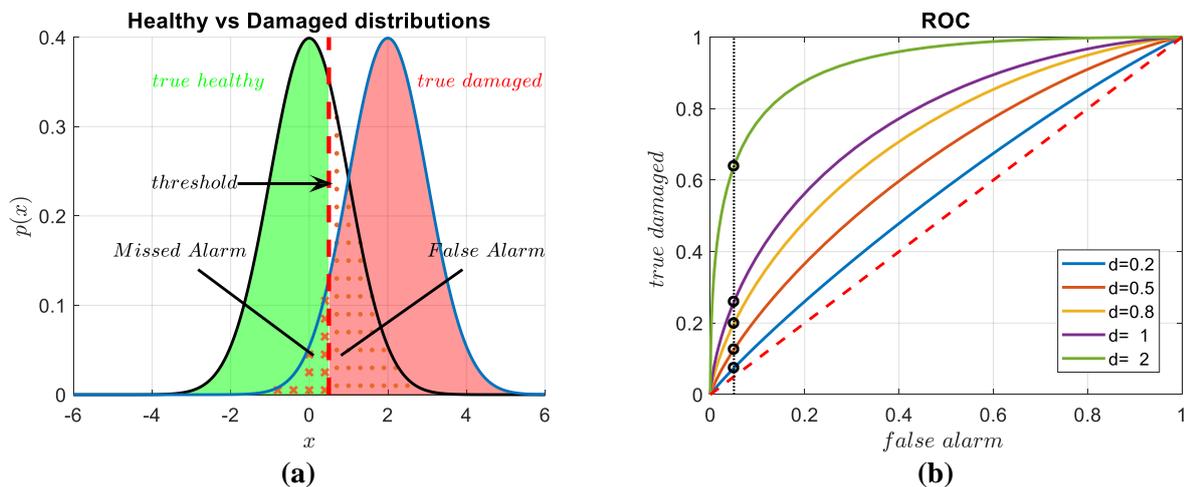


Figure 11. Receiver Operating Characteristic (ROC) as a function of the threshold (Gaussian distributions). **(a)** graphical summary of the table of type I and type II errors in yellow (Table 3). **(b)** ROC for binary classification with different effect sizes d and the position of the 5% false alarm 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.

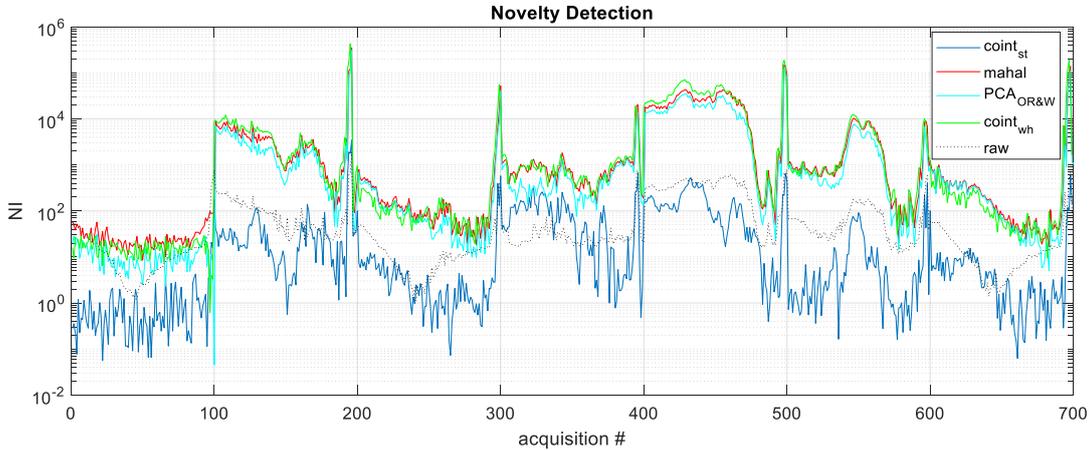


Figure 12. The Novelty Detection with the different NIs. 0-100 samples are the healthy reference, 100-200 corresponds to 1A damage, and so on until 600-700 coming from 6A damage.

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.

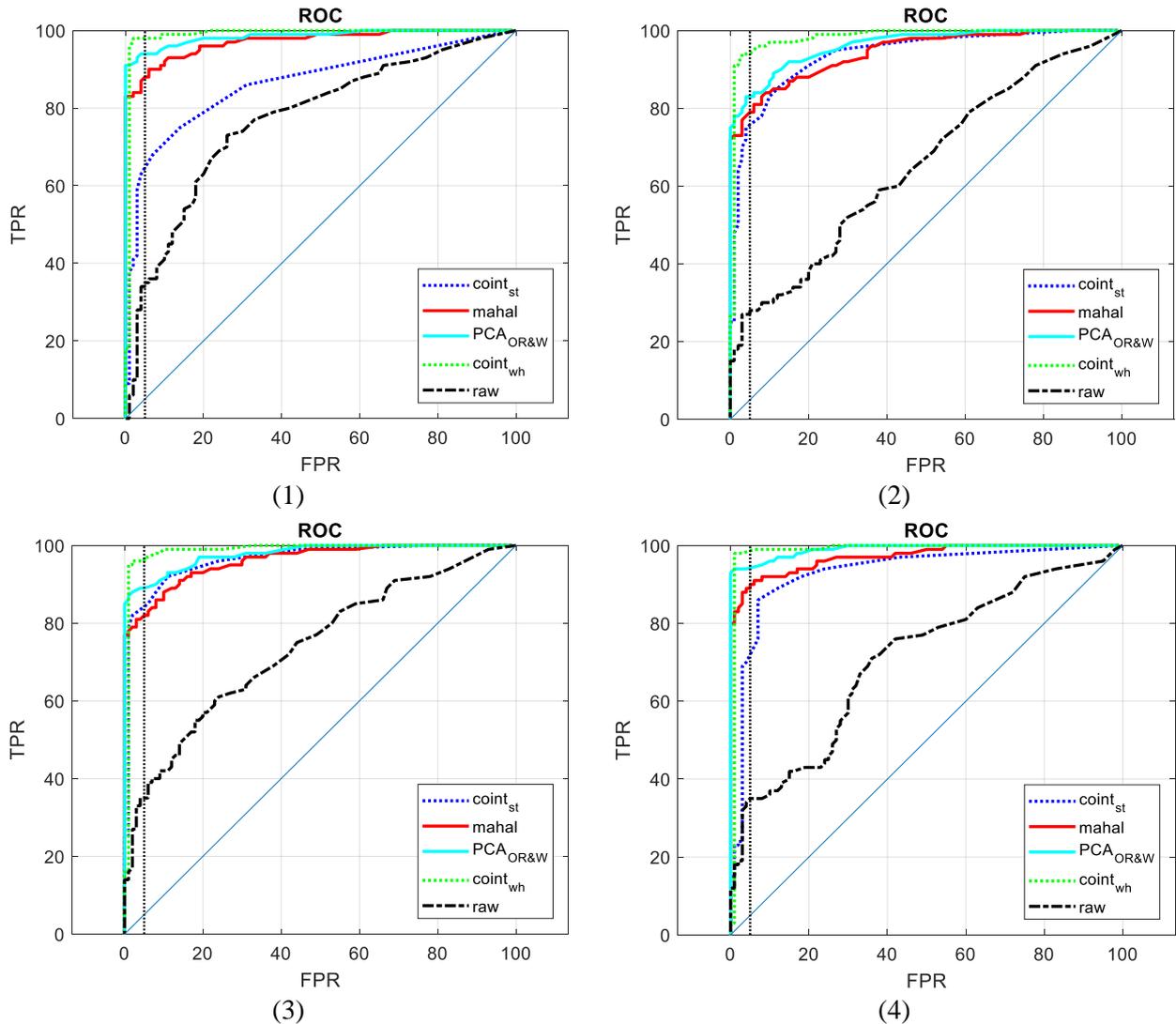


Figure 13. The ROC for Novelty Detection in the 4 loading conditions, respectively 1800, 1400, 1000, 0N

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-Orthogonal 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 by 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. Zhang¹, X. P. Cui¹, M. Salvia^{1*} and O. Bareille¹

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 following: 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

$$A = \frac{U_0 - U_R}{U_R}$$

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.

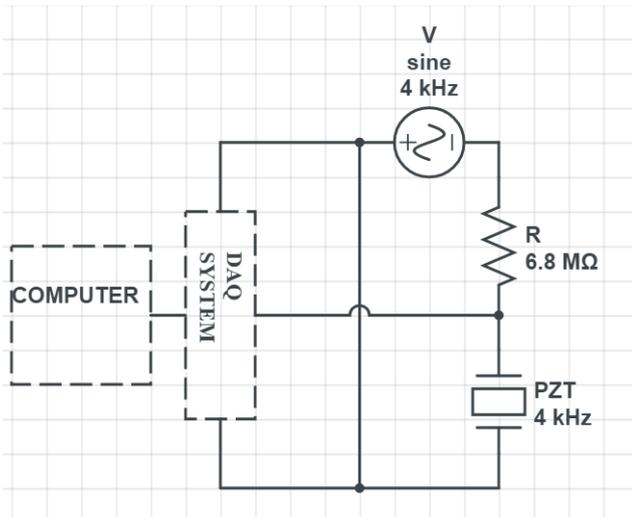


Fig. 1. Circuit diagram of monitoring system

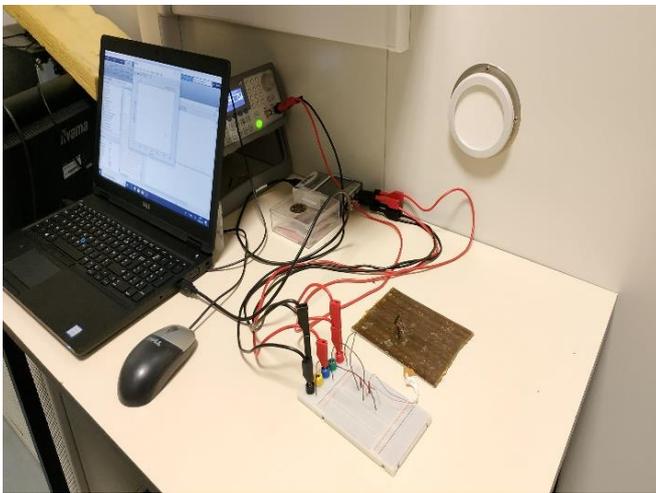


Fig. 2. Monitoring electric circuit and equipment

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/m². 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°C. 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°C



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.

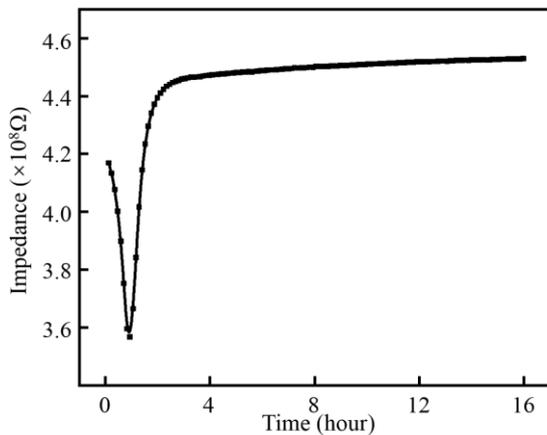


Fig. 5. Impedance of PZT sensor embedded in the specimen of epoxy resin

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 surely 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.

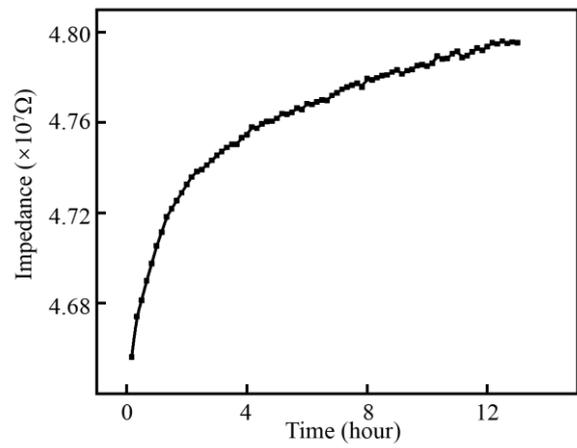


Fig. 6. Impedance of PZT sensor embedded in the specimen of epoxy/flax composite

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°C are shown in Fig. 7. An evident increasing tendency of storage modulus is found, which is associated with polymerization of epoxy.

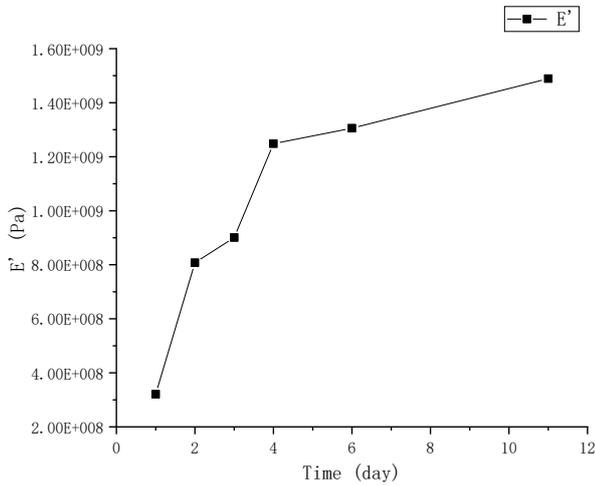


Fig. 7. Storage modulus of epoxy/flax composite measured by DMA tests in a series of days

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

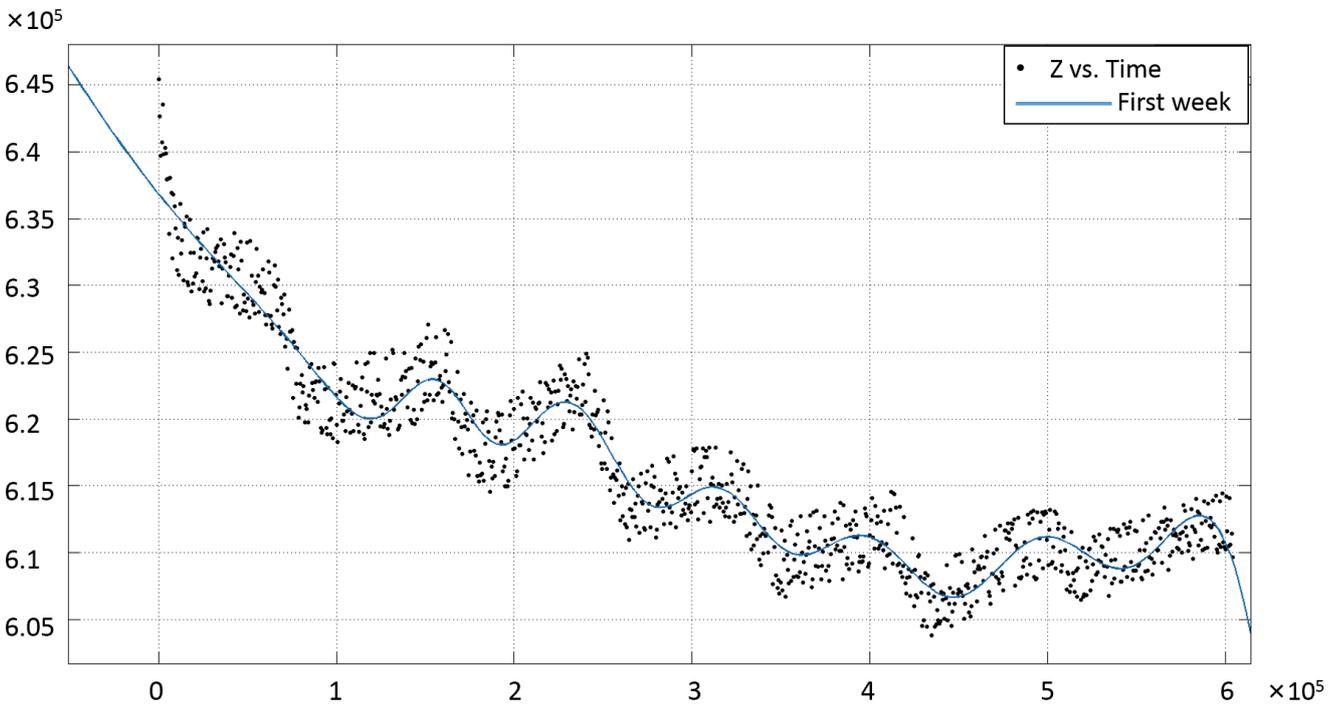


Fig. 8 Impedance of PZT embedded in epoxy/flax composite steeped in water

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 according to 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 varying orientations 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 energy harvesting". *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 LACAZE¹, Simon CHESNE¹, Didier REMOND¹

¹Univ 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.

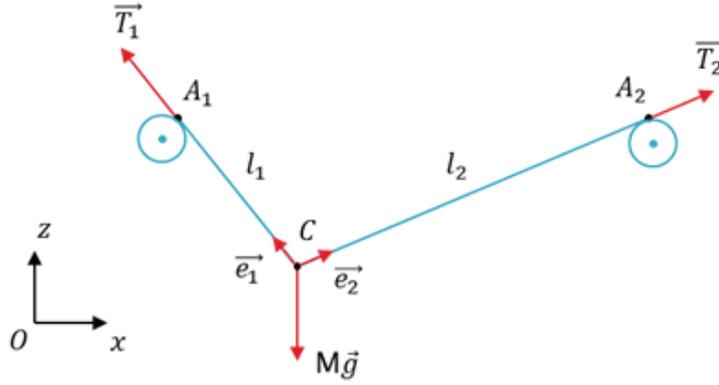


Figure 1: CDPR with 2 cables

Given the Young's modulus $E = 102$ GPa of the cable and its cross-sectional area $A = 1.76e^{-6}$ m², the cable tension can be described as :

$$T = EA \frac{l - l_0}{l_0} \quad (1)$$

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 :

$$\begin{pmatrix} \ddot{x} \\ \ddot{z} \end{pmatrix} = \frac{1}{M} (T_1 \vec{e}_1 + T_2 \vec{e}_2) + \vec{g} \quad (2)$$

Unit vectors state the direction of efforts transmitted through the cables. For the i^{th} cable, we have :

$$\vec{e}_i = \frac{\vec{OA}_i - \vec{OC}}{\|\vec{OA}_i - \vec{OC}\|} \quad (3)$$

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.

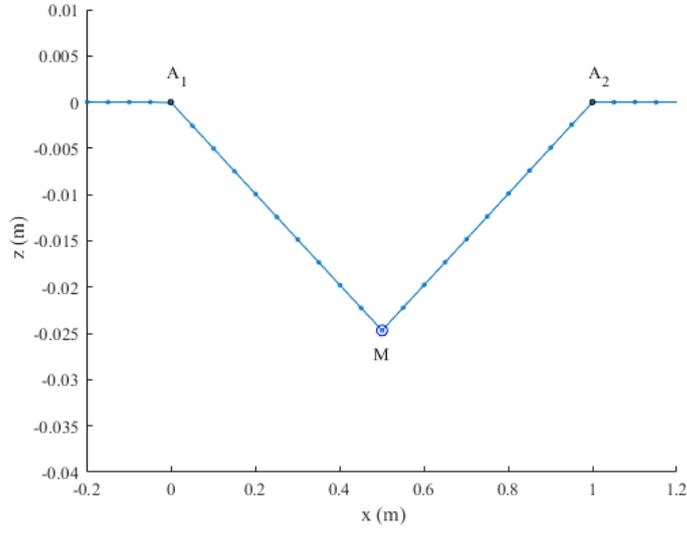


Figure 2: Lumped mass-spring model

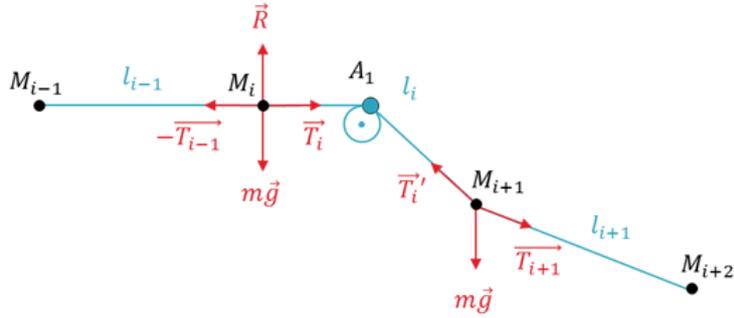


Figure 3: 3 elements of the lumped mass-spring model

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 :

$$\begin{cases} l_{i-1} = \|\overrightarrow{M_{i-1}M_i}\| \\ l_i = \|\overrightarrow{M_iA_1}\| + \|\overrightarrow{A_1M_{i+1}}\| \\ l_{i+1} = \|\overrightarrow{M_{i+1}M_{i+2}}\| \end{cases} \quad (4)$$

The tensions, computed using these lengths, are the following :

$$\begin{cases} T_{i-1} = k(l_{i-1} - l_0) \\ T_i = k(l_i - l_0) \\ T_{i+1} = k(l_{i+1} - l_0) \end{cases} \quad (5)$$

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 :

$$m\ddot{X}_i = \begin{pmatrix} -T_{i-1} + T_i \\ 0 \end{pmatrix} \quad (6)$$

$$m\ddot{X}_{i+1} = \begin{pmatrix} T_i \frac{x_{A_1} - x_{i+1}}{\|\overrightarrow{A_1M_{i+1}}\|} + T_{i+1} \frac{x_{i+2} - x_{i+1}}{l_{i+1}} \\ T_i \frac{z_{A_1} - z_{i+1}}{\|\overrightarrow{A_1M_{i+1}}\|} + T_{i+1} \frac{z_{i+2} - z_{i+1}}{l_{i+1}} \end{pmatrix} - m\vec{g} \quad (7)$$

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-dependent 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 :

$$T = \int_0^l \frac{1}{2} \mu \dot{\vec{r}}^T \cdot \dot{\vec{r}} ds \quad (8)$$

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 :

$$U = \int_0^l \left(\frac{1}{2} EA \varepsilon^2 - \mu g \vec{r}^T \cdot \vec{z} \right) ds \quad (9)$$

where ε is the element strain.

The changing mass of the system can be written as :

$$\delta H = \delta \vec{r}_j^T f_j + \mu \dot{\vec{r}}_1^T \cdot \delta \vec{r}_1 v_1 + \mu \dot{\vec{r}}_2^T \cdot \delta \vec{r}_2 v_2 \quad (10)$$

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 :

$$\int_0^t T dt - \delta \int_0^t U dt + \int_0^t \delta H dt = 0 \quad (11)$$

Eventually, this leads to the dynamic equation, which describes the position \vec{r}_j of the j index point :

$$\mathbf{m}_j \ddot{\vec{r}}_j + \mathbf{c}_j \dot{\vec{r}}_j + \mathbf{k}_j \vec{r}_j = \vec{f}_j + \vec{f}_j^g \quad (12)$$

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}_j^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.

$$\text{step5}(t) = \begin{cases} x_1 & \text{if } t < t_1 \\ x_1 + a \Delta^3(t) (10 - 15\Delta(t) + 6\Delta^2(t)) & \text{if } t_1 \leq t < t_2 \\ x_2 & \text{if } t \geq t_2 \end{cases} \quad (13)$$

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.

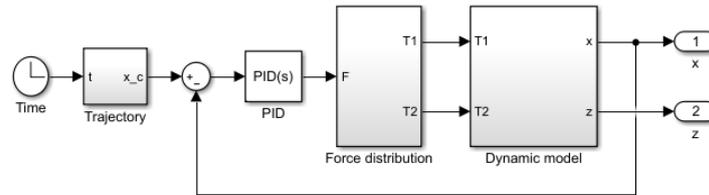


Figure 4: Simulink diagram

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.

$$T_1 \vec{e}_1 + T_2 \vec{e}_2 + M \vec{g} = \vec{0} \quad (14)$$

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.

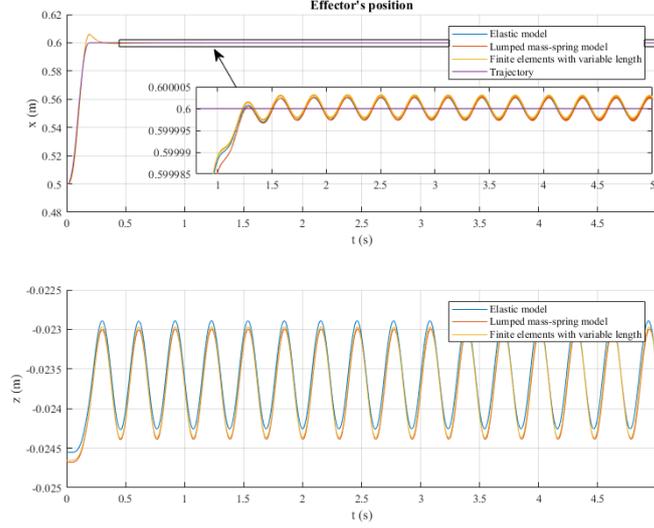


Figure 5: Evolution of x and z effector's position

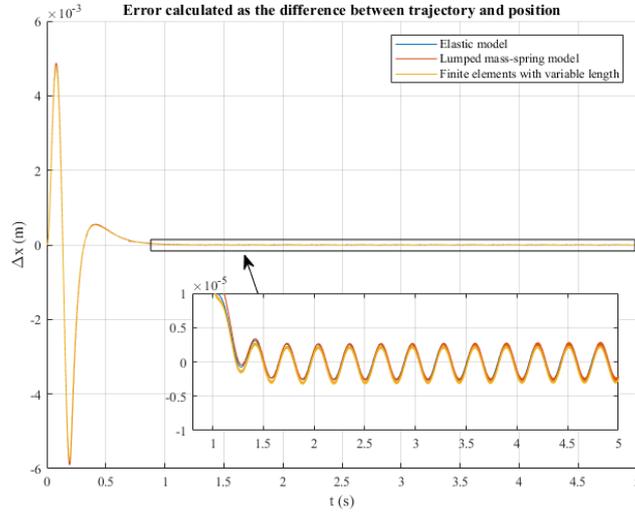


Figure 6: Position error of the CDP

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,

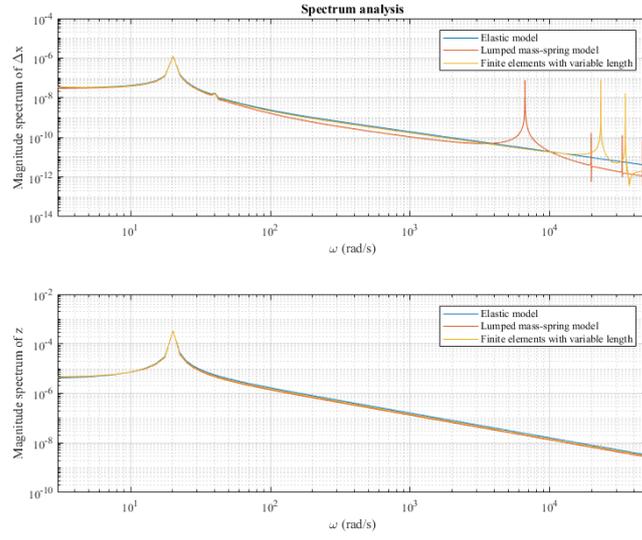


Figure 7: Spectrum analysis of Δx and z

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 :

$$\omega = \sqrt{\frac{T}{M} \left(\frac{1}{x} + \frac{1}{d-x} \right)} \quad (15)$$

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.

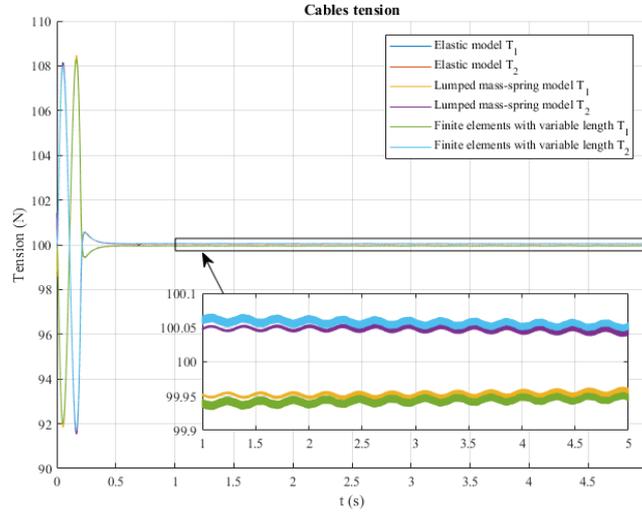


Figure 8: Cables tension T_1 and T_2

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.