INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

SURVEILLANCE
VIBRATIONS
SHOCKS
NOISE

Vibes in action

8-10 JULY 2019
INSA LYON

Proceedings

BOOK OF PROCEEDINGS

SURVISHNO 2019

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

General Chair : J. Antoni, D. Rémond

Table of contents

Monday 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, Aghazadeh 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-Thinh [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 operating 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 simulations, Bachar Lior [et al.] 379

Towards a better understanding of helical gears vibrations – dynamic model validated experimentally, Silverman Nadav [et al.] 382

Modeling and identification of mechanical systems **385**

Localization and quantification of damage by frequency based methods : Numerical 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 tumbling mill, Molina Vicuna Cristián [et al.] 406

Signal processing **428**

Comparison and Improvement of Techniques for Transmission-Path Restoring, Matania 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 flextensionnal 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**

CMBase, 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**

Tuesday

Angular approaches

NUMERICAL AND EXPERIMENTAL LOADS ANALYSIS ON A HORIZONTAL-AXIS WIND TURBINE IN YAW

Davide Astolfi[1], Francesco Castellani[1], Francesco Natili[1], Matteo Becchetti[1]

[1]University of Perugia, Department of Engineering, Via G. Duranti 93 - 06125 Perugia (Italy)
davide.astolfi@unipg.it, francesco.castellani@unipg.it,
francesco.natili@collaboratori.unipg.it, matteo.becchetti@unipg.it

Abstract
The characterization of wind turbines in yawed conditions is one of the most important topics as regards the latest advances in optimizing power production and mechanical behavior in wind farms. The classical wind turbine control strategy consists in keeping the rotor constantly aligned with wind direction: whereas this approach maximizes the power coefficient of each single turbine, it might not be the best solution when, in a wind farm, upwind turbines generate wakes on downwind ones. Considering this, yawing the rotors gives a steer to wakes, improving the flow on downwind turbines. This new kind of control strategy has been attracting the scientific interest not only by an energetic point of view, but also as regards the mechanical behavior of turbines operating not aligned with wind, in particular for what concerns generation of forces and vibrations. On these grounds, the aim of this paper is to study in deep how a wind turbine works on yawed configurations. In order to do this, wind tunnel tests have been performed with yaw angles that range over from $-45°$ to $45°$ on a 2 m. diameter small scale wind turbine. Experimental measurements of forcespower and tower vibrations are then compared with the results of simulations from two different codes. The first, called BEM, is internally developed following the principles of Blade Element Momentum theory and it is used to estimate forces and torque acting on the rotor. The second implemented model is developed using the FAST (Fatigue, Aerodynamics, Structures and Turbulence) software, developed at the National Renewable Energy Laboratory (NREL). FAST simulations provide in output forces, torque and vibrations of tower and blades. Simulations are set up with similar conditions as the wind tunnel tests, with many yaw angles and steady wind speed. One of the main results of this study is that there is a remarkable agreement between simulations and measurements as regards the estimate of the power coefficient C_P in yawed and non-yawed configurations. In spite of this, thrust coefficient C_T is not faithfully estimated when the yaw angles is vanishing. This matter of fact is then explained by the fact that low-fidelity numerical models are not capable in reproducing reliably the effect of the tower blockage, slowing down the air stream in its proximity. As a consequence, when a blade passes close this area of reduced flow speed, the generation of aerodynamic forces decreases. In yawed configurations, this phenomenon is less relevant because of the increased distance between blades and tower on air flow direction.

1 Introduction

The optimization of power production and mechanical behavior of wind turbine through advanced control strategies [1, 2] has been recently becoming one of the main topics in wind energy research. The blade pitch [3, 4] and the yaw management are two very fertile fields of investigation for the research in wind turbine control optimization [5, 6].

The classical approach to control wind turbines nacelle orientation consists in continuously following the wind direction, in order to maintain the rotor axis constantly parallel to air flow. This method guarantees that the single turbine has always the maximum tip-speed ratio and so the maximum energetic production. In spite of this, in wind farm configurations an aspect that has to be considered is that upstream wind turbines generate wakes affecting the downstream ones and this affects the power production [7] and the mechanical loads [8]. By this point of view, new methods for active wind farm control are oriented: the general idea is that a slight decrease of the power produced by the single upwind turbine can optimize the total production of the entire farm. Haces-Fernandes et al. [9] states that a selective turbine deactivation allows an enhancement on wind

farm production and found that improvement is more prominent as the size of the turbines, and so rotor diameter, increases. Another method is the derating [10], that consists on running some turbines with a non-optimal rotational speed in order to catch less energy from air stream but generating less wake effect on downwind turbines: in this case the increase of energy production can range from 1.86% to 6.24%. A novel approach, instead, called wake steering [11, 12, 13], consists in keeping the upwind turbines not aligned with wind direction with the purpose to deviate the wakes and let the downwind turbines to be invested by an air stream with a more energetic content. For a single wind turbine, the power is related to ϕ, the yaw angle between the wind flow and the rotor, with a cosine cube law [14]: as a consequence, the energy production decreases as the yaw angle increases. To understand the net increase of generated power at wind farm level, Archer et al. [12] found that yawing the first row of a turbine array of 20°, the power of the following rows increases profitably (more than the losses of the first row).
To have a wider outlook on wake steering wind farm control, it is necessary to consider not only the effect on energy production, but also the possible side effects on wind turbine structural integrity. For example, Bakshi [15] estimated the reliability of blades in yawed asset, performing a stress analysis in different yaw configurations.

On these grounds, the present study aims at providing a contribution to the experimental analysis and numerical characterization of horizontal-axis wind turbines in yawed conditions. Wind tunnel measurements on a 2 m. diameter turbine are performed, with yaw angles ranging from 0 to $\pm 45°$: forces generated by the rotor, nacelle accelerations, rotational speed and generator power are monitored. Experimental data are then compared to numerical results of simulations performed with two different algorithms. The first is called BEM and has been internally developed according to Blade Element Momentum theory. This code allows an estimation of forces generated by the rotor. The second code, FAST, is developed by NREL, National Renewable Energy Laboratory, Colorado, and is one of the most used software for aeroelastic wind turbine modeling. In FAST, it is possible to obtain in output information concerning power, forces, moments, torque, accelerations and deformations. This software is frequently used to simulate large size wind turbines: by this point of view, one of the purposes of this study is to investigate the reliability of the FAST environment for small wind turbine simulation too. Actually, the critical point is that small wind turbines are strongly affected by fatigue, as a result of their size and the variability of loads , induced by the unsteady wind conditions (especially in urban environment [16]), and modulated by a very high rotational speed [17]. It is therefore interesting to understand the capability of simplified numerical models in reproducing reliably the dynamical behaviour of this kind of devices, especially in yawed conditions.

This paper is organized in the following sections: Section 2 presents the methods and facilities and a discussion on the equipment used. In section 3 the results are presented and examined. Finally Section 4 is devoted to conclusions and future developments.

2 Experimental Set Up and Numerical Models

In this study, experimental tests in wind tunnel and numerical tools are used to characterize the behaviour of a small wind turbine in yawed configuration.

2.1 Wind Turbine and Wind Tunnel for Experimental Tests

The HAWT prototype selected for this work has these main features:

- 40 kg nacelle mass;
- rotor diameter: 2 meters;
- hub height: 1.2 meters;
- hub radius is 0.13 meters;
- minimum chord of the profile: 5 cm. Maximum: 15 cm;
- angle of attack variable between 1.7° and 32°;

- the prototype is equipped with three polymer reinforced with fiberglass blades;
- fixed pitch angle;
- operative rotational speed between 200 and 700 RPM;
- 3 kW of maximum power;
- electric control based on experimental optimal power curve.

In Fig. 1 the test case wind turbine placed inside the wind tunnel is represented; the configuration in the Figure is at 0° of yaw angle but many tests have been performed with yaw angles of up to ±45°.
The wind tunnel used for this research is located at the Department of Engineering at the University of Perugia, Italy (www.windtunnel.unipg.it). The facility consists on a closed loop, open test chamber wind tunnel with a squared cross section of the ducts of 2.2 m. per side. The recovery section is about 2.7 m. x 2.7 m. A 375 kW electric motor puts in rotation a fan that is able to produce variable wind speed in the test section up to 45 m/s. A peculiar characteristic of this tunnel is the extremely low turbulence of the air flow that can be quantified in 0.4%. The wind speed is measured by a Pitot tube and a cup anemometer placed at the inlet section. In Figure 2 a scheme of the wind tunnel is reported.

Figure 1: The small HAWT in the wind tunnel open test section.

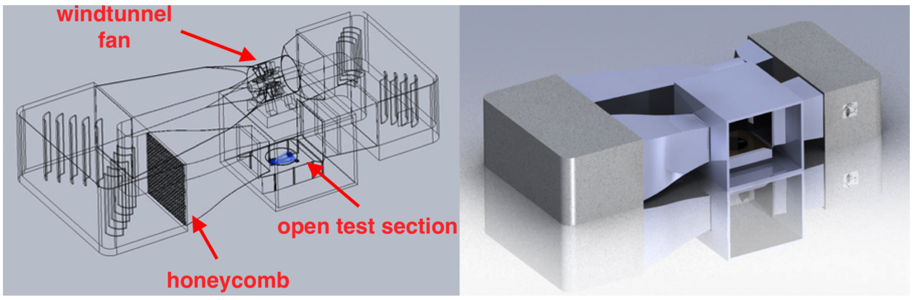


Figure 2: A sketch of the wind tunnel.

As the test section of the wind tunnel is not a free field, it has to be considered a confined environment where the airflow gets modified by the presence of the turbine itself: this phenomenon is called blockage. To consider the blockage, in this discussion wind velocities and thrust or power coefficients will be scaled by a corrective factor, BF (Blockage Factor), estimated following Kinsey and Dumas [18] as in eq.1:

$$BF = \frac{U}{U'}, \tag{1}$$

where U is the free stream wind speed in the wind tunnel with the rotor and U' without the presence of the rotor. Using the blockage factor, it is possible to correct both power and thrust coefficients as expressed by eq.2 and eq. 3:

$$C'_P = C_P \cdot \left(\frac{U}{U'}\right)^3 = C_P \cdot BF^3 \tag{2}$$

$$C'_T = C_T \cdot \left(\frac{U}{U'}\right)^2 = C_T \cdot BF^2, \tag{3}$$

where C'_P and C'_T are the corrected power and thrust coefficient. Previous experimental and numerical studies, in particular Eltayesh et al. [19], have been devoted to the analysis of the blockage factor of the wind tunnel of University of Perugia and the results have been employed for the purpose of this study to correctly estimate the reference free wind speed. According to this, to reliably compare numerical and experimental tests, it should be intended that the C_P and C_T factors obtained from simulations are the corrected ones.
The HAWT has been subjected to steady wind time series having duration of 60 s. During each time series, the yaw angle has a fixed value. The tested yaw angles are:

- 0°;
- ±22.5°
- ±45°.

The selected wind intensity is 10 m/s and some tests have been performed at 8 m/s too.

2.2 The FAST Software

FAST (Fatigue, Aerodynamics, Structures,and Turbulence) is an open-source aeroelastic software developed by NREL (National Renewable Energy Laboratory) and it is used to perform simulations of energetic and mechanical behaviour of horizontal axis wind turbines.
This software offers many alternatives to customize the modeling of turbine components. Electric generator, yaw controller, pitch controller and shaft brake can be modeled in many ways; the most used includes the use of subroutines, look up tables and the interface with external software environments. The number of input files depends on the characteristics of the simulation. In this test the employed input files are:

- Primary: is the main file where simulation parameters can be setted and contains the link to the other files.
- InflowWind: this file describes the wind characteristics. Data about wind speed magnitude,vertical and horizontal components has to be implemented in this file. In addition, it contains the spatial discretization resolution.
- AeroDyn: it includes environment air condition, links to the table of blade airfoils polars, and tower aerodynamic properties.
- ElastoDyn: in this file, the wind turbine mechanical design (pre-cone, tilt angle, masses and inertia) is described. Links to blades and tower shape modes are also included.
- ServoDyn: it manages the behavior of the controllers.Through this file it is possible to implement generator, pitch, yaw and braking models.

Figure 3 is a block chart representing how the files interact during the simulation running.

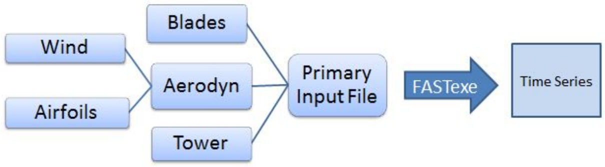


Figure 3: A flowchart of the FAST simulation.

As usual for small wind turbines, the model studied in this paper does not have an active pitch or yaw control, as discussed for example in Scappaticci [20]. To meet market requests and in consideration of the lack of adequate spaces to house actuators, small wind turbines are typically not equipped with advanced control systems. For this reason, in ServoDyn, only the electric generator is modeled. FAST offers many solutions to set up the simulation of generator, in this case, a look up table is considered the best solution. The of external software (like Simulink) is a better choice when PID (Proportional, Integrative, Derivative) controllers have to be implemented, especially for unsteady simulations. In the present paper, instead, simulations are always performed in steady conditions. Neither the default generator model, present in FAST, can be profitably used because it is arranged for large wind turbines. According to this, the choice of look-up table is the most suited among all the possibilities.
Look-up table creates a relationship between the instant rotational speed of the shaft and the resistant torque that has to be applied. Electric generator response is tested experimentally in wind tunnel steady runs with variable wind speed. In this way, once the system reaches the equilibrium, shaft speed and the corresponding torque is collected and than used to create the look up table. It has to be noticed that turbine power controller works according to MPPT (Maximum Power Point Tracking) in order to always find the best performance in terms of power production. FAST allows to impose and keep fixed the the yaw angle in ElastoDyn file. The tested yaw angles are the same as the experimental ones: $0°$, $\pm 22.5°$ and $\pm 45°$.

2.3 The BEM Algorithm

The second numerical framework used to estimate mechanical loads on wind turbine is internally developed on the grounds of the BEM (Blade Element Momentum) theory. Many handbooks on wind turbine aerodynamics explain this mathematical approach; in the following, we refer to a summary by Burton[21]. Drag and lift coefficients are defined in eq.4 and eq.5:

$$C_{lift} = 2\frac{L}{\rho A_{ref} U_\infty^2} \tag{4}$$

$$C_{drag} = 2\frac{D}{\rho A_{ref} U_\infty^2}, \tag{5}$$

where L and D are the aerodynamic lift and drag forces; ρ the air density; A_{ref} the area of wind turbine rotor and U_∞ the free stream wind speed. Moreover, labeled as U_d the wind speed at the disk, it is possible to introduce the axial induction factor a', eq.6:

$$a = \frac{U_\infty - U_d}{U_\infty} \tag{6}$$

To keep in account the rotation effect that the disk imparts to the downstream flow, the a' coefficient is introduced, eq.7

$$a' = \frac{1}{2\Omega}\omega, \tag{7}$$

labeled as ω the angular velocity of the wake imparted by the rotor, whose velocity is Ω. Using the a factor, it is possible to rewrite the axial and tangential speeds as eq.:

$$V_x = U_\infty(1-a) \qquad\qquad V_y = \Omega R(1+a). \tag{8}$$

After the calculation of speed components, the angle of attack(ϕ) on each section of the blades is obtained using polar charts available from aerodynamic simulation software (i.e. Xfoil). Knowing ϕ, the C_x and C_y coefficients (eq. 9) can be computed:

$$C_x = C_l \cos(\phi) + C_d \sin(\phi) \qquad\qquad C_y = C_l \sin(\phi) + C_d \cos(\phi). \tag{9}$$

As stated by Ning[22], it is possible to obtain tip and loss coefficients using eq.10 and eq.11:

$$f_{tip} = \frac{B}{2}\frac{(R-r)}{|sin\phi|} \qquad\qquad F_{tip} = \frac{2}{\pi}acos(e^{-f_{tip}}) \tag{10}$$

$$f_{hub} = \frac{B}{2}\frac{(r-R_{hub})}{R_{hub}|sin\phi|} \qquad\qquad F_{hub} = \frac{2}{\pi}acos(e^{-f_{hub}}), \tag{11}$$

where:

- F_{tip}: tip loss correction;
- B: blade number;
- R: rotor radius;
- r: distance from center of the rotor to root blade section;
- F_{hub}: hub loss correction.

Introducing the solidity σ as eq.12:

$$\sigma = \frac{Bc}{2\pi r}, \tag{12}$$

with c representing the chord length, then one can write eq13:

$$k = \frac{\sigma C_x}{4\sin\phi\sin\phi F} \qquad\qquad k' = \frac{\sigma C_x}{4\sin\phi\cos\phi F}, \tag{13}$$

considering $F = f_{tip}^2$. Many formulations of a are available according to the values of ϕ and k; in particular for $\phi > 0$ and $k > 2/3$ equations 14 and 15 can be used:

$$\gamma_1 = 2Fk - (\frac{10}{9} - F) \qquad\qquad \gamma_2 = 2Fk - F(\frac{4}{3} - F) \qquad\qquad \gamma_2 = 2Fk - (\frac{25}{9} - 2F) \tag{14}$$

$$a = \frac{\gamma_1 - \sqrt{\gamma_2}}{\gamma_3}. \tag{15}$$

If $\phi < 0$ and $k > 1$, the axial induction factor has the following formulation:

$$a = \frac{k}{k-1}. \tag{16}$$

Instead, if $\phi > 0$ and $k < 2/3$, one obtains

$$a = \frac{k}{k+1} \tag{17}$$

For a', a unique formulation is obtained:

$$a' = \frac{k'}{k'+1} \tag{18}$$

From the aforementioned equations, induction flow factors can be estimated in each solution region. When a turbine is yawed, otherwise, it is necessary to consider additional corrective factors: in this case the induction factor with yaw correction is eq.19:

$$a_{yaw} = a(1 + K\frac{r}{R}\sin(\psi)), \tag{19}$$

where ψ is the azimuth angle and K, from Shen[23], is given in eq20:

$$K = \frac{15}{32}\pi \tan\frac{\chi}{2}. \tag{20}$$

χ is known as skew angle and it is obtained from eq.21:

$$\chi = (0.6a+1)\gamma \tag{21}$$

In literature, different formulations for correcting the induction factor considering yaw are available [24]. For example Coleman[25] proposed eq.22:

$$K = \tan(\frac{\chi}{2}). \tag{22}$$

For White and Blake [26] one has eq.23:

$$K = \sqrt{2}\tan(\chi). \tag{23}$$

Moreover Shen[23] proposed eq.24:

$$a_{yaw} = a\left[1+\frac{15\pi}{32}\sqrt{\frac{1-cos\gamma}{1+cos\gamma}}\frac{r}{R}Ksin\psi\right]. \tag{24}$$

Different, additional formulations have been proposed by Ackermann [27] and Bianchi [28]. All these formulations have been compared but noticeable differences that may cause substantial changes to the algorithm have not been found.

3 Results

3.1 Analysis of Power and Thrust Coefficients

In this section, results from simulation codes and experimental measurements are shown. Figure 4 compares the measured and simulated C_p values in different yaw configurations. C_p is computed using eq. 25:

$$C_P = \frac{P}{\frac{1}{2}\rho A_{ref}U_\infty^3}, \tag{25}$$

where P is the generator power. Physically C_p measures the ratio between the power that is produced and the kinetic energy of the flow. The maximum theoretical limit of this coefficient is indicated by the Betz law [29].

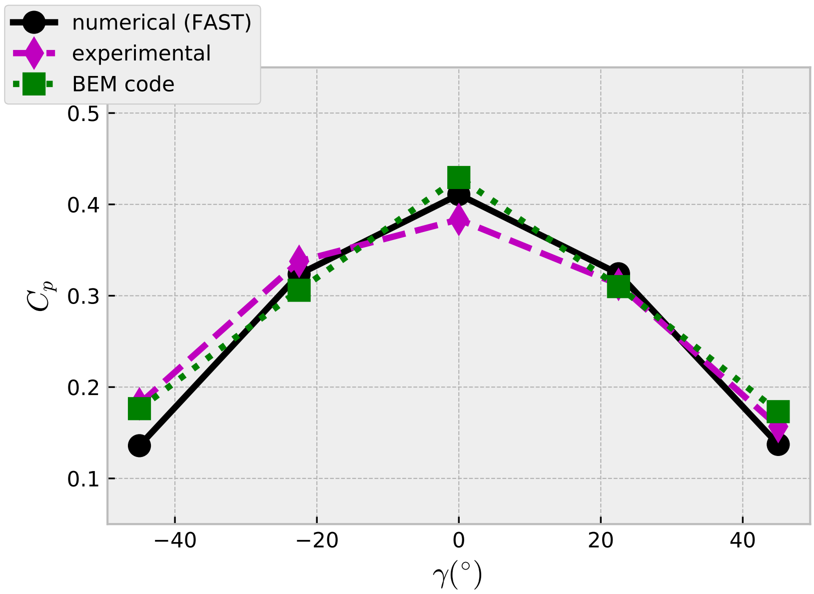


Figure 4: Power coefficients at 10 m/s: experimental vs numerical results.

Similarly to power coefficient, thrust coefficient is used to characterize turbine behavior. Its definition is given in eq.26:

$$C_T = \frac{F}{\frac{1}{2}\rho A U_\infty^2} = 4a(1-a), \tag{26}$$

where F is the thrust force acting on the rotor in the flow direction. In figure 5, the measured and simulated behavior of C_T is shown.

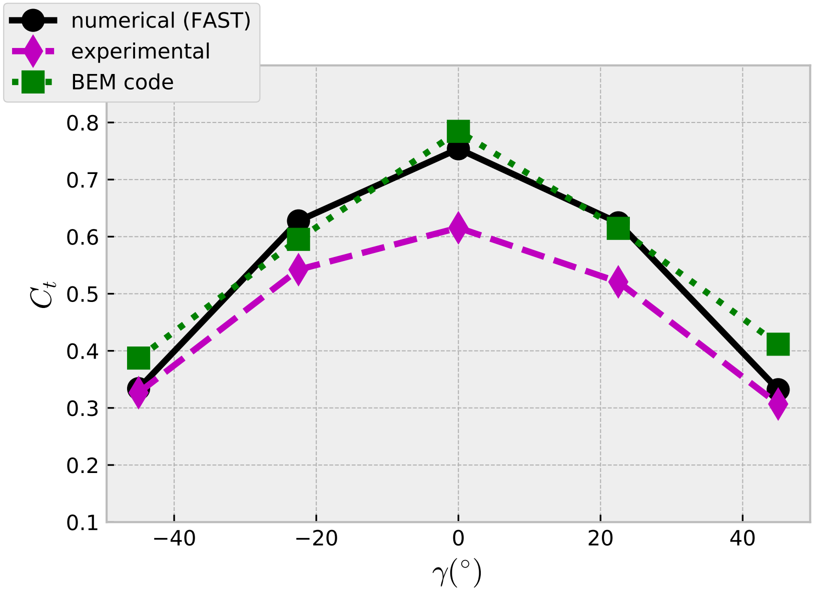


Figure 5: Thrust coefficients at 10 m/s: experimental vs numerical results.

It can be seen that both numerical models reproduce the power coefficient fairly but they largely overestimate the thrust one. The maximum percentage error for C_P is about 5% for the BEM code and 8% for FAST. In spite of this, for C_T errors are up to 25% for BEM and 20% for FAST. In numerical simulations a 95% of generator efficiency has been considered and the small errors on C_P for vanishing yaw angles shows that it can be considered a reliable estimation.

The mismatch between measured and simulated C_T coefficient can be imputable to multiple causes: the most important can be supposed to be the fact that the numerical models do not take into account blades deformations. In section 3.2, it will be discussed how the combined effects of blade deformation and tower blockage are linked to yaw configuration. In fact, the aerodynamic thrust generation depends on the distance between blade and tower in stream direction. When the turbine is yawed, this distance tends to be increased and the blades are affected by a lower tower blockage effect producing more thrust: in this case, the mismatch as regards the C_T, where blockage is not implemented, is negligible.

The slight asymmetry, visible in experimental tests, can be related to wind tunnel layout. In the open test chamber the lateral walls are placed at different distances respect the air stream and the turbine rotor. Because of this, the flow that impacts the rotor at negative or positive yaw angles is slightly different. Anyway, the discrepancies are estimated to be 3% on C_P and 8% on C_T.

3.2 Study of Thrust Cyclic Variation

Because of the critical issues revealed by the previous analysis, it has been considered useful to set up a study devoted to the cyclic variations of the aerodynamic forces during a complete rotation of a blade. The reference angle is denoted as azimuth. From Figure 6, it arises that there is periodic component in correspondence of the first blade passing frequency 3P. This phenomenon is well known and can be interpreted as due to the interaction between tower and airflow, causing cyclic decrease of aerodynamic forces. As previously explained, the intensity of the fluctuations is lower for increasing yaw angles because the blade passes farther from the low velocity air situated close to the tower.

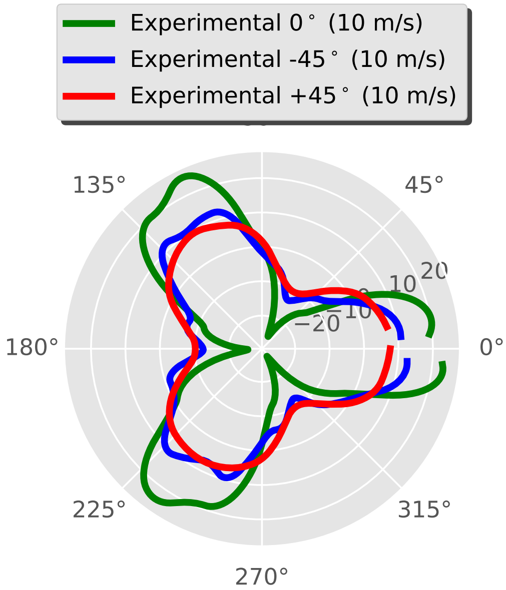


Figure 6: The rotor thrust variation(N), with respect to the average value, as a function of the azimuth angle.

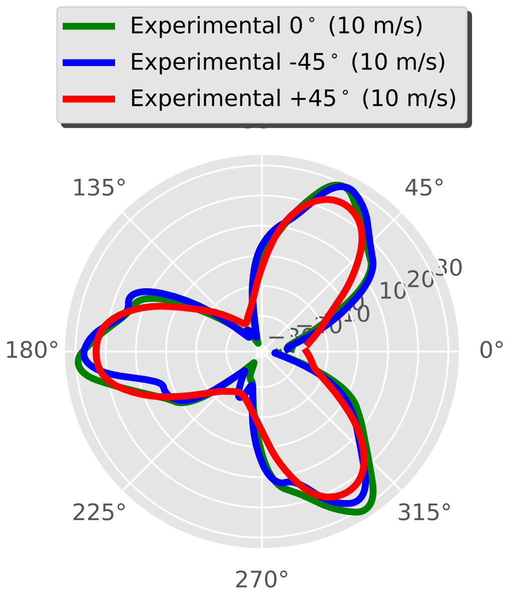


Figure 7: The rotor thrust relative variation(N), with respect to the average value, as a function of the azimuth angle.

The curves of thrust as a function of the azimuth angle have been scaled with respect to the corresponding mean value in fig. 7. It can be seen that there is a consistent overlap: this means that the amplitude of oscillations is not dependant on the yaw angle.

Additional experimental tests with a wind speed of $8m/s$ has been performed to discover the dependence of thrust oscillations with flow characteristics at $0°$ yaw value. The results are reported in fig. 8.

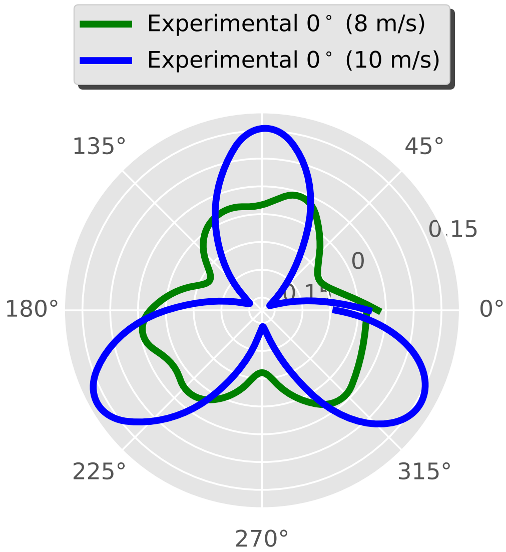


Figure 8: The rotor thrust coefficient cyclic variation in different wind tunnel tests with a wind speed of 8 m/s and 10 m/s.

From the comparison between the $10m/s$ and the $8m/s$ tests, it results that the C_T fluctuations at 10 m/s are more than doubled with respect to 8 m/s. According to this, it can be stated that the tower interference has a less relevant effect as the wind speed tends to be lower. Blade deflections in facts are strictly dependant on the aerodynamics loads and, since they decrease when the wind speed decreases, the space between the deflected blade and the tower increases and therefore the thrust is less affected by blockage. To quantify the blade deflection, it has to be considered that in previous measurements campaigns the blade deflection at the tip has been measured to be around 7% with a wind speed of $32m/s$, so it is expected that at 10 m/s it is of the order of 1% of the rotor radius: it is remarkable that this small value can induce such large tower interaction effects.

3.3 Analysis of Tower Interference on Accelerations

The above results indicate that the tower inference is a non-negligible aspect of the dynamical behavior of the small scale wind turbine considered in this study. Wind tunnel tests have been useful to deeply understand how tower interference effect gets modified by the yawing the turbine: this has been possible thanks to the triaxial accelerometer located on the nacelle, recording the aerodynamic induced vibrations. Fast Fourier Transform (FFT) theory has been used to analyze the spectrum of vibrations and making a comparison between 0° yaw and +45°, fig.9.

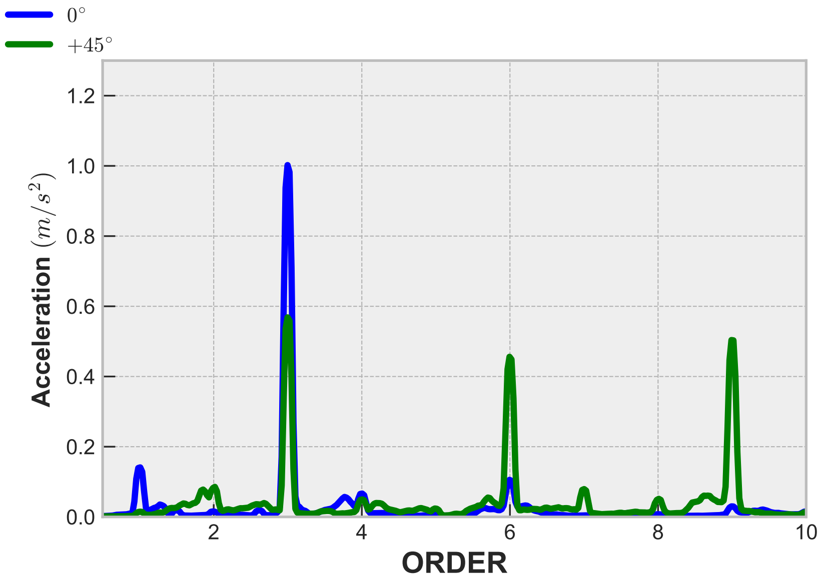


Figure 9: Experimental normalized order spectrum of the acceleration (fore-aft component normalized on the amplitude of the 3P component with zero yaw).

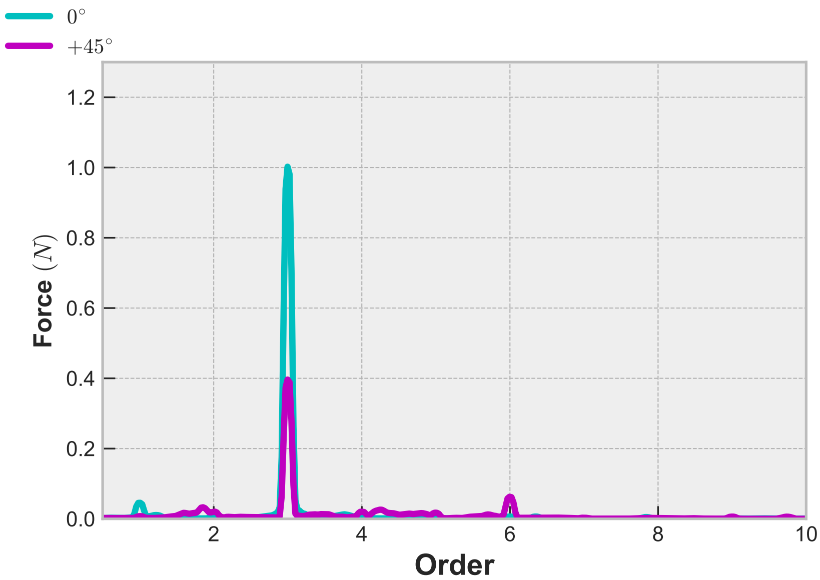


Figure 10: Experimental normalized order spectrum of the forces (fore-aft component normalized on the amplitude of the 3P component with zero yaw).

Figure 10 is the FFT of the thrust force measured by load cell placed between tower top and nacelle. Spectral analysis shows that order 3P, related to tower blockage, undergoes a substantial decrease passing from 0° yaw to 45°. This is an additional proof that gives consistency to the thesis that tower inference is more prominent for vanishing yaw angle and that justifies the thrust overestimation obtained with numerical codes, where it is not possible to account for tower blockage.

4 Conclusion

The objective of this study was the characterization of the mechanical behavior of horizontal axis wind turbines in yaw configuration. This field of study is attracting the scientific interest because yawing turbines allows several types of wind farm cooperative control, as for example the wake steering, that is useful to optimize the energy production.

A small scale horizontal axis wind turbine, with 2 m. of rotor diameter has been tested at the wind tunnel facility of the University of Perugia. The prototype has been equipped with accelerometers, load cell, electric power meter and tachometer in order to collect information about its operative conditions when undergoing different yaw angles. Experimental tests have been performed with a wind speed of 10 m/s and $\pm 45°$,$\pm 22.5°$ and $0°$ of yaw angle; additional tests have been carried out with 8 m/s of wind speed at $0°$ yaw. In addition, two numerical models have been adopted with two different software: an internally developed BEM algorithm and the open source FAST code. The numerical models have been set up in order to reproduce the conditions of the experimental test: the results are then compared in therms of power coefficient C_P and thrust coefficient C_T.

The main result is that the numerical models fairly reproduce the C_P coefficient. There are more critical points as regards the thrust coefficient: whereas for the cases of yawed configurations, the simulations fairly agreed with experimental test, for vanishing yaw angle the discrepancy is remarkable. This fact has motivated further analysis of the experimental data and the interpretation is that the mismatch between simulation and measurements is given by the fact that the tower blockage is a relevant phenomenon that the numerical models employed in this work do not take into account. The tower blockage is related to the streamwise distance between blades and tower, and so it is related to blade deflection too.

The cyclic variation of the thrust as a function of the azimuth angle has been analyzed for different yaw configurations, confirming the presence of a $3P$ periodicity which testifies the presence of a tower induced blockage effect variable respect to yaw angle. Under this circumstance an experimental test, with a wind speed of 8 m/s has been useful to confirm that with lower aerodynamic loads also the blade deflections decrease and so they do the lower thrust oscillations.

In addition, FFT analysis has been used to compare order spectra of nacelle accelerations and thrust for the $0°$ and $45°$ yaw configurations. The accelerations and thrust at 3P order appear diminished when the rotor is yawed. As this order is characteristic of tower interference effect, the order analysis brings an additional argument is support of the fact that blockage phenomenon is strictly correlated to the yawing behavior and cannot be neglected when yaw angle tends to vanish.

The results of this study can be useful to increase the knowledge of the behaviour of small wind turbines in yawed configuration and to evaluate the ability of low-fidelity numerical models predicting loads on yawed rotors. Future improvements of this study regard the possibility of better characterizing the interactions between blade tip and turbine tower, possibly using CFD codes. This study can also be useful for the implementation of wake steering wind farm control where it is expected that turbines will runs in yawed configuration for long periods and so an accurate estimation of loads is a crucial step to guarantee best performances and to assess the fatigue loading of the machine.

References

[1] Torben Knudsen, Thomas Bak, and Mikael Svenstrup. Survey of wind farm controlâpower and fatigue optimization. *Wind Energy*, 18(8):1333–1351, 2015.

[2] Jackson G Njiri and Dirk Soeffker. State-of-the-art in wind turbine control: Trends and challenges. *Renewable and Sustainable Energy Reviews*, 60:377–393, 2016.

[3] Richie Gao and Zhiwei Gao. Pitch control for wind turbine systems using optimization, estimation and compensation. *Renewable Energy*, 91:501–515, 2016.

[4] Bing Han, Lawu Zhou, Fan Yang, and Zeng Xiang. Individual pitch controller based on fuzzy logic control for wind turbine load mitigation. *IET Renewable Power Generation*, 10(5):687–693, 2016.

[5] Dongran Song, Xinyu Fan, Jian Yang, Anfeng Liu, Sifan Chen, and Young Hoon Joo. Power extraction efficiency optimization of horizontal-axis wind turbines through optimizing control parameters of yaw control systems using an intelligent method. *Applied energy*, 224:267–279, 2018.

[6] Aitor Saenz-Aguirre, Ekaitz Zulueta, Unai Fernandez-Gamiz, Javier Lozano, and Jose Manuel Lopez-Guede. Artificial neural network based reinforcement learning for wind turbine yaw control. *Energies*, 12(3):436, 2019.

[7] Rebecca Jane Barthelmie and LE Jensen. Evaluation of wind farm efficiency and wind turbine wakes at the nysted offshore wind farm. *Wind Energy*, 13(6):573–586, 2010.

[8] Soo-Hyun Kim, Hyung-Ki Shin, Young-Chul Joo, and Keon-Hoon Kim. A study of the wake effects on the wind characteristics and fatigue loads for the turbines in a wind farm. *Renewable Energy*, 74:536–543, 2015.

[9] Francisco Haces-Fernandez, Hua Li, and David Ramirez. Improving wind farm power output through deactivating selected wind turbines. *Energy Conversion and Management*, 187:407 – 422, 2019.

[10] Mikel De-Prada-Gil, César Guillén AlÃas, Oriol Gomis-Bellmunt, and Andreas Sumper. Maximum wind power plant generation by reducing the wake effect. *Energy Conversion and Management*, 101:73 – 84, 2015.

[11] Paul A. Fleming, Pieter M.O. Gebraad, Sang Lee, Jan-Willem van Wingerden, Kathryn Johnson, Matt Churchfield, John Michalakes, Philippe Spalart, and Patrick Moriarty. Evaluating techniques for redirecting turbine wakes using sowfa. *Renewable Energy*, 70:211 – 218, 2014. Special issue on aerodynamics of offshore wind energy systems and wakes.

[12] Cristina L. Archer and Ahmad Vasel-Be-Hagh. Wake steering via yaw control in multi-turbine wind farms: Recommendations based on large-eddy simulation. *Sustainable Energy Technologies and Assessments*, 33:34 – 43, 2019.

[13] Paul Fleming, Jennifer R King, Katherine L Dykes, Eric J Simley, Jason M Roadman, Andrew K Scholbrock, Patrick Murphy, Julie K Lundquist, Patrick J Moriarty, Katherine A Fleming, et al. Initial results from a field campaign of wake steering applied at a commercial wind farm: Part 1. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2019.

[14] Shuting Wan, Lifeng Cheng, and Xiaoling Sheng. Effects of yaw error on wind turbine running characteristics based on the equivalent wind speed model. *Energies*, 8(7):6286–6301, 2015.

[15] Roozbeh Bakhshi and Peter Sandborn. The effect of yaw error on the reliability of wind turbine blades. In *ASME 2016 10th International Conference on Energy Sustainability collocated with the ASME 2016 Power Conference and the ASME 2016 14th International Conference on Fuel Cell Science, Engineering and Technology*, pages V001T14A001–V001T14A001. American Society of Mechanical Engineers, 2016.

[16] L Battisti, E Benini, A Brighenti, S DellâAnna, and M Raciti Castelli. Small wind turbine effectiveness in the urban environment. *Renewable energy*, 129:102–113, 2018.

[17] Francesco Castellani, Davide Astolfi, Matteo Becchetti, Francesco Berno, Filippo Cianetti, and Alessandro Cetrini. Experimental and numerical vibrational analysis of a horizontal-axis micro-wind turbine. *Energies*, 11(2):456, 2018.

[18] Thomas Kinsey and Guy Dumas. Impact of channel blockage on the performance of axial and cross-flow hydrokinetic turbines. *Renewable energy*, 103:239–254, 2017.

[19] Abdelgalil Eltayesh, Massimiliano Burlando, Francesco Castellani, and Matteo Becchetti. Experimental and numerical study of the wind tunnel blockage effects on the behaviour of a horizontal axis wind turbine. In *Lecture Notes in Civil Engineering*. Springer, 2019.

[20] Lorenzo Scappatici, Nicola Bartolini, Francesco Castellani, Davide Astolfi, Alberto Garinei, and Marcello Pennicchi. Optimizing the design of horizontal-axis small wind turbines: From the laboratory to market. *Journal of Wind Engineering and Industrial Aerodynamics*, 154:58–68, 2016.

[21] Tony Burton, Nick Jenkins, David Sharpe, and Ervin Bossanyi. *Wind energy handbook*. John Wiley & Sons, 2011.

[22] Andrew Ning, Gregory Hayman, Rick Damiani, and Jason M Jonkman. Development and validation of a new blade element momentum skewed-wake model within aerodyn. In *33rd Wind Energy Symposium*, page 0215, 2015.

[23] Wen Zhong Shen, Wei Jun Zhu, and Jens Nørkær Sørensen. Actuator line/navier–stokes computations for the mexico rotor: comparison with detailed measurements. *Wind Energy*, 15(5):811–825, 2012.

[24] Daniel Micallef and Tonio Sant. A review of wind turbine yaw aerodynamics. In *Wind Turbines-Design, Control and Applications*. InTech, 2016.

[25] Robert P Coleman, Arnold M Feingold, and Carl W Stempin. Evaluation of the induced-velocity field of an idealized helicoptor rotor. Technical report, NATIONAL AERONAUTICS AND SPACE ADMINISTRATION HAMPTON VA LANGLEY RESEARCH CENTER, 1945.

[26] Fred White and Bruce B Blake. Improved method predicting helicopter control response and gust sensitivity. American Helicopter Society, 1979.

[27] MC Ackerman. Yaw modelling of small wind turbines. *Journal of wind engineering and industrial aerodynamics*, 39(1-3):1–9, 1992.

[28] Serena Bianchi, Alessandro Bianchini, Giovanni Ferrara, and Lorenzo Ferrari. Small wind turbines in the built environment: influence of flow inclination on the potential energy yield. *Journal of Turbomachinery*, 136(4):041013, 2014.

[29] A. Betz. *Introduction to the theory of flow machines,*. Oxford Pergamon Press, 1966.

Gears and Bearings faults Detection: from Instrumentation to Classification.

Renaud Bertoni[1], Sylvain Barcet[1], Alexandre Carbonelli[1], Marion Cortes[1]

[1]VIBRATEC
28 Chemin du petit bois, 69130 ECULLY, France
renaud.bertoni@vibratec.fr

Gears and bearings, used in many industrial areas are subject to failure that may lead to costly shutdowns. The current trend is to detect failures (cracks, spall, pitting …) and to identify and control their evolution. Such monitoring leads to a huge amount of data. With a double skill in test and simulation, Vibratec proposes an approach based on measurements coupled with Machine Learning (ML) processing.

This presentation defines the fault detection global approach used by Vibratec, from signal acquisition to the classification of indicators. The methodology is firstly applied on a specific HMS test bench. Then, the machine learning strategy is deployed on a database. The numerical simulations are in good agreement with the measurement results obtained on the test bench, and the machine learning indicators provides encouraging results. In the upcoming months, this complete methodology will be applied on a collaborative project aiming to improve the maintenance of aircraft engines.

Measurement and use of transmission error for diagnostics of gears

Robert RANDALL[1], Zhan Yie CHIN[1], Wade SMITH[1], Pietro BORGHESANI[1]

[1]UNSW Sydney, School of Mech. and Manuf. Engineering, Sydney 2052, Australia
b.randall@unsw.edu.au

Abstract

Transmission error (TE) has long been thought to be a major contributor to gear vibration and noise, but insufficient consideration has been given to the different types of TE and how they generate vibrations. TE is defined as the difference in torsional vibration of two meshing gears, scaled so as to represent linear motion along the line of action. There are three distinct types of TE; 1) Geometric TE (GTE) given by deviations of the (combined) tooth profiles from ideal involute; 2) Static TE (STE) including elastic deformation of the teeth and therefore being load dependent; 3) Dynamic TE (DTE) including inertial as well as stiffness effects, and thus being speed as well as load dependent. It has long been recognized that TE can be measured very accurately by phase demodulation of the signals of shaft encoders rigidly attached to each of the gears in mesh, but only recently realized that all three types can be measured; GTE at low speed and low load, STE at low speed and higher load, and DTE at higher speed and higher load. This paper demonstrates that TE has several advantages over vibration acceleration (or even the raw torsional vibrations) as a diagnostic parameter, being close to the source (the gearmesh) and with “common mode rejection” from the two gears, thus being much less sensitive to operating conditions and rig parameters, including the much greater number of transfer paths, modulations, and resonances in the casing vibration measurements. The measurements in this paper were made on a single stage gearbox, over an input gear speed range from 2 – 20 Hz, and input shaft torque range from 0 – 20 Nm. Earlier measurements on the same gearbox were for soft gears which developed distributed pitting over an operating period of many hours. Unfortunately, the encoders used at that time (actually included in slip rings) had a low torsional resonance frequency, which precluded obtaining TE at higher than 2 Hz shaft speed, so only GTE and STE could be estimated. New results are presented here for ground, hardened gears with a simulated tooth root crack on one tooth. Not only does this illustrate the differences with a local fault, but new encoders were mounted, valid up to a shaft speed of 20 Hz, so that DTE could also be measured.

1 Introduction

Gear transmission error (TE) is defined as the difference in torsional vibration of two gears in mesh, scaled so as to represent linear motion along the line of action, this being common to the two gears. Already in 1996 [1], it was shown that TE could be measured simply and accurately by phase demodulation of the pulse signals from high quality shaft encoders on the free ends of the shafts on which the gears are mounted. The measured torsional vibrations, in terms of angular displacement, are scaled by the respective base circle radii, and subtracted to give relative motion along the line of action. The accuracy of the encoders themselves corresponds to fractions of a micron of TE, and virtually no further error is introduced by the phase demodulation processing by Hilbert transform techniques (as compared with the earlier use of analogue phase meters, or polynomial interpolation between pulses). It is often possible to mount the encoders on the free ends of the gear shafts (the section not transmitting torque) so that they follow the gear motions up to a very high frequency. The proposed application in [1] was to the measurement of TE in design, development and manufacture, to add to information gained from measurements using gear metrology machines, but it has also been proposed as a tool for gear diagnostics in [2]. However, at the time, that was limited by the necessity to mount encoders on the machines.

It is now becoming more common for encoders to be built into machines, to provide valuable information for both control and monitoring of, for example, variable speed machines such as wind turbines, and this will presumably increase with the adoption of the Internet of Things, so it is likely that measurement of TE will become more available as an indicator of gear faults.

Transmission error (TE) has long been thought to be a major contributor to gear vibration and noise, but the relationship between them has not been fully understood. For a start there are three distinct types of TE: 1) Geometric TE (GTE) given by deviations of the (combined) tooth profiles from ideal involute; 2) Static TE (STE) including elastic deformation of the teeth, and therefore being load dependent; 3) Dynamic TE (DTE) including inertial as well as stiffness effects, and thus being speed as well as load dependent.

Measurement and application of these three types of TE as a diagnostic tool were discussed in [3], but it was found that the encoders used there (actually included in slip rings) had a low resonance frequency, which precluded measurements at high enough speed to give DTE. The same test rig has now been equipped with high quality encoders, and the current paper uses new measurements with that system. Another difference is that the old measurements were made with soft gears, run for an extended period so that (uniformly distributed) pitting developed, but no distinct local faults. The current paper uses measurements made with hardened ground gears, but with a simulated tooth root crack seeded in one tooth on the pinion, to give information on local faults, and tooth root cracks in particular, this being one of the most critical faults, and most important to distinguish from less critical faults such as local spalls.

2 Test rig and measurements

The overall layout of the spur gear test rig is shown in Figure 1.

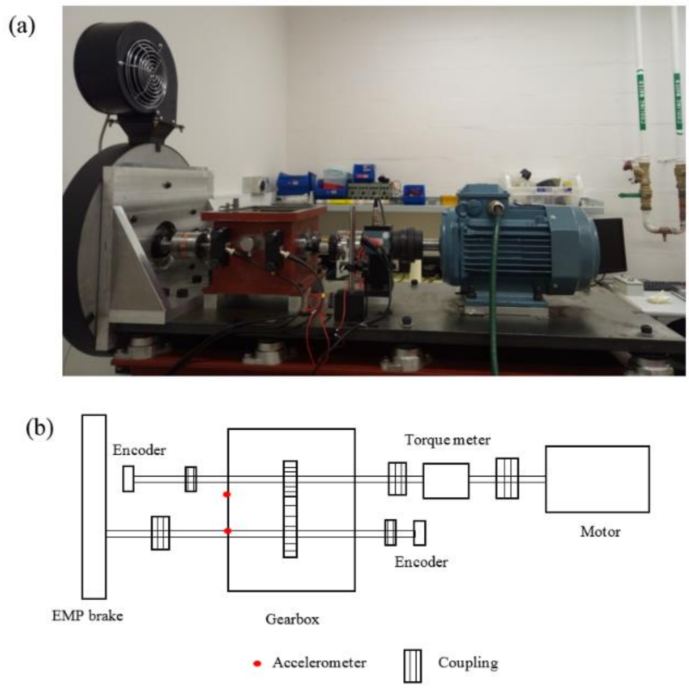


Figure 1: The spur gear test rig at UNSW. (a) photo; (b) schematic diagram

For the original measurements in [3] the reduction ratio was 19:52, and the gears were of mild steel. The original encoders were also slip rings, and had a low frequency resonance so that the highest valid input speed was 2 Hz.

For the new measurements, the reduction ratio was changed to 27:44 (same centre distance) and the gears were of hardened steel to avoid surface distress. The encoders were replaced by Heidenhain type ROD426, with 1000 pulses per rev, as well as a one per rev tacho pulse as a phase marker, and they gave valid results up to at least 20 Hz shaft speed. An EDM-generated half-tooth root crack (a 45° slot across the entire facewidth, 2.86 mm deep, extending to the tooth centreline, and 0.35mm wide) was seeded on one pinion tooth (input gear). Measurements were made at speeds 2, 5, 10, 15, 20 Hz, and loads 0, 5, 10, 20 Nm (all referenced to the input pinion). In addition to the encoder and tacho recordings, accelerometer measurements were made in the vertical direction on the casing above the input shaft at the motor end, and above both shafts at the brake end.

3 Results and discussion

3.1 Earlier results from the spur gearbox

A short summary is given here of the results published in [3], because they contain some findings which are different from those of the more recent measurements, published for the first time here. As mentioned, the gears were of mild steel and were run for a long period (nearly 50 hours) during which time they developed surface pitting fairly uniformly distributed around the gears. This was much more pronounced on the 19 tooth pinion than on the 52 tooth gear, because each tooth had a much greater number of contacts in inverse ratio to the tooth numbers, so only the pinion is discussed here.

Wear was monitored by trending the amplitude of the TE gearmesh harmonics (and the corresponding component of the synchronously averaged TE signal) in two conditions: low speed-low load (GTE) and low speed-high load (STE).

The effect of wear on GTE and STE showed an unexpected trend. The growth of the gearmesh harmonics was more pronounced on GTE during the first 6 hours of operation (mild pitting), and on STE later (severe pitting). The greater sensitivity of GTE in the initial phase was interpreted as being due to the fact that the unloaded GTE would have been dominated by (a few) local high spots at the edges of the pits, which would be easily deformed under relatively light load to give a reduced STE. On the other hand, with severe pitting more continuously distributed along the contact line, high spots would reduce the visibility of wear in GTE, and increased load would tend to give an increase in TE. Figure 2 shows a schematic representation of this interpretation, together with snapshots of the surfaces after about 2.5 and 42.5 hours of operation. For a detailed description of this test campaign the reader is referred to [4].

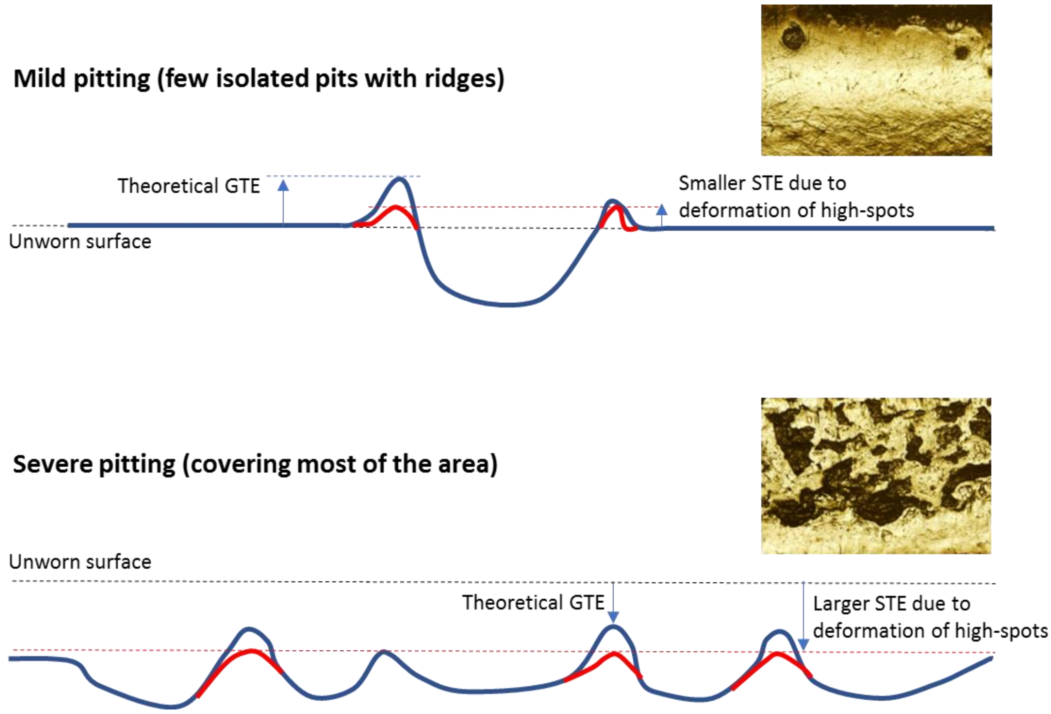


Figure 2: Schematic example of the interpretation of the effect of mild (top) and severe (bottom) pitting on GTE and STE, with corresponding example images of the gear surface.

In simulation models it is quite common to have GTE as a fixed value in series with the toothmesh stiffness. The latter is not always constant, but any nonlinearity is usually taken to correspond just to the extra compliance of the Hertzian component at low load, which still does not give a large difference in the overall stiffness, since the Hertzian component typically only represents about 25% of the total compliance, with the dominant bending stiffness component being almost linear. The above experience with “high points” does seem to indicate that, to obtain a reasonable match between such a simplified model and experiment, it would be better to use a value of GTE measured at a low, but non-zero, load sufficient to negate the effect of the high spots, and giving a more sudden transition to the Hertzian affected section of the stiffness curve.

Another interesting finding from the same study showed that, differently from TE, vibration was almost entirely insensitive to wear in both unloaded and loaded cases, at low speed. This was attributed to the fact that the proportion of the STE due to tooth deflection is still relatively small, but in fact it is only the dynamic tooth load, giving this deflection, which gives rise to vibration. At low speed there is no inertial resistance to rotation, so the driven gear can simply absorb the GTE by relative torsional motion, with almost no change in the GM spring force, even for the loaded case where the static load is almost constant. It could be expected that for DTE the much greater angular accelerations involved might prevent the driven gear from simply “moving out of the way” and thus force tooth deflection and increased vibration. This was actually found in [3] for the higher harmonics of gearmesh. Unfortunately, the encoders mounted at the time of this first test had a low resonance preventing reliable measurements of TE at speeds higher than 2 Hz (i.e. DTE) and their comparison with the vibration.

3.2 New results from the spur gearbox

As mentioned above, the new measurements were for a different gear ratio, and the gears were hardened and ground, to mitigate against surface distress. Moreover, they were reduced in face-width from 20 mm to 5 mm to reduce the gearmesh stiffness proportionately. The tests are to check the effects of the simulated half tooth-root crack described in section 2. It should be noted that the gearbox test rig is non-ideal (and non-typical) because the shafts are relatively long and slender (to give access inside the casing), but this means that the TE tends to be dominated by shaft deflections rather than tooth defections, making it difficult to detect changes in tooth stiffness, such as result from a crack. The tooth stiffness is at least an order of magnitude greater than the shaft stiffness. Both TE and vibration acceleration were measured over a range of speeds and loads, but speeds of 2 Hz and 20 Hz, and loads from zero (nominal) to 20 Nm are presented here. There was a small friction load corresponding to nominal zero, which was sufficient to keep the gears in contact, and allow measurement of the GTE at low speed.

Figure 3 shows the measured TE, synchronously averaged with respect to the pinion, for loads of 0, 5, 10, 15 and 20 Nm, for four different conditions:

1) Original TSA at 2 Hz
2) Original TSA at 20 Hz
3) Filtered TSA at 2 Hz
4) Filtered TSA at 20 Hz

Two (identical) rotational periods are shown. Bandpass filtering was performed to remove the masking effect of the gearmesh (GM) components and the first two harmonics of the input shaft speed, and so shaft harmonics from the 3rd to the 13th were retained in the TE signals. It was checked that the main effect of the crack was additive rather than multiplicative (modulation of the GM harmonics) so the signals were lowpass filtered just under half the GM frequency to enhance additive impulses from the crack, having components above the first two rotational harmonics, but removing modulation sidebands along with the GM harmonics.

Considering first the unfiltered results at low speed in Fig. 3(a), the increasing load gives a corresponding increase in the gearmesh component, but no change in a shaft speed component, which is likely due to a small eccentricity of the pinion. The TE for zero load could be taken as the GTE for this gear. The increasing GM component with load corresponds to the static deflection component of the STE.

For the equivalent results at 20 Hz, in Fig. 3(b), it is seen that the DTE is substantially different from the STE, at least with respect to the GM component. This can be explained by the fact that the GM frequency (540 Hz) is very close to a resonance of the system. This interpretation is also consistent with the fact that the increased GM component is dominated by the first harmonic, whereas that in Fig. 3(a) has many GM harmonics.

The filtered low speed results in Fig. 3(c) reveal the effect of the crack, at about 50 degrees along the scale, although the effect becomes less evident with increasing load. With this knowledge, it will be seen that the crack can also be detected in the unfiltered signal in 3(a), though only at the lowest load.

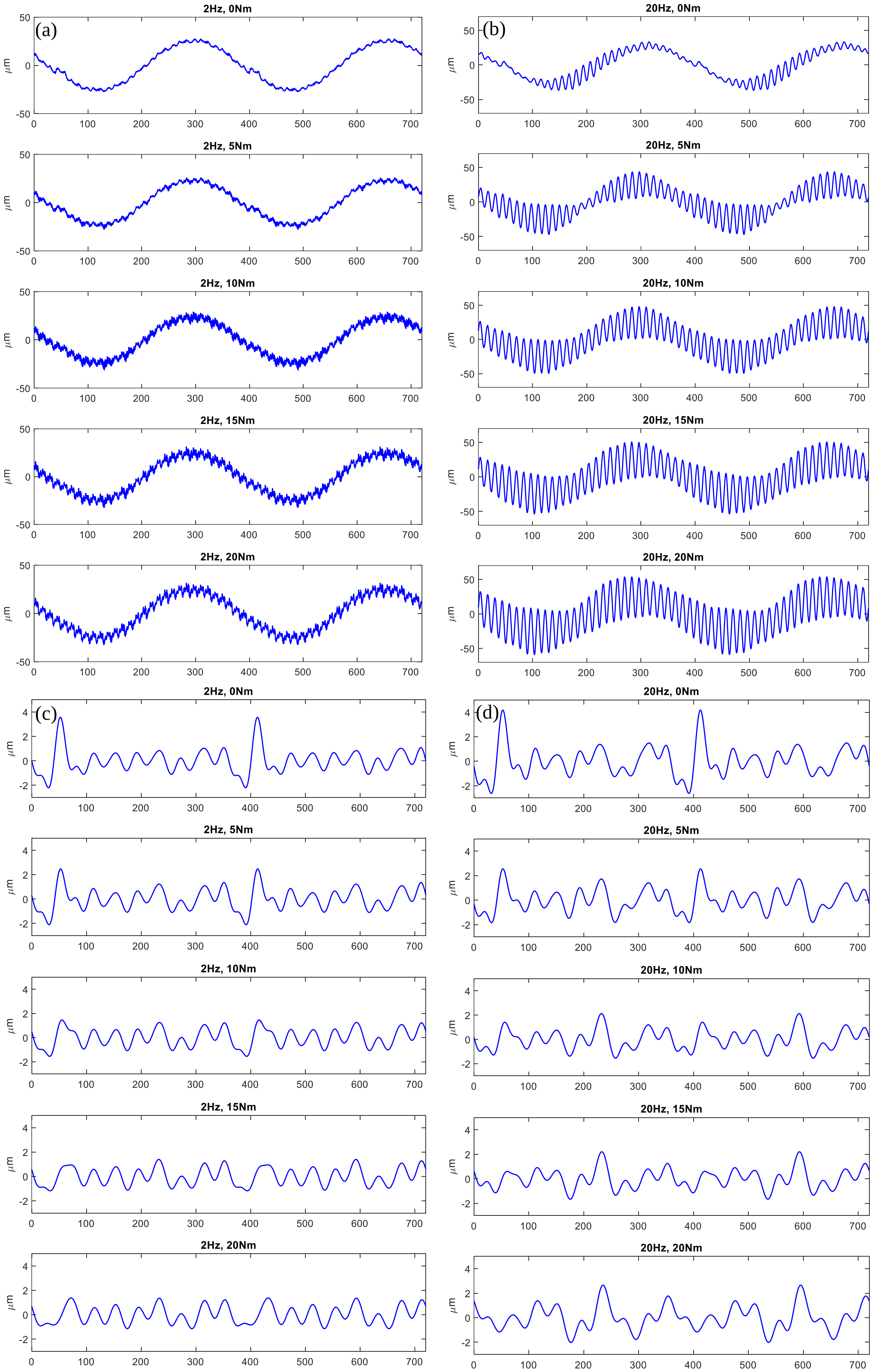


Figure 3: Comparison of original and filtered TE measurements
(a) Original TE, 2 Hz (b) Original TE, 20 Hz (c) Filtered TE, 2 Hz (d) Filtered TE, 20 Hz

The situation is very similar for the high-speed results in Fig. 3(d) (and 3(b)), and it is quite remarkable that once the effect of the resonance on the GM component is removed, the STE of Fig. 3(c) and DTE of 3(d) are very similar, at least for the lowest two loads. This illustrates one of the advantages of TE rather than vibration (including torsional vibration) as a diagnostic parameter, since the effects of operating conditions are greatly reduced.

The unexpected reduction in TE with increase in load gave rise to speculation as to the cause, and it was realised that it must be due to the fact that the “crack” has actually started slightly closed with respect to the undamaged gear, and the effect of increasing load is to counteract this with increasing tooth deflection under load. This is the opposite to what is expected to happen in the case of a genuine natural crack, where it has been demonstrated [5] that there is a tendency for the crack to be permanently open, in the unloaded condition, because of the plastic deformation at the crack tip which is an intrinsic part of crack development. The reason for the “crack” closure in this case is undoubtedly because of relief of residual stresses from heat treatment when the slot was machined, but this should never occur with real crack development, where STE due to loading would be in the same direction as the original GTE.

The change in TE as a result of tooth deflection is not easy to see, even from the filtered results in Fig. 3(c) and (d), but Figure 4(a) and (b) show a zoom of the differential TE in the vicinity of the crack. This represents the difference with respect to the curve at the highest load (20 Nm), but with reversed sign so as to show the increase of deflection with load. This is seen to be monotonic and close to linear. The corresponding linearised compliance can be derived from the deflection vs load curves in Fig. 4(c, d). These differ by only 33%, and indicate that it may be possible to estimate gearmesh stiffness from DTE as well as STE, even where measurements cannot be made at low speed.

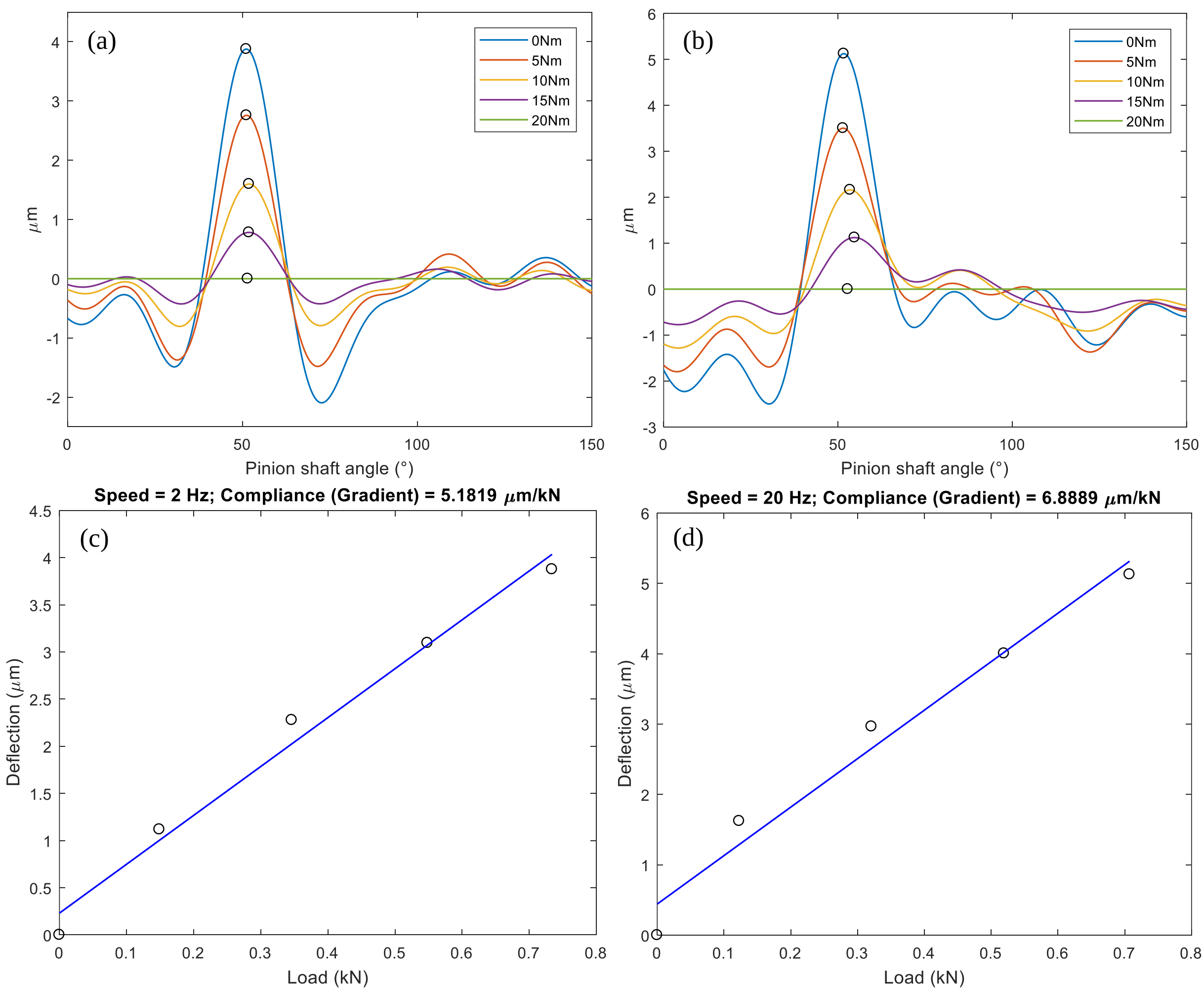


Figure 4: (a, b) Zoom on differential TE in vicinity of crack (c, d) corresponding compliance curves
(a, c) 2 Hz shaft speed (b, d) 20 Hz shaft speed

It is interesting to compare the (differential) compliance values in Fig. 4 with the typical value given for total stiffness by Smith in [6] as “A generally accepted figure for the mesh stiffness of normal teeth is 1.4 $\times 10^{10}$ N/m/m”, which works out in this case to be 7×10^7 N/m, or 14 μm/kN in terms of compliance. This constant value (per unit facewidth) is based only on the bending stiffness component, and is independent of scale for a given shape of tooth since the stiffness varies directly with the cube of the depth, and inversely with the cube of the length. The values in Fig. 4 represent the differential compliance (additional deflection for the same load), which would be 5.18 and 6.89 μm/kN, respectively. In Ref. [7], an estimate is made of the change in stiffness of the toothmesh due to cracks of various sizes, using FEM and an improved simplified method, which agree. For their largest crack, which extends to 48.4% of the tooth thickness, and which has a sharp tip, the increase in compliance is 33% in the single tooth pair zone and 25% in the double tooth pair zone. Considering that the “crack” in the current results has a depth of 50%, and is actually a slot, it is likely the increase in compliance would be greater than those from [7], giving good agreement with the results from Fig. 4.

It is interesting to compare these TE results with those from response accelerations. Figure 5 shows synchronously averaged signals (over two rotation periods) at zero and 20 Nm load, and 2 and 20 Hz input shaft speed. Only the response at highest speed and highest load shows the tooth root crack. Although not shown here, even the responses at 20 Hz and 15 Nm did not show the crack. From Fig. 5(d) it appears that the effect of the crack is mainly multiplicative (local amplitude modulation) so it could be that the resonance near the GM frequency has also amplified the effect of the crack.

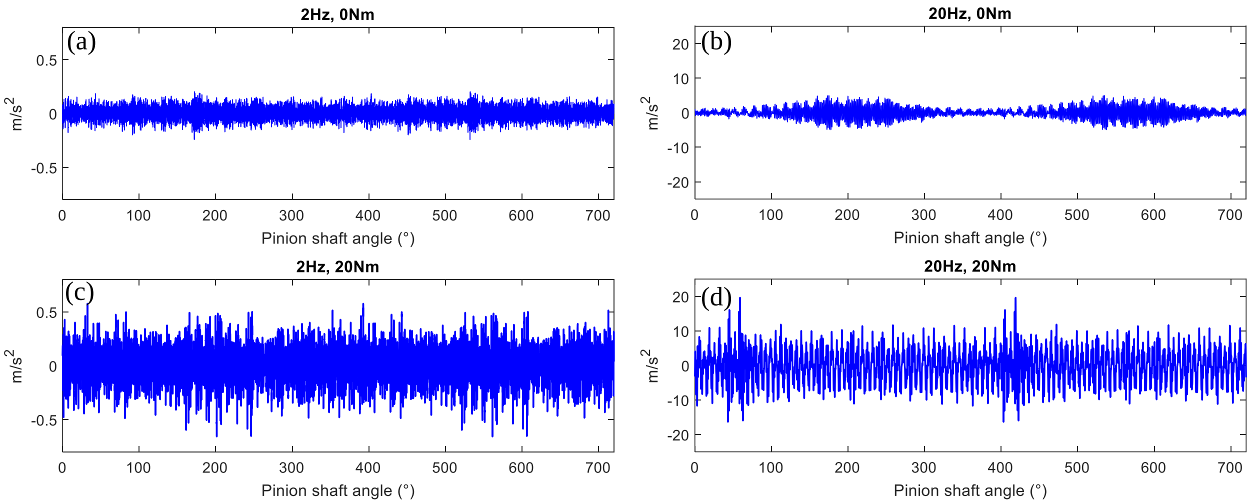


Figure 5: Synchronously averaged acceleration signals for two speeds and two loads
(a, b) 0 Nm (c, d) 20 Nm (a, c) 2 Hz (b, d) 20 Hz

It is quite possible that further signal processing could extract evidence of the crack from more of the response signals, but the main point with respect to this paper is that the TE and vibration responses give quite different information about a tooth root crack, with perhaps the main point being that it only excites a vibration response when teeth are deflected, and therefore not under zero load. The GTE, on the other hand, does show the crack at zero load, in this case because the “slot” had actually closed because of relief of residual stresses. However, in the case of normally developing cracks, they would be partially open because of plastic deformation at the crack tip, and would open further under load, this being detectable by measurement of STE and DTE, the latter at higher speeds, where it would not be possible to measure the GTE.

The fact that information was obtainable, from the measured TE, of toothmesh stiffness, even at higher speed where the GM frequency excited a resonance, emphasises the fact that the TE is measured right at the source, whereas vibration response measurements at different measurement points would all be different, and correspond to different (possibly time-varying) transmission paths.

4 Conclusion

This paper gives a number of examples of how measured gear TE can be useful in gear diagnostics, as an alternative, or supplement, to vibration measurements. It explains how GTE, STE and DTE can be measured if it is possible to run the machine at low speed and low load (GTE), low speed and high load (STE) and high speed and high load (DTE). An earlier paper demonstrated some of the characteristics for generalised distributed wear and pitting of the teeth, giving changes on tooth profiles, whereas the current paper shows a number of advantages, compared with vibration measurement, for the critical case of a tooth root crack. Of particular interest was that it was possible to obtain estimates of the change in toothmesh stiffness (actually compliance) due to the crack, and indirectly of the toothmesh stiffness itself. The latter would probably require comparison with simulations of the cracked tooth, for example with an FE model.

Potential advantages of using TE for gear diagnostics are:

1) The measurement is closer to the source, and less disturbed by transfer function effects than vibration responses, which not only vary considerably between different positions, but can also be time-varying.

2) It is easier to get a good correspondence with simulations, because the torsional parts of simulated systems are simpler, and affected by fewer resonances than lateral vibrations, so model updating should be simpler.

3) The measurement of GTE at different times during the life of a gearbox, as well as giving a more direct measurement of wear, will make possible the inclusion of more accurate versions of this parameter in simulation models, including those giving lateral vibrations as outputs.

The technique does require the mounting of accurate encoders on at least the input and output shafts of the gear transmission, but does not necessarily require them to be mounted on all shafts [3], which can be difficult for internal components. However, the inclusion of such encoders is already implemented in some machines, for operational purposes, and this is likely to increase with the wider implementation of the Internet of Things.

Acknowledgments

The authors would like to thank Mr Haichuan Chang for acquiring the impressions and providing illustrations of the gear tooth surfaces, for the wear tests, at different stages of the wear process.

References

[1] Sweeney, P.J. and Randall, R.B., *Gear Transmission Error Measurement using Phase Demodulation*. Proc. I.Mech.E., Part C, J.Mech.Eng.Sci. 210(C3), 1996, pp. 201-213.

[2] Randall, R.B., Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications. Wiley, Chichester, UK, 2011.

[3] Randall, R.B., Peng, D., Smith, W.A., *Using measured transmission error for diagnostics of gears*, SIRM conference, Copenhagen, February 2019.

[4] Chang, H., Borghesani, P., Smith, W.A., Peng, Z. *Application of surface replication combined with image analysis to investigate wear evolution on gear teeth - A case study*, Wear, 430-431, 2019, pp. 355-368.

[5] Mark, W.D., Reagor, C.P., Static-transmission-error vibratory-excitation contributions from plastically deformed gear teeth caused by tooth bending-fatigue damage, Mechanical Systems and Signal Processing, 21, 2007, pp. 885-905.

[6] Smith, J.D., Gear Noise and Vibration, 2nd Ed., Marcel Dekker Inc., NY, 2003.

[7] Mohammed, O.D., Rantatalo, M., Aidanpää, J.-O., *Improving mesh stiffness calculation of cracked gears for the purpose of vibration-based fault analysis*, Engineering Failure Analysis, 34, 2013. pp. 235–251.

Squared envelope analysis based on the H_∞ filter order tracking: Application for bearing diagnosis

A. Assoumane [1*], D. Abboud [1], E. Sekko [2], M. Elbadaoui [1,3]

[1]Safran Tech, Rue des Jeunes Bois â Châteaufort, 78772 Magny-les-Hameaux, France
[*]amadou.assoumane@safrangroup.com
[2]Univ Orléans, PRISME EA422921, 21 Rue Loigny-La-Bataille, 28000 Chartres, France
[3]Univ Lyon, UJM-St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France

Abstract

The analysis of the squared envelope spectrum (SES) is one among the most used tools for bearing diagnosis. It can easily reveals the characteristic frequencies related to the bearing fault [1, 2]. Actually, the envelope is estimated through a demodulation process in a selected frequency band. The proper choice of the latter is really challenging in a complex environment [3]. In addition to that, the frequency of the bearing fault is likely to be masked by deterministic components. This can jeopardize the efficiency of classical techniques [3, 4, 5]. In this paper, a new approach for bearing diagnostic is proposed. It is based on a recently proposed order tracking technique using the H_∞ filter [7]. In details, the method starts by computing the squared envelope (SE) of the raw signal over the full demodulation band without prior processing. Next, the SE is modeled in a state space using a trigonometric series expansion. Last, an H_∞ estimator is designed to extract the amplitude of each harmonic related to the bearing fault signature. This estimator is well convenient to track the order of bearing faults, particularly in the presence of deterministic components (i.e. the noise). Since this noise is neither white nor Gaussian, the traditional Kalman filter order tracking is compromised [8, 9, 10]. Contrary to the Kalman filter, the H_∞ filter is based on the minimax optimization. The minimax approach leads to the minimization of the estimation error for the worst possible amplification of the noise signal. More interestingly, no prior knowledge about the statistical properties of the noise signals is required [11, 12]. The efficiency of the proposed approach is demonstrated on simulated and real-world vibration signals in nonstationary regimes.

Keywords: H_∞ filter, state space modelling, order tracking, squared envelope, bearing diagnosis, vibration signal, variable speed condition.

1 Introduction

Rolling element bearings are among the most widely used elements in rotating machines. Because of their common role to carry high loads, bearings are likely to be exposed to sudden failures causing system outage. Thus, there has been an increasing interest in developing appropriate techniques for signal denoising and incipient fault detection. Due to their non-invasive nature and their high reactivity to incipient faults, the development of vibration-based techniques has spiked the interest of the scientific community [1]. In this context, envelope analysis has long been recognized as a powerful bearing diagnosis technique. Typically, it consists of a bandpass filtering step in a frequency band wherein the impulsive response is amplified, followed by a demodulation that extracts the signal envelope. The spectrum of the envelope reveals the desired diagnostic information, including the repetition frequency of the fault as well as possible modulations. It has been shown in [2] that it is preferable to use the squared envelope instead of the envelope as the latter is likely to introduce additional interfering components in the envelope spectrum. Since that time, the envelope spectrum was replaced by the squared envelope spectrum (SES) which has become the benchmark technique for bearing diagnostics. A powerful solution to this issue was proposed through the spectral kurtosis [16] (and some derived tools such as the kurtogram [17], the fast kurtogram [3], etc.) which provides an entirely blind way of identifying the best demodulation band according to the impulsivity criterion. Despite its remarkable relevance in machine

signal analysis, the efficiency of the spectral kurtosis is compromised in many situations; for instance, in the presence of energetic deterministic part or the presence of multiple impulsive sources or strongly in nonstationary conditions. This paper comes in this context aiming at providing a new way to address bearing diagnostic based on tracking bearing characteristic orders (i.e. fault frequencies referenced to the shaft frequency) in the squared envelope, without the need of eliminating the deterministic component neither to filter the signal. The method uses the fact that the SE signal comprises a cyclic patterns related to bearing fault. From this observation, the SE signal is described in the state space model using a trigonometric series expansion. Then, an H_∞ filter is designed to track bearing fault order components. This approach is different from the classical Kalman filter based order tracking. The latter is a widespread method used to track sinusoidal components [8, 9, 10], assuming that the exegeneous noises that affect the state model are white and Gaussian with known statistics. In current situations, those assumptions are not valid. Indeed, the meshing components that mask the bearing ones are neither white nor Gaussian. To address this issue, an H_∞ filter is proposed. The latter minimizes the estimation error for the worst possible amplification of the noises. This leads to a minimax optimization where no prior knowledge about the statistical properties of the noises is required [11, 12].
The proposed approach is presented in this paper as follows. In Section 2, the SE of a discrete vibration signal is described in the state space model using a trigonometric series expansion. In Section 3, the methodology to design an H_∞ filter is exposed. In Section 4, the proposed approach is first applied to a simulated vibration signal. Then, it is applied to analyze real-life vibration signals acquired from a wind turbine under nonstationary conditions. Conclusions of this paper are given in Section 5.

2 State space modeling of the squared envelope signal

Condider the discrete measured vibration signal as follows:

$$y[k] = y_r[k] + y_g[k] + b[k] \tag{1}$$

where $y_r[k]$ is the signal related to the bearing vibration, $y_g[k]$ is the meshing signal and $b[k]$ is the signal composed of all the exegeneous vibrations such as the background noise for all $k = 1, \cdots, N$. N is the number of signal samples. The meshing signal, in the case of a tooth crack, exhibits amplitude and phase modulations [6]. The corresponding signal in nonstationary regimes can be written as:

$$y_g[k] = \kappa(\omega[k]) \sum_m a_m[k] e^{j\phi_m[k]} e^{j\theta_m[k]} \tag{2}$$

in which $\kappa(\omega[k])$ is a modulation function depending on the machine regime, $\omega[k] = 2\pi f_r[k]$ is the shaft angular speed and f_r is the machine rotating frequency, a_m and ϕ_m are respectively the amplitude and phase modulations, θ_m is the instantaneous meshing angular displacement and j is the complex number such as $j^2 = -1$. Concerning the bearing vibration signal, it exhibits a series of impulses which can be modelled as [1]:

$$y_r[k] = \kappa(\omega[k]) M[k] \sum_i^d A_i I[k - \lceil T_i f_s \rceil] \tag{3}$$

in which:

- $M[k]$ is the load distribution function for an inner-race under radial load. In stationary conditions, this function is periodic at the shaft rotating period [18];
- A_i is the amplitude of the ith impact so that $A_i = A + \delta A_i$. A is the mean value of the distribution and δA_i is a zero-mean random part with σ_A its standard deviation;
- T_i is the instant of apparition of the ith impact;
- I is the damping response that depends on the damping factor and the resonance frequency of the bearing structure;
- d is the number of impacts resulting from the bearing fault;

- f_s is the sampling frequency;
- $\lceil\cdot\rceil$ stands for the integer part of a decimal number.

Since the bearing's rolling elements are subject to slipping phenomena, the time of occurence from one impact to another is not constant. This time exhibits a random part and, as mentionned in [1], can be modelled in stationary conditions as:

$$T_i = iT + \delta T_i \tag{4}$$

where T is the time isntant between two consecutive impacts and δT_i is a random variable with a Gaussian distribution. This modelling is no longer valid in nonstationary conditions. In this context, Borghesani et *al.* [19] and Abboud et *al.* [20] have written the instant of impact occurrence as follows:

$$T_i = t(i\theta_d + \delta\theta_i) \tag{5}$$

in which θ_d is the angular period of the bearing fault and $\delta\theta_i$ is a zero-mean Gaussian distribution.
The squared envelope of the measured vibration signal, which is of interest in this work and denoted SE, is given as follows:

$$\mathrm{SE}[k] = \mathbb{E}\{y[k]\bar{y}[k]\} \tag{6}$$
$$= \mathbb{E}\{(y_r[k]+y_g[k]+b[k])(\bar{y}_r[k]+\bar{y}_g[k]+\bar{b}[k])\} \tag{7}$$

where $\bar{a}$ is the conjugate of the complex number a and $\mathbb{E}\{\cdot\}$ stands for the expectation symbol.
In this paper, it is assumed that the bearing, the meshing and the noise signals are mutually not correlated. Hence, the squared envelope becomes:

$$\mathrm{SE}[k] = \mathbb{E}\{y_r[k]\bar{y}_r[k]\}+\mathbb{E}\{y_g[k]\bar{y}_g[k]\}+\mathbb{E}\{b[k]\bar{b}[k]\} \tag{8}$$
$$= \mathbb{E}\{y_r[k]\bar{y}_r[k]\}+n[k] \tag{9}$$

where $n[k] = \mathbb{E}\{y_g[k]\bar{y}_g[k]\}+\mathbb{E}\{b[k]\bar{b}[k]\}$ is considered as a noise signal. Otherwise, the SE can be expressed using the autocorrelation function (ACF) denoted by $\mathscr{R}$. The latter, applied to the bearing signal y_r in equation (3), can be written as [20]:

$$\mathscr{R}[k,j] \approx (A^2+\sigma_A^2)\kappa^2(\omega[k])M^2[k]\sum_{i=1}^{d}\mathbb{E}\{g[k-\lceil T_i f_s\rceil, j]\} \tag{10}$$

where $g[k,j] = I[k]I[k-j]$. By writing this function in the angular domain, one gets:

$$\mathscr{R}[k_a,j] \approx (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-\lceil\theta_i N_a\rceil, j]\} \tag{11}$$

with $\tilde{x}$ the angular transformation of the time variable x, k_a the sample index in the angular domain, Na the angular sampling frequency and θ_i the angle instant of the ith impact occurrence. Refering to equation (5), the latter is modelled as $\theta_i = i\theta_d + \delta\theta_i$ [22]. Thus, the ACF becomes:

$$\mathscr{R}[k_a,j] \approx (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-k_{a,i}-\delta k_{a,i}, j]\} \tag{12}$$

where $k_{a,i} \approx \lceil i\theta_d N_a\rceil$ is the angular sample of the ith impact occurrence and $\delta k_{a,i} \approx \lceil\delta\theta_i N_a\rceil$ is a random integer. The above equation of ACF has been proven by Abboud et *al.* in [20]. By taking advantage of this equation, the SE, using the ACF, can be expressed as:

$$\mathrm{SE}[k_a] = \mathscr{R}[k_a, j=0]+\tilde{n}[k_a] \tag{13}$$
$$= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{g}[k_a-k_{a,i}-\delta k_{a,i}, j=0]\}+\tilde{n}[k_a] \tag{14}$$
$$= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\mathbb{E}\{\tilde{h}^2[k_a-k_{a,i}-\delta k_{a,i}]\}+\tilde{n}[k_a] \tag{15}$$

Assume that the random variable $k_a - k_{a,i} - \delta k_{a,i}$ has a probability density function $f[\delta k_{a,i}]$ centered at $k_a - k_{a,i}$ with a constant standard deviation. According to the *law of the unconscious statistician* [15], the above equation is written as:

$$\begin{aligned}
\mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}\sum_{i=1}^{d}\tilde{h}^2[k_a-k_{a,i}-\delta k_{a,i}]f[\delta k_{a,i}]+\tilde{n}[k_a] \qquad (16)\\
&= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}(\tilde{h}^2\circledast f)[k_a-k_{a,i}]+\tilde{n}[k_a] \qquad (17)\\
\mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\tilde{M}^2[k_a]\sum_{i=1}^{d}s[k_a-k_{a,i}]+\tilde{n}[k_a] \qquad (18)
\end{aligned}$$

in which $\circledast$ stands for the convolution symbol and $s[k_a] = (\tilde{h}^2 \circledast f)[k_a]$ is the convolution between the function $\tilde{h}^2$ and f. The function $\tilde{M}^2[k_a]$ is deterministic and can be approximated by a Fourier series such as $\tilde{M}^2[k_a] = \sum_s \lambda_s[k_a]e^{j\psi_s[k_a]}e^{js\theta_r[k_a]}$ where λ_s and ψ_s are respectively the sth variable amplitude and phase of the Fourier series and θ_r is the angular period of the shaft. In same way, the sum in the SE formula can also be expressed by $\sum_{i=1}^{d} s[k_a-k_{a,i}] = \sum_z \rho_z[k_a]e^{j\varphi_z[k_a]}e^{jz\theta_d[k_a]}$ with ρ_z and φ_z respectively the zth variable amplitude and phase of the Fourier series and θ_d the angular period of bearing fault. This leads to:

$$\mathrm{SE}[k_a] = (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\Re\{\sum_{s,z}\lambda_s[k_a]\rho_z[k_a]e^{j(\psi_s[k_a]+\varphi_z[k_a])}e^{j(z\theta_d[k_a]\pm s\theta_r[k_a])}\}+\tilde{n}[k_a] \qquad (19)$$

where $\Re\{.\}$ defines the real part of a complex number. In this paper, all the components related to the shaft angular period are not of interest. Therefore, the SE is written as:

$$\begin{aligned}
\mathrm{SE}[k_a] &= (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\Re\{\sum_{s=0,z}\lambda_0[k_a]\rho_z[k_a]e^{j(\psi_0[k_a]+\varphi_z[k_a])}e^{j(z\theta_d[k_a])}\}+v[k_a] \qquad (20)\\
&\approx \sum_{z=1}^{l}\alpha_z[k_a]\cos(z\theta_d[k_a]+\phi_z[k_a])+v[k_a] \qquad (21)
\end{aligned}$$

in which $\alpha_z[k_a] = (A^2+\sigma_A^2)\tilde{\kappa}^2(\omega[k_a])\lambda_0[k_a]\rho_z[k_a]$, $\phi_z[k_a] = \psi_0[k_a]+\varphi_z[k_a]$ are respectively the zth amplitude and phase of the Fourier series, $v[k_a]$ is the noise comprising the initial noise $\tilde{n}[k_a]$ and all the components related to the shaft angular period $\theta_r[k_a]$ and l is the higher order of the series. The latter defines the number of the bearing component of interest in the estimation procedure. In the state modelling approach that is proposed in this paper, $v[k_a]$ is the so-called *measurement error* or *measurement noise*.

At this stage, the detection of the bearing fault is reduced to the estimation of the amplitude α_z and the phase ϕ_z of the zth order component. This estimation can be done using a linear or a non-linear filtering approach. However, the non-linear approach are subject to a divergence issue. To obtain a linear model, the SE signal is presented in the following form:

$$\mathrm{SE}[k_a] = \sum_{z=1}^{l}\boldsymbol{h}_z^T[k_a]\boldsymbol{x}_z[k_a]+v[k_a] \qquad (22)$$

where:

- $\boldsymbol{h}_z[k_a] = \begin{pmatrix}\cos(z\theta_d[k_a]) & \sin(z\theta_d[k_a])\end{pmatrix}^T \in \mathbb{R}^{2\times 1}$ is the zth measurement vector. $\mathbb{R}$ stands for the ensemble of the real number and $(\cdot)^T$ for the transpose symbol. In the rest of this paper, the lowercase symbols in bold stand for vectors and the uppercase ones in bold stand for matrices;

- $\boldsymbol{x}_z[k_a] = \begin{pmatrix}\alpha_z[k_a]\cos(\phi_z[k_a]) & \alpha_z[k_a]\sin(\phi_z[k_a])\end{pmatrix}^T \in \mathbb{R}^{2\times 1}$ is the zth state variable.

Using equation (22), the estimation of the amplitude and the phase of the Fourier series reduces to the estimation of the state variable $\boldsymbol{x}_z$. In this context, the simultaneous estimation of $\boldsymbol{x}_z$ leads to write the SE as:

$$\mathrm{SE}[k_a] = \boldsymbol{h}^T[k_a]\boldsymbol{x}[k_a]+v[k_a] \qquad (23)$$

in which $\boldsymbol{h}[k_a] = \left(\begin{array}{ccc} \boldsymbol{h}_1^T[k_a] & \cdots & \boldsymbol{h}_l^T[k_a] \end{array} \right)^T \in \mathbb{R}^{2l\times 1}$ and $\boldsymbol{x}[k_a] = \left(\begin{array}{ccc} \boldsymbol{x}_1[k_a] & \cdots & \boldsymbol{x}_l[k_a] \end{array} \right)^T \in \mathbb{R}^{2l\times 1}$ are respectively the *measurement vector* and the *state variable*. Assuming that the angular period of the bearing fault is known, the detection of the latter is reduced to the estimation of the state variable $\boldsymbol{x}[k_a]$. In this context, it is proposed to estimate $\boldsymbol{x}[k_a]$ in a recursive manner using a state space modelling approach.
For this reason, the SE signal is described in a state space. That is to say the dynamic of the state variable has to be defined. In this paper, all parameters include in the state variable $\boldsymbol{x}[k_a]$ are supposed to follow, roughly speaking, a *random walk* so that:

$$\boldsymbol{x}[k_a+1] = \boldsymbol{x}[k_a] + \boldsymbol{w}[k_a] \tag{24}$$

where $\boldsymbol{w}[k_a]$ is a random or deterministic signal with bounded energy. Equations (23) and (24) form the state space model of the *SE* signal. From the latter, an H_∞ filter is designed in the next section for the state variable estimation.

3 H_∞ filter order tracking

Considering equations (23) and (24), the H_∞ filter will be designed to estimate some arbitrary linear combination of the state, say:

$$\hat{s}[k_a] = \boldsymbol{h}^T[k_a]\hat{\boldsymbol{x}}[k_a] \tag{25}$$

where $\hat{x}[k_a]$ satisfies the following recursion:

$$\hat{\boldsymbol{x}}[k_a] = \hat{\boldsymbol{x}}[k_a-1] + \boldsymbol{g}[k_a]\left(\mathrm{SE}[k_a-1] - \boldsymbol{h}^T[k_a-1]\hat{\boldsymbol{x}}[k_a-1]\right) \tag{26}$$

where $\boldsymbol{g}[k_a]$ is the H_∞ gain and $\hat{\boldsymbol{x}}[k_a]$ is the estimate of $\boldsymbol{x}[k_a]$. The state variable is estimated for any $v[k_a]$ and $\boldsymbol{w}[k_a]$ of bounded energy.
Let $e[k_a] = s[k_a] - \hat{s}[k_a]$ be the estimation error, then the H_∞ gain is found by minimizing the following cost function given by [12]:

$$J = \frac{\sum_{k_a=1}^{N} \|e[k_a]\|^2}{\|e[1]\|^2_{\boldsymbol{P}[1]^{-1}} + \sum_{k_a=1}^{N}\left(\|\boldsymbol{w}[k_a]\|^2_{\boldsymbol{Q}^{-1}} + \|v[k_a]\|^2_{R^{-1}}\right)} \tag{27}$$

where $(e[1], \boldsymbol{w}[k_a], v[k_a]) \neq (0,0,0)$, $e[1]$ represents the initial error, $\boldsymbol{P}[1] > 0$, $\boldsymbol{Q} > 0$ and $R > 0$ are positive definite weighting matrices, N is the number of samples and $\|e[k_a]\|_{\boldsymbol{S}} = e[k_a]^T S e[k_a]$. This can be interpreted as the energy gain from the unknown disturbances $\boldsymbol{P}^{-1/2}[1]e[1]$ and $\{\boldsymbol{Q}^{-1/2}\boldsymbol{w}[k_a], R^{-1/2}v[k_a]\}_{k_a=1}^{N}$ to the estimation error $\{e[k_a]\}_{k_a=1}^{N}$. It is quite clear that if the ratio in (27) is small then the estimation is better, and vice versa. However, this ratio depends on the quantities $e[1]$, $\boldsymbol{w}[k_a]$ and $v[k_a]$ which are unknown. In this context, the worst case is considered below:

$$\underset{e[1],\boldsymbol{w}[k_a],v[k_a]}{\mathrm{Sup}} \quad J \leq 1/\gamma \tag{28}$$

where "Sup" stands for the supremum and γ is the performance bound. Otherwise, the goal of the H_∞ problem is to find an estimation $\{\hat{s}[k_a]\}_{k_a=1}^{N}$ that minimizes the worst-case energy. This is equivalent to minimize the following scalar quadratic form:

$$J_f = e^T[1]\boldsymbol{P}^{-1}[1]e[1] + \sum_{k_a=1}^{N} \boldsymbol{w}[k_a]^T\boldsymbol{Q}^{-1}\boldsymbol{w}[k_a] + \sum_{k_a=1}^{N} v[k_a]^T R^{-1} v[k_a] - \gamma \sum_{k_a=1}^{N} e[k_a]^T e[k_a] \tag{29}$$

so that $J_f > 0$ for all vectors $e[1]$, for all nonzero signals $\boldsymbol{w}[k_a]$ and $v[k_a]$ of bounded energy.
Giving the cost function J_f, the worst case minimization is reduces to minimize J_f in respect to $\hat{s}[k_a]$ and to maximize J_f in respect to $e[1]$, $\boldsymbol{w}[k_a]$ and $v[k_a]$ for all $k_a = 1,\cdots,N$. This leads to a minmax optimization formulated in such a way that:

$$\{\hat{s}[k_a]\}_{k_a=1}^{N} = \arg\left(\min_{\hat{s}} \max_{e[1],\boldsymbol{w},v} (J_f)\right) \tag{30}$$

This optimization problem can be solved by the well known Lagrange mutliplier approach. The solution to the above optimization is given by the theorem quoted below [11, 12, 14].

Theorem 1 *Let $\gamma > 0$ be the user-specified performance bound. Then, there exists an H_∞ estimation for $s[k_a]$ if and only if there exists a symmetric positive definite matrix $\boldsymbol{P}[k_a] \in \mathbb{R}^{2l\times 2l}$ that satisfies the following discrete-time Riccati equation:*

$$\boldsymbol{P}[k_a] = \boldsymbol{P}[k_a - 1]\boldsymbol{\Gamma}[k_a - 1] + \boldsymbol{Q} \tag{31}$$

when

$$\boldsymbol{\Gamma}[k_a] = \left(\boldsymbol{I}_{2l} - \gamma \boldsymbol{h}^T[k_a]\boldsymbol{h}[k_a]\boldsymbol{P}[k_a] + \boldsymbol{h}[k_a]R^{-1}\boldsymbol{h}^T[k_a]\boldsymbol{P}[k_a]\right)^{-1} \tag{32}$$

and $\boldsymbol{I}_{2l} \in \mathbb{R}^{2l\times 2l}$ is the identity matrix.
Then, the H_∞ gain $\boldsymbol{g}[k_a] \in \mathbb{R}^{2l\times 1}$ is given by:

$$\boldsymbol{g}[k_a] = \boldsymbol{P}[k_a]\boldsymbol{\Gamma}[k_a - 1]\boldsymbol{h}[k_a]R^{-1} \tag{33}$$

It should be noted that for some weighting matrices $\boldsymbol{P}[1]$, $\boldsymbol{Q}$ and R the performance criterion in (28) is achieved if and only if the performance bound γ satisfies the following inequality:

$$\gamma < R^{-1} \tag{34}$$

Since γ defines the noise level attenuation or the performance bound, it should be as high as possible. And, it has been shown in a previous paper [7] that when γ is greater than its optimal value, the matrix $\boldsymbol{P}$ is not symmetric positive definite. Otherwise, when γ tends to zero, the H_∞ filter is not constrainted. Then, it is equivalent to the standard Kalman filter for which R and $\boldsymbol{Q}$ are defined respectively as the covariance matrix of the measurement noise and the state noise.

4 Application

4.1 Synthetic signal analysis

Here, a synthetic signal is presented to evaluate the performance of the proposed approach in estimating bearing order components. The signal is composed of the bearing, the meshing and the noise signal. They represent respectively 20%, 30% and 50% of the synthetic signal energy. The non-stationary condition is simulated using a non-linear rotating frequency varying between 5 Hz and 30 Hz such as $f_r[k] = 5 + 25\sin\left(\frac{\pi(k-1)}{2(N-1)}\right)$ for $1 \leq k \leq N$ where N is the number of samples. The time duration of the signal is 5 s.
Concerning the meshing signal, sampled at the frequency $f_s = 10$ kHz, it is computed using equation (2) and contains five meshing components. The latter is composed of the 45th and 49th shaft order component with amplitude and phase modulations. The generation of these modulations is presented in details in Appendix A. About the bearing signal, it is generated using equation (3) in which the resonance frequency and the damping factor are respectively equal to 4 kHz and 2000.

The rolling bearing considered is subject to a local defect occurring on the outer-race. Its characteristic order, ball-pass-order on the outer-race denoted BPOO, is equal 8.7 times the rotating frequency with a slight random variation. Last, the additive noise is generated using a white Gaussian noise modulated by the rotating frequency.
The different contributions of the signal are displayed in Figure 1. It can be seen in the SES of the raw signal (Figure 1 (g)) that the bearing fault order is masked by deterministic components. Thus, the proposed approach is applied to track the order component of the bearing fault. The number of order harmonic of interest l is equal to 15. Also, the parameters of the filter take the following values: $R = 1$, $\boldsymbol{Q} = 0 \times \boldsymbol{I}_{2l}$, $\gamma = 0.9R^{-1}$ and the filter initial values are $\boldsymbol{P}[1] = \boldsymbol{I}_{2l}$ and $\boldsymbol{x}[1] = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{2l\times 1}$. The choice of $\boldsymbol{Q}$ is motivated by the fact the coefficients of the trigonometric series expansion are assumed to be constant. This means that the state noise $\boldsymbol{w}[k]$ is null for all $k = 1, \cdots, N$.
The result of the estimation provided by the H_∞ estimator for the choosen parameters is presented on Figure 2. There, the number of peaks estimated by the proposed approach correponds to the expected ones, i.e 15 peaks. Moreover, the deterministic components present in the SES of the raw signal have been greatly attenuated. Besides of that, some peaks, with lower energy level, appear around the bearing order components. This is due to the fact that the H_∞ filter, like all type of filter, don't atteanuate uniformely all the frequencies outside the

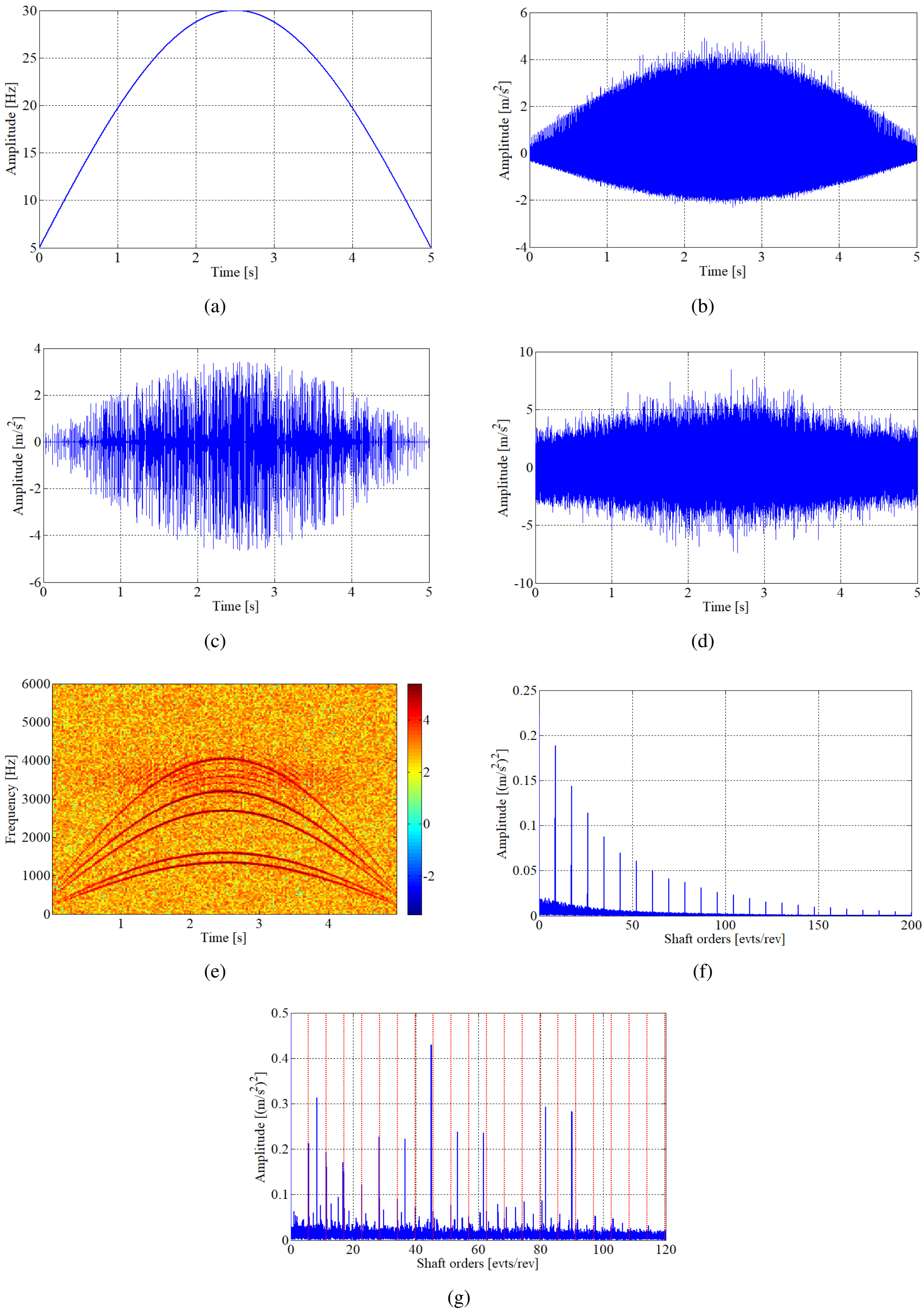


Figure 1 – Synthetic vibration signal: (a) Rotating frequency, (b) Meshing signal, (c) Bearing signal, (d) Complete vibration signal, (e) Time-frequency representation, (f) SES of the bearing signal and (g) SES of the complete signal.

band of interest.
To evaluate the performance of the proposed approach, the signal-to-noise ratio is calculated by the following formula $\text{snr} = 10 \times \log_{10}\left(\frac{\sum_{k=1}^{N} s^2[k]}{\sum_{k=1}^{N}(\hat{s}[k]-s[k])^2}\right)$ where s is the SE related only to the bearing signal and $\hat{s}$ is its estimate. When the bearing signal is totally influenced by the meshing and the noise signal, that is the worst case estimation, the estimation error is $\hat{s}[k] - s[k] = b[k]$. In this case, the snr is equal to -6 dB and defines the lower limit of the performance bound. On this basis, it can be stated that all estimations provided by the proposed approach should have a snr greater than -6 dB. It follows that a higher snr leads to a better estimation. For this first simulation, after a 200 Monte-Carlo simulations, the estimation error leads to a snr equal to 1.48 dB. This value is greater than the performance bound and corroborates the quality of the estimation displayed on Figure 2.

Influence of the fault frequency incertitude

In the results presented above the frequency (or order) of the bearing was exactly known. In real situations, this frequency is known with uncertainty. This can be due (i) to the fluctuation of the shaft rotating frequency, (ii) to the imperfection of the speed sensor or (iii) to the sliping phenomena of the bearing rolling's elements. In this section, the influence of the uncertainty of the bearing fault order on the estimation quality is investigated. Thus, the snr is evaluated for different values of uncertainty on the bearing fault order. This uncertainty varies from 0 % to 5 % of the real value of the order of the bearing fault. The snr obtained in the range of the uncertainty is displayed on Figure 3. There, the snr remains constant when the uncertainty varies from 0% to 2.5%. This is interesting for the bearing health monitoring since the uncertainty on a potential fault frequency can reach 2% in real situations as mentionned by Randall and Antoni [1]. Beyond this value, the snr decreases; thus the quality of the estimation is degraded and the proposed approach is no longer robust to track the bearing order components.

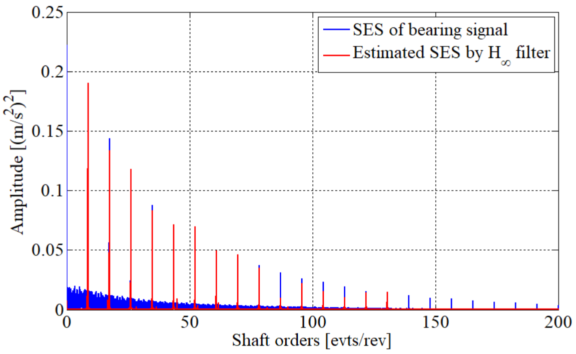


Figure 2 – Estimation of the SES of the bearing signal provided by the H_∞ filter. Here, the uncertainty on the bearing fault order is equal to zero.

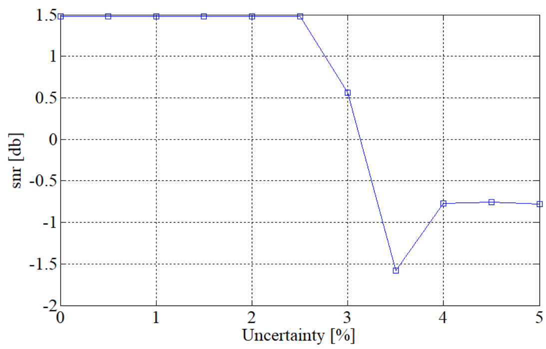


Figure 3 – Signal-to-noise ratio (snr) evolution against the uncertainty on the bearing fault order.

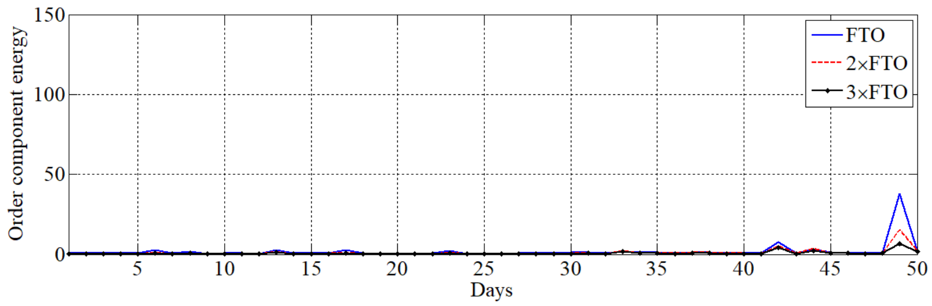
4.2 Experimental signal analysis

This subsection deals with signals acquired from a 2 megawatts wind turbine high speed shaft on which a condition monitoring system is installed. The bearing has an inner race fault which is increasing in severity across the 50-day period. At the end of the test, the bearing was inspected and a crack has been identified in the inner race. Acceleration signals were recorded on a daily basis (one signal per day) together with tachometer signals, over a 6 s duration each with a sampling frequency equal to 97656 Hz. The nominal speed of the bearing shaft is 1800 rpm (30 Hz). Note that the speed variability has reached 15 % of the nominal speed in some records; the reason why the regime is considered nonstationary. The theoretical fault frequencies referenced to the shaft frequency are as follows:

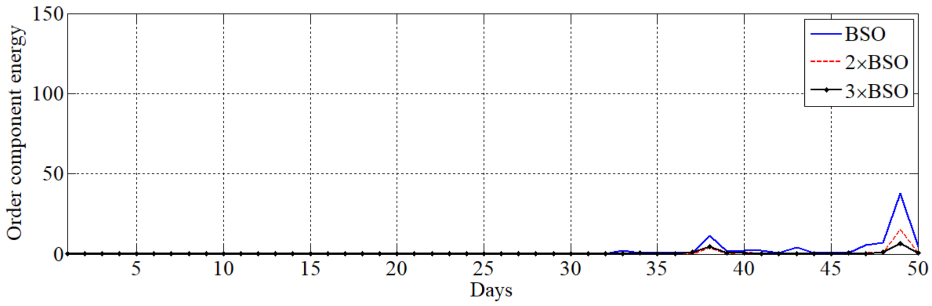
- Ball pass order on outer-race: BPOO = 6.72;
- Ball pass order on inner-race: BPOI = 9.47;

- Ball spin order: BSO = 1.435;
- Fundamental train order: FTO = 0.42.

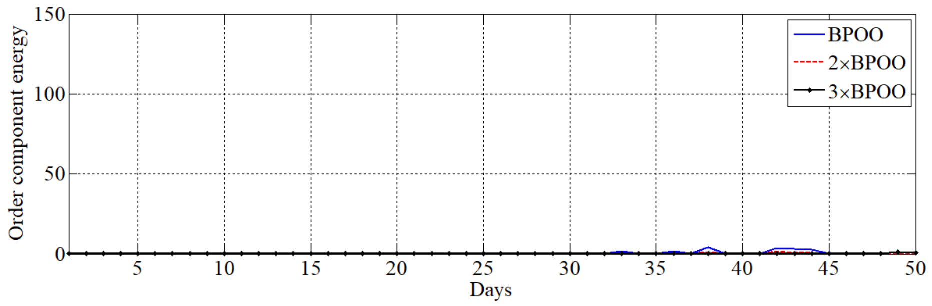
More information can be found in [21]. In this section, the proposed approch is applied to track the order components locolated at $i \times$ BPOO, $i \times$ BPOI, $i \times$ BSO and $i \times$ FTO where $i = 1, 2, 3$. The parameters of the H_∞ filter take the following values: $R = 1$, $\boldsymbol{Q} = 0 \times \boldsymbol{I}_3$ and $\gamma = 0.95$ and the filter initial values are $\boldsymbol{P}[1] = \boldsymbol{I}_3$ and $\boldsymbol{x}[1] = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{6\times 1}$. To monitor the health state of each component of the bearing during the 50-day periods, the energy of each order component is evaluated. Figure 4 shows the evolution of this energy during



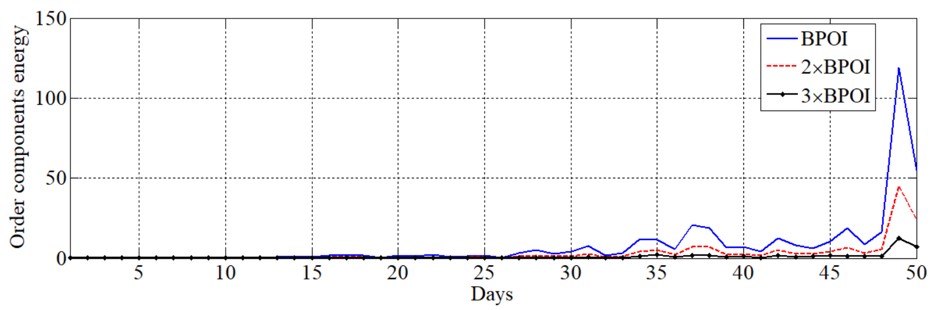
(a) FTO



(b) BSO



(c) BPOO



(d) BPOI

Figure 4 – Evolution of the order component energy for the different type of fault that can occur on the bearing over the 50 days of measurement: (a) cage fault, (b) Rolling elements fault, (c) Outer-race fault and (d) Inner-race fault.

the 50 days. Concerning the order components located at $i \times$ FTO and $i \times$ BSO with $i = 1,2,3$, their energy remains constant till the 49th day where an amplification is observed. This may be related to a degradation of the bearing train and rolling elements at the end of the test. About the BPOO components, their energy is almost constant and negligible compare to that of the FTO and BSO. Therefore, the bearing outer-race is healthy. Otherwise, the energy of the order components located at $i \times$ BPOI increases along the days. This amplification is related to the evolution of the inner-race fault severity. A significant jump in the energy can be seen from the 30th day. According to the inspection done after the 50th day, a crack in the inner-race has been noticed at the 50th day. Thus, the proposed order tracking approach is efficient to detect earlier a bearing fault. Based on the bearing order component estimation provided by the H_∞ filter, different indicators can be designed to monitor bearing health state.

5 Conclusion

In this paper, an order tracking technique was proposed to diagnose a bearing fault under a nonstationary condition. The proposed method consists of estimating a certain number of bearing order components whitout removing the deterministic components. The method described the squared envelope signal in the state space model using a trigonometric series expansion. Then, an H_∞ filter is designed to track bearing fault order components. Firstly, the theoretical foundation of the proposed approach has been described in details. Secondly, a synthetic signal has been generated to evaluate the performance of the proposed approach. It has been shown that the approach was able to track the bearing order components without removing the deterministic components. Moreover, the performance of the proposed approach remains stable for an uncertainty error on bearing orders less than 2.5%. Finally, the efficiency of the proposed approach has been demonstrated with wind turbine vibration signals under a nonstationary condition. The order components related to the bearing fault has been estimated by the proposed approach throughout 50-days of measurement. A fault on the bearing inner-race has been successfully detected earlier at the 30th day. In terms of perspective for this research, the authors will work on the design of a robust H_∞ filter to deal with a large uncertainty on the order of the bearing fault.

Acknowledgements

Acknowledgement is made for the measurements used in this work provided through data-acoustics.com Database.

A Synthetic meshing signal generation

The meshing signal is composed of five components and presented as follows:

$$y_g[k] = \kappa(\omega[k]) \sum_{m=1}^{5} y_{g,m}[k] \tag{35}$$

where $y_{g,m}$ is the mth meshing component. The latter is defined by the expression below:

$$y_{g,m}[k] = A_m[k](1 + a_m[k])e^{o_m \theta[k]} \tag{36}$$

in which:

- a_m is the amplitude modulation of the mth meshing component;
- A_m is the amplitude of the mth meshing component so that it is a random value comprises between 0 and 1 and $\kappa(\omega[k]) = \omega^2[k]/\max(\omega[:])$ is the modulation correponding to the variation of the regime in which ω is the angular speed of the machine;
- $\theta[k] = 2\pi t_s \sum_{c=1}^{k} f_r[c]$ is the instantaneous angular displacement of the shaft rotating at frequency f_r. t_s is the sampling period;

- o_m corresponds to the mth order of the meshing signal. It takes respectively the following values:45, 2×45, 3×45, 1.089×49 and $2\times1.089\times49$.

It is well known that when a fault appears on a gear tooth, the resonance frequency of the gear structure is excited by an impulse [6]. The generated impulse signal modify the shape of the amplitude modulation so that the latter can be described by a series of impulsive. Each impulse is modeled by a narrow bandwidth gausian function. This function is defined by as below:

$$ext[k] = \sum_i e^{-\frac{1}{2}\left(\frac{(k-1)t_s-\mu_i}{\sigma^2}\right)^2} \tag{37}$$

where μ_i is the center of the Gaussian function determined by the time instant for which the impulse occurs on the gear tooth and σ defines the width of the Gaussian function. In this simulation $\sigma = 10^{-3}$. Since the gear attenuates the impulse generated by the fault, a transfert function is included in the model so that the amplitude modulation becomes:

$$a_m[k] = \rho_m \times s[k] \circledast ext[k] \tag{38}$$

in which $s[k] = e^{-\xi^{(g)}(k-1)t_s}\sin\left(2\pi f_{res}^{(g)}(k-1)t_s\right)$ is the transfert function of the gear structure and ρ_m is a random value comprises between 0 and 1. $\xi^{(g)}$ and $f_{res}^{(g)}$ are respectively the damping factor and the resonance frequency of the gear structure. They are respectively equal to 5000 and 4000 Hz.

References

[1] Randall, R. B., and Antoni, J. Rolling element bearing diagnostics-a tutorial. *Mech. Syst. Signal Process.*, 25(2), 485-520, 2011.

[2] Ho, D., & Randall, R. B. (2000). Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals. *Mech. Syst. Signal Process.*, 14(5), 763-788.

[3] Antoni, J. (2007). Fast computation of the kurtogram for the detection of transient faults. *Mech. Syst. Signal Process.*, 21(1), 108-124.

[4] Wang, D., Peter, W. T., & Tsui, K. L. (2013). An enhanced Kurtogram method for fault diagnosis of rolling element bearings. *Mech. Syst. Signal Process.*, 35(1-2), 176-199.

[5] Sawalhi, N., Randall, R. B., & Endo, H. (2007). The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mech. Syst. Signal Process.*, 21(6), 2616-2633.

[6] McFadden, P. D. (1986). Detecting fatigue cracks in gears by amplitude and phase demodulation of the meshing vibration. *Journal of vibration, acoustics, stress, and reliability in design*, 108(2), 165-170.

[7] Assoumane, A., Sekko, E., Capdessus, C., & Ravier, P. (2017). Order tracking using H_∞ estimator and polynomial approximation. *Mechanics & Industry*, 18(8), 808.

[8] Pan, M. C., & Lin, Y. F. (2006). Further exploration of Vold-Kalman-filtering order tracking with shaft-speed information-I: Theoretical part, numerical implementation and parameter investigations. *Mech. Syst. Signal Process.*, 20(5), 1134-1154.

[9] Pan, M. C., and Wu, C. X. Adaptive Vold-Kalman filtering order tracking. *Mech. Syst. Signal Process.*, 21(8), 2957-2969, 2007.

[10] Pan, M. C., Chu, W. C., & Le, D. D. (2016). Adaptive angular-velocity Vold-Kalman filter order tracking-Theoretical basis, numerical implementation and parameter investigation. *Mech. Syst. Signal Process.*, 81, 148-161.

[11] Shen, X. (1995, May). Discrete H_∞ filter design with application to speech enhancement. In *Acoustics, Speech, and Signal Processing*, 1995. *ICASSP*-95., 1995 *International Conference on* (Vol. 2, pp. 1504-1507). IEEE.

[12] Shen, X. M., & Deng, L. (1997). Game theory approach to discrete H/sub/spl infin//filter design. *IEEE Trans. Signal Process*, 45(4), 1092-1095.

[13] Pinkus, A. (2000). Weierstrass and approximation theory. *Journal of Approximation Theory*, 107(1), 1-66.

[14] Banavar, R. N., & Speyer, J. L. (1991, June). A linear-quadratic game approach to estimation and smoothing. In 1991 *American control conference* (pp. 2818-2822). IEEE.

[15] Ross, S. M. (2013). Introduction to stoachastic process. In *Applied probability models with optimization applications* (p. 2). Courier Corporation.

[16] Antoni, J., & Randall, R. B. (2006). The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines. *Mech. Syst. Signal Process.*, 20(2), 308-331.

[17] Antoni, J. (2006). The spectral kurtosis: a useful tool for characterising non-stationary signals. *Mech. Syst. Signal Process.*, 20(2), 282-307.

[18] McFadden, P. D., & Smith, J. D. (1984). Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of sound and vibration*, 96(1), 69-82.

[19] Borghesani, P., Ricci, R., Chatterton, S., & Pennacchi, P. (2013). A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions. *Mech. Syst. Signal Process.*, 38(1), 23-35.

[20] Abboud, D., Antoni, J., Eltabach, M., & Sieg-Zieba, S. (2015). Angle $\setminus$ time cyclostationarity for the analysis of rolling element bearing vibrations. *Measurement*, 75, 29-39.

[21] Bechhoefer, E., Van Hecke, B., & He, D. (2013, October). Processing for improved spectral analysis. In *Annual Conference of the Prognostics and Health Management Society, New Orleans, LA, Oct* (pp. 14-17).

[22] Antoni, J., Bonnardot, F., Raad, A., El Badaoui, M. (2004). Cyclostationary modelling of rotating machine vibration signals. Mechanical systems and signal processing, 18(6), 1285-1314.

Dynamic Characterization of Hydroelectric Turbine with Transient Data Records Using OBMA and Phase-Shift Analysis

Quentin DOLLON[1], Antoine TAHAN[2], Jérôme ANTONI[3], Martin GAGNON[4], Christine MONETTE [5]

[1]ÉTS, Montréal, INSA Lyon
quentin.dollon.1@ens.etsmtl.ca
[2] ÉTS, Montréal, Dpt. Mécanique, H3C 1K3, Montreal, Qc., Canada
antoine.tahan@etsmtl.ca
[3]Univ Lyon, INSA-Lyon, Lab. Vibrations Acoustique, F69621 Villeurbanne, France
jerome.antoni@insa-lyon.fr
[4]Institut de Recherche de HydroQuébec (IREQ), J3X 1S1, Varennes, Qc., Canada
gagnon.martin11@ireq.ca
[5]Andritz Hydro Ltd., H9R 1B9, Pointe-Claire, Qc., Canada
christine.monette@andritz.com

Abstract

The purpose of this paper is to investigate the possibility of estimating Francis hydroelectric turbine modal parameters in transient conditions by focusing on resonance regions generated by the interaction of a structural mode with a frequency-variant harmonic pressure pulsation. Especially when numerous modes are in the same bandwidth, this method separates them by exciting only matching mode shapes. To extract a specified harmonic from the signal, the resonance retrieval is done using Order Tracking method. A classical ambient modal identification algorithm is then used to feature the isolated mode. Furthermore, using the phase-shift between measured locations, modes can be localized and shape determined.

List of Symbols

i	Complex unit
$diag[\boldsymbol{A}]$	Diagonal matrix of vector $\boldsymbol{A}$
$<\boldsymbol{A}\vert\boldsymbol{B}>$	Complex inner product of $\boldsymbol{A},\boldsymbol{B}$
$\boldsymbol{A}^T$	Transpose of $\boldsymbol{A}$
$\boldsymbol{A}^*$	Conjugate of $\boldsymbol{A}$
$\vert\boldsymbol{A}\vert$	Determinant of $\boldsymbol{A}$
δ_r	Kronecker Symbol
N_S	Number of sensors
N_f	Number of frequency samples
$\boldsymbol{\Phi}$	$(N_S \times n)$ Global modal matrix for a system of n modes
MAC	Modal Assurance Criterion
MAC_{thr}	Threshold MAC in E-FDD
$\boldsymbol{\theta}_r$	$(4+N_S \times 1)$ Parameter vector of mode r
$\boldsymbol{\theta}$	$(4+N_S \times 1)$ Modal parameter vector as a variable
$\mathscr{L}(\boldsymbol{\theta})$	Negative Log-Likelyhood Function
ω_k	Discrete pulsation $k \in [\![1, N_f]\!]$
$\widehat{\boldsymbol{X}}_k$	$(N_S \times 1)$ Estimated frequency response in ω_k ($\mathbb{C}$)
$\boldsymbol{X}_k$	$(N_S \times 1)$ Theoretical frequency response in ω_k ($\mathbb{C}$)
$h_{r,k}$	Modal transfer function of mode r in ω_k ($\mathbb{C}$)
p_k	Scaled FFT of modal Excitation in ω_k ($\mathbb{C}$)
$\boldsymbol{\varepsilon}_k$	$(N_S \times 1)$ Scaled FFT of channel noise in ω_k ($\mathbb{C}$)
$\boldsymbol{E}_k$	$(N_S \times N_S)$ Theoretical density matrix in ω_k ($\mathbb{C}$)
$s_{1,k}$	First singular value of E_k
$\boldsymbol{\varphi}_r$	$(N_S \times 1)$ Theoretical mode shape of mode r ($\mathbb{C}$ in FDD)
$\widehat{\boldsymbol{\varphi}_r}$	$(N_S \times 1)$ Estimated mode shape of mode r
β_r	Characteristic real scalar value of mode r
ω_r	Natural pulsation of mode r
ξ_r	Damping Ratio of mode r
S_r	Modal force of mode r
S_{er}	PSD Error for mode r

1 Introduction

Design and exploitation of hydroelectric turbines relies on the knowledge of their dynamic behavior. This enables one to generate and validate models to either get a good assessment of life duration or plan predictive-based maintenance. Two sources of information are useful to properly characterize the mechanical behavior of a structure: numerical simulations and experimental data processing. Giving high computing power, the first source could give a whole and detailed analysis of the behavior in any expected regime through Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) but needs to be validated by the second to be reliable. It is a straightforward consequence of the strong assumptions made to reduce computational burden and model the highly turbulent characteristics of the flow. The second approach relies on in-situ measurements to extract dynamic features [1].

The increase of computational power over years allows getting more accurate simulations for startup regimes [2, 3], no-load or part-load configurations [4, 5, 6] and even hydrodynamic damping estimations [7]. However, the results still show discrepancies in structural parameters due to deviations from real operating conditions: rotating machinery [8], fluid-structure interaction added mass, damping and stiffness [9, 10], cavitation influence [11, 12, 13] or boundary condition sensitivity [14]. On the other hand, experimental characterization is highly fragmented, but in general closer to reality for a given measured operating condition. The features obtained from experimental data relie on statistical models [15, 16], indirect measurements [17], time-frequency analysis [18], but can also be obtained by modal parameter identification using Operational Modal Analysis (OMA) [19] or Experimental Modal Analysis (EMA) [20]. In addition, the experimental hydraulic instability study can be used to compare different computational turbulence models [21, 22, 23]. Typically, the two sources of information (simulations and experiments) are crossed to obtain a hybrid representation of the dynamic behavior, which is used to obtain accurate load levels and allow a better prediction of fatigue [24]. Those predictions are used to assess the runner life duration and reliability of the capacity [25, 26, 27].

One of the problems with experimental analysis is the cost of data acquisition. To reduce financial burden of measurements, the idea is to extract a maximum of information from transient records instead of several stationary records, which would make the measurement less time-consuming. Furthermore, the processing of transient records allows obtaining real structural parameters of highly damaging regimes [23, 28] (what numerical analysis still struggles to perform, as aforementioned). Our goal is to determine whether a signal processing methodology is able to extract precise and suitable features from these transient measurements. For this, a combined methodology using Order Tracking, OMA and Phase-Shift Analysis is implemented and performed on a case study. The case study data come from a medium-head Francis Turbine in Quebec (Canada). The paper first presents the theoretical background, including literature and the different OMA tools to be used. Afterwards, the model is tested on the case study of an operational runner prototype.

2 Resonance Detection Using Phase-Shift Diagrams

Resonances are usually found with the study of experimental correlograms where the amplified regions are treated as Operating Deflecting Shape (ODS). But there is another alternative to detect resonances with more confidence: Phase-Shift Analysis (PSA) [29, 30]. Resonance amplitudes are time-dependent and phases are relative to a reference in experimental data, but the modal phase-shift from one sensor to another is a theoretical time invariant absolute quantity that is specific to each mode. Especially, when a harmonic (time-variant pulsation) and a mode (almost constant pulsation) intersect with the same phase-shift, the observed ODS is very likely the resonance of the only excited mode. This resonance can be extracted and processed with OBMA through a Single Degree of Freedom (SDoF) formulation (Section 3 & 4). Once the mode has been detected, it is possible to feature its shape: the mode shape is assumed to be the nodal diameter that fits the best the modal phase-shift (e.g. [18] in which a self-excited vibration of a hydroelectric runner is studied during load rejection). It is also possible to determine the shape by identifying the pattern of the exciting harmonic, particularly if this last comes from a well-known phenomenon (vortex rope [27, 31, 32, 33, 34], Rotor-Stator Interactions (RSI) [26, 27, 31, 35, 36]).

3 Order Tracking Procedure

Once PSA and resonance mapping is achieved, it is still required to extract accurate damping ratios and frequencies, and eventually other modal properties (modal force *etc.*). In order to do this, identification algorithms are implemented to process multi-channel resonance signals. The first pre-processing step is to extract the resonance component and isolate it from the rest of the signal. This is the purpose of Order Tracking. This class of method gathers all the tools able to extract one harmonic from the signal by shifting the time domain to a harmonic one, called order domain. Orders, measured in times per revolution, are analog to frequencies. Order Tracking is a classical and very used diagnosis tool for rotating machineries. There are four main techniques that are commonly used: direct method using Fourier Transform of a time series (FS), Angular Resampled-based Order Tracking (AD), Time-Variant Discrete Fourier Transforms (TVDFT) and Vold-Kalman filters (VK).

FS extracts the $n-th$ harmonic from a signal by tracking the $\widehat{\boldsymbol{X}}[n\omega_{0,k}]$ response with a short-time Fourier transform at each time step, where $\omega_{0,k}$ is the runner angular velocity at time step k. This procedure is highly biased due to tapering, leakage effects and bandwidth control. In the classical approach with constant time intervals, low rotating frequencies are less accurate than higher ones. If time intervals are non-constant, some power spectral density rescaling issues rise and must be taken into account.

Another technique relies on an adaptative Fourier transform with settable kernel [37]: the kernel of the analytic exponential function tracks the frequency of interest. Consequently, a precise targeted order is extracted. In early works, the kernel orthogonality was lost and a compensation matrix had to be introduced to partly fix the problem. This issue is now easily fixed by introducing a change of variable in the integration domain, and gives a Velocity Synchronous Fourier Transform [38]. Vold-Kalman (VK) Bank Filters can extract orders from a signal with an instantaneous analysis instead of an averaging procedure [39]. Consequently, VK filtering is the most accurate technique in terms of resolution but entails a heavy computational burden, that is irreconcilable with industrial applications.

AD is a resampled-based method that avoids any leakage effect and phase issues [40]. The asynchronous time series are turned into synchronous time series (constant $\Delta\alpha$ instead of Δt) by the means of interpolation and tachometer record (Computed Order Tracking [41]). Then, a short-time Fourier transform is performed on the resampled signal, with intervals corresponding to one runner revolution (so that the spectrum resolution coincides with orders). Intervals are neither overlapped nor windowed. An order spectrum is obtained for each studied revolution. Each of those revolutions is converted into frequency by averaging the rotational speed over the lap. It can be noticed that the lower the studied dynamic, the weaker the quasi-static assumption over a revolution, the higher the response estimation quality. The bias of AD comes from both interpolation and synchronous interval split. Interpolation bias is due to interpolating method (*e.g.* linear, quadratic, splines) and shaft torsion that induces tachometer signal fluctuations. The issue with synchronous interval split is that each interval must represent exactly one revolution, that is not necessarily the case. Most of those biases can be reduced if data are recorded with a extensively high sampling frequency compared to the structure natural frequencies. For the purpose of this paper, the classical COT-based AD will be used, because the data sampling frequency is far higher than the natural frequencies of studied modes.

4 Operational Modal Analysis

Few has been done in the field of OMA for hydroelectric runner dynamic featuring. Gagnon et *al.* used this technique to characterize guide-vane behavior for different operating conditions [19]. The same point is made for EMA for which the study achieved on a runner obtained results that were in well agreement with simulations, but for experimental setting not representative of actual operating conditions [20]. Moreover, in many cases EMA cannot be implemented and when it is possible, suffers from major drawbacks like experimental set-up cost or structural size and complexity. The point of OMA is to extract modal parameters from output-only measurements containing both unknown excitation and response of the system. When those signals are extracted with Order Tracking, the procedure is called OBMA (Order Based Modal Analysis) [40, 42, 43, 44]. OMA is of interest for several reasons: it is fast in terms of computing effort and measurement (mere sensors replace excitation set-up), ambient excitation is appropriate to linearize the dynamic behavior and so on.

OMA techniques are divided into different classes: they can process in the time domain (TD) or frequency domain (FD), and can be parametric or non-parametric [45]. TD approaches are straightforward, and are generally parametric. They mainly study the auto-regression degree with (AR)MA-(X) models [46] and Subspace Identification technics [47] or the output correlations between channels (Polyreference, LSCE, Subspace Identification, ERA) [48]. Frequency approaches can be parametric (Polymax or Polyreference) [49] or non-parametric (Pick- Peaking, (E)-FDD) [50, 51]. Non-parametric methods often rely on Single Degree of Freedom (SDoF) theory, so that a pre-processing step is mandatory to separate modal contributions.

In OBMA, Polymax model has typically been implemented as identification support [40, 42]. However, Polymax does not seem to be the best candidate for such an identification, because parametric models always generate spurious modes (due to noise and numerical bias). Furthermore, the model order is always difficult to define (methods are based on stabilization diagrams or parsimony principle through the minimization of criteria, *e.g.* Akaike and Bayesian Information Criterion, AIC or BIC). In order-tracked signals, it is easy to know in advance the number of excited modes, which allows using non-parametric methods. Thus, the authors propose to perform the following procedure: different modes are decomposed into SDoF responses and bandlimited using a partial E-FDD procedure. Then, each mode is identified using a classical ambient SDoF transfer function with a maximum likelihood estimator.

4.1 SDoF Separation

SDoF separation is performed using a modal coherent criterion applied on the singular vectors of the discrete spectral density matrix. The classical input-output relation of density matrices under the condition of white-noise input, low damping and uncoupled modes, can be developed according to the Heaviside partial-fraction expansion theorem in the vicinity of its modal pulsations [52, 53].

$$\boldsymbol{E}_k|_{\omega_k \approx \omega_r} \approx \boldsymbol{\Phi}^* diag\left[\frac{\beta_r}{(\xi_r\omega_r)^2}\delta_r\right]\boldsymbol{\Phi}^T \tag{1}$$

Eq. 1 shows that the excitation density matrix is diagonal, and thus the output density matrix E is equivalent to a diagonal one. The change of basis is done with the modal matrix $\boldsymbol{\Phi} = (\boldsymbol{\varphi}_1, ...\boldsymbol{\varphi}_n)$. The diagonal matrix contains only one non-zero term expressed as the contribution of the investigated mode through $(\beta_r, \xi_r, \omega_r)$, respectively a characteristic scalar value, the damping ratio and the natural pulsation of mode r. The Kronecker symbol δ_r is 1 for the $r-th$ position of the diagonal matrix, else 0. In the vicinity of a natural pulsation, the associated mode is the only contributor to the global dynamic of the system. The associated mode shape is $\boldsymbol{\varphi}_r$, $r-th$ column of $\boldsymbol{\Phi}$. In other terms, it is shown that the Singular Value Decomposition of the experimental density matrix in the vicinity of a mode returns only one dominant singular value. The associated singular vector in $\omega_k = \omega_r$ (within the limit of frequency resolution) is the mode shape estimator. The set of the first singular values $\{s_{1,k}\}$ is called Complex Modal Identification Function (CMIF) and is a unilateral representation of the previous spectral density functions. The resonance function of each SDoF is identified from the CMIF thanks to a discriminating criterion, the Modal Assurance Criterion (MAC) [54]. MAC varies from 0 to 1 as the modal coherence increases. It compares the degree of agreement of two vectors:

$$MAC(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) = \frac{<\boldsymbol{\varphi}_i|\boldsymbol{\varphi}_j>^2}{<\boldsymbol{\varphi}_i|\boldsymbol{\varphi}_i><\boldsymbol{\varphi}_j|\boldsymbol{\varphi}_j>} \tag{2}$$

This criterion is able to separate two uncoupled close modes and discriminate spike noises. Brinker et *al.* set the threshold to $MAC_{thr} = n/\sqrt{N_S}$, where n is an integer so that MAC_{thr} is close but lower than 1, and N_S the number of studied sensors [55]. If $MAC(\widehat{\boldsymbol{\varphi}_r}, \boldsymbol{\varphi}_j) > MAC_{thr}$, where $\widehat{\boldsymbol{\varphi}_r}$ is the shape estimator and $\boldsymbol{\varphi}_j$ a singular vector of the CMIF, then $s_1[\omega_j]$ belongs to the resonance function of the SDoF. This ensures to select a bandwidth with high modal coherence. Figure 1 shows an example of the use of MAC. The E-FDD theory shifts back in time domain to make the identification. But this procedure is a bad damping estimator, especially in the case of short signals [56]. For this reason, the identification support is different and presented in the next subsection.

A last point can be raised about the E-FDD limits: this procedure is only proper to separate uncoupled signals. In the case of coupled modes, it is unable to differentiate modal contributions. In future works, a Frequency-Domain Blind Source Separation developed by Castiglione et *al.* should be used instead [57]. FD-BSS is able to separate coupled modes with an impressive accuracy, and relies on a more rigorous mathematical approach.

4.2 Identification Using Maximum Likelyhood

After being extracted using AD method and bandlimited with MAC, the N_S experimental frequency responses are concatenated into a vector $\widehat{\boldsymbol{X}}_k$ and are modelled with the classical SDoF response described in eq. (3, 4), where $\boldsymbol{X}_k$ is the theoretical response vector, $h_{r,k}$ is the modal transfer function depending on modal parameters (ω_r, ξ_r) and $p_k, \boldsymbol{\varepsilon}_k$ are the normalized Fourier transforms of excitation and noise respectively.

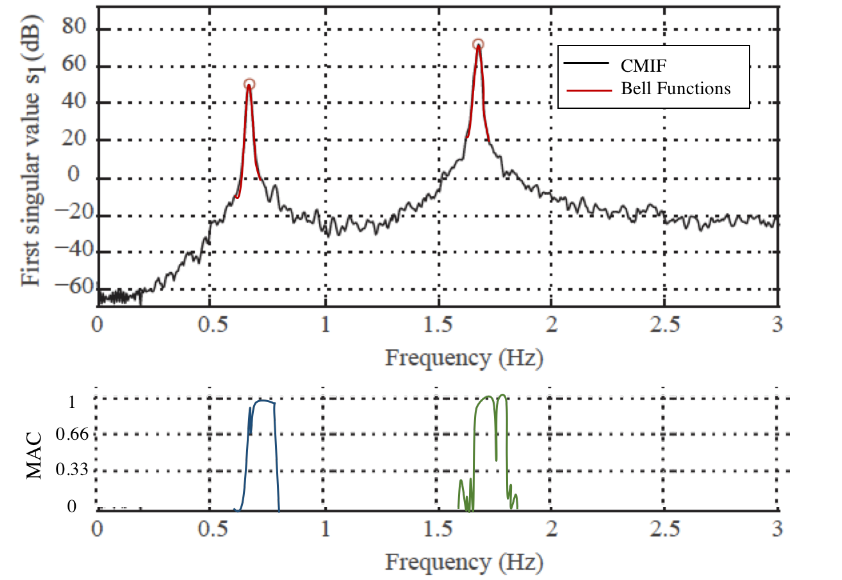


Figure 1: Modal contributions are framed using MAC, which compares the agreement level between two mode shapes.

$$\boldsymbol{X}_k = \boldsymbol{\varphi}_r h_{r,k} p_k + \boldsymbol{\varepsilon}_k \tag{3}$$

$$h_{r,k} = -\frac{1}{\omega_k^2 - \omega_r^2 - 2i\xi_r \omega_r \omega_k} \tag{4}$$

The associated Negative Log-Likelihood Function (NLLF) is given in eq. (5), where N_f is the number of point per channel, $\boldsymbol{E}_k[\boldsymbol{\theta}]$ the theoretical SDoF density matrix arising from eq. (3) and $|\boldsymbol{E}_k[\boldsymbol{\theta}]|$ the determinant of the density matrix ; the analytical determination of both determinant and inverse matrix of $\boldsymbol{E}_k[\boldsymbol{\theta}]$ is far from being trivial, and described in [58]. $\boldsymbol{\theta}_r = (\omega_r, \xi_r, S_r, S_{er}, \boldsymbol{\varphi}_r)$ is the parameter vector, including natural pulsation, damping ratio, modal force, PSD error and mode shape. $\boldsymbol{\theta}$ is the parameter variable, used to estimate $\boldsymbol{\theta}_r$. The NLLF is minimised using a Nelder-Mead algorithm [59]. A such identification method was chosen because it shows the best asymptotic properties.

$$\mathscr{L}(\boldsymbol{\theta}) = N_S N_f ln(\pi) + \sum_{k=1}^{Nf} ln(|\boldsymbol{E}_k[\boldsymbol{\theta})]|) + \sum_{k=1}^{N_f} \widehat{\boldsymbol{X}}_k^* \boldsymbol{E}_k^{-1}[\boldsymbol{\theta}] \widehat{\boldsymbol{X}}_k \tag{5}$$

5 Case Study

The studied measurements come from a vertical medium head Francis hydroelectric runner exploited in Quebec, Canada. This facility was chosen because the turbine was designed and is operated by two partners of the current project. The measurement data were recorded during a slow transient from no-load overspeed to stop. Two blades separated with an angle of 111° were instrumented with strain gauges. Intrados were instrumented with three strain gauge rosettes, located in the band junction to blade leading edge and trailing edge, and in the middle crown-blade weld, as shown in Figure 2. Extrados were instrumented with two uniaxial gauges, one close to the crown, the other close to the band. The locations are the same from one blade to another to ensure a redundant signal. Accelerometers and pressure sensors are located in different points (blade, structure and penstock) and sensors are also installed on the shaft to record torque, flexion and thrust. The rosette and uniaxial gauges are oriented in agreement with the expected strain flow direction, *.e.* in the direction of the principal stresses.

An analysis of experimental correlograms (amplitude of short-time Fourier transforms) and absolute phase-shift spectra between redundant sensors was made first. Some examples are shown in Figures 3 and 4, related to

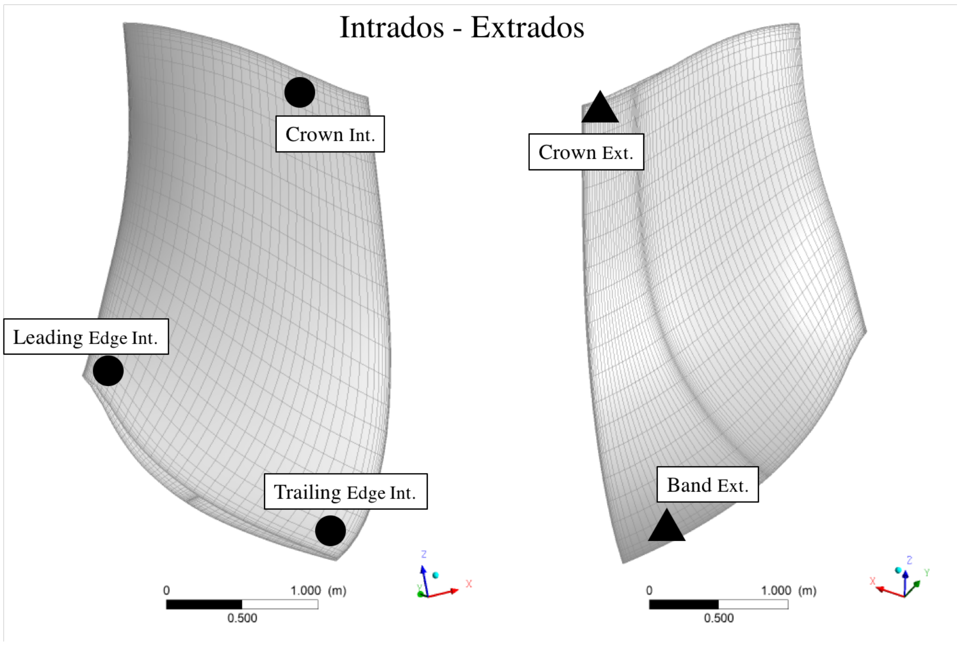


Figure 2: Strain gauge location on the instrumented blades. Circles represent rosette gauges, triangles represent uni-axial gauges.

both sides of the crown (time and frequency axes are empty for the purpose of clarity). When the windowing is long enough, correlograms show five ODS. Several modes can be contained inside. A clear resonance of mode 1 is detected on the intrados in the lower part of the diagram (below blade passing frequency signature, abusively denoted "RSI") ; the other resonances are in the upper part. Phase-Shift spectra are amplitude-filtered, and show only phases associated with a high enough correlogram. They show four single-mode bands and a multi-mode band, due to multiple phase-shift detection. Resonance of mode 2 is detected on both sides, and mode 3 is vaguely detected on the extrados. Only one mode of the multi-shifted band is excited by a harmonic, thus leading to a SDoF resonance. Table 1 summarizes all the detected modes in the range $[0,100]Hz$, and reports the related resonant harmonic index. Phase-shifts are averaged over the region where the mode is found.

In the studied data, all the time series are recorded with an extensively large sampling frequency, and the use of AD technique to extract harmonics should be straightforward. The next subsections present typical case studies based on previous identified modes.

Mode Reference	Frequency [Hz]	Phase-Shift $\times\pi$[rad]	Harm. Order
1	18.0	0	13
2	28.0	$\pm 3/4$	42
3	50.0	$\pm 1/7$	63
4	91.0	$\pm 1/9$	$\emptyset$
5	58.0	$\pm 6/7$	61

Table 1: Experimentally Detected Modes

5.1 Identification Example: Shaft Torsion Mode with $f_0 = 18Hz$

The first mode to be studied, mode 1, is excited by the $13-th$ harmonic of the rotating speed. This corresponds to the blade passing frequency. Such an excitation can come for instance from the spiral case intake or the drat tube elbow that can create a stationary disturbance that is seen by the rotating runner each time a blade passes in front of the intake or the draft tube direction. The investigation of torsion measurements shows that the studied mode is a natural torsion mode of the shaft line. What is observed on blades is only the propagation of shaft natural vibrations. Thus, all the runner is excited with the same phase, and the nodal diameter is 0, that

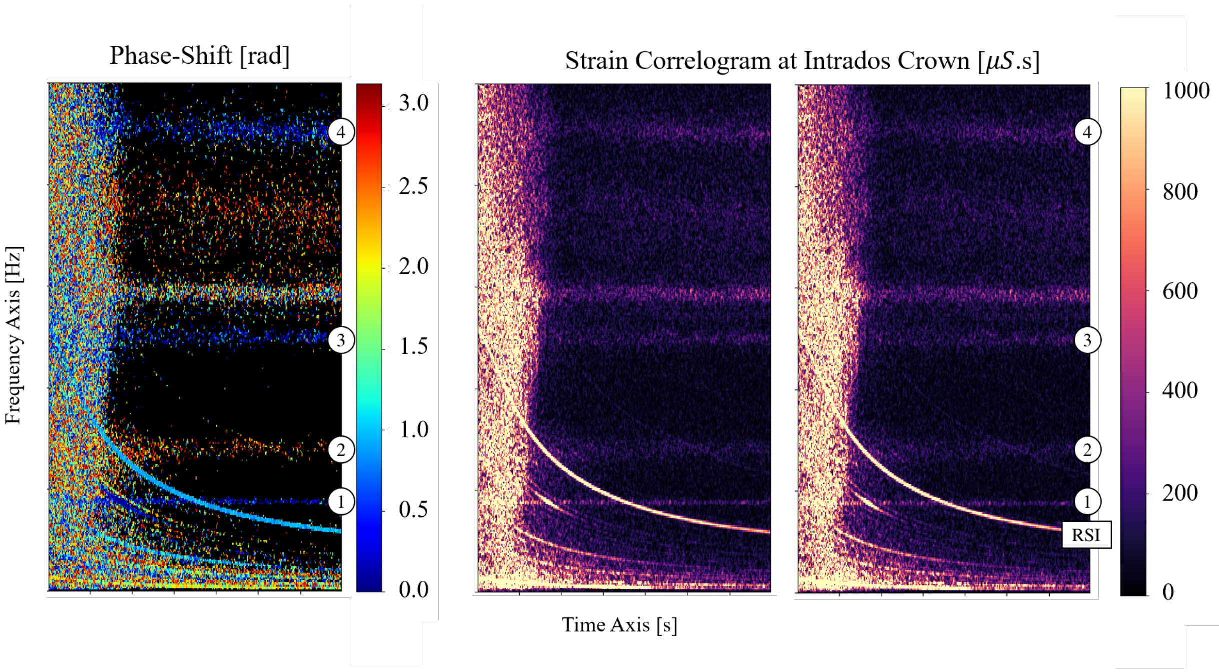


Figure 3: ODS Analysis of principal direction of intrados crown Rosette gauge. On the left, phase-shift spectrum of the redundant gauges. On the right, redundant amplitude spectra.

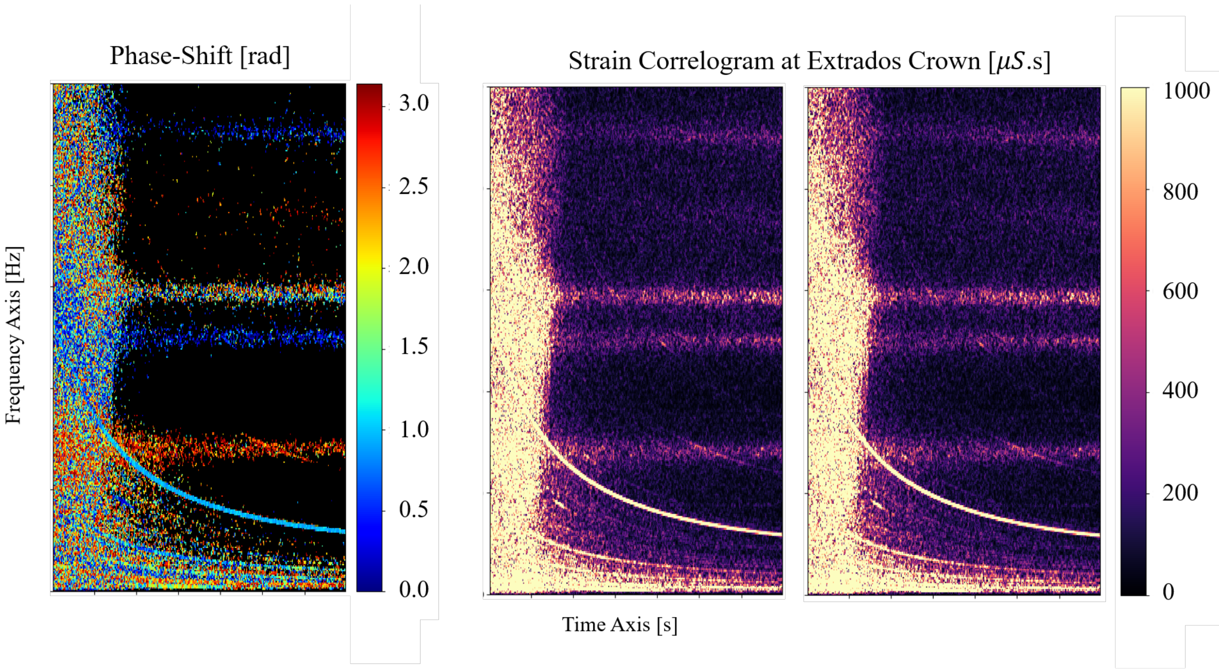


Figure 4: ODS Analysis of extrados crown Rosette gauge.

is confirmed by the absence of phase-shift between blades. An axial thrust pulsation is measured on the shaft, and indicates that the inflow to the runner is not symmetric to the guide vane orientation [31]. Examples of extracted resonances are shown in Figure 5. The *MAC* narrows the bandwidth with a threshold $MAC_{thr} = 0.875$, as depicted in Figure 6. The maximum likelihood estimator raises optimal parameters, shown in Table 2 and Figure 7.a.

Natural Frequency	Damping Ratio	Modal Force	PSD Error
f_0 [Hz]	ξ [%]	S [ms^2/Hz]	S_e [μ S/Hz]
17.43	1.26	2.22E6	1.73E-2

Table 2: Torque Mode Featuring

The shape relative amplitudes are the same on the two blades, as testifies Figure 7.a.. The mode shape is in phase opposition from leading edge to crown, and is not spotted neither on trailing edge intrados or on band extrados signals. That attests a $ND-0$ "in umbrella", as depicted in Figure 7.b. The modal force is very difficult to extract and is likely very biased. The bias on damping ratio mainly depends on experimental data. The leakage and tapering bias due to the windowing is avoided thanks to COT-based AD. However, the global uncertainty level remains likely high because of the unknown excitation.

5.2 Results

Table 3 shows the result of the identification process performed on all detected resonance harmonics. The information presented is: the exciting harmonic (indexed on the rotating frequency), the most likely nodal diameter, the Signal-to-Noise Ratio (SNR), the bandwidth and the associated method (*MAC* or SENS for sensitivity analysis) and the modal parameters. Except for the torsion mode, resonance signals have a low SNR that renders impossible the use of *MAC*, because the singular value spectrum is still buried in noise. Instead, a sensitivity analysis was made on modal parameters as a function of the bandwidth. The selected band corresponds to the parameter convergence. This method gives wider bands (around ten times the width of a *MAC* selected bandwidth), where noise has a significant influence. Rather, *MAC* criterion selects a narrower frequency band with very few noise. The two methods return quite equivalent results. The experimental data reveals five isolated modes numbered from 1 to 4 in Figure 4, and a multi-mode band. Amongst the 4 well-separated modes, only the first 3 are excited by a harmonic and then identifiable. Into the multi-mode band, one mode is excited by a harmonic. It is thus possible to feature it, but the SNR is particularly low. Notice that the first mode of table 3 is the torsion mode featured in table 2.

6 Conclusion

This paper shows that Francis runner structural modes can be identified from ambient vibration data during transient conditions. These modes have been successfully extracted and identified through an enhanced OBMA technique (E-OBMA). E- OBMA combines three existing techniques and takes benefit from the best of each: Order Tracking separation quality, MAC bandlimiting rigor and maximum likelihood accuracy. This work shows that experimental transient data contains accurate frequency information that can be used to assess numerical model validity. The presented results are the first effort in creating OMA strategy tailored for Francis runners. The E-OBMA still has to be validated on an analytical case, which is now being developed. The Order Tracking quality should be evaluated in conjunction with EMA sine-sweep excitation theory. Also, further improvements will make possible the uncertainty quantification which is a major stake in signal processing.

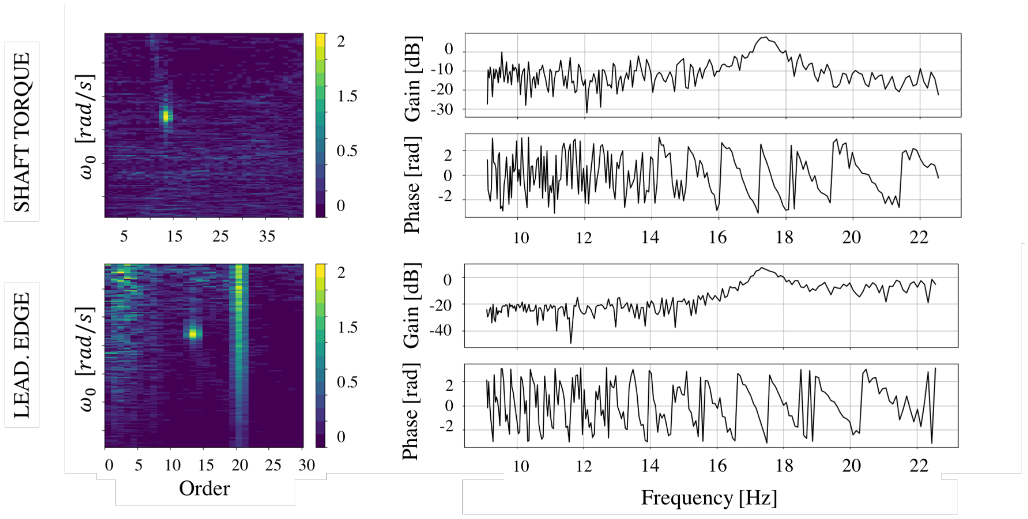


Figure 5: On the left, Order Spectra. All the harmonic contents are on the same line. On the right, the related Bode Diagrams of the harmonic 13.

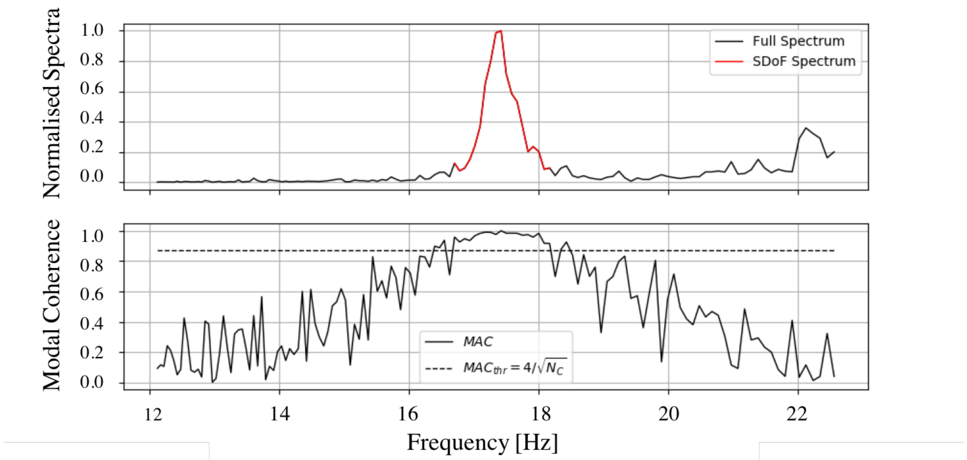


Figure 6: Resonance function extraction from the first singular values spectra (CMIF).

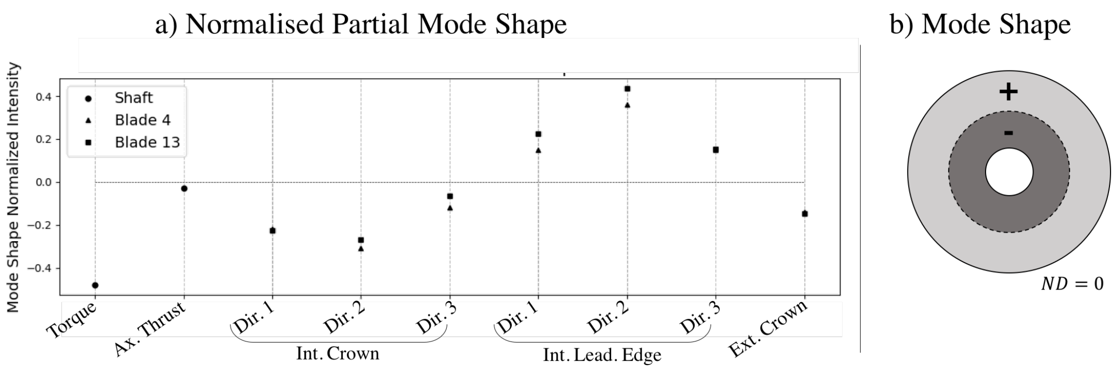


Figure 7: a) Partial mode shape extracted with maximum likelihood. b) Schematics of the observed mode shape.

		Identification BandWidth	Modal Parameters	
Mode	Nodal Diameters	Bandwidth (Hz)	Frequency (Hz)	Modal Force (ms^2/Hz)
Harm. Index	SNR (dB)	Method	Damping ratio (%)	PSD Error (μ s/Hz)
1	0	1.5	17.43	2.22E6
13	7.4	MAC	1.26	1.73E-2
2	1	8	28.71	3.31E5
42	1.8	SENS	3.26	1.18E-3
3	3/5	9	49.84	1.40E6
63	1.9	SENS	1.67	8.00E-4
Multi.	6	12	59.35	4.31E6
61	0.96	SENS	2.30	2.00E-3

Table 3: OBMA Identification Results

References

[1] A. Presas, Y. Luo, Z. Wang, and B. Guo. Fatigue life estimation of francis turbines based on experimental strain measurements: Review of the actual data and future trends. *Renewable and Sustainable Energy Reviews vol. 102*, page 96 to 110, March 2019.

[2] J. Nicolle, J.F. Morissette, and A.M. Giroux. Transient cfd simulation of a francis turbine startup. *IOP Conf. Ser.: Earth Environ. Sci. vol. 15 062014*, October 2012.

[3] J. Nicolle, A.M. Giroux, and J.F. Morissette. Cfd configurations for hydraulic turbine startup. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032021*, September 2014.

[4] T. Krappel, A. Ruprecht, S. Riedelbauch, R. Jester-Zuerker, and A. Jung. Investigation of francis turbine part load instabilities using flow simulations with a hybrid rans-les turbulence model. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032001*, March 2014.

[5] B. Nennemann, J.F. Morissette, J. Chamberland-Lauzon, and C. Monette. Challenges in dynamic pressure and stress predictions at no-load operation in hydraulic turbines. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032055*, 2014.

[6] J.F. Morissette, J. Chamberland-Lauzon, B. Nennemann, C. Monette, A.M. Giroux, A. Coutu, and J. Nicolle. Stress predictions in a francis turbine at no-load operating regime. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072016*, July 2016.

[7] B. Nennemann, C. Monette, and J. Chamberland-lauzon. Hydrodynamic damping and stiffness prediction in francis turbine runners using cfd. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072006*, page 854 to 865, July 2017.

[8] A. Presas, D. Valentin, E. Egusquiza, C. Valero, and U. Seidel. Influence of the rotation on the natural frequencies of a submerged-confined disk in water. *Journal of Sounds and Vibrations vol. 337*, page 161 to 180, February 2015.

[9] C. Trivedi and M.J. Cervantes. Fluid-structure interactions in francis turbines: A perspective review. *Renewable and Sustainable Energy Reviews vol. 68*, page 87 to 101, February 2017.

[10] A. Presas, D. Valentin, E. Egusquiza, C. Valero, M. Egusquiza, and M. Bossio. Accurate determination of the frequency response function of submerged and confined structures by using pzt-patches. *Sensors*, March 2017.

[11] A. Ducoin, J. Andre-Astolfi, and J.-F. Sigrist. An experimental analysis of fluid structure interaction on a flexible hydrofoil in various flow regimes including cavitating flow. *European Journal of Mechanics - B/Fluids vol. 36*, page 63 to 74, December 2012.

[12] A. Lelong, P. Guiffant, and J.A. Astolfi. An experimental analysis of the structural response of flexible lightweight hydrofoils in cavitating flow. *Journal of Fluids Engineering vol. 140*, November 2017.

[13] X. Liu, Y. Luo, A. Presas, Z. Wang, and L. Zhou. Cavitation effects on the structural resonance of hydraulic turbines: failure analysis in a real francis turbine runner. *Energies vol. 11*, September 2018.

[14] D. Valentin, D. Ramos, M. Bossio, A. Presas, E. Egusquiza, and C Valero. Influence of the boundary conditions on the natural frequencies of a francis turbine. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072004.*, November 2016.

[15] M. Gagnon, A. Tahan, P. Bocher, and D. Thibault. On the stochastic simulation of hydroelectric turbine blades transient response. *Mechanical Systems and Signal Processing vol. 32*, page 178 to 187, October 2012.

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

[17] I. Diagne, M. Gagnon, and A. Tahan. Modeling the dynamic behavior of turbine runner blades during transients using indirect measurements. *IOP Conf. Ser.: Earth Environ. Sci. vol. 49 072014*, July 2016.

[18] E. Moisan, D.-B. Giacobbi, M. Gagnon, and F. Leonard. Self-excitation in francis runner during load rejection. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032025*, September 2014.

[19] M. Gagnon, A. Tahan, A. Coutu, and M. Thomas. Oma en presence d'excitations harmoniques : etude de cas sur des composantes de turbine hydroelectrique. *24th Seminar on Machinery Vibration*, October 2006.

[20] D. Valentin, A. Presas, M. Bossio, and M. Egusquiza Montagut. Feasibility to detect natural frequencies of hydraulic turbines under operation using strain gauges. *Sensors*, January 2018.

[21] A. Presas, D. Valentin, E. Egusquiza, and C. Valero. Detection and analysis of part load and full load instabilities in a real francis turbine prototype. *J. Phys.: Conf. Ser. 813 012038*, April 2017.

[22] D. Liu, X. Lui, and Y. Zhao. Experimental investigation of inter-blade vortices in a model francis turbine. *Chinese Journal of Mechanical Engineering vol. 30*, February 2016.

[23] R. Goyal and B.K. Gandhi. Review of hydrodynamics instabilities in francis turbine during off-design and transient operations. *Renewable Energy vol. 116*, page 697 to 709, October 2017.

[24] X. Liu, Y. Luo, and Z. Wang. A review on fatigue damage mechanism in hydro turbines. *Renewable and Sustainable Energy Reviews vol. 54*, February 2016.

[25] A. Gagnon, A. Tahan, P. Bocher, and D. Thibault. Influence of load spectrum assumptions on the expected reliability of hydroelectric turbines: A case study. *Structural Safety, vol. 50*, March 2014.

[26] X. Huang, J. Chamberland-Lauzon, C. Oram, A. Klopfer, and N. Ruchonnet. Fatigue analyses of the prototype francis runners based on site measurements and simulations. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 012014*, 2014.

[27] U. Seidel, C. Mende, B. Hubner, W. Weber, and D. Otto. Dynamic loads in francis runners and their impact on fatigue life. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032054*, 2014.

[28] C. Trivedi, B. Gandhi, and M.J. Cervantes. Effect of transients on francis turbine runner life: a review. *Journal of Hydraulic Research vol. 51-2*, page 121 to 132, 2013.

[29] F. Leonard. Spectrogramme de phase et spectrogramme de frequence. *Traitement de signal vol. 17 nb 4*, page 269 to 286, 2000.

[30] F. Leonard. Phase spectrogram and frequency spectrogram as new diagnostic tools. *Mechanical Systems and Signal Processing vol. 21*, page 125 to 137, January 2007.

[31] P. Dorfler, M. Sick, and A. Coutu. Flow-induced pulsation and vibration in hydroelectric machinery. *Engineer's Guidebook for Planning, Design and Troubleshooting*, April 2012.

[32] G. Blommaert. Etude du comportement dynamique des turbines francis : controle actif de leur stabilite de fonctionnement. *These de l'Université Polytechnique Federale de Lausanne*, 2000.

[33] A. Favrel, C. Landry, A. Muller, and F. Avellan. Experimental identification and study of hydraulic resonance test rig with francis turbine operating at partial load. *IOP Conf. Ser.: Earth Environ. Sci. vol. 15 062064*, August 2012.

[34] A. Favrel, C. Landry, A. Muller, K. Yamamoto, and F. Avellan. Hydro-acoustic resonance behavior in presence of a precessing vortex rope: observation of a lock-in phenomenon at part load francis turbine operation. *IOP Conf. Ser.: Earth Environ. Sci. vol. 22 032035*, September 2014.

[35] C. Nicolet, N. Ruchonnet, S. Alligne, J. Koutnik, and F. Avellan. Hydroacoustic simulation of rotor-stator interaction in resonance conditions in francis pump-turbine. *IOP Conf. Ser.: Earth Environ. Sci. vol. 12 012005*, 2010.

[36] H. Tanaka. Vibration behavior and dynamic stress of runners of very high head reversible pump-turbines. *International Journal of Fluid Machinery and Systems vol.4*, page 289 to 306, June 2011.

[37] J. Blough, D. Brown, and H. Vold. The time variant discrete fourier transform as an order tracking method. *SAE Noise and Vibration Conf. and Exp.*, May 1997.

[38] P. Borghesani, P. Pennachi, S. Chatterton, and R. Ricci. The velocity synchronous discrete fourier transform for order tracking in the field of rotating machinery. *Mechanical Systems and Signal Processing vol. 44*, page 118 to 133, February 2014.

[39] D.B. Stephens and H. Vold. Order tracking signal processing for open rotor acoustics. *Journal of Sounds and Vibration vol. 333*, page 3818 to 3830, August 2014.

[40] E. Di Lorenzo. Operational modal analysis for rotating machines, challenges and solutions. *PhD - KL Arenberg Doct. School and Frederico II Naple Univ.*, April 2017.

[41] K.R. Fyfe and D.S. Munck. Analysis of computed order tracking. *Mechanical Systems an Signal Processing vol. 11*, page 187 to 205, March 1997.

[42] K. Janssens, Z. Kollar, B. Peeters, S. Pauwels, and H. Van der Auweraer. Order-based resonance identification using operational polymax. *ResearchGate Article*, January 2006.

[43] E. Di Lorenzo, S. Manzato, A. Dabizzi, B. Peeters, F. Marulo, and W. Desmet. Industrial applications of advanced modal identification on operational rotating machineries. *Proceeding of ISMA*, page 2833 to 2848, September 2016.

[44] E. Di Lorenzo, A. Palermo, S. Manzato, A. Dabizzi, B. Peeters, W. Desmet, and F. Marulo. Gear dynamics characterization by using order-based modal analysis. *Proceedings of the International Modal Analysis Conference (IMAC)*, page 387 to 404, 2016.

[45] H. Van der Auweraer, P. Mas, B. Peeters, K. Jannsens, and A. Vecchio. Modal and path contribution models from in-operation data: Review and new approaches. *Shock and Vibration vol. 15*, page 403 to 411, 2008.

[46] A.G. Poulimenos and S.D. Fassois. Parametric time-domain methods for non-stationary random vibration modelling and analysis : A critical survey and comparison. *Mechanical Systems and Signal Processing vol. 20*, page 764 to 816, January 2006.

[47] L. Ljung. System identification. *University of Linkoping, Sweden*, 1998.

[48] R Brinker and C.E. Ventura. Introduction to operational modal analysis. *Wiley Editions*, 2015.

[49] B. Peeters, H. Van der Auweraer, P. Guillaume, and J. Leuridan. The polymax frequency-domain method: a new standard for modal parameter estimation? *Shock and Vibration vol. 11*, page 395 to 409, August 2004.

[50] R. Brinker, L. Zhang, and P. Andersen. Modal identification from ambiant responses using frequency domain decomposition. *Proceedings of the International Modal Analysis Conference (IMAC)*, page 625 to 630, February 2000.

[51] R. Brinker, L. Zhang, and P. Anderser. Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures vol. 10*, page 441 to 445, January 2001.

[52] L. Zhang, T. Wang, and Y. Tamura. A frequency-spatial domain decomposition (fsdd) method for operational modal analysis. *Mechanical Systems and Signal Processing vol. 24*, page 1227 to 1239, July 2010.

[53] C.-X. Qu, T.-H. Yi, H.-N. Li, and B. Chen. Closely spaced modes identification through modified frequency domain decomposition. *Measurement vol. 128*, page 388 to 392, November 2018.

[54] M. Pastor, M. Binda, and T. Harcarik. Modal assurance criterion. *Procedia Engineering vol. 48*, page 543 to 548, 2012.

[55] R. Brinker, P. Andersen, and N.-J. Jacobsen. Automated frequency domain decomposition for operational modal analysis. *Proceedings of the International Modal Analysis Conference (IMAC)*, 2007.

[56] M. Danial, A. Hasan, Z.A.B. Ahmad, M. Salman Leong, and L.M. Hee. Enhanced frequency domain decomposition algorithm: a review of a recent development for unbiased damping ratio estimates. *Journal of Vibro-Engineering vol. 20*, page 1919 to 1936, August 2018.

[57] R. Castiglione, J. Antoni, and L. Garibaldi. Separation and identification of structural modes in largely underdetermined scenarios using frequency banding. *Journal of Sound and Vibration, vol. 414*, page 192 to 217, February 2018.

[58] S.-K. Au. Operational modal analysis, modeling, bayesian inference, uncertainty laws. *Springer*, 2017.

[59] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal vol. 7*, page 308 to 313, 1965.

A new method for identifying diagnostic rich frequency bands under varying operating conditions

Stephan Schmidt[1] Alex R. Mauricio[2] P. Stephan Heyns[3] Konstantinos C. Gryllias[4]

[1]Centre for Asset Integrity Management, University of Pretoria, Pretoria, South Africa
stephan.schmidt@up.ac.za

[2]Department of Mechanical Engineering, PMA Division, KU Leuven, Leuven, Belgium
Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium
alex.ricardomauricio@kuleuven.be

[3]Centre for Asset Integrity Management, University of Pretoria, Pretoria, South Africa
stephan.heyns@up.ac.za

[4]Department of Mechanical Engineering, PMA Division, KU Leuven, Leuven, Belgium
Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Belgium
konstantinos.gryllias@kuleuven.be

Abstract

Performing condition monitoring under time-varying operating conditions is challenging. The varying operating conditions impede the ability of conventional fault diagnosis methods to detect damage on rotating machine components such as bearings and gears. This paper investigates a new method for identifying diagnostic rich frequency bands under time-varying operating conditions. This method uses the order-frequency spectral coherence and a feature, which is dependent on the cyclic order of interest and the frequency resolution of the spectral coherence, to decompose the signal into a feature plane. Thereafter, the spectral frequency and the spectral frequency resolution that maximise the feature plane are used to design a bandpass filter. The bandpass filter extracts a diagnostic rich signal, which can be analysed by using the squared envelope spectrum or the synchronous average. The proposed method is compared to the fast kurtogram on a numerical gearbox dataset as well as on an experimental gearbox dataset, with very promising results obtained.

1 Introduction

Effective fault diagnosis techniques are important for expensive assets such as wind turbines, because this can result in early detection of faults, their characteristics can easily be understood (e.g., which component is damaged) and subtle changes in the damage (i.e. deterioration) can be monitored. Many rotating machines inherently operate under time-varying operating conditions, which impede effective fault diagnosis. Hence, it is important to use condition monitoring techniques that are able to diagnose damaged machine components under time-varying operating conditions.

Damaged rotating machine components such as bearings result in periodical excitations of the structure at a rate dependent on the kinematic characteristics of the component (e.g. ball pass order of the outer race, shaft rotation). This angle-dependent periodical excitation of the time-invariant structure generates signals that can be approximated as angle-time cyclostationary [1]. Abboud et al. [1, 2, 3] extended the suite of conventional time and angle cyclostationary techniques to time-varying speed conditions with tools such as the Order-Frequency Spectral Coherence (OFSCoh) being one of the most powerful fault diagnosis techniques for bearings under varying speed conditions.

However, in condition monitoring it is usually desired to utilise simple metrics or representations for making decisions (e.g. a spectrum is preferred instead of a time-frequency spectrum). Hence, the enhanced envelope spectrum and the even more powerful Improved Envelope Spectrum (IES), both calculated from the spectral coherence or the spectral corelation, can be used to diagnose the machine. For the IES, it is very important to select carefully the integration band to ensure that the IES has an optimal signal-to-noise ratio. This means that it is important to be able to identify frequency bands that are rich with diagnostic information. Identifying diagnostic rich frequency bands is also important for calculating the synchronous average and the squared envelope spectrum [4].

The spectral kurtosis and the related kurtogram are effective for identifying frequency bands with much impulsive information [5, 6]. This is very appropriate for diagnostics, because bearing damage [4, 5] and gear damage [7] result in vibration signals containing bandlimited impulses. However, the kurtogram is sensitive to transients not related to the condition of the machine and it is not possible to investigate the optimal frequency band to detect damage associated with a specific cyclic order. Recently, new methods such as the infogram [8] and the IESFOgram [9] have been proposed for identifying frequency bands that are rich with diagnostic information by improving the shortcomings of the kurtogram.

A new method is investigated in this paper that is able to identify a frequency band that contains diagnostic information related to a specific machine component under time-varying operating conditions. This has a significant advantage over conventional methods, because incipient damage components that are normally masked by other dominant signal components and distorted by time-varying operating conditions, can be extracted from the signal and used to diagnose the machine. The performance of this method is compared to the Fast Kurtogram on numerical gearbox data as well as on experimental gearbox data, both acquired under time-varying operating conditions.

The outline of this paper is as follows: In Section 2, the proposed method is presented, whereafter it is investigated on phenomenological gearbox data in Section 3 and experimental gearbox data in Section 4. In the last section, Section 5, some conclusions are extracted and some recommendations are made for future investigations.

2 Methodology

2.1 Overview of the methodology

An overview of the methodology is presented in Figure 1. The measured vibration signal and the corresponding rotational speed (or phase) is given as inputs, whereafter an Order-Frequency Spectral Coherence (OFSCoh) is calculated for a specific window length. A feature is extracted from each frequency band of the calculated OFSCoh. This process is repeated for the set of window lengths under consideration, whereafter a feature plane is constructed. The feature plane contains the value of the feature for different combinations of centre frequencies and window lengths (or frequency resolutions). Thereafter, the feature plane is maximized to obtain the parameters of a bandpass filter. This bandpass filter is used to extract a signal that is rich with diagnostic information from the original signal, whereafter the filtered signal can be analysed to infer the condition of the machine component.

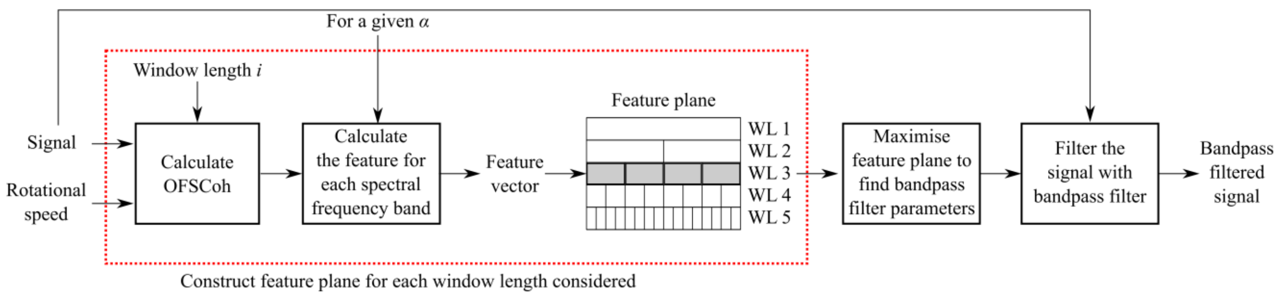


Figure 1: The proposed method for identifying frequency bands that are rich with diagnostic information.

The subsequent sections give detailed information on each step in the proposed method.

2.2 Order-Frequency Spectral Coherence (OFSCoh)

The impulses generated by components such as bearings are periodic in the angle domain, while they manifest in the time-invariant frequency bands. This means that the OFSCoh can be used to identify the resonance bands that are excited at specific cyclic orders. The OFSCoh [2]

$$\gamma_{xx}(\alpha, f) = \frac{S_{xx}(\alpha, f)}{\left(S_{xx}(0, f) S_{x_\alpha x_\alpha}(0, f)\right)^{1/2}} \tag{1}$$

provides a two-dimensional view of the modulating frequencies (i.e. cyclic orders) and their carriers (i.e. spectral frequencies) in the signal *x(t)*. The Order-Frequency Spectral Correlation (OFSC) [2]

$$S_{xx}(\alpha, f) = \lim_{W \to \infty} \frac{1}{\Phi(W)} E\left\{F_W\left(x(t)\right)^* F_W\left(x(t) e^{-j\alpha\theta(t)} \dot{\theta}(t)\right)\right\} \tag{2}$$

is used to calculate the OFSCoh in Equation (1). The expectation operator is denoted *E*, the Fourier transform is denoted F_w and $\Phi(W)$ denotes the phase of the shaft during the measurement time period *W*. The instantaneous phase of the shaft is denoted θ. It is easier to detect non-dominant components by using the OFSCoh as opposed to the OFSC.

Estimators need to be used to calculate the OFSCoh for the measured data, with the Welch estimator as proposed in Ref. [2], used in this work. The Welch estimate of the OFSCoh is denoted $\gamma_{xx}(\alpha, f; \Delta f)$, where Δf is the frequency resolution that is used to obtain the estimate.

2.3 Frequency Band Identification (FBI)

It is possible to use a one-dimensional metric such as the kurtosis to identify the frequency band of interest. However, one-dimensional metrics do not allow different signal components to be distinguished from one another, which may result in a frequency band to be identified that is not necessarily of interest. Hence, a more advanced metric is required.

2.3.1 Feature extraction

Ref. [4] uses a metric to quantify the quality of the Squared Envelope Spectrum (SES). If their metric is large, it means that the diagnostic information is dominant with respect to the noise level in the SES, while a small metric indicates that it could be difficult to detect the cyclic components in the SES. The authors estimated the noise level with the median because the median is robust to outliers generated by the cyclic components in the SES.

We used this metric as inspiration for designing the feature to identify the frequency band of interest, with the following feature obtained for the cyclic order set $\{\alpha_f\}$:

$$\Psi_{xx}(f, \Delta f; \{\alpha_f\}) = \sum_{\{\alpha_f\}} \frac{|\gamma_{xx}(\alpha_f, f; \Delta f)|^2}{median\left(|\gamma_{xx}(\alpha_f, f; \Delta f)|^2\right)} \tag{3}$$

The numerator contains the squared magnitude of the spectral coherence for a specific window length Δf. The denominator contains the median function, which is calculated for the squared magnitude of the spectral coherence and is used to estimate the noise level in the OFSCoh. The following points are important considerations when calculating the feature for practical signals:

1. The analytical cyclic orders may be different from the actual cyclic orders due to slip and therefore the maximum of a range of $[0.9\alpha_f, 1.1\,\alpha_f]$ is calculated to estimate the numerator.
2. The median of the squared magnitude OFSCoh cannot be calculated at $\alpha=\alpha_f$ and therefore it needs to be estimated from the discrete OFSCoh data. Hence, the median of the squared magnitude of the OFSCoh in the range of $[\alpha_f -1, \alpha_f +1]$ is used to estimate the denominator.

This feature also has similarities to the feature used by the IESFOgram [9]. In the latter method the ratio of the signal components in the IES are calculated with respect to the mean of the IES in the predefined bandwidth.

2.3.2 Feature plane construction and maximisation

The feature is calculated for each frequency band in the OFSCoh. The Welch estimator of the OFSCoh depends on a number of parameters, namely, the window length, the window overlap as well as the number of points used to calculate the FFT. It is best to use an overlap longer than 75% of the window length, however, the window length needs to be determined prior to the analysis. It is also necessary to estimate the frequency bandwidth and not only the centre frequency for designing the bandpass filter parameters. Hence, the following procedure is used to simultaneously optimise the centre frequency and frequency bandwidth of the frequency band of interest: Firstly, the OFSCoh is calculated for a specific window length, whereafter the feature is calculated for each spectral frequency band in the OFSCoh. This process is repeated for each window length under consideration, whereafter the feature plane is obtained. The frequency band parameters are identified by finding the centre frequency and frequency bandwidth that maximise the feature plane. This is a very similar procedure to the kurtogram and the infogram, but instead of using the short-time Fourier transform, the OFSCoh is used, and instead of maximising a scalar value (e.g. spectral kurtosis), the maximisation is done for a set of cyclic orders. This allows the optimal frequency band to be determined to detect a set of cyclic orders.

The identified frequency band parameters can be used to calculate the IES or to extract a bandlimited signal. In this work, we used the frequency band parameters to design a bandpass filter, whereafter the bandpass filtered signal is interrogated. The bandpass filtered signal can subsequently be analysed with techniques such as the Synchronous Average (SA) [10] and the Squared Envelope Spectrum (SES) [3].

2.4 Computational aspects

Even though real-time condition monitoring is rarely required in practice, it is still necessary to provide answers in a reasonable time. The Welch-based estimator of the OFSCoh has very good bias and variance properties, but is very expensive to calculate for large datasets, especially for high rotational speed applications. If the cyclic orders of interest are known a priori, it is possible to only estimate the OFSCoh for specific cyclic orders; however, even this may be impractical for complex gearboxes found in wind turbines and helicopters, which may have many cyclic orders of interest. Fortunately, there has been very exciting developments in this field, where fast (and faster) estimators of the spectral correlation are proposed, which could make this method significantly faster to be calculated [11, 12].

3 Numerical gearbox data

In this section, we investigate the method and compare it to the kurtogram on data generated from a phenomenological gearbox model. In the next section, an overview is given of the model and the generated data, whereafter the Fast Kurtogram (FK) is used on the dataset in Section 3.2. The results of the proposed method are presented and discussed in Section 3.3.

3.1 Phenomenological Gearbox Model (PGM)

The Phenomenological Gearbox Model (PGM) proposed in Ref. [3] is used to generate a casing vibration signal. The casing vibration signal

$$x(t) = x_b(t) + x_{rg}(t) + x_n(t) \tag{4}$$

contains a bearing component $x_b(t)$, a random gear component $x_{rg}(t)$ and a broadband noise component $x_n(t)$. The generalised synchronous average can be used to attenuate the deterministic gear components attributed to the meshing of gears as described by Abboud et al. [3] and therefore they are not included in this model. The bearing component is generated by bearing damage on the outer race

$$x_b(t) = M(\omega(t)) \cdot h_b(t) \otimes \sum_{k=1}^{K_b} A_k \cdot \delta(t - T_k) \tag{5}$$

where T_k denotes the time-of-arrival of the kth bearing impulse, which incorporates the varying speed conditions and the slip. The amplitude of the kth impulse, denoted A_k, is sampled from a uniform distribution. The raw bearing impulses are filtered through the structure, which is assumed to have an impulse response function of a single degree-of-freedom system h_b. The modulating function $M(\omega(t)) = \omega^2$ is used to simulate the varying amplitude induced by time-varying operating conditions and is assumed to be the same for all signal components for the sake of simplicity.

The random gear component

$$x_{rg}(t) = M(\omega(t)) \cdot h_{rg}(t) \otimes \left(\varepsilon(t) \cdot \sum_{k=1}^{Krg} B_k \cdot \sin\left(k \cdot \int_0^t \omega(t) dt + \varphi_k \right) \right) \tag{6}$$

is attributed to gear damage and contains the random variable $\varepsilon(t)$ which is sampled from a zero mean, unit variance normal distribution, and B_k and φ_k are, respectively, the amplitude and the phase of the kth harmonic of the component. There are K_{rg} harmonics in the vibration signal. The noise component

$$x_n(t) = M(\omega(t)) \cdot \varepsilon(t) \tag{7}$$

is generated by a zero mean Gaussian distribution with its amplitude dependent on the rotational speed of the system. The natural frequency of the impulse response function of the bearing and the gear components are 7 kHz and 1.3 kHz respectively. The fundamental cyclic order of the distributed gear damage is 1.0 shaft order, while the fundamental cyclic order of the outer race bearing damage component is 4.12 shaft orders.

A single dataset is investigated in this paper with the time-varying speed profile *ω(t)* and the different signal components shown in Figure 2. This system operates under constant load conditions.

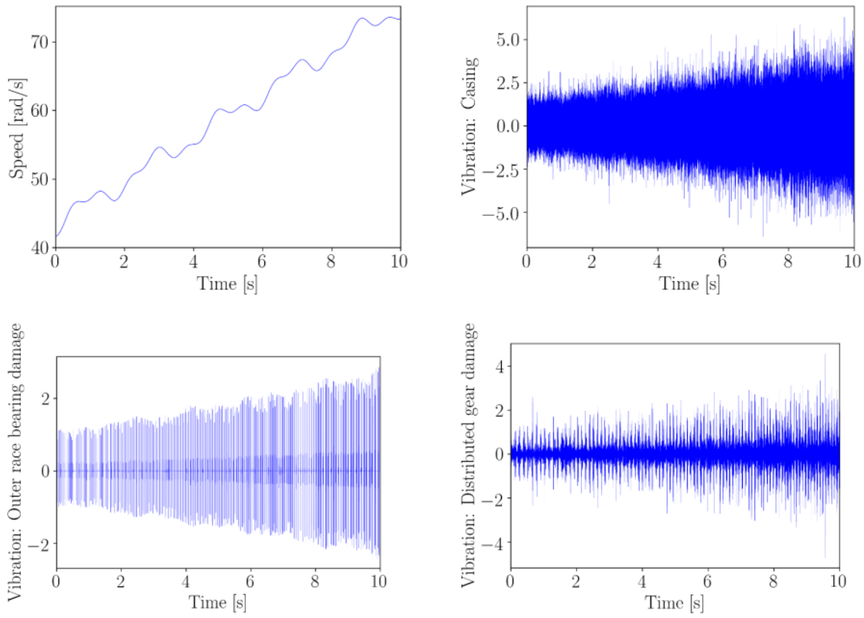


Figure 2: The speed profile, the casing vibration signal and the bearing and random gear components of the phenomenological gearbox model are presented.

The varying speed conditions result in the amplitude and the instantaneous frequency of the signal components to be dependent of time. The relative magnitudes of the components were chosen so that the dominant distributed gear damage component impedes the ability to detect the bearing component. Hence, the focus of the subsequent investigations is to highlight how the proposed method can be used to detect weak components in the presence of dominant components and to show that it is possible to distinguish between the two. In the next section the kurtogram is investigated on the generated dataset.

3.2 Application of the Fast Kurtogram (FK)

The Fast Kurtogram (FK), developed in Ref. [6], is a faster estimator of the kurtogram than the conventional short-time Fourier transform-based estimator and is used in this work. The kurtogram is based on the spectral kurtosis [5], a very useful technique to identify frequency bands that contain transient information (as typically seen by bearing and gear damage). The FK is applied to the casing vibration signal (see Equation (4)) of the PGM with the result shown in Figure 3.

The FK is maximum at a frequency band with a centre frequency of 1328.12 Hz. This is the frequency band associated with the distributed gear damage component. The frequency band of the bearing damage at 7.0 kHz can also be seen in Figure 3; however, its magnitude is significantly smaller than the magnitude of the gear component.

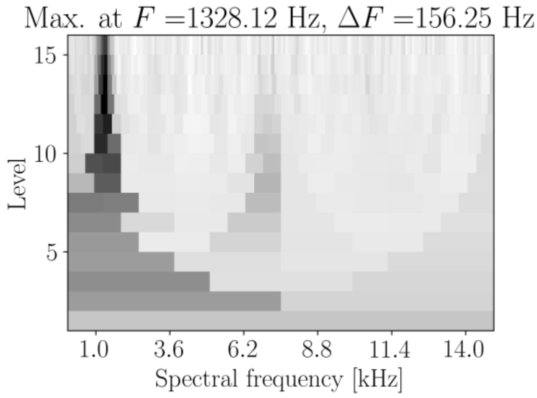


Figure 3: The kurtogram of the PGM's vibration signal.

The implication of this is that without careful consideration, only the dominant impulsive frequency band will be detected by the FK, with a non-dominant frequency band easily missed in the condition interrogation process.

This is corroborated by the results of the Squared Envelope Spectrum (SES) seen in Figure 4. The SES of the raw signal (i.e. without bandpass filtering the signal) and the SES of the filtered signal contains the same information. The fundamental component of the distributed gear damage at one shaft order and its harmonics are clearly seen in both spectra, while the bearing component is not seen.

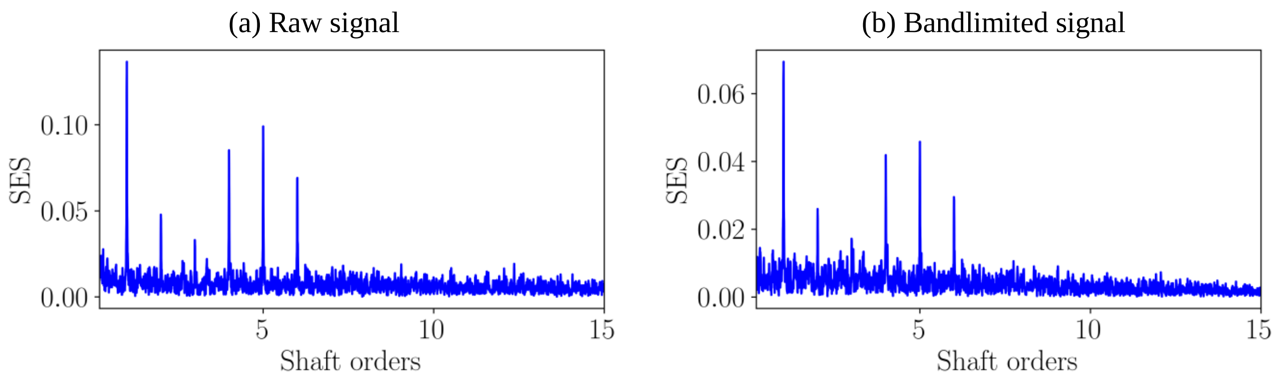


Figure 4: The Squared Envelope Spectrum (SES) of the raw vibration signal and of the bandlimited signal obtained with the Fast Kurtogram (FK) for the PGM.

It is important to emphasise that due to the statistical characteristics of the distributed gear damage component, it is not possible to remove it using cepstrum pre-whitening or the generalised synchronous average [3]. The proposed method is investigated in the next section.

3.3 Application of the proposed method

The proposed method is applied with the procedure discussed in Section 2, with the bearing and gear being monitored for damage. Therefore, the feature, calculated with Equation (3), is calculated for the gear with $\{\alpha_f\}=\{1.0, 2.0, 3.0\}$ (denoted α=1.0 in the figures) and for the bearing with $\{\alpha_f\}=\{4.12, 8.24, 12.36\}$ (denoted α=4.12 in the figures), which result in two feature planes that are maximised independently. The feature plane of the gear and the bearing are shown in Figure 5(a) and Figure 5(b) respectively.

It is evident that the feature plane is clearly very dependent on the cyclic order that is used. Large values are obtained in Figure 5(a) in the region of 1.3 kHz, while large values are obtained in Figure 5(b) in the region of 7 kHz. The optimal value for the gear in Figure 5(b) differs slightly from the analytical value, because the gear component is very dominant, which results in the different blocks to have features with very similar values, i.e. any of the blocks, could be used for detecting the gear.

The SES of the raw and bandlimited signals of the two signal components are shown in Figure 6. The SES of the bandlimited gear signal, presented in Figure 6(b), does not improve the SES of the raw signal, presented in Figure 6(a), because the gear component is already very dominant in the SES.

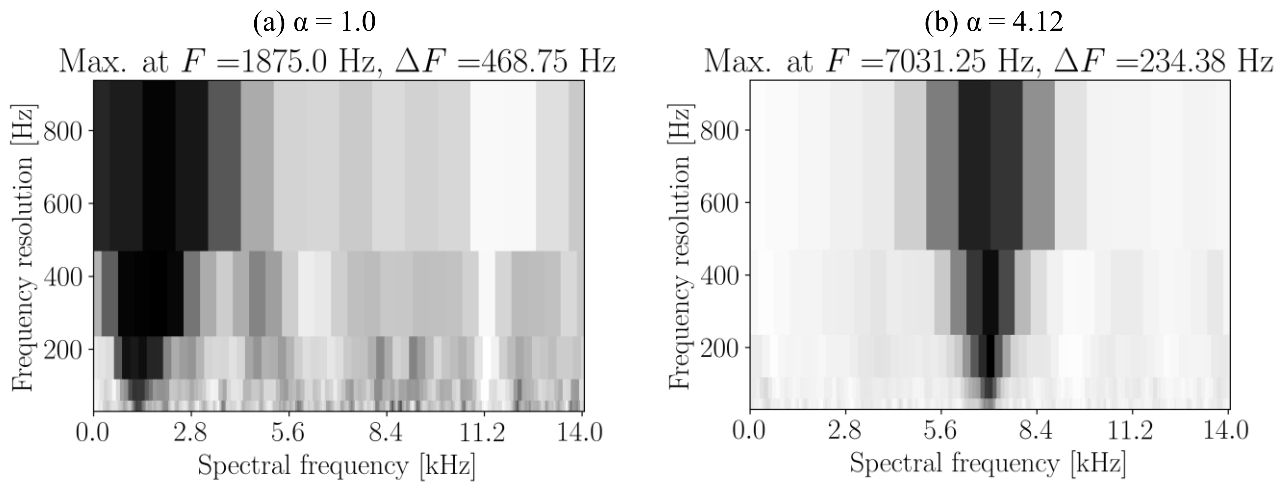


Figure 5: The feature plane obtained with the proposed method for the gear component (a) and the outer race bearing component (b) of the PGM. The colour scales are not the same in the two plots.

A significant improvement can be seen for the SES of the bearing component. The bearing component cannot be detected in Figure 6(c), but after identifying the appropriate frequency band with the proposed method, it is possible to obtain a SES that clearly highlights the damaged bearing component as seen in Figure 6(d).

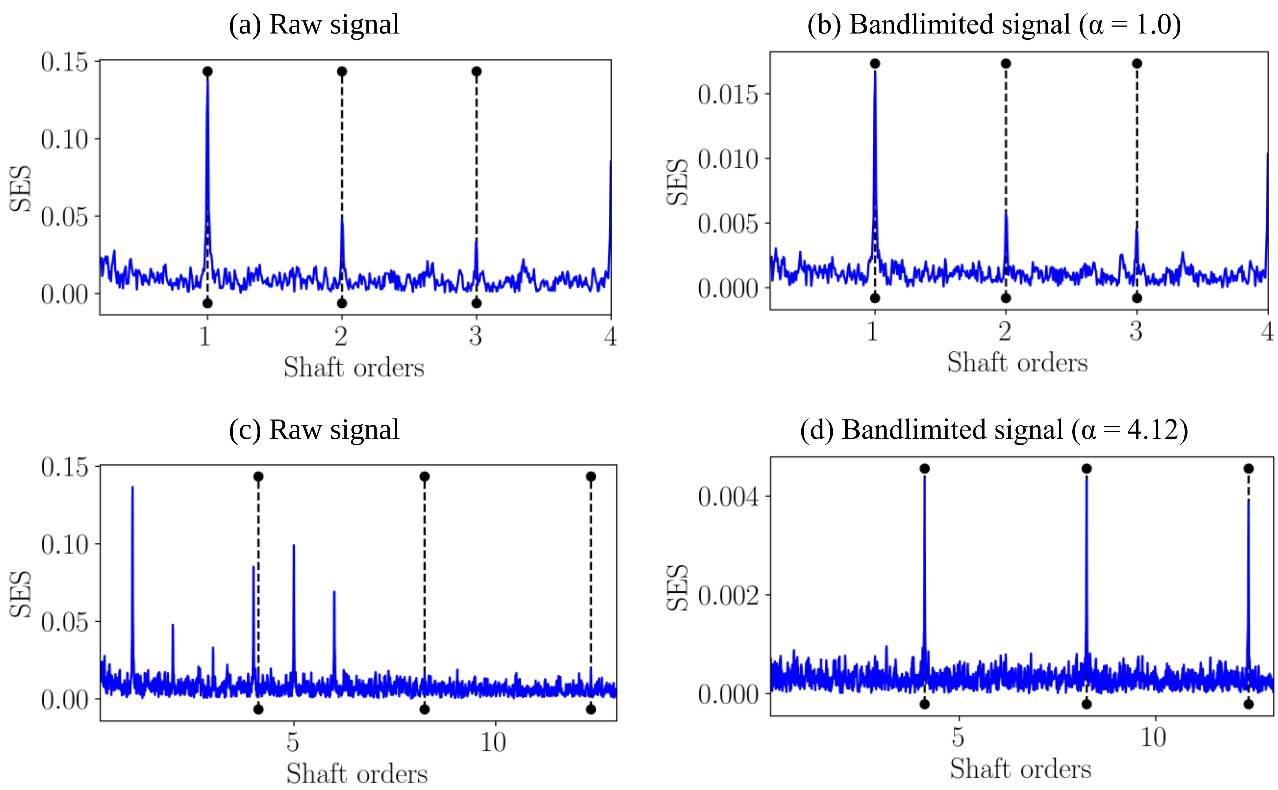


Figure 6: The Squared Envelope Spectra (SES) of the raw and bandlimited signals are shown for the gear component in (a) and (b) and for the outer race bearing component in (c) and (d) for the PGM.

This highlights the benefit of using the proposed method; if the signal component is dominant in the spectrum then the kurtogram can lead to similar results (as seen when comparing the results in Figure 4(b) and Figure 6(b)). However, the proposed method has sufficient flexibility to identify frequency bands for signals with low signal-to-noise ratios as well.

4 Experimental investigation

In this section, the proposed method is investigated on an experimental dataset. A brief overview of the experimental data is given in Section 4.1, whereafter the FK is applied to the dataset in Section 4.2 and the proposed method is investigated in Section 4.3.

4.1 Overview of the experimental dataset

The method is applied and verified in this section on an experimental gearbox dataset that has been acquired in the Centre for Asset Integrity Management (C-AIM) laboratory at the University of Pretoria. The experimental setup contains three helical gearboxes, an alternator and an electrical motor. The alternator and the electrical motor were used to induce the time-varying speed and load conditions shown in Figure 7 on the monitored gearbox. One of the helical gearboxes was damaged with the damaged gear shown in Figure 8(a) and operated for approximately 20 days whereafter the tooth failed as shown in Figure 8(b). A vibration and a tachometer measurement, taken after approximately five days of testing, are used in this paper. The gear rotates at 1.0 shaft order, while the pinion rotates at 1.85 shaft order. More information on the experimental setup can be found in Ref. [13].

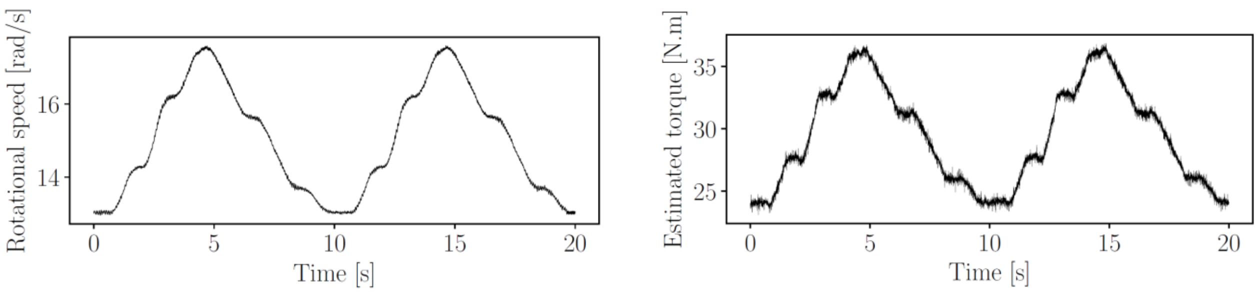


Figure 7: The operating conditions during the measurement period.

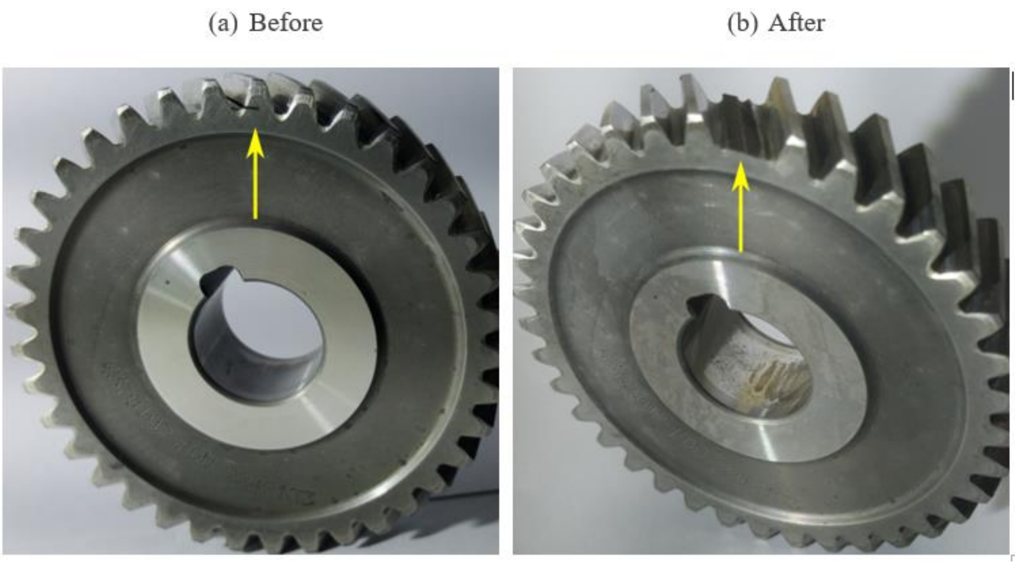


Figure 8: The gear of the helical gearbox with the seeded fault before the fatigue experiment (a) and after the fatigue experiment was completed (b).

4.2 Application of the Fast Kurtogram (FK)

The FK is applied on the dataset with the decomposition shown in Figure 9. Very large values are seen in the higher frequency bands. This is attributed to the presence of bandlimited transients that manifest at the frequency band 8-12 kHz at a cyclic order of approximately 5.5 shaft orders.

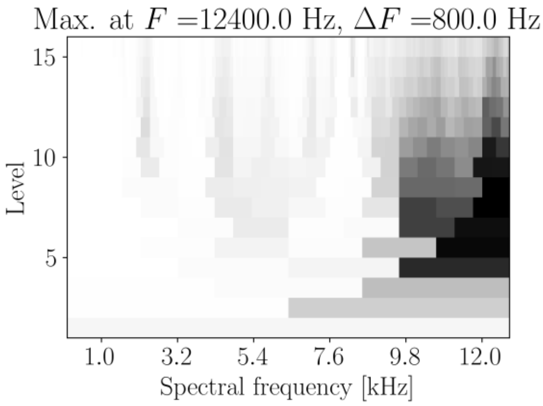


Figure 9: The kurtogram of the experimental gearbox dataset.

The SA is used to interrogate the presence of damage on the gear in Figure 10. The SA of the raw and the bandlimited signals are shown in Figure 10(a) and (b). It is not clear from the raw signal in Figure 10(a) what the condition of the gear is, but the transients that are retained by the bandpass filtering process dominate the synchronous average and make it especially difficult to infer the condition of the machine from the result in Figure 10(b).

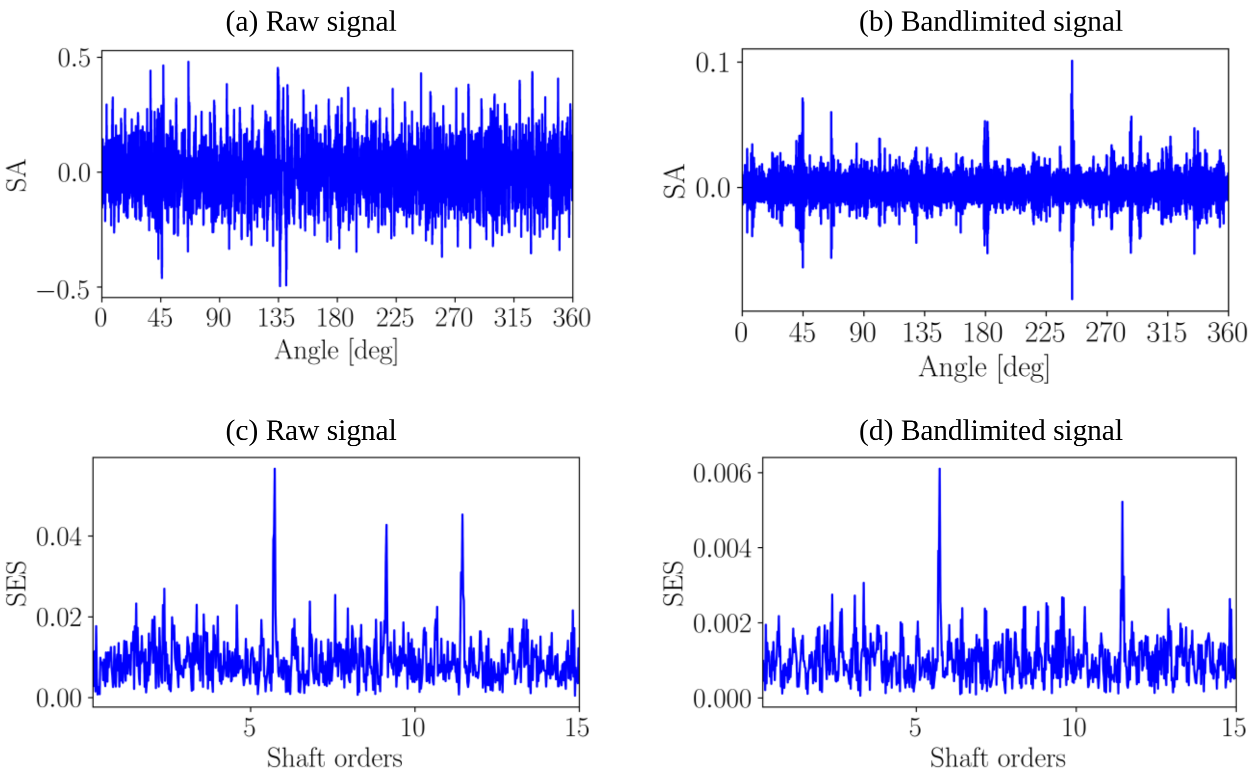


Figure 10: The Synchronous Average (SA) and the Squared Envelope Spectrum (SES of the raw and the bandlimited signals are shown as obtained with the Fast Kurtogram (FK). The damaged gear tooth is located at approximately 135 degrees in the SA plots.

The SES of the raw and the bandlimited signals are also investigated in Figure 10. Three peaks are observed in the SES of the raw signal; the components at 5.72 and 11.44 shaft orders are attributed to the transients in the signal and the component at 9.12 shaft orders is attributed to the alternator's shaft being slightly unbalanced which resulted in periodical excitations. After, the filtering process, only the transient at 5.72 shaft orders and its harmonics are retained. Hence, it is evident from the results that the kurtogram fails to recognise the important frequency band for diagnosing the gear.

4.3 Application of the proposed method

The proposed method is applied on the same signal as investigated in the previous section. The gear and the pinion are monitored and therefore the decomposition is performed for $\{\alpha_f\}=\{1.0, 2.0, 3.0\}$ shaft orders (denoted $\alpha=1$) and $\{\alpha_f\} = \{1.85, 3.7, 5.55\}$ shaft orders (denoted $\alpha=1.85$), respectively. The feature plane is shown in Figure 11 for the two monitored components, where it can be seen that the feature planes are dependent on the cyclic order of interest, however, the identified frequency bands may not necessarily be completely separated. It is completely reasonable that the same cyclic order band is optimal for different mechanical components and therefore care should be taken to interpret the statistics (e.g. kurtosis) of the bandlimited signals.

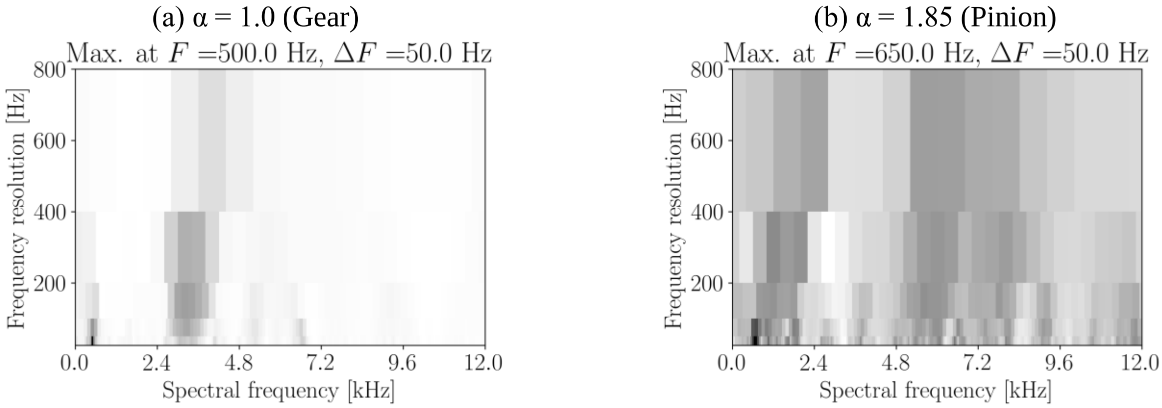


Figure 11: The feature plane obtained with the proposed method. The feature plane of the gear is shown in (a) and the feature plane of the pinion is shown in (b).

The SA in Figure 12 do not clearly reveal damage on either the gear or pinion with only small peaks seen at 135deg for the gear. This is attributed to the fact that the damage is still small and that helical gears are used with large contact ratios. Hence, the synchronous average is ineffective for detecting the incipient gear damage.

The SES of the raw and bandlimited signals in Figure 13 perform significantly better than the SA for the gear and the pinion. It is possible to see that there is a clear 1.0 shaft order component, which is attributed to the damaged gear. In contrast, the SES of the healthy pinion does not contain any dominant components at 1.85 shaft orders, which is indicative that the pinion is healthy. Hence, it is possible to use the proposed method and the SES to detect the incipient gear damage in the presence of dominant frequency components and time-varying operating conditions.

5 Conclusions

In this paper, a new method is investigated for identifying frequency bands that are rich with diagnostic information. The method uses the spectral coherence and a very carefully designed feature to allow specific frequency bands to be detected which can be analysed using the squared envelope spectrum and the synchronous average.

The method is evaluated on two datasets; the first one is a numerical gearbox dataset that simulates bearing damage and gear damage under time-varying speed conditions. The results indicate that it is possible to identify the appropriate frequency band to identify the cyclic components of interest, while the fast kurtogram only identifies the frequency band with the most impulsiveness. Similar results are obtained on the experimental dataset where incipient damage was present on the gear of a helical gearbox. The fast kurtogram maximised on frequency bands with strong impulsive content, with the incipient gear damage only detected by using the proposed method. It was also found that the synchronous average is not very effective for incipient gear damage detection and the squared envelope spectrum performs significantly better.

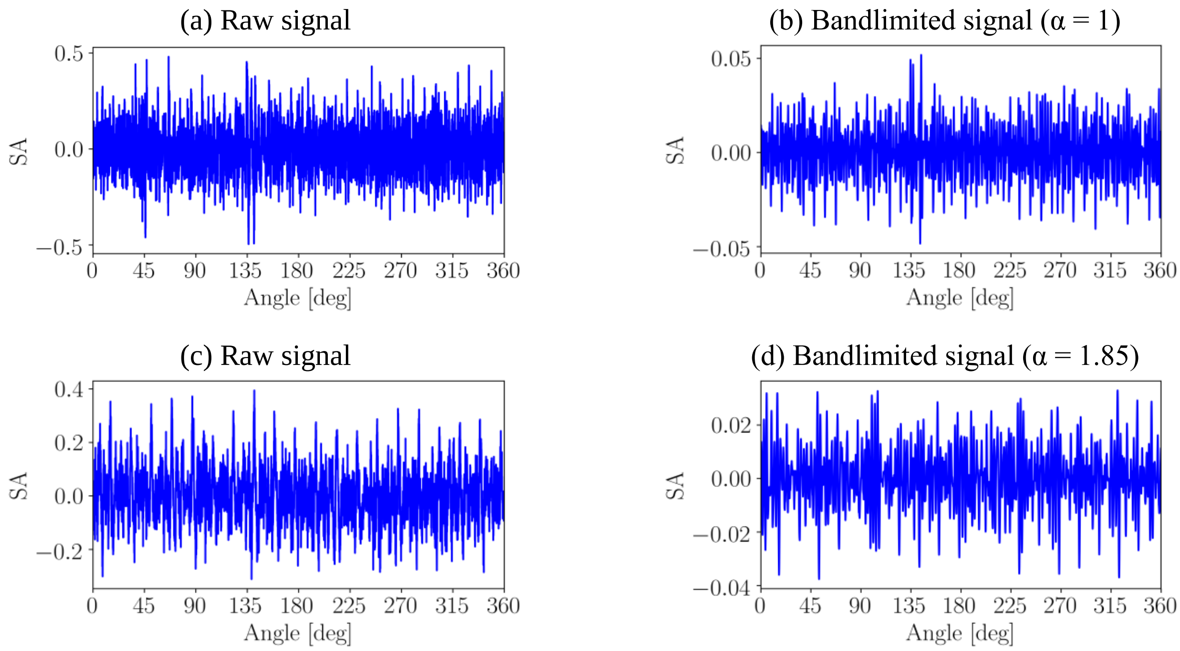


Figure 12: The Synchronous Averages (SA) of the raw and the bandlimited signals, obtained with the proposed method, are shown. The result for the gear is shown in (a) and (b), while the result of the pinion is shown in (c) and (d).

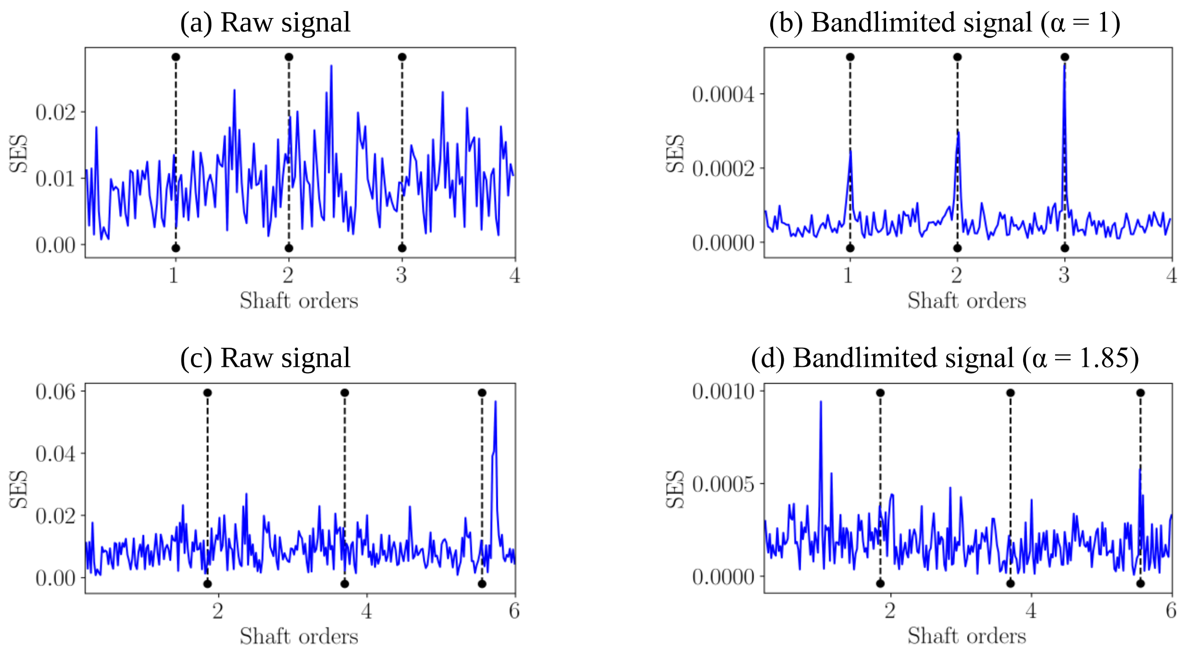


Figure 13: The Squared Envelope Spectra (SES) of the raw and the bandlimited signals, obtained with the proposed method are shown. In (a) and (b) the results for the gear are shown, while the results for the pinion are shown in (c) and (d).

In future investigations, the method will be compared to the more recent developments in the informative frequency band identification field (e.g. infogram) and the suitability of this method for fault diagnosis under time-varying operating conditions will be investigated on more datasets. It is also suggested that the spectral coherence needs to be estimated with the fast or faster spectral correlation instead of the Welch estimator used in this work. This would improve the computational efficiency of the proposed method.

Acknowledgments

The authors gratefully acknowledge the support that was received from the Eskom Power Plant Engineering Institute (EPPEI) in the execution of the research.

References

[1] Abboud, D. & Antoni, J., 2017. Order-frequency analysis of machine signals. Mechanical Systems and Signal Processing, 87, pp.229–258.

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

[3] Abboud, D., Antoni, J., Sieg-Zieba, S. and Eltabach, M., 2017. Envelope analysis of rotating machine vibrations in variable speed conditions: A comprehensive treatment. Mechanical Systems and Signal Processing, 84, pp.200-226.

[4] Schmidt, S., Heyns, P.S. & Gryllias, K.C., 2019. A pre-processing methodology to enhance novel information for rotating machine diagnostics. Mechanical Systems and Signal Processing, 124, pp.541-561.

[5] Antoni, J. & Randall, R.B., 2006. The spectral kurtosis: Application to the vibratory surveillance and diagnostics of rotating machines. Mechanical Systems and Signal Processing, 20(2), pp.308–331.

[6] Antoni, J., 2007. Fast computation of the kurtogram for the detection of transient faults. Mechanical Systems and Signal Processing, 21(1), pp.108–124.

[7] Barszcz, T. & Randall, R.B., 2009. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mechanical Systems and Signal Processing, 23(4), pp.1352–1365.

[8] Antoni, J., 2016. The infogram: Entropic evidence of the signature of repetitive transients. Mechanical Systems and Signal Processing, 74, pp.73–94.

[9] Mauricio A, Qi J, & Gryllias K. 2018. Vibration-based condition monitoring of wind turbine gearboxes based on cyclostationary analysis. ASME. Journal of Engineering Gas Turbines Power.

[10] Stander, C.J., Heyns, P.S. & Schoombie, W., 2002. Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions. Mechanical Systems and Signal Processing, 16(6), pp.1005–1024.

[11] Antoni, J., Xin, G. & Hamzaoui, N., 2017. Fast computation of the spectral correlation. Mechanical Systems and Signal Processing, 92, pp.248–277.

[12] Borghesani, P. & Antoni, J., 2018. A faster algorithm for the calculation of the fast spectral correlation. Mechanical Systems and Signal Processing, 111, pp.113–118.

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

Diagnostics and Dynamic models

Challenging the traditional model of gear vibration signals

E. Hubert[1], P. Borghesani[2], R.B. Randall[2], M. El-Badaoui[1]
[1]SAFRAN Tech, Rue des Jeunes Bois – Châteaufort, 78772 Magny les Hameaux, France
[2]School of Mechanical and Manufacturing Engineering, UNSW Sydney, NSW 2031, Australia

Abstract

Despite rarely being made explicit, signal models (and assumptions) for gear vibration have been fundamental in the development of both condition monitoring and operational modal analysis techniques. The analysis (for condition monitoring) or the removal (for OMA) of the dominant gear-meshing component is in fact dependent on the assumption on the number, location and patterns of the corresponding spectral harmonics. This paper discusses in detail the common modelling choices and their consequences for condition monitoring and OMA. The traditional gearmesh-carrier/shaft-modulated model is analysed and two main limitations of current models are highlighted: the additive assumption on the two gear modulating functions and the regularity of their effect on different gearmesh harmonics. The paper uses experimental gear signals to prove the validity of the newly introduced assumptions and to assess their practical significance.

1 Introduction

Gearbox condition monitoring has often been based on simple signal models. Empirical signal models are used as a first approximation of the vibration signal to justify and guide the development of diagnostic signal processing techniques. For this purpose, they are preferred to more detailed physical models (e.g. FEM or lumped parameters) due to their generality and ease of implementation. Empirical signal models aim at reproducing the overall properties of vibration signals, retaining the main time and frequency features that are observed in real signals and can be used for condition monitoring. Despite not requiring the fine-tuning of structural and geometric parameters typical of detailed physical models, correct assumptions on the phenomena generating the vibration are fundamental in developing appropriate empirical signal models.
In the case of spur gears, the main source of vibration comes from the time-varying meshing force generated at the contact point between the pinion and driven gear teeth. The most widely accepted gear-signal model is therefore represented as [1]

$$y(t) = h(t) \otimes g(t) \tag{1}$$

where the gear-related component $y(t)$ of a measured vibration signal results from the convolution (symbol $\otimes$) of the system impulse response $h(t)$ with the gear-meshing forcing function $g(t)$. As usual in rotating machines [2], the true nature of this signal has a hybrid time-angle definition. However, for nominally constant speed, the approximation of linearity between time and angular domain is often assumed. In this case, the easiest way to represent the signal in the frequency domain is probably to define (with approximation) the system transfer function $H(f) = \mathcal{F}\{h(t)\}$ in an equivalent shaft-order domain, adopting the approximate relationship $f \approx \Omega f_1$ where f_1 is the average (and almost constant) shaft speed of a reference gear (in this study we will always use the pinion/input shaft) and Ω is the order coordinate of the same shaft. In this case we can rewrite eq. (1) in the order domain as:

$$Y(\Omega) = H(\Omega f_1) \cdot G(\Omega). \tag{2}$$

In the case of perfect and healthy teeth, the contact force $g(t)$ is theoretically expected to show a gearmesh fundamental frequency (i.e. for a pinion with Z_1 teeth $G(\Omega) \neq 0$ only for $\Omega = kZ_1$), but in practice even imperfections in the manufacturing stage result in tooth-to-tooth variations. These variations are expected to be even more accentuated in the case of localised gear faults (e.g. tooth crack) and are modelled as gear-synchronous modulations of the gearmesh harmonics.

The excitation g is usually modelled in time or in the angular domain θ of the reference shaft (in this case shaft 1) as an amplitude/frequency modulated signal, where the carrier is represented by the dominant gearmesh harmonics and the two modulating functions are synchronous with the two shafts. Actual explicit mathematical expressions of AM/FM signal models are rare, and many studies focus on the simpler AM case only. In this paper we will mainly focus on the implications that arise for condition monitoring and OMA in considering an AM/FM model. A full analytical discussion will be provided for AM-only models, including limitations of common assumptions and further issues encountered when dealing with actual signals. However, considerations on FM and AM/FM models will be provided, without the explicit formulation of full AM/FM signal models (due to their cumbersome expression), but keeping in mind all the major and minor spectral components arising from the combination of all modulating functions.

2 Secondary sidebands

The simplest model of the gear force (often implicitly considered in many condition monitoring studies) includes two purely amplitude modulation components:

$$g(\theta) = c(\theta) \cdot \{a(\theta) + b(\theta)\} \tag{3}$$

where:

- $c(\theta)$ is the gearmesh-periodic dominant effect of the tooth-meshing

$$c(\theta) = \sum_h C_h e^{jhZ_1\theta} \tag{4}$$

with Z_1 representing the number of teeth of the gear on shaft 1,

- $a(\theta)$ and $b(\theta)$ are a shaft-periodic amplitude modulation functions due to irregularities among the teeth of shaft 1 and 2 respectively (and/or geometric/misalignment issues on the same shaft)

$$a(\theta) = \sum_k A_k e^{jk\theta} \quad \text{and} \quad b(\theta) = \sum_k B_k e^{jk\tau\theta} \tag{5}$$

where $\tau = Z_1/Z_2$ is the gear ratio.
Such AM signal is usually represented as:

$$g(\theta) = \sum_h \sum_k C_h A_k e^{j(hZ_1+k)\theta} + \sum_h \sum_k C_h B_k e^{j(hZ_2+k)\tau\theta}. \tag{6}$$

This formulation shows the main feature of AM gear models: the presence of sidebands around the gearmesh harmonics at orders $hZ_1 + k$ (effect of shaft 1) and $hZ_1 + \tau k$ (effect of shaft 2).

Such a simplified model already poses risks for OMA. In fact, the removal of these harmonics is often considered straightforward by means of established synchronous averaging techniques, using encoders installed on both shafts (or at least on a reference shaft). In most cases, shaft-1 and shaft-2 sidebands are thus removed separately by synchronous averaging over the respective periods. Even in a number of highly rigorous approaches, only one or a few combined periods of the two gears are used for these synchronous averaging operations. This “grand-period” is defined as the interval between the meshing of the same tooth pair, and equivalent to Z_2 periods of shaft 1 or Z_1 periods of shaft 2.

However, considering the physical nature of the AM functions $a(\theta)$ and $b(\theta)$, a multiplicative model is much more justified, i.e. there is no reason why carrier and modulations should be treated differently and a three-term multiplication is more appropriate. This results in the modification of eq. (3) into the following:

$$g(\theta) = c(\theta) \cdot a(\theta) \cdot b(\theta) \tag{7}$$

with a consequent proliferation of sidebands:

$$g(\theta) = \sum_h \sum_k \sum_\ell C_h A_k B_\ell e^{j(hZ_1 + k + \tau\ell)\theta}. \tag{8}$$

Under this modelling assumption – and actually also for model (3) –, the fundamental period of the signal is the "grand-period". However, differently from model (3) the signal shows a vast number of secondary sidebands $hZ_1 + k + \tau\ell$, with $k, \ell \neq 0$ (in addition to the $\ell = 0$ and $k = 0$ primary sidebands present also in the previous model). These secondary sidebands are expected to have (in the spectrum of the excitation) a lower amplitude (as the zero-frequency component of the modulation signals must be dominant to ensure positive-only modulating functions), but they could still significantly compromise OMA attempts based on the assumption that noise dominates the spectrum of the vibration signal, once the primary sidebands are removed.

An experimental test to verify the extent of this effect has been carried out on the UNSW spur-gear test-rig. The test-rig is composed of a speed-reducing spur gear pair ($Z_1 = 27$ and $Z_2 = 44$) powered by an electric drive and connected to a magnetic particle brake. The set of gears (module 2 with 5 mm face width) are built in mild steel, and surface hardened. A gear crack was simulated by means of an artificial slot on the pinion starting at the base of the tooth and reaching the centreline with an angle of 45°. The test rig was operated with constant speed and load (10 Hz /10 Nm on the input shaft) and a vibration signal was measured by means of a B&K4396 accelerometer, installed on the top of the casing in proximity of the DE input-shaft bearing. The signal was sampled at a rate of 100 kSamples/s for a duration of 101 s, sufficient to ensure the observation of 22 "grand-periods". In addition, a phase-reference signal was obtained synchronously to the vibration signal, thanks to an encoder with 1000 pulses/rev installed on the NDE of the pinion shaft.

The vibration signal was order-tracked ensuring that an integer number of samples N_{GM} was taken within a gearmesh period, thus also ensuring integer numbers of samples $Z_1 N_{GM}$ in a revolution of shaft 1 and $Z_2 N_{GM}$ in a revolution of shaft 2. Residuals were obtained following two synchronous averaging (SA) procedures: the first was the traditional approach which removed all harmonics of shaft 1 and shaft 2 (primary sidebands of the gearmesh harmonics), whereas the second used the entire grand-period as reference for the synchronous averaging, thus removing all primary and secondary sidebands.

The results are shown in Figure 1. The raw order-tracked spectrum clearly shows a significant amount of gearmesh harmonics and sidebands, which are only partly removed by the traditional SA approach. Whereas the low-frequency range (up to 10-15 gearmesh harmonics) seems mostly unaffected by the presence of secondary sidebands, in the range from 15 to 30 gearmesh harmonics the removal of the primary harmonics still leaves a large quantity of discrete components, which are identified as secondary sidebands.

(a)

(b)

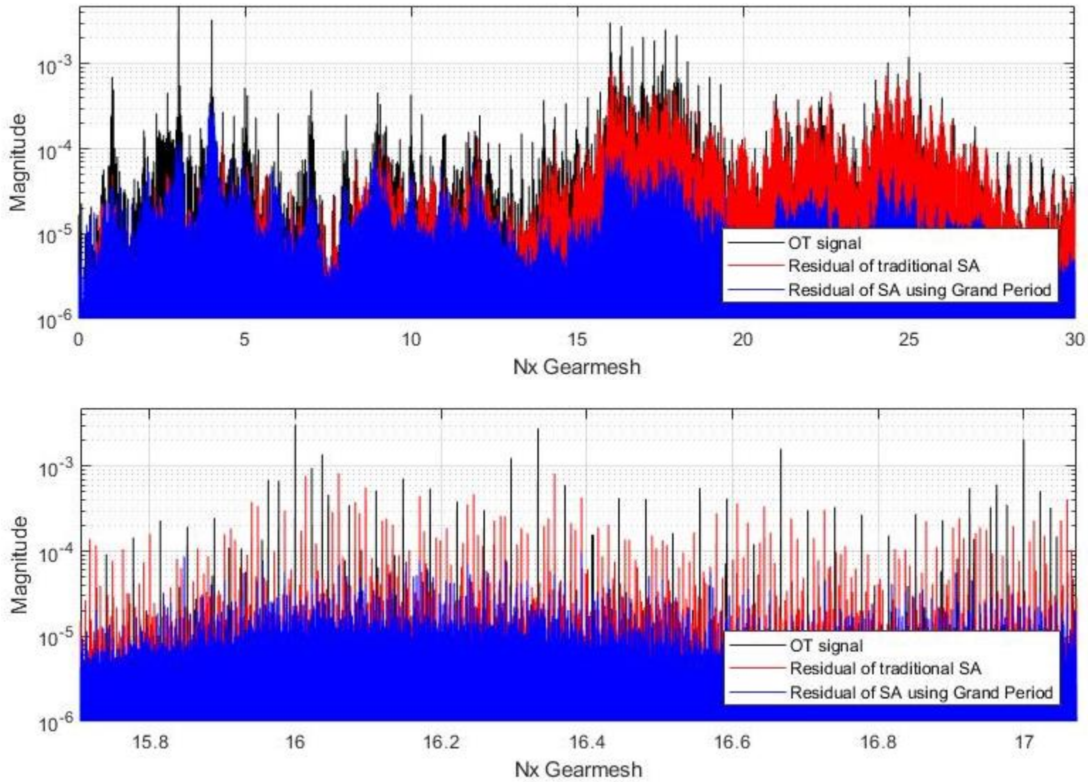


Figure 1. Result of the removal of primary-only (red) and primary + secondary (blue) sidebands from a gear vibration signal (black): (a) frequency range 0-30 gearmesh harmonics, (b) zoom of the most affected area.

Adding frequency modulation to the model, the considerations made so far become even more relevant. Even taking pure frequency modulation, the signal consists of

$$g(\theta) = c\big(\theta + \phi(\theta) + \psi(\theta)\big) \tag{9}$$

where $\phi(\theta)$ and $\psi(\theta)$ are the phase modulations introduced by shaft 1 and 2, respectively:

$$\phi(\theta) = \sum_k \Phi_k e^{jk\theta} \quad \text{and} \quad \psi(\theta) = \sum_k \Psi_k e^{jk\frac{Z_1}{Z_2}\theta} \tag{10}$$

This signal can be expressed as a Fourier Series as:

$$g(\theta) = \sum_h C_h e^{jhZ_1(\theta+\phi(\theta)+\psi(\theta))} = \sum_h C_h e^{jhZ_1\left(\theta+\sum_k \Phi_k e^{k\theta}+\sum_k \Psi_k e^{jk\frac{Z_1}{Z_2}\theta}\right)} \tag{11}$$

The Bessel expansion of such signal (too cumbersome to report in this paper and whose details are of little significance) is composed of a large series of harmonics, at all the multiples of the fundamental frequency obtained from the "grand period". Moreover, whereas the bandwidth of AM sidebands is expected to keep constant for each carrier harmonic, the bandwidth of FM sidebands grows proportionally to the harmonic order of the carrier [3], potentially amplifying this phenomenon at high orders.

A detailed analysis of the motivation (AM, FM or mixed) of the high-frequency location observed for strong secondary side-bands is outside the scope of this paper, but a preliminary sensitivity analysis (varying shaft speed) indicates that the phenomenon has a stable location in the frequency domain (independent of speed), rather than in the order domain. This suggests that the relevance of the secondary harmonics is linked to a dynamic amplification due to the system transfer function, rather than an FM bandwidth problem.

3 Irregularity of the sideband patterns

According to the AM model of eq. (3), each sideband-pattern should repeat identically at each carrier harmonic. This is simply explained by the convolutive nature of the spectrum of an amplitude modulated signal. For instance, simply dividing each set of sidebands $G(hZ_1 + k) = C_h A_k$ by the corresponding carrier harmonic C_h, the following equivalence should be obtained:

$$\frac{G(hZ_1 + k)}{C_h} = \frac{G(h'Z_1 + k)}{C_{h'}} \ \forall h, h' \in \mathbb{Z} \tag{12}$$

This ideal property is explicitly at the basis of Ref. [4], which proposed a multi-carrier demodulation method, but partially and implicitly adopted by most studies which arbitrarily use the first or second harmonic for demodulation.

This concept is challenged in this section using a vibration signal measured on the same gearbox as discussed in the previous section, albeit with healthy 20 mm face width gears and a transmission ratio of 19/52. The test was operated at 20 Hz with a load of 20 Nm (all measured on the pinion/input shaft). The sampling frequency was set at 100 kHz for a total duration of the acquisition of 10 s.

The order-tracked signal was split into frequency bands corresponding to the neighbourhood of the first 4 gearmesh harmonics and the different sidebands patterns were shifted to overlap with each other. All sideband amplitude coefficients were divided by the corresponding gearmesh harmonics amplitude coefficient, in order to compute an amplitude ratio. The amplitude ratio of those patterns is reported in Figure 2.

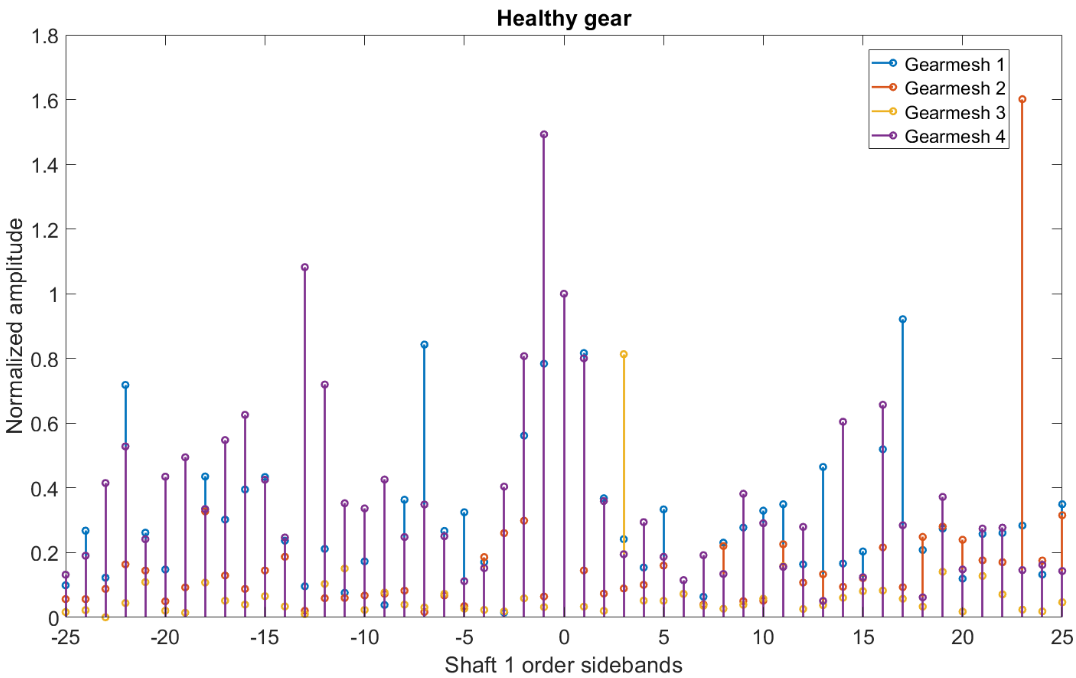
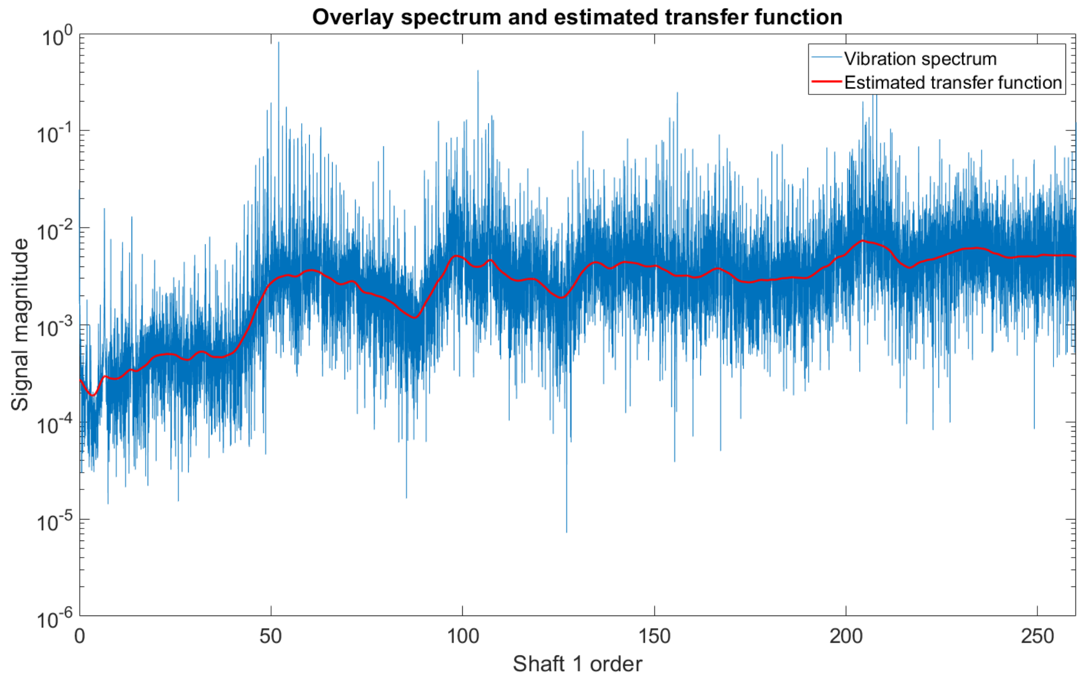


Figure 2. Sideband amplitude patterns of the first 4 gearmesh harmonics. 20 Nm - 20 Hz test.

This result shows how the patterns are massively different even disregarding their phase, which should also coincide after normalisation by the carrier harmonics. Two possible explanations for such behaviour were suggested: amplifications due to the system transfer function, or dominant frequency-modulation effects.

The first option was further investigated trying to remove the transfer function effect by means of cepstral liftering [5], [6]. An exponential lifter was applied, with a cut-off angular quefrency of 0,7 radians,to the original spectrum and the result was used to remove the short-quefrency transfer function effects. The result of the liftering operation and the “normalised” spectrum are shown in Figure 3 (a-b) respectively.

(a
)



(b
)

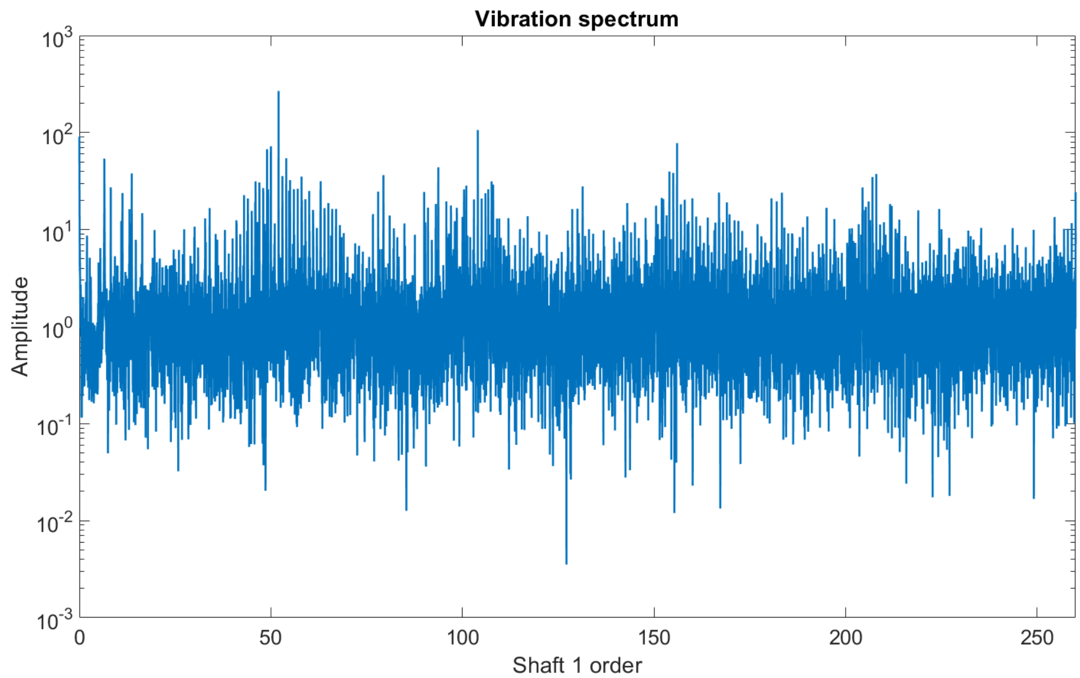


Figure 3. Removal of the TF effect: (a) short-pass liftering operation, (b) normalised spectrum.

Despite the good result in terms of spectral liftering, the problems observed in Figure 2 continue to be as severe in the normalised spectrum harmonics reported in Figure 4.

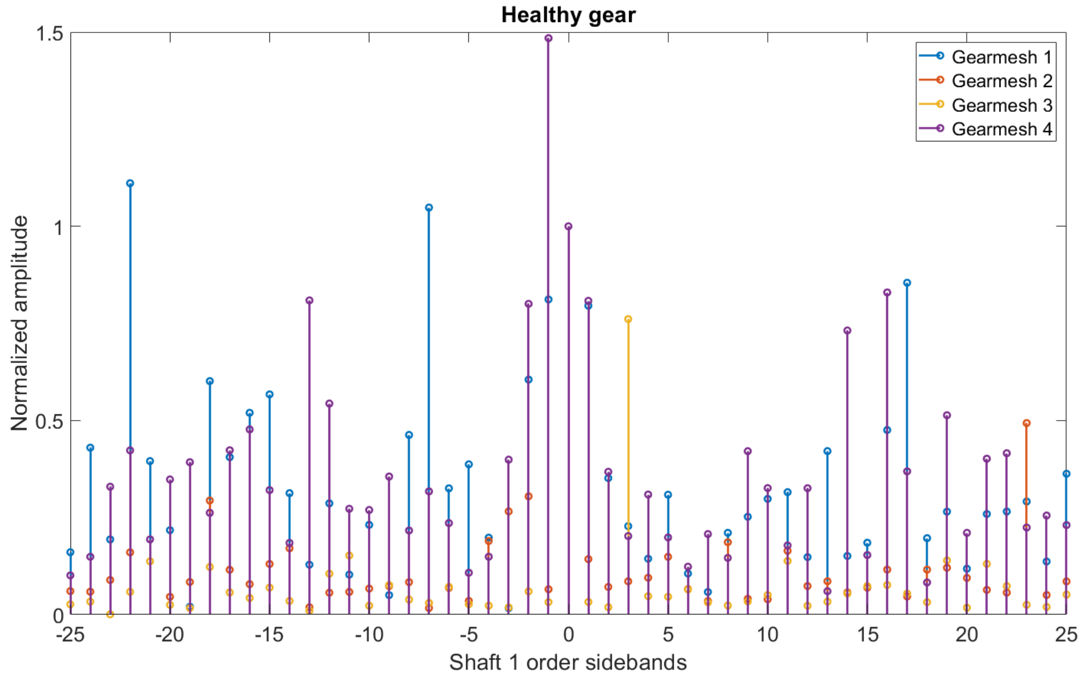


Figure 4. Sideband amplitude patterns of the first 4 gearmesh harmonics after cepstral liftering. ~1.5 Nm - 20 Hz test.

As the shape of the FM patterns of a multi-carrier modulated signal do not seem to match with the ones observed in Figure 2 and Figure 4, the authors are of the opinion that, despite possibly contributing to the difference among the sideband patterns, other phenomena (unmodelled in the current approach) must be influencing the vibration signal. FM multi-carrier modulation in fact usually results in sideband patterns showing a similar "shape", yet with a bandwidth proportional to the carrier harmonic order.

A possible explanation of the differences in the sideband patterns could be found in the different roles played by two different root-cause mechanisms resulting in gear vibration: geometric and static transmission error (each potentially resulting in a separate AM/FM modulated signal, with different carrier and modulation). The first is due to profile irregularities, whereas the second is due to the angular dependence of the gear-meshing compliance under load. An additional test was therefore executed at very low load (~1.5 Nm, just enough to maintain contact between the gear teeth) and speed (2 Hz), where geometric transmission errors were expected to dominate. Since under these operating conditions significant electrical noise was present in the lowest frequency range, in this case harmonics 2-5 were analysed.

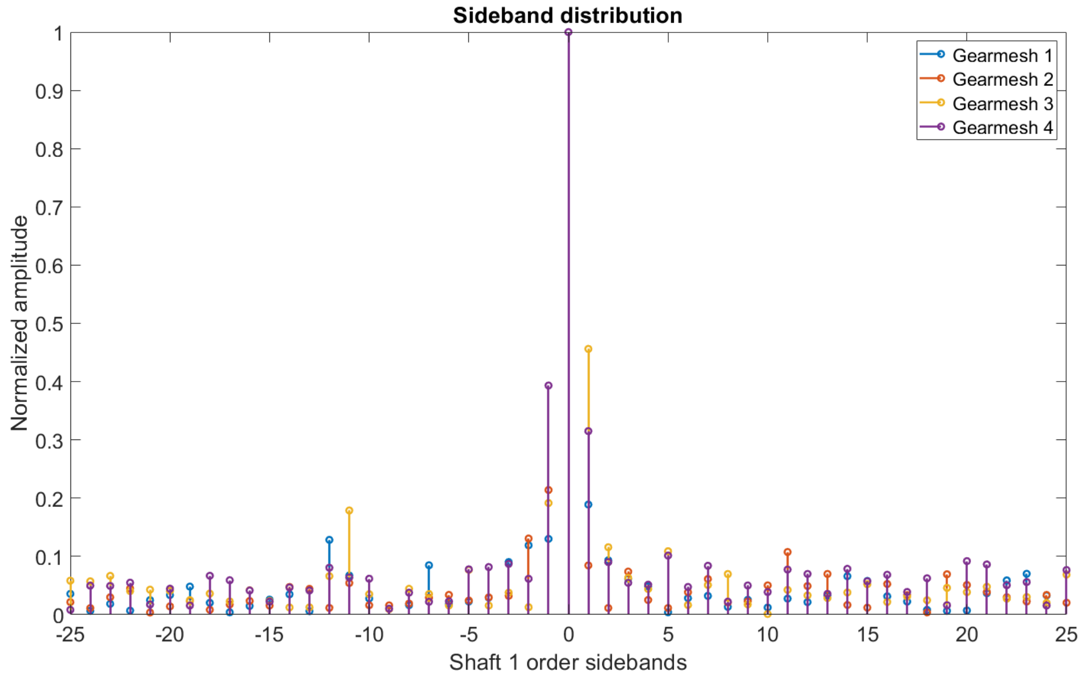


Figure 5. Sideband amplitude patterns of the first 4 gearmesh harmonics. Low load – low speed test.

The patterns shown in Figure 5 are much more consistent, even if discrepancies are still present, thus supporting the idea of a potential two-mechanism root-cause of the observed pattern inconsistency.

In order to investigate more deeply the origin of the pattern inconsistencies observed in the vibration signal, the sideband distribution of the transmission error signal is also studied, in a low-speed and low-load test. As illustrated in Figure 6, the pattern distributions have been plotted for the two cases of a heathy and a faulty gear. The transmission error is computed as the relative difference between the rotation of the input and output shafts. The pattern shown for the sidebands of the healthy gear is almost as consistent as that obtained with the vibration signal, although differences persist in the amplitudes of the sidebands. However, in the case of the faulty gear, the distribution of the sidebands is quite similar for all the sidebands of every gearmesh harmonic.

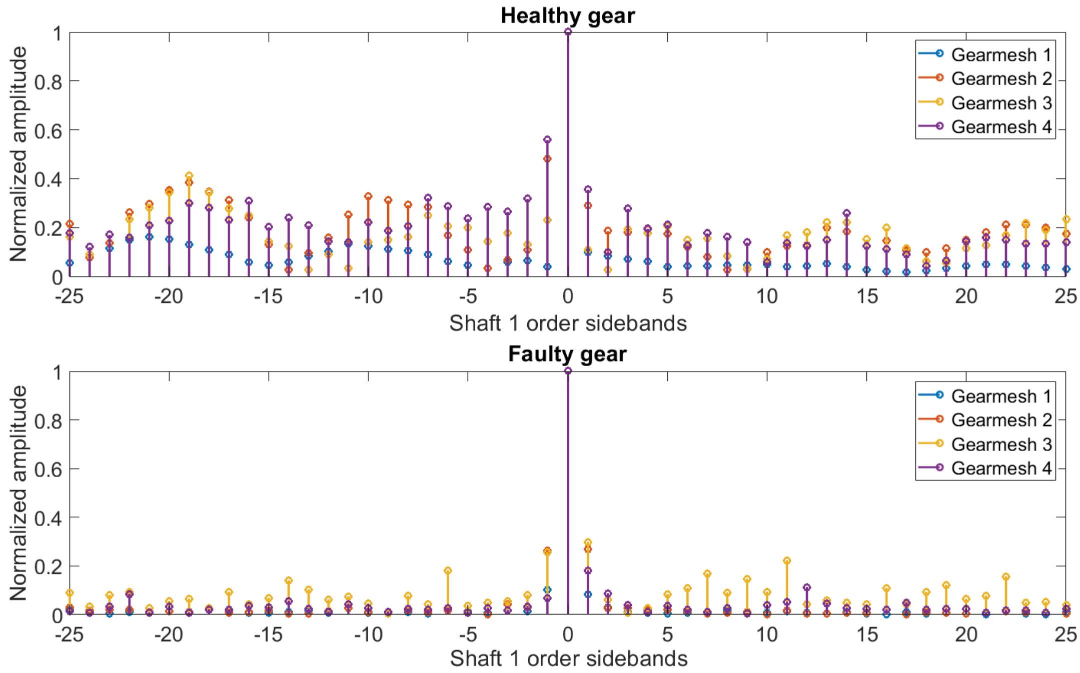


Figure 6: Sideband amplitude patterns of the first 4 gearmesh harmonics of the TE signal. Low load – low speed test.

4 Conclusions

This study has highlighted two major limitations in the current modelling (and assumptions) of gear vibration signals. Neglecting secondary sidebands has been shown to be not always reasonable, and strong discrete components were still observed in the spectrum of a gear signal after removing the primary sidebands by means of traditional synchronous averaging procedures. This problem, which could bias the identification of system transfer functions with OMA approaches, is easily solved if it is possible to observe a sufficient number of "grand-periods". In practice this could be possible for a series of machines operating at reasonably constant speed, but might be impractical for complex transmissions with more than one stage or planetary arrangements (very long grand-period).

Experimental evidence also casts doubts on the validity of AM and even AM/FM models of gear vibration signals, and suggests the possibility of multiple forcing functions (with different spectral distribution) acting simultaneously to create complex modulations. In particular, the geometric vs static transmission error components seem to be potential candidates for future investigations.

Significant investigative efforts are required to clarify these issues and give rise to more reliable models, in turn enabling new and more effective condition monitoring and OMA approaches.

[1] W. D. Mark, "Analysis of the vibratory excitation of gear systems: Basic theory," *The Journal of the Acoustical Society of America*, 2005.
[2] D. Abboud and J. Antoni, "Order-frequency analysis of machine signals," *Mechanical Systems and Signal Processing*, 2017.
[3] P. Borghesani, P. Pennacchi, R. B. Randall, and R. Ricci, "Order tracking for discrete-random separation in variable speed conditions," *Mechanical Systems and Signal Processing*, vol. 30, 2012.
[4] E. Hubert, A. Barrau, and M. E. Badaoui, "New Multi-Carrier Demodulation Method Applied to Gearbox Vibration Analysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 2141–2145.
[5] Y. Gao and R. B. Randall, "Determination of frequency response functions from response measurements - I. Extraction of poles and zeros from response cepstra," *Mechanical Systems and Signal Processing*, 1996.
[6] R. B. Randall, B. Peeters, J. Antoni, and S. Manzato, "New cepstral methods of signal pre-processing for operational modal analysis," in *{P}roceedings of the {I}nternational {C}onference on {N}oise and {V}ibration {E}ngineering {ISMA 2012}*, 2012.

Detection sensitivity study of local faults in spur gears based on realistic simulations

Lior Bachar, Ido Dadon, Renata Klein, Jacob Bortman

Keywords: Gear Model, Vibration Signature, Condition Monitoring, Local Tooth Fault, Spur Gear, Gear Diagnostics

Abstract:

The dynamic response of gear transmissions holds essential information for the recognition of an incipient fault and its propagation. A realistic and validated dynamic model is used to predict the vibration regime of gear transmissions [1]-[2]. This model was validated experimentally for both healthy and damaged conditions [1]. A great virtue of a model is the ability to examine each phenomenon separately and to isolate its contribution to the dynamic response. The model considers the nonlinear behavior of the gear mesh stiffness, integrating the geometric profile errors of the gears. The scattering in the data, which is generated by the random factor of the simulated surface roughness, simulates the reality better than data of an ideal profile. The ability to determine what is possible to monitor for each surface roughness is not trivial and cannot be achieved experimentally, due to the immense span of cases to consider.

This work presents an analysis of spur gears transmissions that can be separated into two integral but still different studies. The first study examines the effect of the operating conditions, including speed load and surface roughness, on the vibration signature of a healthy gearbox. The two main evidences from this study are related to the levels of the gear mesh frequencies (GMF) and to the sidebands (SB's) in the spectra, which are caused by the frequency modulation (FM) of the rotational speed. It was found that there is a strong dependency of the energy at the gear mesh frequency on the applied load. Figure 1.a presents the total GMF's energy for different rotational speeds (R1 is the lowest speed, R3 is the highest speed) under different loads (L1 is the lowest load, L4 is the heaviest load). It is noticeable that under the same rotational speed, the total GMF's energy sharply rises as load increases. On the other hand, it was found that the total spectral energy of the FM sidebands sharply rises as speed increases, but is not affected by load. Figure 1.b shows the spectral energy of the FM SB's for different speeds under different loads around the first six harmonics of the GMF. As for the influence of the surface roughness [3], it was found that a coarser surface roughness tends to obscure the effects of load and speed on the signature due to the gears profile error. Furthermore, the energy level of the FM sidebands sharply rose as the surface roughness got coarser, while the energy level of the GMF's barely changed due to the surface roughness.

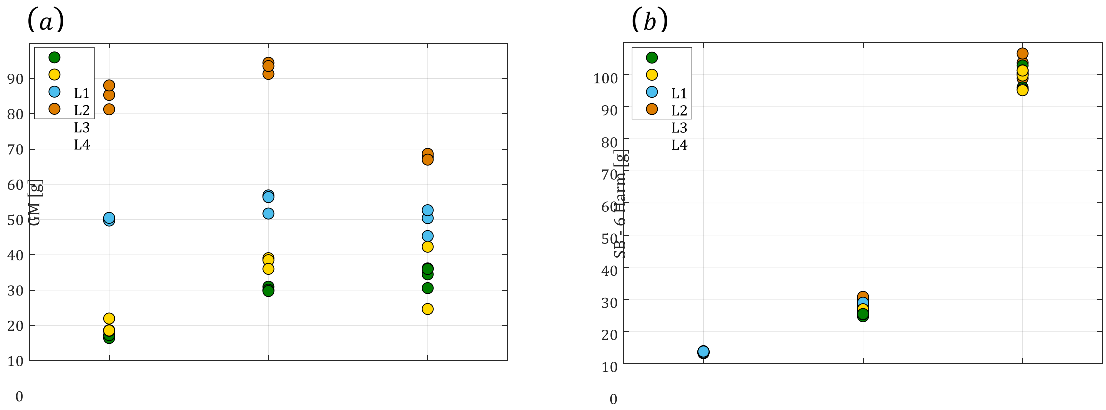


Figure 1: total energy against rotational speed and load of the (a) GMF's (b) FM sidebands around the first six GMF's

The second study examines the expression of local tooth faults in the vibrations signature. Load effects and other AM phenomena including eccentricity and misalignment may obscure the expression of local tooth faults. The comprehensive study of the effects of the operating conditions on the signature was necessary in order to fit a robust and sensitive monitoring process for the local faults detection capability. The optimal process should reflect the expression of the fault in the signature, while extinguishing the effects of the operating conditions, which are not related to the fault itself. The difference signal removes from the synchronized vibrations signal the GMF's components that are strongly affected by load, and components which are related to AM phenomena. Hence, the difference signal let us focus on the fault expression while diminishing the effects of the operating conditions. Figure 2.a shows the RMS level against kurtosis, both of the difference signal, for five different fault severities (where "Fault 1" is the least severe fault and "Fault 5" is the most severe fault). It can be seen that a separation of the faulty conditions from the healthy condition can be achieved for most faults severity levels, due to the significant differences in their locations on the graph. Besides the analysis of the difference signal, we can also utilize the total spectral energy of the FM sidebands. For each random signature, the total energy of the FM sidebands can be calculated and be compared to the healthy condition by statistical distances. The statistical distance may determine whether the examined signature can be attributed to the healthy population or not, meaning that a separation of the faulty condition from the healthy condition may be achieved. It was found that the detection capability is clearer when examining the spectral energy of the FM sidebands around the gear mesh frequencies overlapping the natural frequencies of the gearbox. Figure 2.b presents the Mahalanobis statistical distances (D) of the total FM sidebands energy around the GMF harmony which was found to be the most affected by the natural frequencies of the gearbox, against the level of the fault severity level. A Mahalanobis distance of $D = 10$ was determined to be the threshold for excluding a signature from the healthy population [4]. It can be seen that a separation of most of the local faults was achieved, as well as ranking of the three most severe faults.

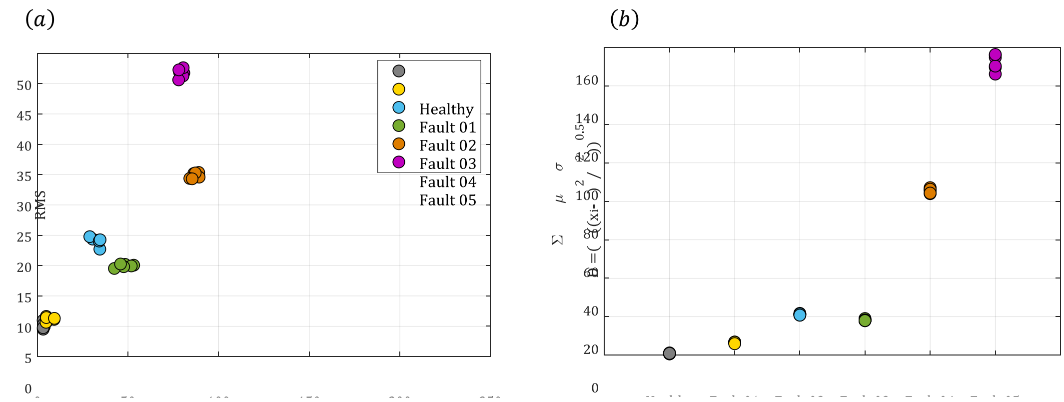


Figure 2: (a) RMS Vs. Kurtosis of the difference signal (b) Statistical distance of FM sidebands around the harmony of the GMF which was found to be the most affected by natural frequencies

References

[1] Dadon, I., Koren, N., Klein, R., & Bortman, J. (2018). A realistic dynamic model for gear fault diagnosis. Engineering Failure Analysis, 84, 77-100.

[2] Koren, N., Dadon, I., Bortman, J., & Klein, R. (2017). Steps towards fault prognostics of gears. International Journal of Condition Monitoring, 7(1), 10-15.

[3] Dadon, I., Koren, N., Klein, R., & Bortman, J. “The Effect of Gear Tooth Surface Quality of Diagnostic Capability .” Surveillance 9, 2017, surveillance9.sciencesconf.org.

[4] Klein, R. (2012, September). Condition indicators for gears. In Annual conference of prognostic and health management society (pp. 1-8).

Towards a better understanding of helical gears vibrations – dynamic model validated experimentally

Silverman, N; Dadon, I; Bortman, J; Klein, R

In order to simulate the vibration signature of gears, an accurate calculation of the gear mesh stiffness (GMS) is required. The time varying GMS, which is the main excitation that determines the dynamic response of transmissions' vibrations, is well understood for spur gears, but that of helical gears was less investigated. Although there is work dedicated to helical gears vibrations, a comprehensive analysis of their GMS compared to spur gears and their time and spectral domains have yet to be made. This paper deals with the dissimilarities and provides a better understanding of helical gears behavior, as they are a key component in many complicated and costly machines. With this new knowledge a more educated approach to diagnostics might be achieved.

The main difference between spur and helical gears is in the contact line pattern. In spur gears the contact line is parallel to the tooth's base and so calculating the GMS in any given moment is rather easy. Helical gears on the other hand have a diagonal line of contact which makes the moment applied by the meshing gear in respect to the tooth's base change along the tooth's width. To overcome this challenge a 'multi slice' method is utilized [1-4], in which the helical tooth is divided into many infinitesimally narrow slices which are treated each as a spur tooth. The total helical tooth stiffness is the sum of all those spur slices.

For the purpose of simulating the vibrations of helical gears a fourteen degree of freedom spur-teeth dynamic model [5] was upgraded to include helical gears a well. The dynamic equations and stiffness calculation were not changed and thus are discussed only briefly. The focus is dedicated to the modeling of the contact line using the multi slice method and other adjustments made to the model.

The challenge with the slice method is determining how many slices are in mesh at every given time, along with determining their mesh "height" (distance from the tooth's base). The solutions found in the literature are rather complicated and require knowing niche data about the gears, such as the transverse operating pressure angle, which are often not provided by the manufacturer. In contrast, the method suggested in this work is based on only a few common parameters such as the gears module, number of teeth and the involute profile.

The model was validated by an experiment conducted on a helical gearbox and recorded with a tri-axial accelerometer. The signals were compared in terms of their load and RPM dependency and exhibited a similar behavior, as can be seen in Figure 1. After obtaining a healthy baseline, a broken tooth case with three severity levels was studied. The fault severities were removal of 25% of the tooth's width in a diagonal line, removal of 50% and a missing tooth (Figure 1). This kind of diagonal material removal was chosen because when helical teeth breaks it happens in a pattern parallel to the contact line. The light and medium fault were challenging, but the missing tooth was seen clearly, mainly in the Kurtosis and Crest Factor of the difference signal. The statistical distances of the SA spectrum around the first and fifth GM harmony proved to provide better sensitivity, and showed clear detection of various fault severities, mainly in the tangential direction. A calculation of the Z-score index around the first GM even showed capability of ranking by fault severity (Figure 2).

Figure 1: Three level of fault severity. Left to right: removal of 25% of the tooth's width, removal of 50%, and a missing tooth.

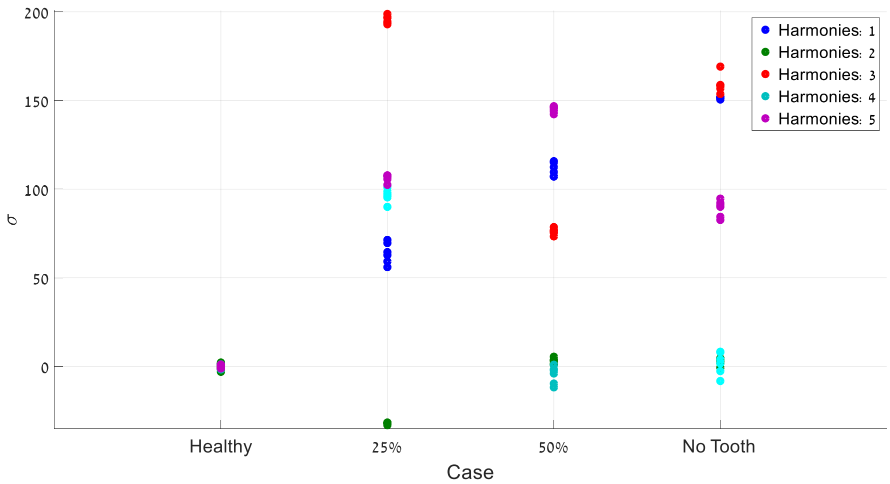


Figure 2: The Z-Score index for each GM harmony. Notice the first harmony which shows fault ranking and the fifth, which shows detection even at the smallest fault.

Keywords: Helical gears, Dynamic model, Multi-slice method, Broken tooth.

Bibliography

1. Smith, J. D. "Estimation of the static load distribution factor for helical gears." Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science209.3 (1995): 193-199.

2. Chen, Kangkang, et al. "Comparison of meshing characteristics of helical gears with spalling fault using analytical and finite-element methods." Mechanical Systems and Signal Processing121 (2019): 279-298.
3. Wang, Qibin, et al. "An improved time-varying mesh stiffness model for helical gear pairs considering axial mesh force component." Mechanical Systems and Signal Processing 106 (2018): 413-429.
4. Wang, Qibin, and Yimin Zhang. "A model for analyzing stiffness and stress in a helical gear pair with tooth profile errors." Journal of vibration and control 23.2 (2017): 272-289.
5. Dadon, I., et al. "A realistic dynamic model for gear fault diagnosis." Engineering Failure Analysis 84 (2018): 77-100.
6. Dadon, I., et al. "A Step Toward Fault Type and Severity Characterization in Spur Gears." Journal of Mechanical Design141.8 (2019): 083301.

Modeling and identification of mechanical systems

Localization and quantification of damage by frequency-based Method: Numerical application on bending vibration beam

Anurag DUBEY[1], Vivien DENIS[1], Roger SERRA[1]

[1]INSA Centre Val de Loire, Laboratoire de Mécanique Gabriel Lamé - EA 7494, Blois, France
anurag.dubey@insa-cvl.fr

Abstract

The sudden growth of damages can cause catastrophic failure of structures or mechanisms that lead to unplanned shutdowns of machines and production lines. If a damage remains undetected and reaches a critical size, sudden collapses and failures can happen. To overcome these problems, it is essential to detect these damages before they reach their critical state. The presence of damages can alter the structure which reduces the bending stiffness and modify the modal parameters and the natural frequencies. One of the most suitable monitoring methods to define the presence of damage and assess the structure is vibration based structure health monitoring (VBSHM). The objective of the work is to localize and quantify the damages with the consideration of eigenfrequencies of healthy and tested structures. Hence, a methodology for damage identification in structure using frequency shift coefficient (FSC) is presented. Numerical finite element models (2D and 3D) are performed and correlated to obtain a damage library for the cantilever beam structure. Based on the cost function, Young's modulus of 2D and 3D models are iteratively updated to closely match the frequencies of the reference beam. The approach also quantified geometry damage with vibration measurements on cantilever beams, which is related to an equivalent bending stiffness reduction by the use of FSC. The effect of severity of the damage is considered. Finally, the result is validated numerically through the identification of geometry damage.

1 Introduction

Damages or cracks are inevitable in aerospace, aeronautical, mechanical and civil structures during their service life. Any changes in the structures such as material, physical or geometrical properties which affects their performance are considered as damages. The study of damages is an important perspective in order to ensure safety or to avoid any serious losses. Sudden occurrence of damages in the structure can cause catastrophic failure and reduction in load carrying capacity. However, it is necessary to improve the durability and reliability of structure as expressed in the design and maintenance specifications. The presence of the damage makes local stiffness vary in the structure and it also affects the mechanical behavior and performance of the structure. However, preventing the formation of damages is almost impossible as they propagate along the structure due to fluctuating stress or fatigue conditions. If these cracks remain undetected and reach a critical size sudden collapse can happen. Indeed, damage identification has significant life safety implications.

Structure Health Monitoring (SHM) is an efficient way for the diagnosis of the constituent's materials or structures. SHM involves the integration of sensors, data transmission, computational techniques, and processing ability to respond the behavior of a structure. Consequently, it aims to provide maintenance services throughout the life of the structure. Nowadays, structural damages are identified by Non-Destructive Testing (radiographic, ultrasonic testing, X-ray, eddy-current etc.) [1]. Vibration based structural health monitoring (VBSHM) is one of these categories based on the fact that a loss of stiffness caused by damages affects the dynamic response of the structure. VBSHM consists of five levels (existence, location, type, extent and prognosis) [2] which are efficient and widely accepted because of their ability to monitor and detect damage from global testing of the structure.

Many researchers from the last few decades, natural frequencies of a damaged structure are found as an identification parameter for both damage location and size. The first study developed by Cawley and Adams [3] depends on the shift of more than one frequency that could yield the location of the damage. In a review of the literature, Salawu [4] found that the natural frequencies are a sensitive indicator to detect the damage in the structure. The important technique is analyzing the changes (shifts) in natural frequencies in a structure with and without damage. Hilmy et al. [5] have presented frequency shifting as a function of damage evolution for a plate structure. The method proves shifting of the natural frequency is greater at higher frequency values and determines the location of the void damage. Messina et al. has proposed Damage Location Assurance Criterion (DLAC) [6] and after extended to Multiple Damage Location Assurance Criterion (MDLAC) [7] to measure the frequency variation due to damage between experimental and numerical values correlation. More recently, a method proposed by Serra et al. [8] demonstrates a correlation of 2D and 3D FE models to identify the typical damages (like hole, crack, notch) based on numerical and experimental study. Masoumi and Ashory [9] presented numerical and experimental studies to localize cracks.

In this paper, an approach for damage identification by using the frequency-shift coefficient is proposed. This method was first introduced by Silva and Gomes [10] for solving the damage detection problem. The method requires numerical models as a function of damage position and size for the frequency shift. First, vibration based strategy is used with detection, localization and classification (Size/Severity/Geometry) of damages. The study is followed by simulating a beam in commercial software (COMSOL, MATLAB) as a numerical case and 2D and 3D FE models are correlated to obtain geometry damage properties (size, location and severity...). Finally, numerical example is validated in order to localize and quantify geometry damage.

2 Cantilever beam bending vibration background

The eigenvalue problems and the analytic formulas concerning the modal parameters of a cantilever beam were described by the partial differential equation of the linear model with viscous damping as:

$$M(x)\frac{\partial^2 v(x,t)}{\partial t^2} + C(x)\frac{\partial v(x,t)}{\partial t} + \frac{\partial^2}{\partial x^2}\left(EI(x)\frac{\partial^2 v(x,t)}{\partial x^2}\right) = F(x,t) \tag{1}$$

where $v(x,t)$ is the transverse deflection, $M(x)$ is the mass per unit length, $C(x)$ is the damping coefficient, $EI(x)$ is the bending stiffness and $F(x,t)$ is the external force per unit length of the beam. The equation of the motion for dynamic systems are easily obtained from Newton's second law. This gives an equation for each degree of freedom within the system. When discretized, the equation of the motion may take the following matrix form:

$$[M]\{\ddot{X}\} + [C]\{\dot{X}\} + [K]\{X\} = \{F\} \tag{2}$$

where $[M]$ is the mass matrix for the system, $[C]$ is the damping matrix and $[K]$ is stiffness matrix, $\{\ddot{X}\}$, $\{\dot{X}\}$, $\{X\}$ are vectors containing acceleration, velocity and displacement in all degree of freedom of the model; and $\{F\}$ contains external forces actuating in the system. If we assume free motion and negligible damping, one possible solution for the equation is:

$$\{x\}_i = \{y\}_i \sin(\omega_i t - \theta_i) \tag{3}$$

y_i are the amplitudes for each mode shape, ω_i are the natural pulsations (in rad.s^{-1}) of vibration for each mode shape and θ_i are phase angles. The natural frequencies (in Hz) are given by $f_i = \frac{\omega_i}{2\pi}$. The following equation is obtained for the healthy case:

$$([K] - \omega_i^2[M])y_i = 0 \tag{4}$$

Damage to the structure changes its dynamic response. Therefore, natural frequencies and natural modes are changed. The equation of a damaged case can be expressed as:

$$([K] - (\omega_i^2)^*[M])y_i^* = 0 \tag{5}$$

2.1 2D Finite element model

The studied model is a cantilever beam, which has two degrees of freedom, a vertical translation y and a rotation θ_z. As can be seen in (Figure 1) this beam is divided into equal size of N elements and $N+1$ nodes.

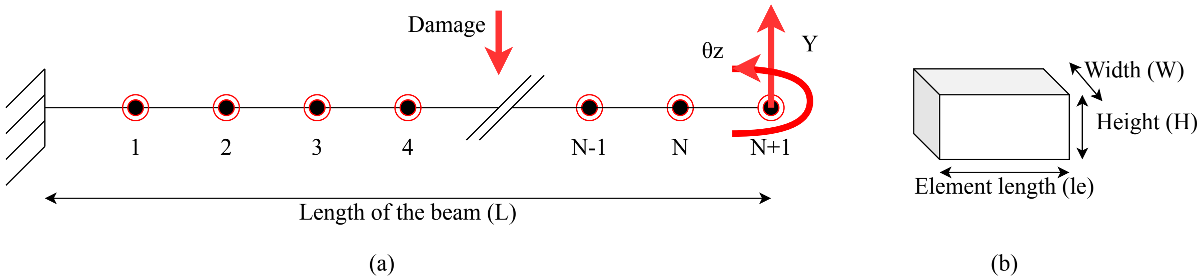


Figure 1 – Cantilever beam model. (a) 2D Finite element model; (b) element properties

In 2D FE Model, the damage is represented by an elemental stiffness reduction coefficient α_i which is the ratio of the stiffness reduction to the initial stiffness. The stiffness matrix of damaged beam is defined as a sum of elemental matrices multiplied by reduction coefficient by the following equation:

$$[K_d] = \sum_{i=1}^{N} (1-\alpha_i)[K_i] \tag{6}$$

where K_d is global stiffness matrix for damaged beam, K_i is elemental stiffness matrix, N is number of elements, and α_i is a reduction coefficient , which varies from 0 to 1 for the damaged structure. The value of α_i = 0 indicates a healthy structure.

2.2 3D Finite element model

Simulation of damaged beam structure is performed using COMSOL multiphysics software. The damage model is built and the mesh is 3D tetrahedron element. The number of mesh is controlled by the software and depends on the shape of the structure, thus it changes with the size of the crack. A high meshing density is applied near the damaged area mainly to have the behavior correctly modeled.

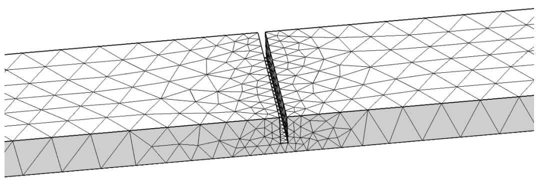


Figure 2 – 3D finite element mesh of the beam with damage

Geometry case (rectangular) is studied in order to quantify the severity of the damage. Figure 2 shows the meshed beam zoomed near the damaged area and width of the crack is set to 0.5 mm while the height is a parameter. The sensitivity of the 3D model is determined by mesh size. As the mesh is finer the model is more sensitive but computing cost is higher.

3 Frequency shift coefficient based strategy

The first type of modal method for damage detection relied on changes in dynamic properties of the structure and particularly natural frequencies. Any changes in the properties of the structure, such as reduction in stiffness will cause changes in the natural frequencies. One of the important advantages of natural frequency is that it can be quickly and easily conducted when measurements required. Classical measurements procedure can be used for the determination of experimental resonant frequencies. In this context, the frequency shift criterion is first presented by Silva and Gomes [10] for damage identification problems. The technique requires experimental measurements or numerical solution for the frequency shifts as a function of size and position of damage The frequency shift coefficient (FSC) is defined as:

$$FSC = \operatorname{argmin}\left(\sqrt{\frac{1}{m}|\sum_{i=1}^{m}\left(\frac{(R_i)_X-(R_i)_A}{(R_i)_X}\right)|}\right) \quad \text{and} \quad R_i=\frac{f_i^u}{f_i^h}, \tag{7}$$

where m is the total number of modes, X refers to the tested case, A refers to the reference case, f_i^u is the unknown beam frequencies, f_i^h is healthy beam frequencies and i is denotes modes indices.

It is well known that the presence of damages modifies dynamic parameters and behavior of the structure. The location, classification and size of damages in the structure are identified by changes in the vibration parameters. At first, a set of reference state frequencies are identified. Numerical correlation of 2D and 3D FE models is performed to fit the frequencies with the references. Then, in the 3D FE model, the damage was materialized as a geometrical discontinuity of rectangular considering the position, type, size, geometry of the damage. At the same time, the damage was materialized as a local reduction of bending stiffness in an element for the 2D FE model. Finally, numerical correlation result will specify the position, size, depth and geometry of the damage from the damage library.

4 Numerical rectangular geometry damage identification

The numerical simulation test is performed to verify the efficiency of the proposed VBSHM strategy. A cantilever steel beam was taken into consideration for the numerical test and beam properties are given below in Table 1. A beam 2D FE model was divided into equal size of 100 elements and each element size is 10 mm.

Beam Properties	Value
Length (L)	1000 mm
Width (W)	24.9 mm
Height(H)	5.3 mm
Young's modulus (E)	210 GPa
Mass density (ρ)	7850 kg/m^3
Poisson's ratio (v)	0.33

Table 1 – Beam dimensions and properties

However, 2D-3D model correlation is done for the damage geometry (rectangular) by using software COMSOL with MATLAB. The modal responses of the structure were generated using FE models before and after the damaged case. The first seven modes are retained. The criterion is employed as a tool for identification of the damage by measuring frequencies. The final goal of correlation is to localize and quantify the severity of geometry damage that can link to the percentage reduction in stiffness of a beam structure.

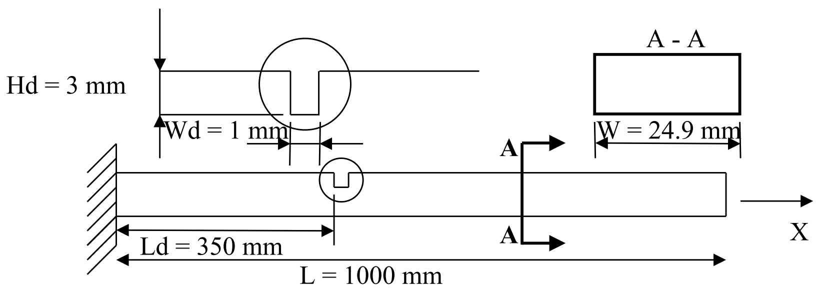


Figure 3 – 2-D clamped free beam plan with rectangular damage

A detailed 2D beam view as shown in Figure 3. The damage case is tested for position 350 mm with width (Wd) and the height (Hd) of damage are 1 mm and 3 mm respectively. Meanwhile, the FSC is computed for every position and severity, in order to illustrate its variations.

In Figure 4, the FSC is shown as a function of tested position and severity where color levels represent the FSC values. The minimum value (coordinates and value of the minimum) allows the identification of given

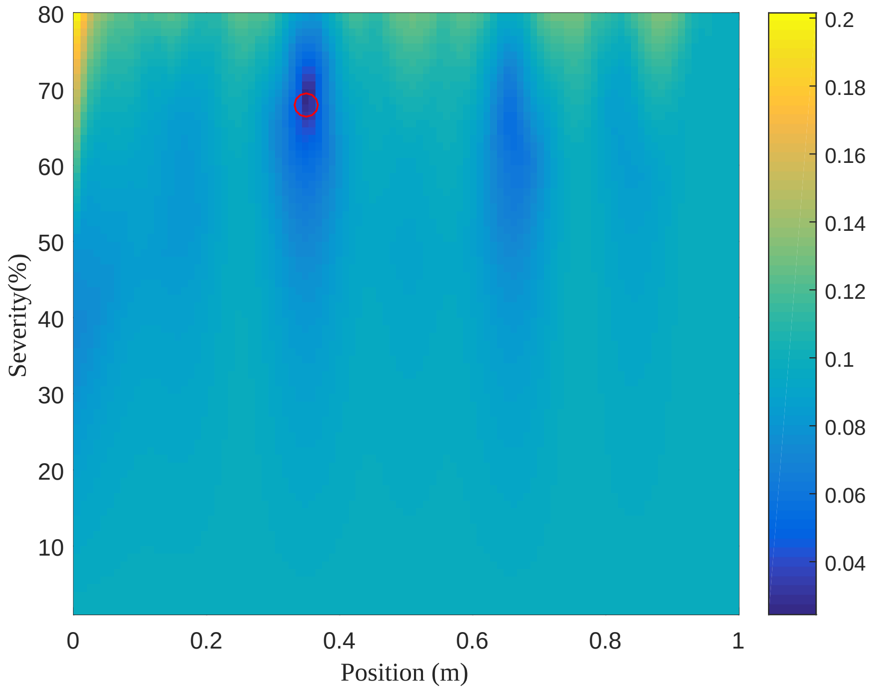


Figure 4 – FSC as a function of tested position and severity. The red circle indicates the severity and position identified by the algorithm.

damage in 2D clamped free beam. In this case, a defect of 68% severity localized at 350 mm is found : the position is thus well identified by the FSC and the identified severity corresponds to the parameters chosen for the rectangular damage. These values relate to the other damages properties and information about the size and type. In addition, Table 2 shows the identified damage properties for this particular case of rectangular damage. This is one item of the damaga library the presented strategy is intended to build.

Type	**Rectangular**
Position	350 mm
Width	24.9 mm
Length	1 mm
Height	3 mm
Severity	68 %

Table 2 – Geometry damage properties

5 Conclusion

This paper presents a method to identify damage in structure by using natural frequencies. The formulation of the method based on stiffness reduction has been validated with the localization and quantification of the rectangular geometry damage in beam like structure. The simulation correlation with COMSOL and MATLAB are presented and the robustness of the present method is examined. A numerical example with 3D geometry damage case is identified. Based on natural frequency, the damage localization and quantification is accurate because of the sensitivity of the frequency shifts to the damage states. Both 2D and 3D models of the beam were used to link the size of damage to the reduction in stiffness. Geometrical damage properties were successfully accomplished by linking FE models. In the future, more experiments and simulations should be investigated in order to validate the methodology in real cases.

Acknowledgements

The work was funded by the Région Centre Val de Loire, France through a doctoral contract. The authors are thankful for the financial support.

References

[1] B. P. Chandra Rao, *Non-destructive Testing and Damage Detection*, In: Aerospace Materials and Material Technologies, Springer, Singapore, 2017. p. 209-228.

[2] A. Rytter, *Vibrational Based Inspection of Civil Engineering Structures*, Department of Building Technology and Structural Engineering, Aalborg University, 1993.

[3] P. Cawley, R.D. Adams, *The location of defects in structures from measurements of natural frequencies*, The Journal of Strain Analysis for Engineering Design, 1979, vol. 14, no 2, pp. 49-57.

[4] O. S. Salawu, *Detection of structural damage through changes in frequency: a review* Engineering Structures, 1997, vol. 19, no 9, pp. 718-723.

[5] I. Hilmy, M. A. Adesta, E. Y. T., and T. Firdaus, *Damage Detection based on the Natural Frequency shifting of a clamped rectangular plate model*, In : Journal of Physics: Conference Series. IOP Publishing, 2015, pp. 012034.

[6] A. Messina, I. Jones, and E. Williams, *Damage detection and localization using natural frequency changes*, In Proceedings of the 1st Conference on Identification, Cambridge, England, UK, 1992, p. 67-76.

[7] A. Messina, E. J. Williams, and T. Contursi, *Structural damage detection by a sensitivity and statistical-based method*, Journal of sound and vibration, 1998, vol 216, no 5, pp. 791-808.

[8] R. Serra, L. Lopez, and G. Gautier, . *Tentative of damage estimation for different damage scenarios on cantilever beam using numerical library*, In : Proceedings of the 23th International Congress on Sound and Vibration, Athens, Greece, 2016, pp. 10-14.

[9] M. Masoumi, and M. R. Ashory, *Damage identification from uniform load surface using continuous and stationary wavelet transforms*, Latin American Journal of solids and structures, 2014, vol. 11, no 5, pp. 738-754.

[10] J. Silva, and A. Gomes, *Crack identification on simple structural elements through the use of natural frequency variations: The inverse problem*, In Proceedings of the 12th International Modal Analysis, 1994, p. 1728.

ARX model for experimental vibration analysis of flexible manipulator during grinding

Quoc-Cuong Nguyen[1], Viet-Hung Vu[1] and Marc Thomas[1*]
[1]Mechanical Engineering Department, École de Technologie Supérieure (ÉTS),
Montréal, Québec, H3C 1K3, Canada
[*]Corresponding author: marc.thomas@etsmtl.ca

Abstract

Using a flexible manipulator for grinding process in situ has become a cost effective engineering service in the recent years, especially for repair and refurbish of mechanical systems and components. In comparison with traditional rigid robot manipulators, the flexible manipulator has proved its efficiency in terms of accuracy and facility. However, because of its compact and flexible structure, concerns arise regarding its dynamic behavior during a grinding process. This paper proposes a method using an ARX (autoregressive with exogenous excitation) model for experimentally analyzing the vibrations of a flexible robot during a grinding operation in different cases: Single Input–Single Output (SISO) and Multi Input–Multi Output (MIMO). Simultaneously, a dynamometer allows for triaxial input force measurement while three accelerometers mounted at the end effector record the vibration outputs. Due to the Operational Modal Analysis (OMA), the dynamical properties of the robot can be identified directly during operation. The results have shown that the ARX model is efficient for analyzing the operational vibration in complex systems with multi degrees of freedom and multi directions. The determination of modal parameters and identified Frequency Response Functions (FRFs) enable to predict the dynamical behavior of the system and to simulate the vibration in real working conditions. Further studies on inverse problem are promising for estimating the excitation forces while these later are not available and not practically measured in industrial applications.

Keywords: Operational modal analysis, flexible manipulator, grinding process, ARX model, transfer functions, force identification.

1. Introduction

Nowadays robots sufficiently conduct manifold manipulation works with a high degree of autonomy and rigorousness. Portable manipulator systems are regarded as an effective and profitable solution for the automation maintenance tasks on large hydroelectric equipment. The SCOMPI (Super COMPact robot Ireq) was developed at IREQ (Hydro Quebec's research institute) and is particularly designed with flexible links and flexible joints for working in the hard-to-reach areas or confined spaces of hydraulic turbines in a hostile environment [1]. Because of its flexible structure, vibration problems of Scompi become crucial since producing chatter and bad surface finish. A numerical simulation [2] has been constructed in MSC/Adams in different configurations included impact force, sinusoidal and operational forces. There is a great number of researches that focus on identifying the modal parameters of the system in order to understand the dynamical behavior of robot [3-7], and estimate the operational forces from the actual accelerations measured on the robot [8]. Knowing a system's frequency response function is a key to many system analysis and control synthesis methods [9]. The main problems are due to the fact that these modal parameters are changing with the robot motion and position and thus a time-varying method is proposed for studying this kind of non-stationary structure [10, 11]. Researchers are particularly interesting to identification of continuous-time system by using discrete data [12].

This paper presents a technique to identify the modal parameters as well as the transfer function of Scompi robot by applying the Autoregressive with eXogenous input (ARX) model [12-14]. This method reveals a convenient and advantageous for Operational Modal Analysis of structures (OMA), which allows for determining operational modal model excited by ambient noise and vibration. The modal parameters are estimated and identified by applying straightforward method such as Ordinary Least Squares (OLS) [12], [15]. The results are validated by another approach based on updated Auto Regressive (AR) model in [4] and shown a great accuracy of identified modal parameters. This study enables us to predict the dynamical behavior of Scompi for identifying excitation forces during operations of grinding and consequently improve the quality of the surface finish.

2. Auto Regressive Exogenous Excitation model (ARX)

The ARX model [12-15] is a primary choice because of its simplicity. It has been applied to numerous practical applications especially in control systems. However, critical motivation for choosing the ARX model, is its correlation to the state space model [16-19] which can be implemented for inverse problem with the aim of reconstructing the excitation forces acting on vibrating structures [8], which is impossible to obtain from direct measurement in the real systems. The ARX model is a convenient model to obtain the general relation between input and output signals for different cases, such as Single Input – Single Output (SISO) or Multiple Input – Multiple Output (MIMO), which can reliably represent the dynamic properties of the system. Figure 1 illustrates the block diagram of ARX model.

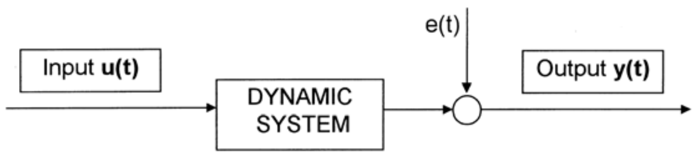


Figure 1. Block diagram of ARX model

This model has a simple structure and strong robustness. It is very efficient when the noise is low. However, when the noise is large, the order of the model must increase to compensate the impact to system identification precision from noise [18].

Examine a c dimensional vector input $\mathbf{u}(t)$ and a d dimensional vector output $\mathbf{y}(t)$ of a Multiple Input and Multiple Output (MIMO) system.

The ARX model can be described as a linear difference equation:

$$\mathbf{y}(t)+\mathbf{A}_1\mathbf{y}(t-1)+\ldots+\mathbf{A}_{n_a}\mathbf{y}(t-n_a) = \mathbf{B}_0\mathbf{u}(t)+\mathbf{B}_1\mathbf{u}(t-1)+\ldots+\mathbf{B}_{n_b}\mathbf{u}(t-n_b)+\mathbf{e}(t) \quad (1)$$

where:

$\mathbf{A}_i$ – are d x d matrices and

$\mathbf{B}_i$ – are d x c matrices.

The general ARX model can be rewritten in the polynomial form:

$$\mathbf{A}(q)\mathbf{y}(t) = \mathbf{B}(q)\mathbf{u}(t) + \mathbf{e}(t) \tag{2}$$

where:

$$\mathbf{A}(q) = \mathbf{I} + \mathbf{A}_1 q^{-1} + \mathbf{A}_2 q^{-2} + \ldots + \mathbf{A}_{n_a} q^{-n_a} \tag{3}$$

$$\mathbf{B}(q) = \mathbf{B}_0 + \mathbf{B}_1 q^{-1} + \mathbf{B}_2 q^{-2} + \ldots + \mathbf{B}_{n_b} q^{-n_b} \tag{4}$$

The model (2) is an ARX model where AR refer to the Autoregressive part **A**(q)**y**(t) and X refer to the extra input **B**(q)**u**(t) called the exogenous input. **y**(t) is considered as the output of the model while **u**(t) is the input to the model and **e**(t) is innovation term at the time *t*. **A**(q) and **B**(q) are polynomials in the delay operator q^{-1} and n_a, n_b are the model order of **A**(q) and **B**(q) respectively. **A**(q) is a matrix whose elements are polynomials in q^{-1}.

This results in Matrix Fraction Description (MFD).

Defining the parameter matrix:

$$\theta = [\mathbf{A}_1 \;\; \mathbf{A}_2 \;\; \ldots \;\; \mathbf{A}_{n_a} \;\; \mathbf{B}_0 \;\; \mathbf{B}_1 \;\; \ldots \;\; \mathbf{B}_{n_b}]^T \tag{5}$$

We may rewrite (2) as a linear regression:

$$y(t) = \theta^T \phi(t) + e(t) \tag{6}$$

If we consider N consecutive values of the responses from *y(k)* to *y(k+N-1)*, the model parameters can be obviously estimated by least square method [15] by minimizing the norm of *e(t):*

$$\Phi = \arg\min\left(\frac{1}{N}\sum_{t=k}^{k+N-1}\left\|e(t)\right\|^2\right) = \arg\min\left(\frac{1}{N}\sum_{t=k}^{k+N-1}\left\|\mathrm{y}(t) - \theta^T\phi(t)\right\|^2\right) \tag{7}$$

After obtaining the measured force and acceleration signals on all channels, the model ARX can be used to fit the data. The ARX model creates a regressive connection between the input vector **u**(t) and the output vector **y**(t) through a residual vector **e**(t). By applying the least square method, the modal parameters matrices **A** and **B** can be estimated. In vibration measurement application, it can be seen that force (input) and acceleration (output) are normally synchronized, thus the two parts may be modeled with the same order $n_a = n_b$.

Once the model parameters of the system are identified, the state matrix can be determined as in the form of autoregressive parameters:

$$A\Big|_{(naxnb)} = \begin{bmatrix} -A_1 & -A_2 & -A_3 & \ldots & -A_p \\ I & 0 & 0 & \ldots & 0 \\ 0 & I & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & I & 0 \end{bmatrix} \tag{8}$$

There is a remarkable coincidence that the poles of model are also the roots of characteristic polynomial of the state matrix. Consequently, the continuous eigenvalues, system natural frequencies and damping rates of the structure can be calculated for each pole by using the subsequent standard equations:

Eigenvalues: $$[V, \lambda] = eig\,(A) \tag{9}$$

Frequencies: $$f_i = \frac{\sqrt{\mathrm{Re}^2(\lambda_i) + \mathrm{Im}^2(\lambda_i)}}{2\pi} \tag{10}$$

Damping rates: $$\xi_i = -\frac{\mathrm{Re}(\lambda_i)}{2\pi f_i} \tag{11}$$

When the modal parameters are estimated, we can construct the transfer function which is regarded as the frequency response function of the system. All the system can be described by linear constant coefficients and represented by transfer functions that are "rational polynomial in q".

$$\mathbf{G}(q) = \frac{\mathbf{B}(q)}{\mathbf{A}(q)} = q^{-n_k} \frac{\mathbf{B}_1 q^{-1} + \mathbf{B}_2 q^{-2} + \ldots + \mathbf{B}_{n_b} q^{-n_b}}{\mathbf{I} + \mathbf{A}_1 q^{-1} + \mathbf{A}_2 q^{-2} + \ldots + \mathbf{A}_{n_a} q^{-n_a}} \tag{12}$$

with n_k is the transport delay.

3. Application to a flexible manipulator during grinding process

3.1 Brief introduction of the SCOMPI robot

The proposed approach is now implemented to the portable robot Scompi. Figure 2 presents the structure of Scompi, which is used for repair tasks in Hydro Quebec power plants, particularly for grinding or welding jobs [1]. Because of its compact and flexible structure, the question is raised up from its dynamical behavior under operating conditions. Hence, the flexibility of the joints and links needs to be taken into consideration, which might affect the stabilization of robot at the end effector during operational process [4]. The aim of Scompi is to achieve both a high Material Removal Rate (MMR) and a polished surface finish with great precision. However, because of the portable and lightweight design, undesired chatter vibrations can appear during machining process which produces an undesirable waviness surface. Therefore, the monitoring of its modal parameters as well as the transfer functions of the structure in the grinding operation are necessary for minimizing vibration at the end effector while controlling chatter phenomenon and improve the quality of grinding surface.

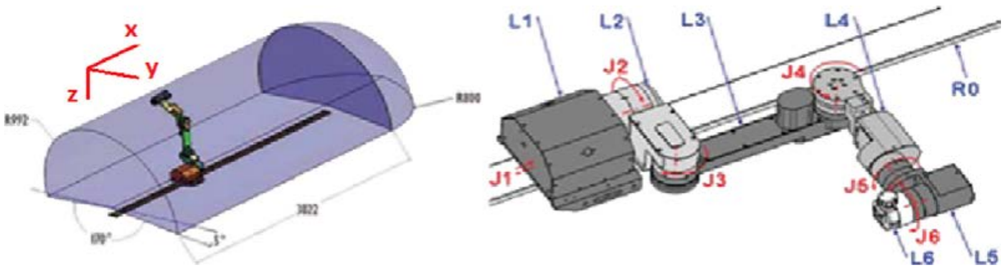


Figure 2. Scompi robot

3.2. Presentation of the Experimental setup

As can be seen from figure 3, a Scompi robot is tested under real grinding operation. Due to the interest in typical dynamic behavior of the robot at the end effector, the Scompi is set to its home configuration. Three accelerometers are mounted at the end effector in triaxial directions X, Y and Z. Meanwhile, a Kistler table dynamometer CH8408 is placed under the work-piece for measuring the forces. The power is set up at 1500 W and grinding motor is rotated at a constant speed of 3225 (rpm) for conducting each single grinding pass within 12 seconds. A multi-component dynamometer is used for measuring the grinding forces in three directions at the tool piece contact point. After obtaining the measured signals from dynamometer and accelerometers, we acquired them to the frequency rate of 512 (Hz) (Figure 4, 5).

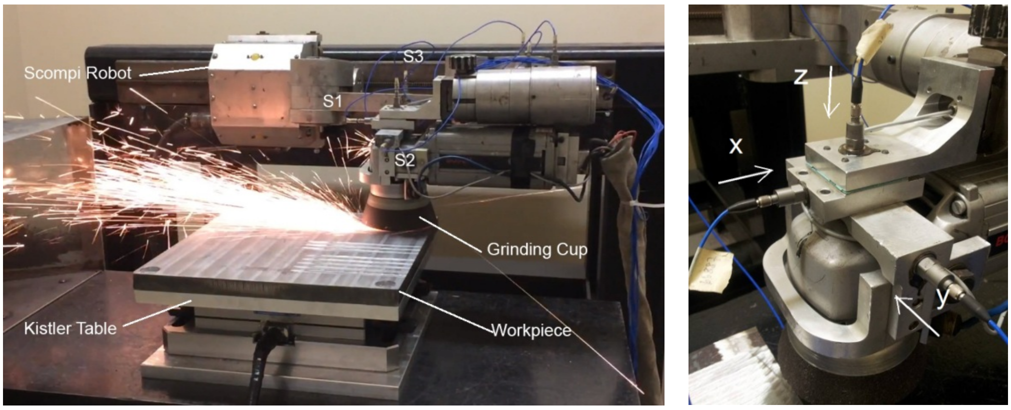


Figure 3. Overall configuration of the experimental setup

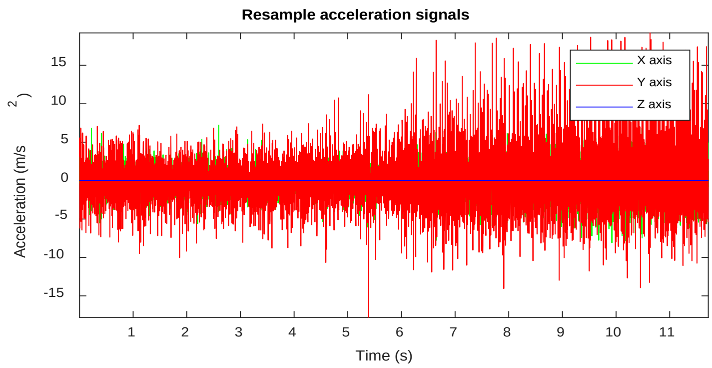


Figure 4. Measured acceleration signals during grinding process.

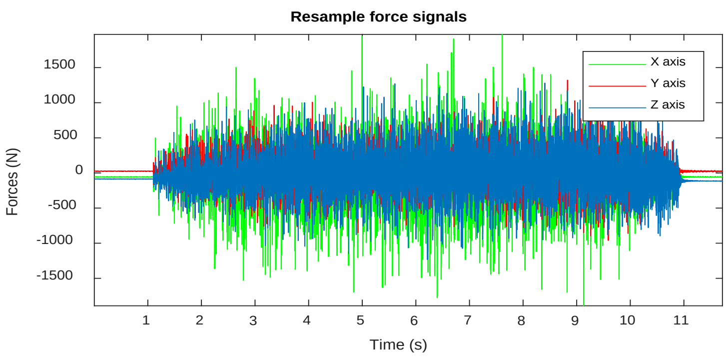


Figure 5. Measured force signals during grinding process.

Taking three measured acceleration signals in X, Y and Z directions, by applied Fast Fourier Transform (FFT) analysis, we can easily see the measured signals in both time and frequency domain as shown in Figure 6. As indicated, there are some significant frequencies in frequency domain such as 53.9 (Hz) - the first harmonic; 93.7 (Hz); 106.2 (Hz) - the second harmonic; 146.8 (Hz) and 200.8 (Hz).

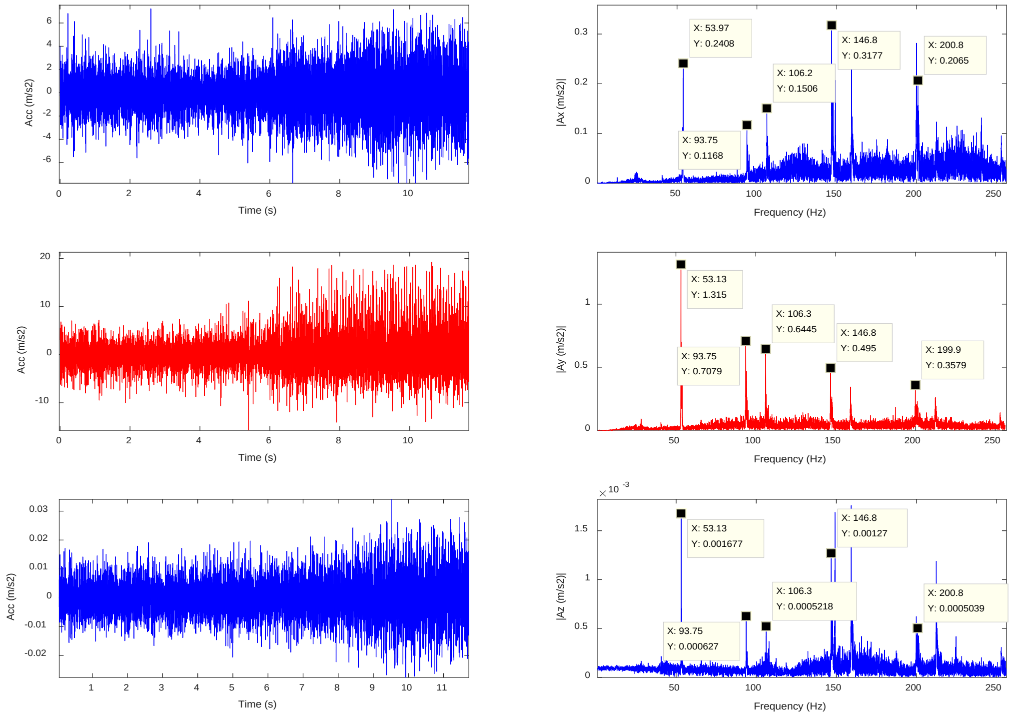


Figure 6. Time domain and frequency domain of the acceleration signals in three directions.

4. Results and discussion

Operating in a tridimensional space, the ARX model is applied on Scompi structure to fit the measured signals on each direction ($S_1 - F_x$); ($S_2 - F_y$); ($S_3 - F_z$) for constructing frequency stabilization in different cases: Single Input – Single Output (SISO) and Multi Input – Multi Output (MIMO). The figures 7–10 demonstrated the frequency stabilization diagrams up to 250 (Hz) with a model order up to 100 where all the interesting frequencies may be observed. The model order is chosen at 100 for computation of the modal parameters with low uncertainties. In addition, another stabilization given in figure 11 is computed by MODALAR based on updated AR model [4] with an aim of validation between two approaches. The 53.75 (Hz) electric frequency of grinding and its harmonics are clearly revealed in the stabilization diagrams.

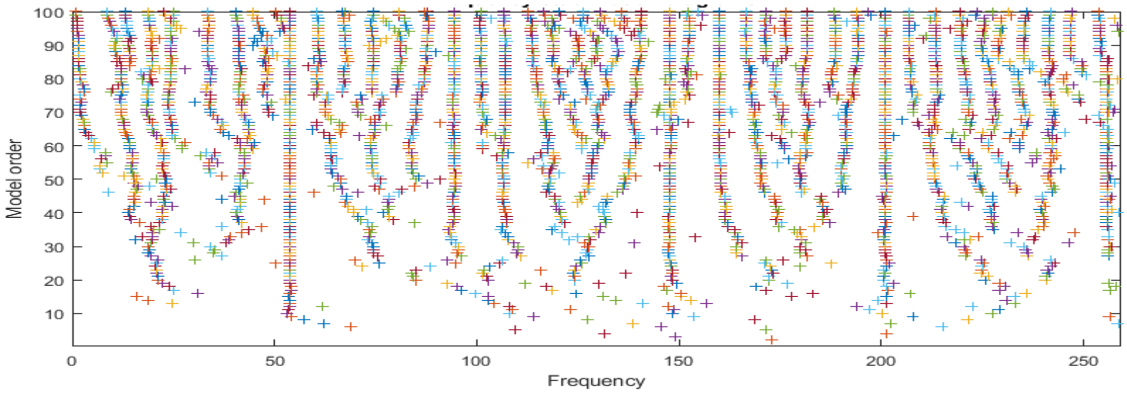


Figure 7. Frequency stabilization diagram on X direction

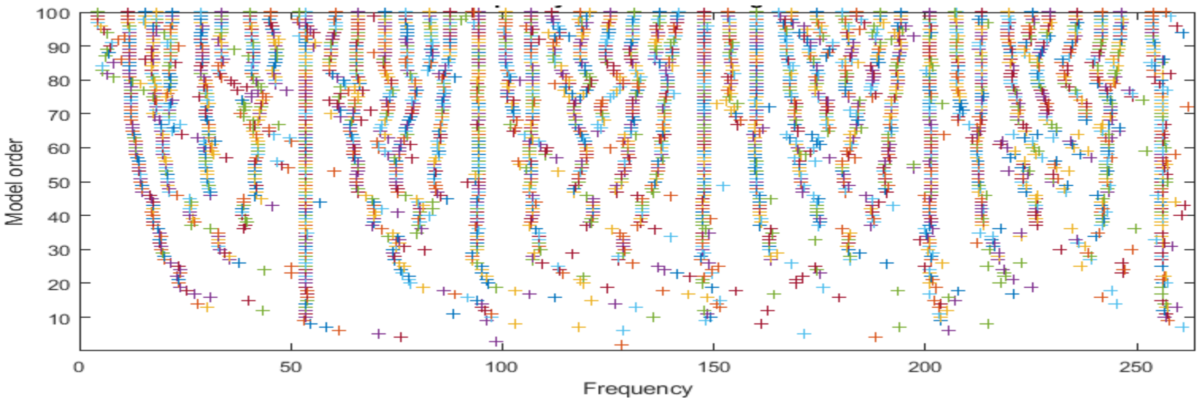


Figure 8. Frequency stabilization diagram on Y direction

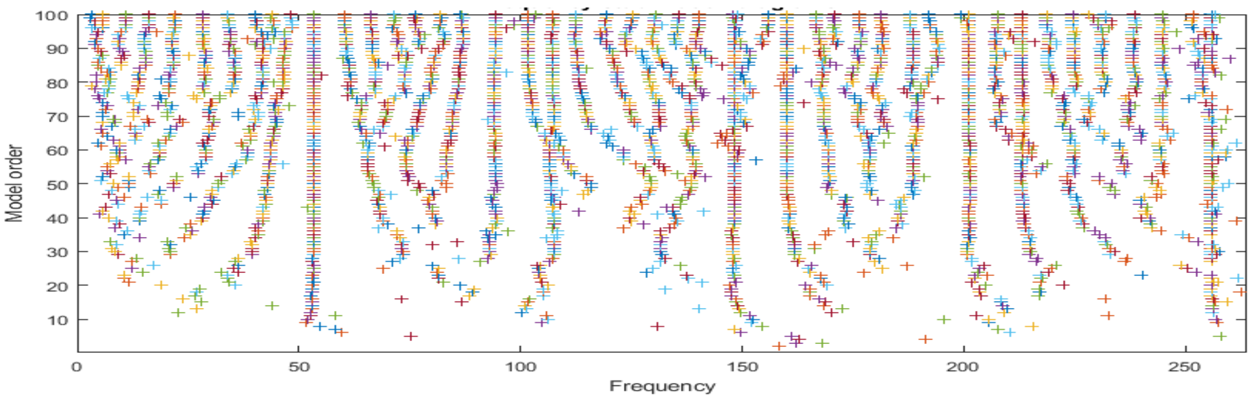


Figure 9. Frequency stabilization diagram on Z direction

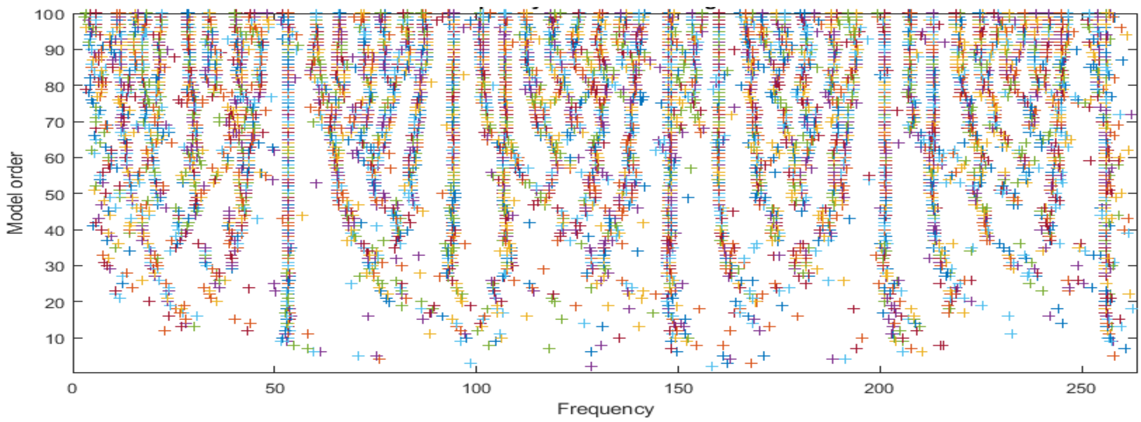


Figure 10. Frequency stabilization diagram in MIMO case (three inputs and three outputs)

Synthetically, the natural frequencies and damping ratios are estimated directly from the frequency stabilization diagram of MIMO case, where all the excited frequencies can observed clearly in multi directions. Figure 12 illustrated the stabilization diagrams of damping ratio with 95% uncertainties. The natural and harmonic frequencies identified by two methods with their damping ratio are given in the table 1. The harmonic frequencies are identified with their damping rates close to zero.

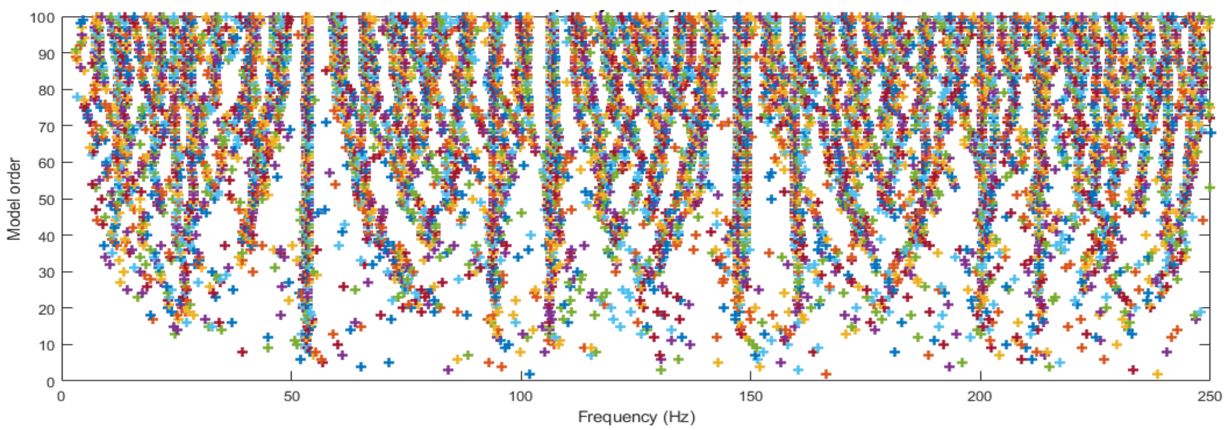


Figure 11. Frequency stabilization diagram by MODALAR

Figures 13-21 present the transfer functions identified by ARX model at the order 100. The identified transfer function from the working condition is crucial for the assessment of the robot dynamics and for further simulations under different loadings.

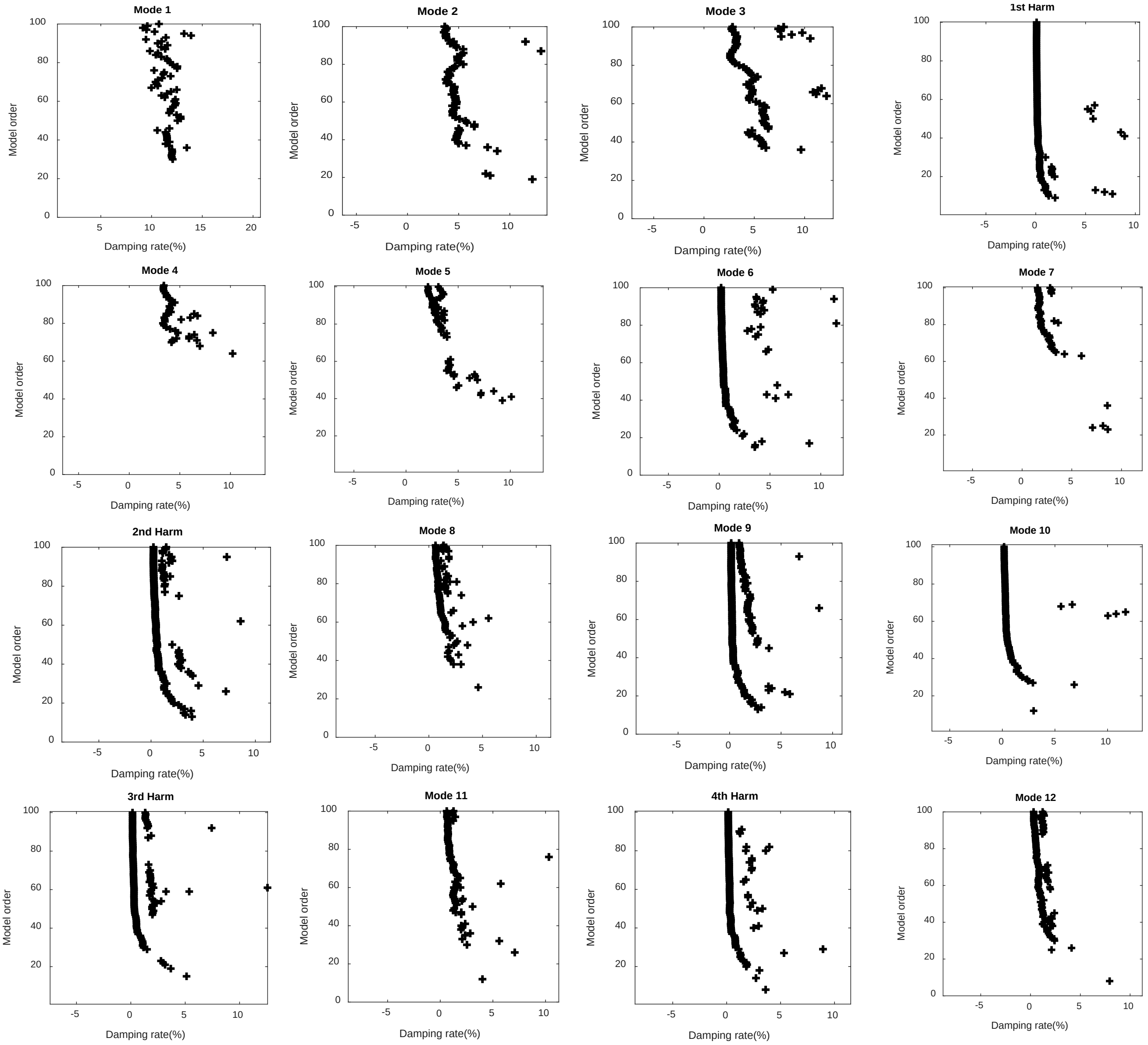


Figure 12. Damping ratio stabilization diagrams

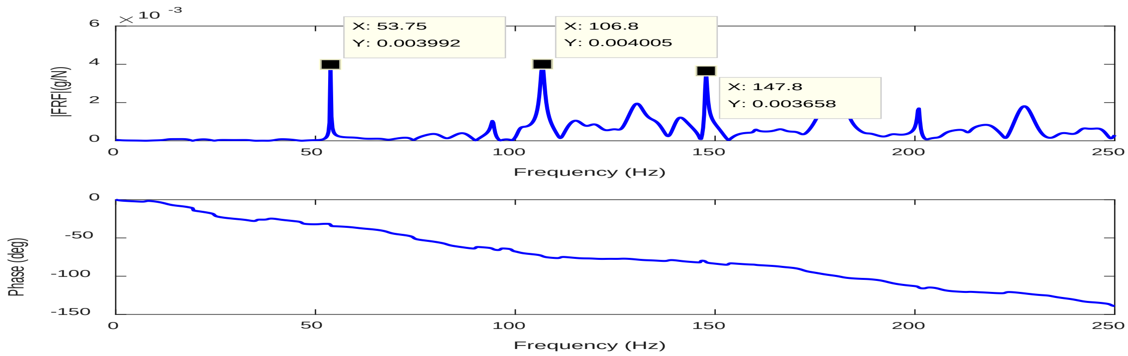


Figure 13. Identified Transfer Function FRFxx

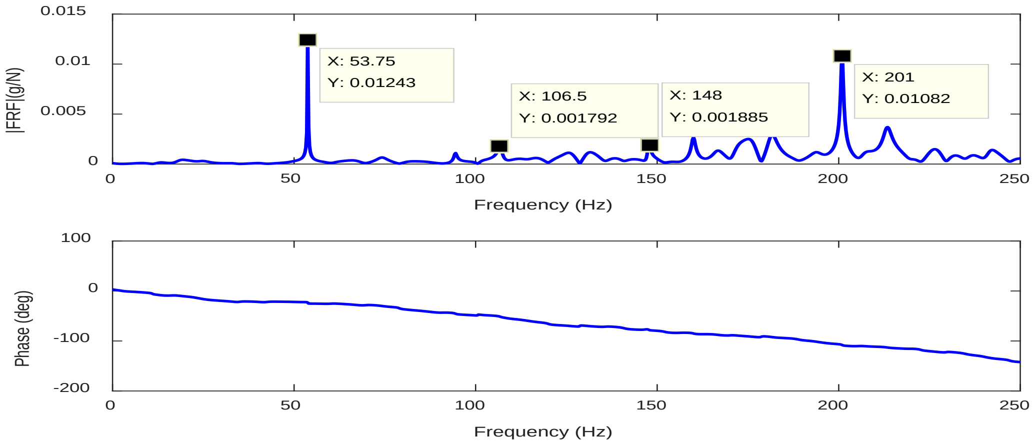


Figure 14. Identified Transfer Function FRFxy

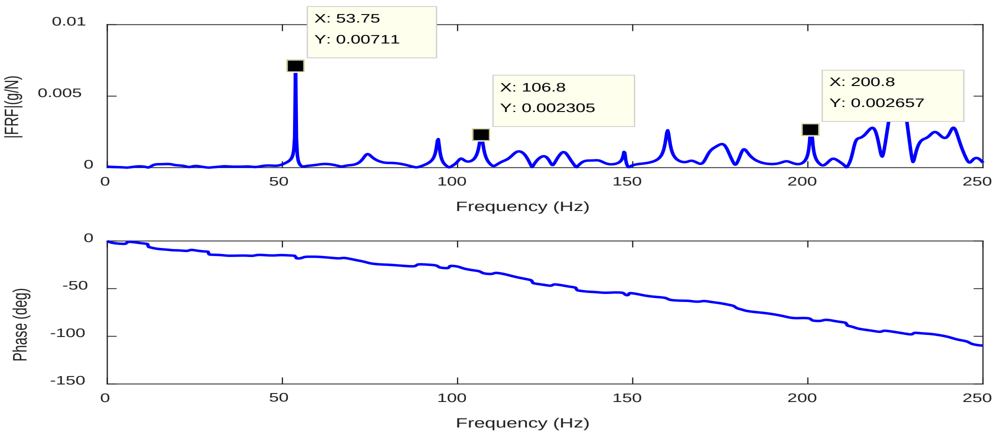


Figure 15. Identified Transfer Function FRFxz

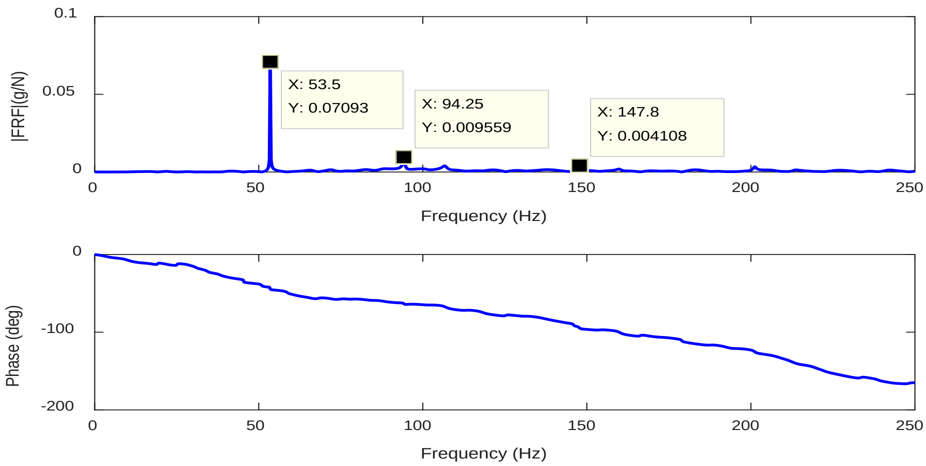


Figure 16. Identified Transfer Function FRFyx

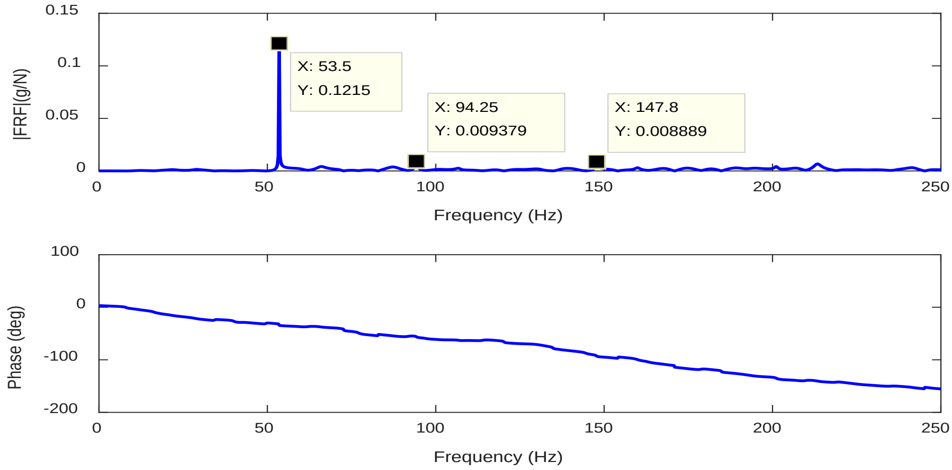


Figure 17. Identified Transfer Function FRFyy

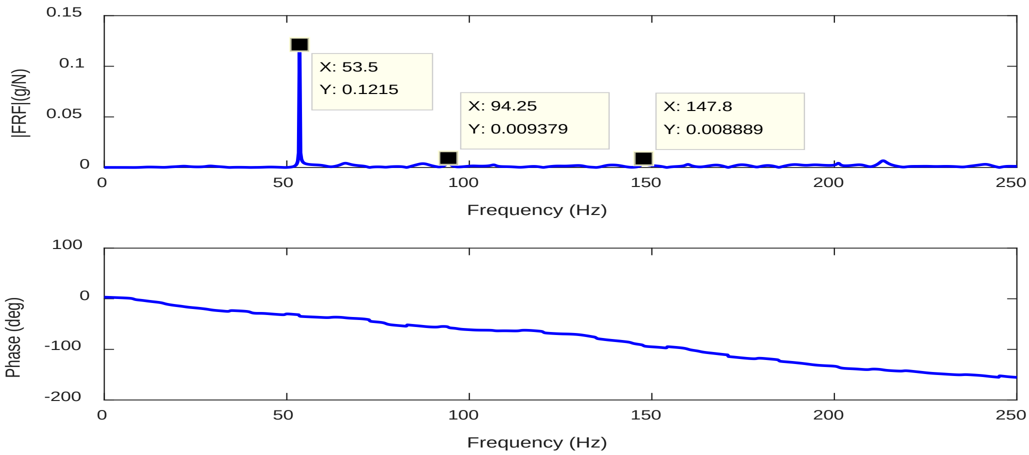


Figure 18. Identified Transfer Function FRFyz

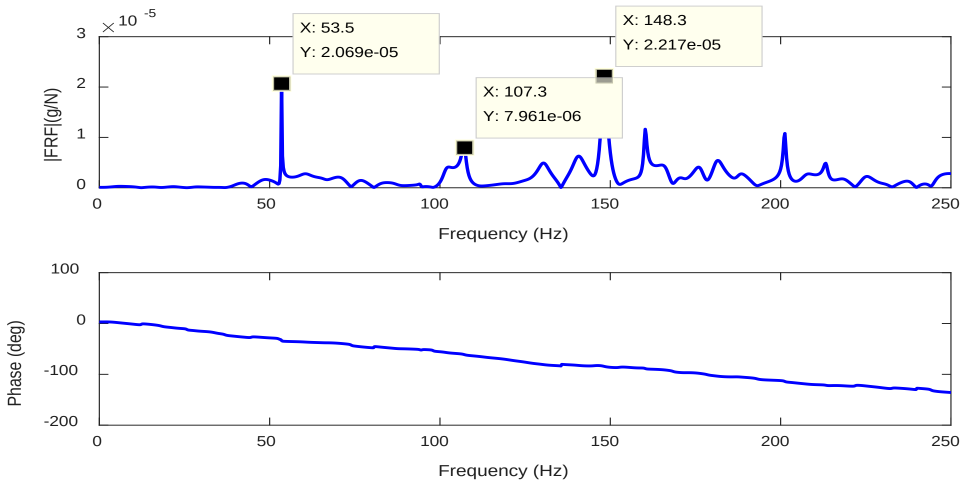


Figure 19. Identified Transfer Function FRFzx

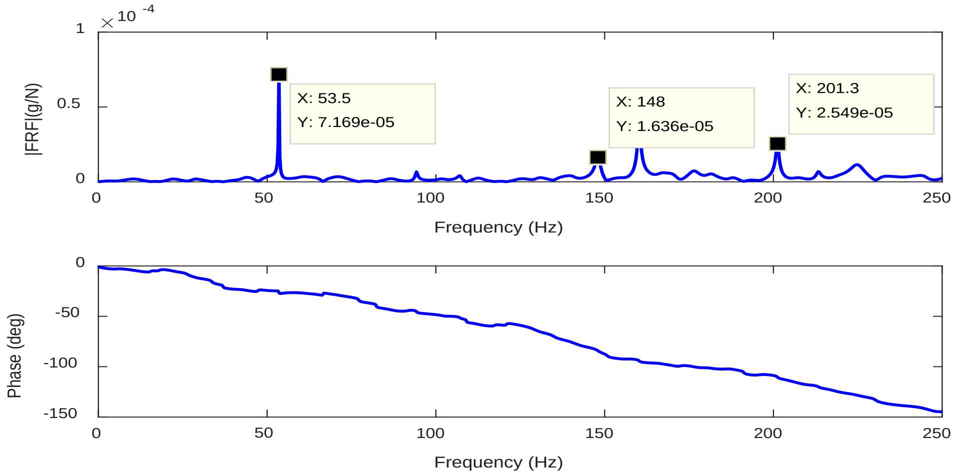


Figure 20. Identified Transfer Function FRFzy

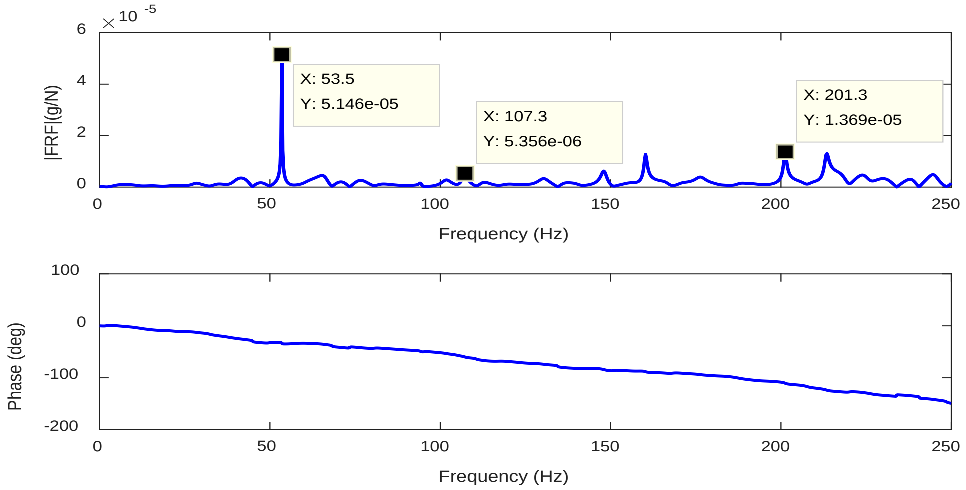


Figure 21. Identified Transfer Function FRFzz

Mode	**M 1**	**M 2**	**M 3**	**1st harm**	**M 4**	**M 5**	**M 6**	**M 7**
Frequencies (Hz) - ARX	14.2	24.4	28.2	53.3	61.2	66.2	94	101.5
Frequencies (Hz) - AR	12.7	24.6	28.2	54.2	60.1	66.2	93.9	101.3
Damping (%) - ARX	11.5	3.7	3.3	0	3.4	2.7	0.2	1.5
Mode	**2nd harm**	**M 8**	**M 9**	**M 10**	**3rd harm**	**M 11**	**4th harm**	**M 12**
Frequencies (Hz) - ARX	106.3	111.7	147.5	149.2	159.5	200.4	212.6	224.6
Frequencies (Hz) - AR	106.3	113.5	147.3	149.1	159.5	201.3	213.8	225.1
Damping (%) - ARX	0	0.6	0.2	0.2	0	0.2	0	0.7

Table 1. Identified frequencies and damping ratios

By comparison to the identified frequencies by MODALAR based on updated AR model [4] shown in table 1, the approach reveals high accuracy identified natural and harmonic frequencies with their damping ratios.

Moreover, frequency response function is directly identified from grinding operation based on ARX model. The results are better observed on the X and Y directions, this can be explained by the configuration of Scompi when working in horizontal surface to perform the grinding task.

5. Conclusion

This work is a part of an ongoing research program on investigating vibration problems of flexible manipulator. The frequencies, damping ratios and operational FRFs can be constructed and most excited modes are revealed during the grinding process. In this paper, operational FRFs of a structure are identified directly from measured signals via an ARX model. The results illustrated the sensibility of the acceleration in the X and Y directions while the contrary is proved in the Z direction with low magnitudes of the FRFs. Furthermore, as damping of the grinding process and equivalent stiffness are in command of cutting stability, so their identification is crucial to predict and avoid detrimental chatter occurrence. In the ongoing research, the inverse of ARX model will be applied in order to estimate the excitation force in the working conditions, with the integration of phase and coupling between directions. The interest lies in the reconstruction of excitation forces that gave rise to measured response signals based on ARX model. This approach is expected to serve for monitoring and vibration control design of the robot during machining operation.

Acknowledgments

The authors would like to thank IREQ (Institut de recherche d’Hydro Québec) for technical support.

References

[1] Hazel B, J. Côté, Y. Laroche and P. Mongenot, A portable, multiprocess, track-based robot for in situ work on hydropower equipment, *Journal of Field Robotics*, vol. 29, no. 1, pp. 69-101, 2012.

[2] Santos T, Z. Liu and B. Hazel, Study and validation testing of the dynamics of a robotic system with flexible links and joints, *IEEE International Conferrence on Mechatronics and Automation, Beijing, China, 2-6 Aug.* 2015, pp. 1891-1897.

[3] Vu V.H, Z. Liu, M. Thomas and B. Hazel, Modal analysis of a light-weight robot with a rotating tool installed at the end effector, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 231, no. 9, pp. 1664-1676, 2015.

[4] Vu V.H, Z. Liu, M. Thomas, W. Li, and B. Hazel, Output-only identification of modal shape coupling in a flexible robot by vector autoregressive modeling, *Mechanism and Machine Theory*, vol. 97, pp. 141-154, 2016.

[5] Li. W, Vu V.H, Z. Liu, M. Thomas and B. Hazel, Extraction of modal parameters for identification of time-varying systems using data-driven stochastic subspace identification, *Journal of Vibration and Control*, vol. 24, nº 20. p. 4781-4796, 2018.

[6] Mokdad. F, Vu V.H, M. Thomas, F. Rafieian, Z. Liu and B. Hazel, Online modal analysis of a flexible robot during grinding, *Proceedings of the International operational modal analysis conference* (IOMAC 2011). Istanbul, Turkey, May-2011, paper 409, 10p.

[7] Rafieian F., Z. Liu, & B. Hazel, Dynamic model and modal testing for vibration analysis of robotic grinding process with a 6DOF flexible-joint manipulator. *In Proceedings of the IEEE International Conference on Mechatronics and Automation* (IEEE ICMA), Changchun, China, 2009.

[8] Vu V.H, Z. Liu, M. Thomas and A.M Tahvilian, Dynamic forces identification and modeling of a robotic grinding process, *Surveillance 8*, Oct. 2015, Roanne, Fr, 15p.

[9] Vu V.H, Z. Liu, M. Thomas and H. Bruce, Identification of frequency response functions of a flexible robot as tool-holder for robotic grinding process, *IEEE, IECON2016, Florence, Italy.*, 6p, Oct. 2016.

[10] Li, W., Vu V.H, Z. Liu, M. Thomas and B. Hazel, Application of adaptable functional series vector time-dependent autoregressive model for extraction of real modal parameters for identification of time-varying systems. *Measurement: Journal of the International Measurement*, vol. 103. p. 143-156, 2017.

[11] Li, W., Vu V.H, Z. Liu, M. Thomas and B. Hazel, Automatic structural mode extraction for time-varying structures via vector time-dependent autoregressive model. In *Surveillance 8*, Roanne, France, Oct. 20-21, 2015.

[12] Soderstrom T, H. Fan, B. Carlsson and S. Bigi, Least squares parameter estimation of continuous time ARX models from discrete-time data, *IEEE Transactions on Automatic Control,* vol. Vol 42, No .5, pp. 659-673, 1997.

[13] Alessandro C., The role of vector autoregressive modeling in predictor-based subspace identification, *Automatica,* vol. 43, no. 6, pp. 1034-1048, 2006.

[14] Da Silva S, M. Dias J. and V. Lopes J, Damage detection in a benchmark structure using AR-ARX models and statistical pattern recognition. *J. Braz. Soc. Mech. Sci. & Eng*. vol.29 no.2, Rio de Janeiro Apr./June 2007.

[15] Ljung L., System Identification: Theory for the User, *Prentice-Hall, 2nd Ed*, 1999.

[16] Juang J-N, State space system realization with input and output data correlation, *NASA Technical Paper 3622,* 1997.

[17] Raath A. D. and C. C. Waveren, A time domain approach to load reconstruction for durability testing, *Engineering Failure Analysis,* vol. Vol.5, No. 2., pp. 113-119, 1998.

[18] Wan Z., T. Wang, and W. Zheng, An improved state space method for force identification based on function interpolation in the presence of large noise, *Journal of Vibroengineering,* vol. 19, no. 2, pp. 751-768, 2017.

[19] Strejc V, State Space Theory of Discrete Linear Control. *IEEE Transactions on Systems*, 12 (5), 1982.

Use of virtual sensors for the analysis of forces exerted by the load inside a tumbling mill

José Venegas Pulgar[1], M. Aníbal Valenzuela[2] and Cristián Molina Vicuña[1]

[1]Laboratorio de Vibraciones Mecánicas, Departamento de Ingeniería Mecánica,
Universidad de Concepción, Chile
{jvenegasp, crimolin}@udec.cl
[2]Laboratorio de Accionamientos Eléctricos, Departamento de Ingeniería Eléctrica,
Universidad de Concepción, Chile
anivalenz@udec.cl

Abstract

In this paper, a methodology is presented to obtain average representative forces exerted by the load inside a tumbling mill on the different faces of the lifters and liners, which are directly related to its power consumption. The methodology is based on the use of virtual sensors included in DEM simulations combined with signal processing and allows obtaining the magnitude of the forces based on the angular position of the lifters as the mill rotates. The methodology is validated by comparing numerical and experimental results obtained from a test bench mill. The variables considered are the power, movement of the load inside the mill, and average forces. The latter are experimentally measured using instrumented lifters specially designed for this task. The results obtained show differences in the magnitude of the average forces in specific angular positions, depending on the operating conditions of the mill. These differences explain the behavior of power consumption with respect to operating conditions reported in the literature.

1 Introduction

Tumbling mills are critical machines of the mining industry. They are used to reduce the size of mineral particles, and their operation has associated high economic costs. The grinding of minerals requires high energy consumption and represents the most expensive stage in the production of metals. The economies of scale and the decrease in the grade of the ores, has led to the development of large grinding mills that currently reach throughput of 80,000 - 100,000 ton/day with powers of up to 28 [MW] [1][2].

The main component of the mill is the horizontal hollow cylinder called drum, inside which the ore is grinded (Figure 1a). The drum rotates on its axis and is delimited at its sides by two ends through which occurs the entry and exit of the material, respectively. Fixed to the inner wall of the drum are the coatings, composed by liners and lifters. The liners protect the drum from wear due to contact with the particles, while the lifters transmit the energy associated with the rotation of the mill to the load, producing its movement inside the mill. The comminution of the mineral occurs due to the contact forces that are generated between different particles and between particles and the internal surfaces of the drum.

The liners and lifters wear out over time as a result of the continuous contact with the particles inside the mill. The replacement of these elements constitute the primary maintenance operations and generate high costs associated not only with the purchase of replacement parts and labor but also with production losses during

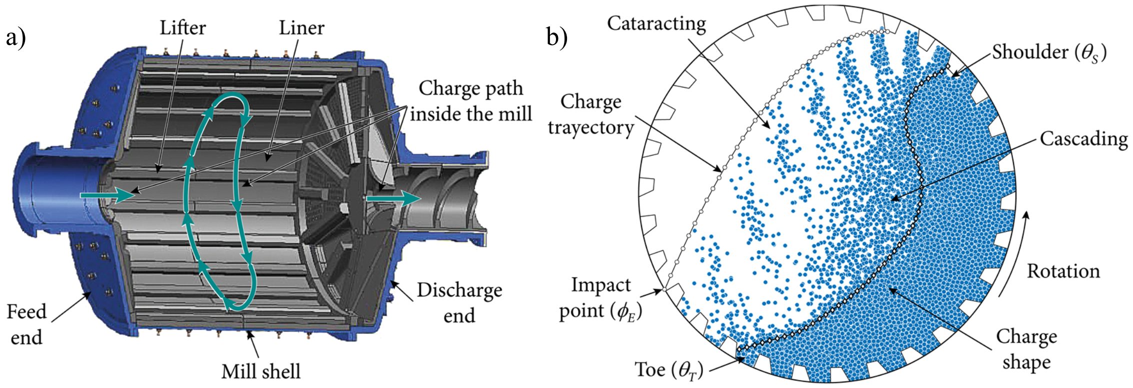


Figure 1: Tumbling mill. (a) General view of the drum. (b) View of the load inside the drum.

maintenance works. In addition, the continuous wear of the coatings produces changes in the power consumption and throughput of the mills.

While rotating, the drum lifts the grinding load along one side of the mill until reaching the point called shoulder of the load, as shown in Figure 1b. In this position, the particles located near the lifters begin to move independently of the movement of the drum and then fall describing free fall movements called cascade type or cataract type. The particles that describe cascade movements from the shoulder characterize by falling continuously, while those that describe cataract movements do so in the form of waves driven by the lifters. In the movement that describes the load inside the mill, a set of relevant angular positions, measured with respect to its axis of rotation, are identified: the position of the shoulder (θ_S), the position of the point of impact (ϕ_E) and the position of the toe (θ_T). The point of impact corresponds to the highest position at which the particles fall from the shoulder on the opposite end of the mill ring. The position of the toe is where the lifters begin to lift the load after it falls from the shoulder.

Figure 1b presents an example of the disposition of the particles inside the mills, showing the shape and trajectory of the load, the position of the toe, of the shoulder, and the point of impact. The shape of the load corresponds to the shape adopted by the set of particles that do not fall in free fall (cascading or cataracting) while the mill rotates, which commonly compares to the shape of a kidney bounded at its ends by the toe and the shoulder. Most of the particles fall from the areas near the shoulder on the internal surface of the shape of the kidney, describing a cascade-type movement, as can be seen in Figure 1. The lifters lift a portion of the particles in the shoulder to higher positions and then fall on the toe or the coatings of the drum near the toe describing a cataract movement. The lifters that leave the position of the shoulder drag small portions of particles that become independent from it gradually, forming waves of particles that are thrown into the free space inside the drum. The formation of these waves then depends on the passage of the lifters out of the shape of the load and, therefore, the impacts of the particles that describe a cataract movement on the area near the toe are not continuous, but linked to the movement of lifters. The trajectory of the load corresponds to the free fall movement that describes the particle that reaches the point of impact.

The Discrete Elements Method (DEM) is a numerical methodology that describes the behavior of granular materials. It allows simulating the movement of each of the particles forming the grinding load inside the mill by modeling the interactions between the different particles and between particles and surfaces, and solving the equations of motion of each particle. DEM has been used by multiple researchers to study tumbling mills focused on, for example, the analysis of the load movement [2][3][4][5][6][7][8][9][10][11][12][13][14][15][16][21][23][26][27][28][32][33][34][45], the study of wear of coatings [2][14][15][30][43][44] and the modeling of the comminution process [16][5][8][17][18][19][20][22][31][35][37][38][39][40][41][42]. This paper focuses on analyzing the power requirements of the mills, a topic that has also been discussed in the literature [13][14][45][49][21][24][25][26][29][33]. Different from other researches, this is done by determining the average forces exerted by the particles on the lifters and liners as a function of the angular position in which they are located. Some studies relate globally the behavior of the load with the contact forces by using different methods [47][48][49][50][51][52][54][55][56]. The correlation between the average forces and the power obtained in this work allows identifying the physical phenomena that explain the observed power variations as a function of the operating conditions.

2 Test bench: SetupD100

The analyses are carried out based on a laboratory scale mill, called SetupD100, shown in Figure 2a. This mill consists of three main components: the ring, a back cover, and a front cover. The ring is a hollow cylinder representing the drum of the mill with lifters mounted on its inside. The ring and lifters are made of technyl. The ring has an internal diameter of 945 [mm], an internal length of 60 [mm] and is delimited at its ends by the back and front cover, respectively. Both covers are made of acrylic. The back cover is gray, while the front is transparent, which allows observing the movement of the load while the mill runs. The mill is connected to the electric drive by a drive shaft in a cantilever arrangement. The drive includes a frequency converter that allows controlling the speed of rotation.

2.1 Lifter geometry

The internal geometry of the mills is one of the main aspects to analyze in order to understand the behavior of the load inside the mill. It is defined mainly by the number of lifters (N_{lft}) and their geometry. Figure 2a shows the dimensions of the lifters installed in the test bench mill.

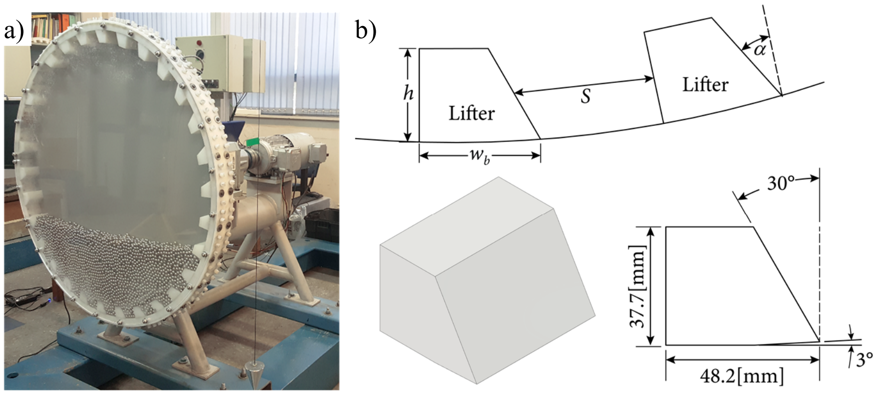


Figure 2: (a) View of the Setup D100. (b) Dimensions of the lifters.

2.2 Operating conditions

The rotating speed and fill level of the mill define its operating condition. The numerical and experimental analyses presented in this paper consider 81 different combinations of speed and fill levels presented in Table 1.

The rotation speed of the mill is defined as a fraction of its critical speed (N). The critical speed of a mill (N_c) corresponds to the speed of rotation from which the load begins to centrifuge, adhering to the internal surfaces of the drum. It is calculated as [47]:

$$N_c = \frac{42.3}{\sqrt{D_M}} \tag{1}$$

The critical speed of the SetupD100 is 4.556 [rad/s], and the 9 speeds analyzed vary from 55% to 95% of it.

The fill level (J_c) is the fraction of the internal volume of the mill that is occupied by the grinding load. The SetupD100 has an internal volume of 0.0396 [m^3] and is filled with 11 [mm] diameter steel balls. The 9 fill levels analyzed range from 25% to 45% of the mill's internal volume, with 2.5% jumps

3 Numerical modeling of the SetupD100

The numerical model of the test bench consists of two main components: a geometric model and a contact model. The geometric model represents the surfaces of the mill with which the particles come into contact, while the contact model describes the interactions between the particles located inside the mill and the components of the geometric model, and between different particles.

3.1 Geometric model of the SetupD100

The geometric model used in this investigation is composed of 3 elements: the ring and the two covers. In the geometric model, the ring and the lifters are considered as a single element. Figure 3a shows the three components of the geometric model, and Figure 3b shows a view of the load inside the mill together with the cartesian system used as a reference for the analyses.

Fraction of the critical speed	Rotation speed, [rad/s]
55%	2.506
60%	2.734
65%	2.961
70%	3.190
75%	3.418
80%	3.645
85%	3.874
90%	4.101
95%	4.329

Fill level	Mass, [kg]
25%	46.25
27.5%	50.87
30%	55.49
32.5%	60.12
35%	64.74
37.5%	69.37
40%	73.99
42.5%	78.61
45%	83.24

Table 1: Operating conditions.

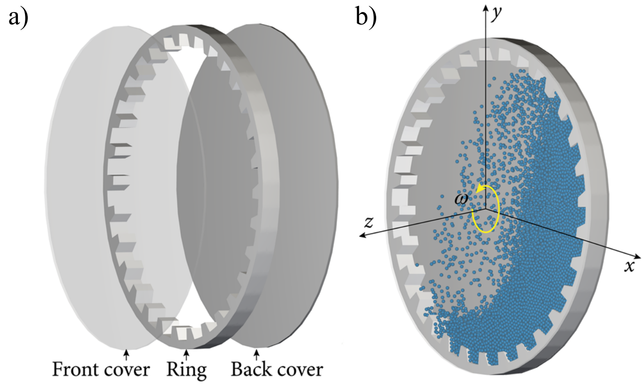


Figure 3: (a) Components of the geometric model. (b) Cartesian reference system.

3.2 Contact model

The particles inside the mill can contact other particles, the covers, or the ring. This means the existence of three different contact types: steel-steel contact between the steel balls, steel-acrylic contact between the steel balls and the acrylic covers, and steel-technyl contact between the steel balls and the ring. The contact models used in DEM allow calculating the forces associated with the contacts to which all the particles are subject, but for this, it is necessary to define the physical parameters that characterize all possible contacts. These parameters are the static friction coefficient, the dynamic friction coefficient, the coefficient of restitution, and the coefficient of rolling resistance. Table 2 shows the values of the contact parameters used in the DEM simulations.

3.3 Power due to the movement of the load inside the mill

Considering the SetupD100 operating as shown in Figure 3b, it can be noted that in any instant of time, not all particles are in contact with the internal surfaces of the mill. It is also noted that a particle can be in contact with the internal surfaces of the mill in more than one point (maximum three) and that there is a given number of contacts (n) between particles and internal surfaces of the mill.

Now, consider a particle that is in contact with one of the internal surfaces of the mill, and that this contact i occurs in a position $\vec{r}_i$ with respect to the axis of rotation of the mill. Let $\vec{F}_i$ be the force exerted by the particle on the point of contact i, as shown in Figure 4a:

$$\vec{r}_i = r_{x,i}\hat{i} + r_{y,i}\hat{j} \tag{2}$$

$$\vec{F}_i = F_{x,i}\hat{i} + F_{y,i}\hat{j} \tag{3}$$

The torque exerted by the particle on the surface at contact i is, thus, given by:

$$\vec{T}_i = \vec{r}_i \times \vec{F}_i \tag{4}$$

Finally, taking into account the n existing contacts between particles and internal surfaces of the mill during a time instant, the torque associated with the movement of the particles inside the mill is given by:

$$\vec{T}_{Mill} = \sum_{i=1}^{n} \vec{T}_i \tag{5}$$

The torque allows to calculate the power due to the operation of the mill at speed ω:

Contact	Coefficient of rolling resistance (μ_r)	Static friction coefficient (μ_s)	Dynamic friction coefficient (μ_k)	Coefficient of restitution (c_R)
	0.01	-	-	-
Steel–Steel	-	0.43	0.36	0.91
Steel–Acrylic	-	0.42	0.38	0.44
Steel – Technyl	-	0.23	0.22	0.69

Table 2: Contact parameters used in the DEM simulations.

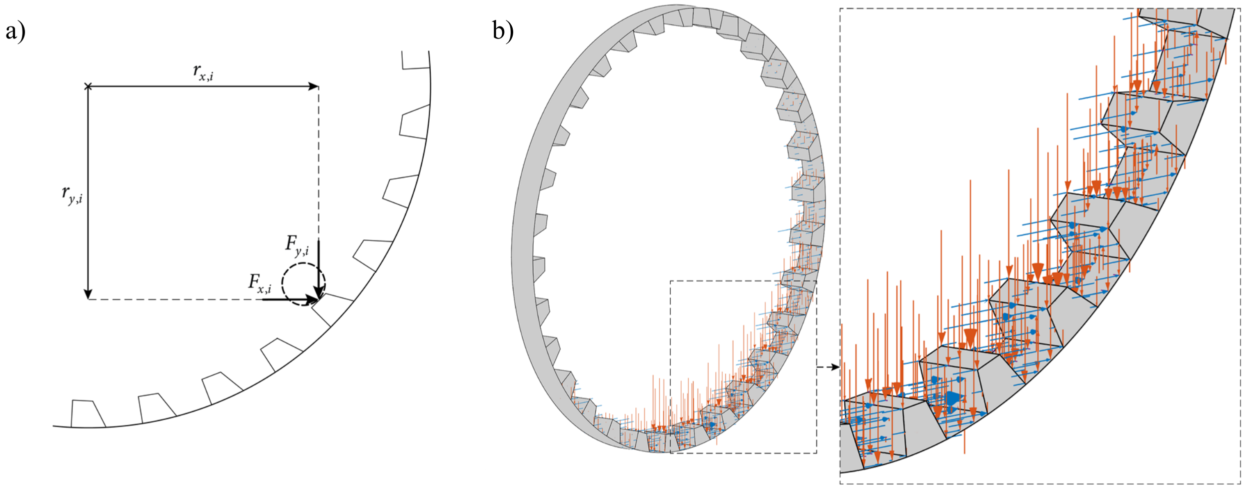


Figure 4: (a) Contact forces of a single particle in contact with the ring. (b) All contact forces acting on the ring with a close view of the contact forces.

$$P_{Mill} = T_{Mill}\omega \tag{6}$$

Thus, the power associated with the movement of the mill originates from the contact between the particles and their internal surfaces. The contacts between the particles contribute to the power indirectly by action/reaction forces, and the Discrete Element Method (DEM) is based on calculating the forces associated with the contacts between particles and between particles and surfaces.

3.3.1 Forces acting on the ring and covers

Figure 4b shows the components of the nodal forces acting on the ring at a given time instant, with a close view of a section of the ring. The blue arrows represent the components of the forces on the *x*-axis; the red arrows represent the components of the forces on the *y*-axis. There are also small purple arrows representing the force components on the *z*-axis. The size of the arrows is scaled based on the magnitude of the forces they represent. As observed, the components of the forces on the *z*-axis are much smaller than the other two components, being so small that they are not recognizable in Figure 4b. The latter, added to the fact that these forces do not contribute to the power, allows treating the analysis of the forces as bidimensional.

Consider the disposition of particles during a given time instant shown in Figure 5a and the forces excerted by these particles on the ring, shown in Figure 5b. It can be seen that the magnitude of the forces exerted by the particles impacting the ring before the toe has a low magnitude and that the amplitude of the forces grows as they come into contact with the ring in the zone between the toe and the shoulder. Also, it is possible to notice that the vertical components of the nodal forces act mainly in the negative direction while the forces in the horizontal direction are positive or negative depending on the face of the lifter in which they act. The sense of the horizontal forces determines the sense of the torque they generate.

Figure 6 shows the forces of Figure 5 decomposed in the ring tangential direction. These components are those that contribute to the torque associated with the movement of the particles. It can be seen that the forces acting on the back face of the lifters (purple arrows) generate torque in the same direction as the rotation of the

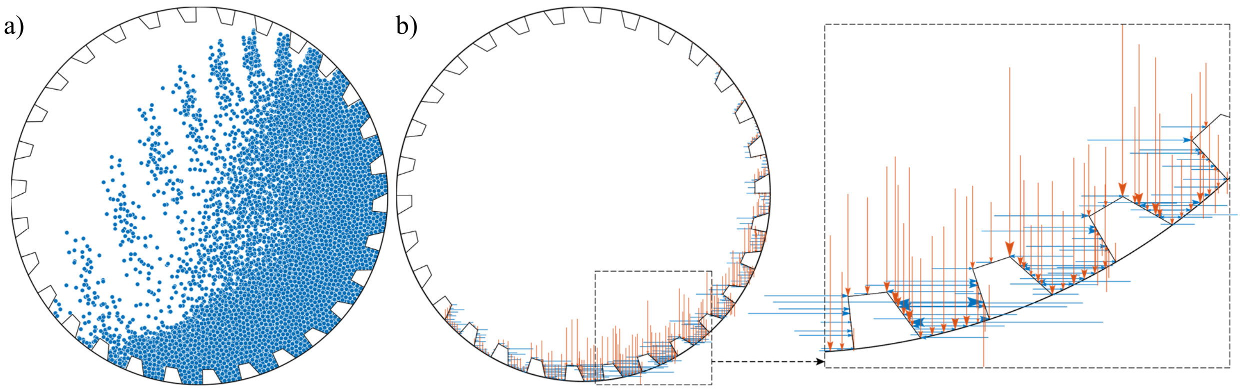


Figure 5: (a) Disposition of particles at a given time instant. (b) Forces excerted by the particles on the lifters and liners at a given instant of time.

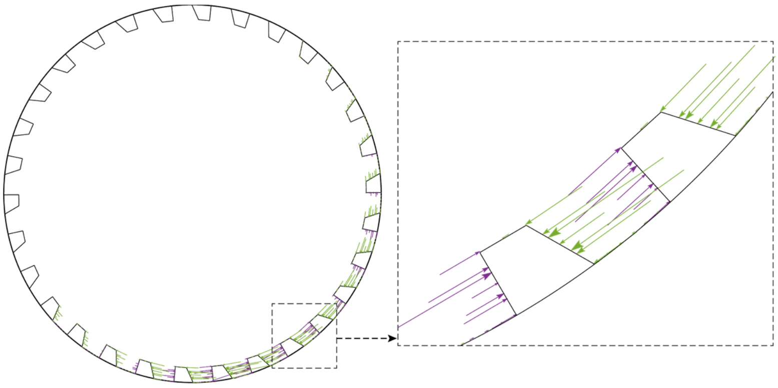


Figure 6: Decomposition of forces in tangential direction.

mill, while the forces acting on the front face of the lifters (green arrows) generate torque in the opposite direction. It is also possible to notice that the components of the forces that contribute to the torque and that act on the liners or the top faces of the lifters are of very low amplitude when compared with the components that act on the front and back faces. Finally, it can be seen that the magnitude of the tangential forces varies significantly with the angular position. In this work, the power associated with the forces exerted by the particles on the ring and obtained using DEM is called $P_{DEM,Ring}$.

The main consequence of this analysis is that, depending on the face of the lifter in which the forces act, they can contribute or oppose the total torque exerted by the load, which determines the required driving torque. The evolution of these forces, depending on the internal geometry of the mill and the operating conditions, have a significant influence on the power consumption associated with its operation.

Figure 7a shows the components of the forces in *x*-axis and *y*-axis acting due to friction on one of the acrylic covers of the mill. Figure 7b shows the tangential decomposition of these forces, where the size of the arrows is according to their magnitude. As observed, the forces on the covers that generate torques are of similar magnitude to those acting on the ring. The magnitude of the power associated with the back cover and front cover is similar and, therefore, both covers are treated indistinctly in this work.

The power associated with the movement of both covers is defined as $P_{DEM,Covers}$ and is calculated by adding the powers associated with each cover.

3.3.2 Total power due to the movement of the load inside the mill

The total power associated with the operation of the mill obtained by modeling the SetupD100 using DEM corresponds to the sum of the powers associated with each of its components:

$$P_{DEM,Mill} = P_{DEM,Ring} + P_{DEM,Cover1} + P_{DEM,Cover2} \tag{7}$$

Most of the power associated with the movement of particles inside industrial scale mills comes from the contact forces acting on the liners and lifters. In the case of the power associated with the operation of the SetupD100, both the power associated with the ring and the covers are of similar magnitude because the length of the ring is too short for its diameter.

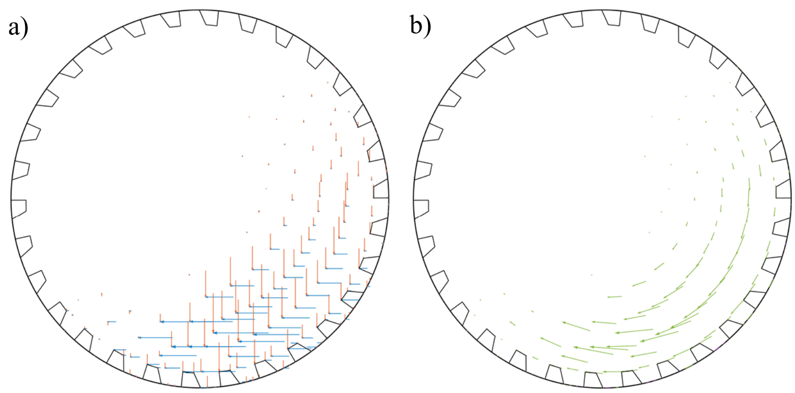


Figure 7: Forces acting on one cover (a) in *x* and *y* direction. (b) Tangential decomposition of the forces.

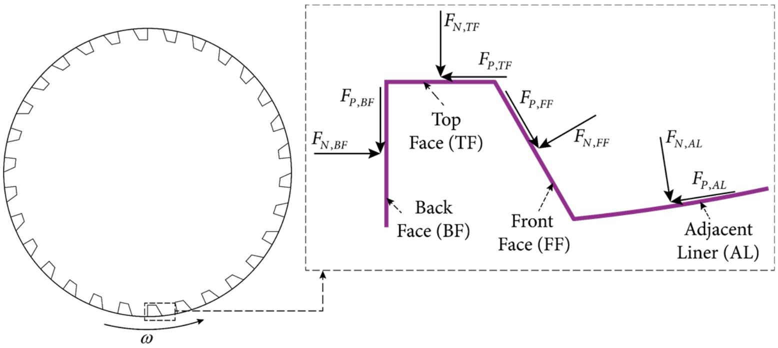


Figure 8: Normal and parallel forces acting on the 3 faces of the lifter and adyacent liner.

4 Model of forces acting on the lifters

As explained in section 3.3.1, the magnitude of the forces exerted by the particles on the ring when the mill rotates depends on the angular position of the liners and lifters, and on the lifter face on which they act. Motivated by this, a force model is developed, based on DEM simulation results. The model allows obtaining the magnitude of the average forces acting on the different faces of the lifters and liners as a function of its angular position θ while the mill operates. This model groups the contact forces exerted by the particles on the liners and each face of the lifters of the mill. The internal geometry of the mill divides into four faces: the front face (FF), the back face (BF), the top face (TF) and the adjacent liner (AL). The forces acting on the ring are divided into 8 forces that act normal and tangential to each face, as shown in Figure 8.

4.1 Modification of the geometric model: implementation of virtual sensors

To obtain the magnitude of the average forces exerted by the particles on the 3 faces of the lifters and adjacent liners as a function of their angular position, a set of three virtual sensors is incorporated in the geometrical model of the DEM model of the SetupD100, as shown in Figure 9a. The virtual sensors (VS) consist of 4 independent sheets (one for each face of the lifter plus the adjacent liner) fixed to the lifters they cover. The fact that the sheets are independent geometries allows obtaining the resultant force exerted by the particles on each sheet as a function of time separately.

Figure 9b shows an example of the forces acting on one of the sheets covering the front face (FF) of the lifter of one of the virtual sensors, as directly obtained from the DEM software, that is, in terms of the *xy*-coordinates. Since there are 3 virtual sensors with their respective 4 sheets, a total of 12 *x* and *y* force signals as a function of time is obtained from each simulation.

a)

b)

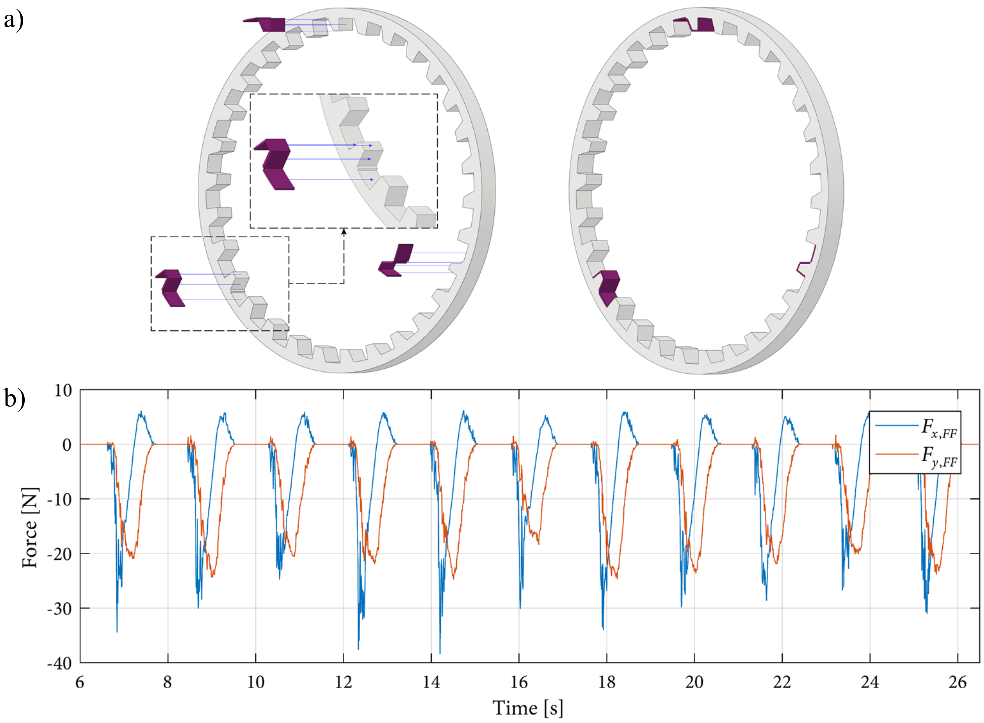


Figure 9: (a) Set of virtual sensors included in the geometric model. (b) Example of time history of force signal in *xy* coordinate for the front face of one lifter.

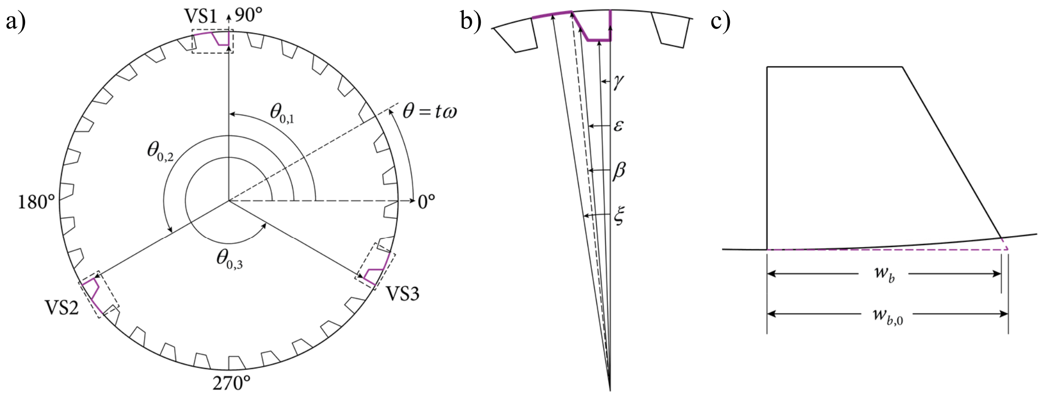


Figure 10: a) Starting position of the virtual lifters. b) Angles used to define the center position of the lifter faces and adjacent liner. c) Projected width of lifter.

4.2 Transformation of forces into local coordinates

The first step in the processing of the forces is to transform the forces captured by the virtual sensors, based on the global *xy* coordinates (Figure 9b), in terms of the local coordinates defined for each face according to Figure 8. The global coordinates are fixed in space so that the mill rotates with respect to the *xy* plane. On the other hand, the virtual sensors move with the lifters on which they are mounted and, consequently, rotate with the mill. Because of this, all local coordinates, defined for each of the faces of the VS, also rotate with the mill. Hence, to transform the forces from global to local coordinates, it is necessary to know the location of each of the faces of the VS as a function of time, based on the rotation of the mill.

4.2.1 Angular displacement equations of virtual sensor faces

In order to know the location of the different faces of the virtual sensors with respect to the axis of the mill, the equations describing the angular displacement of the centers of the 3 lifter faces and adjacent liner are considered. Figure 10a shows the starting position of the three virtual sensors (VS), where the back face of VS1 is located at 90° with respect to the horizontal, that is, $\theta_{0,1} = \pi/2$. The back face of the VS2 and VS3 are located at 120° and 240° from the back face of the SV1, respectively. The angular position of the centers of the other faces is defined relative to the back face of the respective virtual sensor, as shown in Figure 10b. The angles shown in the figure are defined according to the geometry of the lifters and the ring. Due to how the lifters are installed in the ring (Figure 10c) it is convenient to define a projected width of the lifter $w_{b,0}$, which represents the width of the base of the lifter if the curvature of the ring did not exist. Doing so allows defining the angles between the centers of the faces with respect to the back face of the lifter as:

$$\gamma = \tan^{-1}\left(\frac{w_{b,0} - h\tan(\alpha)}{D_M - 2h}\right) \tag{8}$$

$$\varepsilon = \tan^{-1}\left(\frac{2w_{b,0} - h\tan(\alpha)}{D_M - h}\right) \tag{9}$$

$$\beta = 2(\varepsilon - \gamma) \tag{10}$$

$$\xi = \frac{\pi}{N_{lft}} + \frac{\beta}{2} \tag{11}$$

For the geometry and number of lifters of the SetupD100, the angles are $\gamma = 1.836°$, $\varepsilon = 4.762°$, $\beta = 5.853°$ y $\xi = 8.927°$. After locating the different faces of the VS with respect to their corresponding back faces, it is possible to describe the movement of each of the faces of the 3 VS as a function of time, based on the rotation speed of the mill as follows:

$$\theta_{BF,i}(t) = \theta_{0,i} + \omega t \tag{12}$$

$$\theta_{TF,i}(t) = \theta_{BF,i} + \gamma \tag{13}$$

$$\theta_{FF,i}(t) = \theta_{BF,i} + \varepsilon \tag{14}$$

$$\theta_{AL,i}(t) = \theta_{BF,i} + \xi \tag{15}$$

Where ω is the rotation speed of the mill, t corresponds to time, the index $i=1,2,3$ refers to the virtual sensor and $\theta_{0,i}$ represents the initial position of the back face of the virtual sensor i.

4.2.2 Transformation of forces into local coordinates

Once the location of the 4 faces of the 3 virtual sensors is known as a function of time, it is possible to transform the forces obtained from the virtual sensors from global xy coordinates into local coordinates (normal and parallel) by using the following equations:

For forces acting on the front face:

$$F_{N,FF,i}(t)=F_{x,FF,i}(t)\sin\left(\alpha+\theta_{FF,i}(t)\right)-F_{y,FF,i}(t)\cos\left(\alpha+\theta_{FF,i}(t)\right) \tag{16}$$
$$F_{P,FF,i}(t)=F_{x,FF,i}(t)\cos\left(\alpha+\theta_{FF,i}(t)\right)-F_{y,FF,i}(t)\sin\left(\alpha+\theta_{FF,i}(t)\right) \tag{17}$$

For forces acting on the adjacent liner:

$$F_{N,AL,i}(t)=F_{x,AL,i}(t)\cos\left(\theta_{AL,i}(t)\right)+F_{y,AL,i}(t)\sin\left(\theta_{AL,i}(t)\right) \tag{18}$$
$$F_{P,AL,i}(t)=F_{x,AL,i}(t)\sin\left(\theta_{AL,i}(t)\right)+F_{y,AL,i}(t)\cos\left(\theta_{AL,i}(t)\right) \tag{19}$$

For forces acting on the back face:

$$F_{N,BF,i}(t)=-F_{x,BF,i}(t)\sin\left(\theta_{BF,i}(t)\right)+F_{y,BF,i}(t)\cos\left(\theta_{BF,i}(t)\right) \tag{20}$$
$$F_{P,BF,i}(t)=F_{x,BF,i}(t)\cos\left(\theta_{BF,i}(t)\right)+F_{y,BF,i}(t)\sin\left(\theta_{BF,i}(t)\right) \tag{21}$$

For forces acting on the top face:

$$F_{N,TF,i}(t)=F_{x,TF,i}(t)\cos\left(\theta_{TF,i}(t)\right)+F_{y,TF,i}(t)\sin\left(\theta_{TF,i}(t)\right) \tag{22}$$
$$F_{P,TF,i}(t)=F_{x,TF,i}(t)\sin\left(\theta_{TF,i}(t)\right)-F_{y,TF,i}(t)\cos\left(\theta_{TF,i}(t)\right) \tag{23}$$

Figure 11a shows the forces of Figure 9b, acting on the front face of one of the virtual sensors, after the change of coordinates. After the change of coordinates, it is possible to notice that the magnitude of the normal forces acting on the faces of the lifters is significantly larger than that of the parallel forces.

4.1 Separation in force pulses

As observed in Figure 11a, the magnitude of the forces changes from zero to non-zero in consecutive time intervals, a situation that repeats with every rotation of the mill. This is because the VS is not in contact with the particles when it is located between the shoulder and the toe or the point of impact (whatever occurs first) and, therefore, is not subjected to forces during this time interval. Contrary, when the VS is in contact with particles, the force is non-zero. This process of loading and unloading the VSs is approximately periodic so that lifters and liners are subject to similar forces as they pass through the same angular position.

Based on the above, the forces are subjected to a process of pulse separation. A pulse corresponds to a set of the 8 forces (two for each face, normal and parallel) exerted by the particles on one of the virtual sensors

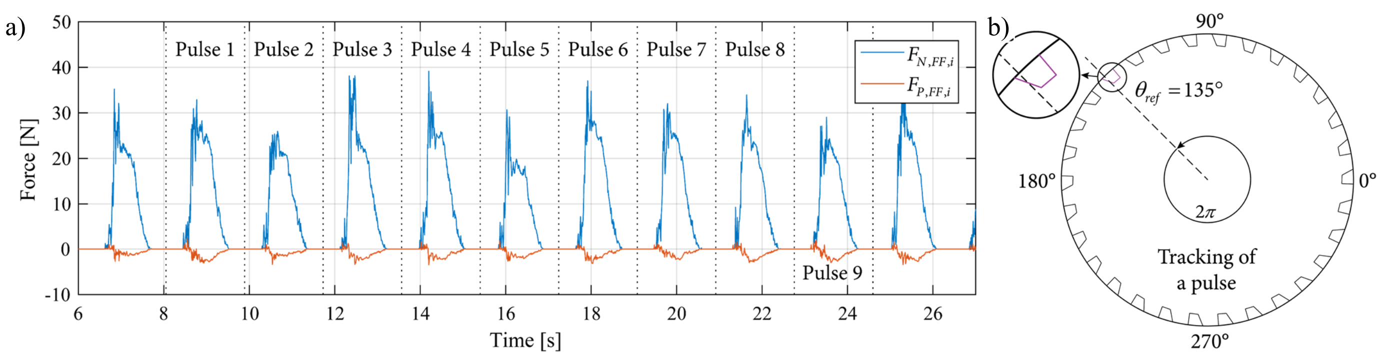


Figure 11: (a) Forces acting on the front face of a SV. Dotted lines mark the beginning and end of the force pulses. (b) Path followed by a SV for pulse separation with reference position of 135°.

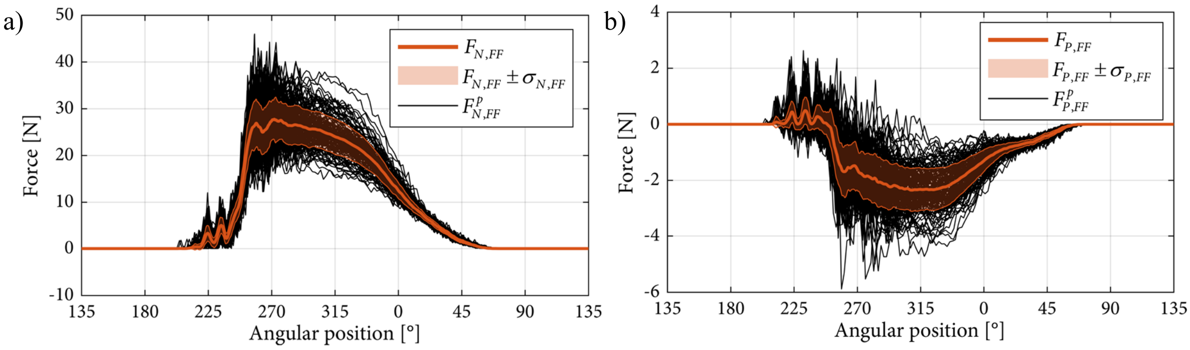


Figure 12: Averaged (red) and individual (black) pulse forces acting on the front face of the lifter. (a) Normal force. (b) Parallel force.

for a complete revolution of the mill around its axis of rotation. The pulse separation allows relating the magnitude of the forces exerted by the particles on the virtual sensors as a function of the angular position instead of time.

The pulse separation is carried out using an algorithm that defines a reference position (θ_{ref}) and identifies the instants of time where the front face of the VS pass through said position. Figure 11b shows the path followed by a VS during which one pulse is obtained, with respect to the reference position (135° in this case). Thus, the pulses correspond to the set of forces exerted by the particles during one rotational period of the mill based on a reference position. In Figure 11a, the vertical dotted lines represent the initial and final time instants of the different pulses. The first pulse is not considered representative because the movement of the load develops only after the first seconds of the simulation.

4.2 Averaging process of pulse forces

Once separated into pulses, the forces are subjected to a process of averaging for deterministic/random separation. The process consists of averaging the magnitude of the pulse forces of the respective faces of the SV when they are in corresponding angular positions. The result of the process is the set of 8 average pulse forces that act on each of the faces of the lifters as a function of the angular position: $F_{N,FF}$, $F_{P,FF}$, $F_{N,AL}$, $F_{P,AL}$, $F_{N,BF}$, $F_{P,BF}$, $F_{N,TF}$ y $F_{P,TF}$. Figure 12a and Figure 12b show, respectively, the average pulse forces $F_{N,FF}$ and $F_{T,FF}$ superimposed on the individual pulses from which they are calculated. Figure 13 shows the forces of Figure 11 (in blue), along with the average force pulses (in red) and the random part of the forces (in yellow). In general, the magnitude of the random part of the forces is low compared to the magnitude of the average forces, especially for normal forces. The average forces represent the stationary forces acting on any of the lifters as

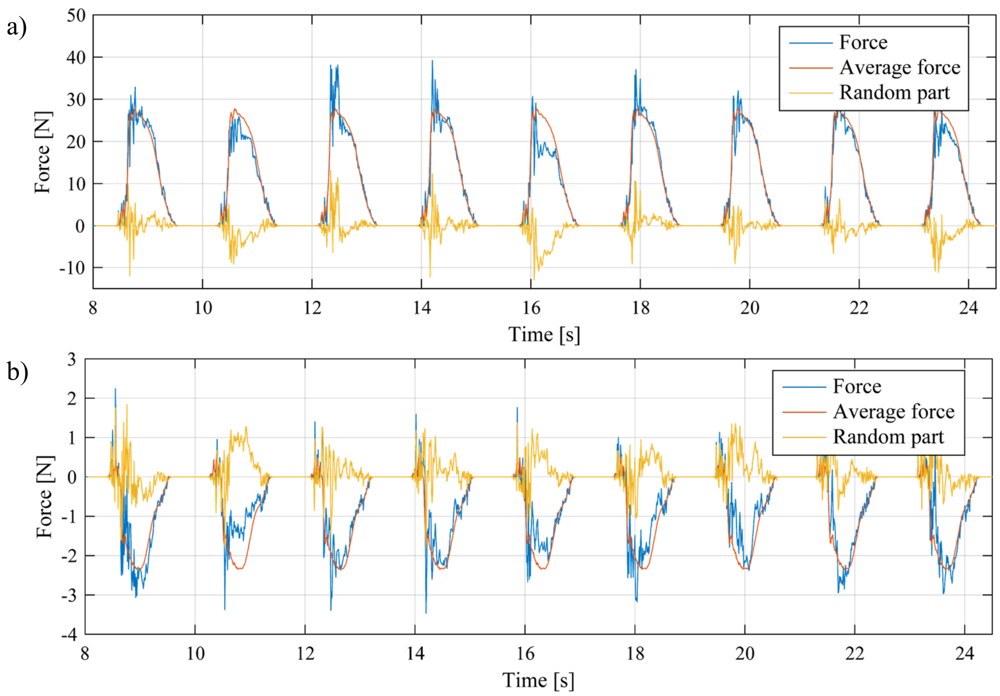


Figure 13: Average/random decomposition of the forces acting on the front face of the lifter. a) Normal force. b) Parallel force.

a function of the angular position, being the random part responsible for deviating the magnitude of the average forces until reaching the magnitude of the individual pulses.

4.3 Calculation of torques and power from the forces acting on the virtual sensors

4.3.1 Time delay of virtual sensor forces

The ring of the mill is composed of a set of liners and lifters, of identical geometry, located in different angular positions. As the VS collect the forces that act on the 3 faces of a lifter and its adjacent liner, it is possible to calculate the power associated with the movement of the ring from the forces they register, assuming that the average forces acting on a lifter *i* is equal to the forces acting on the reference lifter, when its located in the same angular position.

It is considered that the stationary force acting on each face of a lifter *i* is equal to the force of the respective face of the reference lifter, including the time delay according to the difference between their angular positions. That is:

$$F_{i,j}(t) = F_{1,j}\left(t - (i-1)T_{lft}\right) \tag{24}$$

Where:

$$T_{lft} = \frac{2\pi}{\omega N_{lft}} \tag{25}$$

N_{lft} is the number of lifters, j represents the force (i.e., face, normal or parallel), i represents the lifter for which the force is being expressed based on the reference lifter on which the stationary force $F_{1,j}$ acts, t is time and T_{lft} is the lifter pass period. Figure 14 illustrates the time delay process, where the forces acting on the rest of the lifters are determined from the forces acting on the lifter 1. Figure 14a shows the result of the phase shift process for 3 lifters. Figure 14b shows the forces acting on all lifters at a given instant, calculated from the average forces acting on the reference lifter.

4.3.2 Calculation of the driving torque and power

Of the eight forces shown in Figure 8, five of them have tangential components and, therefore, exert torques in the direction of rotation of the mill. This set of forces is determinant in the driving torque and, thus, in the

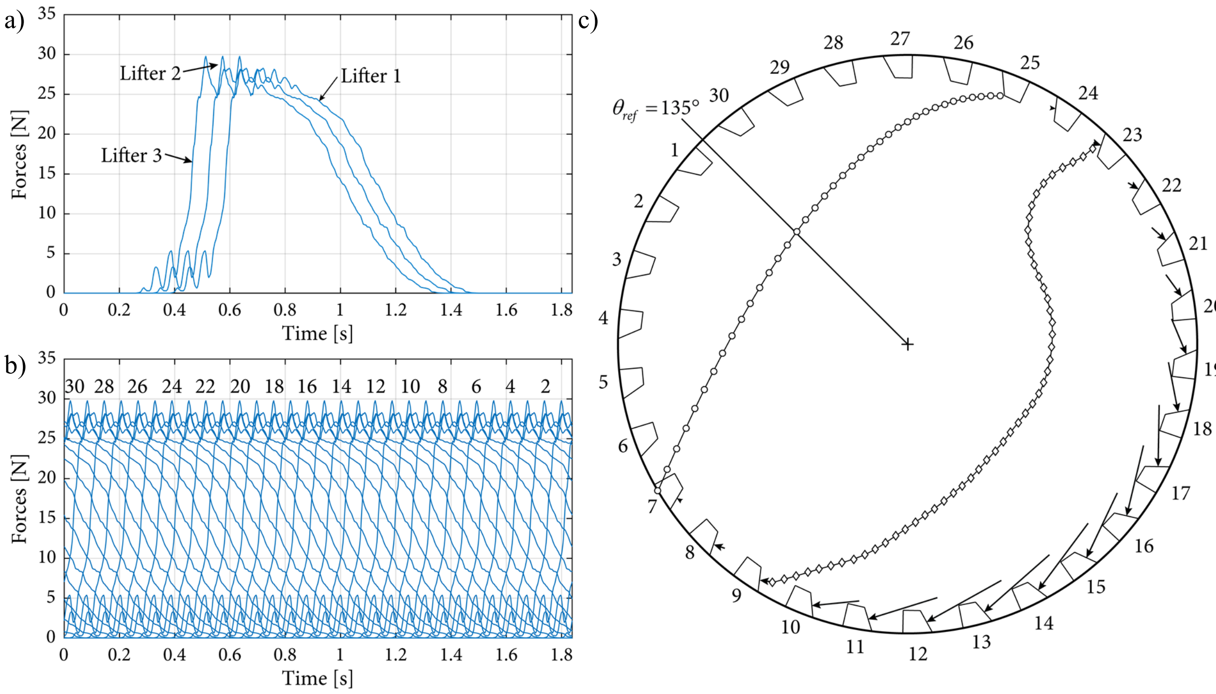


Figure 14: Illustration of the time delay process for determining the forces acting on all lifters and liners. (a) Time delay for three lifters. (b) Time delay for all lifters. (c) Global representation of the forces.

power consumption. Of these five forces, four exert resistant torques, as shown in Figure 6. The force $F_{N,BF}$, acting on the back face of the lifter, exerts torque in the rotation direction, thus favoring the mill's movement and diminishing the required driving torque. As the forces are treated separately for each face, it is possible to obtain the torque associated with the front face, top face, back face, and adjacent liner independently. These torques consider the total number of lifters, due to the time delay process explained in the previous section.

The torques exerted by the different forces acting on any face of a single lifter are given by:

$$\vec{T}_{i,j} = \vec{r}_j \times \vec{F}_{i,j} \tag{26}$$

Where $T_{i,j}$ is the torque exerted by force $F_{i,j}$ acting on the face j of the lifter i, and r_j is the arm of this force. Hence, the torques associated with the different forces acting on each of the faces of a lifter i are given by:

For normal and tangential forces acting on the front face of the lifter (FF):

$$T_{FF,i} = \left(F_{N,FF,i}\cos(\alpha) + F_{P,FF,i}\sin(\alpha)\right)\left(0.25D_M\left(1+\cos(\beta)\right) - h\right) \tag{27}$$

For the tangential force acting on the adjacent lifter (AL):

$$T_{AL,i} = 0.5F_{P,AL,i}D_M \tag{28}$$

For the normal force acting on the back face of the lifter (BF):

$$T_{BF,i} = -0.5F_{N,BF,i}\left(D_M - h\right) \tag{29}$$

For the tangential force acting on the top face of the lifter (TF):

$$T_{TF,i} = F_{P,TF,i}\left(0.5D_M - h\right) \tag{30}$$

And the total torque contribution per face is obtained by considering the respective torques of all N_{lft} lifters:

$$T_{FF} = \sum_{i=1}^{N_{lft}} T_{FF,i};\; T_{AL} = \sum_{i=1}^{N_{lft}} T_{AL,i};\; T_{BF} = \sum_{i=1}^{N_{lft}} T_{BF,i};\; T_{TF} = \sum_{i=1}^{N_{lft}} T_{TF,i} \tag{31}$$

The total torque per face of the lifter and adjacent liner includes the effect of the magnitude change of the forces and its varying angular position as the mill rotates. The total torques associated with each face, obtained from the set of forces acting on the ring, are shown in Figure 15 as a function of time. The figure also indicates the average values.

Finally, the total power due to the movement of the load inside the mill is given by the sum of the average torques associated to each face:

$$P_{Forces} = \left(\overline{T}_{FF} + \overline{T}_{AL} + \overline{T}_{BF} + \overline{T}_{TF}\right)\cdot\omega \tag{32}$$

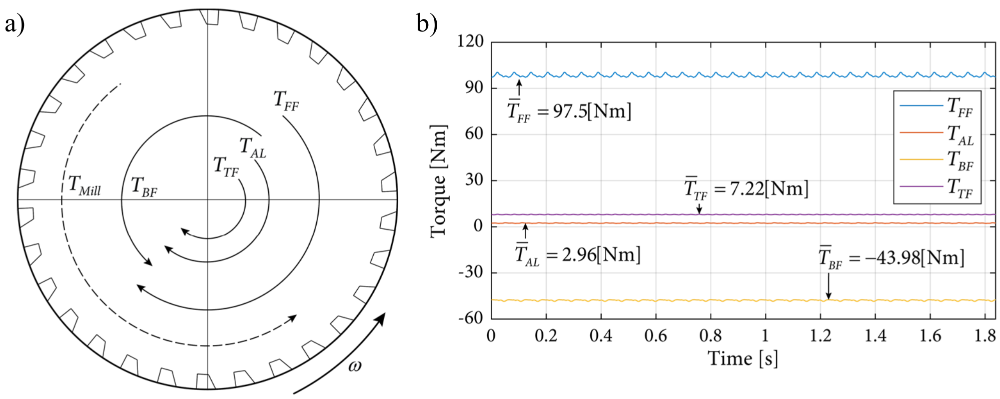


Figure 15: (a) Illustration of the components of the total torque per face of the lifter and adjacent liner. (b) Results of total torque per face as a function of time (average values indicated).

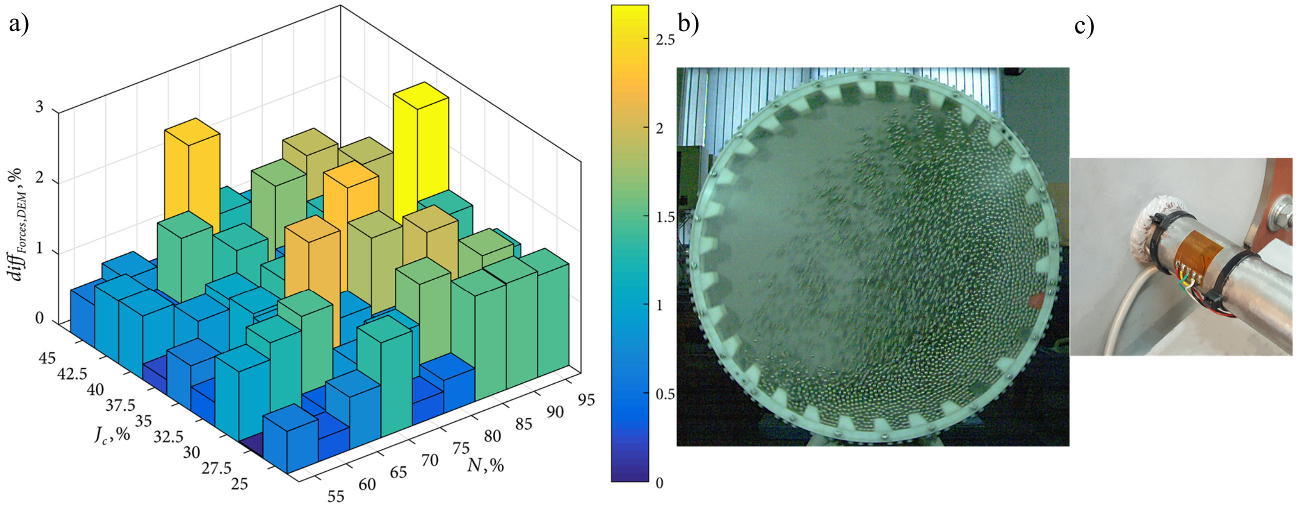


Figure 16: (a) Power differences for different operating conditions. (b) Front view of the mill while operating. (c) Strain gage for experimental measurement of driving torque.

4.3.3 Comparison between the power calculated from the forces and the power obtained from DEM

In order to validate the methodology, the power P_{Forces} is compared to the power $P_{DEM,Ring}$, which is obtained directly from the DEM model of the SetupD100.

Figure 16a shows the difference between the two powers calculated as:

$$diff_{Forces,DEM} = \frac{\left|P_{Forces} - P_{DEM,Ring}\right|}{P_{DEM,Ring}} \tag{33}$$

As observed, the differences between the power obtained directly from the DEM software and the power calculated from the forces are small (maximum 2.69%), thus validating the method to obtain the set of 8 forces as a function of the angular position using the virtual sensors.

5 Instrumentation used in SetupD100

The instrumentation used in SetupD100 focuses on analyzing the movement of the load inside the mill, analyzing the interaction between the lifters and the particles, and the experimental measurement of the power associated with the operation of the mill.

5.1 High-speed camera

One of the main features of the SetupD100 is that it is possible to observe the movement of the particles inside while operating, thanks to its transparent front cover. A high-speed camera (200 fps) is used to record the movement of the load inside the mill. Figure 16b shows an example of the images captured by the camera.

5.2 Driving torque

The measurement of the driving torque is made by using strain gages in full-bridge configuration installed on the mill shaft, between the mill and the bearing, Figure 16c.

The strain measurement in this position allows determining the torque due to the movement of the particles only, eliminating the need to estimate losses in other elements of the powertrain such as couplings and gear transmission. By multiplying the motor torque by the angular speed of the mill, the experimental power associated with the movement of the mill P_{Exp} is obtained, which is equivalent to the power $P_{DEM,Mill}$.

5.3 Instrumented lifter and tachometer

In order to validate the average forces obtained numerically, instrumented lifters resembling the virtual sensors are built. The instrumented lifters are capable of sensing the interaction between the particles and two of the lifter faces: front face (FF) and back face (BF).

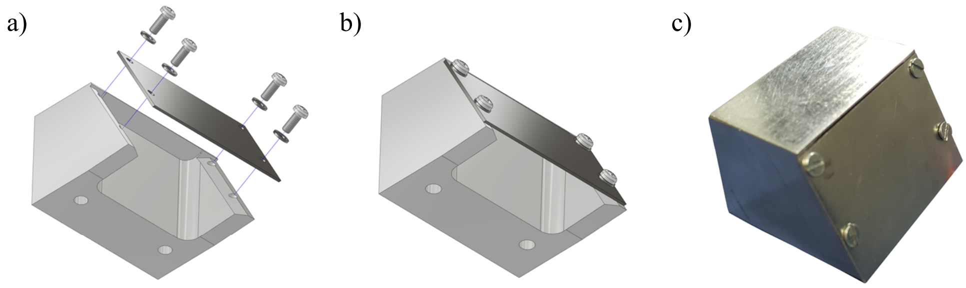


Figure 17: Instrumented lifter for measurement of front face interactions (LI FF). (a) and (b) Assembly. (c) Actual lifter.

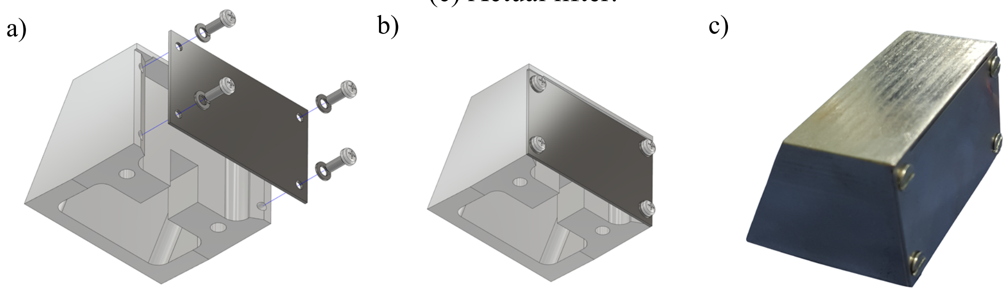


Figure 18: Instrumented lifter for measurement of back face interactions (LI FF). (a) and (b) Assembly. (c) Actual lifter.

The instrumented lifters are composed of two parts: a base block and a thin sensing plate. The sensing plate is made of steel and is fixed to the base block using 4 bolts at its side ends, as shown in Figure 17a. When the instrumented lifter contacts the particles, the sensing plates deflect. The deflection is measured by a pair of bi-axial strain gauges in half bridge configuration installed in the back of the plate. Figure 17 shows the assembly of the instrumented lifter for the measurement of front face interactions (LI-FF); whereas Figure 18 shows the same for measurement of back face interactions (LI-BF).

Figure 19 shows the results of the measurements with the two instrumented lifters for a rotation speed of 75% of the critical speed and a fill level of 30%. As in the case of the forces measured with the VS, the strain measured by the instrumented lifters are separated in pulses. The angular position of the instrumented lifters is determined from the reference signal provided by a photo-tachometer. Subsequently, the experimental pulses (EP) are averaged to obtain an estimator of the stationary component of the pulse forces, Figure 20.

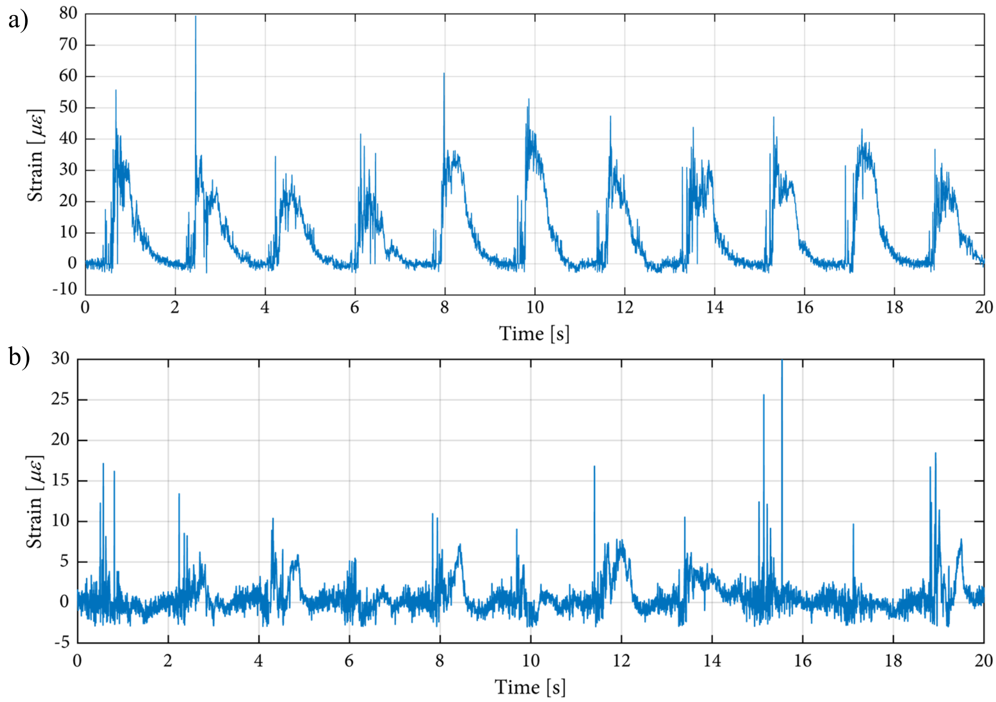


Figure 19: Strain waveform measured by the instrumented lifters. (a) Front face (LI-FF). (b) Back face (LI-BF).

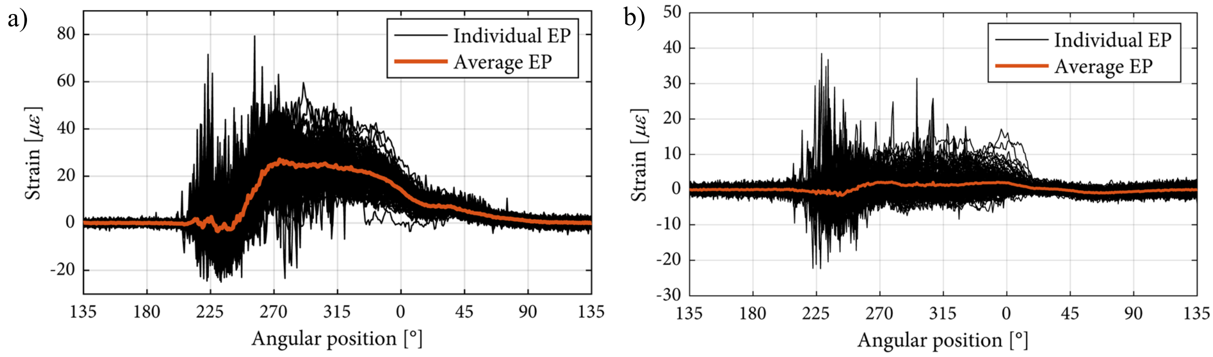


Figure 20: Experimental average pulses obtained with the instrumented lifters. (a) Front face (LI-FF). (b) Back face (LI-BF).

6 Results and discussion

The presented methodology allows a detailed analysis of the behavior of the load inside the mill, which in turn defines the overall behavior of the machine. Of particular interest is the behavior of the power consumption depending on the operating conditions. This behavior has been described in the literature based on the observation of the phenomenon, but a physical explanation has not yet been presented.

This section presents experimental results that validate the proposed methodology. Afterwards, the methodology is used to provide the physical explanation mentioned.

The average forces associated to each face as a function of the angular position, and the corresponding torques they exert, allows analyzing the behavior of the power associated with the movement of the ring from the viewpoint of the interaction of the particles with its internal surfaces.

6.1 Average forces

Figure 21 shows the average forces obtained after processing the forces from the virtual sensors included in the DEM model of the SetupD100, for the mill operating at 75% of its critical speed and a fill level of 30%. As can be observed, the magnitude of the parallel forces is much smaller than the magnitude of the normal forces. Five of the forces acting on the lifter faces are tangential to the ring. From these, the normal forces acting on the front face (FF) and back face (BF) are the most significant in the total torque associated with the rotation of the ring. The 3 remaining forces are parallel forces of magnitude considerably lower than that of normal forces. Based on this, the torque components T_{FF} and T_{BF} are the most relevant in the total torque and thus in the power consumption.

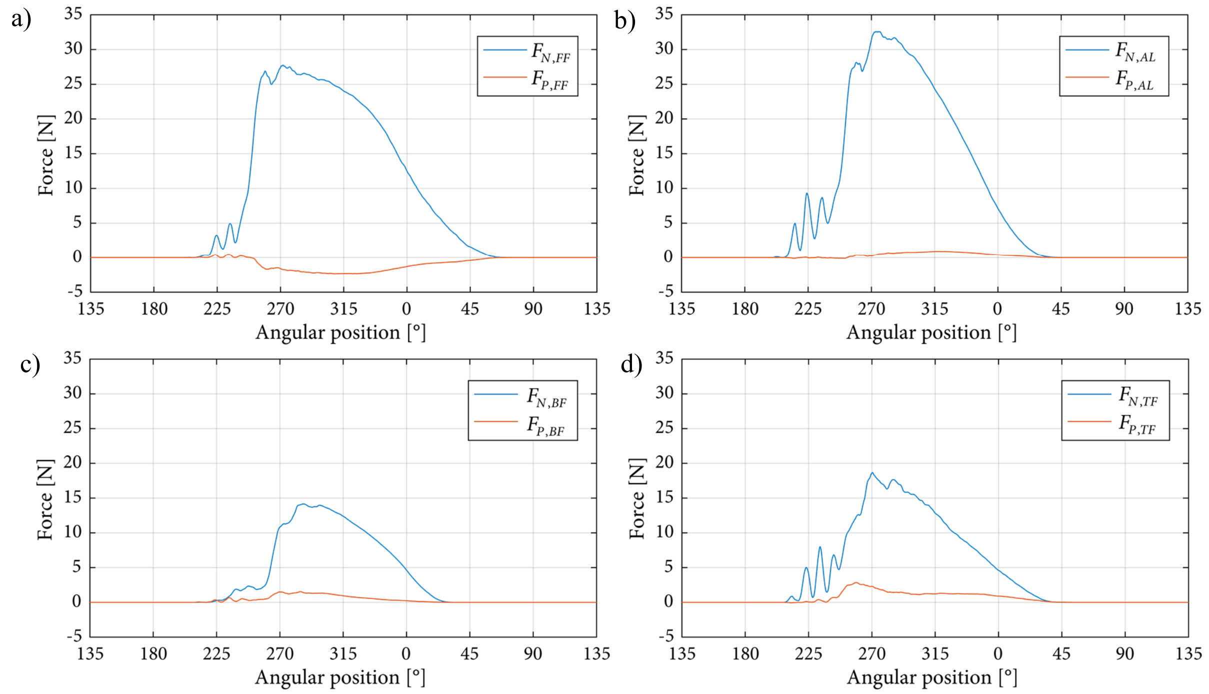


Figure 21: Normal and parallel average forces from the virtual sensors. (a) Front face. (b) Adjacent lifter. (c) Back face. (d) Top face.

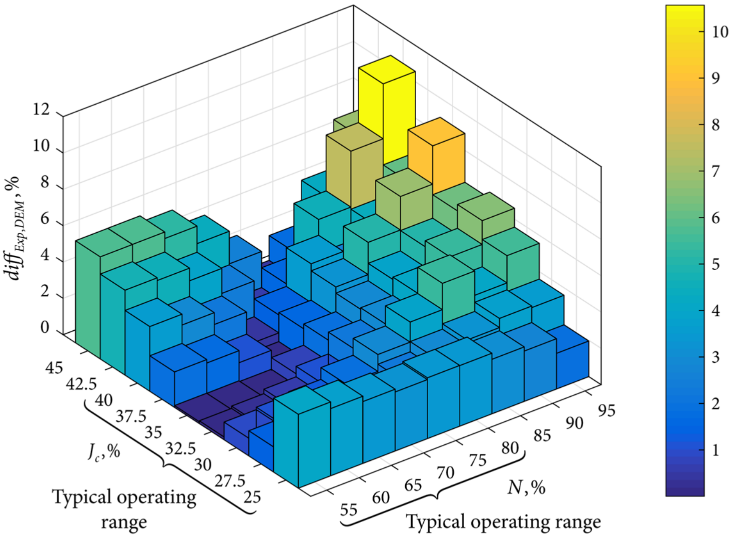


Figure 22: Difference between P_{Exp} and $P_{DEM,Mill}$ for each of the conditions analyzed.

6.2 Experimental validation

The numerical model of the SetupD100 is validated by comparing the numerical results obtained by processing the data extracted from the DEM simulations against their experimental equivalents obtained using the instrumentation of the test bench. The comparison is made first in terms of the power associated with the movement of the mill obtained by DEM ($P_{DEM,Mill}$) and the power obtained experimentally by the strain gauge installed on the drive shaft of the mill (P_{Exp}). As a second approach, the average forces obtained from the virtual sensors and the average experimental pulses obtained with the instrumented lifters are considered. Finally, the disposition of the particles inside the mill obtained by DEM and by the high-speed camera are contrasted.

Figure 22 shows the difference between P_{Exp} and $P_{DEM,Mill}$, calculated as:

$$diff_{Exp,DEM} = \frac{\left|P_{DEm,Mill} - P_{Exp}\right|}{P_{Exp}} \tag{34}$$

The maximum difference is 10.5% for the mill operating at 95% of its critical speed and 42.5% fill level. In mining industries, grinding mills typically operate with speeds ranging between 55% to 80% of the critical speed and fill levels between 25% to 40% of the internal volume. Within this range of operating conditions, the difference between the numerical and experimental power is less than 5%.

Figure 23 shows the comparison between the forces $F_{N,FF}$ and $F_{N,BF}$, and the experimental average pulses obtained with the LI-FF and the LI-BF, for different operating conditions. The differences observed are due to the fact that the instrumented lifters do not measure the contact forces directly, but the strain caused by these forces. Even though the strain is a consequence of these forces, there is no direct relation to the resultant force magnitude, because the force distribution has an influence too. It can also be seen that the magnitude of the measurements made with the LI-BF is much lower than those obtained with the LI-FF. Despite this, there is a clear agreement between the angular intervals in which the forces and the average experimental pulses are non-zero. Also, the fact that the average experimental pulse obtained with LI-BF is mostly positive confirms that the forces exerted by the particles on the BF compress it, generating torques in the sense of rotation of the mill.

Figure 24 compares the disposition of particles inside the mill obtained numerically and experimentally for 3 different operating conditions, showing good correlation.

Based on the similarities observed between the numerical and experimental results, it is reasonable to assume that the numerical model of the SetupD100 and the proposed methodology are valid.

6.3 Effect of rotating speed on the power, torque, and forces

Figure 25a shows the behavior of the power $P_{DEM,Ring}$ as a function of the speed, for the mill operating with different fill levels. Figure 25b shows the same for the torque $T_{DEM,Ring} = P_{DEM,Ring}/\omega$. This torque is equivalent to the sum of the torques $\overline{T}_{FF}$, $\overline{T}_{AL}$, $\overline{T}_{BF}$ and $\overline{T}_{TF}$. It can be seen that the power increases steadily, for all fill levels,

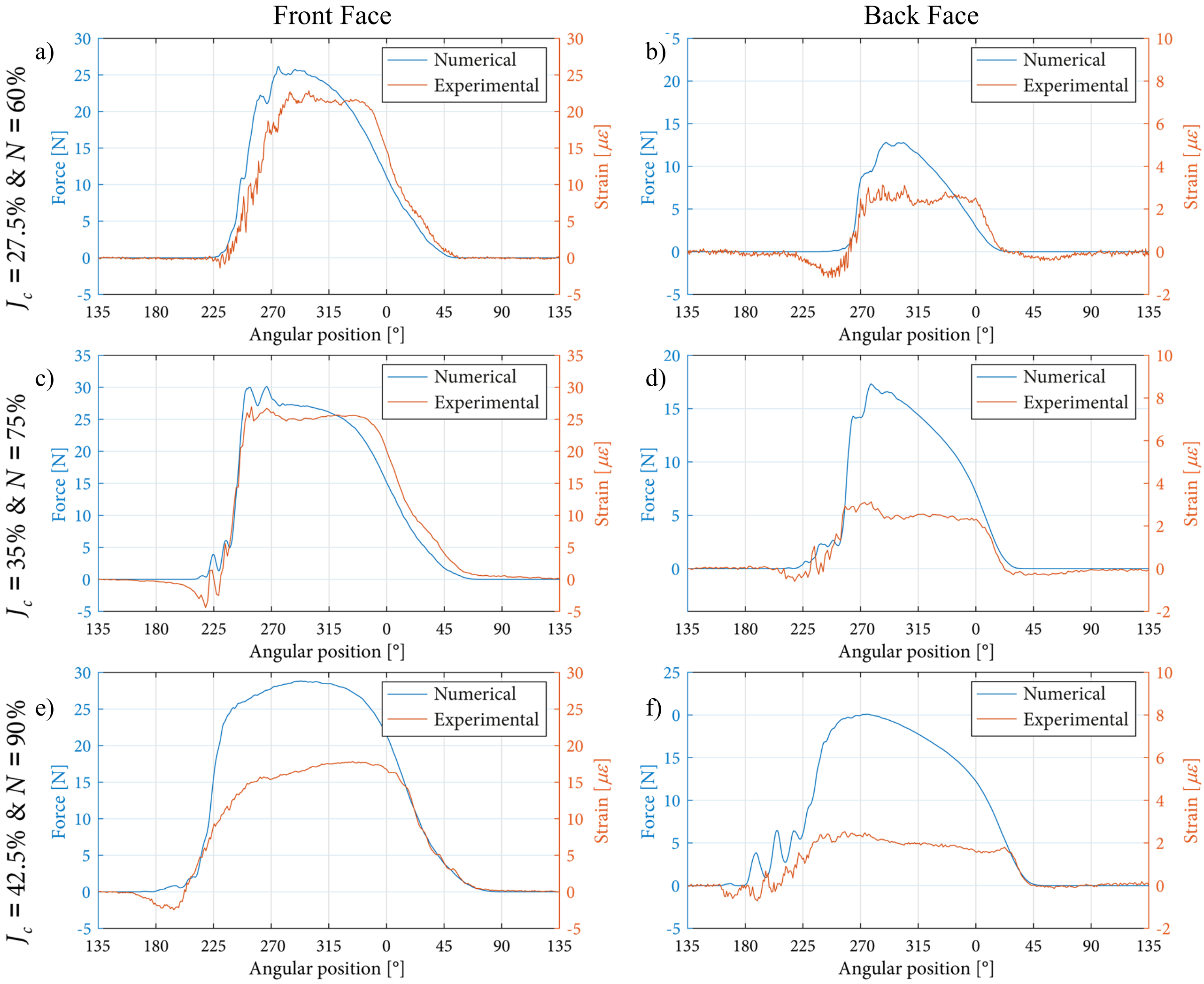


Figure 23: Comparison of numerical and experimental average pulses for the front face and back face, and different operating conditions.

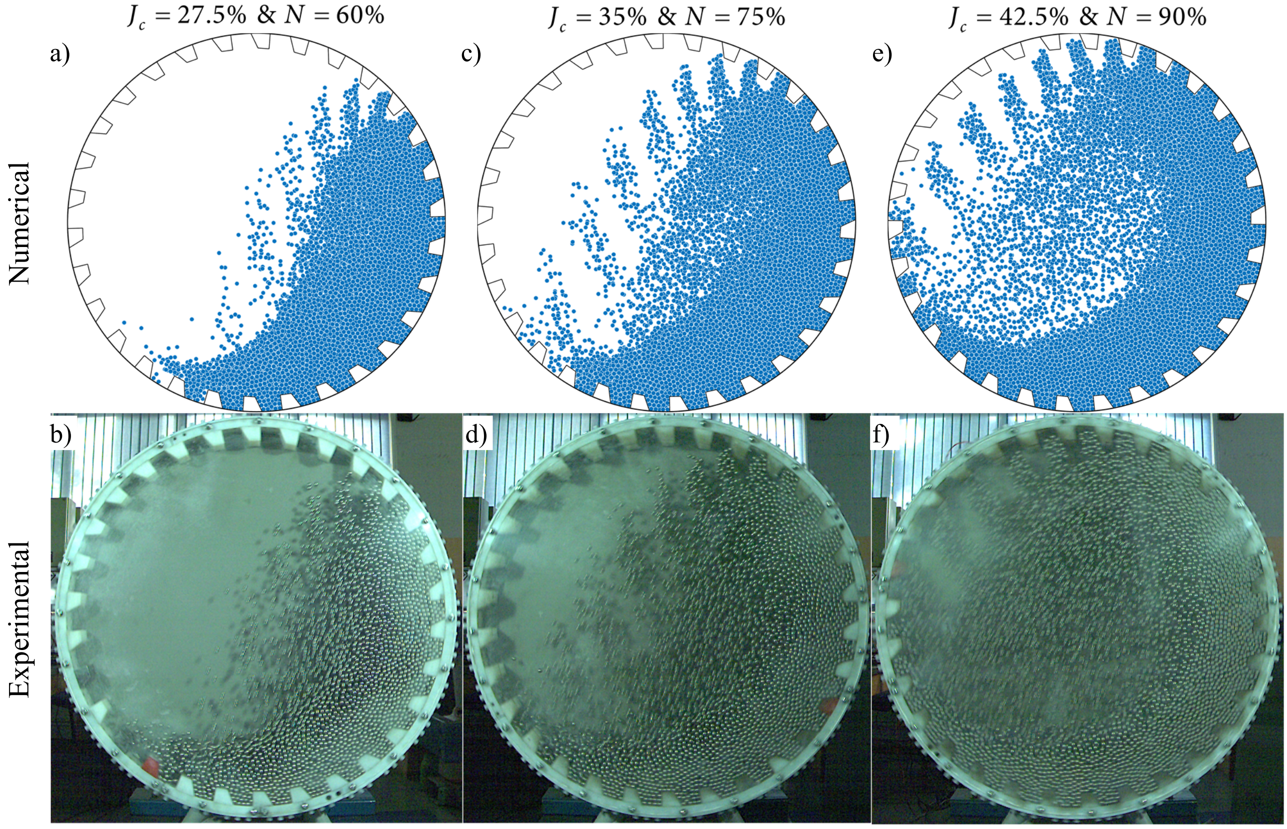


Figure 24: Disposition of particles inside the mill (a),(c),(e) Numerical results. (b),(d),(f) Expermiental results.

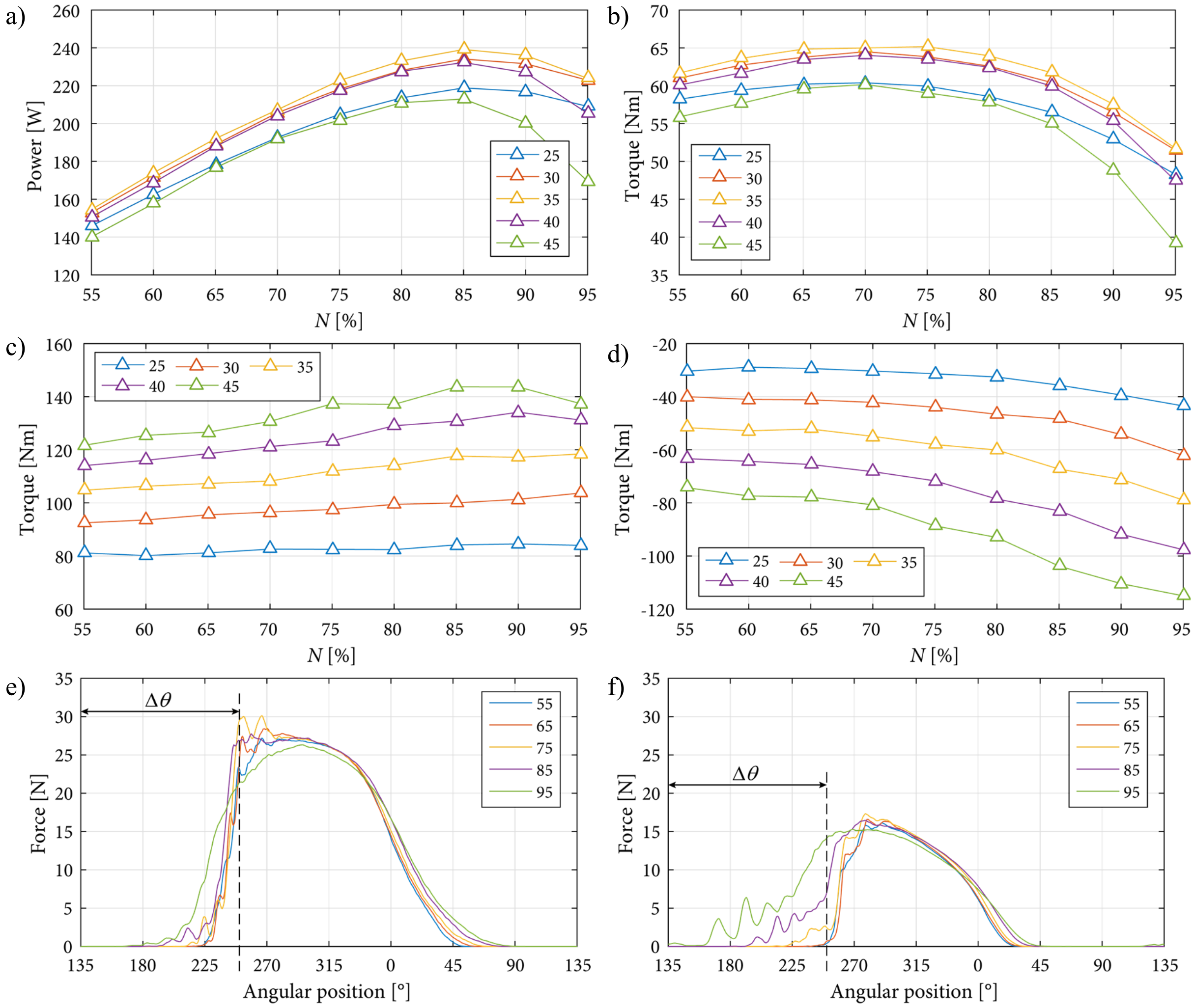


Figure 25: (a) Power $P_{DEM,Ring}$ as a function of speed. (b) Torque $T_{DEM,Ring}$ as a function of speed. (c) Torque T_{FF} as a function of speed. (d) Torque T_{BF} as a function of speed. (e) Force $F_{N,FF}$ as a function of the angular position. (f) Force $F_{N,BF}$ as a function of the angular position.

from 55% to 80% of the critical speed. For higher speeds, the power decreases at a rate that becomes more pronounced as the fill level increases. The torque $T_{DEM,Ring}$ increases up to 70% of the critical speed. From this point on, it begins to decrease. This behavior has been reported in the literature [13][14][21][57], and in order to explain it, the torques and the average forces associated with the different faces of the lifters are further analyzed.

Figure 25c and Figure 25d show the behavior of the torques T_{FF} and T_{BF} as a function of the speed. It can be seen that in the range between 55% and 70% of the critical speed, the increase rate in the magnitude of T_{FF} is slightly higher than the increase rate in the magnitude of T_{BF}, which explains the increase in the magnitude of $T_{DEM,Ring}$ within this speed range. For speeds above 70% of the critical speed, it can be noted that the magnitude of T_{BF} increases at a higher rate than the magnitude of T_{FF}, and that this is more pronounced for higher fill levels. This indicates that the drop in the torque $T_{DEM,Ring}$ for speeds higher than 70% of the critical speed observed in Figure 25b originates because the magnitude of the force $F_{N,BF}$ increases more significantly than the magnitude of the force $F_{N,FF}$ as the rotating speed increases.

Figure 25e and Figure 25f show, respectively, the magnitude of the forces $F_{N,FF}$ and $F_{N,BF}$ as a function of the angular position for the mill operating at 35% of fill level and speed from 55% to 95% of the critical speed with 10% jumps. It can be noted that the magnitude of $F_{N,FF}$ and the angular interval where this force is non-zero increase slightly as a function of the speed of rotation, which produces the slight and sustained increase in the magnitude of T_{FF} observed in Figure 25c. On the other hand, the force $F_{N,BF}$ remains approximately constant between 55% and 65% of the critical speed but begins to increase in the angular range $\Delta\theta$ between 135° and 250° for fill levels above 75%. This behavior originates the increase suffered by the magnitude of

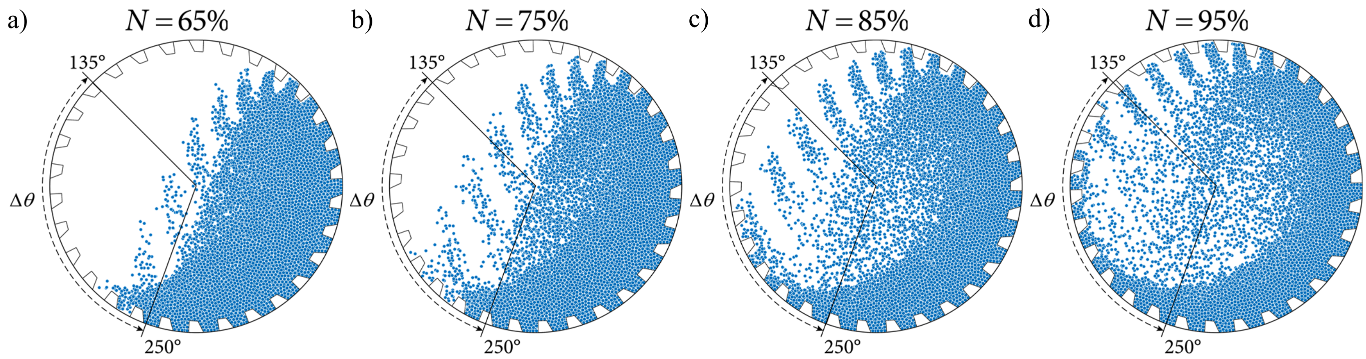


Figure 26: Disposition of particles inside the mill operating with 35% fill level and rotating speed of (a) 65%, (b) 75%, (c) 85% and (d) 95% of the critical speed.

T_{BF} between 70% and 95% of the critical speed, and that causes the decrease of $T_{DEM,Ring}$ in this speed range. The physical phenomenon that gives rise to the forces exerted by the particles in the angular interval between 135° and 250°, for speeds higher than 70% of the critical speed—and that is accentuated with speed--, corresponds to the impact of the particles describing cataract movements and that fall over the toe position, as shown in Figure 26. The particles impacting the internal surface of the mill over the position of the toe exert forces on the back face of the lifters and not on the front face, due to their orientation relative to the path of the particles. These forces exert torques in the sense of rotation of the mill, thus diminishing the total torque requirement. This phenomenon becomes more significant as the rotation speed increases, because both the number and speed of the particles falling on the back face of the lifters above the toe position, increase.

6.4 Conclusions

A methodology is proposed to obtain the magnitude of the average forces exerted by the particles inside the mill on the different faces of the lifters and liners of the ring as a function of their angular position, and the torques associated with each of these faces for the complete set of lifters in the ring. The results obtained from the numerical modeling are in agreement with the corresponding experimental measurements obtained in the test bench, thus validating the methodology.

The forces and torques provide a useful tool to understand the behavior of the power consumed by the mill depending on its geometry and operating conditions. These variables provide the link between the behavior of the load inside the drum and the resulting behavior of the power consumption of the mill. This is especially relevant considering that the energy costs associated with the operation of tumbling mills and the evolution of their internal geometry due to wear, represent a significant fraction of the costs associated with the refinement of minerals.

The results of the analyses carried out in this work show that the decrease in power consumption observed when the speed of the mill increases is due to the impacts of the particles falling on the zone of the ring above the toe of the load. These impacts exert forces on the back face of the lifters, thus increasing the torque component associated with this face, which acts in the sense of rotation. At the same time, as the speed increases, the torque associated with the front face of the lifters also increase, but its magnitude grows at a lower rate. This difference in the growth rates of both torques provides the physical explanation of the power consumption drop at higher speeds.

The methodology seems to be a promising tool in view of gaining further insight about the behavior of tumbling mills and could be useful, for example, to develop strategies to reduce O&M costs and increase efficiency.

References

[1] 'The world's largest sag mill', Outotec, South East Asia Pacific Output, Dec. 2012, page 6.

[2] Guerrero, V. & Pontt, J. 2011. 'Oscillatory Torque Caused by Dead Time in the Current Control of High Power Gearless Mills', *IECON 2011* - 37th Annual Conference on IEEE Industrial Electronics Society, 2011, pp. 1966 – 1970.

[3] Martins, S., Li, W., Radziszewski, P., Faucher, A. & Makni, S. 2013. 'Experimental and simulated instrumented ball in a tumbling mill - A comparison'. *Minerals Engineering*, vol. 43-44, pp.79-84.

[4] Hosseini, P., Martinsm S, Martin, T. Radziszewski, P. & Boyer, F. 2011. 'Acoustic emissions simulation of tumbling mills using charge dynamics'. *Minerals Engineering*, vol. 24, pp. 1440-1447.

[5] Cleary, P., Morrison, R. & Morrell, S. 2003. 'Comparison of DEM and experiment for a scale model SAG mil'. *International Journal of Mineral Processing*, vol. 68, pp. 129-165.

[6] Cleary, P. & Morrison, R. 2016. 'Comminution mechanisms, particle shape evolution and collision energy partitioning in tumbling mills'. *Minerals Engineering*, vol. 86, pp. 75-95.

[7] Cleary, P. 2015. 'Prediction of coupled particle and fluid flows using DEM and SPH'. *Minerals Engineering*, vo. 73, pp. 85-99.

[8] Owen, P. & Clearly, P. 2015. 'The relationship between charge shape characteristics and fill level and lifter height for a SAG mill'. *Minerals Engineering*, vol. 83, pp. 19-32.

[9] Cleary, P. & Morrison, R. 2011. 'Understanding fine ore breakage in a laboratory scale ball mill using DEM'. *Minerals Engineering*, vol. 24, pp. 352-366.

[10] Kalala, J., Breetzke, M. & Moys, H. 2008. 'Study of the influence of liner wear on the load behaviour of an industrial dry tumbling mill using the Discrete Element Method (DEM)'. *International Journal of Mineral Processing*, vol. 86, pp. 33-39.

[11] Powell, M. & McBride, A. 2004. 'A three-dimensional analysis of media motion and grinding regions in mills'. *Minerals Engineering*, vol. 17, pp. 1099-1109.

[12] Mishra, B. 2003. 'A review of computer simulation of tumbling mills by the discrete element method: Part I-contact mechanics'. *International Journal of Mineral Processing*, vol. 71, pp. 73-93.

[13] Cleary, P. 2001. 'Charge behaviour and power consumption in ball mills: Sensitivity to mill operating conditions, liner geometry and charge composition'. *International Journal of Mineral Processing*, vol. 63, pp. 79-114.

[14] Cleary, P. 1998. 'Predicting charge motion, power draw, segregation and wear in ball mills using discrete element methods'. *Minerals Engineering*, vol. 11, pp. 1061-1080.

[15] Mishra, B. & Rajamani, R. 1992. 'The discrete element method for the simulation of ball mills'. *Applied Mathematical Modelling*, vol. 16, pp. 598-604.

[16] Kalala, J & Moys, M. 2004. 'Discrete element method modelling of liner wear in dry ball milling'. *Journal of The South African Institute of Mining and Metallurgy*, vol. 104, pp. 597-602.

[17] Morrison, R. & Cleary, P. 2004. 'Using DEM to model ore breakage within a pilot scale SAG mil'. *Minerals Engineering*. Vol. 17, pp. 1117-1124.

[18] Jiménez-Herrera, N., Barrios, G. & Tavares, L. 2017. 'Comparison of breakage models in DEM in simulating impact on particle beds'. *Advanced Powder Technology*, vol. 29, pp. 692-706.

[19] Cleary, P., Sinnott, M. & Pereira, G. 2015. 'Computational prediction of performance for a full scale Isamill: Part 1 - Media motion and energy utilisation in a dry mill'. *Minerals Engineering*, vol. 79, pp. 220-238.

[20] Tuzcu, E. & Rajamani, R. 2011. 'Modeling breakage rates in mills with impact energy spectra and ultra fast load cell data'. *Minerals Engineering*, vol. 24, pp. 252-260.

[21] Cleary, P. 2001. 'Recent advances in DEM modelling of tumbling mills'. *Minerals Engineering*, vol. 14, pp. 1295-1319.

[22] Cleary, P. 2001b. ‘Modelling comminution devices using DEM’. *International Journal for Numerical and Analytical Methods in Geomechanics*, vol. 25, pp. 83–105.

[23] Cleary, P.W., Sinnott, M. & Morrison, R., 2008. ‘DEM prediction of particle flows in grinding processes’. *International Journal for Numerical Methods in Fluids*, vol. 58, pp. 319–353.

[24] Cleary, P. 2015. ‘A multiscale method for including fine particle effects in DEM models of grinding mills’. *Minerals Engineering,* vol. 84, pp. 88–99.

[25] Djordjevic, N. 2003. ‘Discrete element modelling of the influence of lifters on power draw of tumbling mills’. *Minerals Engineering,* vol. 16, pp. 331–336.

[26] Djordjevic, N., 2005. ‘Influence of charge size distribution on net-power draw of tumbling mill based on DEM modelling’. *Minerals Engineering*, vol. 18, pp. 375–378.

[27] Geng, F., Gang, L., Wang, Y., Li, Y. & Yuan, Z., 2015. ‘Numerical investigation on particle mixing in a ball mill’. *Powder Technology*, vol. 292, pp. 64–73.

[28] Govender, I., McBride, A. & Powell, M. 2004. ‘Improved experimental tracking techniques for validating discrete element method simulations of tumbling mills’. *Experimental Mechanics*, vol. 44, pp. 593–607.

[29] Govender, I., Cleary, P. & Mainza, A., 2013. ‘Comparisons of PEPT derived charge features in wet milling environments with a friction-adjusted DEM model’. *Chemical Engineering Science,* vol. 97, pp. 162–175.

[30] Hlungwani, O., Rikhotso, J., Dong, H., Moys, M.H., 2003. ‘Further validation of DEM modeling of milling: effects of liner profile and mill speed’. *Minerals Engineering,* vol. 16, pp. 993– 998.

[31] Kalala, J., Bwalya, M & Moys, M. 2005. ‘Discrete element method (DEM) modelling of evolving mill liner profiles due to wear Part II. Industrial case study’. *Minerals Engineering,* vol. 18, pp. 1386–1391.

[32] Makokha, A & Moys, M. 2006. ‘Towards optimising ball-milling capacity: effect of lifter design’. *Minerals Engineering* ,vol. 19, pp. 1439–1445.

[33] Mayank, K., Malahe, M., Govender, I. & Mangadoddy, N. 2015. ‘Coupled dem-cfd model to predict the tumbling mill dynamics’. *Procedia IUTAM*, vol. 15, pp. 139–149.

[34] Mishra, B. & Rajamani, R. 1994a. ‘Simulation of charge motion in ball mills. Part 1: experimental verifications’. *International Journal of Mineral Processing,* vol. 40, pp. 171–186.

[35] Mishra, B. & Rajamani, R. 1994b. ‘Simulation of charge motion in ball mills. Part 2: numerical simulations’. *International Journal of Mineral Processing*, vol. 40, 187–197.

[36] Morrison, R. & Cleary, P. 2008. ‘Towards a virtual comminution machine’. *Minerals Engineering*, vol. 21, pp. 770–781.

[37] Mulenga, F. & Moys, M. 2014. ‘Effects of slurry filling and mill speed on the net power draw of a tumbling ball mill’. *Minerals Engineering*, vol. 56, pp. 45–56.

[38] Powell, M. & McBride, A. 2006. ‘What is required from DEM simulation to model breakage in mills?’ *Minerals Engineering*, vol. 19, pp. 1013–1021.

[39] Powell, M., Weerasekara, N., Cole, S., LaRoche, R. & Favier, J. 2011. ‘DEM modelling of liner evolution and its influence on grinding rate in ball mills’. *Minerals Engineering,* vol. 24, pp. 341–351.

[40] Rezaeizadeh, M., Fooladi, M., Powell, M., Mansouri, S. & Weerasekara, N. 2010. ‘A new predictive model of lifter bar wear in mills’. *Minerals Engineering*, vol. 23, pp. 1174–1181.

[41] Toor, P., Franke, J., Powell, M., Bird, M. & Waters, T. 2013. ‘Designing liners for performance not life’. *Minerals Engineering*, vol. 43, pp. 22–28.

[42] Wang, M., Yang, R. & Yu, A. 2012. ‘DEM investigation of energy distribution and particle breakage in tumbling ball mills’. *Powder Technology*, vol. 223, pp. 83–91.

[43] Weerasekara, N., Liu, L. & Powell, M. 2016. ‘Estimating energy in grinding using dem modelling’. *Minerals Engineering*, vol. 85, pp. 23-33.

[44] Xu, L., Luo, K. & Zhao, Y. 2018. ‘Numerical prediction of wear in SAG mills based on DEM simulations’, Powder Technology, vol. 329, pp. 353-363.

[45] Qui, Xx, Potapov, A., Song, M., & Nordell, L. 2001. ‘Prediction of Wear of Mill Lifters Using Discrete Element Method’, In SAG 2001, University of British Columbia, Vancouver, B.B, Canada, pp. 2–7.

[46] Boemer, D & Ponthot, J. 2017. ‘DEM modeling of ball mills with experimental validation: influence of contact parameters on charge motion and power draw’, *Computational Particle Mechanics*, vol. 4, pp. 53-67.

[47] Wills, B. & Napier-Munn, T. 2006. ‘Wills' Mineral Processing Technology’, 7th edn, Elsevier Science & Technology Books.

[48] Tano, K., Pålsson, B. & Sellgren, A. 2005. ‘On-line lifter deflection measurements showing flow resistance effects in grinding mills’, *Minerals Engineering*, vol. 18, pp. 1077–1085.

[49] Owen, P. & Cleary, P. W. 2015. ‘The relationship between charge shape characteristics and fill level and lifter height for a SAG mill’, *Minerals Engineering*. vol. 83, pp. 19–32.

[50] Cleary, P. W. & Owen, P. 2018. ‘Development of models relating charge shape and power draw to SAG mill operating parameters and their use in devising mill operating strategies to account for liner wear’, *Minerals Engineering*. vol. 117, pp. 42–62.

[51] Kalala, J., Bwalya, M. and Moys, M. 2005. ‘Discrete element method (DEM) modelling of evolving mill liner profiles due to wear. Part I: DEM validation’, *Minerals Engineering*, vol. 18, pp. 1386–1391.

[52] Kalala, J., Bwalya, M. & Moys, M. 2005. ‘Discrete element method (DEM) modelling of evolving mill liner profiles due to wear. Part I: DEM validation’, *Minerals Engineering*, vol. 18, pp. 1386–1391.

[53] Jonsén, P., Pålsson, B. & Häggblad, H. 2012. ‘A novel method for full-body modelling of grinding charges in tumbling mills’, *Minerals Engineering*, vol. 33, pp. 2–12.

[54] Moys, M., van Nierop, M., van Tondero, J. & Glover, G. 2000. ‘Validation of the discrete element method (DEM) by comparing predicted load behaviour of a grinding mill with measured data’, *Developments in Mineral Processing,* vol. 13, pp. C3-39-C3-44.

[55] Tano, K. 2005. ‘Continuous Monitoring of Mineral Processes with Special Focus on Tumbling Mills – A Multivariate Approach’, Doctoral Thesis, Luleå University of Technology.

[56] Tano, K., Pålsson, B. & Sellgren, A. 2005. ‘On-line lifter deflection measurements showing flow resistance effects in grinding mills’, *Minerals Engineering*, vol. 18, pp. 1077–1085.

[57] Bian, X., Wang, G., Wang, H., Wang, S. & Lv, W. 2017. ‘Effect of lifters and mill speed on particle behaviour, torque, and power consumption of a tumbling ball mill: Experimental study and DEM simulation’. *Minerals Engineering*, vol. 105, pp. 22-35.

Signal processing

Comparison and Improvement of Techniques for Transmission-Path Restoring

Authors: Omri Matania, Renata Klein, Jacob Bortman

Keywords: vibration signal, transmission-path, pre-whitening, adaptive clutter separation (ACS), Cepstrum-liftering, AR model, restoration process.

Abstract:

Condition based monitoring by vibration sensors is a widely spread technique for monitoring the status and condition ("health") of rotating machines. In most cases, the monitoring is based on the ability to isolate specific elements of the vibration signals, generated by the different rotating components.

The generated signals are propagating through various transmissions paths of the machine, that distort the original signals, hence affect the assessment of the machine's condition. While these effects are usually ignored by most vibration analysis techniques, first steps towards mitigating this problem have been taken place during the last years. These techniques used pre-whitening methods, which usually served to separate the signal from its background, as well as to reduce the transmission path effects.

In this study, we aim to go a step further in transmission path restoring through deepening our understanding of their effects on the vibration signals. We start by reviewing three main pre-whitening methods: liftering low quefrencies at the Cepstrum (Cepstrum-liftering), adaptive clutter separation (ACS), and pre-whitening using auto-regressive (AR) models.

We first show that signal's pre-whitening by the AR model has large errors and therefore is less adequate for transmission path restoration processes. Through several simulations, we show that the AR model succeeds in extracting the background spectrum only where the noise is significantly larger than any other component in the PSD. For all other cases, there are large inaccuracies at the restoring background process. We propose a theoretical explanation for this phenomenon, and then strengthen our argument that AR in its current use is less adequate for this purpose.

We then propose a theoretical approach to adjust parameters for ACS and Cepstrum-liftering techniques and examine them and their sensitivity through quantitative methods. By tying the theoretical adequate windows parameters of Cepstrum-liftering and ACS, we show that it is possible to predict the sizes of these two adequate windows. Furthermore, we suggest an adaptive algorithm that is based on the theoretical calculation to reach more accurate values for the parameters of the window in real cases.

After adjustment of the parameters, we compare ACS and Cepstrum-liftering through a variety of simulations and receive slightly different quantities results of the transmission path estimation using ACS and Cepstrum liftering than the former paper [1].

A new technique to restore the transmission path background and its phase is proposed (Figure 1). The restoration of the phase is an important feature due to its deep implications on the ability to correctly restore the original signal in the time domain. The new technique is based on AR model and artificial noise colorization to restore both the background spectrum and its phase. Furthermore, it also improves the magnitude of the restored background, compared to ACS and Cepstrum-liftering techniques.

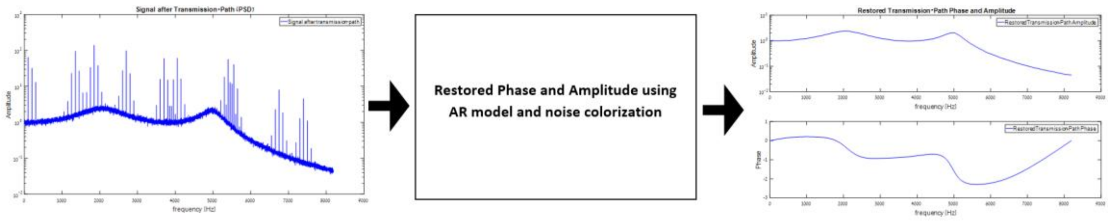


Figure 1 – restored Transmission-Path phase and amplitude

We also suggest exploiting the advantage of the new technique to restore the background in the frequency domain by converting the signal to the order domain and restore it there. The technique is highly beneficial in cases where the signal is smeared in the frequency domain but is sharp in the order domain, cases which the system rotating speed varies during the time.

References:

[1] R. Klein, *Comparison of methods for separating excitation sources in rotating machinery*, The International Conference SURVISHNO 7, 2013.

[2] R.B Randall, B. Peeters, J. Antoni, S. Manzato, *New cepstral methods of signal pre-processing for operational modal analysis*, ISMA2012-USD2012, Leuven, Belgium, pp. 775-764, 2012.

[3] R.B. Randall, W.A Smith, *Cepstrum-based operational modal analysis revisited: A discussion on pole-zero models and the regeneration of frequency response functions*, Article in Mechanical Systems and Signal Processing (MSSP), 2016.

[4] R.B. Randall, W.A Smith, *New cepstral techniques for operational modal analysis*, Conference Paper, 2017.

[5] N. Sawalhi and R.B. Randall, *Spectral Kurtosis Enhancement using Autoregressive Models*, ACAM2005 conference, Melbourne, 2005.

[6] R. Castiglione, J. Antoni, L. Garibaldi, *Separation and identification of structural modes in largely underdetermined scenarios using frequency banding*, Journal of Sound and Vibration 414 pp. 192-217, 2018.

Influence of Gaussian Signal Distribution Error on Random Vibration Fatigue Calculation.

Yuzhu WANG[1], Roger SERRA[1] et Pierre ARGOUL[2]

[1]INSA Centre Val de Loire, Laboratoire de Mécanique G. Lamé, Campus de Blois, Equipe DivS, 3 rue de la chocalaterie, 41000 Blois, France

[2]IFSTTAR, Laboratoire Mast – EMGCU, 14-20 Boulevard Newton, Cité Descartes, 77447 Marne-la-Vallée Cedex 2

Abstract

In the study of random vibration problems, Gaussian vibration and non-Gaussian vibrations are usually classified according to the excitation signal. The skewness and kurtosis are usually used to distinguish. Here we discuss a non-strict Gaussian signal, which is the error that exists in skewness and kurtosis and usually unavoidable in actual experiments or signals analysis. Through experiments and simulation calculations, the influence of this error on the traditional fatigue calculation method is discussed. The PSD approach will be discussed primarily, and time domain signals based on the rain-flow counting method will be recorded and verified. Total nine calculation model studied in this process. Finally, through a threshold, the range of skewness and kurtosis is indicated, that within this range, Gaussian signal-based calculations can be continued. By comparing the performance of different methods, a better method for signal adaptability can be obtained.

Keywords: Random vibration fatigue, Damage cumulative calculation and Gaussian random vibration.

1. Introduction

For many mechanical components, the working load is in the form of random vibrations. In the fatigue design project, the load cycle of the structure is usually obtained by the rain flow counting method according to the conventional time domain signal, with the material property, the damage of the structure could be obtained by using the Miner's Law, and then prediction the life of the structure. But the acquisition of time domain signals relies on a large number of experimental records, which obviously sounds expensive. Later, according to the stochastic theoretical method, the power spectral density was used to characterize the random vibration characteristics, and the method of inferring the rain flow counting result of response stress was proposed.[1][2] This theory generally assumes that the load is subject to a Gaussian distribution. The more successful method is the narrow-band approximation method proposed by Bendat in 1968.[3] Later, due to the efforts of more scholars, the broadband approximation method was also proposed. At the same time, various improvement schemes were proposed to improve the accuracy of the approximate results. [4][5][6]

However, since most of the actual loads are non-Gaussian distributions, it is obvious. Therefore, when the load is a non-Gaussian signal, the original Gaussian-based frequency domain damage analysis method may be used, which may cause poor deviation. Therefore, it is necessary to further discuss the influence of signal non-Gaussian on stress distribution. The research method based on the frequency domain signal discusses the overall distribution of signals in the frequency domain, and the non-Gaussian signal that satisfies this condition is not unique. This leads to the use of frequency domain method to study the distribution of non-Gaussian signals. Large deviations, when calculating damage using this distribution result, pose a significant risk to product design and life estimation.

In this paper, a reference based on kurtosis judgment will be proposed to select the method of fatigue damage calculation. The rain-flow count analysis is performed on the response stress of non-Gaussian load, and the difference between the Gaussian signal and the non-Gaussian signal at the same level is obtained, and an allowable value is obtained, that is, in this range, even if it is not a Gaussian signal, the PSD method also could be used. The fatigue damage results obtained by the method are still within error tolerance. Beyond this range, the non-Gaussian signals must be considered with the special method.

2. Non-Gaussian signal and Kurtosis control

Generally, a signal whose probability density distribution obeys a Gaussian distribution is called a Gaussian signal and is mainly judged by the skewness and kurtosis of the signal. This indicator indicates the distribution of data within the data range. The skewness refers to the zero offsets of the centre, which is represented by S. The kurtosis can be understood as the specific gravity in the central region, denoted by K. The greater the kurtosis, the greater the accumulation of data in the centre. Usually, the Gaussian distribution has a kurtosis of 3 and skewness of 0.

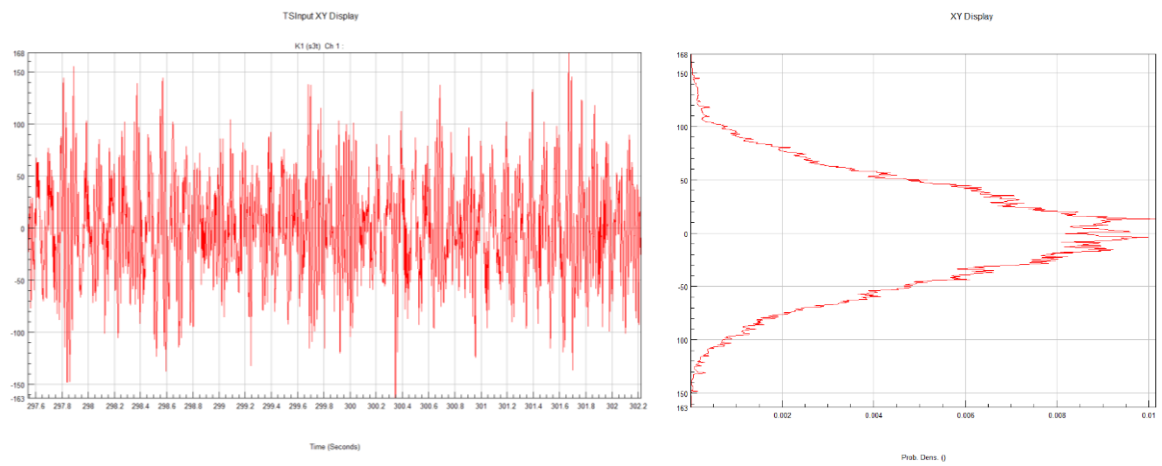


Figure 1 Gaussian signal and probability density distribution

Thus σ is the standard deviation, μ_n is the nth central moment

$$S = \frac{\mu_3}{\sigma^3} \tag{1}$$

$$K = \frac{\mu_4}{\sigma^4} \tag{2}$$

When the kurtosis or skewness has a condition that does not satisfy the Gaussian distribution, the signal is called a non-Gaussian distribution. At the same time, according to the index, when K>3, it is called leptokurtic, and when k<3, it is called platykurtic. According to the probability density distribution of the non-Gaussian signal, it can be found that the kurtosis reflects the distribution of acceleration in the middle region.[7]

In order to obtain an acceleration signal of a non-Gaussian distribution, it is usually obtained by Gaussian signal transformation. There are many methods used, Hermit polynomial, Gaussian mixture model, Phase selection method, Power-law model, Exponential method, Non-parametric method. In this paper, Steinwolf's phase selection method is used to modulate non-Gaussian signals with specific skewness and kurtosis.[8]

3. Phase selection method

By fast Fourier transform, an acceleration time domain signal can be described as a superposition of harmonics in the frequency range.

$$\mathrm{x(t)} = \sum_{n=1}^{N} A_n \cos(2\pi n\Delta f t + \varphi_n) \tag{3}$$

The amplitudes of the harmonics are obtained

$$A_n = \sqrt{2\Delta f S(n\Delta f)} \tag{4}$$

Then the nth centre moment could be obtained

$$M_z = \lim_{T\to\infty} \frac{1}{T}\int_0^T \{x(t)\}^z dt = \frac{1}{T}\int_0^T \{x(t)\}^z dt,\ \ z > 2 \tag{5}$$

After the kurtosis formula could be written as

$$K = \frac{M_4}{M_2^2} = 3 - \frac{3}{2}\left(\frac{\sum_{n=1}^{N} A_n^4}{\left(\sum_{n=1}^{N} A_n^2\right)^2}\right) + \left(\frac{1}{2}\sum_{n=1}^{N} A_n^2\right)^{-2} \{\frac{3}{2}\sum_{\substack{i+2j=k\\ i\neq j}} A_i {A_j}^2 A_k \cos\left(\varphi_i + 2\varphi_j - \varphi_k\right) + \frac{3}{2}\sum_{\substack{i+j=2k\\ i<j}} A_i {A_k}^2 A_j \cos\left(\varphi_i + \varphi_j - 2\varphi_k\right) + 3\sum_{\substack{i+j=k+m\\ i<j,k<m,i<k}} A_i A_j A_k A_m \cos\left(\varphi_i + \varphi_j - \varphi_k - \varphi_m\right) + 3\sum_{\substack{i+j+k=m\\ i<j<k<m}} A_i A_j A_k A_m \cos\left(\varphi_i + \varphi_j + \varphi_k - \varphi_m\right) + \frac{1}{2}\sum_{i=3j} A_i {A_j}^3 \cos\left(\varphi_i - 3\varphi_j\right)\} \tag{6}$$

In the case of ensuring that the mean and RMS of the non-Gaussian signal are not changed, only the distribution is changed. So to make the K fitted for the experiment by modulating the specific φ. The resulting non-Gaussian distribution acceleration time series with specific K could be found. It should be noted that as the bandwidth increases, the amount of calculation becomes very large and also affecting the amount is the sampling frequency and data length. In general, the required non-Gaussian data is obtained in a combination of several methods.

4. Rain-flow count and damage calculation

The rain-flow counting method was proposed in the 1950s by two British engineers, M. Matsuishi and T. Endo.[9] The main function of this counting method is to simplify the measured load history into several load cycles for fatigue life estimation and fatigue test load spectrum. It is based on the two-parameter method and considers two variables of dynamic strength (magnitude) and static strength (mean). The rain flow counting method is mainly used in the engineering field, and is widely used in the calculation of fatigue life.

Through the rain flow counting method, the response stress is cyclically counted, and the stress amplitude-cycle number curve is obtained, which is the ***p(s)*** curve required for the fatigue damage calculation. The ultimate goal of the frequency domain based PSD method is also to approximate the fitting through various frequency models through the frequency domain information, and finally obtain the probability density curve ***p(s)*** of the stress response.

The S-N curve of the material expresses the number of life cycles of the material under different stresses. Usually expressed in the Basquin model.[10]

$$S^m N = C \tag{7}$$

And then according Miner's Law[11], calculation formula of damage in unit time could be written

$$D = v_p C^{-1} t_0 \int_0^{\infty} S^m p(s) ds \tag{8}$$

Then the prediction life could be obtained by D

$$T = \frac{1}{D} \times t_0 \tag{9}$$

5. Case study

In order to study the influence of the probability distribution of the response stress on the structural damage calculation, using the same frequency range, the 0 mean, the RMS is the same, the skewness is 0, and the different time series with the kurtosis between 2.9-10 are applied to the same double notched specimen, the effect of kurtosis on the damage calculation is obtained by comparing the damage conditions.

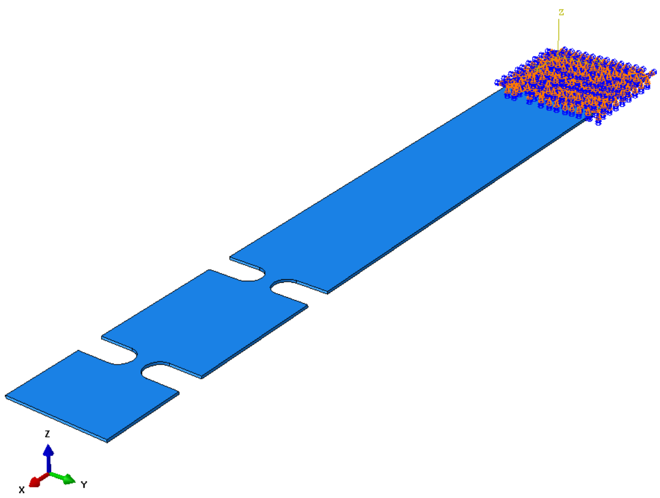


Figure 2 Two notch specimen & excitation

To meet the requirement, 10s time series of acceleration which allow non-Gaussian distribution were generated with different kurtosis. The bandwidth selection is based on the modal analysis result of the specimen, the second-order bending mode is 51.2 Hz, and 15-95 Hz is selected as the frequency range of the excitation signal. The sampling ratio is 1024Hz according to the band width.

Frequency	PSD (g^2/Hz)	(m^2/s^4)/Hz
15	0.25	12.03
95	0.25	12.03
RMS	4.472136g	31.02m/s^2

Table 1 Frequency domain information (g=9.81 m/s^2)

The simulation part is finished by Abaqus. Before taken simulation, the modal analysis was used to check FE model to do verify and validation according mode frequency and stress distribution. Though the modal dynamic, the response stress of different times series could be obtained.

No.	1	2	3	4	5	6	7	8
RMS(m/s^2)	31.44	31.32	31.27	31.31	31.10	30.84	31.54	31.44
Kurtosis	3.01	3.50	4.01	4.53	5.02	8.04	10.24	2.85
No.	9	10	11	12	13	14	15	16
RMS(m/s^2)	31.44	31.44	31.44	31.44	31.44	31.44	31.40	31.52
Kurtosis	2.92	2.80	3.06	2.96	3.12	2.90	3.15	3.21

Table 2 The RMS and Kurtosis of acceleration

After, the response stress result was show that the specimen is not linear structure, the time series of response stress was as below. The direction opposite to the gravitational acceleration is the Z-axis forward direction, and the nodal stress at the left side of the upper surface of the model is selected as the research object.

The overall research strategy is shown below. The focus is on the difference in stress cycling and damage values after rain flow counting. In general, the infinite fatigue stress of the material is selected as the threshold value of the rain flow count, that is, in this stress cycle and below, the material has an infinite life. However, in order to compare the results of the data, this threshold is set to zero.

6. Results

According the result of response stress, it could be found that the kurtosis were influence by structure. Moreover, as the acceleration kurtosis increases, the kurtosis of the response stress decays more significantly.

No.	1	2	3	4	5	6	7	8
RMS(MPa)	28.89	29.00	28.88	29.74	29.39	30.21	31.77	29.07
Kurtosis	3.07	3.30	3.52	3.92	3.70	6.11	6.49	2.85
No.	9	10	11	12	13	14	15	16
RMS(MPa)	28.89	29.09	29.17	28.34	28.89	28.45	29.00	28.93
Kurtosis	3.02	2.78	2.90	3.07	3.88	3.14	3.04	3.18

Table 3 The RMS and Kurtosis of response stress

By comparing the results of time-domain and frequency-domain methods with different kurtosis, it can be found that as the kurtosis increases, the stress distribution is more concentrated in the low-stress region, but the maximum value of the stress amplitude is significantly increased. The key to damage deviation. Moreover, there is a possibility that the maximum value of the stress amplitude is greater than the allowable stress directly causing structural failure.

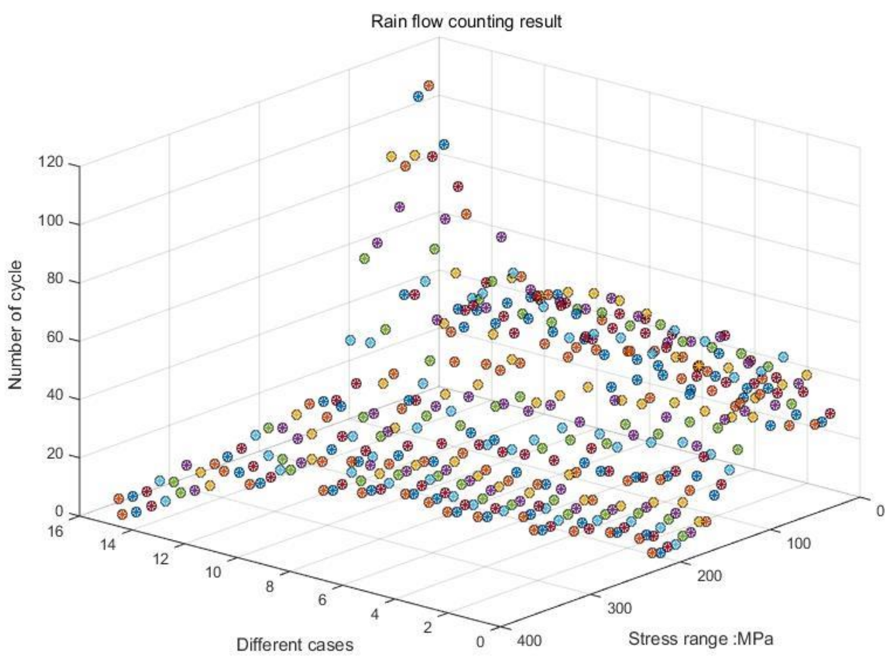


Figure 3 Rain flow counting result

The DIRLIK method is selected to obtain the stress expectation based on the Gaussian distribution of the frequency domain method, which is used as a reference object and compared with the results under the time domain signal.

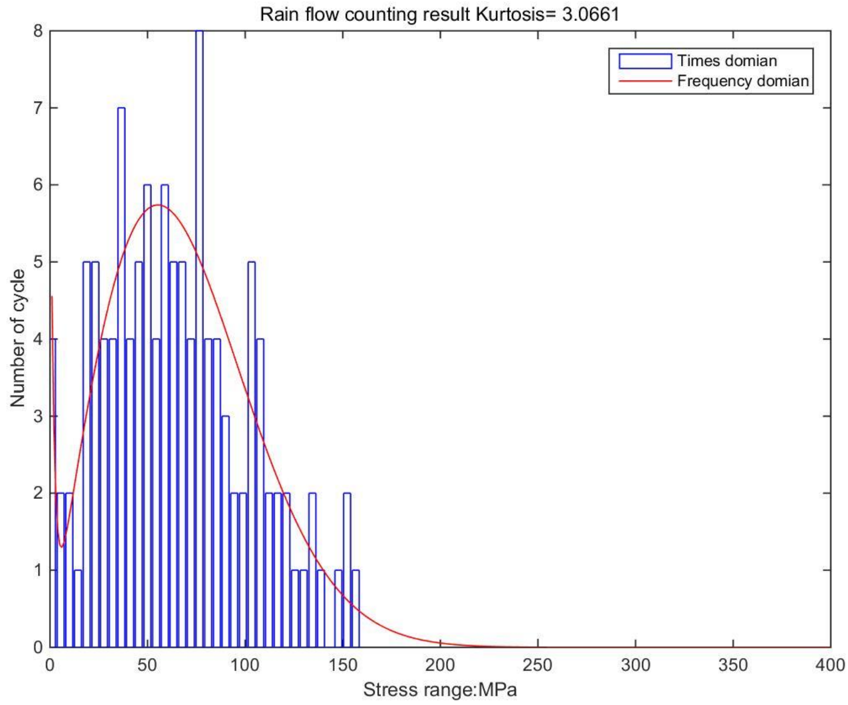


Figure 4 Rain flow counting of K=3.07

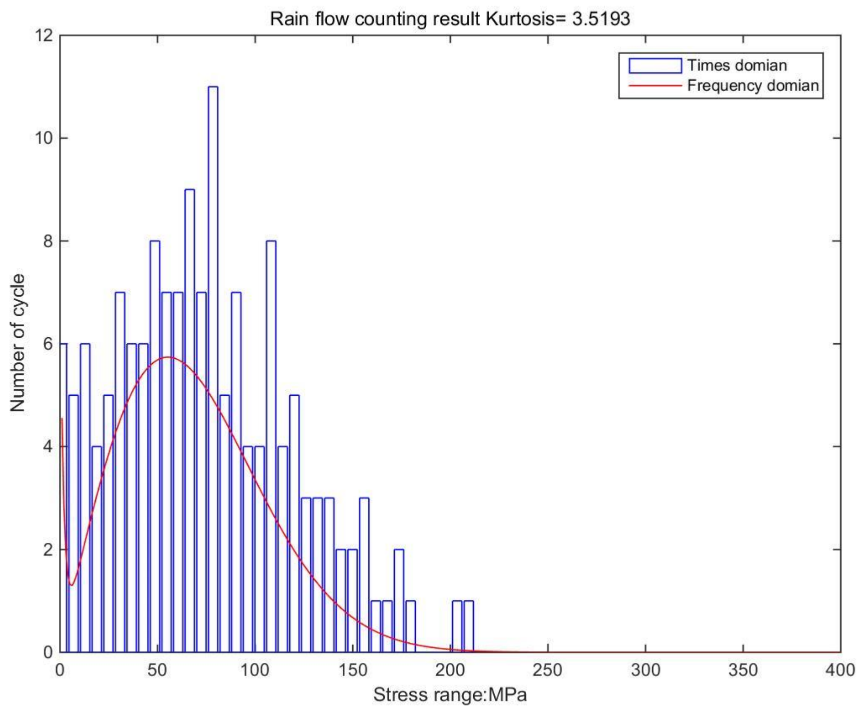


Figure 5 Rain flow counting of K=3.52

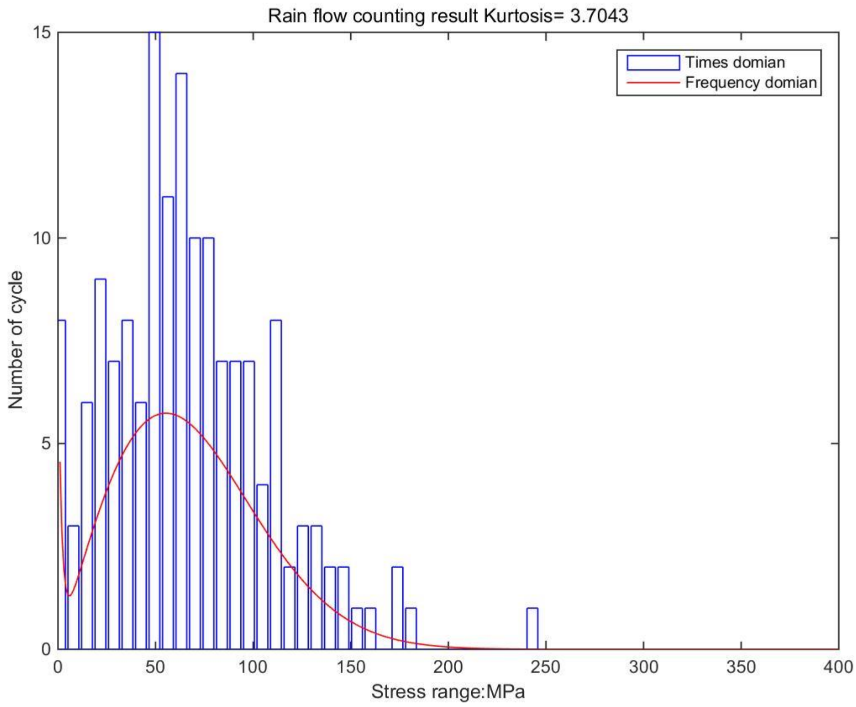


Figure 6 Rain flow counting of K=3.70

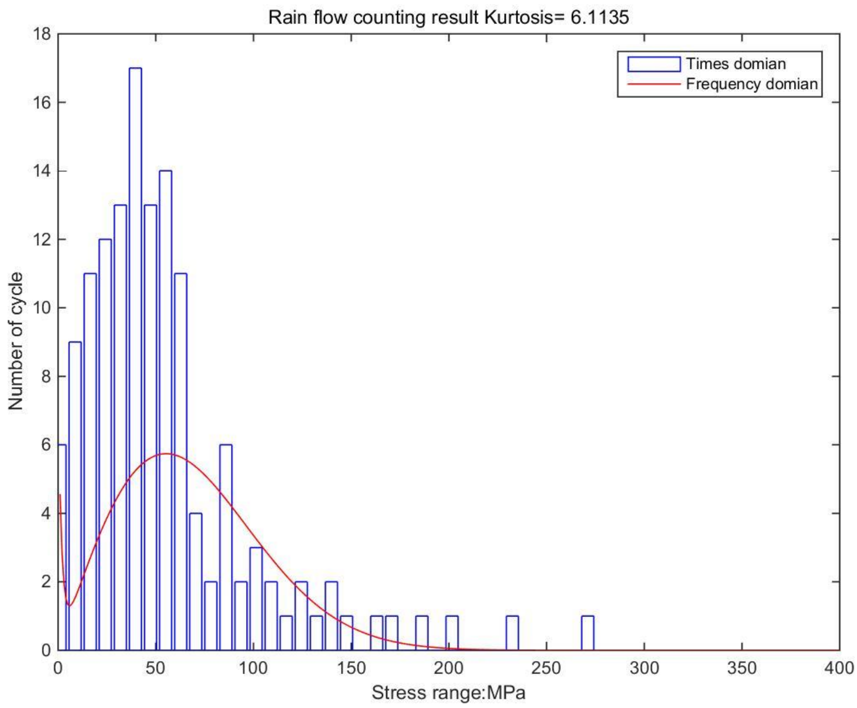


Figure 7 Rain flow counting of K=6.11

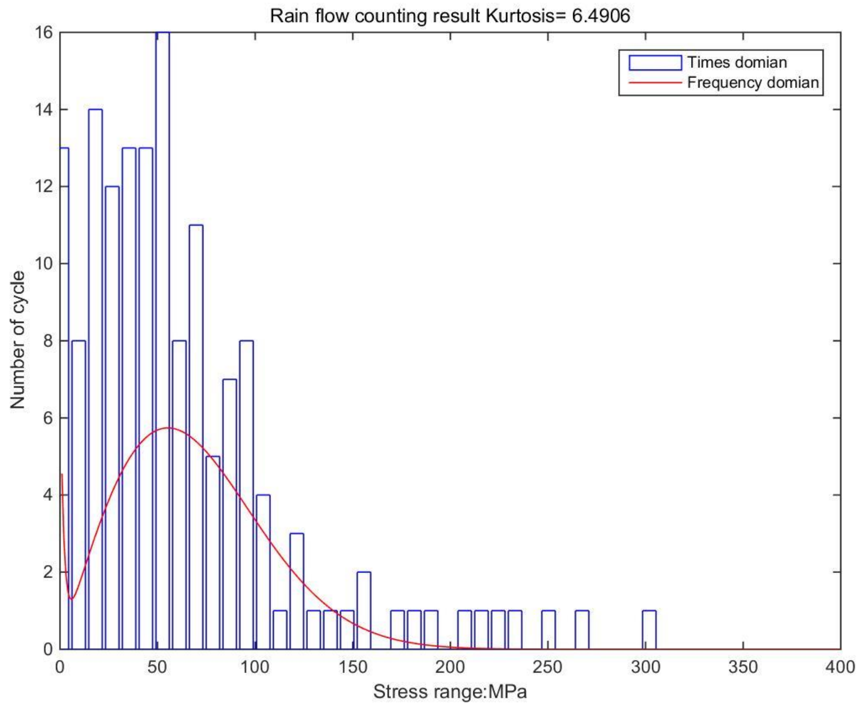


Figure 8 Rain flow counting of K=6.49

Though Miner's Law, the linear cumulative damager could be calculation,

$$D = \frac{n}{N} \tag{10}$$

For the different kurtosis time series data, the error control γ could be written ,

$$\gamma = 1 - \frac{D_K}{D_{PSD}} = \frac{\sum_0^{S_{max}} \cdot N_k(s) \cdot s^m}{v_p \int_0^\infty s^m p(s) ds} \tag{11}$$

When $\gamma \leq 30\%$, this can accept the results of using a PSD-based Gaussian signal for non-strict Gaussian vibration signals. Beyond this range, fatigue calculation methods based on non-Gaussian vibration signals must be used. Generally, when the kurtosis is less than 3.5, the result is within an acceptable range. Of course the accuracy of this result is limited by the length of the data. The longer the data, the more obvious the distribution. Smaller samples are relatively more affected by randomness.

7. Conclusion

Based on the analysis results, it can be found that the data difference is affected by the material property $\boldsymbol{m}$. At the same time, due to the structural relationship, the actual excitation kurtosis is greater than the threshold, but the response stress after attenuation can be calculated according to the Gaussian distribution.

Because of the high kurtosis, the response stress exceeds the limit of material strength, and should be avoided in industrial design. For response stresses with high kurtosis, there is a possibility that the maximum value exceeds the strength limit. This should be taken seriously.

8. Acknowledgement

This contribution has been elaborated under the **China Scholarship Council**

References:

[1] Y. WANG, R. SERRA *Life assessment of structures using random vibration testing*, ENVIRORISK Bourges 3ème edition June 2018

[2] R. SERRA, L. KHALIJ et al. *Effets de l'environnement vibratoire sur la durée de vie d'une éprouvette en aluminum*. SF2M, 35 èmes Journées de Printemps, Mai 2016.

[3] L. Crow, Edwin & S. Bendat, J & G. Piersol, A. (1968). Measurement and Analysis of Random Data. Technometrics. 10. 869. 10.2307/1267469.

[4] H. Wirsching, P & C. Light, Mark. (1980). Fatigue under wide band random stress. ASCE J Struct Div. 106.

[5] D.Benasciutti R.Tovo & A.Cristofori, (2013), Analogies between spectral methods and multi axial criteria in fatigue damage. Probabilistic Engineering Mechanics.

[6] Dirlik, Turan. (1985). Application of computer in fatigue analysis.

[7] Benasciutti, Denis & Tovo, Roberto. (2006). Fatigue life assessment in non-Gaussian random loadings. International Journal of Fatigue. 28. 733-746. 10.1016/j.ijfatigue.2005.09.006.

[8] Steinwolf, Alexander. (2015). Vibration testing of vehicle components by random excitations with increased kurtosis. International Journal of Vehicle Noise and Vibration. 11. 10.1504/IJVNV.2015.067983.

[9] Matsuichi, M & Endo, T. (1968). Fatigue of Metals Subjected to Varying Stress. Proceedings of the Kyushu Branch of Japan Society of Mechanics Engineering. 102.

[10] Basquin, O.H.. (1910). The exponential law of endurance testing. 19. 625-630.

[11] Miner, M.A.. (1945). Cumulative damage in fatigue. J. Appl. Mech.. 12. A159-A164.

Fault diagnosis and prognosis for rolling bearings

Synchronous fitting for deterministic signal extraction in non-stationary regimes: Application to helicopter vibrations

Dany ABBOUD[1], Yosra MARNISSI[1], Amadou ASSOUMANE[1], Mohammed ELBADAOUI[1,2]

[1] Signal & Data Analytics Lab, SafranTech, Rue des Jeunes Bois – Châteaufort 78772 Magny –les-Hameaux, France
[2]Univ Lyon, UJM St-Etienne, LASPI, EA3059, F-42023 Saint-Etienne, France

Dany.abboud@safrangroup.com

Abstract

Deterministic-random separation is crucial in machine signal processing. The synchronous average is a widely used tool that separates the deterministic contribution from the random one. This tool consists on averaging the cycles of the vibration signal. In fact, it uses the fact that, for a given location in the cycle, the associated samples have a constant mean. This makes possible to estimate the signal mean through synchronous averaging, i.e. by averaging the samples associated with each position in the cycle. However, in many practical applications, the cycle-to-cycle statistics can change according to many factors such as the speed, torque, load, etc. The resulting signal is widely referred in the literature as cyclo-non-stationary. This means that the mean signal is not periodic anymore, thus jeopardizing the synchronous average technique. This paper addresses this issue by proposing a new generalization of the synchronous average. The proposed method takes advantage of the smoothness of the statistics (in particular the mean) variation with respect to cycles. Instead of computing the average of the samples located at a given angular location, the time-varying mean is computed by optimally fitting the data with an appropriate curve. This defines the synchronous fitting idea, being a mean estimator of cyclo-non-stationary signals. Two solutions are proposed to solve the fitting problem. Whereas the first seeks for a global solution, the second adopts a local solution inspired from Savitzky–Golay filter. These two approaches are tested and compared on numerical and real signals captured from a helicopter engine operating under a runup regime. Overall, the results have asserted the superiority of the local approach over the global one.

1. Introduction

The theory of cyclostationary processes has proven to be effective in describing and processing rotating machine signals [Antoni 2009]. The vibration components generated by mechanical sources can be mainly classified into first and second (or higher) order cyclostationary classes. First order cyclostationary components are those deterministic, principally consisting of a set of sinusoids corrupted with stationary noises. Those are described by the (quasi-) periodicity of their mean. Examples of first-order phenomena can include gear meshing vibrations, shaft unbalance and misalignment, fan rotations and others. Second-order cyclostationary components are random in nature, meaning that their mean equals zero or, equivalently, their spectrum does not exhibit clean harmonics. The periodicity of these components is hidden and can be revealed through the instantaneous power or, more generally, the auto-covariance function. In practice, those components are generated by different kind of mechanical phenomena subjected to some randomness. A typical example is the vibrations generated by a local fault in a rolling element bearing wherein the randomness is due to the presence of a slippage in the motion of the rolling elements [Ho 2000]. For this reason, the cyclostationary analysis offers an efficient way to detect and characterize the presence of clear or hidden periodicities in the signals through a thorough differential diagnosis. Obviously, the separation of first and second-order components is crucial for an accurate analysis of the signal. Nowadays, the corresponding state of the art comprises a set of supervised and unsupervised signal processing tools that deal with this issue. Among these methods, one of the most widely used is the synchronous average (SA) [Braun 1975]. The latter simply consists of cutting the signal into slices of the same length, being equal to the fundamental period of the extracted component and averaging them together. As it will be shown later, this paper deals with its generalization.

The cyclostationary modelling assumes the (hidden-) periodicity to be stable in time, which in turn requires a constant speed. Such a condition is however hard to obtain as the speed often undergoes some fluctuations. This jeopardizes the effectiveness of the SA even if the magnitude of the speed fluctuations is low. Since repetitive patterns in rotating machines are intrinsically locked to specific angular positions, it totally makes sense to rather process the signal in the angular domain. In this case, the cyclostationary property holds in the angle domain and, consequently, the SA is applied on the angular signal, either obtained by angular sampling or resampling [Antoni 2004].

In the case of large speed fluctuations, signals are subjected to significant distortions that jeopardize the effectiveness of the SA. These distortions are basically introduced by (i) variations of the machine power intake and (ii) the effect of linear time-invariant (LTI) transfers. Whereas the former essentially results in amplitude modulation, the latter also induces phase modulation. Non-periodic modulations obviously invalidate the (angle-) CS assumption and call for a more general description of nonstationary signals. Accordingly, the principle of cyclo-non-stationarity was proposed to formalize this specific type of signals. The consideration of cyclo-non-stationary signals requires the extension of the cyclostationary signals. This paper is particularly concerned in extending the synchronous average. Many previous works have addressed this issue. Reference [Coats 2009] proposed the improved synchronous average, being based on resampling the signal with a virtual tachometer signal synthesized via the demodulated phase. Another attempt to generalize the SA was proposed in Ref. [Daher 2010] through a parametric approach. In details, the authors used the Hilbert space representation of the deterministic component in which they decomposed the deterministic components onto a set of periodic functions multiplied by speed-dependent functions apt to capture long-term evolution over consecutive cycles. In ref. [Abboud 2016], the authors proposed a non-parametric approach based on averaging the signal cycles that belong to a given regime, defined by its central speed and a pre-defined width.

This paper proposes a different approach to generalize the SA based on a synchronous curve fitting of the data. The theoretical backgrounds of the proposed technique is exposed in section 2, while its performances are evaluated through numerical simulations in section 3. In section 4, the efficiency of the technique is tested on real vibration signals recorded under a varying speed condition.

2. Description of the synchronous fitting technique

In this section, the fundamentals of the proposed method are provided. First, a mathematical model for CNS signals is reviewed. Then, a global solution for the first order CNS estimation, which corresponds to the one proposed in [Daher 2009], is presented. Finally, the newly proposed technique based on a local solution is introduced.

2.1. General

Let $x[n]$ be a first-order CNS signal with a characteristic period N (i.e. cycle and $1/N$ the normalized frequency) and a length L. One can model such a signal as follows:

$$\forall n \in \{1, \dots, L\} \quad x[n] = d[n] + w[n] = \sum_k d_k[n] e^{\frac{j2\pi kn}{N}} + w[n] \tag{1}$$

where k is an integer, $d_k[n] \in \mathbb{C}$ are deterministic smooth functions (whose real and imaginary part are continuous and differentiable) and whose bandwidths, noted B_k, are much smaller than the half the fundamental frequency i.e. : $\forall k, B_k \ll 1/2N$ and $w[n]$ is a random noise. The discrete-time Fourier transform (DTFT) of (1) reads:

$$\forall f \in]-1/2; 1/2] \quad X[f] = \sum_k D_k[f] * \delta(f - k/N) + W[f] \tag{2}$$

where $D_k[f]$ and $W[f]$ are respectively the DTFTs of $d_k[n]$ and $w[n]$. Since $d_k[n]$ are deterministic smooth functions, and according to the Weirstrass theorem, they can be approximated through a P-order polynomial function, i.e.:

$$\forall k \quad d_k[n] \approx \sum_{p=0}^{P} d_k^p n^p \tag{3}$$

where $d_k^p \in \mathbb{C}$. By inserting Eq. (3) into the expression of $d[n]$, one obtains:

$$\forall n \in \{1, \dots, L\} \quad d[n] = \sum_{p=0}^{P} c_p[n] n^p \tag{4}$$

where $c_p(n) = \sum_k d_k^p e^{j2\pi kn/N}$ is a periodic function of period N. Equation (4) indicates that the deterministic component can be approximated by a sum of periodic functions multiplied with the polynomial basis: it is actually a polynomial with periodic coefficients.

Let's first define $\bar{n} = \lfloor (n-1)/N \rfloor + 1$ as the sample location within the period N ($\lfloor a/b \rfloor$ denotes the remainder of the division of a by b) . Since $c_p(n)$ is periodic with period N, we have $c_p[\bar{n}] = c_p[\bar{n} + (q-1)N]$ for all integer $q = 1, \dots, Q$ (Q is the number of cycles). Thus, Eq. (4) can be equivalently written as follows:

$$\forall q \in \{1, \dots, Q\} \ \forall \bar{n} \in \{1, \dots, N\} \quad d[\bar{n} + (q-1)N] = \sum_{p=0}^{P} c_p[\bar{n}](\bar{n} + (q-1)N)^p \tag{5}$$

Using the binomial theorem ($(\bar{n} + (q-1).N)^p = \sum_{i=0}^{p} C_i^p N^i (\bar{n} - N)^{p-i} q^i$ where C_i^p is the binomial coefficient), one can deduce from Eq. (5) that the samples associated with the same location $\bar{n}$ in the period, $s_q[\bar{n}] = d[\bar{n} + (q-1).N]$ for all integer $q \in \{1, \dots, Q\}$, defines a polynomial of order P with constant coefficient, i.e.:

$$\forall q \in \{1, \dots, Q\} \ \forall \bar{n} \in \{1, \dots, N\} \quad s_q[\bar{n}] = \sum_{p=0}^{P} b_p[\bar{n}] q^p \tag{6}$$

where $b_p[\bar{n}] = N^p \sum_{j=p}^{P} C_p^j (\bar{n} - N)^{j-p} c_j[\bar{n}]$. Note that $b_p[\bar{n}]$ is parametrized by $\bar{n}$.

2.2. A global LMS solution

In the case of a noisy signal $x[n]$, a good estimate of the deterministic component $d[n]$ is then to find the best fit of the curve $\boldsymbol{s}[\bar{n}] = \left[s_1[\bar{n}], \dots, s_Q[\bar{n}]\right]^T$ for each $\bar{n} \in \{1, \dots, N\}$ which reduces to find an estimate of $\boldsymbol{b}[\bar{n}] = \left[b_1[\bar{n}], \dots, b_P[\bar{n}]\right]^T$ for each $\bar{n} \in \{1, \dots, N\}$ for a given polynomial of order P. A common way to do this is to find the curve which minimizes the least mean square error, i.e.:

$$\begin{aligned} \forall \bar{n} \in \{1, \dots, N\} \quad \hat{\boldsymbol{b}}[\bar{n}] &= \operatorname{argmin}\left(\sum_{q=1}^{Q} w[\bar{n} + (q-1)N]^2\right) \\ &= \operatorname{argmin}\left(\sum_{q=1}^{Q} \left(d[\bar{n} + (q-1)N] - x_q[\bar{n}]\right)^2\right) \\ &= \operatorname{argmin}\left(\sum_{q=1}^{Q} \left(\sum_{p=0}^{P} b_p[\bar{n}] q^p - x_q[\bar{n}]\right)^2\right) \end{aligned} \tag{7}$$

where $x_q[\bar{n}] = x[\bar{n} + (q-1)N]$. Let's define the $Q \times (P+1)$ matrix $\boldsymbol{\Phi}$ such that $\boldsymbol{\Phi}_{q,p} = q^{p-1}$ (with $q \in \{1, \dots, Q\}$ and $p \in \{1, \dots, P+1\}$, and $\boldsymbol{x}[\bar{n}] = \left[x_1[\bar{n}], \dots, x_Q[\bar{n}]\right]^T$. One can rewrite the Eq. (7) as:

$$\forall \bar{n} \in \{1, \dots, N\} \ \hat{\boldsymbol{b}}[\bar{n}] = \operatorname{argmin} \|\boldsymbol{\Phi} \boldsymbol{b}[\bar{n}] - \boldsymbol{x}[\bar{n}]\|^2 \tag{8}$$

whose solution expresses as follows:

$$\forall \bar{n} \in \{1, \dots, N\} \ \hat{\boldsymbol{b}}[\bar{n}] = \left(\boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}^{\mathbf{T}} \boldsymbol{x}[\bar{n}] \tag{9}$$

Once the coefficients are calculated, one can find the estimated deterministic component located at $\bar{n}$ in the form:

$$\forall \bar{n} \in \{1, \dots, N\}\;\; \hat{\boldsymbol{s}}[\bar{n}] = \boldsymbol{\Phi}\hat{\boldsymbol{b}}[\bar{n}] = \boldsymbol{\Phi}\left(\boldsymbol{\Phi}^{\mathbf{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathbf{T}}\,\boldsymbol{x}[\bar{n}] \tag{10}$$

The deterministic signal can then be deduced as follows:

$$\forall n \in \{1, \dots, L\}\;\; d[n] = s_q[\bar{n}] \;\; \text{where} \;\; \bar{n} = \lfloor (n-1)/N \rfloor + 1 \;\; \text{and} \;\; q = 1 + (n - \bar{n})/N \tag{11}$$

2.3. A local LMS solution

This subsection describes the proposed method. The basic idea is excerpted from the "Savitzky-Golay filter" which is a widely known method to smooth or fit the data based on the least mean square solution of local polynomial fitting [Savitzky 1964]. Precisely, for every $q \in \{1, \dots, Q\}$, let's consider the data set $x_q[\bar{n}]$ being a function of q and parametrized by $\bar{n}$; we try to find the best LMS polynomial fit, with a fixed order P at the point $\bar{n}$, from the $2M+1$ subset centered at q, i.e. $\{x_{q-M}[\bar{n}], \dots, x_{q+M}[\bar{n}]\}$. That being said, this problem can be stated in a similar way as the previous subsection, i.e.:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\}\;\; \hat{\boldsymbol{b}}^{(\boldsymbol{q})}[\bar{n}] = \operatorname{argmin}\left\|\mathbf{J}\,\boldsymbol{b}^{(\boldsymbol{q})}[\bar{n}] - \boldsymbol{x}^{(q)}[\bar{n}]\right\|^2 \tag{12}$$

where $\boldsymbol{x}^{(q)}[\bar{n}] = \left[x_{q-M}[\bar{n}], \dots, x_{q-M}[\bar{n}]\right]^{\mathrm{T}}$ represents the q^{th} subset, $\boldsymbol{b}^{(\boldsymbol{q})}[\bar{n}] = \left[b_0^{(\boldsymbol{q})}[\bar{n}], \dots b_P^{(\boldsymbol{q})}[\bar{n}]\right]^{\mathrm{T}}$ are the $P+1$ polynomial coefficients associated with the q^{th} subset, and $\mathbf{J}$ the $(2M+1)\times(p+1)$ matrix such that $\forall m \in \{1, \dots, 2M+1\}\, \forall p \in \{1, \dots, P+1\}\;\; \mathbf{J}_{m,p} = (m - M + 1)^{p-1}$. The $(2M+1)$-length curve that best fits the q^{th} subset writes:

$$s_m^{(q)}[\bar{n}] = \textstyle\sum_{p=0}^{P} b_p^{(q)}[\bar{n}]\,.\,(m - M + 1)^p \tag{13}$$

The Savitzky-Golay method suggests to estimate the deterministic component at the q^{th} data point by retaining the value of the polynomial at the central point i.e. at $m = M+1$:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\}\;\; \hat{d}[\bar{n} + (q-1).N] = \hat{s}_{M+1}^{(q)}[\bar{n}] = b_{M+1}^{(q)}[\bar{n}] \tag{14}$$

Following the same lines as for Eq (9), one can show that the coefficients of the polynomial write:

$$\forall\, q \in \{1, \dots, Q\}\, \forall \bar{n} \in \{1, \dots, N\}\;\; \hat{\mathbf{b}}^{(\boldsymbol{q})}[\bar{n}] = \boldsymbol{H}\,\boldsymbol{x}^{(q)}[\bar{n}] \tag{15}$$

with $\boldsymbol{H} = \left(\mathbf{J}^{\mathbf{T}}\mathbf{J}\right)^{-1}\mathbf{J}^{\mathbf{T}}$ a matrix of size $(P+1)\times(2M+1)$ whose elements are independent of $\bar{n}$ and q. The $(M+1)^{th}$ element of the above vector namely $\hat{b}_{M+1}^{(q)}[\bar{n}]$ is actually a linear combination of $\boldsymbol{x}^{(q)}[\bar{n}]$ with the $2M+1$ elements of the $(M+1)^{th}$ row, $\boldsymbol{h}^{\boldsymbol{T}} = [h_{-M}, \dots h_M]$, of $\boldsymbol{H}$ being independent of q and $\bar{n}$:

$$\hat{b}_{M+1}^{(q)}[\bar{n}] = \boldsymbol{h}^{\boldsymbol{T}}\boldsymbol{x}^{(q)}[\bar{n}] \tag{16}$$

Considering equations (14) and (15), one can write the estimate of the deterministic component

$$\begin{aligned} \hat{d}[\bar{n} + (q-1).N] &= \textstyle\sum_{m=-M}^{M} x_{q-m}[\bar{n}]\, h_m \\ &= \textstyle\sum_{m=-M}^{M} x[\bar{n} + (q-1).N - m.N]\, h_m \\ &= \textstyle\sum_{i=-MN}^{MN} x[\bar{n} + (q-1).N - i]\, \tilde{h}_i \end{aligned} \tag{17}$$

where $\tilde{\boldsymbol{h}}^T = [\tilde{h}_{-MN}, \dots, \tilde{h}_{MN}]$ is obtained by zero-padding $\boldsymbol{h}$ as follows:

$$\begin{cases} \tilde{h}_i = h_m & if \ \ i = mN, \ \forall - M \leq m \leq M \\ \tilde{h}_i = 0 & elsewhere \end{cases} \tag{18}$$

It becomes obvious that the estimated deterministic component turns to a LTI filtering of the original signal $x[n]$ with the $(2MN+1)$-length filter $\tilde{\boldsymbol{h}}_i$:

$$\hat{d}[n] = \sum_{i=-MN}^{MN} x[n-i]\, \tilde{h}_i \tag{19}$$

3. Numerical evaluation

In this section, the performance of the synchronous fitting techniques are tested and compared on a synthetic signal. The deterministic signal is modelled as a sum of four speed-varying sinusoids whose envelopes and phases are functions of the cyclo-non-stationary $\lambda[n]$ (which can be in practice the torque, load, speed, etc.):

$$d[n] = \sum_{k=1}^{4} A_k[n] \sin(2\pi kn/N + \Phi_k[n]) \tag{20}$$

Where:

- $A_k[n]$ and $\Phi_k[n]$ are functions of $\lambda[n]$ (see Fig. 1);
- $N = 100$ is the fundamental period;
- $L = 15000$ is the signal length.

The deterministic signal is exposed in Fig 2 together with its noisy version constituted by adding a white Gaussian noise such that the initial signal to noise ratio is equal to -3 dB.

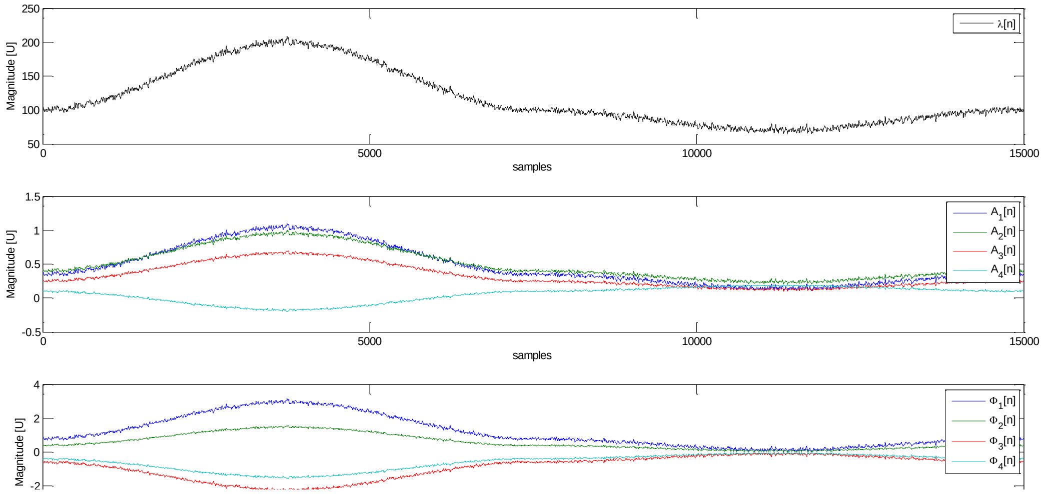


Figure 1: The plot of the cyclo-non-stationary agent $\lambda[n]$ (top), the 4 amplitudes $A_k[n]$ (middle) and the 4 phase modulations $\Phi_k[n]$ (bottom) associated with the sinusoids of the synthetic signal.

In the following, the global and local approaches are applied to the noisy signal with respect to the cycle N with the aim of recovering the deterministic component, being here the signal of interest. For the global approach, the degree of polynomial was set to 30, knowing that the results were stable for higher polynomial degrees. For the local approach, the window length was set to 49 (i.e. $M = 24$) and the polynomial order to 3. The obtained results are exposed in Fig. 3 together with the error signal obtained by subtracting the estimated signal from the actual one. Both approaches tend to estimate with good accuracy the deterministic signal, with a clear superiority of the newly proposed local approach over the global one: the estimation error of the latter is almost twice larger than the former.

Eventually, the performance of these methods are compared for different signal-to-noise ratios (SNR). For this purpose, the relative error defined as the energy of the error normalized by the signal energy:

$$\epsilon_n[n] = 10\log_{10}\left(\sum_{n=1}^{N}\left(d[n]-\hat{d}[n]\right)^2 / \sum_{n=1}^{N} d[n]^2\right). \tag{21}$$

The obtained results are exposed in Fig. 4. The SA returns poor and consistent results as this latter only estimates the average periodic component existing in the signal. The reason is that the number of average is big so it was slightly affected by the SNR: the average periodic part was almost the same for all SNR. When it comes to the synchronous fitting techniques, the local approach evidences better estimation performances. In fact, the local approach returns an estimation error less than -6 dB when the SNR greater than -6 dB, while the global approach needs a SNR greater than 10 dB to get this accuracy. The results highlight the effectiveness of the local approach as compared with the global one. The reason is that the local approach assumes that the mean is locally smooth which is a more accurate assumption.

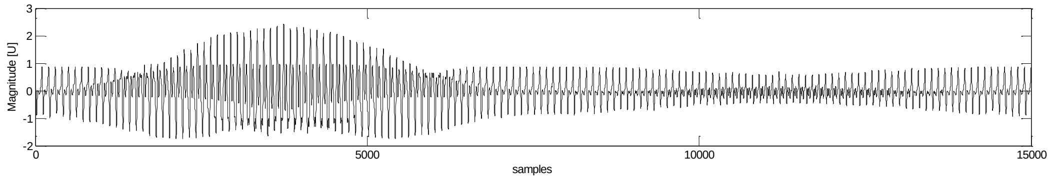


Figure 2: The deterministic component (top plot) and the noisy signal (bottom plot) constituted by adding a white Gaussian noise whose standard deviation equals twice of the former.

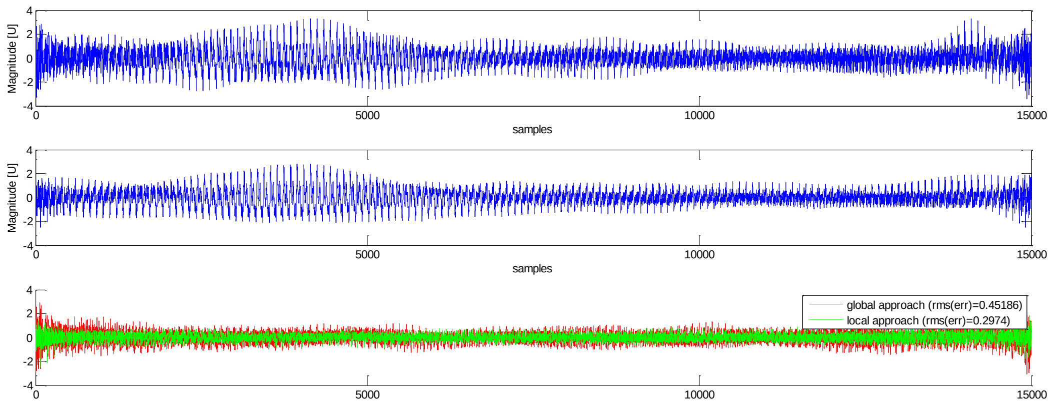


Figure 3: The deterministic component (top plot) and the noisy signal (bottom plot) constituted by adding a white Gaussian noise whose standard deviation equals twice of the former.

4. Application: a helicopter engine

In this section, the proposed approach is applied on real vibration signals captured from the gas generator of a helicopter engine. The aim is to extract the component related to the centrifugal compressor while the engine speed operates under a runup regime. An encoder is also present to measure the shaft location and to provide an accurate estimation of the engine speed. The encoder signal is used to resample the signal in the angular domain and the synchronous fitting techniques are both applied with respect to the blade pass period of the centrifugal compressor. It is worth noting that the blade pass period equals the shaft period divided by the compressor blade number. The blade number as well as the signals magnitude are not given for confidentiality reason. Figure 5 exposes the raw acceleration signal, the synchronous fitting estimations via the global and local approach. It is obvious that the signal associated with the local approach is much more accurate presenting a clear resonance starting at 10s. The related spectrograms are exposed in Figure 6 wherein the speed-varying harmonics of the centrifugal compressor are clearly shown. Though the extraction seems good in both techniques, it was hard to compare the performance of the extraction techniques. The order spectrum of the signals are computed by the Welch estimator applied to the angular resampled signals. The obtained spectra are exposed in Fig. 6 for comparison. Whereas the latter show better noise rejection in the global solution case especially for the noise floor, the close-ups clearly evidence the superiority of the local approach in accurately estimating the peaks. The global approach tends to loose accuracy as the frequency (order) gets high, this is because the global interpolation tends to confuse high frequency sinusoids with high frequency noises. Overall, the newly proposed local approach evidences better performances than the global one.

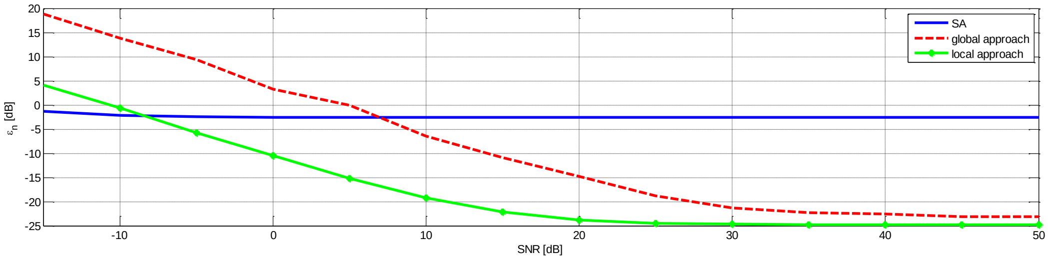


Figure 4: Performance of the SA, the synchronous fitting with the global and the local approach.

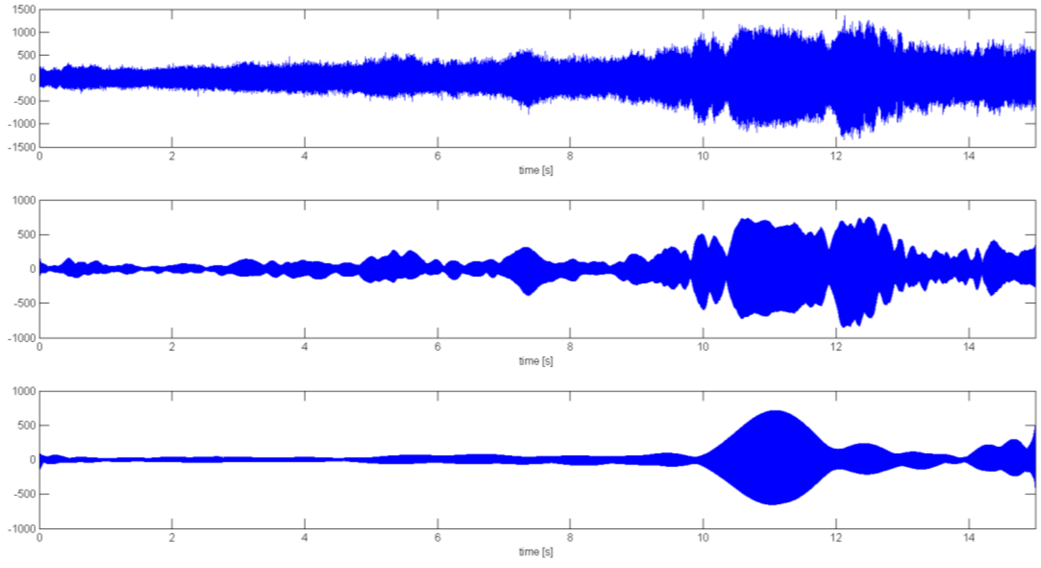


Figure 5: The raw acceleration signal (top), the synchronous fitting with the global approach (middle) and local approach (bottom).

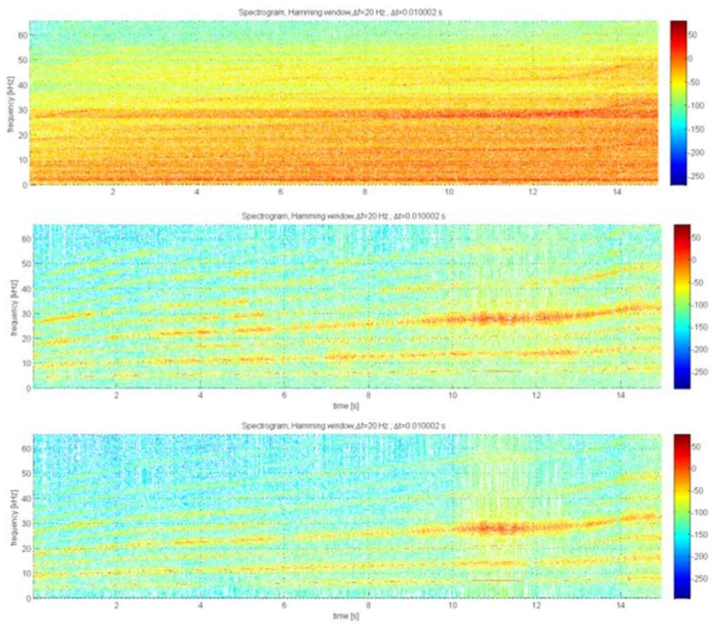


Figure 6: Spectrograms of the raw acceleration signal (top), the synchronous fitting with the global approach (middle) and local approach (bottom).

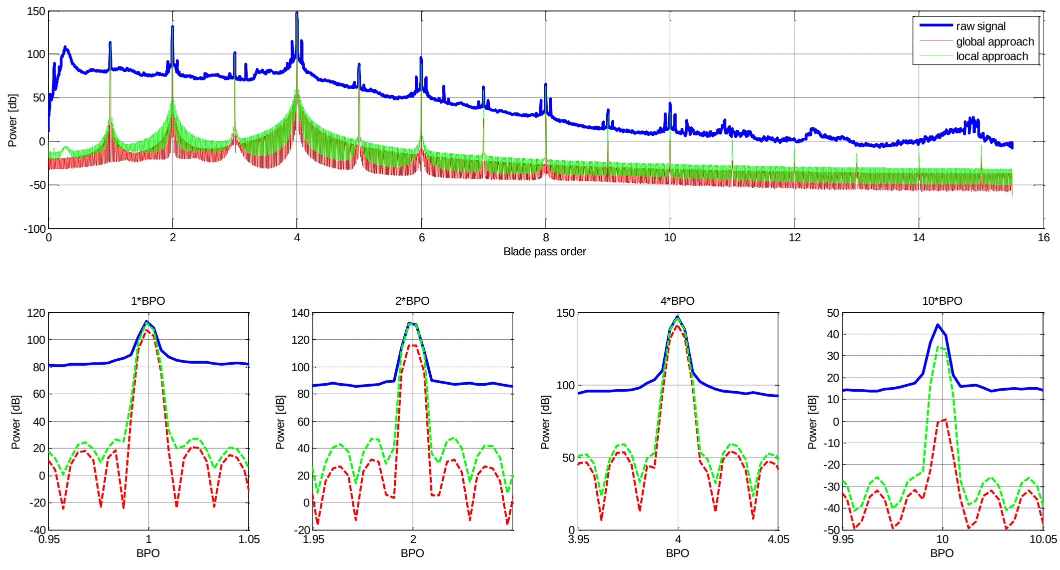


Figure 7: Order spectra of the raw acceleration signal (blue continuous line), the synchronous fitting with the global approach (red dotted line) and local approach (green dotted line).

5. Conclusion

This paper proposes a new technique for the extraction of a deterministic component in variable regimes. It can be seen as an extension of the classical synchronous averaging. In fact, instead of computing the cyclic mean via the synchronous average, the latter is computed via a synchronous fitting. This leads to two ways to tackle the issue. The first way seeks a global solution and compute the mean, for the data associated with a given position in the cycle, by finding the best polynomial that minimizes the least mean square error. It turns out that the solution of this problem was previously proposed in a previous publication. However, the second way is original and addresses differently the same problem by seeking a local solution based on the Savitzky-Golay filter. Numerical simulations are conducted showing a clear superiority of the newly proposed local approach over the global one. In fact, the local approach returns an estimation error less than -6 dB when the SNR greater than -6 dB, while the global approach needs an SNR greater than 10 dB to get this accuracy. The results highlights the effectiveness of the local approach as compared with the global one. An additional advantage of the local approach over the global one is the computational cost. As the first turns to a linear-time-invariant convolution, its implementation is much easier than solving a global least mean square problem which requires a matrix inversion. Eventually, both techniques are successfully tested on a real vibration signal measured on a helicopter engine under a runup condition with the aim of extracting the vibratory component emitted by the centrifugal compressor of the gas generator. Both techniques were able to extract the component of interest, yet the local approach evidences much better extraction than the global one.

References

[Antoni 2009] J. Antoni, Cyclostationarity by examples, Mech. Syst. and Sign. Proc., 23 (2009) 987–1036

[Ho 2000] D. Ho, Randall R. B., Optimization of bearing diagnostic techniques using simulated and actual bearing fault signals. Mech. Syst. and Sign. Proc. 14, 763-788 (2000)

[Braun 1975] S. Braun, The extraction of periodic waveforms by time domain averaging, Acta Acustica United with Acustica, vol. 32, no. 2, pp. 69-77, 1975.

[Antoni 2004] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, *Cyclostationary modelling of rotating machine vibration signals*. Mech. Syst. And Sig. Proc. 18, 1285–1314 (2004)

[Coats 2009] M.D. Coats, N. Sawalhi, R.B. Randall, Extraction of tacho information from a vibration signal for improved synchronous averaging, Proc. Acoust. (2009).

[Daher 2010] Z. Daher, E. Sekko, J. Antoni, C. Capdessus and L. Allam, Estimation of the synchronous average under varying rotating speed condition for vibration monitoring, Proceedings of ISMA 2010.

[Abboud 2016] D. Abboud, J. Antoni, S. Sieg-Zieba, M. Eltabach, *Deterministic-Random separation in non-stationary conditions*, Journal of Sound and Vibration, Volume 362, 3 February 2016, Pages 305-326

[Savitzky 1964] Abraham Savitzky et Marcel J. E. Golay, « Smoothing and Differentiation of Data by Simplified Least Squares Procedures », Analytical Chemistry, vol. 8, no 36, 1964, p. 1627–1639 (DOI 10.1021/ac60214a047)

High Frequency Demodulation Technique for Instantaneous Angular Speed Estimation

Frédéric Bonnardot[1], Khalid Lizoul[1], Saad Errafik [1], Hugo André[1], François Guillet[1]

[1]Univ Lyon, UJM Saint Etienne, LASPI, EA 3059
F-42334, IUT de Roanne, France.
{frederic.bonnardot}@univ-st-etienne.fr

Abstract

This paper adapts the super-heterodyne method from telecommunication field to speed estimation based on optical encoder signal.

This method uses an analog frequency shifting before sampling of the signal. Therefore, the required sampling frequency is reduced and linked with the speed fluctuation frequency. Spectrum of optical encoder signal is analysed and used to explain how to set up the frequency shift and limitations. Finally, a comparison is made with elapsed time.

1 Introduction

Many vibration monitoring methods use an accelerometer sensor in order to diagnose the systems. Instantaneous Angular Speed (IAS) has appeared to provide a new source of information on the rotating machine and it is commonly used to diagnose faults such as bearing faults[1], gearboxes [2], ...

Classically, IAS use an optical encoder as a speed sensor. This sensor is available in many rotating machines (especially for speed and position control). It provides a square / sine wave with a frequency proportional to the rotation speed. Thus, this signal is frequency modulated by the system speed variations. There are two major methods for recording IAS signal: timer / counter technic [3], and ADC- based methods [4]. The latter is limited by the capability of the ADC-board to collect the data (sampling frequency) which could restrict the use of higher encoder resolution and introduces spatial aliasing.

In this study, we are interested in IAS estimation using a super-heterodyne like demodulation technique. The idea is to be able to acquire the speed signal with relatively higher resolution without using a higher sampling frequency. The IAS signal is first analogously shifted in frequency domain in order to be acquired at a lower sampling rate. Then, the sampled signal is further treated to get IAS.

A first part presents the measurement principle, a second part discuss about limitations, finally a last part show a comparison with elapsed time technique.

2 Measurement principle

In frequency demodulation technique, the signal is recorded at a high sample rate in order to exploit only the spectral content between $[f_{cod} - B/2; f_{cod} + B/2]$. The idea of super-heterodyne technique illustrated by figure 1 that come from telecommunication is to shift analogically the optical encoder signal from the band $[f_{cod} - B/2; f_{cod} + B/2]$ to the band $[\varepsilon; \varepsilon + B]$ by using an analog multiplier (ε is a small margin). Therefore, the recorded signal could be recorded with a sampling frequency greater than $2 \times B$. Since the band B (a few kiloHertz) is smaller than optical encoder frequency, it becomes possible to use high resolution (4096 lines per revolution) with a classical data acquisition device. The demodulation effect is simply compensated after computing instantaneous speed by adding a constant frequency.

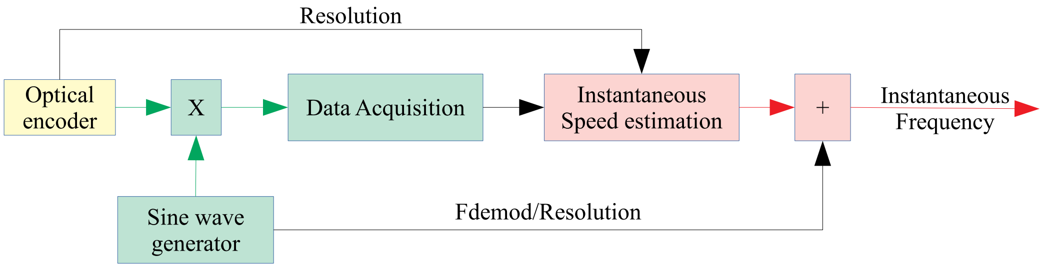


Figure 1: Principle of High frequency Demodulation Technique.

In order to illustrates the analog frequency shifting part, the sine wave generator, optical encoder signal and product signal has been acquired. Due to the limitation of data acquisition device, an optical encoder resolution of 256 lines has been used (to visualise original optical encoder signal). The sine wave generator has been set to $F_{mod} = 4\,kHz$. The spectrum of each signal is presented in figure 2.

The first spectrum is the spectrum of the sine wave generator. On top of the peak at $4\,kHz$, a third harmonic $68\,dB$ lower (i.e. amplitude 2400 lower) indicate the quality of the generator. It will also ask the question latter of the influence of a third harmonic on low-cost generators. The low noise in the spectrum is produced by the engine inverter as well as the generator.

The second spectrum corresponds to the optical encoder and was already studied in the first part. The effect of multiplication of encoder signal by sine wave results in a displacement of each component of the encoder spectrum by F_{mod} on the left and on the right as indicated by the arrow. Each component is identified by a colour to help analysis after demodulation.

The last spectrum shows the demodulated signals by using the analog multiplier. Each optical encoder component has been shifted in both left and right direction. The amplitude ratio of $20\,dB$ between 1^{st} and 3^{rd} optical encoder harmonic was kept after demodulation (on the bottom spectrum).

Unfortunately, components at f_{cod}, $3.f_{cod}$ (marked in gray) and F_{mod} are present after demodulation. It means that the product $p(t)$ signal is:

$$
\begin{aligned}
p(t) &= [encoder(t) + \varepsilon_e].[sin(2\pi F_{mod}t) + \varepsilon_s] && (1)\\
&= encoder(t).sin(2\pi F_{mod}t) + \varepsilon_e.\sin(2\pi F_{mod}t) + \varepsilon_s.encoder(t) && (2)
\end{aligned}
$$

Where :

- $encoder(t)$ is a zero mean version of encoder signal obtained by a high pass filter,
- ε_e is an DC component (introduced by electronic),
- ε_s is an DC component (introduced by electronic despite the fine tune of the sine wave generator).

The attenuation between the gray signal at f_{cod} and the original red signal is $41\,dB$, it means that $\varepsilon_s = 0.009$. By looking at the sine wave the value $\varepsilon_e = 0.001$ could be identified.

The values ε_e and ε_s are in fact small, and, reduce them necessitate higher precision offset adjustment.

In practice the last spectrum corresponds to the only acquired signal. The only interesting part of the signal is the left red pattern corresponding to the demodulated f_{cod} part. This part could be acquired by an acquisition with a sampling frequency of $8\,kHz$.

After acquisition, a classical instantaneous frequency scheme is applied (band pass filtering, analytic signal extraction, differentiation). A simple addition of F_{mod}/N_{ppt} is required to compensate the frequency shift of F_{mod}.

It should be noted that in real application the optical encoder frequency F_{cod} should be higher (for example $102.4\,kHz$ at $1500\,rpm$ and a resolution of 4096 lines).

This example helps to view some limitation of the method.

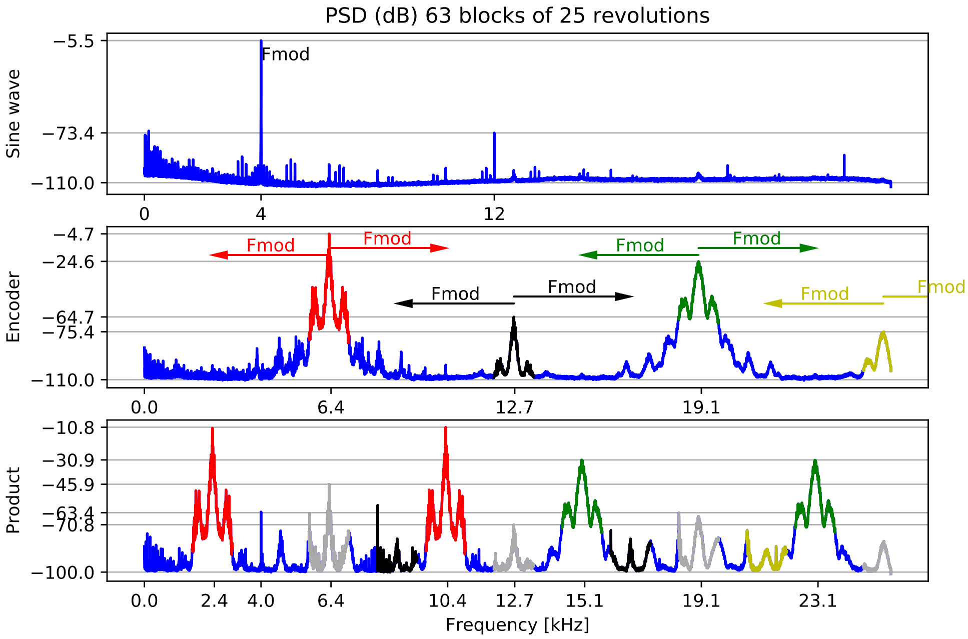


Figure 2: Illustration of High frequency Demodulation Technique.

3 Limit : Frequency overlapping

Let be β the highest fluctuation frequency in Hz (like in the previous section). By looking at the bottom spectrum of figure 2, it is possible to identify the following content :

- $[f_{cod} - f_{mod} - \beta, f_{cod} - f_{mod} + \beta]$: left shifted encoder frequency used for instantaneous frequency estimation,
- $[k.f_{cod} \pm f_{mod} - \beta, k.f_{cod} \pm f_{mod} + \beta]$: all left and right shifted encoder frequencies,
- f_{mod} : modulation frequency (should not be here if $\varepsilon_e = 0$),
- $[k.f_{cod} - \beta, k.f_{cod} + \beta]$: encoder frequencies (should not be here if $\varepsilon_s = 0$).

In order to have a good estimation, the left shifted encoder frequency should not overlap with the other area. Otherwise, the analytic signal will no longer be a sine wave and its phase will have a different meaning.

The frequency shifted red pattern should be kept in positive frequency. It leads to :

$$\begin{aligned} f_{cod} - f_{mod} - \beta &> 0 \qquad (3)\\ f_{mod} &< f_{cod} - \beta \qquad (4)\end{aligned}$$

In order to access the first harmonic of the encoder, the demodulation frequency f_{mod} is lower than f_{cod}.

If ε_e is not 0, no overlapping between the left frequency shifted red pattern and f_{mod} implies that :

$$\begin{aligned} f_{cod} - f_{mod} + \beta &< f_{mod} \qquad (5)\\ f_{mod} &> \frac{f_{cod} + \beta}{2}. \qquad (6)\end{aligned}$$

If ε_s is not 0, no overlapping between the left red frequency and the original one implies that :

$$\begin{aligned} f_{cod} - f_{mod} + \beta &< f_{cod} - \beta \qquad (7)\\ f_{mod} &> 2\beta \qquad (8)\end{aligned}$$

If $\varepsilon_s = 0$ and $\varepsilon_e = 0$, no overlapping between frequency left and right shifted version implies that :

$$\begin{aligned} f_{cod} - f_{mod} + \beta &< f_{cod} + f_{mod} - \beta & (9) \\ f_{mod} &> \beta & (10) \end{aligned}$$

To conclude, the modulation frequency f_{mod} should be lower than $f_{cod} - \beta$ and depending on whether ε_e, ε_s, or no ε are take into account, f_{mod} should be respectively greater than $\frac{f_{cod}+\beta}{2}$, 2β or β.

Another way of thinking should be to find the demodulation frequency f_{modop} that maximises the highest observable fluctuation frequency β :

$$f_{modop} = \frac{2}{3}.f_{cod} \quad , \quad \beta_{max} = \frac{1}{3}.f_{cod} \quad \text{considering } \varepsilon_e \text{ and } \varepsilon_s, \tag{11}$$

$$f_{modop} = \frac{f_{cod}}{2} \quad , \quad \beta_{max} = \frac{f_{cod}}{2} \quad \text{if } \varepsilon_e = \varepsilon_s = 0. \tag{12}$$

If offsets are not compensated, they limit β_{max} to order $N_{ppt}/3$ instead of $N_{ppt}/2$. The table 1 shows examples of demodulator setup.

Rotation frequency (rpm)	1500				3000	6000
Optical encoder resolution	256	512	1024	4096	4096	1024
Optical encoder mean frequency (kHz)	6.4	12.8	25.6	102.4	204.8	102.4
Optimal f_{modop} frequency (kHz)	4.3	8.5	17.0	68.3	136.5	68.3
Theoretical maximum order	85	170	341	1365	1365	341

Table 1: Example of demodulator setup

Since the sampling frequency of the acquisition device should be twice the f_{modop} frequency in this particular case, this table is not compatible with a $51,2\,kHz$ maximum sampling frequency. To adapt it the constraint $f_{mod} > f_{cod} + \beta - f_s/2$ should be added. This constraint leads to :

$$f_{modop} = \max(\frac{2}{3}.f_{cod}, f_{cod} - \frac{f_s}{4}) \quad , \quad \beta_{max} = min(\frac{f_{cod}}{3}, \frac{f_s}{4}) \tag{13}$$

or by dividing β by f_{rot} the maximum order o_{max} is :

$$o_{max} = \min(\frac{N_{ppt}}{3}, \frac{f_s}{4.f_{rot}}). \tag{14}$$

It means that the demodulation frequency f_{mod} should be increased in order to shift more on the left the red pattern in figure 2. In this case, the needed sampling frequency become f_s.

Rotation frequency (rpm)	1500				3000	6000
Optical encoder resolution	256	512	1024	4096	4096	1024
Optical encoder mean frequency (kHz)	6.4	12.8	25.6	102.4	204.8	102.4
Optimal f_{modop} frequency (kHz)	4.3	8.5	17.0	89.6	192	89.6
Theoretical maximum order	85	170	341	512	256	128

Table 2: Example of demodulator setup to use not more than $f_s = 51.2\,kHz$

These adjustments shows that the sampling frequency will limit the bandwidth in order. It should be kept in mind for comparison that theoretical order correspond to half of the equivalent optical encoder resolution.

4 Comparison with elapsed time

In order to compare both methods, the full resolution optical encoder was recorded using elapsed time technique and high frequency demodulation technique. The frequency f_{mod} is fixed to $90\,kHz$ according to table 2.

The figure 3 compares two results for the two approaches. The instantaneous speed obtained with high frequency demodulation was resampled in angular domain and manually shifted in order to be compared to elapsed time speed. The band β is chosen to be $3.f_{mesh}$ (a lower band gives bad results). The comparison is valid only locally for the first graph. The blue curve corresponds to elapsed time, and the red curve to high frequency demodulation.

The two instantaneous speed curves are superposed. The high frequency method is smoother than elapsed time one due to quantification. The variation in the difference of $\pm 0.06\,Hz$ is not so big be compared to the quantification step of $0.03\,Hz$ for elapsed time method.

The spectrum is also close till order 30. After it seems that an additive noise is present in the high frequency demodulated signal : high peaks are still present but not smallest one. The filtering effect due to the band of $3.f_{mesh}$ (order 69) becomes easily visible near order 100.

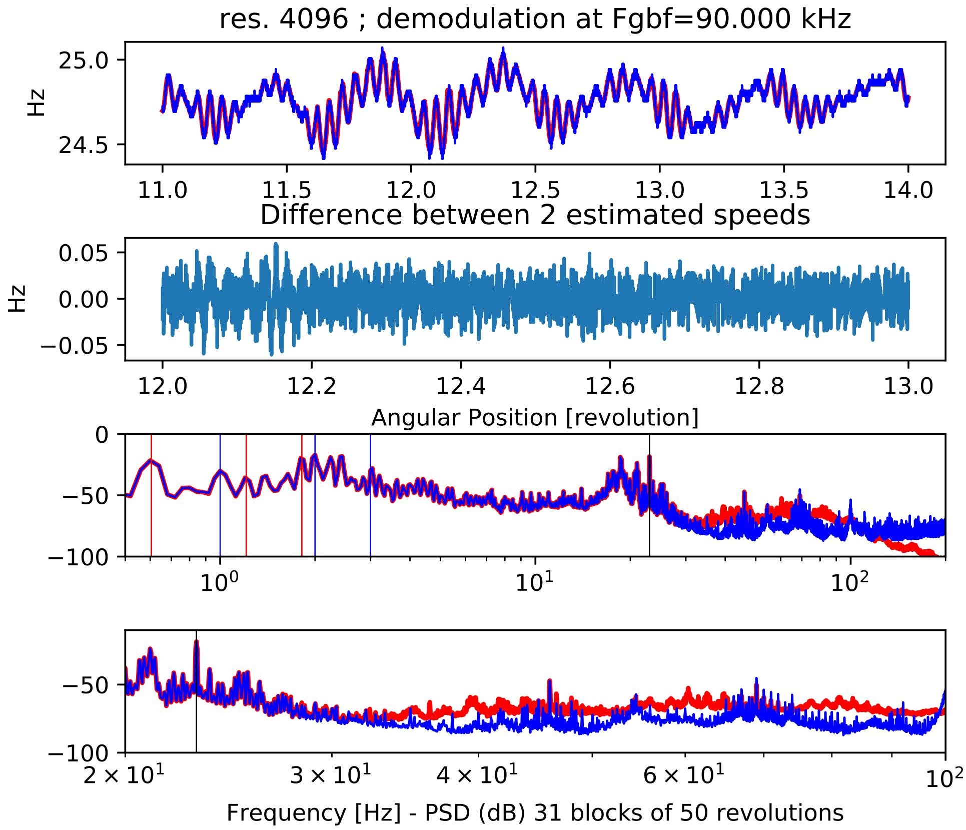


Figure 3: Comparison of Elapsed time and high frequency demodulation Technique.

5 Conclusion

We have proposed a high frequency demodulation technique that performs an analog demodulation prior to acquisition and enable to use the full optical encoder resolution. This technique known in telecommunication as super-heterodyne enables to inspect a restricted bandwidth in high frequency with a low sampling rate. The study of this new technique explains how to choose intermediary frequency for demodulation and some limitations due to the imperfection of electronic components.

A test on a bench to compare this new method with elapsed time shows that the results are similar to the beginning of the spectrum and that an additive noise is present in high frequency. Therefore, this method seems to be interesting when no elapsed time system is present, or it is required to do simultaneous temporal sampling of other signal (acceleration, ...), or, if it is necessary to record very high frequency optical encoder signal.

Acknowledgement

The authors would like to thank Lucien Périsse that design this test bench for us.

References

[1] L. Renaudin, F. Bonnardot, O. Musy, J.B. Doray, and D. Rémond. Natural roller bearing fault detection by angular measurement of true instantaneous angular speed. *Mechanical Systems and Signal Processing*, 24(7):1998 – 2011, 2010. Special Issue: ISMA 2010.

[2] Ming Zhao, Xiaodong Jia, Jing Lin, Yaguo Lei, and Jay Lee. Instantaneous speed jitter detection via encoder signal and its application for the diagnosis of planetary gearbox. *Mechanical Systems and Signal Processing*, 98:16 – 31, 2018.

[3] H. André, F. Girardin, A. Bourdon, J. Antoni, and D. Rémond. Precision of the ias monitoring system based on the elapsed time method in the spectral domain. *Mechanical Systems and Signal Processing*, 44(1):14 – 30, 2014. Special Issue on Instantaneous Angular Speed (IAS) Processing and Angular Applications.

[4] K. Lizoul, H. André, and F. Guillet. Improvement of instantaneous angular speed estimation based on frequency demodulation method under non stationary conditions. In *Proceedings of ISMA2018 and USD2018*, Leuven, 17-19 September 2018.

[5] F. Bonnardot and J. Al Fallah. Non-uniform (re)sampling for optical encoder. In *First World Congress on Condition Monitoring (WCCM)*, London, 13-16 June 2017.

[6] M. El Badaoui and F. Bonnardot. Impact of the non-uniform angular sampling on mechanical signals. *Mechanical Systems and Signal Processing*, 44(1-2):199–210, February 2014.

[7] Heidenhain. *Rotary Encoders brochure*, 2018.

[8] F. Bonnardot, M. El Badaoui, R.B. Randall, J. Danière, and F. Guillet. Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation). *Mechanical Systems and Signal Processing*, 19(4):766 – 785, 2005.

Development of a vibration monitoring strategy based on cyclostationary analysis for the predictive maintenance of helicopter gearbox bearings

V. Camerini L. Macchi F. Champavier
G. Naccarato

HUMS department, Airbus Helicopters, 13725 Marignane Cedex, France

Abstract

The scope of this paper is the development of a fault detection and diagnosis method aimed to helicopter gearbox bearings vibration monitoring in an operational context. Bearings are critical components in the gearbox, and their monitoring allows for failure anticipation capabilities, leading to increased safety and improved maintenance planning. Deploying a monitoring strategy for helicopter gearboxes necessitates the development of a methodology which can provide reliable information under varying operating conditions, dealing with a noisy vibration environment and simultaneously considering acquisition system constraints, such as limited acquisition duration and sampling frequency, and operational needs, such as low rate of false alarms and minimal workload for the analyst. The approach proposed in this paper is based on the cyclostationary signals theory and relies on a two-steps procedure of detection and diagnosis. First, bearing fault detection indicators are devised on a statistical basis, leveraging on the theoretical properties of the envelope method. Then, a diagnosis based on the computation of the averaged cyclic periodogram is performed to assess the damage in the eventuality of an alarm. The developed methodology is validated on real helicopter data collected over about twenty thousand flight hours, including four bearings from different machines for which in-service spalling initiation occurred. The fault detection performance is evaluated on the basis of the achieved false alarm rates and the improvement in fault anticipation with respect to chip detectors, whereas the capability of isolating the fault-related signals using cyclostationary signal separation methods is shown for the diagnosis stage.

1 Introduction

Aircraft operations always pose the problem of guaranteeing at any time a compliant level of safety of the machine, achieved through an adequate maintenance plan, without significantly compromising its availability. In helicopters, the drive train sub-system is responsible for transferring power from the engines to the rotors, and represents a critical sub-system for the machine due to non-redundant load paths and the high variability of the dynamic loads acting on the components [1]. As to ensure aircraft airworthiness, the system needs to be maintained following a prescribed preventive maintenance program, resulting in a burden to operating costs and aircraft availability. Searching for an optimal trade-off between keeping the machine operational and reducing safety risk within acceptable hazard levels calls for making as informed as possible decisions. In the last decades, the helicopter industry worked on implementing technical solutions for increasing safety and reducing maintenance costs by enabling Condition Based Maintenance (CBM) [2, 3]. The effort resulted in the widespread adoption of Health and Usage Monitoring Systems (HUMS). Owing to the mechanical degradation of drive train components often resulting in specific vibration symptoms, and considering the widespread availability of vibration measurement systems, helicopter HUMS mostly rely on vibration analysis as a monitoring mean.

A structured breakdown of the failure mechanisms that may affect a helicopter transmission is given in [4], mostly based on [5, 6]. Failure modes may be divided into gear failure modes, bearing failure modes and shaft failure modes. From an operational point of view, the potential of vibration monitoring in driving maintenance operations toward condition based can probably at most be realized by improving the gear and bearing monitoring procedures. Gearbox inspections are expensive and require long-term grounding of the machine. Therefore, timely detecting an impending mechanical degradation is a great advantage which results in improved maintenance planning and increased machine availability.

In this paper, a two-steps procedure for rolling element bearing fault detection and diagnosis is proposed, with the aim of obtaining a reliable operational procedure able to cope with the monitoring of a fleet of helicopters. First, cyclostationary analysis is recalled as a tool to describe and characterize the characteristic signature of a faulty bearing in section 2. Then, a procedure based on automated statistical bearing fault detection and successive fault diagnosis, and tailored on the specific features of helicopter gearbox vibration environment, is proposed in section 3. The devised strategy is validated in section 4 using data collected from a fleet of operating commercial helicopters. Finally, conclusions are drawn in section 5.

2 Theoretical background

This section is a summary of existing literature on the subject of cyclostationary methods for bearing fault detection, and does not contain original material, except for section 2.4.3.

2.1 Cyclic spectral analysis

Rotating machinery vibration signals have been in the last decades successfully modelled as cyclostationary processes [7–11]. Cyclostationarity is a property characterizing stochastic processes whose statistics vary periodically with respect to some variable (for rotating machinery, typically time or shaft's angular position) [12]. Due to this generality, it is particularly fit to describe rotating machinery signals [13]. The impact forces generated by rolling elements interacting with a local defect on the race are not repeating perfectly periodically due to slippage of the elements in normal operating conditions. Also, the transfer path to the accelerometer varies depending on the relative position of the sensor and the source of the impact. This is the case, for example, for a defect localized on any rotating element in the bearing, where the impact location varies periodically with respect to the transducer position. Such phenomena can be described by their periodic statistics, and therefore the class of cyclostationary signals is suitable to represent the associated excitation. In this paper, the second order cyclostationary descriptors are used to characterize the bearing fault signature. The main quantities of interest when dealing with cyclostationary processes in rotating machinery vibration monitoring are the cyclic spectral correlation (SC) and its normalized version, the cyclic spectral coherence (SCoh). Those quantities are bi-spectral representations, containing information related to the correlation between spectral frequency bands spaced apart by a so-called cyclic frequency α. By considering a signal $y(t)$ recorded in the time T and its Fourier transform $Y_T(f)$, its cyclic spectral correlation can be expressed as:

$$S_{yy}(f,\alpha) = \lim_{T\to\infty} E\left\{Y_T\left(f+\frac{\alpha}{2}\right)Y_T\left(f-\frac{\alpha}{2}\right)^*\right\} \tag{1}$$

whereas the (squared magnitude) cyclic spectral coherence reads:

$$|\gamma_x(f\alpha)|^2 = \frac{|s_{yy}(f,\alpha)|^2}{s_{yy}(f+(\alpha/2))s_{yy}(f-(\alpha/2))}, \tag{2}$$

A practical estimator of the cyclic spectral correlation can be obtained using the averaged cyclic periodogram method [14]. Based on [14], the averaged cyclic periodogram for the N-length discrete sequence $x[n]$ sampled with sampling frequency F_s, computed using K (possibly overlapping) windows $w[n]$ of length N_w can be computed as:

$$\tilde{S}_{xx}^{(N)}(f;\alpha) = \frac{1}{KF_s\|w\|^2}\sum_{k=0}^{K-1} X_{N_w}^{(k)}\left(f+\frac{\alpha}{2}\right)X_{N_w}^{(k)}\left(f-\frac{\alpha}{2}\right)^*, \tag{3}$$

where:

$$X_{N_w}^{(k)}\left(f\pm\frac{\alpha}{2}\right) = \sum_{n=kR}^{kR+N_w-1} w_k[n]x[n]\mathrm{e}^{\pm j\pi wn/F_s}\mathrm{e}^{-\mathrm{j}2\pi fn/F_s} \tag{4}$$

is the DFT of the kth windowed sequence $w_k[n]x[n]\mathrm{e}^{\pm j\pi\alpha n/F_s}$. Practically, the selection of the window length, the window function and the cyclic frequency resolution can be optimized as to minimize the computational time, minimize the cyclic leakage and find the proper trade-off between frequency resolution on the spectral axis f and variance reduction of the estimator [15]. Despite being computationally heavy, the estimator of equation (3) provides reliable results, thanks to its statistical properties, well characterized in [15]. Albeit faster algorithms were developed to estimate the cyclic spectral correlation, e.g. [16], the average cyclic periodogram method still remains a benchmark in terms of estimation accuracy and estimation variance properties.

2.2 Envelope analysis

An important relationship that can be exploited for characterizing a second order cyclostationary process is that connecting the cyclic spectral correlation of the process with its envelope spectrum. It holds from [10] that marginalizing the spectral correlation on the cyclic frequency axis, by integrating out the spectral frequency yields the squared envelope spectrum of the signal. The envelope spectrum has indeed been used in rotating machinery long before the cyclostationary framework was introduced [17]. However, the work in [10] allows to explain the efficiency of the envelope spectrum as an analysis tool for second order cyclostationary processes, framing the technique in the solid theoretical framework of cyclostationary analysis. Other than simplifying the analysis (albeit at the price of losing information on the spectral frequency distribution of the investigated process), the envelope spectrum can be easily estimated from a digitalized realization of the stochastic process by making use of the discrete Hilbert transform and the Fast Fourier Transform, and it is therefore a computationally very convenient quantity. Being equivalent to the integration of the cyclic coherence along the frequency axis, the squared envelope spectrum (SES) as a function of the cyclic frequency α, can be obtained from the N-samples discrete sequence $x[n]$ sampled with sampling frequency F_s as [14]:

$$
\begin{aligned}
\mathrm{ICC}_x^{(N)}(\alpha) &\propto \Big|\frac{1}{N}\sum_{n=0}^{N-1}|x[n]*g[n]|^2\mathrm{e}^{-\mathrm{j}2\pi n\alpha/F_s}\Big|^2 \\
&= \left|\mathrm{DFT}\left\{|x[n]*g[n]|^2\right\}\right|^2 \qquad , \\
&= \mathrm{SES}_x^{(N)}(\alpha)
\end{aligned}
\tag{5}
$$

where the convolution with $g[n]$ accounts for whitening of the signal (necessary to have the power normalization leading to cyclic coherence, in place of cyclic correlation); analytic signal transformation; and band-pass filtering in a band comprised between the frequencies F_1, F_2, normally to be chosen as to filter the signal in a band in which the fault symptoms are prominent with respect to the background vibration and the interfering sources. More recently, the logarithm of the envelope spectrum (LES) for a discrete sequence $x[n]$ with $n=\{1,...,N\}$ was introduced in [18] as:

$$
LES_x[\alpha] = \left|\frac{\sum_{n=0}^{N-1}\log\left(x[n]^2\right)\mathrm{e}^{-2\pi jn\alpha/F_s}}{N}\right|^2, \tag{6}
$$

The LES is an interesting quantity to be considered in an automated detection framework, thanks to its advantageous statistical properties demonstrated in [18].

2.3 Statistical tests for cyclostationarity

The problem of detecting the second order cyclostationarity is formulated as the decision between the two alternative hypotheses:

$$
\begin{aligned}
&\mathrm{H}_0 : \text{"The signal does not contain a CS2 component at the cyclic frequency } \alpha\text{"} \\
&\mathrm{H}_1 : \text{"The signal contains a CS2 component at the cyclic frequency } \alpha.\text{"}
\end{aligned}
\tag{7}
$$

2.3.1 Testing the SES for cyclostationarity

A rigorous statistical test for the presence of a cyclostationary component at frequency α was given in [14] and is based on the cyclic coherence. By exploiting the link between the cyclic coherence and the SES, a practical statistical test on the SES can be obtained, with the advantage of allowing to work on a simpler, faster to compute quantity. Namely, for a discrete signal $x[n]$ of length N, the following result is obtained in [14] by extending the statistical test on the cyclic coherence, and in [19] following a direct analysis of the discrete SES:

$$\text{"Reject } H_0 \text{ if:"} \mathrm{SES}_x^{(N)}(\alpha) \geqslant \frac{\sigma_{x*g}^4}{2N} \frac{F_s}{F_2 - F_1} f(\alpha) \cdot \chi^2_{1-p,2}\text{"}, \tag{8}$$

being p the significance level of the test, σ_{x*g} the standard deviation of the filtered signal $x[n] * g[n]$ and:

$$f(\alpha) = \begin{cases} 1 - |\alpha| / (F_2 - F_1), & |\alpha| < F_2 - F_1 \\ 0 & \text{elsewhere} \end{cases} \tag{9}$$

It is important to underline that the optimality of the test is obtained under the assumption of white noise signal for a healthy component. An analysis of the effects of CS1, CS2 components and colored noise on the SES of the signal is exhaustively performed in [19]. The relevant points are summarized below:

- The effects of a set of M additive multi-harmonic CS1 components of frequencies λ_m, $m = \{1, ..., M\}$ are that of biasing the SES at the difference frequencies $\{\Delta\lambda\} = \{\lambda_m - \lambda_n\}$, $m, n = \{1, ..., M\}$; and that of amplifying the variance in large frequency bands.
- The effect of an additive CS2 component in the signal is that of introducing a bias in the estimator of the SES, which is stronger when the average power of the CS2 carrier is dominating over the background noise.
- The generalization to colored noise implies estimating the variance of the signal at each frequency bin, resulting in a statistical threshold which is no longer a linear function of the frequency.

2.3.2 Testing the LES for cyclostationarity

For a white noise, discrete signal $x[n]$ of length N, the distribution of the LES at a cyclic frequency α is given in [18] as:

$$\frac{\mathrm{LES}_x[\alpha]}{\pi^2/4N} \sim \chi^2_2 \tag{10}$$

Therefore, the LES test for cyclostationarity at significance level p reads:

$$\text{"Reject } H_0 \text{ if:"} \mathrm{LES}_x^{(N)}(\alpha) \geqslant \frac{\pi^2}{4N} \cdot \chi^2_{1-p,2}\text{"}, \tag{11}$$

The LES allows estimating the CS2 components in the signal with the following advantages with respect to the SES:

- The estimator is unbiased by the presence of cyclic components of a frequency different than the considered one.
- Under the white noise assumption, the variance of the estimator is independent from the variance of the noise in the signal.

The second point is more a matter of mathematical rigor for long, whitened noise signals, whereas the first point constitutes an important advantage of the LES when it comes to defining automated tests for the presence of cyclostationary components at a frequency of interest. Additionally, the LES was shown to yield better statistical performance in presence of impulsive noise [20]. This last characteristic is not surprising in light of its being unbiased from CS2 components, considering the existing relation between the CS2 components in the signal and its Kurtosis [21].

2.4 Signal pre-processing

In order to leverage on the optimality of the statistical tests discussed in section 2.3, it is necessary to bring the analyzed signal's statistics as close as possible to the white noise conditions. First, it is a good practice to remove CS1 components from the signal, as they have a biasing effect as explained in section 2.3. The removal of CS1 components can be performed, e.g., through estimation and subtraction. Such an estimation can be performed in different ways, depending on whether the fundamental cycle of interest is known or not. In the case it is not, it can be based on blind estimators, as the linear adaptive enhancer (ALE), or the self-adaptive noise canceller (SANC) and its more efficient frequency domain formulation [22–25]. Generally, blind estimators performance is negatively affected by signal to noise ratio. Moreover, blind filters require a proper parameter tuning which may not be trivial in every case. When the cycle of the signal is known, a popular estimator of the periodic mean is the synchronous average (SA) operator, which is also known as Time Synchronous Average (TSA) due to its original formulation in time domain [26]. In order to obtain the periodic mean in the case of a quasi-cyclostationary signal, the SA must be applied for each of the fundamental cycles which are present in the signal, and then the extracted periodic components need to be summed together [13]. Once the deterministic part of the signal is removed, it is necessary to obtain a flat frequency spectrum for the signal, resembling white noise statistics. In order to do so, there are mainly two strategy: one is selecting a narrow-band frequency region and filter it out; the other is to apply any method to "flatten" the spectrum, such as cepstrum pre-whitening (CPW) [27–29]. On the other hand, the estimation bias resulting from the exogenous CS2 components discussed in section 2.3 cannot be simply corrected for, due to its statistical nature. As a summary, two main steps shall be performed before analyzing the signal, i.e. removal of CS1 components and pre-whitening of the residual. In this work, two techniques were found particularly useful for the scope: the angular domain synchronous average and the cepstrum pre-whitening. The first technique is preferred as the cycles of the main additive deterministic components are known for a given gearbox, whereas the second one is preferred over filtering, as it allows to consider the full-band signal in the analysis, avoiding a further optimization step to select a narrow-band filter which is able of isolating the fault signature (e.g., Spectral Kurtosis [30] is a popular tool that can be used for the scope). Also, if compared to other pre-whitening techniques, the CPW excels for the simplicity of use and the lack of configuration parameters to be properly selected.

2.4.1 Synchronous average removal

Under the assumption of cycloergodicity [13], SA is indeed a practical estimation of the periodic mean of a CS signal, which is its first order cyclostationary part of cycle equal to the fundamental period used for averaging. The equation for the SA of a signal $x(\theta)$ of fundamental cycle Θ reads in angle domain [31]:

$$SA[x(\theta)]_\Theta = \frac{1}{N}\sum_{i=0}^{N-1} x(\theta + i\Theta) \tag{12}$$

Synchronous averaging is thus equivalent to applying a comb filter to the signal [13], which extracts the multiples of the reference harmonic. The number of averages controls the bandwidth of the lobes, the amount of noise rejection and the position of the notches of the filter. In order to remove multiple cycles linked to different harmonic families, equation (12) can be applied multiple times to extract the harmonic family of interest and then subtract it from the original signal. It is worth mentioning that an "order tracking" or "angular resampling" step has to be performed to correct for small speed fluctuations, expressing therefore the measured vibration signal in the angle-domain form of equation (12) [27]. This angular resampling step is typically performed using a synchronization signal acquired from an external measurement system, as e.g. a magnetic pick-up sensor mounted on a reference shaft [31].

2.4.2 Cepstrum pre-whitening

The pre-whitening operation consists in setting a zero value for the whole real cepstrum (except possibly at zero quefrency), then, once transformed back to the frequency domain, the obtained signal is recombined with the phase of the original signal and inverse transformed to time domain [29]. Considering a signal x, its pre-whitened version can be simply computed as [29]:

$$x_{cpw} = IFT\left\{\frac{FT(x)}{|FT(x)|}\right\} \tag{13}$$

2.4.3 Remarks on signal pre-processing

The two steps of signal pre-processing carry with them some hidden difficulties which is worth pointing out. First, it is necessary to observe that the synchronous average removal requires angular resampling of the signal, and when it is performed using the computed order tracking (COT), it involves interpolating the signal in order to obtain its angle-domain values. In [32], a discussion of interpolation methods is given. Interpolating acts as a low-pass filter in the frequency domain. Therefore, it is important to keep into account that any order tracking step has the effect of distorting the signal's spectrum by attenuating the high-frequency components. As a consequence, spectral flattening shall always be performed after order tracking, when envisaging the use of the statistical tests of section 2.3. Furthermore, the cepstrum pre-whitening enhances the sensitivity of the squared envelope results to phase correlation. This last issue is formalized below by using a simple, unit amplitude complex harmonic signal as an example. Consider the following:

$$x(t) = e^{j2\pi\tilde{f}t} \tag{14}$$

then its Fourier transform reads:

$$X(f) = \delta(f - \tilde{f}) \tag{15}$$

being $x(t)$ complex analytic, and noting $\overline{x}(t)$ its complex conjugate, then the envelope spectrum can be computed as:

$$\mathrm{ENV}_x(f) = FT(x(t)\cdot\overline{x}(t)) \tag{16}$$

if a finite-length, discrete signal $y[n] = x[n]\cdot w[n]$, $n = \{1,...,N\}$ is considered (neglecting the sampling step from the notation for simplicity), being $w[n]$ a rectangular, causal observation window of length N, and the Discrete Fourier Transform is used to compute the Fourier transform, it holds:

$$\mathrm{ENV}_y[k] = \frac{1}{N}\sum_{n=0}^{N-1} y[n]\cdot\overline{y}[n]\cdot\exp\left(-j2\pi\frac{nk}{N}\right) \tag{17}$$

then by the convolution theorem, equation (17) can be rewritten as:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} Y[r]\cdot\overline{Y}[r-k] \tag{18}$$

again, for the convolution theorem, the sifting property of the Dirac's delta distribution $\delta(\cdot)$, and considering the Fourier transform of the rectangular window $w[n]$, equation (18) can be expressed as:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} \frac{\sin(\pi(r-\tilde{k}))}{\sin(\pi(r-\tilde{k})/N)} e^{-j\pi(r-\tilde{k})\frac{N-1}{N}} \cdot \frac{\sin(\pi(r-k-\tilde{k}))}{\sin(\pi(r-k-\tilde{k})/N)} e^{j\pi(r-k-\tilde{k})\frac{N-1}{N}} \tag{19}$$

where $\tilde{k}$ denotes the discrete frequency index corresponding to $\tilde{f}$. After the cepstrum pre-whitening operation of equation (13), only the phase terms are left from the DFT of the original signal and equation (19) becomes:

$$\mathrm{ENV}_y[k] = \sum_{r=k}^{N/2} e^{-j\pi(r-\tilde{k})\frac{N-1}{N}} \cdot e^{j\pi(r-k-\tilde{k})\frac{N-1}{N}} = e^{-j\pi k\frac{N-1}{N}}\left(\frac{N}{2}-k\right) \tag{20}$$

From equation (20), it can be observed that the phase correlation from residual periodic components has a harmful effect on the classical envelope spectrum of the pre-whitened signal, which is amplified when the amplitude spectrum of the signal is equalized to the unit value.

2.5 Bearing fault signature

Healthy bearings vibration does not typically bring a significant contribution to the vibration generated by a helicopter gearbox. On the other hand, a defective bearing generates a characteristic vibration signature, characterized by repeated impacts occurring each time that a bearing element contacts the defective surfaces [5, 33–35]. Typically, four characteristic frequencies can be identified: ball pass frequencies on the outer and inner races (respectively BPFO and BPFI), typically linked to localized defects on one of the races; fundamental train frequency (FTF), generally linked to cage defects; and ball spin frequency (BSF), normally related to localized defects on the rolling elements surface. By indicating with f_i and f_e respectively the inner and outer race rotation frequency, with N_b the number of rolling elements in the bearing, by α_0 the initial contact angle, by d the rolling element diameter and by D the bearing pitch diameter, these characteristic frequencies read [36]:

$$\begin{aligned}
BPFI &= \frac{N_b\,|f_e - f_i|}{2}\left(1 + \frac{d}{D}\cos(\alpha_0)\right) \\
BPFO &= \frac{N_b\,|f_e - f_i|}{2}\left(1 - \frac{d}{D}\cos(\alpha_0)\right) \\
BSF &= \frac{D\,|f_e - f_i|}{2d}\left(1 - (\frac{d}{D}\cos(\alpha_0))^2\right) \\
FTF &= \frac{1}{2}\left(f_e\left(1 + \frac{d}{D}\cos(\alpha_0)\right) + f_i\left(1 - \frac{d}{D}\cos(\alpha_0)\right)\right)
\end{aligned} \tag{21}$$

Letting $h_j(t)$ be the impulse response to a single impact measured by the sensor located at position j, $q(t)$ the periodic modulation owing to load distribution (or periodic changes in the loading conditions, or sensor orientation/position with respect to the impact point) [37, 38], and letting T be the fundamental impact periodicity (which can be computed by inverting the frequency of interest from equation (21)); then the measured response $x_j(t)$ related to the defective bearing was given in [14, 39–41] as:

$$x_j(t) = \sum_{i=-\infty}^{+\infty} h_j(t - iT - \tau_i)\,q(iT)A_i + n_j(t), \tag{22}$$

where $n_j(t)$ includes the additive background noise and all eventual interference sources, the subscript i indicates the ith impact, τ_i represents the mentioned uncertainty on pulse arrival time and A_i the random amplitude of the impact. Both the variables are modeled in [14] as mutually independent, white, stationary random sequences with respectively zero and unity mean. Those idealized assumptions allow, according to the literature, to gain sufficient insight into the described phenomenon. The fault signature appears as a pseudo-periodic excitation consisting of pulses which are separated by a period close to that of the fault frequency, but affected by a small, random variation typically of the order of one percent of the fundamental period. Such slight fluctuations results practically in destroying the discrete, harmonic structure that would arise if the random fluctuations were neglected as in [34], giving raise to an essentially random vibration signal in the frequency range of interest [14]. The main difference between the model found in [34] and that of equation (22) is that in the latter, the harmonic structure produced by the fault-related impacts rapidly turns into a random signal. As a consequence, the bearing fault signature in the spectrum is likely to be localized in the low-frequency region, and therefore subject to masking from the background noise and from other possibly existing interference sources, as e.g. gear mesh harmonics. Hence, the model can explain the reason for which classical spectral analysis may fail in detecting rolling-element bearing faults, making this signal representation closer to the reality of the phenomenon. A practical solution to this issue resides in making use of the second-order cyclostationary tools presented in section 2.1 in order to isolate the bearing vibration signature from the rest of the measured signal. The cyclostationary approach was actually shown very successful for bearing diagnostic problems in several works, as in [10, 14, 19, 21, 33, 41]. From equation (1), adopting the proper normalization to obtain the cyclic power spectrum from the spectral correlation, it can be shown that for the signal model of equation (22), it holds [10]:

$$S_{x_j}(f,\alpha) \simeq \frac{1}{T} H_j\left(f+\frac{\alpha}{2}\right) H_j\left(f-\frac{\alpha}{2}\right)^* \left(\Phi(\alpha)\left(1+\sigma_A^2\right) - \Phi\left(f+\frac{\alpha}{2}\right)\Phi\left(f-\frac{\alpha}{2}\right)^*\right) \\ \times \sum_{k,l=-\infty}^{+\infty} Q_l \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right] + \delta[\alpha] S_n(f) \quad , \tag{23}$$

in which $\Phi(f)$ stands for the Fourier transform of the probability density function of the random variable τ associated to the impacts jitter, P for the load variation characteristic period, $H_j(f)$ for the transfer function from the impact point to the measurement location j, obtained as the Fourier transform of the impulse response function $h_j(f)$, σ_A the standard deviation of the random variable A representing the random impact amplitude, and the Q_l are coefficients of the Fourier transform of the modulating function $q(t)$. The weak harmonic contribution was neglected, being it highly attenuated in the high frequency region as an effect of the random jitter of the impact times. From equation (23), the discrete structure of the bearing signature is finally evident in the cyclic power spectrum plane, with continuously distributed values along the spectral frequency lines appearing at multiples of the fundamental impact frequency (along axis α). Also, the values are higher for those spectral frequency bands where the fault signature is dominating. It follows for the cyclic coherence [14]:

$$\left|\gamma_{x_j}(f,\alpha)\right|^2 \simeq \left|\frac{\mathrm{SNR}(f)}{1+\mathrm{SNR}(f)}\right|^2 |\Phi(\alpha)|^2 \sum_{k,l=-\infty}^{+\infty} \left|\frac{Q_l}{Q_0}\right|^2 \delta\left[\alpha - \frac{k}{T} - \frac{l}{P}\right], \tag{24}$$

where $SNR(f)$ represents the signal-to-noise ratio of the fault. Consequently, as a function of the spectral frequency f, equation (24) shows an increased coherence for increasing amplitude modulation randomness, impact frequency and load modulation intensity, whereas as a function of the cyclic frequency α, it shows a discrete structure consisting of harmonics of the fault signature separated by the characteristic impact frequency, with decreasing amplitude depending on the low-pass filtering function $\Phi(\alpha)$. Therefore, the highest the multiple of the fundamental impact frequency, the lower the intensity of the observed bearing signature in the cyclic coherence. Equation (3), along with the proper normalization, can be used to estimate the quantity appearing at the LHS of equation (24), yielding an efficient diagnostic representation able of highlighting the bearing fault signature according to the structure of equation (24).

3 Proposed monitoring procedure

Operationally, it is desirable to control the risk of false alarms from the health monitoring system, in addition to providing the earliest possible warning. In order to achieve those targets, a two-steps procedure is proposed in this work. First, statistical indicators leveraging on the cyclostationary theory are designed in order to attain a specified false alarm rate. Secondly, a diagnostic step based on the analysis of the cyclic coherence is taken each time that an alarm is raised, in order to confirm that the threshold exceedance is actually due to a mechanical defect. The importance of the first step of the procedure is that of providing an easy-to-read scalar indicator, with known statistical behavior, which can be employed to guarantee a given false alarm rate as low as not to overload the analysts, maintaining contextually an acceptable detection performance. The second step implies confirming the alarms raised in the detection phase before performing any maintenance, as a measure to avoid unnecessary grounding of the concerned helicopters.

3.1 Fault detection stage

With the aim of deriving statistically reliable monitoring indicators, the cyclostationary signal theory is adopted in this work within the frame of a procedure similar to that proposed in [42]. The envelope spectrum is here computed through the methods described in section 2.2. Both SES and LES can be considered. The necessity of an automated bearing monitoring gives raise to the following challenges:

1. The false alarm rate shall be kept under control in order to avoid unnecessary grounding of the machine;
2. The actual fault frequency cannot be accurately predicted using the simplified kinematics relations;
3. Interfering, exogenous components may mask the bearing fault signature.

As to cope with the first point, the statistical tests presented in section 2.3 for LES and SES can be exploited. First, a signal pre-whitening step consisting of the removal of the periodic (CS1) components through synchronous averaging followed by cepstrum pre-whitening for spectral flattening is performed. This allows to get rid of interferring CS1 sources and to bring the signal's statistics closer to those of white noise. A statistical threshold can then be derived from the white-noise envelope spectrum statistics as in [18, 43]. For a number of bearing fault harmonics to be configured, a narrow-band frequency range around the fault frequencies of an extent to be configured, can be defined. This is done in order to allow some margin in considering eventual shifts of the fault frequency from its predicted nominal value, addressing the second problem in automatizing the algorithm. At the same time, if the considered cyclic frequency range is too wide, there is the risk of exogenous CS2 components leaking into the analysis band, leading to incorrect diagnosis. This issue has to be carefully addressed when tuning the algorithm's parameters. Any number of fault harmonics can be considered in the algorithm. However, according to equation (24), best results are obtained for the low bearing harmonics. For each defined range, the values of the SES or LES are compared to the statistical thresholds of equations (8) and (11). If any statistically significant value is present, a fault detection alarm is raised. The advantages of the adopted procedure are two-fold: on one hand, the envelope spectrum is computed in a computationally efficient way; on the other hand, the computation of the theoretical threshold after pre-whitening according to [18, 43] provides solid grounds for statistical testing. For a given signal to be processed, the algorithm can be summarized in the following steps:

1. Define the desired false alarm rate $P\tilde{F}A$ for the indicator according to the operational needs;
2. Calculate the fault frequency of interest F_F (according to equation (21)) in units of the sampling frequency F_s;
3. Set the number of fault harmonics N_h to be monitored;
4. Set a tolerance band ψ as a percentage of the fault frequency of interest, in order to account for the uncertainty on the actual fault frequency;
5. Remove known CS1 components using synchronous average removal;
6. Apply spectral flattening using cepstrum pre-whitening;
7. Compute the full-band SES/LES of the pre-whitened signal;
8. Compute the statistical threshold p according to the defined desired false alarm rate, based on the white noise assumption;
9. For each considered fault harmonic: find the maximum value of SES/LES in the defined tolerance band;
10. Compute the indicator value as the mean of the statistically significant values with respect to the defined threshold (if no significant value is found for any of the considered harmonics of the fault, the indicator value is set to zero).

The expected false alarm rate $P\tilde{F}A$ corresponds to the probability of one value of the (squared or logarithmic) envelope spectrum within the considered range being higher than its statistical threshold computed through equations (8) and (11). Therefore, to attain a desired level for $P\tilde{F}A$, it is necessary to calculate the significance level p as to satisfy:

$$P\tilde{F}A = 1 - p^{\sum_{h=1}^{N_h} r_h} \tag{25}$$

with N_h being the number of the considered fault harmonics and:

$$r_h = \lceil h \frac{\psi}{100} \frac{F_F}{F_s} N \rceil \tag{26}$$

where N is the discrete signal length and F_s its sampling frequency expressed in inverse units of the sampling step. The value of p obtained from equation (25) allows for calculating the statistical thresholds of equations (8) and (11) so that the probability of having one exceedance of the envelope spectrum in the computational range is equal to $P\tilde{F}A$. It is important to point out that where the false alarm rate can be kept under control, the detection performance cannot be predicted a priori. Typically, the higher the tolerance band ψF_f, the lower the detection performance for a fixed $P\tilde{F}A$; the higher the admissible alarm rate $P\tilde{F}A$, the higher the detection performance that can be expected.

3.2 Fault diagnosis stage

Each time an alarm is triggered, a fault diagnosis step is performed by the analyst. It consists of computing the cyclic spectral coherence around the fault frequencies for which the detection algorithm raised an alert and visually assessing the existence of CS2 components compatible with the expected signature from a bearing fault. The spectral coherence is estimated according to equation (3), where the number of averages and the window length are set according to the guidelines discussed in [15]. Within this assessment step, the analyst can additionally assess the presence of diagnostic side-bands carrying supplementary information on the nature of the fault.

4 Results

In this section, the proposed monitoring strategy is applied on in-service helicopter HUMS data in order to assess its performance in terms of reliability and detection. LES and SES indicators, along with different pre-processing treatments are compared, stressing the importance of properly pre-whitening the signal before carrying on with the analysis.

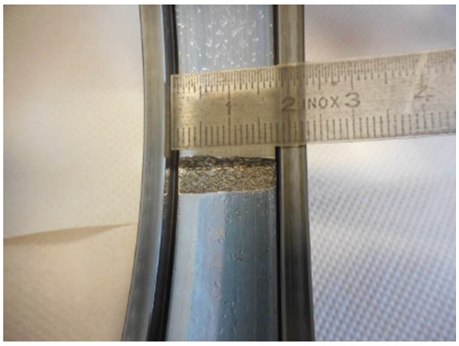
4.1 Data description

In order to validate the proposed procedure, a comprehensive data-set consisting of vibration data recorded from fourteen machines over about twenty-thousand flight hours (FH) is considered. The data-set includes four bearing in-service degradation cases that were detected by the HUMS: two of them concern roller bearings, and the other two concern ball bearings. The acquisitions were performed in various operating conditions, involving different regimes for the rotational speed of the rotor and for the transmitted torque from the engines. Main gearbox (MGB) and accessory gearbox (AGB) acquisitions comprise signals from seven accelerometers and two keyphasor signals. The two keyphasor signals provide respectively one pulse per revolution of the main rotor and of the tail rotor shaft. Accelerometers are typically mounted on the gearbox casing, close to the monitored components. Generally, acquisitions are divided in groups. Each acquisition group is launched when specific flight conditions are matched and consists of a synchronized acquisition from a set of sensors, performed with a configured sampling frequency for a configured duration. The available keyphasor signals are always sampled with the same sampling frequency of the accelerometer signals. Signals were sampled synchronously from all the MGB accelerometers with a sampling frequency of 50 kHz, for one second duration. The four documented bearing fault cases occurred on different machines during the monitoring period. The faults were anticipated by the HUMS in the spalling initiation phase, allowing for timely maintenance. Table 1 summarizes the four selected fault cases, whereas figure 1 show the inspected bearings after component removal. All the cases involve outer race spalling which occurred whether on a roller or on a ball bearings. The HUMS, through the deployed monitoring strategy, triggered regular inspections of the chip detectors and allowed in each case to anticipate the chip warning coming from the oil metal chips detectors. Additionally, for the fault case 4, HUMS had gained enough confidence to trigger the removal without the need of waiting for the metal particles in the chip detector to be out of criteria. The geometrical parameters of the concerned bearings are not reported for proprietary reasons.

Table 1 – Fleet selected bearing fault cases summary

Fault case ID	Machine ID	Damaged bearing
1	1	Roller bearing (bearing 1)
2	2	Roller bearing (bearing 1)
3	3	Ball bearing (bearing 2)
4	3	Ball bearing (bearing 2)

(a)



(b)

(c)

(d)

Figure 1 – Damaged bearings after inspection. Outer race presenting spalled area at the time of removal – a) Bearing fault case 1 ; b) Bearing fault case 2; c) Bearing fault case 3 ; d) Bearing fault case 4.

4.2 Fault detection performance

The bearing monitoring strategy proposed in section 3 is applied on the operational data of this section. Bearing anomaly indicators are computed routinely on all the fleet data, and then the detected exceedances are analyzed using cyclostationary analysis in order to complete the diagnosis. Although bearing anomaly indicators are computed for each monitored bearing in the gearbox, the results here presented are restricted to those for which the faults were observed, allowing to validate both detection and the reliability performance. From the bearing geometrical properties, the theoretical characteristic frequencies could be computed according to equation (21). Table 2 reports the computed, nominal bearing defect frequencies for the considered bearings 1 and 2, expressed in orders of the rotational speed of the shaft to which they are attached. According to the nominal design parameters, the fault frequencies are very close (for the outer race defect frequency, the difference between bearing 1 and 2 is less than two percent), posing a challenge in discriminating which one is the faulty bearing in the event of a detection. Both SES-based and LES-based indicators are evaluated, computing the envelope spectra on the pre-whitened signal after a synchronous average removal step and cepstrum pre-whitening (SES-CPW and LES-CPW indicators). Performance on the signal after the synchronous average removal step only are also reported for comparison (SES-SA and LES-SA indicators). The indicators were configured such as to allow to separate the two fault frequencies, but allowing some slippage through one-percent width analysis bands. Only the first harmonic of the outer race fault frequency was considered. With reference to section 3, care was taken when performing the OT steps in the SA removal and when resampling the signal to order domain before the cepstum pre-whitening.

Table 2 – Bearing 1 and 2 characteristic fault frequencies in order domain.

Parameter	Description [dimension]	Bearing 1 Value	Bearing 2 Value
BSF	Roller Spin Frequency [Hz]	3.64	3.67
FTF	Fundamental Train Frequency [Hz]	0.44	0.45
$BPFO$	Ball Pass Frequency Outer race [Hz]	9.69	9.85
$BPFI$	Ball Pass Frequency Inner race [Hz]	12.31	12.15

Figure 2 shows the actual false alarm rate against the expected one for the SES and LES indicators for bearing 1 and bearing 2 on the whole collected healthy fleet data. It can be seen that the actual alarm rates for both the bearings do not match well the expected results in the case of the SA indicators. This is attributable to two main factors: first, the SA procedures do not remove all of the periodic components originally present in the signal, leaving some residual, biasing CS1 component; secondly, the spectrum of the measured vibration is far from resembling white noise, leading to the statistical thresholds computed using equations (8) and (11) being inaccurate. On the other hand, figures 2c and 2d show better results for the indicators computed after the cepstrum pre-whitening step. From the distribution check it can be seen that the LES-based indicator provides very accurate results for both the considered bearings: the actual false alarm rate agrees very well with the predicted one. Conversely, despite the spectral flattening and the SA removal, the SES indicator yields higher false alarm rate than expected. This fact might be explained by the SES statistics being affected by exogenous CS2 components, differently from those of the LES. In order to evaluate the detection performance of the devised statistical indicators, an anticipation over removal metrics is introduced. For each fault case, the removal anticipation (RA) achieved by HUMS alert and expressed in acquisitions number, was computed as a function of the actual false alarm rate on the fleet. Typically, the removal is triggered by an inspection of the particles captured in the chip detector and matching some criteria on size and composition. The RA figure allows to judge the achievable trade-off between detection capability and global false alarm rate performance of an indicator. Figures 3a and 3b show the results respectively for the fault cases 1 and 2, for the SES-SA and LES-SA outer race fault indicators, whereas figures 4a and 4b report the same results for the SES-CPW and LES-CPW outer race fault indicators. Concerning the fault cases 3 and 4, related to the ball bearing, the RA is constant for each considered indicator, despite pre-processing differences and equal respectively to 66 acquisitions and 34 acquisitions. This depends actually on two distinct facts: first, in the considered dataset, acquisitions for machine 3 begins already in a relatively advanced bearing degradation stage, where all indicators detect very clearly. Secondly, the second degradation produced very early, strong CS2 symptoms which were as well detectable in a robust manner from all of the considered indicators. Also, for the second ball bearing degradation (fault case 4), the HUMS alerting system was already deployed, and guaranteed the detection of the incipient bearing degradation, along with an optimized, planned maintenance intervention. Therefore, the time to removal metrics is not representative of the anticipation over the chip detector alert for case 4. For the roller bearing degradation cases, evidently the SES-CPW and LES-CPW achieve a better anticipation over removal for a given false alarm rate, implying their better performance in terms of early detection with respect to the SES-SA and LES-SA indicators in both the cases. The comparison between LES-CPW and SES-CPW shows that they perform very similarly for the first fault case (figure 4a), whereas the SES-CPW indicator shows better detection performance with respect to the LES-CPW in the second fault case (figure 4b).

4.3 Fault diagnosis stage

In this section, the diagnostic charts based on the cyclic coherence are shown for each detected bearing degradation. The defined operational procedure only requires to compute such quantities when an alert is raised, so to confirm the actual occurrence of a mechanical degradation. In figures 5 to 8, for fault cases respectively 1 to 4, the diagnostic charts and the associated LES and SES spectra, together with their statistical threshold computed for the 0.1 percent significance level according to equations (8) and (11), are shown in three conditions: before the beginning of the bearing degradation, during the bearing degradation and after gearbox replacement. In each case, the spalling manifests as a high-frequency excitation, at a cyclic order which is slightly different than that predicted from the theoretical calculations. Consequently, the actual loading conditions encountered in operations have an impact on the exact determination of the fault frequency, owing to the simplified kinematics assumptions being inadequate to describe the bearing dynamic behavior. By comparing figure 5 to figure 6, it can be seen that the symptoms of the spalling appears much more evident for the second fault case (compare, e.g., the statistical threshold to the value of the emerging peak in the two cases). Coherently, figures 3 and 4 show that the second fault case is predicted with a higher anticipation time from both the SES and LES indicators. This can be explained by looking at figure 1. It can be noticed from figure 1a and figure 1b that the shape of the surface degradation is consistently different in the two cases. In fault case number 2, the spalling area extends across the full span of the race, creating a slot. Conversely, in fault case number 1 the spalling area is restricted to part of the width of the race. It can be expected that for the second degradation

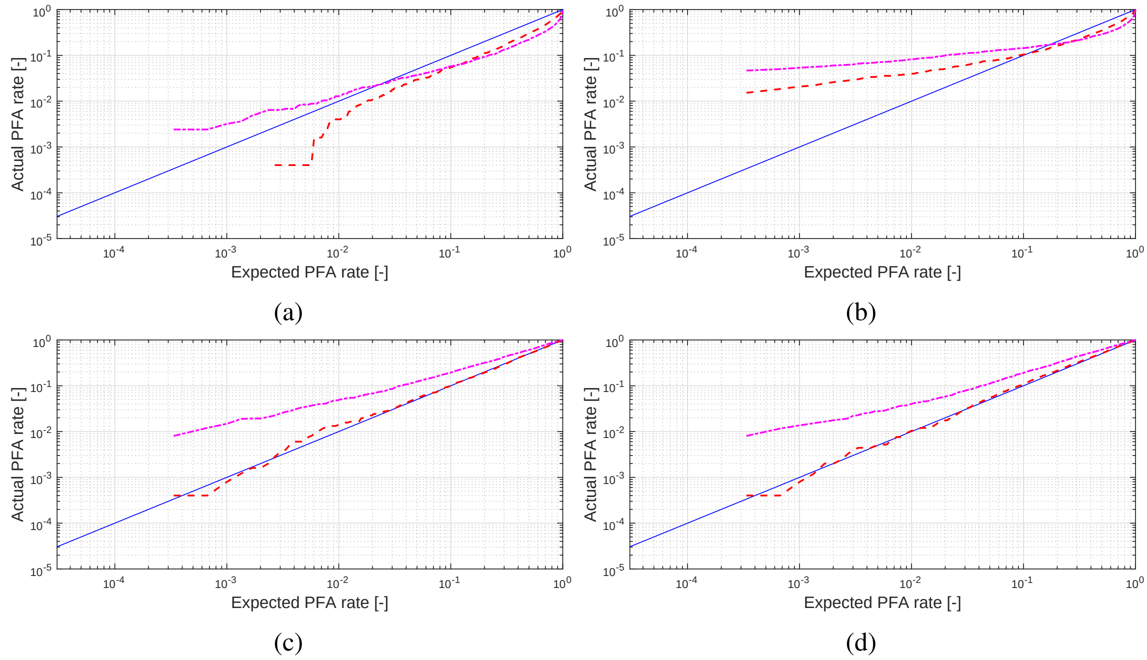


Figure 2 – Healthy fleet data, actual vs. expected probability of false alarm rate of the SES and LES outer race fault detection indicators for: a) Bearing 1, SA indicators; b) Bearing 2, SA indicators; c) Bearing 1, CPW indicators; d) Bearing 2, CPW indicators. Blue: theoretical relation; red dashed: LES indicator; magenta dot-dashed: SES indicator.

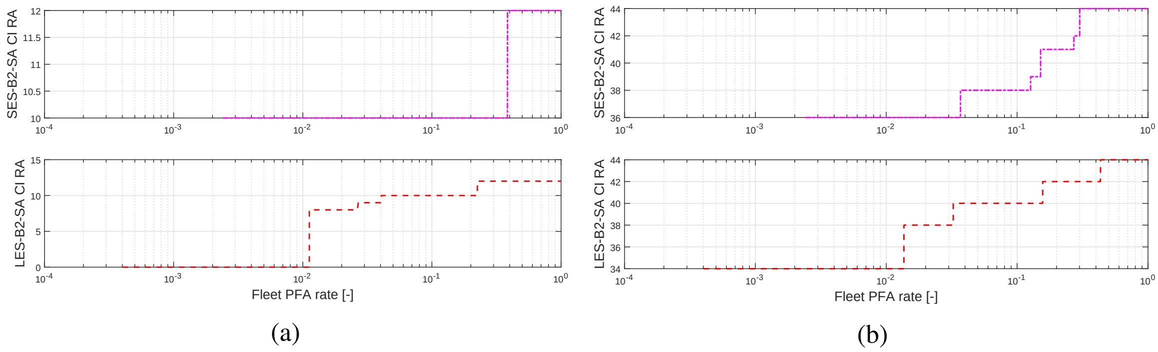


Figure 3 – Removal anticipation (RA) time in acquisitions vs. fleet false alarm rate for SES-SA (upper row, magenta dot-dashed) ad LES-SA (lower row, red dashed) indicators – a) Fault case number 1; b) Fault case number 2.

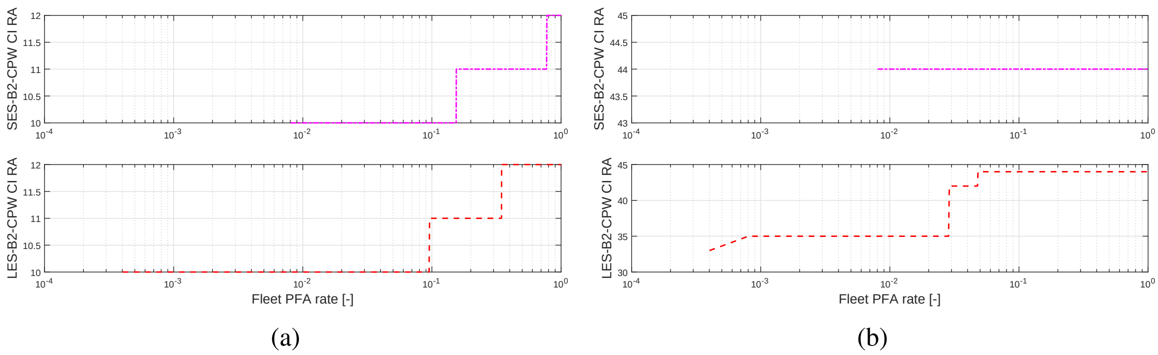


Figure 4 – Removal anticipation (RA) time in acquisitions vs. fleet false alarm rate for SES-CPW (upper row, magenta dot-dashed) ad LES-CPW (lower row, red dashed) indicators – a) Fault case number 1; b) Fault case number 2.

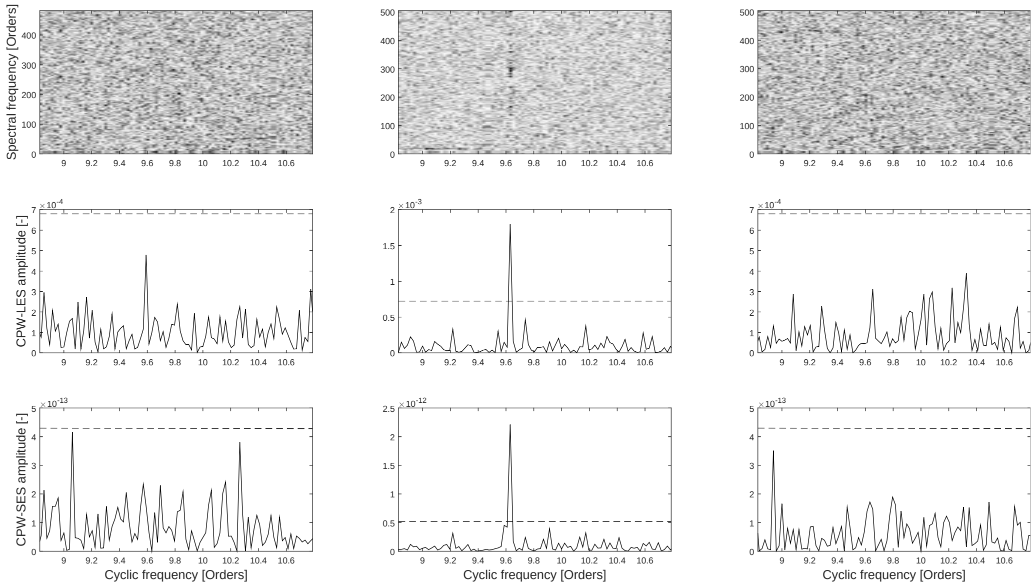


Figure 5 – Diagnostic charts for fault case 1. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

case, impacts will occur, exciting the bearing resonances and resulting in stronger CS2 symptoms with respect to those produced in case number 1, where the quasi-periodic change in the dynamic response is more likely due to the change of the load distribution within the contact line of the rolling elements rolling over the defective surface. As a matter of fact, the detection performance depends, among the other factors, from the evolution of the mechanical degradation. This does not appear to be the case for the ball bearing degradation, where the symptoms are of relatively comparable magnitude, despite the case number 3 presenting a more advanced degradation at the time of removal (figure 1). The fact that the contact taking place in ball bearings between the races and the elements can better be described as a point contact interaction could explain these results. In fact, any geometry of the degradation would almost surely provoke mechanical impacts between the elements and the race. The analysis of the diagnostic charts is an important step to confirm the alarms: the bearing fault signature presents itself as a clear cyclic excitation localized in the high spectral frequency band, and allows the analysts to reliably confirm whether the fault detection alert is effectively related to a mechanical fault. In fact, by observing the cyclic spectral coherence representation, it is possible in an operational scenario to rule out both the occurrence of false alarms and those false indications coming from corrupted measurements.

4.4 Impact of the operating conditions

In order to assess the sensitivity of the bearing monitoring indicators to the different operating conditions, the correlation of the indicator values to the torque and rotor speed values was studied. Figures 9 and 10 show such correlation for the LES-CPW outer race indicators related respectively to the roller (bearing 1) and to the ball bearing (bearing 2). Similar results were obtained for the other indicators, therefore only figures 9 and 10 were reported for brevity. As indicator value, the magnitude of the detected peak in the envelope spectrum was reported. This is a consistent indication, being the LES statistical distribution independent from the cyclic frequency. From the results, it can be observed that there is no significant correlation between the values of the indicator and the contextual parameter, whenever the indicator takes its nominal values. However, the higher values related to degraded conditions are mostly localized in correspondence of the high-torque region, for a rotor speed of around 97 percent of the nominal speed. This may indicate a slight sensitivity of the indicators to the operating conditions. However, it has to be pointed out that the majority of the recorded acquisitions occurred in the low rotor speed, high torque conditions. Being the acquisitions density in the region higher than in the rest of the operating spectrum, the likelihood of acquiring in those conditions during the progression of the fault is also higher. Finally, relatively high values of the indicator during the fault progression can be observed also in other regions of the operating spectrum, confirming its detection robustness to the different flying regimes of the helicopter.

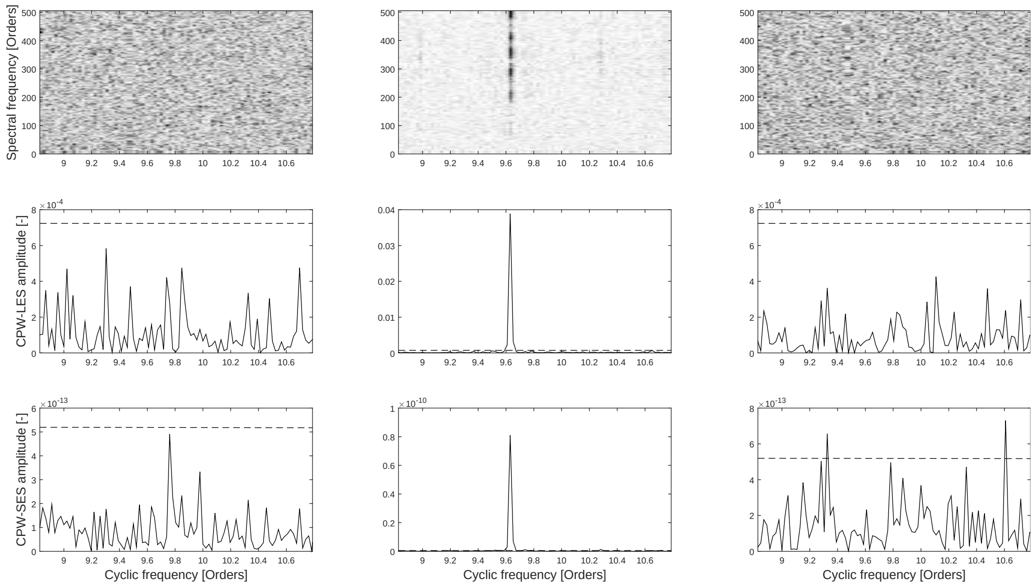


Figure 6 – Diagnostic charts for fault case 2. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

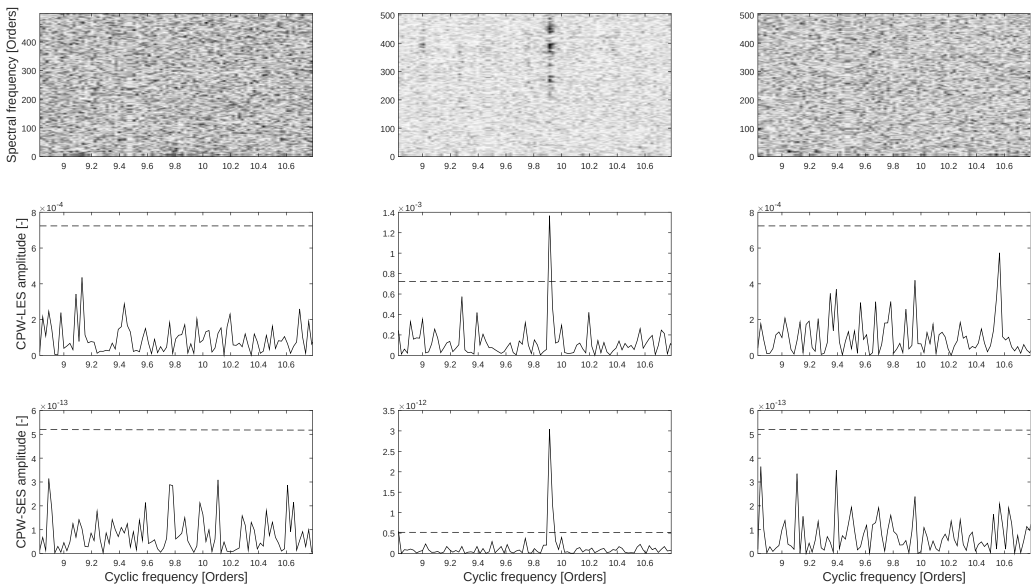


Figure 7 – Diagnostic charts for fault case 3. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

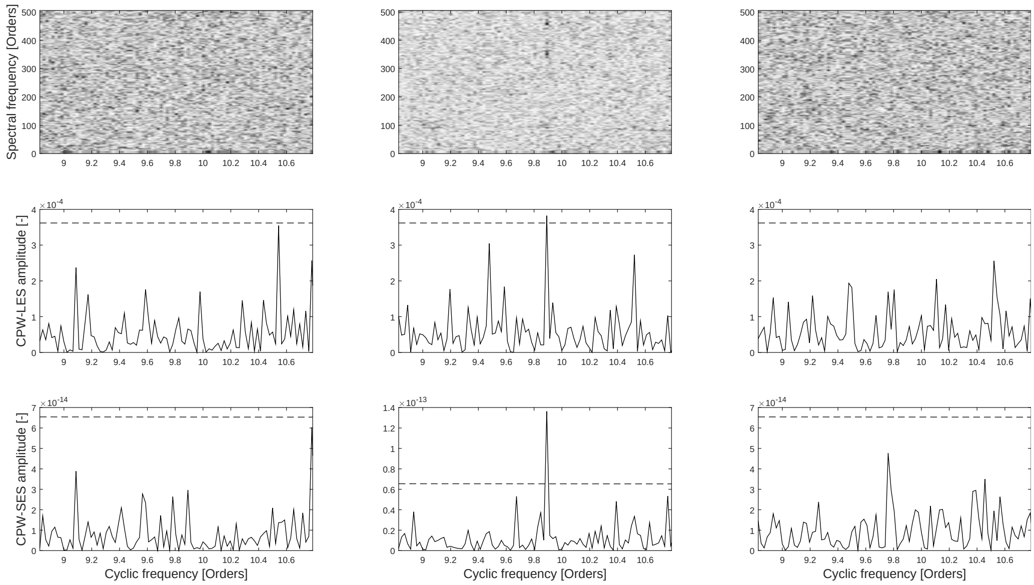


Figure 8 – Diagnostic charts for fault case 4. Upper row: cyclic coherence (darker areas corresponds to higher values); middle row: LES-CPW spectra; lower row: SES-CPW spectra. First column: healthy bearing; second column: detected outer race spalling; third column: gearbox replaced. Dashed line: 99.9 percentile threshold.

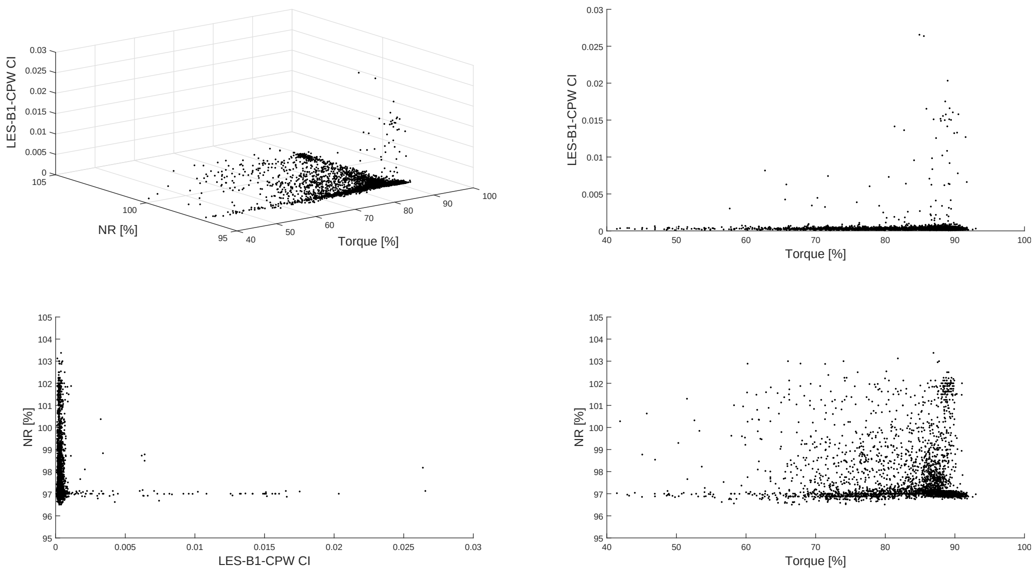


Figure 9 – Impact of the operating conditions on the bearing 1 outer race fault LES-CPW indicator values. Top-left: joint distribution of torque, rotor speed and indicator values; top-right: indicator values vs. torque; bottom-left: rotor speed vs. indicator values; bottom-right: rotor speed vs torque.

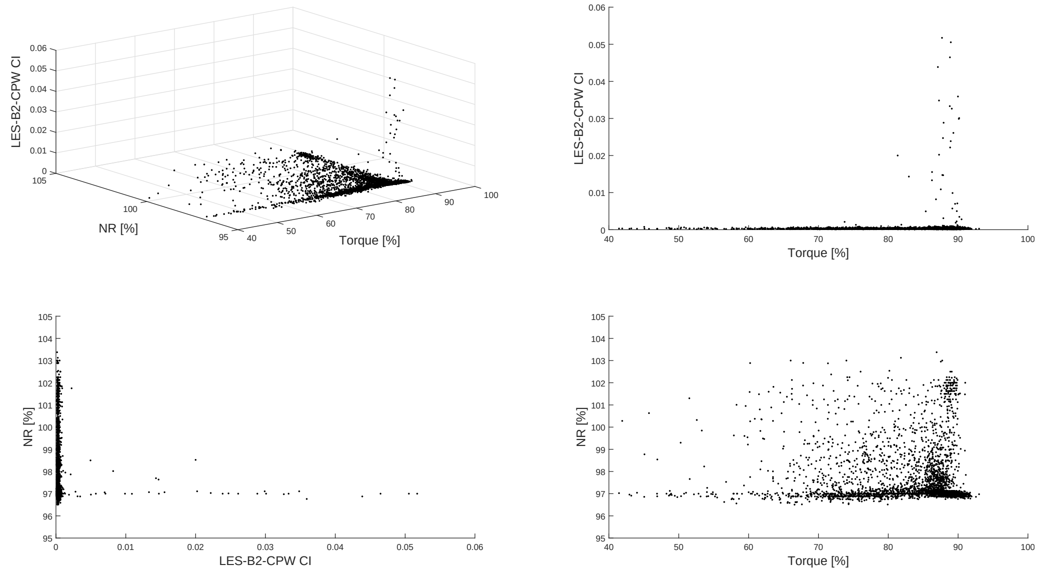


Figure 10 – Impact of the operating conditions on the bearing 2 outer race fault LES-CPW indicator values. Top-left: joint distribution of torque, rotor speed and indicator values; top-right: indicator values vs. torque; bottom-left: rotor speed vs. indicator values; bottom-right: rotor speed vs torque.

5 Conclusions

A predictive maintenance strategy for the monitoring of rolling element bearings in the context of helicopter operations was developed. The proposed methodology relies on a two-step fault detection and diagnosis process. The first step consists of devising reliable statistical indicators, allowing to attain a given false alarm rate. The second step consists of confirming eventual alarms through refined diagnostic analysis based on the estimation of the cyclic spectral coherence of the concerned signals. The devised procedure was validated on comprehensive, in-service helicopter fleet data set, comprising high-frequency acquisitions from fourteen machines flying according to several different profiles. Statistical indicators based on the logarithm envelope spectrum and on the squared envelope spectrum were compared in terms of both reliability in providing a given false alarm rate and ability to promptly anticipate four in-service bearing degradation cases. Two pre-processing algorithms based on synchronous average removal and cepstrum pre-whitening were considered, and some of their properties investigated. It was shown that indicators based on the logarithm envelope allow for a very fine tuning of the desired false alarm rate, together with providing acceptable detection performance in the considered cases. Conversely, squared envelope based indicators proved less reliable in actual operations. The results were shown to be consistent across the full range of considered operating conditions. The diagnostic step was shown to be able of accurately disclosing the faulty bearing signature, proving to be an effective discrimination mean to avoid unnecessary grounding of the machine in the occurrence of a HUMS alarm. At the same time, the burden of performing the diagnosis is restricted to those cases in which an alarm from the detection step actually occurs. The impact of different mechanical degradation shapes was also considered: it was shown that for the roller bearing outer race degradation cases, the fault shape has a significant impact on the HUMS detection performance. The developed approach allowed to deploy a statistically efficient, operationally valid procedure to monitor the rolling element bearings in helicopter mechanical transmissions within an in-service context, characterized by harsh mechanical environment, acquisition constraints and multiple different operating conditions. Remarkably, it could be used to deploy an effective, semi-automated monitoring for helicopter bearings which guarantees an effective condition based maintenance of the monitored components.

References

[1] A.R.S. Bramwell, D. Balmford, and G. Done. *Bramwell's helicopter dynamics*. Butterworth-Heinemann, 2001.

[2] A. Bayoumi et al. “Conditioned-Based Maintenance at USC-Part I: Integration of Maintenance Management Systems and Health Monitoring Systems through Historical Data Investigation”. In: *Proceedings of the AHS International Specialists Meeting on Condition Based Maintenance, Huntsville, AL*. 2008.

[3] B.D. Larder and M.W. Davis. “HUMS Condition Based Maintenance Credit Validation”. In: *American Helicopter Society 63rd Annual Forum*. 2007.

[4] B.D. Forrester. “Advanced vibration analysis techniques for fault detection and diagnosis in geared transmission systems”. PhD thesis. 1996.

[5] I. Howard. *A Review of Rolling Element Bearing Vibration Detection, Diagnosis and Prognosis*. Tech. rep. Defence Science and Technology Organization Canberra (Australia), 1994.

[6] R.J. Drago. *Fundamentals of Gear Design*. Butterworths, 1988.

[7] C. Capdessus, M. Sidahmed, and J.L. Lacoume. “Cyclostationary processes: application in gear faults early diagnosis”. In: *Mechanical systems and signal processing* 14.3 (2000), pp. 371–385.

[8] L. Bouillaut and M. Sidahmed. “Cyclostationary approach and bilinear approach: comparison, applications to early diagnosis for helicopter gearbox and classification method based on HOCS”. In: *Mechanical Systems and Signal Processing* 15.5 (2001), pp. 923–943.

[9] A.C. McCormick and A.K. Nandi. “Cyclostationarity in rotating machine vibrations”. In: *Mechanical systems and signal processing* 12.2 (1998), pp. 225–242.

[10] R.B. Randall, J. Antoni, and S. Chobsaard. “The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals”. In: *Mechanical systems and signal processing* 15.5 (2001), pp. 945–962.

[11] J. Antoni, J. Daniere, and F. Guillet. “Effective vibration analysis of IC engines using cyclostationarity. Part I: a methodology for condition monitoring”. In: *Journal of sound and vibration* 257.5 (2002), pp. 815–837.

[12] G.B. Giannakis. “Cyclostationary signal analysis”. In: *Digital Signal Processing Handbook* (1998), pp. 17–1.

[13] J. Antoni et al. “Cyclostationary modelling of rotating machine vibration signals”. In: *Mechanical systems and signal processing* 18.6 (2004), pp. 1285–1314.

[14] J. Antoni. “Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions”. In: *Journal of Sound and vibration* 304.3-5 (2007), pp. 497–529.

[15] J. Antoni. “Cyclic spectral analysis in practice”. In: *Mechanical Systems and Signal Processing* 21.2 (2007), pp. 597–630.

[16] J. Antoni, G. Xin, and N. Hamzaoui. “Fast computation of the spectral correlation”. In: *Mechanical Systems and Signal Processing* 92 (2017), pp. 248–277.

[17] P.D. McFadden and J.D. Smith. “Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review”. In: *Tribology international* 17.1 (1984), pp. 3–10.

[18] P. Borghesani and M.R. Shahriar. “Cyclostationary analysis with logarithmic variance stabilisation”. In: *Mechanical Systems and Signal Processing* 70 (2016), pp. 51–72.

[19] P. Borghesani et al. “Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals”. In: *Mechanical Systems and Signal Processing* 40.1 (2013), pp. 38–55.

[20] P. Borghesani and J. Antoni. “CS2 analysis in presence of non-Gaussian background noise–Effect on traditional estimators and resilience of log-envelope indicators”. In: *Mechanical Systems and Signal Processing* 90 (2017), pp. 378–398.

[21] P. Borghesani, P. Pennacchi, and S. Chatterton. “The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings”. In: *Mechanical Systems and Signal Processing* 43.1-2 (2014), pp. 25–43.

[22] J. Zeidler et al. “Adaptive enhancement of multiple sinusoids in uncorrelated noise”. In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26.3 (1978), pp. 240–254.

[23] R.B. Randall and Y.J. Li. “Diagnostics of planetary gear bearings in the presence of gear vibrations”. In: *IMECHE Conference Transactions*. Vol. 5. Mechanical Engineering Publications. 1995, pp. 73–80.

[24] J. Antoni and R.B. Randall. “Unsupervised noise cancellation for vibration signals: part I—evaluation of adaptive algorithms”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 89–101.

[25] J. Antoni and R.B. Randall. “Unsupervised noise cancellation for vibration signals: part II—a novel frequency-domain algorithm”. In: *Mechanical Systems and Signal Processing* 18.1 (2004), pp. 103–117.

[26] P.D. McFadden. “A revised model for the extraction of periodic waveforms by time domain averaging”. In: *Mechanical Systems and Signal Processing* 1.1 (1987), pp. 83–95.

[27] R.B. Randall, N. Sawalhi, and M. Coats. “A comparison of methods for separation of deterministic and random signals”. In: *International Journal of Condition Monitoring* 1.1 (2011), pp. 11–19.

[28] N. Sawalhi and R.B. Randall. “Signal pre-whitening using cepstrum editing (liftering) to enhance fault detection in rolling element bearings”. In: *Proceedings of the 24th international congress on condition monitoring and diagnostic engineering management*. 2011, pp. 330–336.

[29] P. Borghesani et al. “Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions”. In: *Mechanical Systems and Signal Processing* 36.2 (2013), pp. 370–384.

[30] J. Antoni and R.B. Randall. “The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines”. In: *Mechanical Systems and Signal Processing* 20.2 (2006), pp. 308–331.

[31] V. Camerini et al. “Impact of pulse time uncertainty on synchronous average: statistical analysis and relevance to rotating machinery diagnosis”. In: *Mechanical Systems and Signal Processing* 129C (2019), pp. 308–336.

[32] P.D. McFadden. *Interpolation techniques for the time domain averaging of vibration data with application to helicopter gearbox monitoring*. Tech. rep. AERONAUTICAL RESEARCH LABS MELBOURNE (AUSTRALIA), 1986.

[33] R.B. Randall and J. Antoni. “Rolling element bearing diagnostics—A tutorial”. In: *Mechanical systems and signal processing* 25.2 (2011), pp. 485–520.

[34] P.D. McFadden and J.D. Smith. “Model for the vibration produced by a single point defect in a rolling element bearing”. In: *Journal of sound and vibration* 96.1 (1984), pp. 69–82.

[35] P.D. McFadden and J.D. Smith. “The vibration produced by multiple point defects in a rolling element bearing”. In: *Journal of sound and vibration* 98.2 (1985), pp. 263–273.

[36] B.J. Hamrock and W.J. Anderson. *Rolling element bearings*. Tech. rep. 1983.

[37] R.B. Randall. “Applications of spectral kurtosis in machine diagnostics and prognostics”. In: *Key Engineering Materials*. Vol. 293. Trans Tech Publ. 2005, pp. 21–32.

[38] N. Tandon and A. Choudhury. “A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings”. In: *Tribology international* 32.8 (1999), pp. 469–480.

[39] J. Antoni and R.B. Randall. “A stochastic model for simulation and diagnostics of rolling element bearings with localized faults”. In: *Journal of vibration and acoustics* (2003).

[40] D. Ho and R.B. Randall. “Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals”. In: *Mechanical systems and signal processing* 14.5 (2000), pp. 763–788.

[41] J. Antoni and R.B. Randall. “Differential diagnosis of gear and bearing faults”. In: *Journal of Vibration and Acoustics* 124.2 (2002), pp. 165–171.

[42] S. Kass, A. Raad, and J. Antoni. “Self-running Fault Diagnosis Method for Rolling Element Bearing”. In: *Mechanism, Machine, Robotics and Mechatronics Sciences*. Springer, 2019, pp. 127–140.

[43] J. Antoni. “Cyclostationarity by examples”. In: *Mechanical Systems and Signal Processing* 23.4 (2009), pp. 987–1036.

A new indicator designed from the spectral coherence, proposition and application to bearing diagnosis

Souhayb KASS[1,2], Amani RAAD[2], Jérôme Antoni[1]

[1] Laboratoire Vibrations Acoustique, Univ Lyon, INSA-Lyon, LVA EA677, F-69621 Villeurbanne, France, e-mail : ***jerome.antoni@insa-lyon.fr***

[2] Engineering faculty, Lebanese University, Tripoli, Lebanon

1 Abstract

In vibration-based diagnosis of rolling element bearings, the complexity of the signals requires an expert to use advanced signal processing tools and to interpret the results based on his/her experience. Recently, a few autonomous methods have been proposed to alleviate the demand on the user's expertise, yet they have been mainly focused on fault detection. They ideally track certain properties in the signal, whose occurrence is correlated with the symptom of a fault. This paper follows a similar direction but with wider objectives: it aims to develop an indicator that is sensitive to both non-stationarity, non-Gaussianity and to the modification of the acoustic signature of the vibratory signal. The indicator is based on the recently developed Fast Spectral Coherence, a key tool of the theory of second-order cyclostationary processes. It condenses the whole information initially displayed in three dimensions into a scalar. it initially addresses the case where the faults frequencies are unknown. In addition, the proposed indicator is able to return information for different levels of damages in both stationary and non-stationary operating conditions. A new pre-processing step is provided to ensure an efficient and constant statistical threshold. The proposed indicator is intended to be used in an autonomous process without the need for visual analysis and human interpretation. The proposed indicator is compared with a recent indicator based on the Envelop Spectrum, in terms of classification and detection performance. Several applications using real and benchmarked data eventually illustrate the capability for self-running diagnosis.

Keywords: *spectral coherence, Gaussianity, Stationarity, Acoustic signature, Indicator.*

2 Introduction

Roller bearings (REBs) are one of the essential components of rotating machines, hence the demand for their efficient and reliable condition monitoring (CM). Condition monitoring ensures maximum production, prevents accidents and serious damage and helps to detect failures at an early stage by keeping the system in good conditions. Over the last few decades, considerable research has been conducted on diagnostics based on REB vibrations and acoustics signals leading to the development of some condition's indicators. Therefore, several strategies have been adopted. This is derived from the fact that the occurrence of numerous faults induces changes in signal characteristics that can be described as (i) a deviation from Gaussianity, and/or (ii) a shift in the statistical behavior of the signal from stationary to non-stationary, accompanied by (iii) a change in the machine's acoustic signature.
The crest factor [1], the peak-to-peak, the entropy [2]–[4], the form factor, the third-order central moment (skewness) and the fourth-order central moment (kurtosis) [5], [6], or any higher-order moments or cumulants [7], [8] are all typical examples of the non-Gaussianity measure. They are dedicated to characterizing the non-Gaussian behavior in the form of impulsivity of machine signals. The most traditional and probably the most widely used is the kurtosis, or its combination with other indicators. In the past, they have been used mainly because of their simple calculation formulas, and their relatively

short computation times. This is despite its uncorrelated values with the fault symptoms in numerous situations, reported in many studies[9], [10]. This argument has become obsolete thanks to modern computing capabilities.
The roughness indicator, as traditionally calculated in the time domain from the Aures model [11], is an example of the psychoacoustic parameters used to monitor the existence of a fault based on an alteration of the machine's acoustic signature. It aims to mimic the ability of the human auditory system to detect high-frequency modulation, as evidenced by the faulty rotating machines. The researchers have developed various mathematical models [12]–[15] to estimate roughness but none of them have been normalized.
The degree of cyclostationarity [16] and the indices of cyclostationarity [17] based on the 2nd-order cyclical cumulants or any higher-order cyclostationarity indices based on higher-order cyclical cumulants [18] are typical examples of measures of the non-stationarity that characterizes the cyclostationarity introduced by the fault existence. Despite its importance in diagnosis, the real-time use of CS indices can be hindered in practice by its high cost, especially in real-time applications [19].

This paper aims to fill in these gaps by proposing an indicator sensitive to both non-stationarity, non-Gaussianity and to the modification of the acoustic signature of the vibratory signal. To do so, the frequency domain instead of time domain is used since it better extracts cyclic repetition from a signal produced by a repetitive fault and also reduces the noise impact. The cyclostationary framework is then our subject of interest since it has been reported that rotating machine signals are cyclostationary [20]. Advantage is taken of the availability of a recently proposed fast algorithm to calculate the spectral coherence [19], on which the proposed conditioning indicator is based. The spectral correlation is a three-dimensional distribution of all modulation patterns existing in a signal as a function of the carrier frequency in Hertz and the modulation frequency (also called cyclic frequency) in machine order, which generalizes the SC to nonstationary operations. It is thus considered optimal for revealing bearing fault signatures under stationary and nonstationary speed regimes [21], [22].
The idea is to condensate the whole information initially displayed in the spectral coherence into a scalar after an appropriate weighting performed to select the audible frequencies range from about 20 Hz to 20 kHz and the audible modulations range from about 15 Hz to 200 Hz. This selection can be easily made using the weight $\omega_{k,p}$ designed as a bandpass filter, used to accentuate or reduce certain frequency components in order to model the bandpass characteristic of the roughness on the modulation frequency.
This approaching, the spectral coherence which is sensitive to both non-stationarity and non-Gaussianity, also becomes sensitive to the acoustic signature of the vibratory signal. A new preprocessing step is provided in order to eliminate any possible bias (as typically produced by transient disturbances in the signal or the presence of unexplained nonstationarities) in the estimated spectral coherence. This original step is necessary to produce a pivotal statistic by forcing the spectral coherence to have a constant probability distribution with respect to the dual frequencies plan.
The proposed indicator is statistically consistent, i.e. its variance converges to zero when the signal length increases. On the opposite, Aures' roughness does not involve any time average and is therefore prone to significant estimation errors. Contrary to the kurtosis, the proposed indicator separates impulsivity from non-stationarity, allowing the identification of the type of deviation from normality. A non-nonparametric hypothesis test is also provided in order for this indicator to be credible and possibly implemented in an automated monitoring system. The capacity of the proposed indicator is validated on real data and benchmarked with the kurtosis to extract meaningful conclusions. It is found to return higher performance in terms of detecting faulty bearings.

3 Indicator

3.1 Preliminary steps

This steps briefly resumes the statistical methodology proposed by *kass et al.*[22] to design the test statistics. The starting point is to describe the health of the system under investigation by two alternative hypotheses, H_0 and H_1, which correspond respectively to the healthy and the faulty states, respectively. The principle is to consider the spectral coherence, $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, as the random quantity of interest rather than its squared magnitude.

In principle, Under H_0, the random field $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, seen as a function of the two frequency variables α_l and f_k, can be shown to have zero probability of being nil at any position $(\alpha_l\ ,\ f_k)$ even though having small values. Under the alternative hypothesis H_1, the difference is that the random field will have higher magnitudes along parallel lines, discretely located at cyclic frequencies associated with the fault frequencies. The objective is to keep only these values and to zero all the other ones. To do so, a statistical threshold, defined as a high percentile, is needed to differentiate between information and background noise. Hence, the presence of a possible bias (as typically produced by transient disturbances in the signal or the presence of unexplained nonstationarities) in the $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ compromises the efficiency of the latter threshold. As result, the noise baseline is not uniformly distributed along the frequencies axis. It is therefore impossible to establish a fixed threshold to distinguish between information and noise. To correct this situation, the following empirical steps are proposed. The first step is to standardize the EES in order to force it to have a constant probability distribution with respect to the cyclic order α, under the null hypothesis H_0. In principle, The transformation reads

$$\gamma_X^{(2)}(\alpha_l\ ,\ f_k)\ = \frac{\gamma_X^{(1)}(\alpha_l\ ,\ f_k) - \mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}}{\sqrt{\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\} - \mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}^2}} \tag{1}$$

where $\mathrm{E}\{\cdots|H_0\}$ stands for the ensemble averaging operator taken under H_0. One issue is to replace $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}$ and $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\}$ in the above equation by estimates obtained from a realization of $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ which may either pertain to H_0 or H_1. Since the difference in $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ under the null and alternative hypotheses is essentially marked by the presence of parallel lines, it is proposed to estimate $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)|H_0\right\}$ from a running median of $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$, called $\mu_{MED}(\alpha_k)$, and $\mathrm{E}\left\{\gamma_X^{(1)}(\alpha_l\ ,\ f_k)^2|H_0\right\}$ from the running median of the absolute deviation, called $\sigma_{MAD}(\alpha_k)$. The rationale for using a running median is to leave unaffected informative peaks in the spectral coherence. Therefore, the $\gamma_X^{(2)}(\alpha_l\ ,\ f_k)$ reads

$$\gamma_X^{(2)}(\alpha_l\ ,\ f_k)\ = \frac{\gamma_X^{(1)}(\alpha_l\ ,\ f_k) - \mu_{MED}(\alpha_l, f_k)}{\sigma_{MAD}(\alpha_l, f_k)} \tag{2}$$

which returns a “pivotal” statistics (i.e. whose probability distribution does not depend on any unknown parameter). Briefly, the presence of the possible bias is firstly removed by subtracting a running median, then the results are standardized by dividing it with a running median of the absolute deviation. These two steps can be though as a normalization of the H_1 statistics with respect to the H_0 statistics.

3.2 Proposition

The first step is to preserve the informative values, i. e. the spectral lines parallel to the frequency axis f_k along α_l, expected during the fault’s existence i. e. under H_1.To do so, this step consists of setting all non-significant peaks in $\gamma_X^{(2)}(\alpha_l\ ,\ f_k)$ to zero which are found below a given threshold. A reasonable choice is to define the threshold as a high percentile p_c (e.g. $p_c = 0.9$)[23], which means that only the $100(1 - p_c)\%$ highest values will be kept. Thanks to the standardization step, the threshold is constant over the full frequency plane $(\alpha_l\,,\, f_k)$. And the $\gamma_X^{(3)}(\alpha_l\,,\, f_k)$ is expressed as

$$\gamma_X^{(3)}(\alpha_l\,,\, f_k) = \gamma_X^{(2)}(\alpha_l\,,\, f_k).\,\mathbb{I}_{\left[\gamma_X^{(2)}(\alpha_l\,,\, f_k) > p_c\right]} \tag{3}$$

where the symbol $\mathbb{I}_{\left[\gamma_X^{(2)}(\alpha_l\,,\, f_k) > p_c\right]}$ denotes the indicator function defined on the frequencies plane $(\alpha_l\,,\, f_k)$ having the value 1 for all elements of $(\alpha_l\,,\, f_k)$ satisfying the condition $\gamma_X^{(2)}(\alpha_l\,,\, f_k) > p_c$ and the value zero otherwise.

This is a crucial step as the next step involves modelling the bandpass characteristics of the roughness by accentuating or reducing certain frequency components, thus preventing the occurrence of misleading peaks.

The next step is to performs an appropriate weighting of the $\gamma_X^{(3)}(\alpha_l\,,\, f_k)$ so as to select the audible frequencies range from about 20 Hz to 20 kHz and the audible modulations range from about 15 Hz to 200 Hz in order to approach the roughness measurement proposed by Aures. This selection can be easily realized based on the weight $\omega_{k,p}$ designed as a bandpass filter, used to accentuate or reduce specific frequency components in order to model the roughness bandpass characteristic on modulation frequencies. In other words, $\omega_{k,p}$ resembles the distribution matrix of the weighting functions for each *Bark* channel. The weighting of $\gamma_X^{(3)}(\alpha_l\,,\, f_k)$ is achieved as follows:

$$\gamma_X^{(4)}(\alpha_l\,,\, f_k) = \gamma_X^{(3)}(\alpha_l\,,\, f_k).\,\omega_{k,p} \tag{4}$$

The third step is divided into two sub-steps, the first is to integrating the $\gamma_X^{(4)}(\alpha_l\,,\, f_k)$ over the cyclical frequency axis α_l, which condenses the whole information initially displayed in three dimensions into a two-dimensional representation,

$$I_X^{(5)}(\,f_k) = \frac{1}{F_1}\left[\sum_{l \in F_1} \gamma_X^{(4)}(\alpha_l\,,\, f_k)\right] \tag{5}$$

While the second sub-step consists of dividing the frequency axis into a *Bark* filter bank to estimate the modulation depth per auditory channel which is spaced by 1 *Bark* representing a psychoacoustic scale for the bandwidths of the hearing filters. The latter is a frequency band established by Zwiker [24], it is divided into 24 critical bands ranging from 0 to 15500 Hz.

As a final step, the roughness dependence with respect to the carrier frequency is introduced into the model by multiplying $I_X^{(5)}(\,f_k)$ by a weighting function $g(f_k)$ with factors ranging from 0.6 to 1.1 in accordance with the dependency of the roughness to the carrier frequency of the amplitude modulated tones. The values of the weighting function with respect to the channel number are shown in Figure 1.

Finally, the proposed indicator I_R is obtained by integrating $I_X^{(5)}(\,f_k)$ over the frequency axis f_k.

$$I_R = \frac{1}{F_2}\sum_{k \in F_2} I_X^{(5)}(\,f_k)\,.\,g(f_k) \tag{6}$$

By analogy with the result given by the connection between the kurtosis and the sum of the squared envelope spectrum [25], it was shown in [23] that the proposed indicator might be interpreted as a kurtosis, yet sensitive only to cyclostationary components.

The next section describes how the fault detection will be done using a statistical hypothesis test using the proposed indicator. In this context, alternative strategies to statistical testing can also be used. For example, the proposed indicator can also be used as an input parameter for an SVM classifier (machine vector support) or a neural network. This statement is based on the observation of the results obtained by applying the proposed indicator to the various databases.

3.3 Hypothesis testing, design

The interpretation of the proposed indicator could vary from one application to another, depending on several parameters (such as the noise level related to transient perturbations in the signal or to the presence of unexplained non-stationarity as well as the vibration level and the interfering contribution of other second-order components emitted by other sources). In the majority of the literature [16]–[18], the provided methods give their results as scalar. The latter indicates the presence of a fault in some applications while in another, and for the same value, the fault will be considered as absent. This is why a threshold is needed for decision-making. To do so, two thresholds are provided in this paper. The lower threshold I_L is defined as the indicator value of the randomized version of the vibration signal under which the fault is absolutely absent. In detail, I_L is equal to the indicator value when the signal is randomized. The randomization of the signal is defined as a circular permutation of its elements, it can be performed using the MATLAB function called "randperm". The latter returns a new version of the signal containing a random permutation of its values. On the other hand, the link between the kurtosis and the spectral correlation makes it possible to reach the upper threshold, I_U, beyond which the fault presence is declared with high certainty.

After the I_U is calculated, a comparison is then made with the proposed indicator to detect the fault presence. If I_R has a value greater than I_U, the fault exists.

The null hypothesis test relative to our case originating from the comparison between I_R and I_U can be written as:

“Reject the null hypothesis H0 if:

$$I_R \geq I_U \tag{7}$$

where $I_U \geq 2 \times I_L$.

It easily allows performing a statistical test: according to the decision rule, any value of I_R that is greater than the I_U-threshold will indicate that the signature of the fault is detected. The proof of proposition Eq. (7) is based on observing that under the null hypothesis test H_0 the quantity asymptotically follows a nonparametric distribution that has a constant bias and variance all over the cyclic order axis. It also remembers that this test is true almost everywhere.

The complete flow diagram for the algorithm described in this section is shown in Fig.3

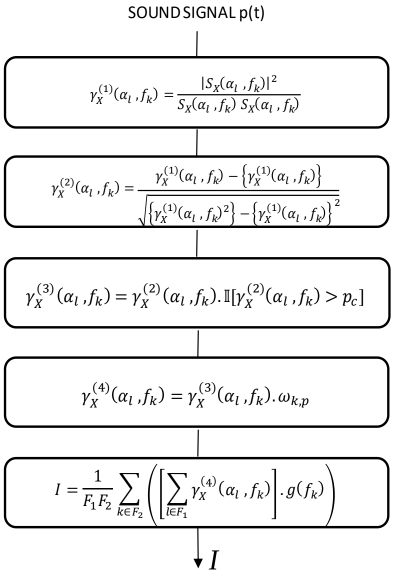


Figure 1.Complete algorithm flowchart

It is worth noting that I_R is very similar to the Aures' roughness measure used in psychoacoustics. One difference is that Aures' roughness is based on a decomposition of the signal through a Bark filter bank whereas a narrow-band decomposition is used in this paper, yet this is more or less transparent after integration over the frequency plane $(\alpha_l\ ,\ f_k)$. Another difference is that Aures' roughness does not involve any time average and is therefore prone to significant estimation errors. On the contrary, the indicators introduced in this work are statistically" consistent" (i.e. their variances converge to zero when the signal length increases).

4 Experimental Validation

The ability of any method in detecting a bearing fault must be validated on real signals. In the present paper, four benchmarks are used. The first is provided by the Case Western Reserve University (CRWU) bearing data center [26], while the second is an industrial database provided by SOMFY-Cluse. These databases are widely used to test new algorithms by comparing their efficiency with existing techniques [27]. The CRWU's database provides multiple fault types, i.e. rolling element, cage inner-race, and outer-race fault, and it is used to illustrate the proposed method and to compare the proposed algorithm to those existing in the literature. The industrial database is used to illustrate the diagnosis of bearings in a real industrial world signal.

4.1 Algorithm illustration and comparison with kurtosis

To illustrate the proposed algorithm, we consider a real industrial signal. This analyzed signal is provided by Somfy and includes an industrial fault. It is provided as supplementary material of the article. This may be used as a general source of bench mark data for research on diagnosis of industrial faults under constant speed operation. The comparing the results of the proposed indicator with those given by kurtosis. The duration of signals is 20 s with a sampling frequency of 50 kHz.

As explained above, the first begins with the calculation of the fast estimator of the spectral coherence for the resampled time domain signal $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$. In what follows, the window length in the Fast-OFSC is set to $Nw = 2^9$ in order to achieve a frequency resolution of about 100 Hz and the cyclic range $\alpha_{max} =$ 750 Hz. The next step is to standardize $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ in order to force it to have a constant probability distribution with respect to the frequency plane, under H_0. The statistical threshold will be defined as a high P_c percentile ($P_c = 0.9$), which means that only the $100(1 - P_c)\%$ of the highest values will be preserved. It is noteworthy that this method perfectly preserves the diagnostic information that nicely appears with a significant overrun of the 0,1% statistical threshold. This signal is easily diagnosable and it should, therefore, be considered as a preliminary test for the proposed algorithm. The visual inspection of the spectral coherences presented in Figure 2 (a) and (b) shows a series of symptomatic pulses at the fault frequency - spectral lines parallel to the f-*axis* discretely located at cyclic frequencies associated with the fault frequencies- as expected by the model given by Somfy during under H_1. In Figure 2 (b) it is obvious that some frequency components disappear from the $\gamma_X^{(1)}(\alpha_l\,,\, f_k)$ while retaining only informative peaks and eliminating noise-related components found below the chosen threshold.

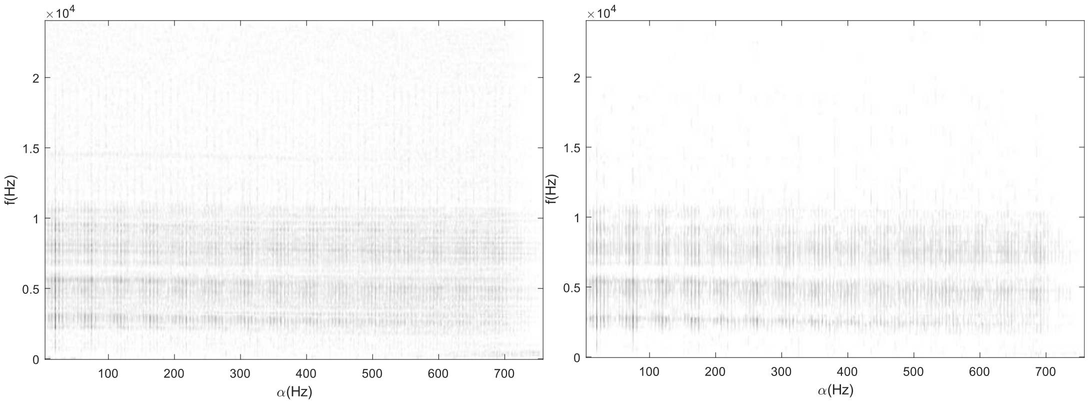


Figure 2. For the raw time signal a) the $\gamma_X^{(1)}(\alpha_l\,,\, f_k)$, and b) the $\gamma_X^{(3)}(\alpha_l\,,\, f_k)$

The audible frequencies and modulations range are selected by using the weight $\omega_{k,p}$, shown in Figure 3(a), so as to select the audible frequencies range from about 20 Hz to 20 KHz and the audible modulations range from about 15 Hz to 200 Hz. The weighted version of spectral coherence $\gamma_X^{(4)}(\alpha_l\,,\, f_k)$ is presented in Figure 3 (b). As shown in Figure 3 (b), certain frequency components are emphasized or reduced so as to model the band-pass characteristic of the roughness over the modulation frequency.

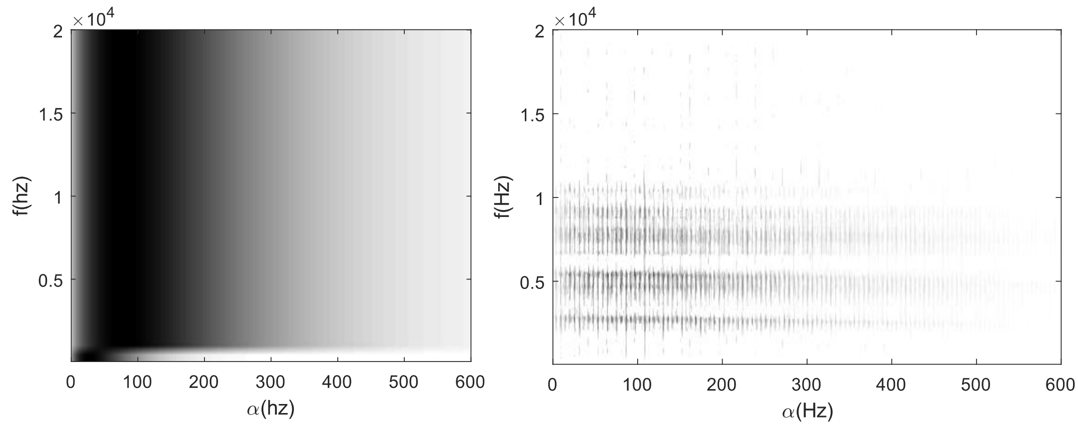


Figure 3 (a) The weight $\omega_{k,p}$ used to select the audible frequencies and modulations ranges, (b) the $\gamma_X^{(4)}(\alpha_l\,,\, f_k)$.

In the next step and as mentioned in section 3.2, the integration of $\gamma_X^{(4)}(\alpha_l\,,\, f_k)$ over α_k will be performed. The two-dimensional representation $I_X^{(5)}(\,f_k)$ which condenses the three-dimensional information is shown in Figure 4 (a). Then, the modulation-depth per auditory channel is estimated by

dividing the f-axis into the 24 bands of Bark filter. $I_X^{(5)}(f_k)$ is then multiplied by a weighting function $g(f_k)$. Finally, the roughness indicator I_R is obtained by integrating the weighted $I_X^{(5)}(f_k)$ over the f-axis. The latter has in this case a value of 2.2624.
The signal is now randomized in order to calculate the upper and the lower threshold. Figure 4 (a) et (b) shows both the raw time signal and its normalized version. When both signals are visually inspected, it is evident that cyclostationary symptoms are lost when the signal is randomized, which is consistent with the auditory test performed using MATLAB's sound function (). More precisely, when performing a hearing test, the original signal exhibits periodic behavior that is produced with each cycle. This periodic symptom is no longer heard after the randomization of the signal. The value of the indicator in this case $I_L = 0.27109$ is negligible compared to the case of existing fault, matching both the visual and the hearing inspection. I_R is equal to about 10 times I_L. According to the decision rule provided by Eq. (7), the fault existence is reported.

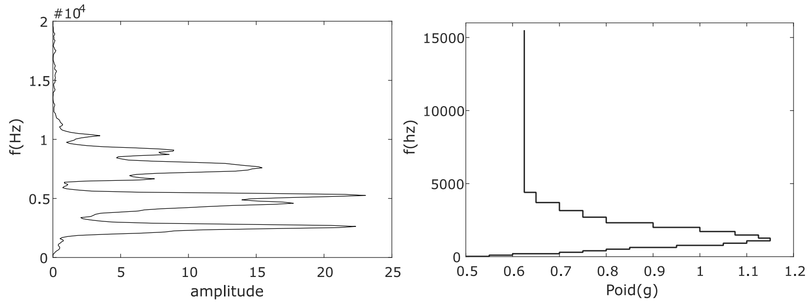


Figure 4. (a) the integration of $\gamma_X^{(4)}(\alpha_l\ ,\ f_k)$ over α_l , (b) the weighting function g

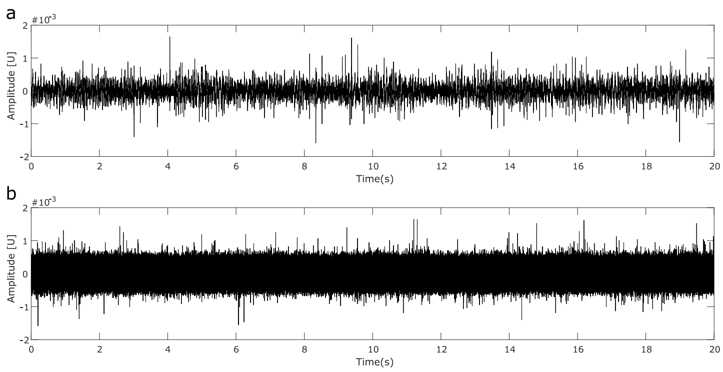


Figure 5. (a) the raw time signal and (b) its randomized version

Surprisingly, and contrary to both visual and auditory inspection tests, the kurtosis of the original and randomized signal gives a very low value of 3.0088 reporting the fault absence in both cases.
After an appropriate filtration of the raw time signal, the kurtosis value is now 20.8589, corresponding to a very high value -7 times the value of a normal case - indicating the failure's presence. The success of kurtosis after proper filtering highlights its limitation when analyzing a signal with a low signal-to-noise ratio and at the same time demonstrating the superiority of the indicator designed from a cyclostationary method for the detection of fault symptoms.
Now the filtered signal is randomized to study kurtosis response in this case and presents another superiority of the proposed indicator over kurtosis. In detail, the same kurtosis value for both the randomized signal and the filtered signal is obtained indicating that the fault is detected in these cases. The obtained results are expected since kurtosis is defined as follows

$$k = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} \quad (8)$$

This equation shows that regardless the sequence of this signal summation, results will be the same since the permutation disorganizes only the sequence of the original signal. The kurtosis in this case cannot therefore indicate whether the signal is cyclostationary or stationary but not Gaussian. On the other hand, the proposed indicator can perform this distinction giving a cyclic roughness value of 2.2424 for the filtered signal and 0.25819 for its randomized version.

4.2 Performance Evaluation in the CWRU database

The performance of the proposed indicators is now evaluated on the bearing signals provided by the CWRU database. The CWRU database has been used in many references (e.g. [19], [22], [27], [28]) and can be considered as a reference to test newly proposed algorithms and compare them against the state-of-the-art. The experimental setup consists of a 1.4914 kW, reliance electric motor driving a shaft on which a torque transducer and encoder are mounted. Torque is applied to the shaft via a dynamometer and electronic control system. Four types of vibration signals are collected (normal, ball fault, inner-race fault, and outer-race fault), acquired by accelerometer sensors under different operating loads and speeds. More details about the test bench as well as the description of its vibration signals can be found in the reference source [27]. In this study, the drive end data-set category with sampling frequency 48 kHz have been analyzed. Information for all 64 data sets used are shown in table 1. The capacity of the proposed indicator is evaluated using different faults types. The $\gamma_X^{(1)}(\alpha_l\ ,\ f_k)$ parameters are as given in the previous section.

Table 1. The 48K drive end bearing faults data sets used.

Fault types	Data sets name
Inner-race	110,111,112,174,176,177,14,215,217
Outer-race (centered)	135,136,137,138,201,203,204,238,239,240,241
Normal data	97,98,99,100

Table 2 and table 4 collect the results of the proposed method. Included in these tables are the I_R as well as the *kurtosis* values, of each raw time signal and its randomized version.
As shown in Table 3, for the original signals, I_R is close to 0.15, while for its randomized version, it is about 0.1 ($I_L = 0.1$). In all these cases and in according to the decision rule in equation (7), the fault is declared missed. The kurtosis in these cases is approximately 3. From Table 3, nearly the same increasing or decreasing behavior of the proposed indicator values are detected compared to the values provided by kurtosis and by the roughness indicator provided in commercial psychoacoustic software. In all these cases and in according to the decision rule in equation (7), the fault is declared presented. According to the obtained results, all faults detected by the human visual inspection of [27] in the inner ring and outer ring are also detected by the proposed indicator. It is clearly proven that the distinction between healthy and defective bearings can be made using these indicators. Unfortunately, given that the proposed indicator provides overlapping values when applied to the different types of bearing faults, the existence of the fault can be detected but not identified.
In conclusion, the objective has been achieved and the proposed indicator can identify the fault even if its frequencies are unknown.

Table 2. analysis results of the healthy bearing; Kurtosis, I_R

	Raw signal	Randomized signal

Dataset	*kurtosis*	I_R	*kurtosis*	I_L
100	2.9572	0.1354	2.9572	0.0999
97	2.7642	0.1471	2.7642	0.1011
98	2.9306	0.1422	2.9306	0.1002
99	2.9306	0.1422	2.9306	0.0977

Table 3. analysis results of the faulty bearing; Kurtosis, I_R

Inner-race faults				
	Raw signal		Randomized signal	
Dataset	*kurtosis*	I_R	*kurtosis*	I_L
110	7.36862	0.494821	7.36862	0.101008
111	7.51043	0.467891	7.51043	0.103501
112	6.82779	0.439782	6.82779	0.104584
174	11.6689	0.504829	11.6689	0.10279
176	20.1511	0.861307	20.151	0.117898
177	14.9400	0.823425	14.940	0.114356
214	3.93724	0.507309	3.93724	0.100236
215	3.65185	0.444887	3.65185	0.099826
217	3.65185	0.444887	3.65185	0.09756
Outer-race faults (Centered)				
	Raw signal		Randomized signal	
Dataset	*kurtosis*	I_R	*kurtosis*	I_L
135	6.7415	0.5032	6.741461	0.1047
136	6.8648	0.5293	6.864769	0.1027
137	6.9370	0.5295	6.937026	0.1017
138	7.3917	0.5154	7.391694	0.1030
201	3.7676	0.2039	3.767617	0.1031
203	6.0172	0.2108	6.017181	0.1015
204	2.9993	0.2104	2.999284	0.0996
238	20.7910	2.3374	20.79105	0.1169
239	20.5436	2.7322	20.54365	0.1148
240	21.0265	0.5032	21.02651	0.1205
241	19.7818	0.5293	19.78182	0.1199

5 Conclusion

This paper aims introduces an autonomous method of bearing diagnosis. It relies on the introduction of a new scalar indicator. The indicator results from a post-processing of the spectral coherence, as computed by the fast algorithm.

The factors that are likely to impede the autonomous diagnosis have been addressed; a new standardization of the estimated spectral coherence to remove any possible bias and frequency dependence in the estimation variance. The method comes with a robust hypothesis test, which is crucial for decision making.

The proposed method has been validated on several databases, where it has been checked to be able to systematically replace both the human intervention or the classical conditioning indicator to efficiently complete the diagnosis of bearings.

6 References

[1] J. Igba, K. Alemzadeh, C. Durugbo, and E. T. Eiriksson, “Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes,” *Renew. Energy*, vol. 91, pp. 90–106,

2016.

[2] R. Yan and R. X. Gao, “Approximate entropy as a diagnostic tool for machine health monitoring,” *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 824–839, 2007.

[3] H. Hong and M. Liang, “Fault severity assessment for rolling element bearings using the Lempel–Ziv complexity and continuous wavelet transform,” *J. Sound Vib.*, vol. 320, no. 1–2, pp. 452–468, 2009.

[4] J. Antoni, “The infogram: Entropic evidence of the signature of repetitive transients,” *Mech. Syst. Signal Process.*, vol. 74, pp. 73–94, 2016.

[5] A. Rai and S. H. Upadhyay, “A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings,” *Tribol. Int.*, vol. 96, pp. 289–306, 2016.

[6] C. Pachaud, R. Salvetat, and C. Fray, “Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces,” *Mech. Syst. Signal Process.*, vol. 11, no. 6, pp. 903–916, 1997.

[7] M. Cocconcelli, G. Curcurú, and R. Rubini, “Statistical evidence of central moment as fault indicators in ball bearing diagnostics,” in *The International Conference Surveillance 9*, 2017.

[8] H. R. Martin, “Statistical moment analysis as a means of surface damage detection,” in *Proceedings of the 7th international modal analysis conference*, 1989, vol. 1, pp. 1016–1021.

[9] J. Antoni and P. Borghesani, “A statistical methodology for the design of condition indicators,” *Mech. Syst. Signal Process.*, vol. 114, pp. 290–327, 2019.

[10] R. B. Randall, J. Antoni, and K. Gryllias, “Alternatives to kurtosis as an indicator of rolling element bearing faults,” in *Proceedings of ISMA2016 International Conference on Noise and Vibration Engineering*, 2016, pp. 2503–2516.

[11] W. Aures, “Ein berechnungsverfahren der rauhigkeit,” *Acta Acust. united with Acust.*, vol. 58, no. 5, pp. 268–281, 1985.

[12] P. Daniel and R. Weber, “Psychoacoustical roughness: Implementation of an optimized model,” *Acta Acust. united with Acust.*, vol. 83, no. 1, pp. 113–123, 1997.

[13] P. N. Vassilakis, “Auditory roughness as a means of musical expression,” *Sel. Reports Ethnomusicol.*, vol. 12, no. 119–144, p. 122, 2005.

[14] M. Leman, “Visualization and calculation of the roughness of acoustical musical signals using the synchronization index model (SIM),” in *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-00), December*, 2000, pp. 7–9.

[15] R. Duisters, “The modeling of auditory roughness for signals with temporally asymmetric envelopes.” Master’s thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2005.

[16] G. D. Živanović and W. A. Gardner, “Degrees of cyclostationarity and their application to signal

detection and estimation," *Signal Processing*, vol. 22, no. 3, pp. 287–297, 1991.

[17] P. PRIEUR and G. D'URSO, "Des indices de cyclostationnarité pour la surveillance des engrenages," in *15° Colloque sur le traitement du signal et des images, FRA, 1995*, 1995.

[18] A. Raad, J. Antoni, and M. Sidahmed, "Indicators of cyclostationarity: Theory and application to gear fault monitoring," *Mech. Syst. Signal Process.*, vol. 22, no. 3, pp. 574–587, 2008.

[19] J. Antoni, G. Xin, and N. Hamzaoui, "Fast computation of the spectral correlation," *Mech. Syst. Signal Process.*, vol. 92, pp. 248–277, 2017.

[20] J. Antoni, "Cyclic spectral analysis in practice," *Mech. Syst. Signal Process.*, vol. 21, no. 2, pp. 597–630, 2007.

[21] D. Abboud, "Vibration-based condition monitoring of rotating machines in nonstationary regime," 2015.

[22] S. Kass, A. Raad, and J. Antoni, "Self-running bearing diagnosis based on scalar indicator using fast order frequency spectral coherence," *Measurement*, vol. 138, pp. 467–484, 2019.

[23] J. Antoni, P. Borghesani, S. Kass, A. Raad, and K. Gryllias, "Methodologies for designing new condition indicators," in *Proceedings of ISMA 2018-International Conference on Noise and Vibration Engineering and USD 2018-International Conference on Uncertainty in Structural Dynamics*, 2013, pp. 883–890.

[24] E. Zwicker and H. Fastl, "Psychoacoustics: Facts and Models. 1990." Springer-Verlag.

[25] P. Borghesani, P. Pennacchi, and S. Chatterton, "The relationship between kurtosis-and envelope-based indexes for the diagnostic of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 43, no. 1–2, pp. 25–43, 2014.

[26] "The Case Western Reserve University Bearing Data Center Website." [Online]. Available: https://csegroups.case.edu/bearingdatacenter/.

[27] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study," *Mech. Syst. Signal Process.*, vol. 64, pp. 100–131, 2015.

[28] W. Li, M. Qiu, Z. Zhu, F. Jiang, and G. Zhou, "Fault diagnosis of rolling element bearings with a spectrum searching method," *Meas. Sci. Technol.*, vol. 28, no. 9, p. 95008, 2017.

Multi band integration on the cyclostationary bivariable methods for bearing diagnostics.

Alexandre MAURICIO[1,2], Konstantinos GRYLLIAS[1,2]

[1]KU Leuven, Faculty of Engineering Science, Department of Mechanical Engineering, Division PMA, Celestijnenlaan 300 B, B-3001, Heverlee, Belgium

[2]Dynamics of Mechanical and Mechatronic Systems, Flanders Make
alex.ricardomauricio@kuleuven.be

Abstract

Rolling element bearings are critical parts of rotating machinery, as they support the loads applied to the rotating components. Therefore, continuous monitoring of the health state of the operational bearings is applied in order to detect early damages before any unexpected breakdown of the rotating machinery occurs. Bearing diagnostics is a field of intensive research, focusing nowadays mainly in complicated machinery (e.g. planetary gearboxes, multi-stage gearboxes, etc.) operating under varying conditions (e.g. varying speed and load), as they still provide challenges in terms of accuracy and time of detection/diagnosis. One of the most common methods for bearings diagnostics is the Envelope Analysis. A filter is usually applied around an excited frequency band (by impulsive damage) and the signal is enveloped, thus obtaining the Squared Envelope Spectrum. For the detection of the filtering frequency band, several band selection tools have been proposed in the past that extract the optimal band in a semi-autonomous or fully autonomous manner. The most widely used tool for band selection is the Kurtogram, where the band that returns the highest Spectral Kurtosis value is selected as the optimal band for demodulation. However, as the bearing damage may excite several frequency bands simultaneously, band-pass filtering around only one frequency band may not be sufficient for the detection of the bearing fault under the presence of noise. One proposed method to circumvent this case is to filter around several bands that carry the Signal of Interest (bearing damage signature). Recently, multi-band filtering based on the Autogram feature values, used as a pre-step in order to extract the Combined Squared Envelope Spectrum (CSES) has been presented, providing better detection performance of faulty bearings compared to the extraction of the SES after filtering over a single optimal band returned by the Autogram. Recently, a particular interest had been target to the Cyclic Spectral Correlation (CSC) and to the derived methods, due to their effectiveness in describing second-order cyclostationary signals. One of such methods is the Cyclic Spectral Coherence (CSCoh) which is a normalized version of the CSC bivariable map. Both methods are represented in the frequency-frequency domain. It has been shown that the integration of the bivariable functions over discrete spectral frequency bands is analogous to band-pass filtering. The IESFOgram has been proposed as a band selection tool, based on either the CSC or CSCoh, in order to extract the optimal frequency band. The integration on the frequency band of the bivariable map further enhances the detectability of faulty bearings on the resulting Improved Envelope Spectrum (IES). However, the method has been proposed with the integration of one single band. In this paper the method is extended towards the extraction of the Combined Improved Envelope Spectrum (CIES), performing a multi-band integration of the bivariable map around multiple resonant frequencies that are carriers of the bearing damage signature. The proposed method is applied, tested and evaluated on experimental data and the results are compared with other state-of-the-art band-selection tools.

1 Introduction

Rolling element bearings are critical components of rotating machinery and their failure can cause sudden breakdown of the system, leading to time-loss and increased costs. Condition monitoring is the field where rotating machinery is analysed, including bearings and gears and damages that may be present on the structures can be detected. Therefore, maintenance and faulty component repair can be performed before breakdown. The diagnostics of bearings continues to be a challenge however, as their signatures are usually masked under noise and other stronger component signatures (e.g. gears). Specifically, condition monitoring of complex machinery has seen increased research, due to their wide application on critical mechanisms and to their high difficulty to diagnose caused by their many components signatures.

One of the most well established methods is the Envelope Analysis, where the signal is demodulated after band-pass filtering around the resonant frequencies excited by the damage impulses, obtaining in the end a filtered Squared Envelope Spectrum (SES). The main idea is to obtain an optimal filter band which presents a high Signal-to-Noise ratio (SNR) leading to a SES after demodulation where the fault harmonics are enhanced [1, 2]. The selection of this frequency band for demodulation is a frequency and continuous topic present in the field. The main reason is because some sort filtering processing is common to most of condition monitoring applications, and the band can either be selected by engineering knowledge, or by a methodology that selects the band in a (semi-)automated manner. The most widely used of these band selections tools is the Fast Kurtogram (FK) [3], which is an automated band selection tool based on the maximum kurtosis level. Aside from this tool there are other band selection tools that have been developed to obtain the SES. The Optimised Spectral Kurtosis (OSK) [4] selects the band with the maximum kurtosis as well, while retaining a narrow bandwidth in order to by-pass electro-magnetic interference noise on the signals. The Sparsogram [5] is based on the sparsity level on different bands based on the wavelet-packet, and the Infogram is utilizes the negentropy as a feature to detect the impulsive bands of the signal for demodulation. Moshrefzadeh and Fasana proposed the Autogram [6], a tool also based on the maximum kurtosis, but unlike the FK, it is calculated from the unbiased autocorrelation of the squared envelope of the demodulated signals. Instead of a classical filter (e.g. Butterworth filter), the undecimated wavelet packet transform (MODWPT) is used instead to split the signal in a series of frequency bands. The band of the autocorrelated squared envelope with the highest Kurtosis is selected as the optimal one, and is shown to have higher diagnosis performance than other stat-of-the-art band selections tools. One conclusion of the method is that band selection tools select only one node as the optimal, and often other unused nodes may contain useful information that is negleted. As such, they introduce the concept of multi-band integration to the Autogram, where several filtered SES, corresponding to the highest kurtosis of each level, are all combined into one spectrum denominated as Combined Squared Envelope Spectrum (CSES).

The Cyclic Spectral Correlation (CSC) and the Cyclic Spectral Coherence (CSCoh) have been proposed in the last two decades as an alternative for the SES-based methods [7, 8, 9]. The main advantage of this method falls on its ability to reveal hidden periodicities of second-order cyclostationarity, like bearing signals that are masked under stronger signals. They are represented in bi-variable maps in the frequency-frequency domain, from which its spectral axis can be integrated to obtain either the Enhanced Envelope Spectrum (EES) or the Improved Envelope Spectrum (IES). These spectra, the EES and IES, have been seen to improve the detection of cyclostationary faulty signals. However, to obtain the optimal band of demodulation for the CSC or CSCoh, its bi-variable map needs to be analysed in order to select the optimal band for integration along the spectral axis. The IESFOgram [11] as been previously proposed as a band selection tool to be applied on the bi-variable maps of CSC or CSCoh, in order to take advantage of its good perfomance in extracting the cyclistationry information of the signals. It also displays a color-mapped 1/3 binary tree like the FK and is seen to provide an optimal band of integration resulting in an IES allowing the detection of the fault frequency harmonics. However, as information of the damage can be present in other bands beside the optimal band, the authors propose an approach to combine the spectra of different bands into one Combined Improved Envelope Spectrum (CIES).

The objective of this paper is the proposal of extending the IESFOgram methodolgy by adding the information of other bands into a combined spectrum, the CIES. The methodology is validated on real signals of two datasets (one with roller bearing damage under variable speed and load conditions [10] and one from a planetary gearbox with electromagnetic interference [4]). Furthermore, the performance of the methodology is compared with the band pass filtering selection based on the Fast Kurtogram-based SES and the Autogram-based CSES. Figure 1 depicts a scheme of the used methodologies. The rest of the paper is outlined as follows. In Section 2, the background theory deemed needed for the application of the proposed method is detailed. In Section 3 the proposed methodology itself is presented. In Section 4, the methodology is tested, validated and compared with state of the art methodologies. The paper closes in Section 5 with some conclusions.

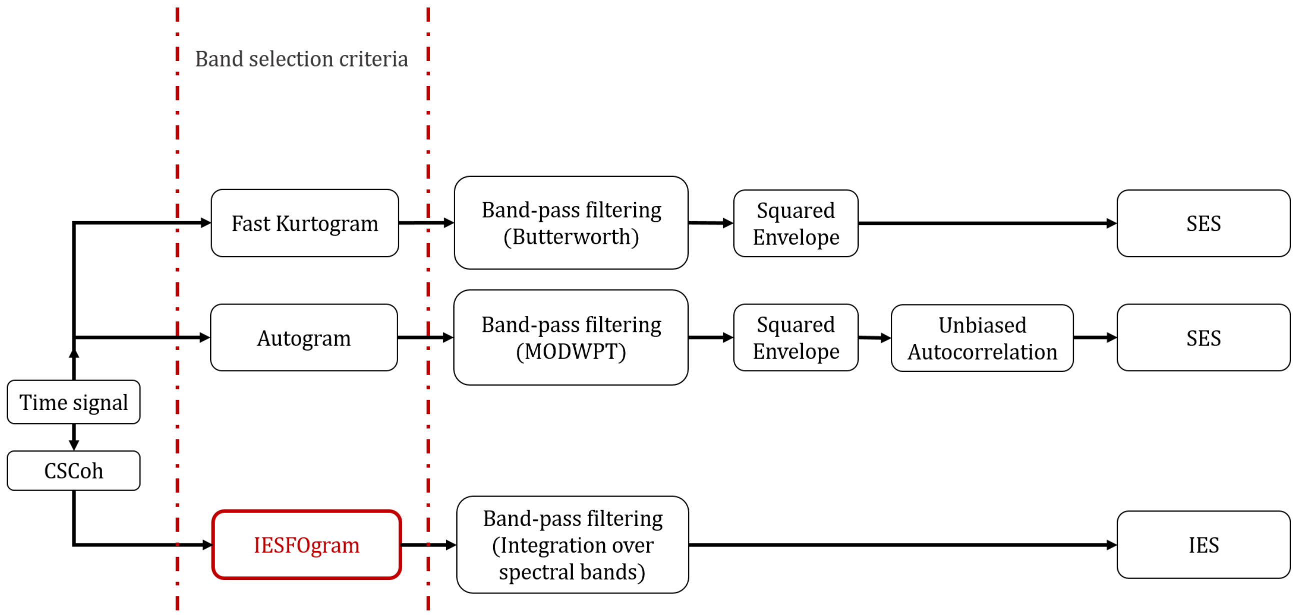


Figure 1: Schematic description of diagnostic methods based on Fast Kurtogram, Autogram, and IESFOgram

2 Cyclostationary signals

Rotating mechanical components are likely to generate cyclic transient signatures which are periodic in nature if the rotational speed is kept constant during the acquisition of signals. These signals often carry information on the health of its components, and signal processing and feature extraction are widely used in order to track the health condition of its components. Following the cyclostationary theory, the signals of interest acquired from rotating machinery can be defined into two orders of cyclostationary signals. Signals of first order of cyclostationarity (CS1) are signals whose first-order statistical moment is a periodic function of T that complies with the condition of Eq. 1.

$$C_{1x}(t) = \mathbb{E}\{x(t)\} = C_{1x}(t+T) \tag{1}$$

where $\mathbb{E}$ denotes the ensemble averaging operator, and t stands for time. In rotating machinery, CS1 vibrations signals are periodic waveforms related to components phase-locked with the rotor speed(*e.g.* shaft misalignment, spalling on meshing gears, etc). A second-order cyclostationary (CS2) signal is a signal whose second order statistical moment is periodic [13]. In particular, if its autocorrelation function is periodic with period T as described in Eq. 2.

$$C_{2x}(t,\tau) = \mathbb{E}\{x(t)x(t-\tau)^*\} = C_{2x}(t+T,\tau) \tag{2}$$

where τ corresponds to the time-lag variable. Bearing vibration signals are often described as CS2, due to having a hidden periodicity related to the shaft speed. Finally, an nth- order cyclostationary (CSn) is a signal whose nth-order statistical moment is periodic, but signal with higher order than CS2 are not taken into account, as CS1 and CS2 describe well the signals of interest generated by rotating machinery.

The Cyclic Spectral Correlation (CSC) is a tool in which the CS1 and CS2 signals are well described in the frequency-frequency domain. The method is represented as a distribution function of two frequency variable: the *cyclic frequency* α linked to the modulation; and the *spectral frequency* f linked to the carrier signal. The tool can be described also as the correlation distribution of the carrier and modulation frequencies of the signatures present in the signals, defined in Eq. 3.

$$CSC(\alpha, f) = \lim_{W\to\infty} \frac{1}{W}\mathbb{E}\{\mathscr{F}_W[x(t)]\mathscr{F}_W[x(t+\tau)]^*\} \tag{3}$$

where $\mathscr{F}_W[x(t)]$ stands the Fourier transform of the signal $x(t)$ over a finite time duration of W. Processing the CSC results in the bi-variable map which reveals the hidden modulations, making it a robust tool for detecting the cyclostationarity in vibration signals [8, 9].

In order to minimize uneven distributions, a whitening operation can be applied to the CSC. This extended tool, named the Cyclic Spectral Coherence (CSCoh), describes the spectral correlations in normalized values between 0 and 1, and is defined as in Eq. 4 :

$$CSCoh(\alpha, f) = \frac{CSC_x(\alpha, f)}{\sqrt{CSC_x(0, f) CSC_x(0, f+\alpha)}} \tag{4}$$

Both the CSC and the CSCoh bi-variable maps can be integrated along the spectral frequency axis in order to obtain a regular spectrum, resulting in a one dimension spectrum function of the $cyclic frequency \alpha$. The band of spectral frequencies to be integrated can be defined as the full available band, from zero to the Nyquist frequency, resulting in a spectrum that exhibits all modulations present in the signal. On the other, the band can be defined as the one that maximizes the cyclic characteristic frequency of interest while minimizing the background noise and the other frequency components that may mask the frequency of interest. In this manner, the integration over a specific band on the bi-variable map can improve the detection rate of the characteristic frequency related to the present damage on the signal. The resulting spectrum is then named Improved Envelope Spectrum (IES) and it is obtained from the frequency-frequency domain according to Eq. 5 :

$$IES(\alpha) = \frac{1}{F_2 - F_1} \int_{F_1}^{F_2} |CSCoh_x(\alpha, f)| df \tag{5}$$

3 Proposed methodology

Diagnosis using the bi-variable maps requires a deep understanding of the map in order to exploit its information. Analysis of one dimensional spectra is far more widely applied in the academia and industry and easier to anaylise. It has been seen that integration of the bi-variable function along its spectral variable results in a one dimensional spectrum, which would be a good tool on itself for diagnostics purposes. On the other hand, the diagnostics information could still be masked under the noise and other components signatures. Integration of the specific band that carries the signal of interest can further enhance the spectrum and increase the performance in the detection of the frequencies of interest. The detection of the optimal band of integration on the bi-variable map is not always straight forward.

The Improved Envelope Spectrum via Feature Optimization-gram (IESFOgram)[11] is one such band selection tool to be applied in the bi-variable map as placed in the scheme depicted Fig. 1.

The proposed method tries to optimize a Normalized Diagnostic Feature (NDF) based on the cyclic characteristics of interest (e.g. rolling element bearing characteristic fault frequencies/orders) on the demodulated spectrum resulting from the integration of the bi-variable map. The method is thought to be general enough to be applied to either the CSC or the CSCoh. The scheme representing the IESFOgram procedure and the extraction of the NDF is shown in Fig. 2, and step-by-step details for its extraction as described as follows.

Step 1: In the first step, the bi-variable map is extracted from the signal. The estimators of the $CSC(\alpha, f)$ can be based on the Averaged Cyclic Periodogram, Cyclic Modulation Spectrum or any other numerical method [14] to extract the CSC bi-variable map previously described in Eq. 3. The CSC can also be in its normalized version $CSCoh(\alpha, f)$. The user can define on its own discretion which method to use in order to obtain the bi-variable map $CSC(\alpha, f)$. The reader is forwarded to the references [15, 16], suggested as providers of the numerical implementation of the CSC in the Order-Frequency domain, and the reference [8] to the Frequency-Frequency domain CSC, if it is applied to the order tracked signal as a function of angle results in the Order-Order domain CSC.

Step 2: The next step consists in dividing the map along the spectral axis f according to the $1/3$-binary tree that is also applied to the Fast Kurtogram [3]. Each is defined by a series with a decreasing bandwidth bw and incremental steps of center frequency cf which define the upper and lower limit f_1 and f_2 described in the integration of Eq. 5. Each band is then integrated and results in a demodulated spectrum $IES_{cf,bw}(\alpha)$.

Step 3: From each processed $IES_{cf,bw}(\alpha)$, one Diagnostic Feature $DF(cf, bw)$ is extracted. This feature is based on the cyclic fault frequency/order of interest. Therefore, to calculate this feature, as well as the IESFOgram, this cyclic component needs to inserted into the method as input. The feature $DF(n)$ is defined as the sum of the N-harmonics of the characteristic fault frequency/order α_{fault} normalized by the noise level estimated in a bandwidth $2 \times f_b$, as described in Eq. 6.

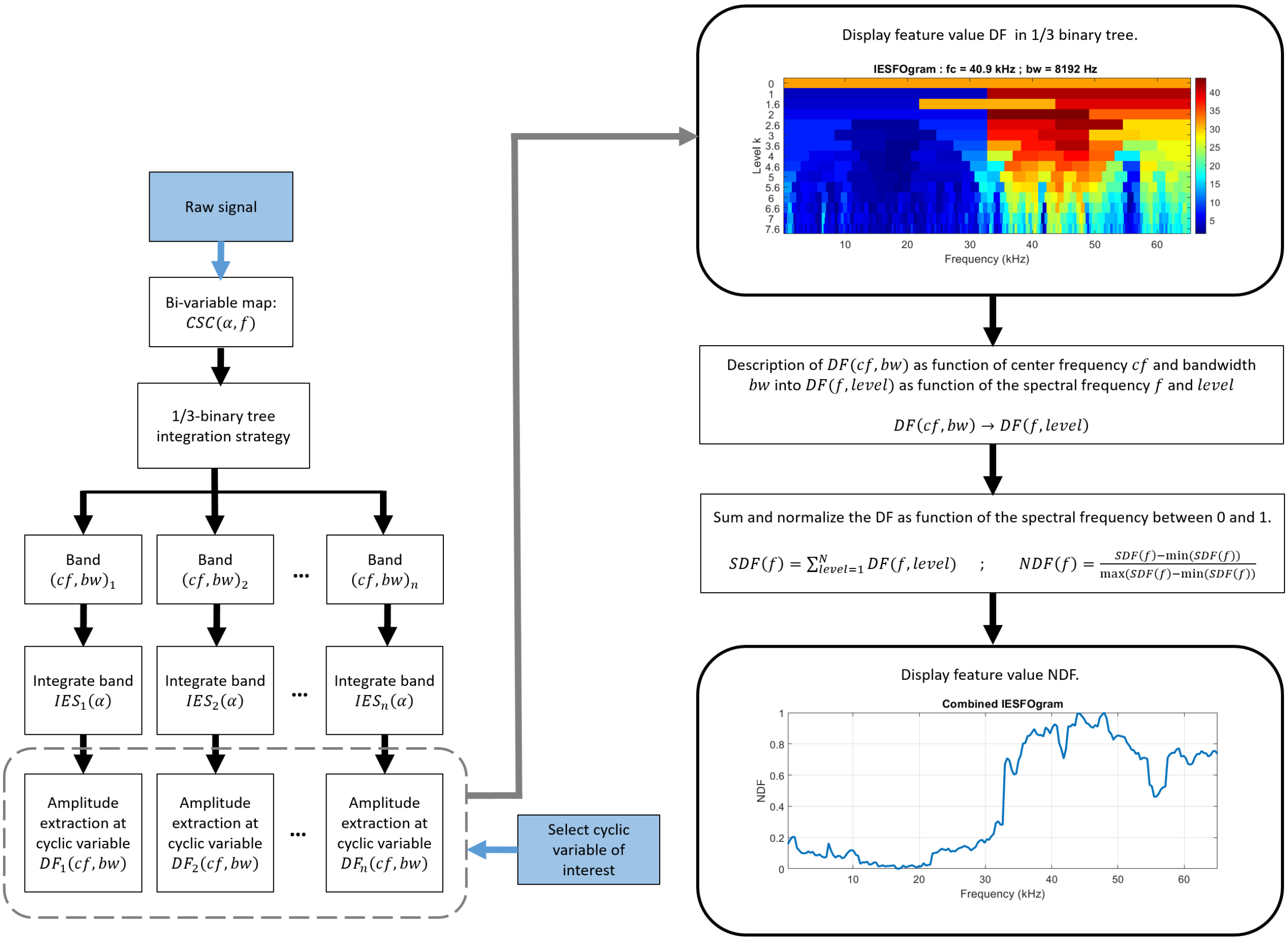


Figure 2: Schematic description of the IESFOgram procedure for extraction of CIES.

$$DF(cf, bw) = \sum_{k=1}^{N} \frac{IES_{cf,bw}(k \times \alpha_{fault})}{\frac{1}{2f_b}\left[\int_{kf_{fault}-f_b}^{kf_{fault}+f_b} IES_{cf,bw}(\alpha)d\alpha - IES_{cf,bw}(k \times \alpha_{fault})\right]} \tag{6}$$

The normalization procedure is important to be taken into account. This is due to some bands having high peak values of noise, and the direct absolute value at the fault frequencies can be higher than at the optimal band. Normalizing with the background noise level at the peaks solves this problem, making high values of the DF to correspond to bands where the frequency peaks of interest are present.

Step 4: The objective of this step is to find the the weight value to be applied on each band for integration of the bi-variable. To quantify the presence of a cyclic component in each band, the library of features $DF(cf, bw)$ is used. The higher the value of DF, the higher the presence of the component of interest is present. Thus, the optimal band OB is identified as the arguments which maximizes $DF(cf, bw)$, as described in Eq. 7.

$$OB = \underset{cf,bw}{\arg\max}[\, DF(cf, bw)\,] \tag{7}$$

Colormap presentation of the values of DF as function of (cf, bw) in a 1/3-binary tree is called the IESFOgram, and its maximum value corresponds to the selected optimal band for integration.

Step 6: The representation of $DF(cf, bw)$ as function of center frequency cf and bandwidth bw is transformed to a representation function of the spectral frequency f and the $level$, as $DF(f, level)$. To be used as a proper weight on the bi-variable map, the $DF(cf, bw)$ is summed along its level ($SDF(cf, bw)$), and then normalized between 0 and 1 ($NDF(cf, bw)$):

$$SDF(f) = \sum_{level=1}^{N} [\, DF(f, level)\,] \tag{8}$$

$$NDF(f) = \frac{SDF - min(SDF)}{max(SDF) - min(SDF)} \tag{9}$$

Step 7: The final step is to integrate the bi-variable map along the spectral axis on and weight each band with normalized diagnostic feature $NDF(f)$ in order to obtain the Combined Improved Envelope Spectrum $CIES(\alpha)$. This step can be considered not be part of the IESFOgram procedure, but as the extraction of the CIES with highest Signal-to-Noise Ratio (SNR) for diagnostic purposes.

$$CIES(\alpha) = \sum_{f=0}^{fs/2} CSCoh(\alpha, f) * NDF(f) \tag{10}$$

To define if the extracted amplitude value on the spectra of the cyclic component are statistically relevant, a threshold is also calculated, and visualized on all spectra of this paper. The threshold is the same as the one presented by the authors in [13], based on 3 times the Moving Absolute Deviation (MAD) of its spectra. The window defined on all spectra corresponds to the total number of samples of each corresponding spectrum divided by 2^7. All values above the threshold are considered to be statistically relevant for detection of the frequency.

4 Experimental application and results

In order to test and validate the proposed methodology, vibration data captured from two separate test rigs are used as case studies. One dataset corresponds to damage on the roller of a bearing under variable speed and load conditions. The second dataset corresponds vibration signals with damage in the outer race of the bearing with high electromagnetic interference present in the signals.

4.1 Case 1 - Roller damage under variable speed and load conditions

The first studied case of this paper is performed on the rolling bearing testing rig developed at the DIRG lab of Politecnico di Torino, where a high-speed spindle drives a shaft supported by a couple of identical bearings. Different damage conditions have been imposed on one of the bearings, and accelerations have been recorded in different positions and directions, under various load and speed conditions. The evaluation of the local damage on a roller has also been evaluated by monitoring the bearing under the same speed and load conditions for about 230 hours. A selection of the acquired records can be downloaded from `ftp://ftp.polito.it/people/DIRG_BearingData/` [10] .

The test rig is composed of three bearings (B1, B2 and B3) mounted inline on the output shaft of a high-speed spindle motor and a precision sledge applying radial load on the bearing on the middle, as depicted in Fig. 3.

The speed of the spindle is set through the control panel of an inverter but can not be actively controlled: not only the spindle has no keyphasor transducer or tachometer to detect its actual speed but also there is no feedback to the controller of the inverter. As a direct consequence, the actual speed of the shaft is always lower than the ideal one and the difference increases with the applied load. A static load cell allows for measuring the resulting force, whose direction is purely radial.

The main geometrical properties of the three bearings, specifically manufactured for this high speed aeronautical applications, are listed in Table 1.

Bearing reference	Pitch diameter D (*mm*)	Rollers diameter d (*mm*)	Rolling elements
$B1$ & $B3$	40.5	9.0	10
$B2$	54.0	8.0	16

Table 1: Main properties of the roller bearings [10].

The dataset corresponding to the variable conditions has 7 cases: three with indentation on the inner ring of the bearing; three with indentation on a roller of the bearing; and one with no damage (healthy). Each damage corresponds to indentation damages diameter of $450, 250$ and $150\mu m$. The case depicted in this paper

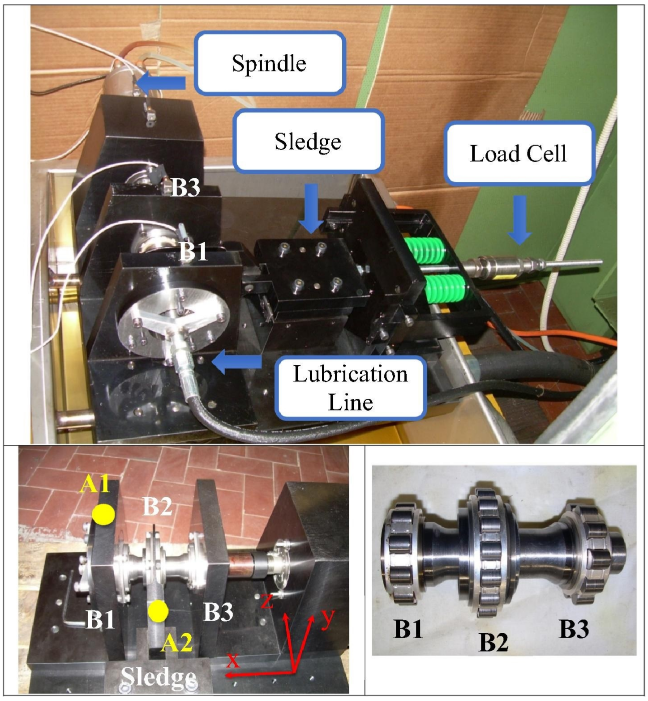


Figure 3: Test rig (up) general view, (left) position of the two accelerometers, (right) shaft with the three accelerometers [10].

corresponds to the damage on the roller with a diameter of $150\mu m$. Each case contains signals acquired during 20 seconds at 51.2 kHz under speeds of: 100, 200, 300, 400 and 500 revolutions per second. Furthermore, the acquired signals are under 4 radial load conditions: 1000 N, 1400 N, 1800 N, and no load. The damaged bearing corresponds to bearing B1, and its characteristic frequencies under the different speed conditions are described in Table 2.

Motor speed (Hz)	FTF (Hz)	2xBSF (Hz)	BPFO (Hz)	BPFI (Hz)
100	38.9	427.8	388.9	611.1
200	77.8	855.6	777.8	1222.2
300	116.7	1283.3	1166.7	1833.3
400	155.6	1711.1	1555.6	2444.4
500	194.4	2138.9	1944.4	3055.6

Table 2: Characteristic frequencies of bearing B1 under different steady speed conditions.

From the two triaxial accelerometers, the signals used to diagnose the roller damage are on the radial direction of the accelerometer mounted on the damaged bearing B1, or in other word, the radial output of accelerometer A1 is used.

The speed reference is provided along with the signals, but the description details that the speed is in reality lower than the provided one. Upon analyzing the signal in the bivariable map from the CSCoh, it is defined that the shaft speed frequency is clear at around 90% of the theoretical one and can be extracted from the spectra under null applied load. Figure 4 shows the CSCoh map exemplifiying the clear shaft frequency at 288 Hz, for the case of speed 300 Hz and a load of 0 N.

When load is applied to the bearing, the FTF harmonics of bearing B1 become the prominent ones, as can be seen from the CSCoh map in Fig. 5 for the case of speed 300 Hz and a radial load of 1000 N.

These peaks at 90% of the given speed and the FTF were used as reference for the real speed of the test rig. With this, the frequencies at the 2xBSF can be correctly defined to determine if they are indeed present in the spectra.

The frequency at 2xBSF and its two next harmonics are then used as inputs to calculate the IESFOgram, and the weight of the different bands that will enhance the peak extraction at those frequencies, as shown in Fig. 6.

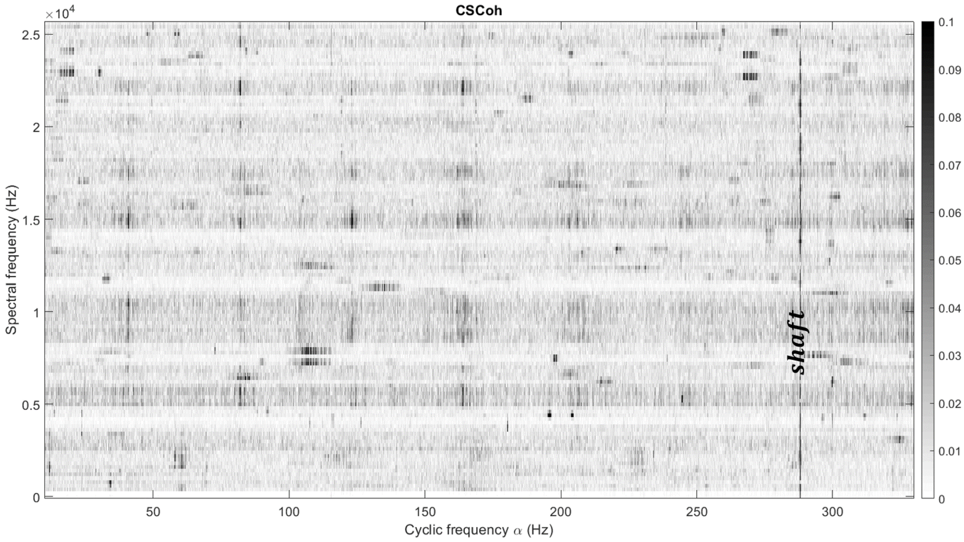


Figure 4: CSCoh bi-variable map around the first 3 FTF harmonics of the damaged bearing and shaft frequency speed of signal under no radial load.

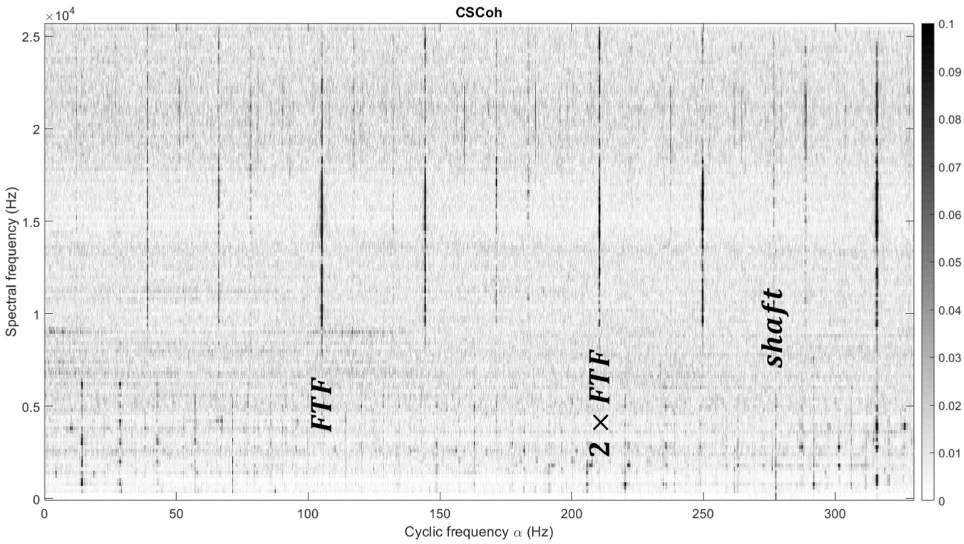


Figure 5: CSCoh bi-variable map around the first 3 FTF harmonics of the damaged bearing and shaft frequency speed of signal under 1000 N of radial load.

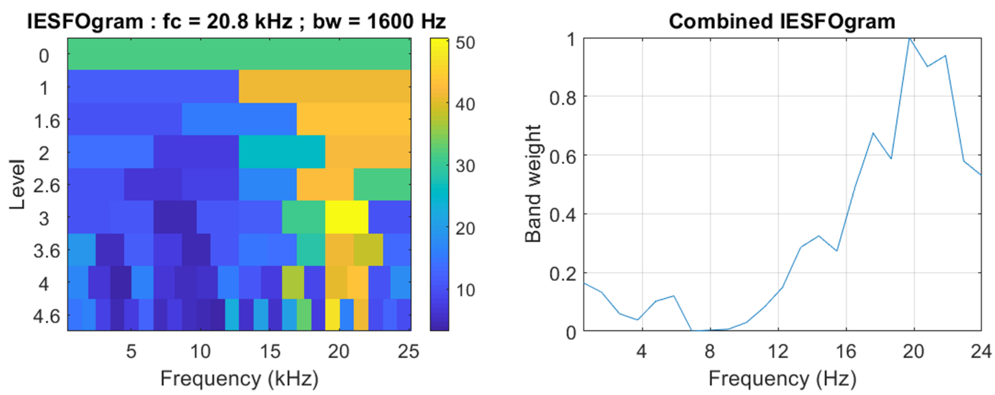


Figure 6: Signal under 300 Hz speed and 0 N load: (left) IESFOgram, (right) Combined IESFOgram.

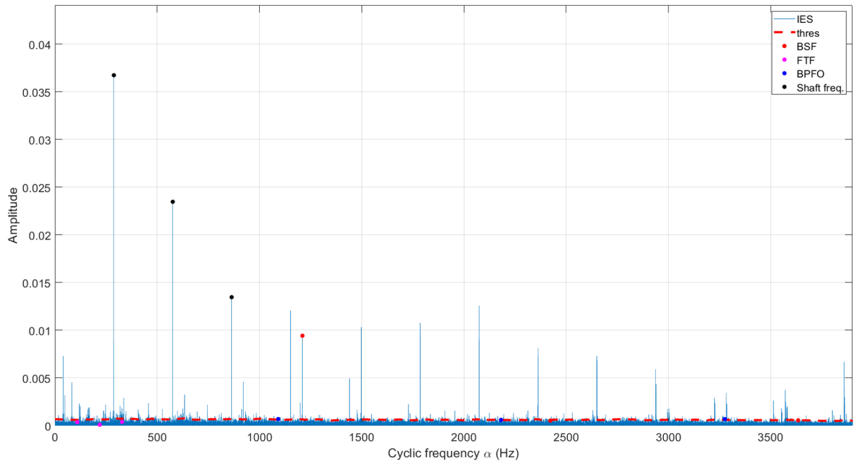


Figure 7: CIES for the signal under 300 Hz speed and 0 N load.

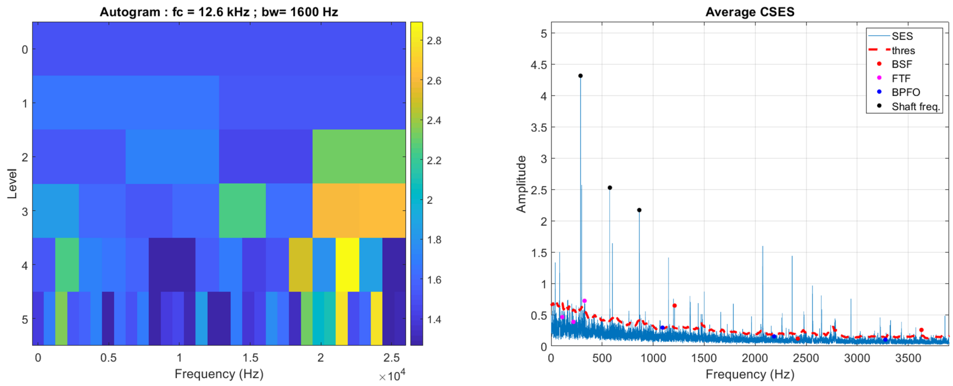


Figure 8: CSES for the signal under 300 Hz speed and 0 N load.

The IESFOgram shows high feature values at high frequencies, mainly around 20 kHz. Applying the resulting weighted function (Combined IESFOgram) to the integration of the CSCoh, the 2xBSF can be clearly extracted from the CIES seen in Fig. 7.

The Autogram applied to the same signal has a similar result, with high feature values for around high frequency bands. The Combined SES based on the Autogram is successful in extracting the peak at the 2xBSF above the threshold, along with the shaft frequency harmonics, as shown in Fig. 8.

The Fast Kurtogram detects the center frequency 10.1 kHz with a bandwidth of 1067 Hz with the highest kurtosis. The resulting SES based on the Fast Kurtogram does not provide any valuable diagnostic information, as illustrated by Fig. 9.

Following to the case of radial load of 1000 N at 300 Hz of shaft speed, the IESFOgram shows high feature values around 12 kHz. Applying the weighted function based on the IESFOgram, the 2xBSF and its harmonics are clearly detected from the Combined IES seen in Fig. 10.

For this case, the radial increases the impulses due to the roller damage, making it easier to detect the fault frequencies. The harmonics of the FTF that modulate the roller characteristic frequency also became dominant under the radial load.

The Autogram applied to the same signal has a similar result, with high feature values for around high frequency bands. The CSES based on the Autogram is also successful in extracting the peak at the 2xBSF above the threshold, as well as the FTF harmonics, as shown in Fig. 11.

In the cases with high load applied, the FTF and 2xBSF harmonics are so dominant on the signals that both the FK-based SES as well as the classical SES with no filtering detect these above the threshold. As a final remark, for the other speeds, the same pattern is found, where in high radial loads the damage related frequencies are dominant while in the no load cases they are masked in the noise level. For avoiding redundancy

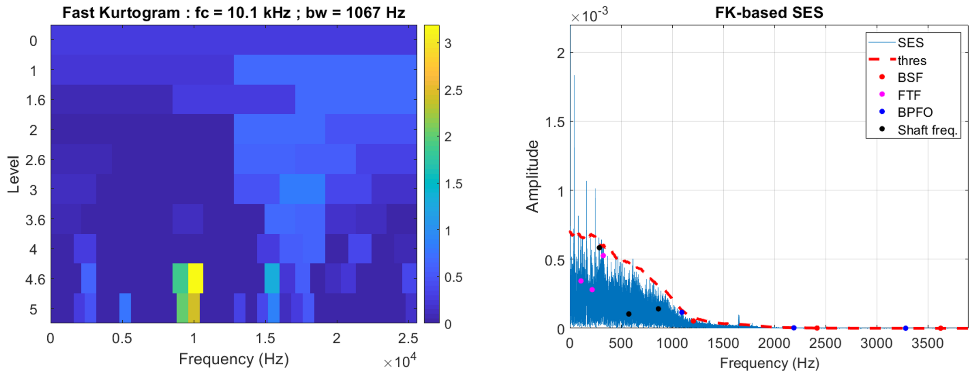


Figure 9: FK-based SES for the signal under 300 Hz speed and 0 N load.

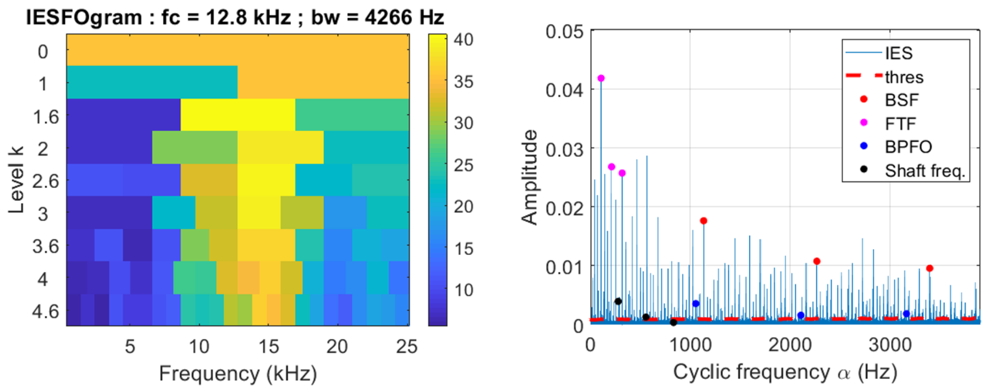


Figure 10: (left) IESFOgram and (right) CIES for the signal under 300 Hz speed and 1000 N of radial load.

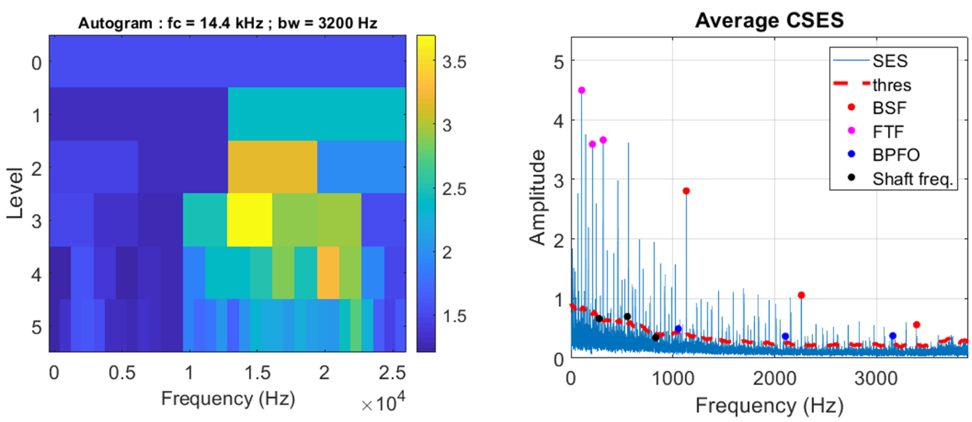


Figure 11: (left) Autogram and (right) CSES for the signal under 300 Hz speed and 1000 N of radial load.

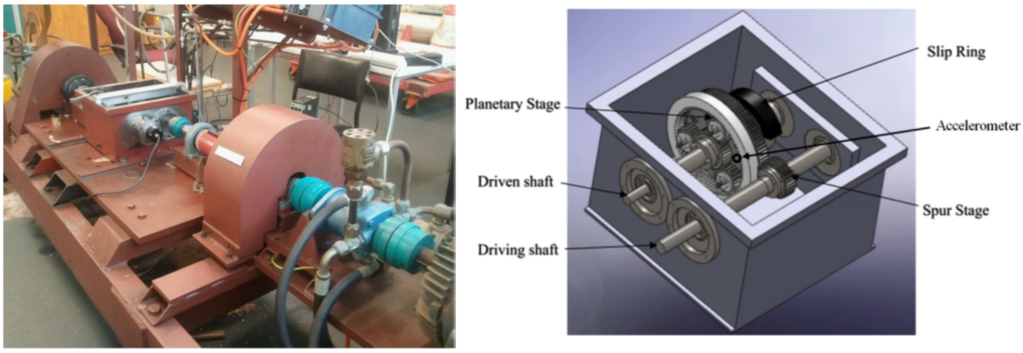


Figure 12: Test rig (left) inside view, (right) schematic representation [4].

on the results, these figure results are not shown here.

4.2 Case 2 - Outer race damage under electromagnetic interference (EMI)

The following case vibration data was acquired from the planetary gearbox test rig with applied torque is presented in Fig. 12. The gearbox torque is provided by a hydraulic system driven by a three-phase induction motor. A torque transducer is attached to measure the applied torque on the gear set while the speed of the driving shaft is controlled by a VFD. The gear ratio of the planetary stage is 1:3 (speed up) consisted of a 40 tooth sun gear, three 20 tooth planetary gears and a 80 tooth ring gear. The planet carrier is the input of the planetary stage, the sun gear is the output while the ring gear is fixed. The overall transmission ratio of the test rig is approximately 1:1 as an initial 90:32 reduction stage is attached. An accelerometer is mounted on the planet carrier to measure acceleration in the axial direction focusing towards the investigation of internal vibration measurements. The vibration signal is finally transmitted to the signal conditioner by the use of a slip ring. Faults have been seeded in the inner race and the outer races of the planet gear bearings using spark erosion. needle roller bearings are used, containing 15 rollers of 2 mm diameters and a pitch diameter of 18 mm. The depth of the faults is 0.4 mm while the width is 1.2 mm for the outer race and 1.0 mm for the inner race respectively. The measurements have been realised for each type of defect (inner and outer race) at a constant input shaft speed of 6 Hz for three torque loads 30, 50 and 70 Nm. The sampling frequency has been selected equal to 131,072 Hz, the switching frequency of the VFD was set at 14 kHz, and the control frequency of the VFD was 24 Hz (giving a nominal 6 Hz shaft speed for the 8-pole motor). The PWM carrier and the PWM message are equal respectively to 14000 Hz and 24 Hz. Based on the geometry of the bearing and its speed, the BPFO is equal to 55 Hz and the BPFI is equal to 69 Hz, but only the BPFO case is demonstrated in this paper.

Applying the Autogram to the signal with outer race damage and with a applied torque of 50 Nm, the band with the maximum Kurtosis with center frequency of 44 kHz and a bandwidth of 2048 Hz is selected as the optimal. This is the correct band with the carrier of the outer race damage. However, the impulsive bands of the carrier of the EMI have also high values of kurtosis, and the resulting CSES shows the peaks at noise level and bellow the threshold, as shown in Fig. 13.

This is the principal obstacle in detecting bearing damage with EMI noise, as both signatures have a impulsive nature which are represented with high kurtosis. The Fast Kurtogram selects the bands related to EMI as the ones with the highest kurtosis, as it as been concluded on this signal by the authors in [4, 12]. In this case, the SES based only on one band with the maximum kurtosis level of the Autogram would provide better performance.

The IESFOgram and its CIES are presented in Fig. 14, and the peaks of the BPFO are seen to be detect with clarity, well above the threshold. The case of BPFI not shown here the IESFOgram extracts the peaks of BPFI, however the CSES version also allows detection of fault.

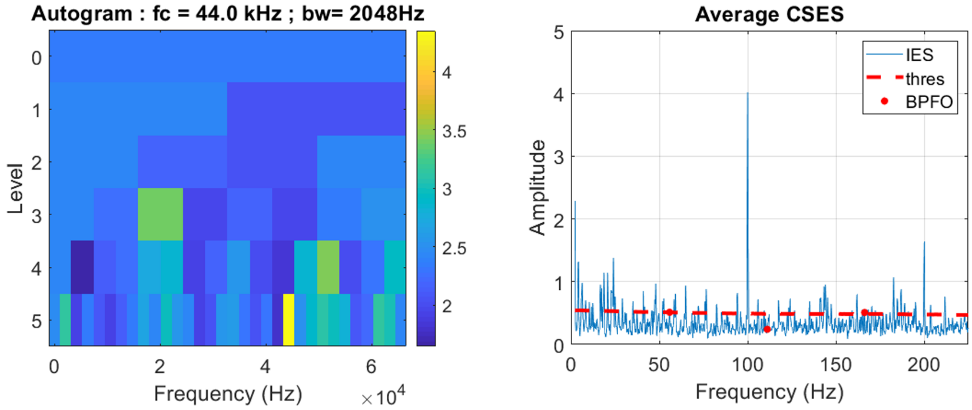


Figure 13: (left) Autogram and (right) CSES for the signal under 50 Nm of torque.

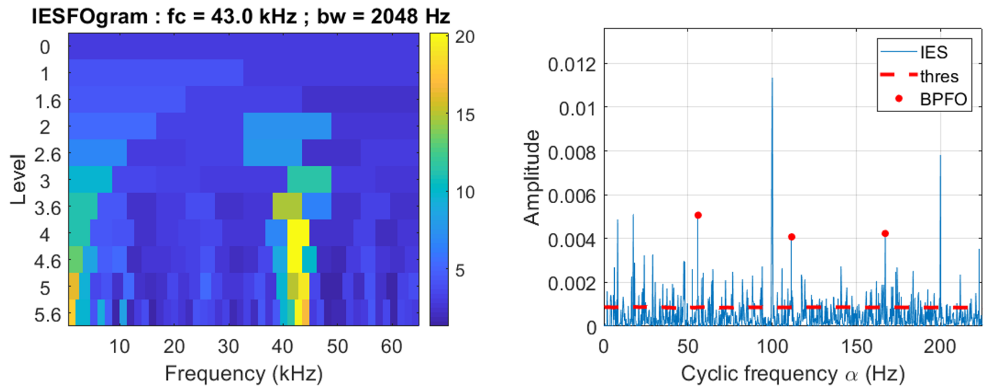


Figure 14: (left) IESFOgram and (right) CIES for the signal under 50 Nm of torque.

5 Conclusion

This paper proposes a new method for demodulation on multi-bands of frequencies using the bi-variable maps based on the Cyclic Spectral Correlation and Coherence. Initially, the IESFOgram related to the damaged bearing characteristic frequency is calculated in order to extract the spectral frequency bands that have high information content on the fault. The Normalized Diagnostic Feature as a function of frequency is defined as the normalized sum of feature values of the IESFOgram. The last step is to perform a multi band integration with the enhanced bands with high Normalized Diagnostic Feature values, and finally obtaining the Combined Improved Envelope Spectrum that allows a correct bearing diagnosis.

The methodology shows good performance in detecting the characteristic frequencies when they are present on the signals, by enhancing the bands with more relevant information on the fault and performing a multi band integration with higher weigthed values on those bands. The method was compared mainly with the Combined Squared Envelope Spectrum based on the Autogram, which also showed higher performance than the Fast Kurtogram on the detection of the bearing characteristic frequencies by also performing a multi band filtering procedure. When the signals contain a high impulsive signal with the same nature as a bearing damage, such as electromagnetic interference, the CSES-based on the Autogram shows to keep the peak frequencies of interest bellow the noise level in case of high impulsive noise spread over several bands. The CIES shows to be able to extract clearly the fault frequency peaks above the noise level while ignoring the bands dominated by the impulsive noise. This is shown to be the case because the IESFOgram is targeted at the frequencies of interest, while the Autogram is blind. Indeed, the main disadvantage of the IESFOgram compared to the Autogram and the Fast Kurtogram is that it is not a blind method.

The methology was tested and validated on two experimental datasets: the first under variable speed and load condition, and the second on signals with high electromagnetic interference contamination. The results show the method can diagnose with confidence bearing damages on vibration signals.

Acknowledgements

The authors would like to thank Ali Moshrefzadeh and Alessandro Fasana for the sharing the roller bearing damage vibration data and the Autogram publicly, Jerome Antoni for sharing the Fast Kurtogram publicly, and Wade Smith and Robert Randall for providing vibration signals with electromagnetic interference. K. Gryllias gratefully acknowledges the Research Fund KU Leuven.

References

[1] Randall, R., Antoni, J., *Rolling Element Bearing Diagnostics - A Tutorial*, Mechanical Systems and Signal Processing, 2012, pp.485-520 (25).

[2] Feldman, M., *Hilbert transform applications in mechanical vibration*, John Wiley and Sons, 2011.

[3] Antoni, J., *Fast computation of the kurtogram for the detection of transient faults*, Mechanical Systems and Signal Processing, 2007, pp.108-124 (21).

[4] Smith, W., Fan, Z., Pheng, Z., Li, H., Randall, R., *Optimised Spectral Kurtosis for bearing diagnostics under electromagnetic interference*, Mechanical Systems and Signal Processing, 2016, pp.371-394 (75).

[5] Tse, P. W., *The sparsogram: A new and effective method for extracting bearing fault features*, Prognostics and System Health Managment Confernece, 2011, pp.1-10).

[6] Moshrefzadeh, A., Fasana, A., *The Autogram: An effective approach for selecting optimal demodulation band in rolling element bearings diagnosis*, Mechanical Systems and Signal Processing, 2017, pp.294-318 (105).

[7] Randall, R., Antoni, J., Chobsaard, S., *The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals*, Mechanical Systems and Signal Processing, 2001, pp.945-962 (15).

[8] Antoni, J., Xin, G., Hamzaoui, N., *Fast computation of the spectral correlation*, Mechanical Systems and Signal Processing, 2017, pp.248-277 (92).

[9] Antoni, J., *Cyclic spectral analysis in practice*, Mechanical Systems and Signal Processing, 2007, pp.597-630 (21).

[10] Daga, A. P., Fasana, A., Marchesiello, S., Garibaldi, L., *The Politecnico di Torino rolling bearing test rig: Description and analysis of open access data*, Mechanical Systems and Signal Processing, 2019, pp.252-273 (120).

[11] Mauricio, A., Smith, W., Qi, J., Randall, R., Gryllias, K., *Cyclo-non-stationarity based bearings diagnostics of planetary bearings*, CMMNO, Santander, Spain, 2018, pp.1-10.

[12] Mauricio, A., Qi, J., Sarrazin, M., Janssens, K., Smith, W., Randall, R., Gryllias, K., *Cyclostationary-based bearing diagnostics under electromagnetic interference*, ICSV25, Hiroshima, Japan, 2018, pp.1-8.

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

[14] Borghesani, P., Antoni, J., *A faster algorithm for the calculation of the fast spectral correlation*, Mechanical Systems and Signal Processing, 2018, pp.113-118 (111).

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

[16] Abboud, D., Antoni, J., *Order-frequency analysis of machine signals*, Mechanical Systems and Signal Processing, 2017, pp.229-258 (87).

Data Mining Classification & Machine Learning methods

Deep Learning Protocol for Condition Monitoring & Fault Identification in a Rotor-Bearing System from Raw Time-Domain Data

Sonkul Nikhil Avinash Singh Jasdeep Vyas Nalinaksh. S.

Department of Mechanical Engineering
Indian Institute of Technology, Kanpur 208016, India
nasonkul@gmail.com, vyas@iitk.ac.in

Abstract

A Deep Learning protocol is developed for identification of typical faults occurring in rotating machinery. In past Support Vector Machines (SVMs), Clustering, Artificial Neural Networks (ANNs) and other algorithms have been used for this purpose. However, these algorithms require the raw time-domain data, from the sensors on the machine, to be first processed and handcrafted into parameters like Fast Fourier Transform (FFT) coefficients, Statistical Moments, etc. before being fed as inputs. ANN with back-propagation is a popular algorithm in many domains. It however suffers from the vanishing gradient problem and not adequate enough when subjected to raw sensor response as input. Deep Learning or Deep Neural Network is understood as a form of neural network with a large number of layers. Convolutional Neural Network (CNN) architecture is commonly used in a deep neural network for image recognition. A Deep Learning CNN architecture has been developed, employing the analogy of an RGB image, to directly work upon the raw time-domain signals obtained from sensors on a rotor-bearing system. The analogous RGB channels are vibration data from different sub-systems of the complete rotor-bearing assembly. The Deep Learning Network effectively recognizes all kinds of faults that were investigated.

1 Introduction

Data collection, feature extraction and fault identification are typical steps in fault diagnosis of a mechanical system. Conventional approaches in fault diagnosis extract features from time and frequency domain of raw signals. Statistical parameters and Fast Fourier Transform coefficients are widely used features. Machine learning techniques are used with these handcrafted features to identify faults. Fuzzy Logic, Wavelets, Clustering, Decision Trees, Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are some of the techniques used in past for fault identifications [1]–[7]. Statistical moments of raw time domain signal and its derivatives were used as features of ANN by Vyas [7] for fault diagnosis of rotor-bearing system. Efforts to automate the feature extraction using deep learning techniques are being made in recent studies [8]. Janssensa [9] used Discrete Fourier Transform coefficients as input to Convolutional Neural Network (CNN) for bearing fault classification. Features learnt from CNN network combined with time domain features were used by Xie [10] to train SVM model. Guo [11] trained a CNN network with input as continuous wavelet transform scalogram of rotor machinery.

ANNs are not adequate to abstract features from raw time domain data due to large dimension of the data and vanishing gradients. The method proposed in this study, directly works upon raw time domain data eliminating the pre-requisite of extracting features.

Deep Neural Network, like CNN is capable of learning abstract features from large and multi-dimensional data like images and audio signals. CNN preserves the topology of the input and has lesser number of learning parameters than a neural network of the same depth, which also makes learning faster.

2 Methodology and Training Data

CNN is chosen as primary model in this study. As a multi sensor model is necessary to identify faults in large systems, a Multi-channel Convolution Neural Network (McCNN) is incorporated in the proposed model.

2.1 Network Architecture

Input layer, Convolution layer, Pooling layer and Fully Connected layer form the building blocks of a CNN. The architecture is described in Figure 2.1. The input layer is a multi-dimensional array like an RGB image. Convolution is the dot product of a Kernel with a part of input or previous layer. The topology of a Kernel is smaller than previous layer. Kernels are weight parameters of CNN and they traverse over the input space and generate a feature map. A Kernel of size m x n after convoluting over input of size M x N will result into feature map of size (M-m+1) x (N-n+1). Number of Kernels and Size of Kernel are the hyperparameters for particular convolutional layer. The Feature Map is then activated by an Activation Layer. ReLU Activation Function is used in CNN as it reduces vanishing gradients. The Feature Map is sometimes padded with zeroes to control the topology of further feature maps. Batch Normalization Layer is added between Convolution and ReLU Activation Layer. It normalizes the feature map after Convolution layer, firstly by subtracting the mini-batch mean from each of its inputs and then dividing by their standard deviation for each channel; and secondly by scaling the new obtained featured map by γ and then shifting by β, where γ and β are learnable parameters. Batch Normalization accelerates deep network training [12]. Pooling layer summarizes the response over a neighbourhood. It reduces the output size and makes features invariant to small input noise. Max Pooling, Average Pooling are widely used pooling operations. The size of the region is the hyperparameter. Finally, softmax function is used for fault classification.

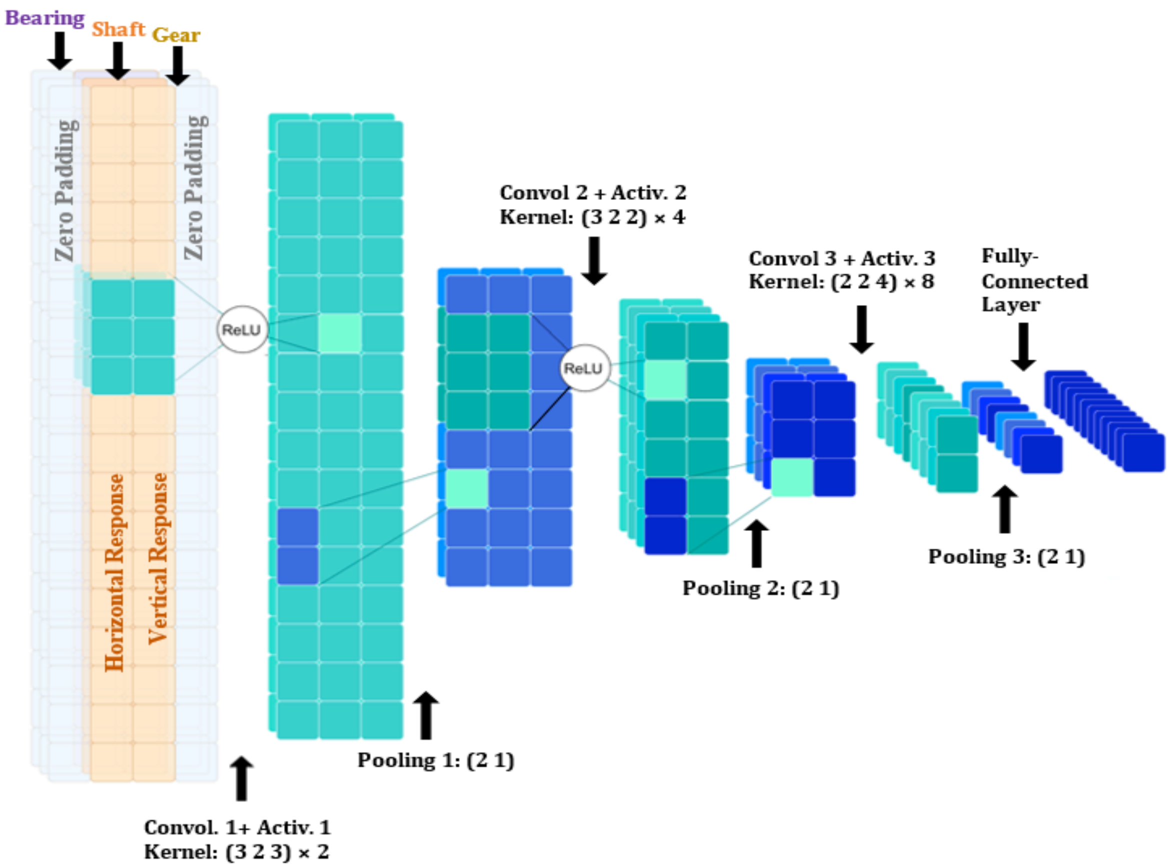


Figure 2.1: CNN Architecture: Input matrix is made with responses from three sub-systems corresponding to three channels with two sensors response (horizontal & vertical) in each channel.

2.2 Experiment Setup & Training Data

Experiments were performed by Jasdeep Singh [13] on a rotor test rig (Machine Fault Simulator Figure 2.2), which consists of a shaft supported in two roller bearings and driven by a DC motor. A flexible coupling is used to connect rotor shaft to that of motor. At one end of the shaft there is a sheave, which is connected to a reciprocating mechanism through a belt drive and a gearbox. Two discs mounted on the shaft.

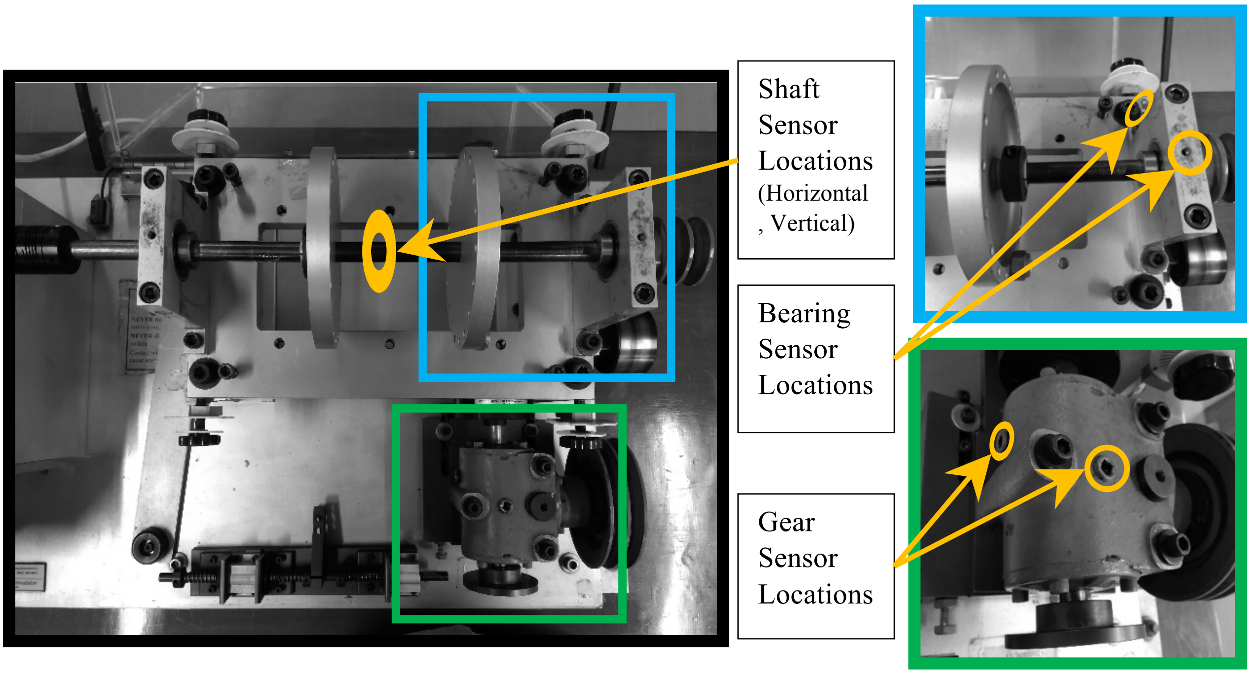


Figure 2.2: The rotor system (Machine fault Simulator) and sensor locations

The set-up comprises subsystems - (i) Gearbox (ii) Bearing and (iii) Shaft-Rotor Disc. Accelerometers were used at (i), (ii) and Proximity pickups at (iii). Each location has two sensors oriented mutually perpendicular to each other (horizontal & vertical). Data was collected at 2560 samples/sec. For an experiment data is collected for duration of 1.6 sec i.e. 4096 sample points in time domain. Variety of faults (Table 2.1) were introduced at a constant operating frequency of 40 Hz. Typical vibrations sensed for one of the faults introduced in the rotor are shown in Figure 2.3. Twenty such signals were collected for each sensor for the fault introduced.

	Faults (Set A)
1	No Fault with Attached Belt
2	Loose Belt
3	Tight Belt
4	Missing Tooth
5	Loose Gear
6	Unbalance 6.82 gram
7	Eccentric Rotor
8	Cocked Rotor
9	Bearing Outer Race Defect
10	Bearing Inner Race Defect
11	Ball Spin Fault
12	Combined Bearing Fault

	Faults (Set B) *Combined with missing tooth*
13	Unbalance 6.82 gram
14	Eccentric Rotor
15	Cocked Rotor
16	Bearing Outer Race Defect
17	Bearing Inner Race Defect
18	Ball Spin Fault
19	Combined Bearing Fault
20	Loose Gear

Table 2.1: Type of faults in dataset (Set A). Extra combined faults and (Set A) form (Set B)

Twelve faults (Set A) are described in Table 2.1. Set B is a combination of Set A and combined faults. Set A has 240 and Set B has 400 experimental responses. The datasets are divided between - training (60%), validation (20%) and test (20%) datasets.

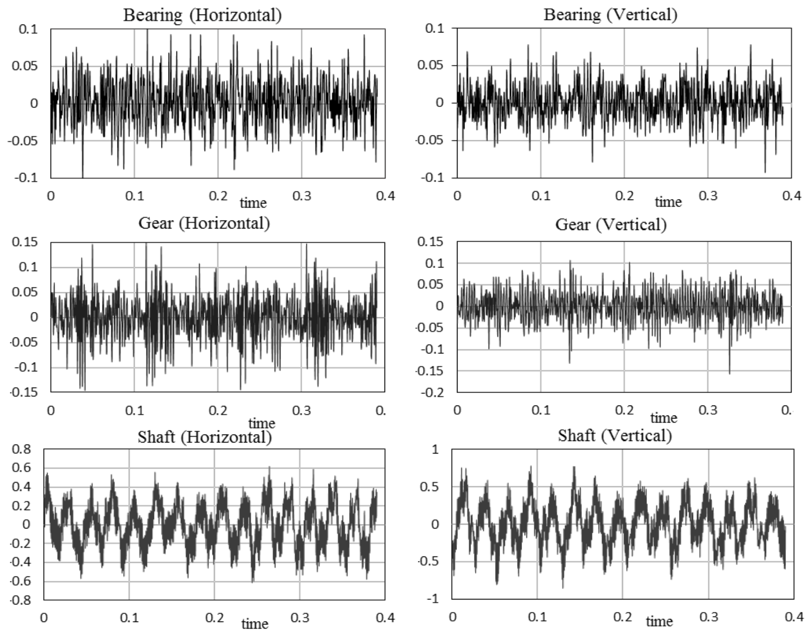


Figure 2.3: Typical sensor data

2.3 Network Topology

As described earlier (Figure 2.1), the response of all six sensors is mapped between normalised values of 0 and 1, on a McCNN, a multi-channel input similar to an RGB image. The two (horizontal and vertical) sensors of a sub-system comprise a channel and three such sub-systems complete the multi-channel input to the CNN. The topology of input, therefore is, 4096x2x3. Each channel is analogous to an RGB channel of a colored image. Each channel is padded with zero columns on both sides as shown in Figure 2.1. This input is passed down to further layers as listed in Table 2.2.

#	*Layer*	*Kernel Size/ Pooling Region*	*# of Kernels*	*Padding*
1	Input	-	-	-
2	Convolution 1	40 × 2	8	Yes
3	Batch Normalization + ReLU	-	-	-
4	Pooling 1	3 × 1	-	-
5	Convolution 2	20 × 2	16	Yes
6	Batch Normalization + ReLU	-	-	-
7	Pooling 2	4 × 1	-	-
8	Convolution 3	4 × 2	32	No
9	Batch Normalization + ReLU	-		-
10	Pooling 3	3 × 1	-	-
11	Convolution 4	4 × 2	32	No
12	Batch Normalization + ReLU	-	-	-
13	Pooling 4	2 × 1	-	-
14	Convolution 5	2 × 2	32	No
15	Batch Normalization + ReLU	-	-	-
16	Pooling 5	2 × 1	-	-
17	Convolution 6	2 × 1	32	No
18	Batch Normalization + ReLU	-	-	-
19	Pooling 6	2 × 1	-	-
20	Fully Connected	12 Neurons	-	-

Table 2.2: CNN architecture and hyperparameters

Network training is performed with stochastic gradient descent algorithm with momentum. The network was limited to a maximum of 800 iterations. Learning rate was 0.01. Mini-batch size was kept as 64 for Set A training and 128 for Set B. Both, max and average pooling were used in training for comparison.

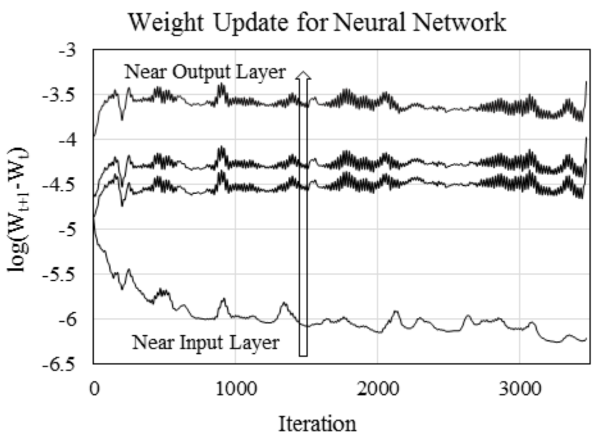
3 Training and Validation

Training of McCNN is performed, both, on Set A and Set B. The architecture classified all training, validation and test dataset successfully for both Set A (individual faults) and Set B (combined faults). The accuracy increases as we increase number of layers hence depth of network, (Table 3.1). Variation of accuracy with pooling function is also studied. Average Pooling network required less number of layers in comparison to Max Pooling to achieve zero error.

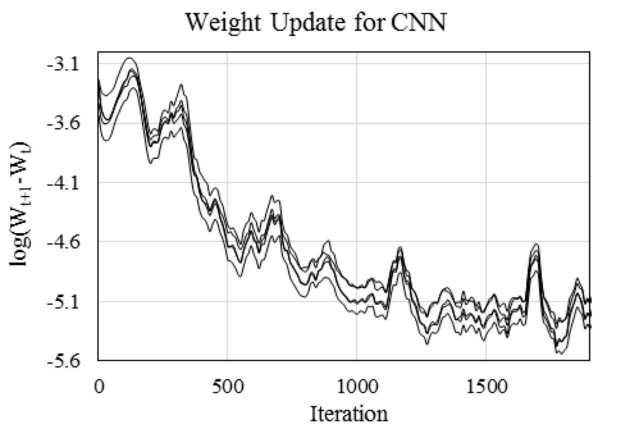
Network Depth	*Set A (Test Accuracy %)*		*Set B (Test Accuracy %)*	
	(Max. Pooling)	(Avg. Pooling)	(Max. Pooling)	(Avg. Pooling)
Layer [1 to 4] + 20	0.00	39.00	5.00	33.00
Layer [1 to 7] + 20	50.00	29.00	23.00	16.00
Layer [1 to 10] + 20	81.00	95.00	97.00	92.00
Layer [1 to 13] + 20	95.00	100.00	98.00	98.00
Layer [1 to 16] + 20	100.00	100.00	98.00	100.00
Layer [1 to 19] + 20	-	-	100.00	100.00

Table 3.1: Accuracy variation with depth and pooling function

A comparison between traditional Neural Network and McCNN is also made to study the weights updation with training iteration. From Figure 3.1 it can be seen that traditional Neural Network shows the effect of vanishing gradient, while in McCNN the weight updation of layers distant from output layer are of nearly same order compare to proximate layers. Complex and large systems can be diagnosed in McCNN with significant depth without facing vanishing gradient.



(a) traditional ANN



(b) McCNN

Figure 3.1: Weight updation

It is seen that better results are obtained when training is carried out with data from all sub-systems (row number 4 in Table 3.2) in comparison to those obtained through training by isolated data from an individual sub-system (row numbers 1-3 in Table 3.2). This observation is visible for Set B, which contains more and combined faults, which increase the complexity of task, and underlines the fact that for fairly large and complex system we need sensor data from multiple critical locations.

Network Input	*Input Layer Topology*	*% Error (Set A)*			*% Error (Set B)*		
		Training	Validation	Test	Training	Validation	Test
Bearing	4096 × 2 × 1	0.00	0.00	0.00	0.00	7.50	3.75
Gear	4096 × 2 × 1	0.00	0.00	0.00	0.00	1.25	1.25
Shaft	4096 × 2 × 1	0.00	0.00	0.00	0.00	0.00	1.25
Bearing + Gear + Shaft	4096 × 2 × 3	0.00	0.00	0.00	0.00	0.00	0.00

Table 3.2: Training with input from individual Sub-Systems and all sub-systems combined

Figure 3.2, shows features obtained after each ReLU activation layer, after they are transformed to two dimensions using t-SNE dimensionality reduction algorithm. The fault classification process can be seen to be consolidated at each ReLU layer where fault data points are getting clustered more compactly and distant from other faults. The distinction and separation between faults increases as we move toward output layer.

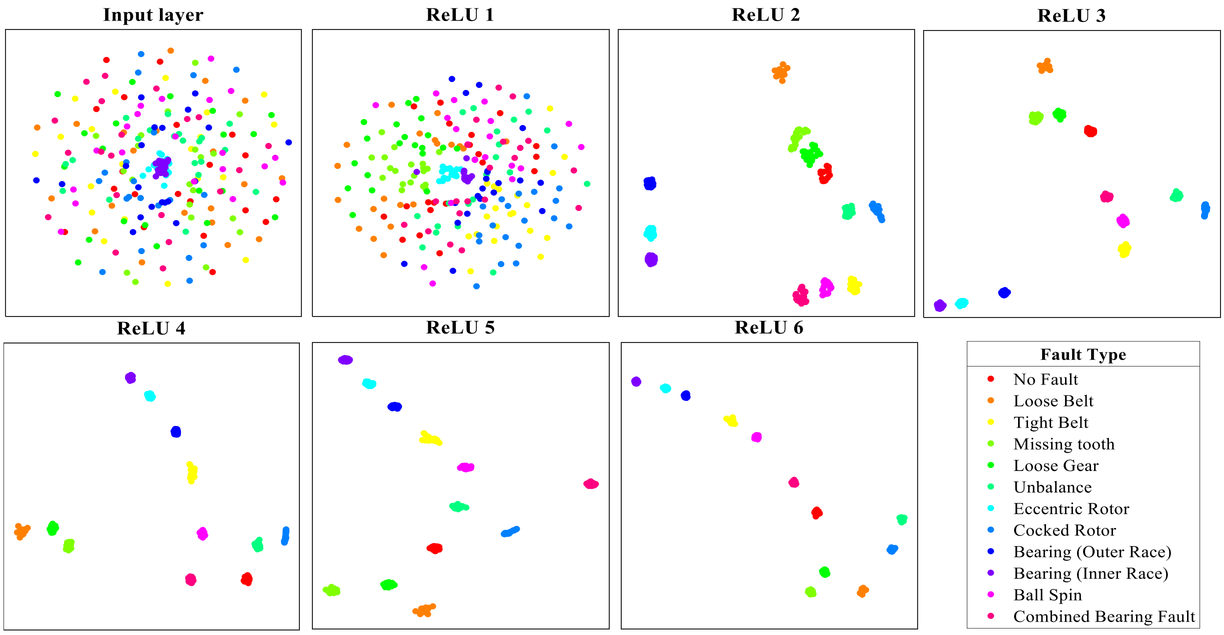


Figure 3.2: Feature map after each ReLU layer, plotted t-SNE (Set A).

4 Conclusion

The present study introduces an approach for fault identification of a rotor-bearing system using Convolutional Neural Network. The need to extract features like statistical parameter or to convert time domain data to FFT domain for training is eliminated in present study. Raw time domain sensor response is used without any pre-processing or feature engineering for constructing input layer. A generic method of fault diagnosis for big complex systems with distant and multiple subsystems is presented, which merges sensor responses from demanding subsystems into distinct channels of input matrix to form the input layer of McCNN architecture. Patterns in raw time domain data, difficult to be comprehended manually, are perceived by the McCNN network.

References

[1] M. Boumahdi, J. P. Dron, S. Rechak, and O. Cousinard, *On the extraction of rules in the identification of bearing defects in rotating machinery using decision tree,* Expert Syst. Appl., vol. 37, no. 8, pp. 5887–5894, 2010.

[2] Y. Lei, Z. He, Y. Zi, and X. Chen, *New clustering algorithm-based fault diagnosis using compensation distance evaluation technique,* Mech. Syst. Signal Process., vol. 22, no. 2, pp. 419–435, 2008.

[3] R. Yan, R. X. Gao, and X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications,* Signal Processing, vol. 96, no. PART A, pp. 1–15, 2014.

[4] M. Systems, *Objective Machinery Fault Diagnosis Using Fuzzy Logic,* Mech. Syst. Signal Process., vol. 12, pp. 855–862, 1998.

[5] C. T. Yiakopoulos, K. C. Gryllias, and I. A. Antoniadis, *Rolling element bearing fault detection in industrial environments based on a K-means clustering approach,* Expert Syst. Appl., vol. 38, no. 3, pp. 2888–2911, 2011.

[6] L. B. Jack and A. K. Nandi, *Support vector machines for detection and characterization of rolling element bearing faults,* Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci., vol. 215, no. 9, pp. 1065–1074, 2001.

[7] N. S. Vyas and D. Satishkumar, *Artificial neural network design for fault identification in a rotor-bearing system,* Mech. Mach. Theory, vol. 36, no. 2, pp. 157–175, 2001.

[8] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, *Deep Learning and Its Applications to Machine Health Monitoring: A Survey,* vol. 14, no. 8, pp. 1–14, 2016.

[9] O. Janssens et al. *Convolutional Neural Network Based Fault Detection for Rotating Machinery,* J. Sound Vib., vol. 377, pp. 331–345, 2016.

[10] Y. Xie and T. Zhang, *Fault Diagnosis for Rotating Machinery Based on Convolutional Neural Network and Empirical Mode Decomposition,* Shock Vib., vol. 2017, pp. 1–12, 2017.

[11] S. Guo, T. Yang, W. Gao, and C. Zhang, *A novel fault diagnosis method for rotating machinery based on a convolutional neural network,* Sensors (Switzerland), vol. 18, no. 5, 2018.

[12] S. Ioffe and C. Szegedy, *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,* In ICML 2015.

[13] J. Singh, *Condition Monitoring of Geared Rotor Systems using Neural Networks,* Indian Institute of Technology Kanpur, 2006.

WIND TURBINE GEARBOXES FAULT DETECTION THROUGH ON-SITE MEASUREMENTS AND VIBRATION SIGNAL PROCESSING

Francesco Castellani[1], Davide Astolfi[1], Luigi Garibaldi[2], Alessandro Paolo Daga[2]

[1]University of Perugia, Department of Engineering, Perugia (Italy)
[2]Politecnico di Torino, Department of Mechanical and Aerospace Engineering, Torino (Italy)
francesco.castellani@unipg.it, davide.astolfi@unipg.it,
luigi.garibaldi@polito.it, alessandro.daga@polito.it

Abstract

Condition monitoring of gear-based mechanical systems undergoing non-stationary operation conditions is in general very challenging. In particular, this issue is remarkable as regards wind turbine technology because most of the modern wind turbines are geared and gearbox damages account for at least the 20% of unavailability time. For this reason, wind turbines are often equipped with condition monitoring systems (CMS), processing vibration measurements collected at meaningful subcomponents of the gearbox. In this work, a novel approach for the diagnosis of gearbox damages is proposed: the turning point is that vibration measurements are collected at the tower, instead that at the gearbox and can be performed also for machine not provided with specific CMS. This implies that measurement campaigns are quite easily performed and repeatable, also for wind turbine practitioners, and that there is no impact on wind turbine operation and power production. A test case study is discussed: it deals with a wind farm owned by Renvico, featuring 6 wind turbine with 2 MW of rated power each. The vibration measurements at two wind turbines suspected to be damaged and at reference wind turbines are processed through a multivariate Novelty Detection algorithm in the feature space. The application of this algorithm is justified by univariate statistical tests on the time-domain features selected and by a visual inspection of the dataset via Principal Component Analysis. Finally, the novelty indices based on such time-domain features, computed from the accelerometric signals acquired inside the turbine tower, prove to be suitable to highlight a damaged condition in the wind-turbine gearbox, which can be then successfully monitored.

1 Introduction

The diagnosis of gears and bearings faults of gearbox systems [1] is a very important topic, especially if the gear-based mechanical system of interest undergoes non-stationary operation conditions.

The technology of most of the modern wind turbines is based on the transformation of the slow rotor rotational speed (order of 10 revolutions per minute) into the fast generator rotational speed through a gearbox. It is estimated [2] that the unavailability time of a large wind turbine operating in an industrial wind farm is of the order of the 3% and at least the 20% of this quantity is due to gearbox damages. For this reason, therefore, the improvement in gearbox condition monitoring is a crucial step for the target of 100% availability of wind turbines. Therefore, commonly, megawatt-scale wind turbines are equipped with condition monitoring systems, elaborating the vibration measurements collected at meaningful sub-components of the gearbox.

Nevertheless, in the wind energy practitioners community, gearbox vibration data are often under-exploited because of the complexity of the analysis techniques that are required in order to interpret them. Therefore, often it happens that oil particle counting and operation data analysis (especially temperatures, as in [3]) are employed as condition monitoring techniques, despite they provide a late stage fault diagnosis, with respect to vibration analysis.

Therefore, two can be important direction targets as regards wind turbine gearbox condition monitoring through vibration analysis: on one hand, the precision of the diagnosis and on the other hand the simplicity of the methods. As regards the former aspect, there are several recent studies. In [4], data mining algorithms and statistical methods are applied to analyze the jerk data obtained from monitoring the gearbox of a wind turbine: the failed stages of the gearbox are identified in time-domain analysis and frequency-domain analysis. In [5], the proposed techniques are based on three models (signal correlation, extreme vibration, and RMS intensity) and have been validated with a time-domain data-driven approach using condition monitoring data of wind turbines in operation. The results of that study support that monitoring RMS and extreme values serves as a leading indicator for early detection. In [6], the focus is on separating the bearing fault signals from masking signals coming from drivetrain elements like gears or shafts. The separation is based on the assumption that signal components of gears or shafts are deterministic and appear as clear peaks in the frequency spectrum, whereas bearing signals are stochastic due to random jitter on their fundamental period and can be classified as cyclo-stationary [7]. In [8], order analysis is individuated as a useful technique for condition monitoring the planetary stage of wind turbine gearbox. The approach takes advantage of angular resampling to achieve cyclo-stationary vibration signals and lessen the effects due to speed changes. In [9], the objective is condition monitoring of the planetary stage of wind turbine gearboxes: the proposed technique is resampling vibration measurements from time to angular domain, identification of the expected spectral signature for proper residual signal calculation and filtering of any frequency component not related to the planetary stage.

On the grounds of this brief literature survey, it arises that the techniques for the analysis of cyclo-stationary signals are the most employed for an accurate condition monitoring. The type and the quality of data that are requested for this kind of analysis confines the subject mainly to the scientific community and at present discourages the collaboration between industry and academia. Actually, the commercial condition monitoring systems, that are mostly adopted in most operating wind turbines exploited at the industrial level, record vibration measurements only when some trigger events occur and, most of all, don't stock the raw data (commonly, Fourier transforms and-or simple statistical indicators are stocked).

On these grounds, there is a growing demand of vibration-based gearbox condition monitoring techniques that could be easily repeatable, without impacting on the wind turbines operation (i.e. without intruding in the gearbox), and whose interpretation could be sufficiently simple and powerful. One remarkable study by this point of view is [10], where sound and vibration measurements collected at the wind turbine towers are employed for condition monitoring of generators. Tower vibration signals are analyzed using Empirical Mode Decomposition (EMD) and the outcomes are correlated with the vibration signals acquired directly from the generator bearings. It is shown that the generator bearing fault signatures are present in the vibrations from the tower.

This study is devoted to the test case of two multi-megawatt wind turbines sited in Italy, owned by Renvico (a company managing around 340 MW of wind turbines in Italy and France, www.renvicoenergy.com). The wind turbines are not equipped with gearbox condition monitoring systems and they have been diagnosed of gearbox damages (of different severity) through the analysis of oil particle counting. Before the gearboxes replacement intervention, a measurement campaign has been conducted by the University of Perugia. The idea is measuring vibrations at the tower: the measurements are collected on the target damaged wind turbines and on one (or more) reference undamaged wind turbines. Subsequently, the data are processed through a multivariate Novelty Detection algorithm in the feature space. The application of this algorithm is supported by statistical analysis on the time-domain features selected. Finally, the novelty indexes based on such time-domain features prove to be suitable to diagnose a damaged condition. It should be noticed that the obtained results allow distinguishing between the two target wind turbines and the corresponding different severity of the gearbox damages.

The manuscript is organized as follows: in Section 2, the test case wind farm, the measurement techniques and equipment and the obtained data sets are described. Section 3 is devoted to the data analysis, feature extraction and results discussion. Finally, in Section 4 some concluding remarks and further directions of this study are indicated.

2 The on-site measurements and the data sets

The wind farm is composed of six multi-megawatt wind turbines and it is sited in southern Italy. The layout of the wind farm is reported in Figure 1, where the damaged wind turbines are indicated in red. The lowest inter-turbine distance on site is of the order of 7 rotor diameters.

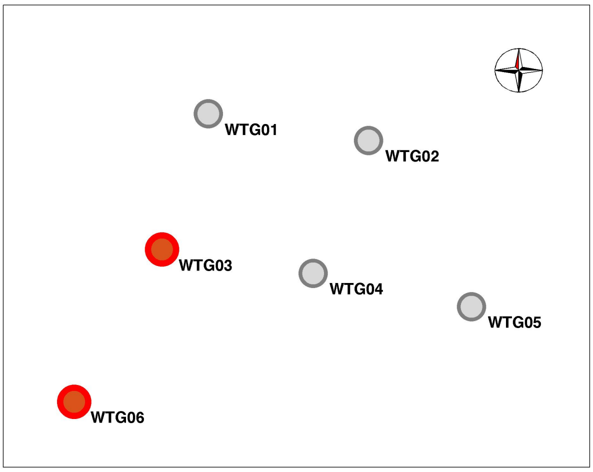


Figure 1: The layout of the wind farm. In red, the damages wind turbines (WTG03 and WTG06) are indicated.

It should be noticed that the damages to WTG03 and WTG06 have different level of severity: actually, the damage at WTG06 was detected through oil particle counting some days before the measurement campaign, while the damage at WTG03 can be considered at incipient stage.

The measurements are conducted as follows: accelerometers are mounted inside the tower of the wind turbine. They measure the longitudinal (x-axis) and transversal (y-axis) vibrations, as displayed in Figure 2. An overall set of four accelerometers (respectively two on the superior level 7 m above ground and two at the inferior level 2 m above ground) and a microphone (on the inferior level) were used for the acquisition. Each acquisition therefore consists of 4 channels sampled at 12.8 kHz for 2 minutes.

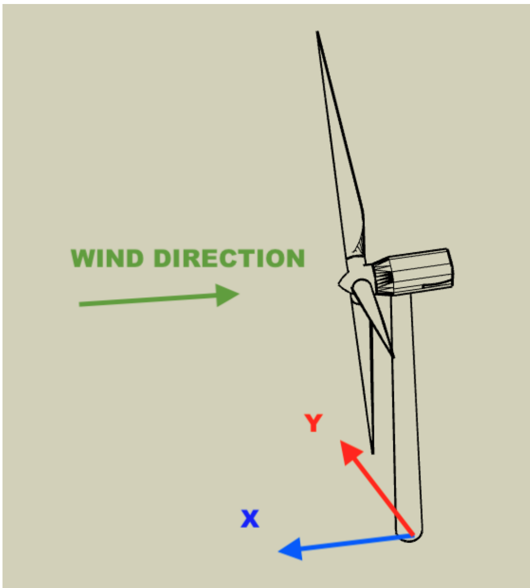


Figure 2: Definition of the reference frame for the longitudinal and the transversal directions.

Operational data have been provided by the wind turbine manufacturer in real time during the measurement campaign, with a sampling time of the order of the second. These have been used to assess the similarity of the wind and operation conditions at different wind turbines at the same time.

The vibration time series have been organized as indicated in Tables 1 and 2. The WTG01 time series are not labelled with more details (as for example the recording time) because they can be interchanged and the results of the following analysis don't sensibly change.

TS number	Wind turbine	Wind turbine status	Use
1	WTG01	healthy	reference - calibration
2	WTG01	healthy	reference - calibration
3	WTG01	healthy	validation
4	WTG03	damaged	validation

Table 1: The data set for WTG03 damage detection

TS number	Wind turbine	Wind turbine status	Use
1	WTG01	healthy	reference - calibration
2	WTG01	healthy	reference - calibration
3	WTG01	healthy	validation
4	WTG06	damaged	validation

Table 2: The data set for WTG06 damage detection

The information regarding the state of health of the wind turbine must be extracted from these data. In order to highlight it, some features can be extracted from the raw data set. Obviously, the choice of these characteristic parameters is strongly affecting the ability to perform a damage detection, so that they should be selected wisely. A simple choice is to use common time-domain statistics such as root mean square, skewness, kurtosis, peak value and crest factor (peak/RMS). These are usually quite sensitive to the operational and environmental conditions and are very fast to compute [11]. To ensure the statistical significance of the results, many measurement points are necessary. These features will be then extracted on short, independent (no overlap) chunks of the original signals. In particular, each acquisition is divided in 100 sub-parts on which the five features are computed. The considered data sets X results then to be a $n \cdot d$ matrix, where $n = 20$ is the number of channel and feature combinations, while $d = 400$ is the number of samples from the 4 acquisitions of Tables 1 and 2 placed one after the other.

3 Analysis and results

The results about the feature extraction are reported in Figures 3 and 4. The samples 0-200 are referred to the training data set for the wind turbine WTG01, the samples 201-300 are referred to the validation data set for the wind turbine WTG01 and, finally, the samples 301-400 are referred to the validation data set for the wind turbine WTG03 (WTG06, respectively). In the Figures, the training - calibration data set is separated from the validation data set by a black line. The validation data set for the damaged wind turbine is separated from the rest of the data sets through a red line.

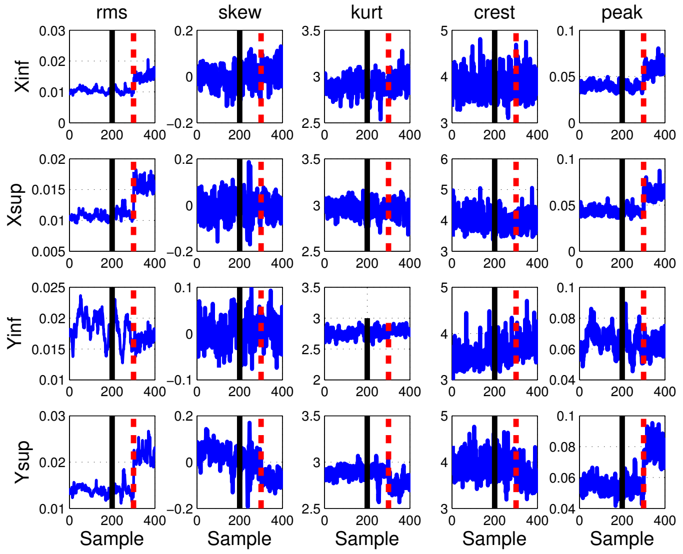


Figure 3: The extracted features for the data sets in Table 1.

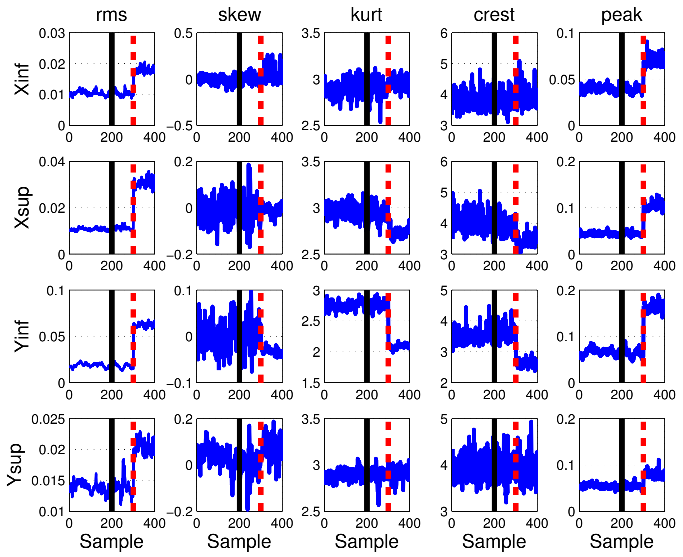


Figure 4: The extracted features for the data sets in Table 2.

A statistical approach is used in this paper to test if some diagnostic information can be obtained from the data, basically assessing the goodness of the selected features. The study starts with a univariate Analysis Of Variance (ANOVA), able to infer from the data the hypothesis that no statistical difference is detected among the groups, meaning that all the groups come from the same distribution.

The ANOVA is a statistical tool to test the omnibus (variance based) null hypothesis H_0: all the considered groups populations come from the same distribution, meaning that no significant difference is detectable. This hypothesis will be accepted or rejected according to a statistical summary $\hat{F}$ which, under the assumptions of independence, normality and homoscedasticity of the original data, follows a Fisher distribution:

$$\hat{F} = \frac{\frac{\sigma^2_{bg}}{G-1}}{\frac{\sigma^2_{wg}}{N-G}} \simeq F(G-1, N-G), \tag{1}$$

where

$$\sigma^2_{bg} = \sum_{j=1}^{G} \frac{n_j}{N} (\bar{y} - \mu_j)^2, \tag{2}$$

$$\sigma_{wg}^2 = \frac{1}{N}\sum_{j=1}^{G}\sum_{i=1}^{n_j}(\bar{y}_{ij}-\mu_j)^2, \tag{3}$$

with G being the number of groups of size n_j, N being the global number of samples with overall average $\bar{y}$, σ_{bg}^2 being the variance between the groups, σ_{wg}^2 being the variance within the groups (basically the average of the variance computed in each group) [12, 13]. The null hypothesis H_0 will be accepted with a confidence level $1-\alpha$ if the summary $\hat{F}$ is less extreme than a critical value $F^{\alpha}(G-1,N-G)$. A corresponding p-value can also be computed: it coincides with the probability of the summary to be more extreme than the observed $\hat{F}$, assuming H_0 to be true. If this value is less than α (typically, 5%), H_0 is rejected. The concepts of critical value and p-value are summarized in Figure 5.

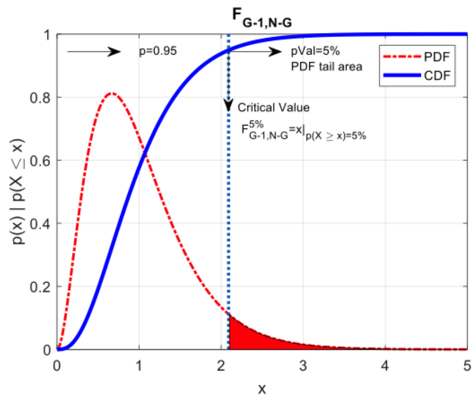


Figure 5: $F(G-1,N-G)$ distribution, with highlighted the 5% critical value and the concept of p-value.

In this analysis, the data sets are divided in 2 groups: the healthy one contains the first 300 samples (time series 1 to 3), while the last 100 samples, coming from the damaged and turbines WTG03 and WTG06 (time series 4 and $\bar{4}$), are labelled as damaged. The assumption of normality can be considered verified with enough confidence. The same does not hold for the homoscedasticity (equal variance in the different groups), but the ANOVA is commonly considered robust to such violations, so that the trustworthiness of the results will not be affected. It is relevant to point out that in this case, which uses 2 groups only, the ANOVA reduces to a Studentâs t-test. Furthermore, the ANOVA is a univariate technique, so it will be repeated per each channel and feature combination (20 times). The results are reported in Tables 3 and 4.

Feature / Channel	Xinf	Xsup	Yinf	Ysup
RMS	$< 10^{-32}$	$< 10^{-32}$	$9 \cdot 10^{-13}$	$< 10^{-32}$
Skewness	$< 10^{-8}$	0.4	0.4	$< 10^{-32}$
Kurtosis	$< 10^{-12}$	0.6	$< 10^{-12}$	$< 10^{-32}$
Crest	0.005	0.8	$< 10^{-19}$	$< 10^{-22}$
Peak	$< 10^{-32}$	$< 10^{-32}$	0.01	$< 10^{-32}$

Table 3: ANOVA p-values for the data sets in Table 1. The red cells are used to highlight the acceptance of H_0 (p-value$> 5\%$), which implies a more difficult damage detection.

Feature / Channel	Xinf	Xsup	Yinf	Ysup
RMS	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$
Skewness	$< 10^{-13}$	0.2	$< 10^{-32}$	$< 10^{-32}$
Kurtosis	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	0.01
Crest	0.001	$< 10^{-32}$	$< 10^{-32}$	0.6
Peak	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$	$< 10^{-32}$

Table 4: ANOVA p-values for the data sets in Table 2. The red cells are used to highlight the acceptance of H_0 (p-value$> 5\%$), which implies a more difficult damage detection.

The Principal Component Analysis (PCA) is a technique widely used in multivariate statistics, in particular for the purpose of allowing the visualization of multi-dimensional data sets using projections on the first 2 or 3 principal components. This dimension reduction is not really advisable for diagnostic purposes, as the condition-information may, in principle, be hidden in the neglected principal components, making the detection more challenging. In any case, it is used in this analysis as a qualitative visualization of the data set under a different point of view, resulting from the transform produced by the technique. The PCA uses an orthogonal space transform to convert a set of correlated quantities into the uncorrelated variables called principal components. This transform is basically a rotation of the space in such a way that the first principal component will explain the largest possible variance, while each succeeding component will show the highest possible variance under the constraint of orthogonality with the preceding ones. This is usually accomplished by eigenvalue decomposition of the data covariance matrix, often after mean centering.

The PCA transform has been applied to the reference data set: the statistical features matrix extracted from the WTG01 time series 1 and 2 of Tables 1 and 2. Subsequently, the validation data sets have been separately projected to the space generated by the first two principal components of the reference data set. The results are reported in Figures 6 and 7, from which it arises that the data set of WTG06 is more easily distinguishable with respect to the calibration data set than the data set of WTG03. As regards Figure 7, the indication is that the visual inspection based on the first two principal components can be sufficient for detecting an anomaly.

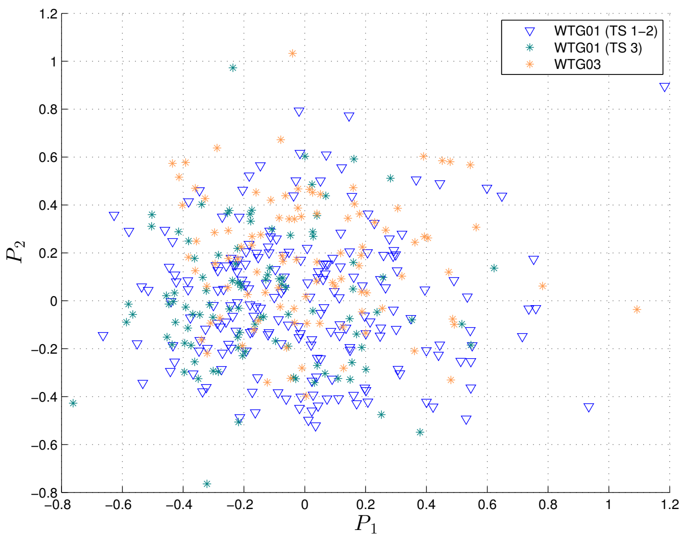


Figure 6: Projection of the data from Figure 3 to the space generated by the two principal components of the calibration data set (first 200 samples in Figure 3)

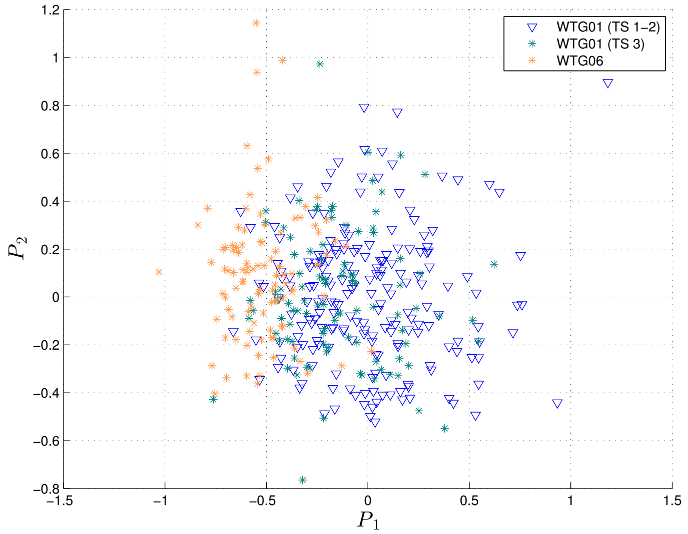


Figure 7: Projection of the data from Figure 4 to the space generated by the two principal components of the calibration data set (first 200 samples in Figure 4)

In statistics, the detection of anomalies can be performed pointwise, looking for the degree of discordance of each sample in a data set. A discordant measure is commonly defined outlier, when, being inconsistent with the others, is believed to be generated by an alternate mechanism. The judgment on discordance will depend on a measure of distance from the reference distribution, usually called Novelty Index (NI) on which a threshold can be defined [14]. The Mahalanobis distance is the optimal candidate for evaluating discordance in a multi-dimensional space, because it is non-dimensional and scale-invariant, and takes into account the correlations of the data set. The Mahalanobis distance between one measurement y (possibly multi-dimensional) and the x distribution, whose covariance matrix is S, is given by

$$d_M(y) = \sqrt{(y - \bar{x})\, S^{-1}\, (y - \bar{x})}. \tag{4}$$

In the following, the reference x distribution is selected as the statistical features matrix extracted from the WTG01 time series 1 and 2 of Tables 1 and 2. The target y is selected as the statistical features matrix extracted from respectively time series 3 (WTG01), 4 (WTG03), $\bar{4}$ (WTG06).

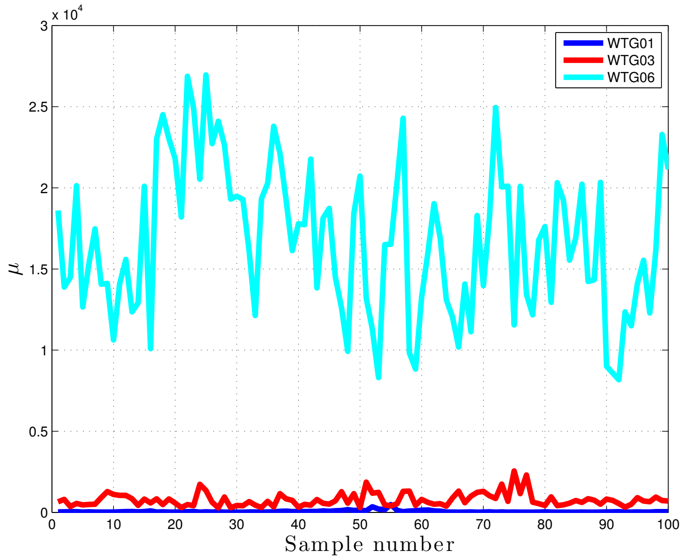


Figure 8: The Mahalanobis distance μ with respect to the calibration WTG01 data set: WTG01, WTG03, WTG06.

From Figure 8, it is possible to clearly distinguish between wind turbine WTG01 and wind turbines WTG03 and especially WTG06. The Mahalanobis distance therefore qualifies to be particularly responsive for novelty detection issues.

4 Conclusions

A novel approach for damage detection of a wind turbine gearbox was proposed in this study. One main novelty is that the accelerometric acquisitions were performed inside the tower of the wind turbines of interest because, despite the distance with respect to the gearbox, it is easily accessible by the turbine practitioners without shutting down the wind turbine. This measurement technique is a distinctive part of the outcome of the present work. One reference healthy wind turbine and two wind turbinse affected by different damage severity have been selected as test cases for the measurement campaigns proposed in the present study.

Subsequently, a Novelty detection procedure was set up, based on the calculation and the elaboration of common time domain features like RMS, Skewness, Kurtosis, Crest factor and Peak value. The analysis started with an ANOVA and a PCA, two fundamental tools in univariate and multivariate statistics. Both techniques proved that the damages can be detected. Finally, the Mahalanobis Novelty detection showed optimal results in detecting the possible damage, given the large margin which separates the supposedly damaged wind turbines from the healthy wind turbine. This algorithm also proved to be a good unsupervised damage detection technique considering the quickness, the simplicity and the full independence from human interaction, which makes it suitable for real time implementation. Overall, the whole gearbox vibration monitoring methodology can be considered validated by the test. The simple, non-invasive measurement system composed of just 2 biaxial accelerometers placed in accessible locations at 2 levels inside the tower of the wind turbine, together with the Novelty detection algorithm applied on the common time-domain features extracted, demonstrated indeed to provide a robust monitoring system, which can be easily integrated in existing installations.

This system can, in principle, enable to monitor also the damage evolution in time, establishing the foundations for further works on prognostics: this is supported by the responsiveness of the proposed methods (especially the Mahalanobis distance analysis) with respect to the severity of the damages (Figure 8). The straightforward further direction of the present work is therefore the analysis of the evolution in time of the same test case.

Acknowledgements

The authors would like to thank Ludovico Terzi, technology manager of the Renvico company. This research was partially funded by Fondazione âCassa di Risparmio di Perugiaâ through the research projects OPTOWIND (Operational Performance and Technical Optimization of WIND turbines) and WIND4EV (WIND turbine technology EVolution FOR lifecycle optimization).

References

[1] J. Rafiee, F. Arvani, A. Harifi, & M.H. Sadeghi, *Intelligent condition monitoring of a gearbox using artificial neural network*, Mechanical systems and signal processing, 21(4), 1746-1754.

[2] P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. Tameghe, & G. Ekemb, *Wind turbine condition monitoring: State-of-the-art review, new trends, and future challenges*, Energies, 7(4), 2595-2630.

[3] D. Astolfi, F. Castellani, & L. Terzi, *Fault prevention and diagnosis through SCADA temperature data analysis of an onshore wind farm*, Diagnostyka, 15.

[4] Z. Zhang, A. Verma, & A. Kusiak, *Fault analysis and condition monitoring of the wind turbine gearbox*, IEEE transactions on energy conversion, 27(2), 526-535.

[5] J. Igba, K. Alemzadeh, C. Durugbo, & E.T. Eiriksson, *Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes*, Renewable Energy, 91, 90-106.

[6] C. Peeters, P. Guillaume, & J. Helsen, *Vibration-based bearing fault detection for operations and maintenance cost reduction in wind energy*, Renewable Energy, 116, 74-87.

[7] J. Antoni, *Cyclic spectral analysis in practice*, Mechanical Systems and Signal Processing, 21(2), 597-630.

[8] F. Yanhui, L. Jiawei, Q. Yingning, Y. Wenxian, & D. Infield, *Study on order analysis for condition monitoring wind turbine gearbox*, 3rd Renewable Power Generation Conference (RPG 2014).

[9] G.A. Skrimpas, T. Ursin, C. Sweeney, K. Marhadi, N. Mijatovic, & J. Holboell, *Residual signal feature extraction for gearbox planetary stage fault detection*, Wind Energy, 20(8), 1389-1404.

[10] E. Mollasalehi, D. Wood., & Q. Sun, *Indicative fault diagnosis of wind turbine generator bearings using tower sound and vibration*, Energies, 10(11), 1853.

[11] A. P. Daga, L. Garibaldi, A. Fasana, & S. Marchesiello, *ANOVA and other statistical tools for bearing damage detection*, International Conference Surveillance 9, Fez (Morocco), 22-24 May 2017.

[12] H. Scheffe, *The analysis of variance*, John Wiley Sons, 1959.

[13] H. Sahai, & M.I. Ageel, *The analysis of variance: fixed, random and mixed models*, Springer Science Business Media, 2012

[14] K. Worden, G. Manson, & N.R. Fieller, *Damage detection using outlier analysis*, Journal of Sound and Vibration, 229(3), 647-667.

Gears and bearings defaults: from classification to diagnosis using machine learning for SURVISHNO Conference 2019

Sylvain BARCET[1], Valentin BARON[2], Alexandre CARBONELLI[1]

[1]VIBRATEC, F-69130 Ecully, France
[2]MICRODB, F-69130 Ecully, France

sylvain.barcet@vibratec.fr
valentin.baron@micro-db.fr
alexandre.carbonelli@vibratec.fr

Abstract

Gears and bearings are more and more used in every industrial area mainly due to their strong reliability. Nevertheless as every mechanical transmission system, failures appear during time life. It induces critical damage, time cost for maintenance services to repair the fault potentially on duty. A wide part of work in the scientific community already provides a large quantity of features to follow health status of these systems (e.g., RMS, kurtosis, crest factor, FM0) in order to detect the fault as soon as possible.

Since few years, methods developed in signal post-processing are coupled with Machine Learning (ML). ML allows ability to detect novelty or fault based on a trained algorithm. According to the literature [1], to identify the type of damage, a supervised algorithm is needed. Consequently an accurate diagnosis implies labelled data which are often difficult to obtain practically.

The aim of this paper is to provide keys, based on our knowledge about features in Structural Health Monitoring (SHM), to get higher information level in classification by adding a qualitative analysis (type of damage) without label or information about the type of fault.

Work carries on a measurement database. The assumption is made about two classes “healthy” / “faulty” using a supervised algorithm. The contribution of our work brings a new step in the default analysis by adding a probability for a defect case to be identified. Indeed, by combining some sensitive features selected for their relevance to describe a type of fault, a probability to have this particular default can be given. This classification is tested against three fault classes: bearing, gear generalized, gear localized.

Results show that a probability for having bearing fault can be identified using this method contrary to the gear generalized and localized fault which are more complex to characterize. This new step enables to help maintenance services to focus more efficiently on the incriminated faulty part of the system, inducing a reduction of time to repair for maintenance services, a shorter out of order time leading to a significant productivity gain.

[redacted] Introduction

Gears are used in a huge quantity of mechanical systems. As a consequence, monitoring their possible faults the most accurately possible is a major issue in the field of Structural Health Monitoring (SHM) as they can provoke critical damages.

With the increasing use of ML techniques, different algorithms have emerged to deal with this problem. Support Vector Machine, Neural Network, Random Forest are examples of ML methods currently used to classify faulty and healthy sample.

Looking at industrial maintenance services requirements, the needs in terms of monitoring may be resumed as:

1- Find efficient condition indicators (CI) to monitor their systems,
2- Use ML algorithms to allow a continuous monitoring and an high efficiency of faulty detection,
3- Have a minimal cost and time to repair the faulty equipment.

The first and the second point are already addressed in literature. This paper proposes to industrials a method to complete their process by the third part: a qualitative analysis of fault, leading to a reduction of cost and time for maintenance services.

[redacted] Presentation of the study case

PHM Society proposes a challenge for monitoring and fault detection. They provide a measurement database (measured on a test bench). Students, researchers and companies can participate. Each one proposes their own method to classify the given database. This work is based on the database provided for the 2009 challenge. Figure 1 presents the test bench used to build this measurement database.

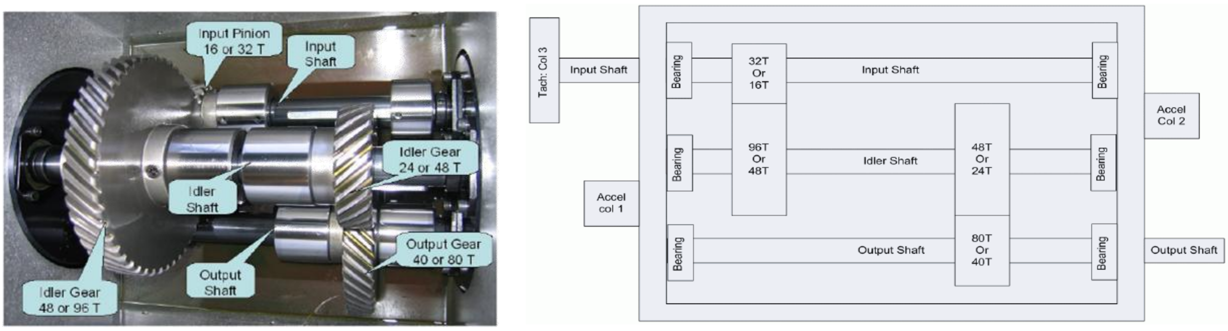


Figure 1: Presentation of the test bench used for measurements coming from PHM Society 2009 Challenge

This test bench is built with two gear stages mounted on three shafts with six bearings. Two gear geometries are used: one using spur gears, the other one using helical gears.

Table 1 presents the gear parameters for both configurations.

Shaft	Gear	Spur gear	Helical gear
Input shaft	input gear	32	16
Idler shaft	1st idler gear	96	48
Idler shaft	2nd idler gear	48	24
Output shaft	output gear	80	40

Table 1 : Gear description for the two geometries: spur and helical

Whatever the gear geometry, the gear ratio between each gear stage is the same, leading to keep the same global gear ratio. Thus, from input to output the gear reduction ratio is 5 to 1 reduction:

$$gear\ ratio = \frac{16}{48} \times \frac{24}{40} = \frac{1}{5}\,. \quad (1)$$

The instrumentation is composed of a limited number of sensors with two accelerometers mounted on the housing and one tachometer on the input shaft (see Figure 1). The tachometer delivers 10 tops by rotation. The sampling frequency is the same for the three sensors and fixed at 66666.67 Hz. Each sample of the database is composed of three raw data columns of one second length.

Several configurations are listed such as:
- 14 gear cases:
    - 8 configurations on spur gear [7 faulty and 1 healthy],
    - 6 configurations on helical gear [5 faulty and 1 healthy],
- 5 rotational speeds [30 Hz, 35 Hz, 40 Hz, 45 Hz, 50 Hz],
- 2 load cases [high, low].

This leads to 140 different configurations. Each configuration is repeated four times to give at least 560 measurement inputs.

[redacted] Method

The methodology developed and exposed here is composed of three main steps:
- CI computing,
- Classification using ML,
- Qualitative analysis using relevant selected features for each chosen default.

[redacted] Condition indicators computing

Figure 2 and Figure 3 present the post-processing used to build a matrix with all indicators.

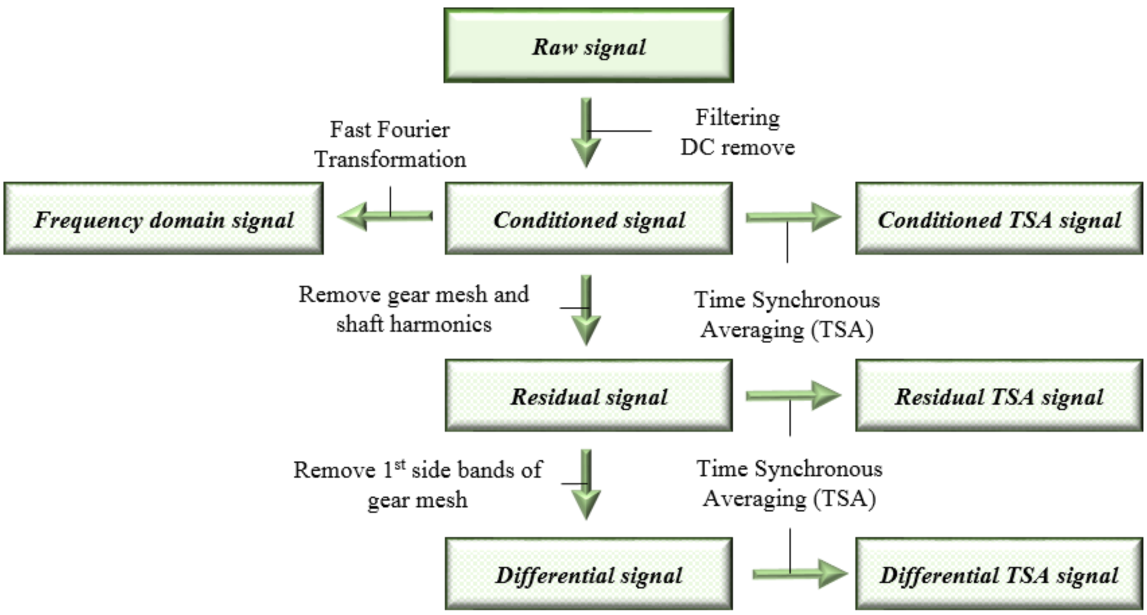


Figure 2 : Signal post-processing used to access to the different needed types of signals

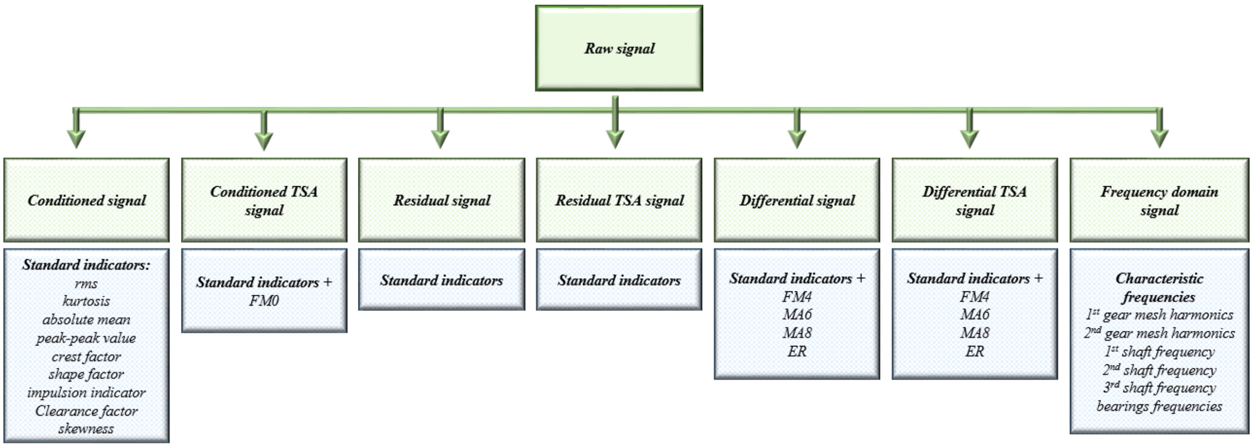


Figure 3 : List of indicators calculated on each signal

From raw signals, based on literature on bearings and gearboxes monitoring, some CI are extracted. Taking into account the difficulty to add a lot of different sensors, most of usual CI used in industry are directly computed from accelerometers raw signals (conditioned signal) such as RMS, kurtosis, crest factor. Based on [3], some new indicators, using residual or differential signals, requiring a tachometer information allow to increase strongly the efficiency of fault detection.

In order to complete this list of CI, some other indicators are computed from the frequency domain, such as gear mesh harmonics, shaft rotational frequencies and specifics bearings frequencies.

All these frequencies are calculated and added in the panel of features using the following equations:

$$Ball\ Pass\ Frequency\ of\ Inner\ ring: BPFI = \frac{N_b}{2}\, f\left(1 + \frac{D_b}{D_p}\cos(\alpha)\right), \qquad (2)$$

$$Ball\ Pass\ Frequency\ of\ Outer\ ring: BPFO = \frac{N_b}{2}\, f\left(1 - \frac{D_b}{D_p}\cos(\alpha)\right), \qquad (3)$$

$$Ball\ Spin\ Frequency: BSF = \frac{D_p}{2\,D_b}\, f\left(1 - \left(\frac{D_b}{D_p}\right)^2 (\cos(\alpha))^2\right), \qquad (4)$$

with f the number of revolutions per second, D_b the ball diameter, N_b the ball number, D_p the pitch diameter and α the contact angle.

$$Gear\ mesh\ frequency = f_{sh(i)} \times n_{teeth(i)}, \qquad (5)$$

with $f_{sh(i)}$ the rotational speed frequency of the shaft “i” and $n_{teeth(i)}$ the number of teeth on the gear mounted on the shaft “i”.

Considering the two accelerometers, a total of 298 indicators are extracted.

[redacted] Classification using Machine Learning

Based on all our indicators, the second step of the methodology consists in the classification of the 560 signals in two classes: healthy or faulty. Although the labels are not given in the database from the PHM Society, in [2], authors give the label of all healthy samples they have classified. Based on this, an approach using supervised algorithms is possible for the classification.

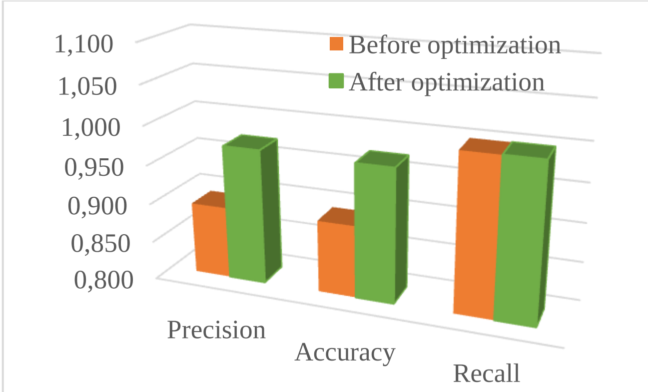
A wide range of supervised algorithms exists in the literature. Among them, a classification is proposed using four of them implemented within the Scikit-learn Python module [5]:

- Nearest Neighbors classifier (KNN) [6],
- Random Forest classifier (RFC) [7],
- Support Vector classifier (SVC) [8],
- Multi Layer Perceptron Classifier (MLPC) [9].

Each algorithm works with a specific method to classify. SVC determines a boundary between the two classes using only the data of each class which are close one to the other, the so-called support vectors. KNN looks at the same class nearest samples of a particular observation to build the boundary between classes. RFC builds decision trees and it combines them together to give its final classification. Finally MLPC relies on a trained neural network to decide that class the tested observation belongs to.

The considered input signals for this step are the 560 one second duration signals provided by the PHM database. So an observation for the following of the contribution refers to a vector of dimension 298 gathering all the features computed from one of these 560 signals. These 560 signals are splitted in two categories: a train set and a test set. 80 % of the database is used for the train set (448 signals) and 20 % for the test set (112 signals).

The efficiency of these algorithms without any optimization is around 90% of good classification. An efficient solution to increase the performance of a ML algorithm is to optimize some hyper parameters. Table 2 presents results before and after optimization on SVC algorithm.



SVC algorithm	Before optimization	After optimization
Accuracy	0.893	0.973
Precision	0.893	0.971
Recall	1.000	1.000

Table 2 : Results of Support Vector Classifier before and after hyper parameters optimization

The optimization phase enables to increase significantly the accuracy of the considered algorithm from 89.3 % to 97.3 %. After an optimization of the hyper parameters on every algorithm, each individual algorithm gives interesting results with a percentage of good classification between 90 % and 97.1 %.

Finally a method to increase strongly the performances relies on combining different algorithms with different approaches. This method called “ensemble learning” reduces individual weaknesses of each algorithm. In the present case, the precision reaches 100 % using ensemble learning on the three best algorithms, namely RFC, SVC and MLPC.

[redacted] Qualitative analysis

The classification in healthy or faulty cases enables to alarm maintenance service immediately after or ideally a short time before the failure. However, no information is given about the type of fault.

The methodology proposed in this work consists in adding a qualitative information in the classification.

The work presented here proposes a method to estimate the probability for a faulty observation to correspond to one of these three types of faults:

- Bearing faults,
- Gear generalized faults,
- Gear localized faults.

The wide quantity of indicators to monitor bearings and gearboxes given in the first section does not describe the same input signals' features. Consequently they are not equally sensitive to the different types of faults [4].

Where RMS represents the energy of the signal leading to an accurate indication about the general state of the complete system, the peak-peak value is in opposition, very sensitive to any localized phenomenon on the signal, enabling to discriminate a localized fault such as crack on tooth or more critical case like a missing tooth. A solution based on the physics described by each indicator is investigated to predict the type of fault.

3.3.1 Feature selection

The selection proposed is composed of:

- 70 indicators for generalized faults:
    - Rms on TSA signal:
        - Rms on conditioned TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft,
        - Rms on residual TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft,
        - Rms on differential TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - Absolute mean:
        - Absolute mean on conditioned signal,
        - Absolute mean on conditioned TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft,
        - Absolute mean on residual signal,
        - Absolute mean on residual TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft,
        - Absolute mean on differential signal,
        - Absolute mean on differential TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - MA6:
        - MA6 on differential signal,
        - MA6 on differential TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - Gear mesh harmonics:
        - 1^{st} gear mesh frequency, harmonics 1 to 5,
        - 2^{nd} gear mesh frequency, harmonics 1 to 5.
- 14 indicators for localized faults:
    - FM0:
        - FM0 on conditioned TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - Peak-peak value on residual signal:
        - Peak-peak value on residual signal,
        - Peak-peak value on residual TSA signal on the 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
- 18 indicators for bearing faults:
    - BPFO:
        - BPFO on 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - BPFI:
        - BPFI on 1^{st} shaft, 2^{nd} shaft and 3^{rd} shaft.
    - BSF:

- BSF on 1st shaft, 2nd shaft and 3rd shaft.

From the 298 indicators, a total of 102 indicators are extracted for this qualitative analysis, they are summarized in Table 3.

Type of fault	Type of CI	Number of indicators (for 2 accelerometers)
Generalized	RMS on residual signals	18
	MA6	8
	Absolute mean	24
	Gear mesh harmonics	20
Localized	FM0	6
	peak-peak value on residual signal	8
Bearing	BPFO	6
	BPFI	6
	BSF	6

Table 3: List of CI used for the qualitative analysis

3.3.2 Characterization method

Once feature selection is performed according to their relevance for each of the investigated fault, the method to highlight the cases associated with a particular fault relies on the Principal Component Analysis (PCA) method [10]. The goal of PCA is to find, in a point cloud, the direction on which the projection of its point cloud has a maximum variance. This direction is called the first principal component and the next ones are the orthogonal directions of this first one that again explain the maximum variance. In this contribution each feature selection for each fault gives a point cloud of high dimension:

- Dimension 70 for the gear generalized faults,
- Dimension 14 for the gear localized faults,
- Dimension 18 for the bearing faults.

Consequently the PCA is used in this case to perform a dimensionality reduction in order to be able to represent in a two-dimensional space data which came from these high-dimensional spaces. From this representation the observations which represent a particular fault are expected to be significantly far away from the healthy and the other default cases.

Therefore once the PCA is performed on each case, the probability density function (PDF) derived from the healthy cases is estimated. According to isolines corresponding to specific probability values three categories are differentiated:

- A category with a low probability to have the studied fault type whose data are close to the healthy ones after PCA,
- A category with a medium probability to have this fault which are a little bit further from the healthy data,
- A category with a high probability to have this fault which are far away from the healthy data.

[redacted] Results and analysis

4.1.1 Bearing defaults

Once the 18 features which should represent well the bearing faults are selected, the PCA is performed and Figure 4 is obtained.

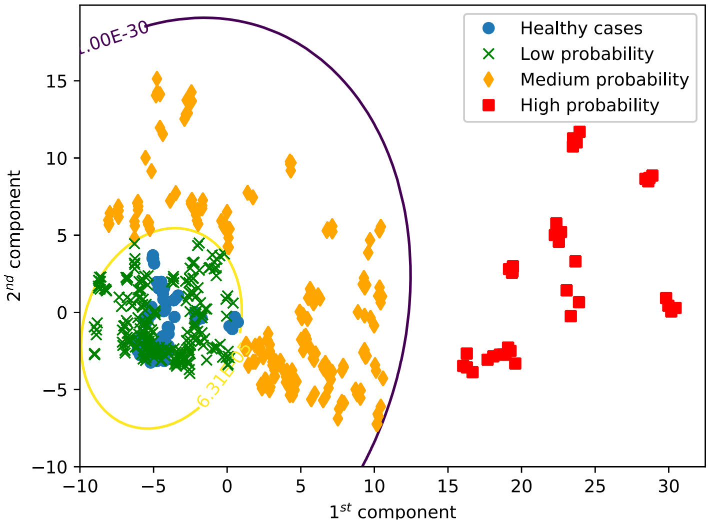


Figure 4: PCA result with the bearing faults feature selection. Yellow and purple lines represent respectively the $10^{-4.2}$ and 10^{-30} isolines of the probability density function derived from the healthy data. Healthy cases are given by blue dots, cases with a low / medium / high probability to have a bearing faults are given by green crosses / orange diamonds / red squares.

It clearly shows a distinct group further than the 10^{-30} probability isoline that gives candidates for having bearing fault. Then in between the two probability isolines the observations are classified as possible to have bearing fault. Finally data which are lying among the healthy cases are not likely to have bearing fault at all.

Consequently the feature selection made beforehand has allowed a qualitative analysis of the data. Indeed, thanks to their position compared to the healthy cases some observations can be classified within a bearing fault category with a given probability.

4.1.2 Localized and generalized defaults

Concerning gear localized and generalized faults, the results are more difficult to interpret as it is shown in Figure 5 with the PCA result of the localized fault case and in Figure 6 with the generalized fault case.

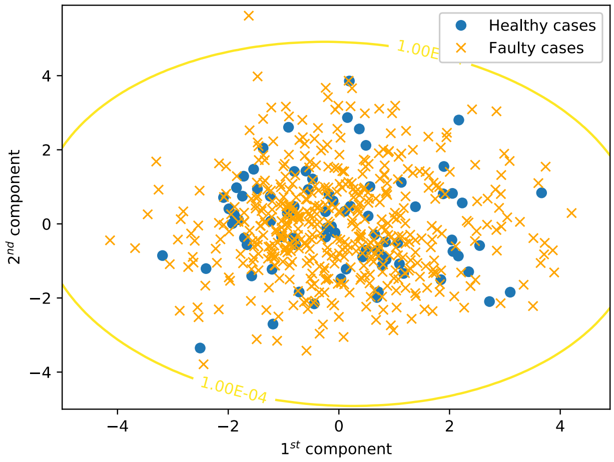


Figure 5: PCA result with the gear localized fault feature selection. The yellow line represents the 10^{-4} isoline of the probability density function derived from the healthy data. Healthy cases are given by blue dots and faulty cases are given by orange crosses.

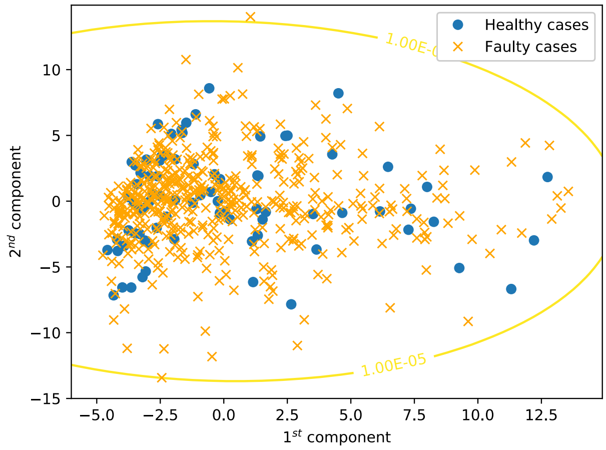


Figure 6: PCA result with the gear generalized fault feature selection. The yellow line represents the 10^{-5} isoline of the probability density function derived from the healthy data. Healthy cases are given by blue dots and faulty cases are given by orange crosses.

In these cases the feature selection has not allowed to distinguish different categories within the fault cases. It means that the selected features are not a relevant set enough to predict both gear localized or generalized faults. Some work is in progress to find better sets of features in order to be able to reproduce the results obtained with the bearing fault identification.

[redacted] Conclusion

This work presents the full method to diagnose and follow the healthy condition of a rotating equipment: from the conditions indicators to the classification. From 298 relevant conditions indicators obtained from accelerometer signals, a first supervised classification step enables to decide whether or not a considered observation is healthy or faulty using ensemble learning with three combined ML algorithms. To add a qualitative analysis of the faulty cases, a new step is performed using principal component analysis on a feature subset of the 298 ones. Three different faults are studied, bearing faults and gear localized and generalized faults, bringing three different subset of features of dimension 18, 14 and 70, respectively. The dimensionality reduction obtained using the principal component analysis allows to represent in a two-dimensional space, corresponding to the two first principal components, both healthy and faulty cases. From this representation a probability density function of the healthy cases is determined and the faulty cases can be marked as having a low or high probability to indeed, having this specific fault from their position compared to this function.

Results show that bearing faults can be identified using this methodology with three different categories highlighted: low, medium and high probability for having this fault type. However the feature selection made for the gear localized and generalized faults have not allowed to distinguish clearly between healthy and faulty cases and consequently there is no specific observation which can be identified with these fault types.

Work is in progress on the feature selection to be able to reproduce the bearing fault results and other methods are investigated to perform a better dimensionality reduction such as manifold learning techniques. These two ideas could bring the missing part to be able to give a complete qualitative assessment of the fault types.

References

[1] K. Worden and al., *The fundamental axioms of structural health monitoring*, 2007

[2] H. Al-Atat, and al., *A systematic methodology for gearbox health assessment and fault classification*, Journal of Prognostics and Health Management, 2011

[3] A. S. Sait, Y. I. Sharaf-Eldeen, *A Review of Gearbox Condition Monitoring Based on vibration Analysis Techniques Diagnostics and Prognostics*, 2011

[4] V. Sharmaa, A. Pareya*, *A review of gear fault diagnosis using various condition indicators, 12th International Conference on Vibration Problems*, ICOVP 2015

[5] F. Pedregosa et al., *Scikit-learn : Machine Learning in Python*, JMLR 12, pp. 2825-2830, 2011.

[6] R. Duda et al., *Pattern Classification*, 2nd Edition, John Wiley & Sons, 2012.

[7] L. Breiman, *Random Forests, Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001 January 01.

[8] C. Cortes and V. Vapnik, *Support-vector networks, Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995 September 01.

[9] D. Rumelhart et al., *Learning representations by back-propagating errors*, Nature, vol. 323, no. 6088, pp. 533-536, 1986 October.

[10] I. Jolliffe, *Principal Component Analysis*, 2nd Edition, Springer Series in Statistics, 2002.

Vibration Feature for Detecting Eccentric Workpiece/Runout Faults During Continuous Gear Grinding Processes

Agusmian Partogi OMPUSUNGGU[1], Yann VONDERSCHER[2], Daniel MOTL[2]

[1]Flanders Make, Gaston Geenslaan 8, 3001 Leuven, Belgium
agusmian.ompusunggu@flandersmake.be

[2]VCST Industrial Products, Schurhovenveld 3 025, 3800 Sint-Truiden, Belgium
yann.vonderscher@gmail.com, daniel.motl@vcst.be

Abstract

Continuous gear grinding is a well-established and widely used process in the industry for large-scale production gears. It offers an economic/efficient process for finishing gears, which shapes the micro-geometry of the gear tooth flank and improves its surface quality. The resulting quality of ground gears depends on several factors, namely the tool performance, the machine stability as well as the correct clamping/positioning of the workpiece. The grinding step is very crucial since it has a direct impact on the operating quality of gears and in particular on the running noise behaviour of the end product. The potential of online vibration based gear grinding monitoring has been explored and demonstrated in the previous work [1] as a means of quality control that could lead to the overall reduction of production losses and to the prevention of sending defective parts to customers. A number of features which could be used to monitor the grinding processes and to identify a specific type of defects have been proposed & experimentally validated to some extent. The types of faults include *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. However, a further investigation on a new test campaign revealed that none of the features developed in [1] was sensitive and robust enough to detect eccentric workpieces during the grinding process. It is worth mentioning here that an eccentric workpiece fault is unlikely to happen, but it is analogous to a runout on the incoming workpiece quality. In this paper, a qualitative model to predict the vibration signature due to eccentric workpieces/runouts is developed and discussed. Based on the qualitative understanding, a novel feature to detect eccentric workpieces/runouts during gear grinding processes based on vibration signals has been developed. The newly developed feature has been validated on real vibration signals captured during the emulation of process malfunctions on an industrial gear grinding machine. The experimental results show that the novel feature is sensitive and robust for detecting workpiece eccentricities of about 40 microns. It is also shown in this study that the feature is insensitive to other types of gear grinding faults, which is important for diagnostics/root-cause analysis purposes.

1. Introduction

A new industrial revolution, known as Industry 4.0, has been taking place worldwide for less than a decade. However, Europe is still at the beginning of this era and is currently in an effort to reindustrialise and to increase total value added from the manufacturing sector to a targeted 20%. The EU supports industrial change through its industrial policy and through research and infrastructure funding in the frames of H2020. The key success to Industry 4.0 is digitalisation and data transformation into business insight, namely how to gather, filter, analyse, store and retrieve data and extract useful information from it. The information is then subsequently shared and used to drive, control and monitor processes. This direction is supported by the continuous need for lower production cost, higher quality, more flexibility, better safety and more environmentally friendly, as well as the ever-increasing per currency computing power of microprocessors and the ever-decreasing cost of sensors and measuring platforms.

One concrete example of Industry 4.0 applications is the monitoring of manufacturing processes for improved efficiency and productivity. The monitoring application in manufacturing can contribute to the machine health monitoring, to the tool condition monitoring (e.g. wear, breakage etc), to the workpiece inspection and quality (e.g. geometry) as well as to the process monitoring [2]. As a result, the manufacturing processes monitoring could from one side reduce the downtime and possible repair costs, and from the other side, be used as a product quality control, leading to the overall reduction of production losses and to the prevention of sending defective products to customers. Since manufacturing processes and equipment are numerous and vastly different, a plethora of sensors and signal processing tools have been proposed in the literature being usual adapted to the specific application. Among other approaches, online vibration monitoring seemed to be promising and could be used in order to monitor manufacturing processes and consequently accurately, on time and online malfunctions and defects can be identified, detected and diagnosed.

In gear manufacturing processes, continuous generating gear grinding is a well-established and commonly used process in the industry for large-scale gears production because it offers an economic/efficient process for finishing gears. This process is a dominating hard-fine finishing process, especially in the field of automotive gears [3]. Due to its high process efficiency, continuous generating gear grinding has replaced other grinding processes such as profile grinding in batch production of small- and middle-sized gears. Figure 1 illustrates a continuous generating gear grinding process that involves three movements, namely 1) Infeed – along with the tooth depth, 2) feed – along with the workpiece axis and 3) shift – along with the grinding wheel axis. The three movements serve as the process parameters determining the final quality of gears.

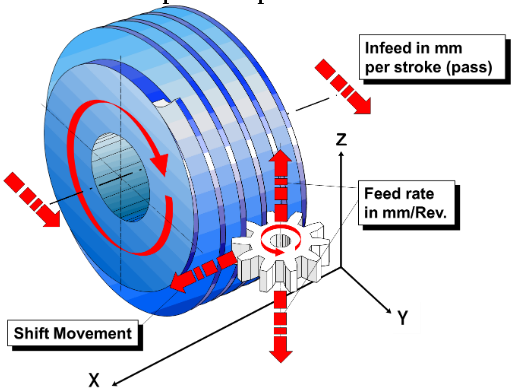


Figure 1: Illustration of a continuous generating gear grinding [4]

Continuous generating gear grinding is characterized by a high stock removal rate and is thus suited for high productivity batch processing. One of the main challenges is the determination of the abrasive/grinding forces due to their significant influence on the dynamics of the grinding process. During the grinding, there are multiple points of contact between the grinding worm and the gear, however, the number of these contacts changes continuously. This involves also a continuous change of the excitation forces. Thus, optimizing the cutting forces can lead to an increased quality of ground gears and a minimised wear behaviour of the grinding worm. Despite its wide industrial application, the knowledge of the generating grinding process is limited. The process design is based on experience along with time- and cost-intensive trials and research is based mostly on empirical studies. To maintain the high quality of ground gears, a reliable and robust online monitoring solution for gear grinding processes is therefore necessary.

The potential of online vibration based gear grinding monitoring has been explored and demonstrated in the previous work [1] as a means of quality control that could lead to the overall reduction of production losses and to the prevention of sending defective products to customers. A number of vibration features which could be used to monitor the grinding processes and to identify a specific type of defects were proposed & experimentally validated to some extent. The types of gear grinding faults include *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. However, a further investigation on a new test campaign revealed that the previously developed features [1] were not sensitive and robust enough to detect eccentric workpiece during the grinding process.

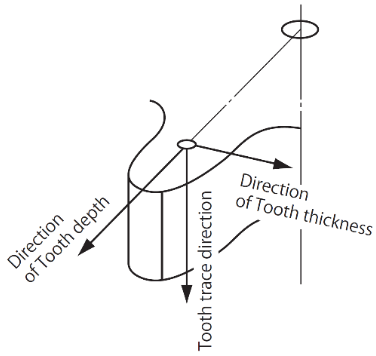
To fill the gap discussed earlier, the main focus of this study is therefore on the development of a novel feature that is reliable and robust for detecting eccentric workpieces/runout faults during gear grinding processes based on vibration signals measurement on the grinding machines. Some quality parameters of the ground gears were also measured that can be used as objective measures for the quality assessment. The remainder of the paper is organised as follows. In Section 2, an introduction to eccentric workpiece/runout is presented. In Section 3, qualitative modelling to predict the vibration signature of eccentric workpiece/runout is presented. In Section 4, the signal processing and feature extraction algorithm is presented. In Section 5, the experimental setup and the design of experiments are discussed. In Section 6, the results of the data analysis are discussed. Finally, some conclusions and recommendations for future work are summarised in section 7.

2. Eccentric Workpiece/Runout as A Type of Gear Errors

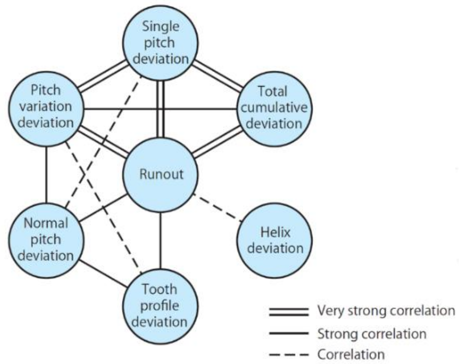
Gear errors can be classified into individual and composite errors. As shown in Figure 2(a), an individual error is, in fact, a three-dimensional (3D) error occurring in the direction of:

1) Tooth depth referring to the shape of tooth profile and length of tooth depth.
2) Tooth trace referring to the inclination and unevenness of tooth trace.
3) Tooth thickness referring to the thickness of tooth and tooth space.

These three types of individual errors are measured by taking apart a three-dimensional error into a two-dimensional error. However, these individual errors are correlated and the extent of correlation differs between the methods of production and measurement. In general, the correlations of these individual errors are shown in Figure 2(b) [5].



(a)



(b)

Figure 2: (a) Theoretical three-dimensional deviations, (b) Correlation with an individual deviation of ground spur gears [5]

Runout is one of the gear errors and defined as the amount of off-centre measured between gear and the axis. It is a characteristic of gear quality that results in an effective centre distance variation. Figure 2(b) shows how the runout deviation matters for gear quality and in fact affects every other characteristic of gear quality, such as involute or tooth form, index or pitch variation, lead or tooth alignment variation, etc. Therefore, a good practice for the manufacturing or inspection/monitoring of gears requires the control of runout. Runout results in accumulated pitch variation, and this causes non-uniform motions, which affect the function of gears [5]. Although runout is a radial phenomenon, the resulting accumulated pitch variation affects the tangential functionality that causes transmission error. In gear quality metrology, the runout error Fr is measured by indicating the position of a pin or ball inserted in each tooth space around the gear and taking the largest difference as schematically shown in Figure 3.

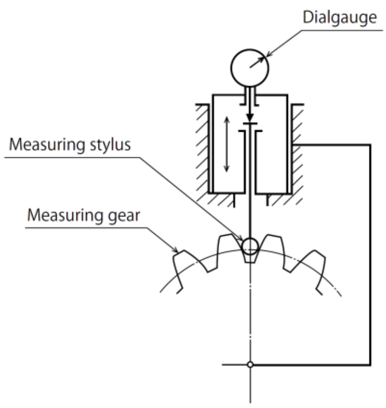


Figure 3: Measurement of runout [5].

When a workpiece is not centred on the spindle rotating axis as illustrated in Figure 4 (i.e. eccentric workpiece), the grinding of the eccentric workpiece will lead to a variation in the infeed penetration (i.e. the direction of tooth depth) once per gear revolution. Eventually, this variation will result in a runout error on the workpiece as will be shown later in the gear quality measurement results discussed in Section 6. Note that an eccentric workpiece fault is unlikely to happen, but a runout deviation might occur on incoming workpiece quality prior to grinding process.

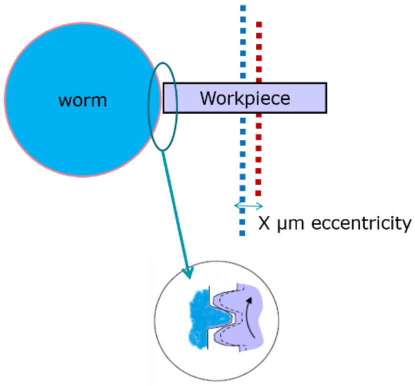


Figure 4: Illustration of an eccentric workpiece fault [1].

3. Qualitative Prediction of Grinding Vibration Signature/Feature Resulting From Eccentric Workpiece/Runout Faults

Figure 5 illustrates the kinematic model between a grinding worm and an eccentric workpiece. Without loss of generality, let assume that the workpiece is mounted on a rigid shaft, while the grinding wheel is mounted on a flexible shaft. Let r_1 be the radius of workpiece and r_2 be the radius of the grinding wheel. For a concentric workpiece as illustrated by the dashed circles in the figure, the rotation centre of the workpiece O coincides with the centre of the workpiece geometry and the rotation centre of the grinding wheel is at point $\hat{O}$. However, for an eccentric workpiece, the centre of the geometry O_1 does not coincide with the rotation centre O and the distance between the point O_1 and O determines the eccentricity level ϵ. As a result of the eccentricity on the workpiece, the rotation centre of the grinding worm (which is coinciding with the centre of the geometry) shifts from $\hat{O}$ to O_2.

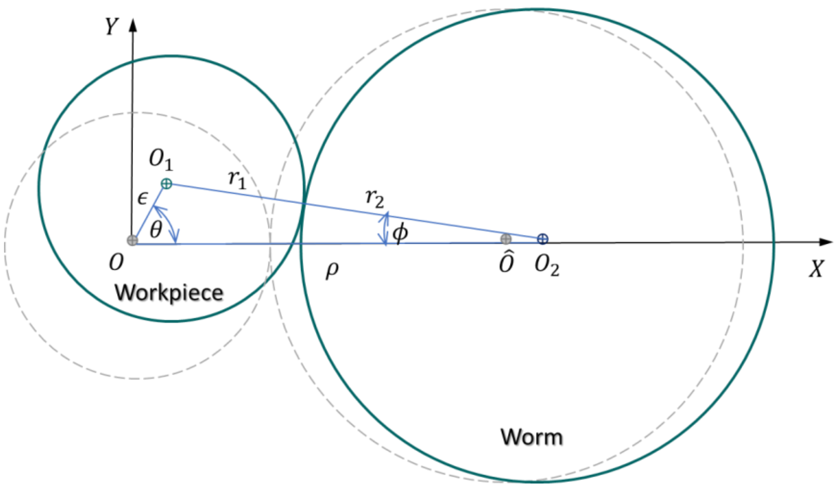


Figure 5: Kinematic model of a grinding wheel and an eccentric workpiece (gear).

With the aid of Figure 5, the distance between the rotation centres of the workpiece and the grinding wheel ρ can be derived as follows:

$$\rho = \epsilon \cos(\theta) + R \cos(\phi) \tag{1}$$

With $R = r_1 + r_2$, and the angle θ denoting the angular displacement of the workpiece which can be determined by integrating the angular speed of the workpiece over time $\theta = \int_o^t \omega_g \, dt$. As the angular speed is typically constant, thus $\theta = \omega t$. The angle ϕ can be inferred by using the sine rule as follows:

$$\frac{\epsilon}{\sin(\phi)} = \frac{R}{\sin(\theta)} \tag{2}$$

Eq (2) can be further rearranged as follow:

$$\phi = \arcsin\left(\delta \sin\left(\omega_g t\right)\right) \tag{3}$$

With $\delta = \frac{\epsilon}{R}$. In practice, the eccentricity is a lot smaller than the sum of the workpiece and the grinding wheel radii, $\epsilon \ll R \rightarrow \frac{\epsilon}{R} = \delta \approx 0$, so the angle ϕ is approximately zero, $\phi \approx 0$.
Hence, the distance ρ can now be approximated as follows:

$$\rho \approx R \times \left[1 + \delta \cos\left(\omega_g t\right)\right] \tag{4}$$

As theoretically and experimentally shown in [3,6], the amplitude of the gear grinding force in the normal direction F_n is proportional to the grinding depth a_e, the grinding width b_D and the feed rate v_w. Mathematically, the grinding force amplitude F_n can be expressed as follows:

$$F_n = C \times a_e^{f_a} \times b_D^{f_b} \times v_w^{f_v} \tag{5}$$

where C, f_a, f_b and f_v are constants which should be determined experimentally.

Eq. (4) shows that the distance ρ changes in time with the amplitude modulated with the same frequency as the workpiece angular speed. This suggests that the grinding depth (infeed penetration) a_e also varies in time with the same frequency as the workpiece angular speed. This implies that:

$$a_e \propto \rho \propto \left[1 + \delta \cos\left(\omega_g t\right)\right] \tag{6}$$

Hence, by substituting Eq. (6) to (5) and by assuming the other terms are constant, the gear grinding force amplitude can be written as follows:

$$F_n \propto \left[1 + \delta \cos\left(\omega_g t\right)\right] = \hat{C} \times \left[1 + \delta \cos\left(\omega_g t\right)\right]^{f_a} \tag{7}$$

where $\hat{C}$ is a constant. Since $\delta \ll 1$, Eq. (7) can thus be approximated by using the Taylor expansion as follows:

$$F_n = \hat{C} \times \left[1 + f_a \times \delta \cos\left(\omega_g t\right)\right] \tag{8}$$

It is obvious from Eq. (8) that the normal force amplitude F_n is modulated by the rotational speed of the workpiece $\omega_g = 2\pi f_g$.

According to [7], the dynamic mesh forces between two meshing gears have spectral content at (at least) the harmonics of mesh frequency. By assuming the grinding wheel and the workpiece have similar properties as two meshing gears, the dynamic mesh forces between the grinding wheel and workpiece can be expressed as follows:

$$F_{mesh}(t) = \sum_{k=1}^{K} F_{nk}(t) \sin(2\pi k f_m t + \psi_k(t)) \tag{9}$$

With $f_m = z_g \times \omega_g = z_w \times \omega_w$ denoting the gear meshing frequency between the grinding wheel and the workpiece, the index $k = 1, 2, \dots, K$ denoting the harmonic order of the meshing frequency, $F_{nk}(t)$ denoting the normal force amplitude at the harmonic order k and $\psi_k(t)$ is the phase which probably varies in time. By substituting Eq. (8) into (9), one can show that the latter can be rewritten as follows:

$$F_{mesh}(t) = \sum_{k=1}^{K} A_k \underbrace{\left[1 + \varepsilon_k \sin(2\pi f_g)\right]}_{\text{Related to gear eccentricity}} \times \sin(2\pi k f_m t + \psi_k(t)) \tag{10}$$

Eq. (10) suggests that workpiece eccentricity/runout errors modulate the gear mesh force amplitude with the frequency the same as the workpiece rotational speed. Hence, one can expect that the resulting vibration signals will also contain some amplitude modulation with the frequency of the workpiece rotational speed around the gear mesh frequency (i.e. the carrier frequency) and possibly the higher order harmonics.

4. Proposed Signal Processing and Feature Extraction Algorithm

In practice, depending on the application, gear grinding process is repeated for a given set of gears (multiple stroke passes). For such applications, it is important to localise the measured vibration signal corresponding to each stroke pass. This way, the physical effects of different passes on the measured vibration signal that are not relevant to the fault detections are eliminated.

During the first pass, the eccentricity will generate a variation of the amount of stock removal around one rotation of the gear. But during the following passes, the amount of stock removal will be constant all around the gear because the first pass will have corrected the variation of stock with respect to the eccentricity. Hence, only the first stroke pass (Pass#1) is relevant for detecting an eccentric workpiece during grinding processes

Based on the aforementioned reasoning and the qualitative model described in Section 3, the workflow of the signal processing and feature extraction for eccentric workpiece/runout fault detection has been established as visualized in Figure 6. The details of each step are described in the subsequent paragraphs.

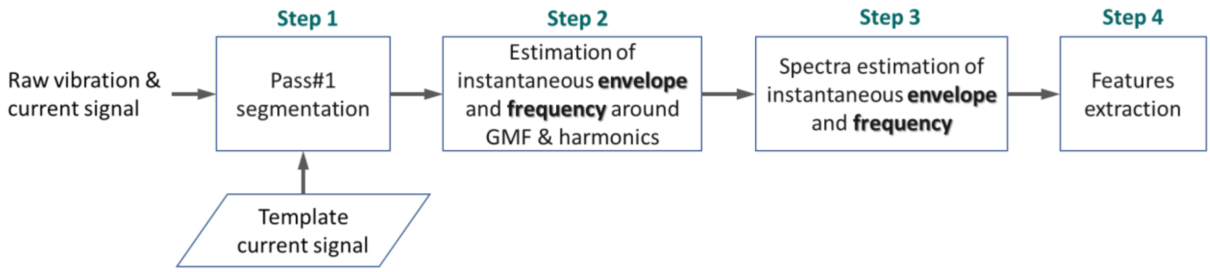


Figure 6: Workflow of the signal processing and features extraction for eccentric workpiece/runout faults detection.

Step 1

As shown in Figure 6, the ingredients to complete this first step are (*i*) the vibration signal recorded on a continuous gear grinding process which is synchronously acquired together with the worm motor current signal and (*ii*) a template current signal corresponding to Pass#1 that needs to be pre-determined by expert supervision. Pass#1 is determined by using the worm motor current signal $I_w(t)$ and the template current signal $I_t(t)$. As the current signal of the worm motor is typically noisy, low-pass filtering on this signal is recommended. The key process in this step is to align the measured current with template current signals. The alignment is attained by first computing the cross-correlation of the two current signals. Later on, the time delay is computed by detecting the lag corresponding to the peak in the resulting cross-correlated signal as described in [8]. Subsequently, the time delay is used to align the two current signals. Once aligned, the sample indices corresponding to Pass#1 can be determined. Finally, the sample indices are used to segment the measured vibration signal corresponding to Pass#1. Note that the segmentation of the signal can be further refined, for example, by chopping off the segmented signal around the maximum current signal. The workflow of this step is schematically illustrated in Figure 7.

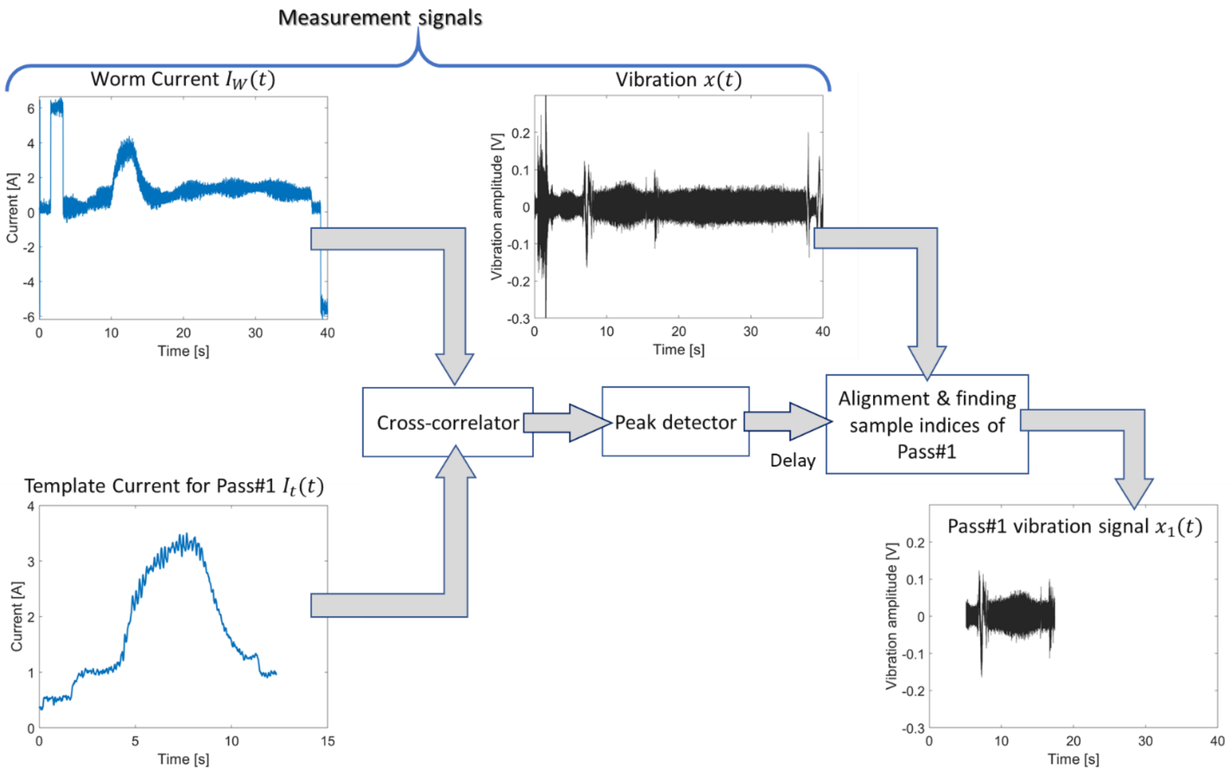


Figure 7: Block-diagram of the vibration signal segmentation of Pass#1.

Step 2

Once the raw vibration signal corresponding to Pass#1 has been segmented following the procedure in Step 1, the segmented signal is then fed into a band-pass filter. The central frequency of the band-pass filter is at the gear mesh frequency (GMF) between the workpiece and the worm, while the bandwidth of the filter is fixed about 4 times of the rotational speed of the workpiece.

The instantaneous envelope (IE) and the instantaneous frequency (IF) of each band-pass filtered signal are estimated using the Hilbert Transformation [9]. For an arbitrary signal $x(t)$, its Hilbert transform is defined as

$$H[x(t)] = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} d\tau \tag{11}$$

And the analytical signal $z(t)$ of $x(t)$ is determined as

$$z(t) = x(t) + jH[x(t)] = A(t)e^{j\phi(t)} \tag{12}$$

In this case, the IE $A(t)$ is given by

$$A(t) = |z(t)| = \sqrt{x^2(t) + H^2[x(t)]} \tag{13}$$

And the instantaneous phase $\phi(t)$ is calculated as follows

$$\phi(t) = \arg[z(t)] = \arctan\left[\frac{H[x(t)]}{x(t)}\right] \tag{14}$$

Furthermore, the IF $\upsilon(t)$ can be computed by taking the time derivative of the instantaneous phase

$$\upsilon(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt} \tag{15}$$

Step 3

The frequency spectra of the IE and IF are determined by applying the Fourier transform to $A(t)$ and $v(t)$ as described in the following equations:

$$X_{IE}(f) = \mathcal{F}[A(t)] = \int_{-\infty}^{+\infty} A(t)\, e^{-j2\pi ft} dt \tag{16}$$

$$X_{IF}(f) = \mathcal{F}[v(t)] = \int_{-\infty}^{+\infty} v(t)\, e^{-j2\pi ft} dt \tag{17}$$

With $X_{IE}(f)$ and $X_{IF}(f)$ denoting the frequency spectra of the IE and IF respectively.

Step 4

The IE and IF features are defined as the average magnitude of their corresponding spectrum within a given frequency range ($[f_l \quad f_u]$). Based on this definition, the IE and IF features can be mathematically expressed as the following equations:

$$\text{IE Feature} = \frac{1}{f_u - f_l} \int_{f_l}^{f_u} X_{IE}\, df \tag{18}$$

$$\text{IF Feature} = \frac{1}{f_u - f_l} \int_{f_l}^{f_u} X_{IF}\, df \tag{19}$$

The frequency range ($[f_l \quad f_u]$) should be selected such that a few harmonics of the *workpiece rotational speed* are included. In this study, f_l is set at 2 Hz and f_u is set at 20 Hz.

5. Experimental Study

5.1 Experimental Setup

An industrial grinding machine depicted in Figure 8(a) has been used to emulate different grinding faults, namely *i*) high feed rate, *ii*) high infeed, *iii*) non-flat workpiece, and *iv*) eccentric workpiece. Two triaxial accelerometers were mounted on the machine, the first one on the grinding worm holder and the second one on the base of the workpiece spindle. From the previous investigation [1], it can be concluded that one accelerometer mounted on the grinding worm holder is sufficient for gear grinding monitoring purposes, which is the focus of this study. Note that the axes orientations of the triaxial accelerometer in this study are different from the ones in the previous study [1]. The accelerometer axes orientations of this study are the following:

- X-axis: *Parallel to the worm axis*
- Y-axis: *Direction of Infeed* (and small component in direction of Feed)
- Z-axis: *Upward, the direction of Feed* (and small component in direction of Infeed)

(a)

(b)

Figure 8: (a) Experimental setup of the grinding machine, (b) Emulation of an eccentric workpiece.

5.2 Design of Experiment

In this study, three types of gear grinding faults are considered, namely *i*) *high feed rate*, *ii*) *high infeed* and *iii*) *eccentric workpiece*. Moreover, two severity levels of each fault type are considered as summarised in Table 1. The eccentric workpiece is emulated by mounting the clamping arbour eccentrically as shown by Figure 8(b). To seed the eccentric fault, the clamping arbor was mounted with an eccentricity. This can be done by playing with 3 centring bolts at the base of the clamping arbor. The actual eccentricity was checked using a dial indicator as shown in the photograph.

Test Classification	**Settings**
Baseline	Grinding with the nominal (optimised) parameters
Feed 1	The feed rate is 75% higher than the nominal feed rate
Infeed 1	The infeed is 80% higher than the nominal infeed
Excenter 1	Nominal grinding parameters. The workpiece is mounted eccentrically at 20 µm
Feed 2	Feedrate is 150% higher than the nominal feed rate
Infeed 2	The infeed is 150% higher than the nominal infeed
Excenter 2	Nominal grinding parameters. The workpiece is mounted eccentrically at 40 µm

Table 1. The settings of each test condition

For each type of fault, five measurements are captured sequentially by grinding five gears (workpieces) in a row. Prior to performing tests in a faulty state, five measurements are realised by grinding five gears with the nominal parameters to set up the baselines. Moreover, the grinding wheel is redressed prior to the baseline tests. The complete test sequence is illustrated in Figure 9, so in total there are 60 tests.

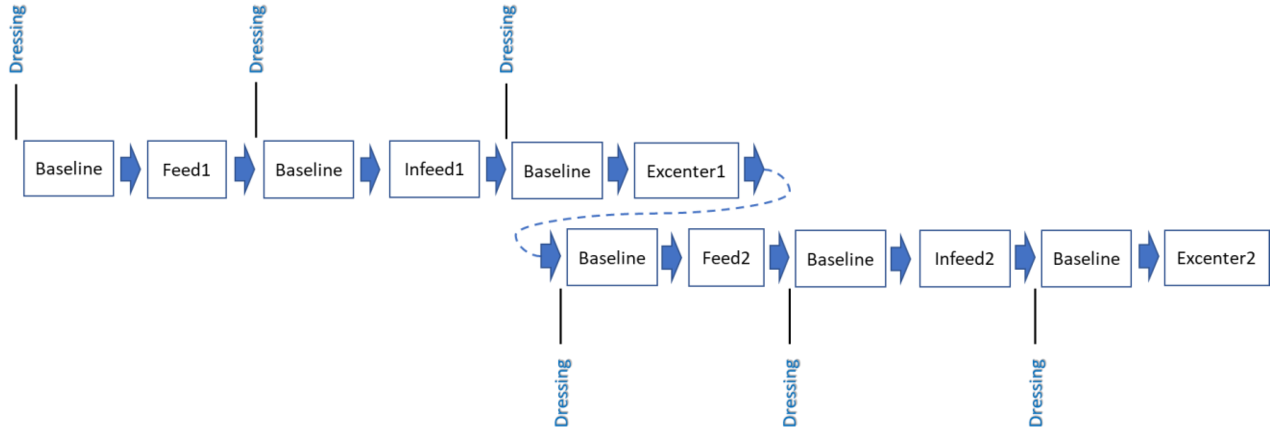


Figure 9: Overview of the test sequence

The grinding program is realised in two-stroke passes. As mentioned earlier in the previous section that only the first pass (Pass#1) is relevant for detecting eccentric workpieces/runouts during grinding processes. The duration of the signal is approximately about 40 seconds and the sampling frequency was selected at 25.6 kHz.

6. Results and Discussion

After completing the test campaign, a number of gear quality parameters of individual ground gear were measured using a dedicated gear metrology machine. For the analysis of this particular eccentric workpiece/runout fault test data, only the runout error Fr is relevant as discussed earlier. Figure 10 shows the boxplot of the measured runout error for each test class. As seen in the figure, only the *Excenter#2* (i.e. eccentricity of 40 µm) test class leads to significantly increased runout errors. Moreover, one can clearly see that the *Excenter#2* test class is very well separated with the *Baseline* test class and the other test classes including *Feed#1*, *Infeed#1*, *Excenter#1*, *Feed#2*, and *Infeed#2*. Notably, the other fault types (high infeed and high feed rate) do not lead to significantly increased runout errors.

Despite the fact that all the gears are ground with the nominal (optimised) parameter in the *Baseline* tests, there still exists some residual runout error on the ground gears. As can be seen in the figure, the averaged runout error of the gears in the baseline tests is about 8 µm. The residual runout errors could be caused by existing runout errors on the incoming workpiece due to imperfect pre-grinding processes.

Although the eccentricity introduced on the workpiece in the *Excenter#2* tests is about 40 µm, the resulting averaged runout error is about 65 µm, which is higher than the workpiece eccentricity. This suggests that the grinding processes seem to add more runouts on the ground gears. On the other hand, the averaged runout error resulting from the *Excenter#1* test class is about 10 µm, which is lower than the workpiece eccentricity of 20 µm. It is not clear yet till now, why for the *Excenter#1* test the resulting runout error is lower than the workpiece eccentricity. A possible explanation is that the emulation method for introducing an eccentric workpiece fault of 20 µm was not accurate enough, see Figure 8(b).

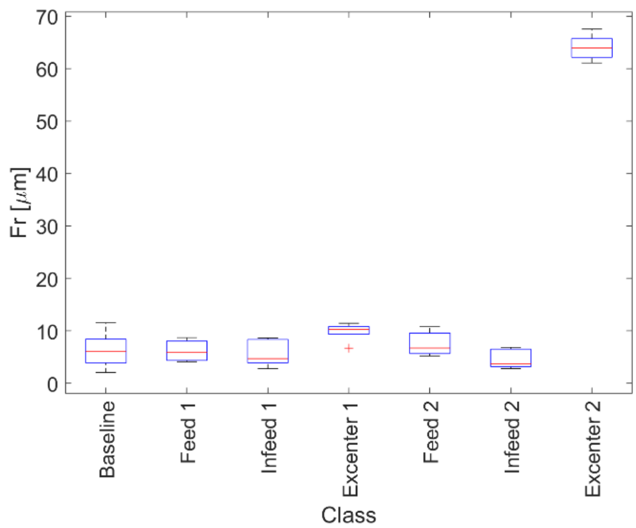
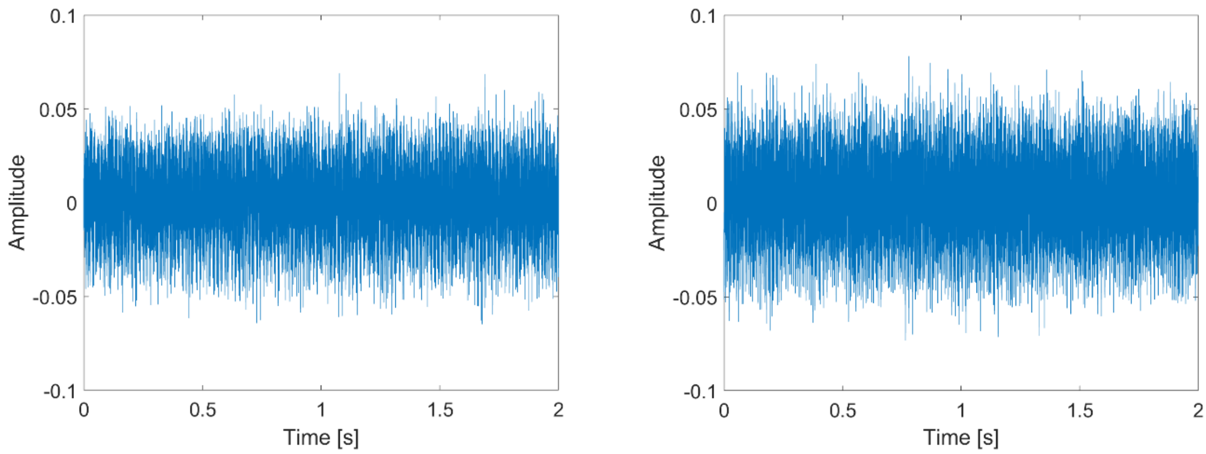


Figure 10: The measured runout error Fr of different test classes.

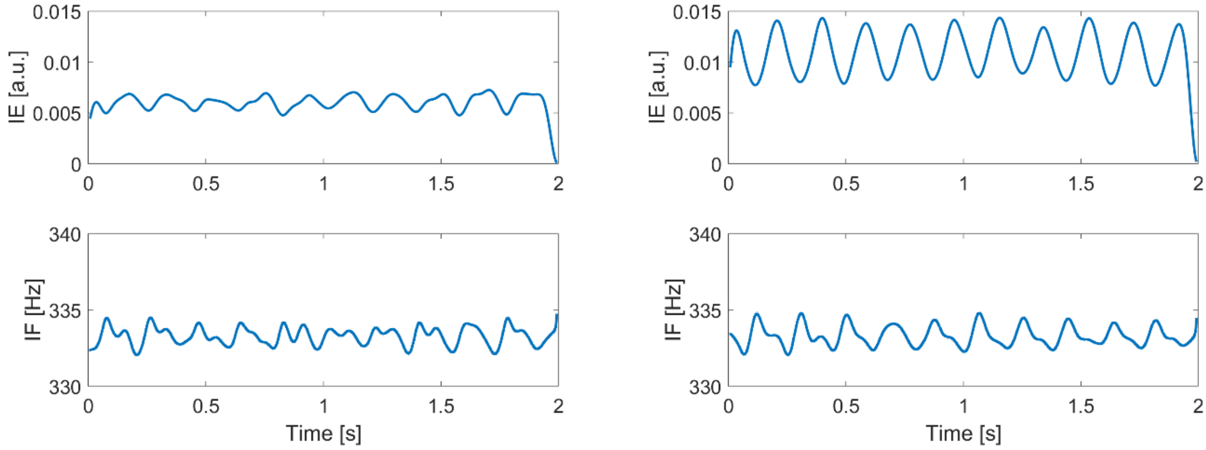
The measured vibration data of each test class has been processed and analysed using the algorithms proposed in Section 4. Initially, the raw vibration signal is segmented according to the event corresponding to Pass#1. The segmentation steps of the vibration signal corresponding to Pass#1 from the raw vibration signal have been visualised in Figure 7. Figure 11 shows the resulting signals of a *Baseline* test and an eccentric workpiece test (*Excenter#2*) after applying the Pass#1 segmentation algorithm. Visually, it is seen that the overall vibration amplitude of the eccentric workpiece test is slightly higher than that of the *Baseline* test. However, at this point, it is still difficult to justify the difference between the two signals.



(a) (b)

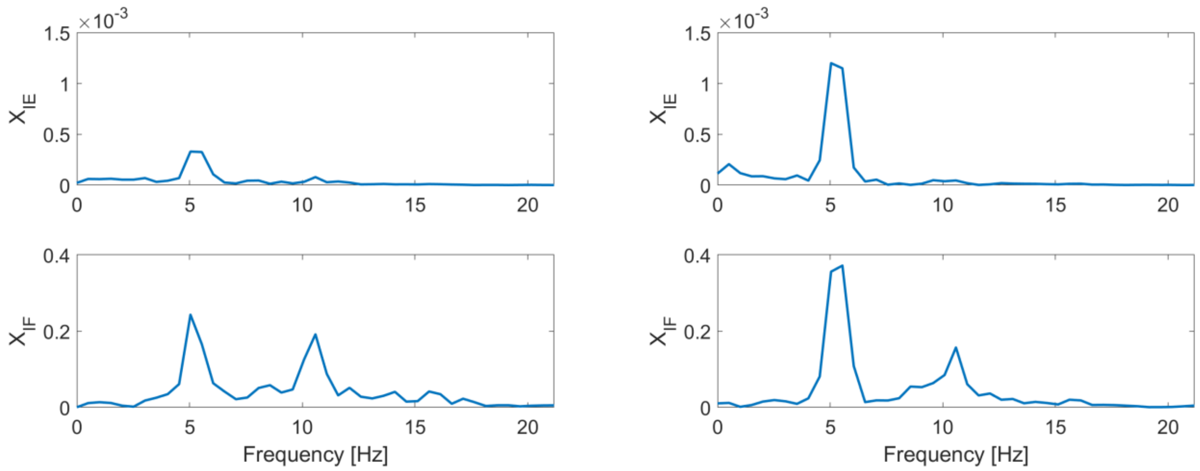
Figure 11: Typical vibration signals segmented from Pass#1 measured in the *infeed direction*/Y-axis of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test.

Once the vibration signal corresponding to Pass#1 has been segmented, the instantaneous envelope (IE) and instantaneous frequency (IF) are then estimated using Eq. (13) and (15), respectively. Figure 12 shows the resulting IE and IF signals calculated from the segmented signals of the baseline and eccentric workpiece test. As seen in the figure, the IE varies periodically in time and the IF around the fundamental gear mesh (GMF), ≈ 333 Hz, also varies periodically in time. It becomes more apparent from the figure that the proposed signal processing steps for computing the IE and IF help us to highlight the difference between the eccentric workpiece (*Excenter#2*) test and the *Baseline* test. Furthermore, the IE and IF are transformed into the frequency domain as shown in Figure 13. The difference between the two test conditions is even more highlighted by comparing the magnitudes of the IE and IF spectra, respectively X_{IE} and X_{IF}, around the workpiece rotational speed and the magnitudes of the higher order harmonics.



(a) (b)

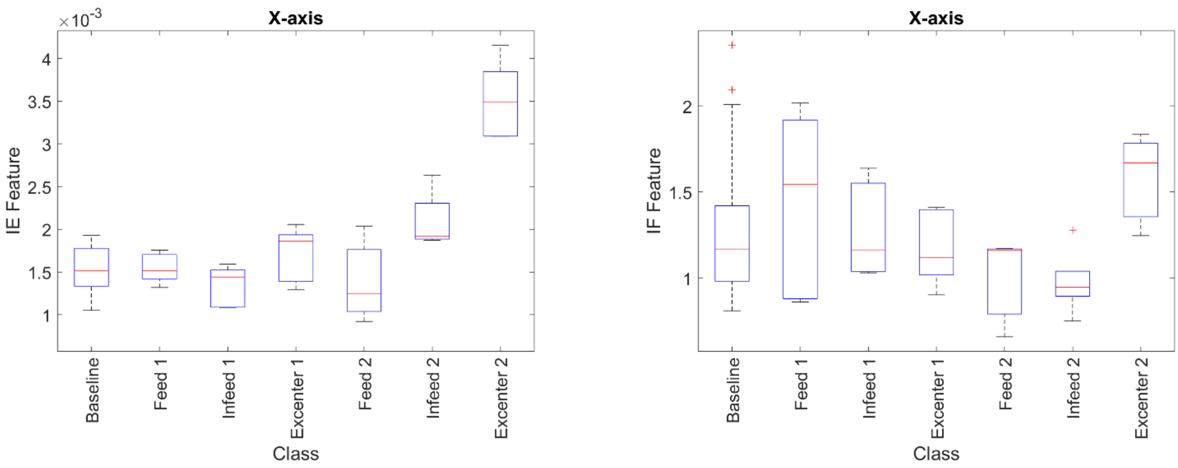
Figure 12: The estimated IE and IF of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test.



(a) (b)

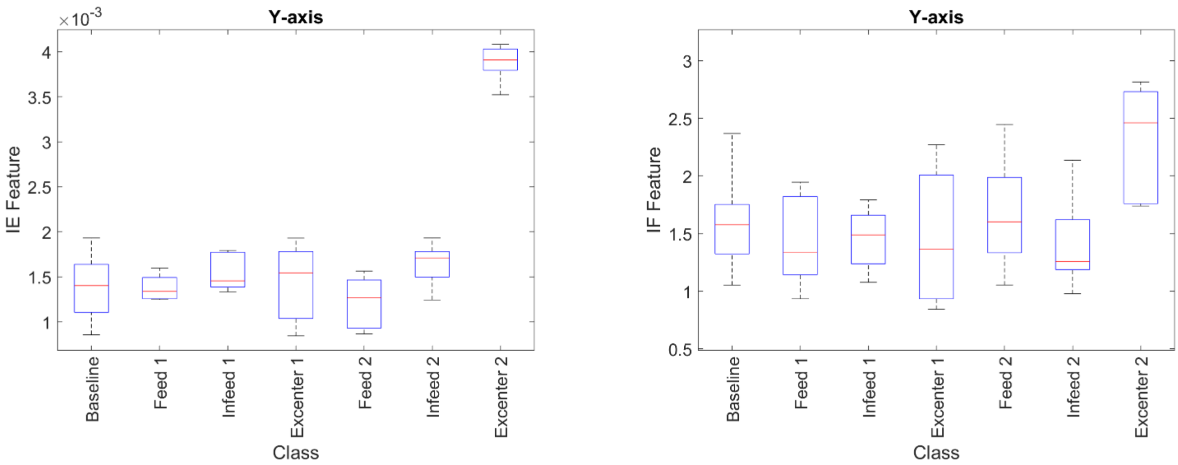
Figure 13: The frequency spectra of IE and IF of (a) *Baseline* test, (b) eccentric workpiece (*Excenter#2*) test. Note that the fundamental frequency of 5 Hz corresponds to the workpiece (gear) rotational speed f_g.

The final step is to compute the IE and IF features from the frequency spectra based on Eqs. (18) and (19). Figure 14, Figure 15 and Figure 16 show the IE and IF features extracted from the X-axis, Y-axis and Z-axis signal, respectively. As seen in the figures, in general, the IE feature shows a strong correlation with the measured runout error shown in Figure 10. On the other hand, the IF feature shows a weak correlation with the measured runout error. Notably, the strongest correlation of the IE feature with the runout error can be observed on the measurement data in Y-axis, which is in the *infeed* direction. This is expected since the workpiece eccentricity/runout errors affect significantly the grinding forces in the infeed direction, as discussed earlier in Section 3. A rather weak correlation between the IE feature and the runout error can be observed on the measurement data of Z-axis, which is in the *feed* direction.



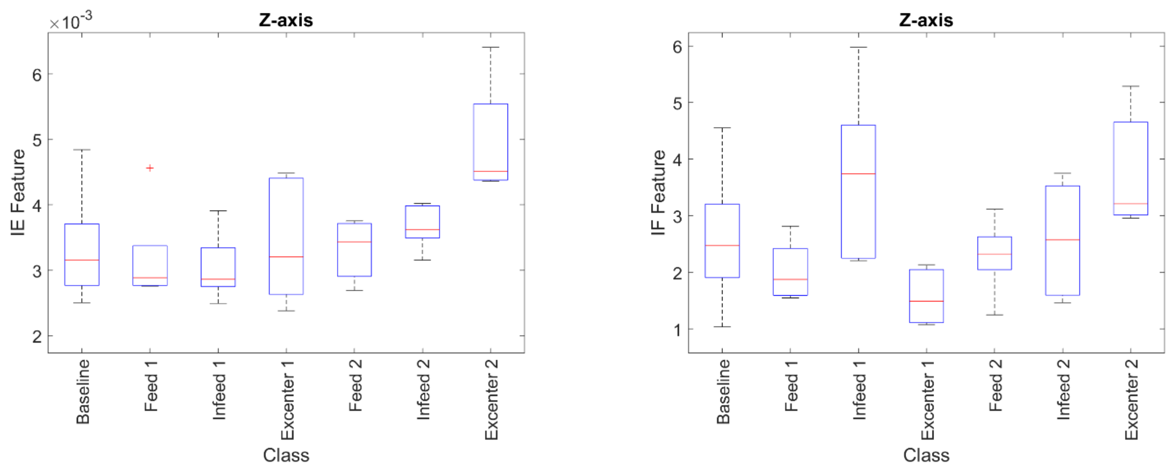
(a) (b)

Figure 14: (a) The IE feature and (b) the IF feature of the **X-axis** signal from all test conditions.



(a) (b)

Figure 15: (a) The IE feature and (b) the IF feature of the **Y-axis** signal from all test conditions.



(a) (b)

Figure 16: (a) The IE feature and (b) the IF feature of the **Z-axis** signal from all test conditions.

7. Conclusions and Outlook

Eccentric workpiece/runout error is one of the gear errors affecting the operating quality of gears and in particular the running noise behaviour of the end product. For gear manufactures, it is important to detect this fault to prevent sending defective gears to customers. A qualitative model developed in this paper has enabled us to derive a novel vibration feature for detecting eccentric workpiece/runouts. The feature extracted from the vibration signals measured at different test conditions/other fault types shows that it is sensitive and robust for detecting workpiece eccentricity of 40 microns. It is also shown in this study that the feature is insensitive to other types of gear grinding faults, which is important for diagnostics/root-cause analysis purposes. Furthermore, it is also shown in the study that the most sensitive measurement direction to detect workpiece eccentricity/runout errors is in the *infeed* direction.

Future work will be on the investigation of the minimum workpiece eccentricity/runout error that can still be detected by the proposed vibration feature during gear grinding processes.

Acknowledgements

This research was supported by both Flanders Make vzw, the strategic research centre for the manufacturing industry, and VLAIO in the frame of the VIBMON_ICON project.

References

[1] K. Gryllias, B. Kilundu, S. Devos, Y. Vonderscher, B. Vandewal, *Condition monitoring of gear grinding processes*, Surveillance 9, Fez (Morocco), 2017.

[2] B, Karpuschewski and I. Inasaki, Monitoring Systems for Grinding Processes, *Springer London*, 2006, 42,83-107.

[3] C. Brecher, M. Brumm, F. Hubner, *Approach for the calculation of cutting forces in generating gear grinding*, 9th CIRP Conference on Intelligent Computation in Manufacturing Engineering, Procedia CIRP 33, pp. 287 – 292, 2015.

[4] https://slideplayer.com/slide/6077265/ (Accessed on 9th of May 2019)

[5] https://www.kggear.co.jp/en/gear-technical-data (Accessed on 9th of May 2019)

[6] B.W. Kruszynski, S. Midera, Forces in Generating Gear Grinding – Theoretical and Experimental Approach, Annals of the CIRP Vol. 47, 1998.

[7] R.G. Parker, S.M. Vijayakar, T. Imajo, *Non-linear dynamic response of a spur gear pair: Modelling and Experimental Comparisons*, Journal of Sound and Vibration, 273(3), pp. 435-455, 2000.

[8] J. Ianniello, *Time delay estimation via cross-correlation in the presence of large estimation errors*, in IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 30, no. 6, pp. 998-1003, 1982.

[9] M. Feldman, Hilbert Transform Applications in Mechanical Vibration, *John Wiley & Sons*, Ltd., 2011.

Toward the quality prognostic of an aircraft engine workpiece in Inconel Alloy 625: case study and proposed system architecture

Antoine PROTEAU[1], Marc THOMAS[1] and Antoine TAHAN[1]

[1]École de technologie supérieure, Montréal, QC, Canada
antoine.proteau.1@ens.etsmtl.ca

Abstract

Manufacturing companies are under a constant pressure due to multiple factors: new competition, disruptive innovations, cost reduction request, etc. To survive, they must strive to innovate and adapt their business model to improve their productivity. Recent developments based on the concept of Industry 4.0 such as big data, new communication protocols and artificial intelligence provide several new avenues to explore. In the specific context of machining, we are working toward the development of a system capable of making the prognostic of the quality (in terms of dimensional conformance) of a workpiece in real time while it is being manufactured. The goal of this paper is to showcase a prototype of the data acquisition aspect of this system and a case study presenting our first results. This case study has been conducted at our industrial partner facility (Quebec, Canada) and is based on the manufacturing of an aircraft component made from Inconel alloy 625 (AMS5666). The proposed prototype is a data acquisition system installed on a 5 axis CNC machines (GROB model G352) used to acquire and to contextualize the vibration signal obtained from the CNC machine sensor. The contextualization of the data is a key component for future work regarding the development of a prognostic system based on supervised machine learning algorithms. In the end, this paper depicts the system architecture as well as its interactions between the multiple systems and software already in place at our industrial partner. This paper also shows preliminary results describing the relationship between the workpiece quality (in terms of respect toward the dimensional requirements) and the extracted features from the sensors signals. We conclude that it is now possible to do the diagnostic of a cutting operation. Additionally, with the same information we show that it is possible to quickly do the general diagnostic of the health state of the machine. Future work regarding this project will include data acquisition from a wider range of products (i.e. different shapes, materials, processes, etc.) and the development of a machine learning based prognostic model.

1 Introduction

Fuelled by the rapid evolution and introduction of new technologies and new philosophies such as Industry 4.0, the manufacturing industry is quickly transforming. This new manufacturing era brings a lot of possibilities to an industry that is under constant pressure for cost reduction and better quality caused by a global competition [1]. In the specific context of machining; automation and methodologies such as lean manufacturing were the go to solutions to decrease process cost and improve quality output. However, in this new age, possibilities brought by artificial intelligence, more affordable technologies such as sensing technologies and collaborative robotic offer new improvements directions.

In this context, the objective of our research project is to see if we can connect the operational information of a machining process to the physical phenomenon happening during the machining of a workpiece on a CNC machine in order to be able to predict the quality of this workpiece in real-time. Thus, the objective of this paper is to propose a data acquisition system architecture based on the prototype we built, showcase that it is now possible to do the diagnostics of a cutting operation with this system and that we are now able to put in relationship the quality, in terms of the conformity towards a workpiece's G&DT specifications, and the physical phenomenon happening during the machining process.

In a general manufacturing context, attempts have been made to try to predict the quality of a production process. For instance, Wang [2] tried to predict the quality of a chemical batch process operation. However, their results are based on simulated data and not industrial data such as what we propose. Closer to the machining industry, through our exploration of the literature we have not yet found authors who have proposed a methodology to predict the quality of a whole machining process and the produced workpiece. Nevertheless, we can find articles related to the prognostic of some aspect of a machining process such as predicting the surface roughness. In that context, Benardos, Vosniakos [3] propose a review of the works that have been done in that domain and more recently, Balamurugamohanraj et al. [4] used a machine learning approach and data from an accelerometer to predict the surface roughness in terms of its Ra value.

Even though we have not found many publications with industrial application of prognostic methodology related to the quality of a workpiece, we clearly see an interest for the concept of prognostic in the manufacturing industry. Reviews and publications by authors such as Vogl et al. [5], Wang [6], Peng et al. [7],Lee et al. [8] are all dedicated to the state of the prognostic concept or the proposal of a framework related to manufacturing. Thus, we are not the only one with interest in applying these concepts to a manufacturing context. Still, one of the biggest challenge to the industrial application of such concept and the development of prognostic methodologies is the access to data of good quality and in sufficient quantity. The foremost challenge is addressed in this article.

We also see that, in our research domain, the interest related to applying prognostic methodologies is strong in fields related to tool wear prediction and condition-based maintenance. For instance, Proteau et al. [9] showed that it is possible to predict the tool wear with a Long Short-Term Memory (LSTM) neural network. Balan, Epureanu [10, 11] and Aghazadeh et al. [12] also proposed different methodologies based on artificial intelligence approaches to monitor and predict the cutting tool condition. Related to condition-based maintenance, Waqar, Demetgul [13] and Aydin, Guldamlasioglu [14] also suggested methodologies based on artificial intelligence to predict the state of an equipment or a component (e.g. bearing, gears, etc.).

To improve the state of this research domain and to make a step toward the industrial application of prognostic methodologies, this paper will present our most recent work to show that it is now possible to put the workpiece quality in relationship with the physical phenomenon happening during the machining process. We also want to prove that we are now making a step forward to go from being able to diagnose a cutting operation toward being able to predict the quality of that process. To do so, this article is structured as follows: section 2 will introduce our research partner as well as our research environment and equipment. Then, in section 3, we propose a data acquisition (DAQ) system architecture and describe the dataset built. In section 4, we present our signal processing methodologies and the different features that we extracted from the acquired signals. In section 5, we detail our results and show that it is now possible to do the diagnostic of a cutting operation as well as working toward the prognostic of the quality of a workpiece. Finally, in section 6, we make our conclusions.

2 Research environment

To pursue this research project, we are collaborating with an industrial partner: APN Inc.[1] APN is a leader of the machining industry as well as at the forefront of the Industry 4.0 movement in Quebec, Canada. They are specialized in the machining of complex products in exotic material (i.e. titanium, Inconel, etc.) for the aerospace and high-tech industry.

In our research context, our work was done on a 5 axis CNC machine made by GROB, model G352 (see Figure 1). This machine was acquired in 2017.

[1] More information here : http://apnglobal.ca/en/apn/

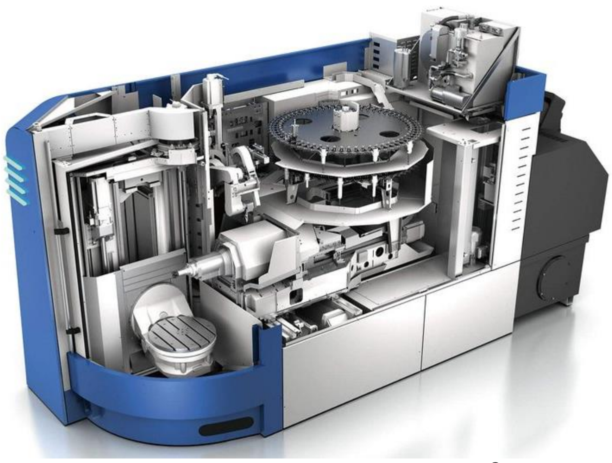


Figure 1 GROB G352 CNC machine[2]

It is also important to state that, to be able to acquire a vibration signal, we worked with the GROB employees to have access to the accelerometer already installed into the machine's spindle. Thus, the signal was acquired through an IFM VSA004[3] accelerometer on which the signal was amplified with a Phoenix Contact signal conditioner model MACX MCR-UI-IU[4]. From Figure 2, we can see where the accelerometer was installed by the manufacturer (as indicated by the bubble #1). This information was provided by the GROB documentation available at APN Inc. In the next section, we present our DAQ system architecture.

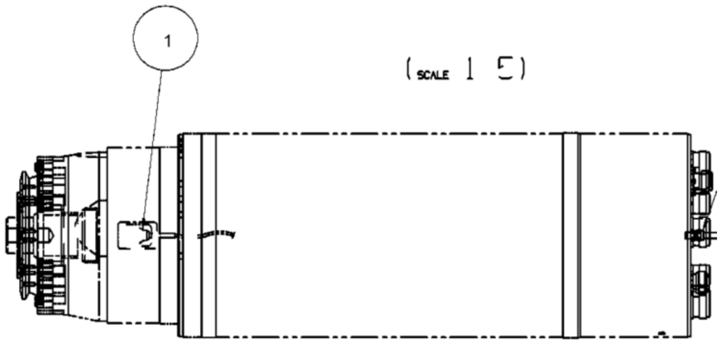


Figure 2 Accelerometer location in the spindle

3 Data acquisition system architecture

One of our hypothesis is that, in order to be able to predict the quality of a workpiece based on the physical information of the CNC machine, we must contextualize the signals acquired from sensors. Therefore, we developed a data acquisition system to automatically execute this operation. Figure 3 shows the contextualization of the data in terms of its relationships.

[2] Picture source : https://www.grobgroup.com/en/products/product-range/universal-machining-centers/milling-centers/g350/
[3] Specifications : https://www.ifm.com/ca/en/product/VSA004
[4] Specifications : https://www.phoenixcontact.com/online/portal/ca?uri=pxc-oc-itemdetail:pid=2811446&library=caen&tab=1

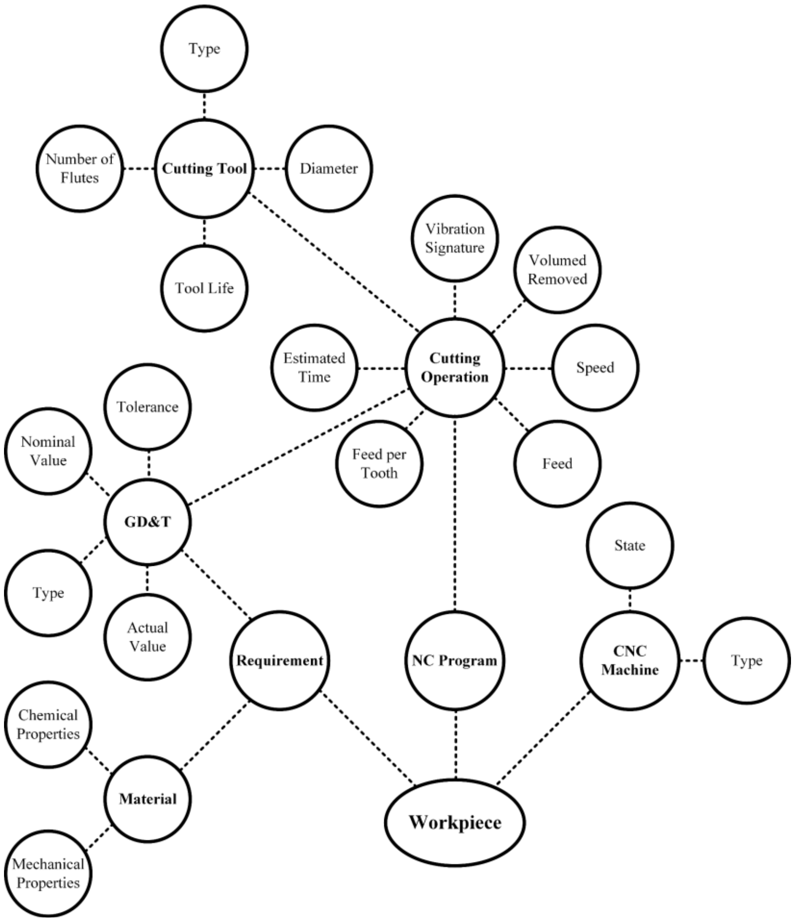


Figure 3 Relationships between data sources

Whit this figure, we can see that, through this system, it is possible to create a relationship that goes from the workpiece requirements (including the actual measurements made on a finished workpiece) up to the vibration signature of a specific cutting operation. This means that at every moment during the machining process, we can know which cutting operation was being executed, its vibration signature, what was the cutting tool and its cutting parameters as well as which GD&T was influenced. To achieve these relationships, multiple data sources must be integrated. Table 1 shows the source of each data types.

Data types	**Sources**
Workpiece Requirements	APN Quality System (From the technical drawing)
Actual Measurements	APN Quality System
Material Properties	APN Quality Documents System
CNC Machine information	Machine controller through OPC Protocol
Cutting Operation	CAM Software
Cutting Tool	CAM Software
Vibration signal	Accelerometer and National Instruments Card

Table 1 Data sources by data types

To automatically integrate these data sources and create the relationships between the data, we developed an acquisition system that had to take into account the state of the machine (cutting or not, on idle, etc.). To illustrate our acquisition system, Figure 4 shows an overview of the acquisition process. On this figure, we can also see the isometric view of the workpiece as well as the quality data flow.

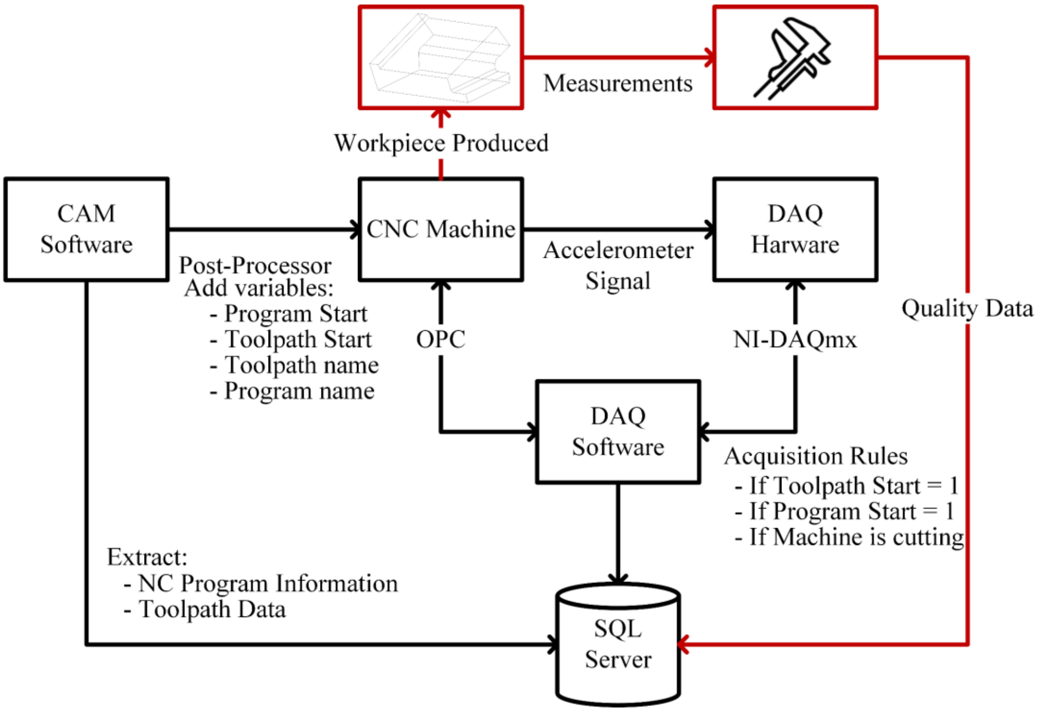


Figure 4 DAQ System architecture

We worked with our industrial partner to modify their post-processor program to add four variables: when the NC Program start/stop, when a cutting operation start/stop, the cutting operation name and the NC Program name for reference. This modification allowed us to control the behaviour of the National Instruments data acquisition card by sampling the signal only when the machine was actually cutting the workpiece. However, since we are in a production environment, different machine states can also arise: the machine is in idle during a cutting operation, an alarm is raised, etc. Therefore, in our acquisition rules, we added some logic based on variables extracted in real-time from the CNC machine controller through the OPC communication standard. The reader can refer himself to the Siemens Sinumerik 840D SL documentation for a complete list of all available variables.

During the acquisition process, the raw signal is thus contextualized and attached to the current workpiece and the current cutting operation being executed on the CNC machine. This, consequently, gives us a contextualized vibration signal suited to model the relationships between the physical and operational data (the model input) and the actual measurement in terms of GD&T (the model output). The next section presents an overview of the data collected.

3.1 Dataset overview

To prove the concept and functionalities of our proposed system, we conducted a first acquisition process. The acquisition was made during the machining of an Inconel 625 (AMS5666) workpiece intended for the aerospace industry. During our acquisition process, we were able to cover the entirety of the machining process which means:

- 22 different cutting tools;
- 140 cutting operations of multiple types;
- 135 GD&T to be respected for a workpiece to be considered conform.

The acquired data covers five finished workpieces which represents approximately 13 hours of machining process. Unfortunately, due to industrial constraints, we were not able to gather more workpiece. However, this information is sufficient to prove our concept and start our analysis. In this article, we focused on one specific operation where measurements were made for every workpiece to showcase our results. Information related to this operation can be found in Table 2. Information regarding the cutting tool used during the cutting operation can be found in Table 3. The cutting tool was new at the beginning of the machining process and was not changed during the machining of the five workpieces. For visualization, Figure 5 shows the cutting operation strategy obtained from the CAM software and Figure 6 shows the difference between the finished workpiece and the raw material used. Due to confidentiality, the 3D model shown in this paper have been redesign to showcase the overall shape of the workpiece and not the actual geometry.

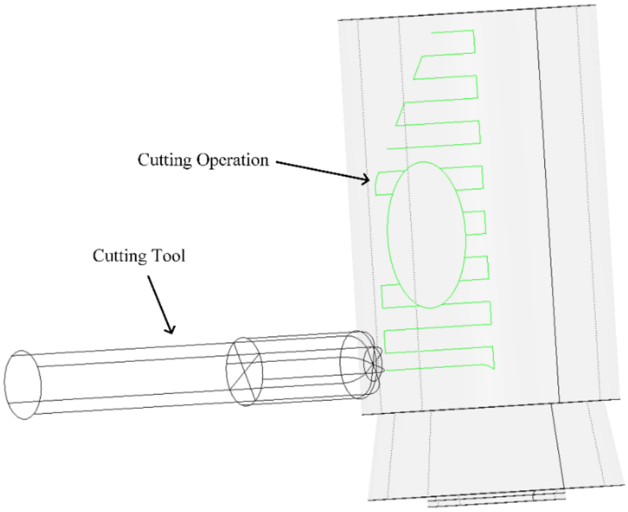


Figure 5 Cutting operation *OP_510*

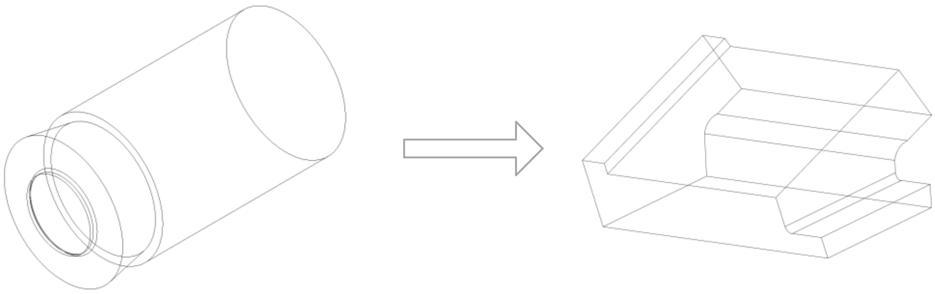


Figure 6 Difference between the raw material and the finished workpiece

Name	**Type**	**Spindle speed [RPM]**	**Feed [mm/s]**	**Est. cutting time [min]**	**Volume removed [cm^3]**	**Coolant pressure [MPa]**
OP_510	Face Milling	1047	233.934	2.5	0.078	6

Table 2 Cutting operation information

Name	**Type**	**Diameter [mm]**	**Radius [mm]**	**Number of flutes**
EMR0.500R0.125L4LG1.100	Radius End Mill	12.7	3.175	4

Table 3 Cutting tool information

By connecting our system to the quality system of APN, we are able to associate the cutting operation with the specifications (GD&T) influenced by that operation. These associations are made by expert employees at APN. Thus, for this project, we assumed that the associations are good. Through these associations, we know that the *OP_510* operation influences the specification #15. Details about this specification is found in Table 4.

Number	**GD&T type**	**Minimum value [µm]**	**Maximum value [µm]**	**Severity**	**Illustration [µm]**	**Inspection tool used**
15	Flatness	0	25.4	Critical	⏥ 25.4	CMM

Table 4 Details of the GD&T specification #15

The interpretation of this type of GD&T was made according to the standard Y14.5 [15]. The reader can refer to ASME [15] for further details.

For the five workpieces, Figure 7 shows each actual measurements made by the operator after each workpiece was produced.

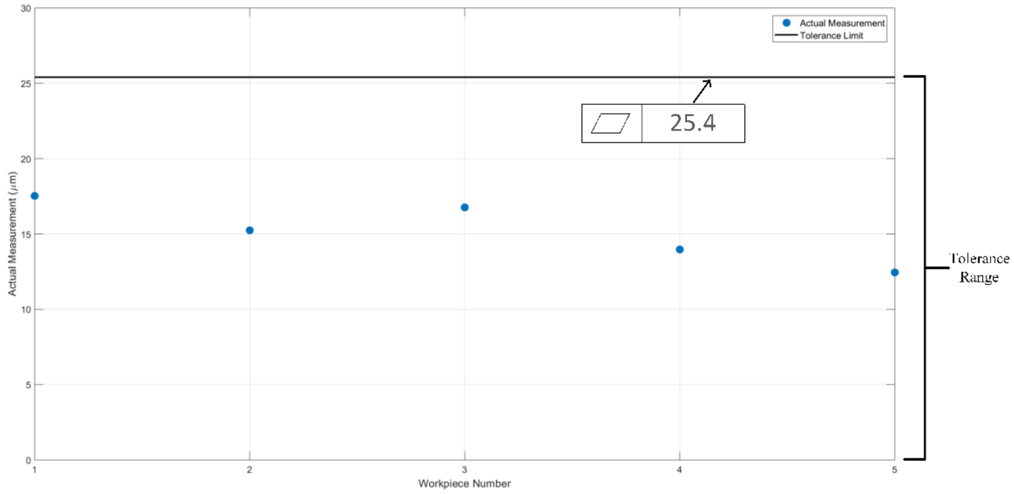


Figure 7 Actual measurement per workpiece for the specification #15

The next section will present the signal processing methodology applied to the data acquired.

4 Signal processing methodology

We have shown in the previous section that we can now acquire a signal that is well contextualized. The objective of this section is to describe our signal processing methodology in order to be able to do the diagnostic of the machining process as well as the production equipment itself; taking a step toward a predictive methodology.

Our methodology is segmented in two sections: a time domain methodology and a frequency domain methodology. Once acquired, each signal sampled file is cleaned and has several features extracted. The signal features used are described below and are chosen according to the work of Lei et al. [16], Elattar et al. [17] and Abellan-Nebot, Romero Subirón [18].

4.1 Time domain

To describe the signal in the time domain, we used equation (1) to equation (5) which refer respectively to the Root Mean Square (RMS), the Kurtosis (K), the Peak value ($Peak$), the Peak-to-Peak value (PTP) and the Crest Factor (CF).

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{1}$$

$$K = \frac{\sum_{i=1}^{N}(x_i - \bar{x})^4}{\frac{1}{N}\left[\sum_{i=1}^{N}(x_i - \bar{x})^2\right]^2} \tag{2}$$

$$Peak = \max(x) \tag{3}$$

$$PTP = \max(x) - \min(x) \tag{4}$$

$$CF = \frac{Peak}{RMS} \tag{5}$$

Where x is a signal of N samples, x_i is the value of i^{th} sample and $\overline{x}$ is the average of x.

4.2 Frequency domain

In the frequency domain, we are interested in following the evolution, through time, of the cutting tool frequency (A_{CT}). This is because the degradation of the tool is one of the major cause of the degradation of a machining process. To do so, we used equation (6). This equation is similar to the RMS equation in the sense that we could interpret its result as being the energy content of the signal around a specific frequency.

$$A_{CT} = \sqrt{\sum_{i=a}^{b} A_i^2} \tag{6}$$

Where A_i is the amplitude of the signal at the i^{th} frequency and a and b are the two corners of the window. In our case, we assigned $a = 64$ Hz and $b = 75$ Hz, which correspond to ±5 Hz around the frequency of interest (f_i). We are also interested in making the same measurement at the different harmonics of the cutting tool frequency, thus we also applied the same equation at the 2nd, 3rd, 4th and 5th harmonics ($2f_i$, $3f_i$, $4f_i$ and $5f_i$).

To improve the physical meaning of this feature we use the work of Proteau et al. [9]. In Proteau et al. [9], the authors proposed an adaptation of the specific cutting energy (SCE, k_c) metric first established by Debongnie [19]. In their work, the authors used their version of the SCE to show that it can adequately represents tool wear degradation. Their version is defined by equation (7).

$$k_c = \frac{P_{Tool}[\mathrm{W}]}{Q[\mathrm{cm}^3\mathrm{s}^{-1}]} = \frac{[\mathrm{Js}^{-1}]}{[\mathrm{cm}^3\mathrm{s}^{-1}]} = [\mathrm{Jcm}^{-3}] \tag{7}$$

Where P_{Tool} is the power [W] consumed by the cutting tool and Q is the material removal rate express in $\mathrm{cm}^3\mathrm{s}^{-1}$.

SCE is therefore defined in terms of the energy required to remove and keep a certain rate of material removal in a specific material (aluminum, Inconel, etc.). The reader can refer himself to Proteau et al. [9] for the details. Since we do not have the actual power transmitted to the cutting tool, we can estimate this value by using the energy contained in the signal at the frequency related to the cutting tool (A_{CT}). For the same material and a constant material removal rate, k_c should be constant. In case of tool wear, k_c increase through time. The next section will present the results of our analysis.

5 Results and discussion

We first present the results of our analysis in the time domain. Figure 8 shows the values of the RMS, $Peak$ and PTP values through time for each workpiece.

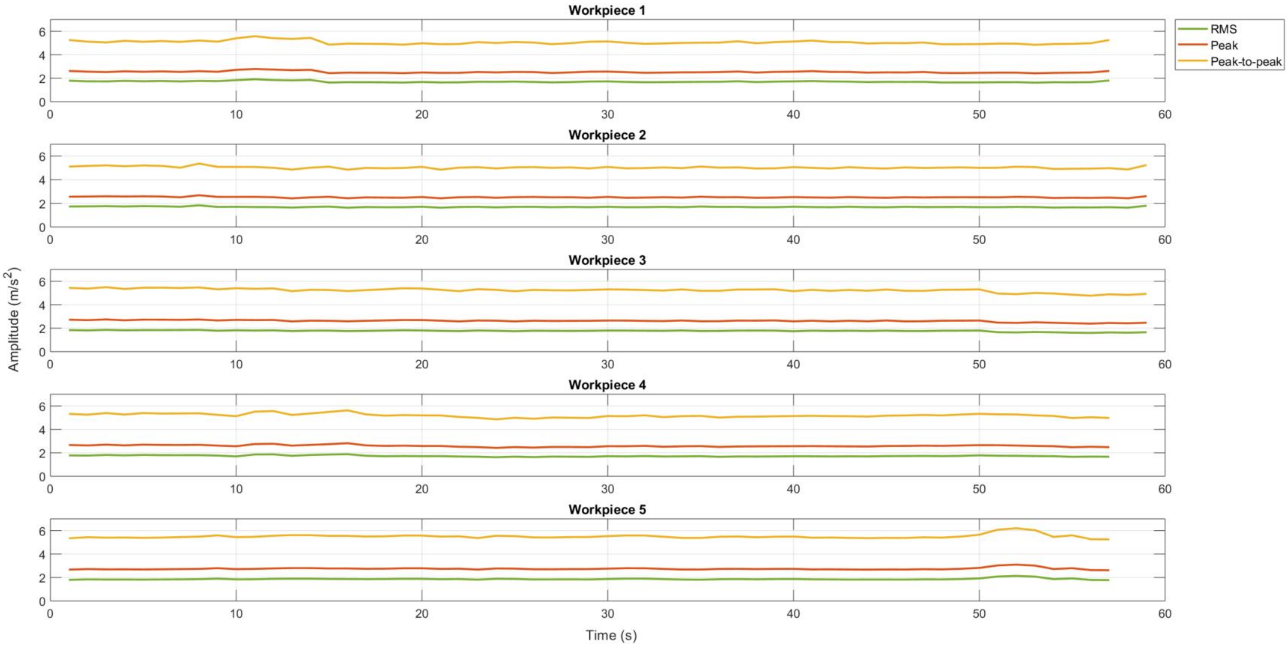


Figure 8 RMS, $Peak$ and PTP values through time per workpiece

Then, Figure 9 presents the average values for the RMS, $Peak$ and PTP value per workpiece. We also included the evolution of the actual measurement per workpiece for the specification #15 to see if there is a direct relationship between the evolutions of the two phenomenon that could be visually witnessed.

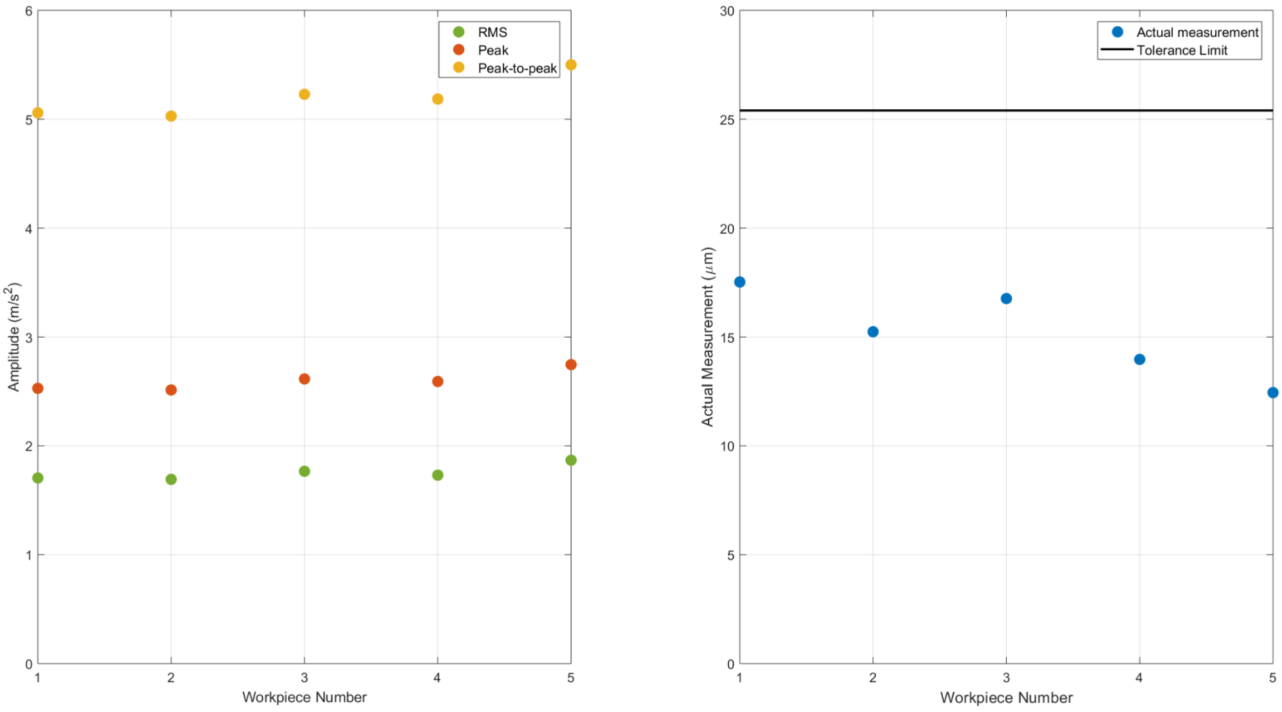


Figure 9 RMS, $Peak$ and PTP values per workpiece and actual measurement per workpiece

We did the same analysis with the kurtosis and the crest factor. Results are shown in Figure 10 and Figure 11.

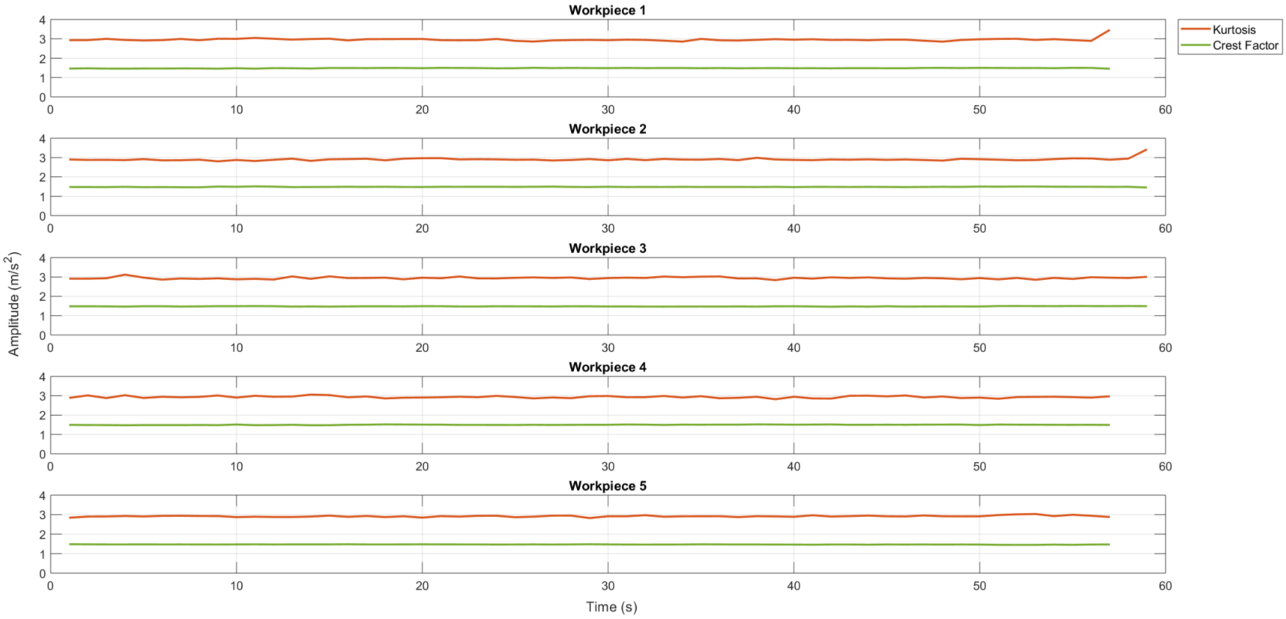


Figure 10 K and CF values through time per workpiece

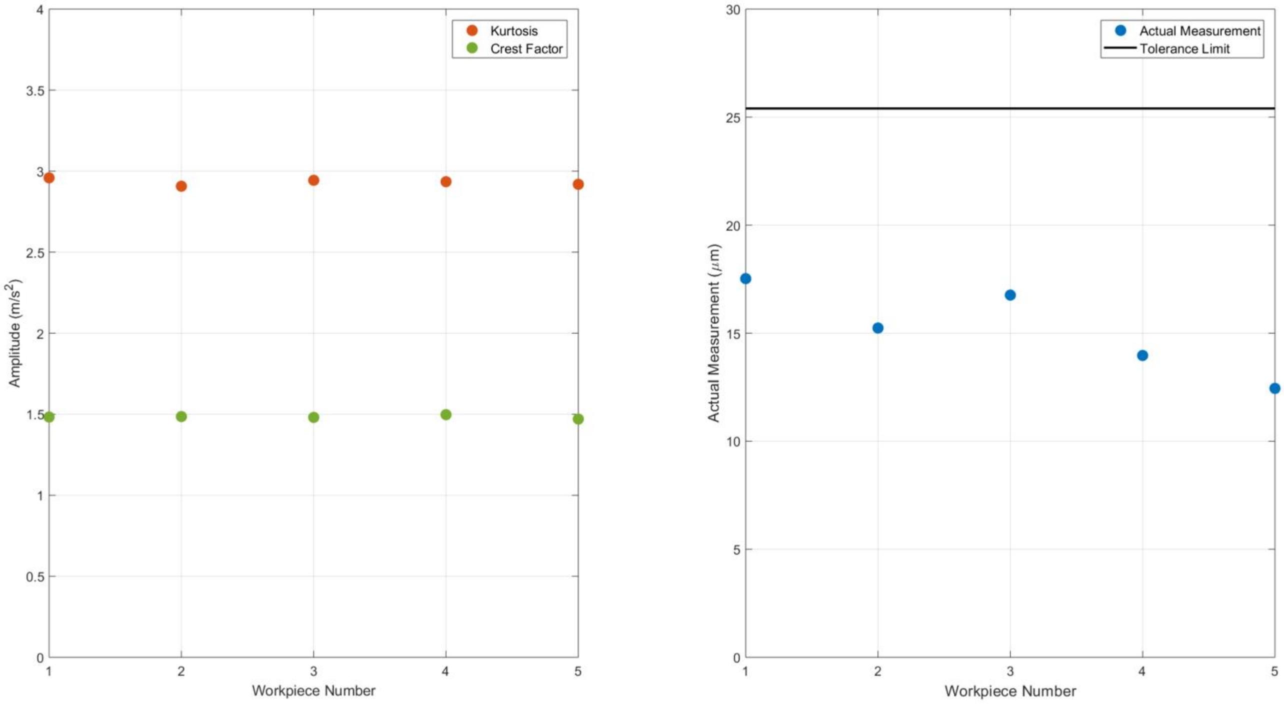


Figure 11 K and CF values per workpiece and actual measurement per workpiece

We then did the same analysis in the frequency domain. Results of the SCE values for each harmonics are shown in Figure 12 and Figure 13.

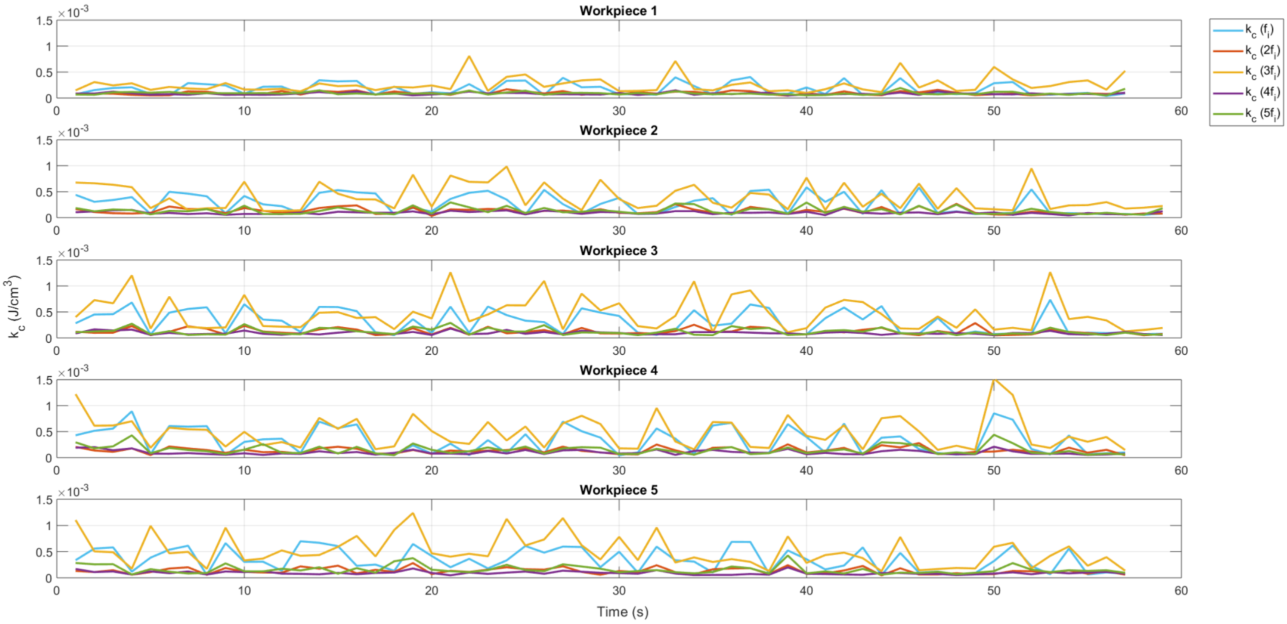


Figure 12 k_c values for each harmonics of f_i through time per workpiece

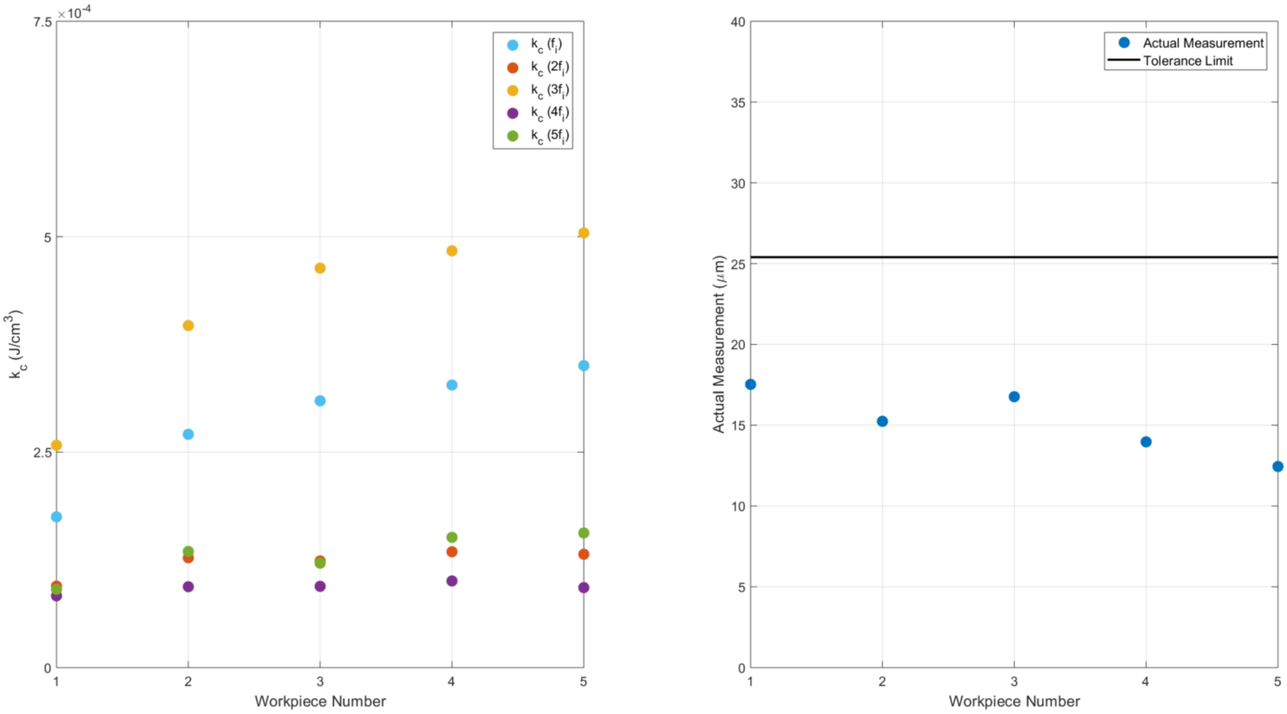


Figure 13 k_c values for each harmonics of f_i per workpiece and actual measurement per workpiece

Finally, we also did a time-frequency analysis where we looked at the frequency domain of the signal through time. Figure 14 presents the spectrogram we obtained. The white lines represent the separation between each workpiece; starting to the left with workpiece 1 up to the right with workpiece 5. In a) we gave the spectrogram for the frequencies between 0 and 400 Hz and in b) for the frequencies 400 to 1000 Hz. Most frequencies of interest are located in the range of 0 to 1 kHz.

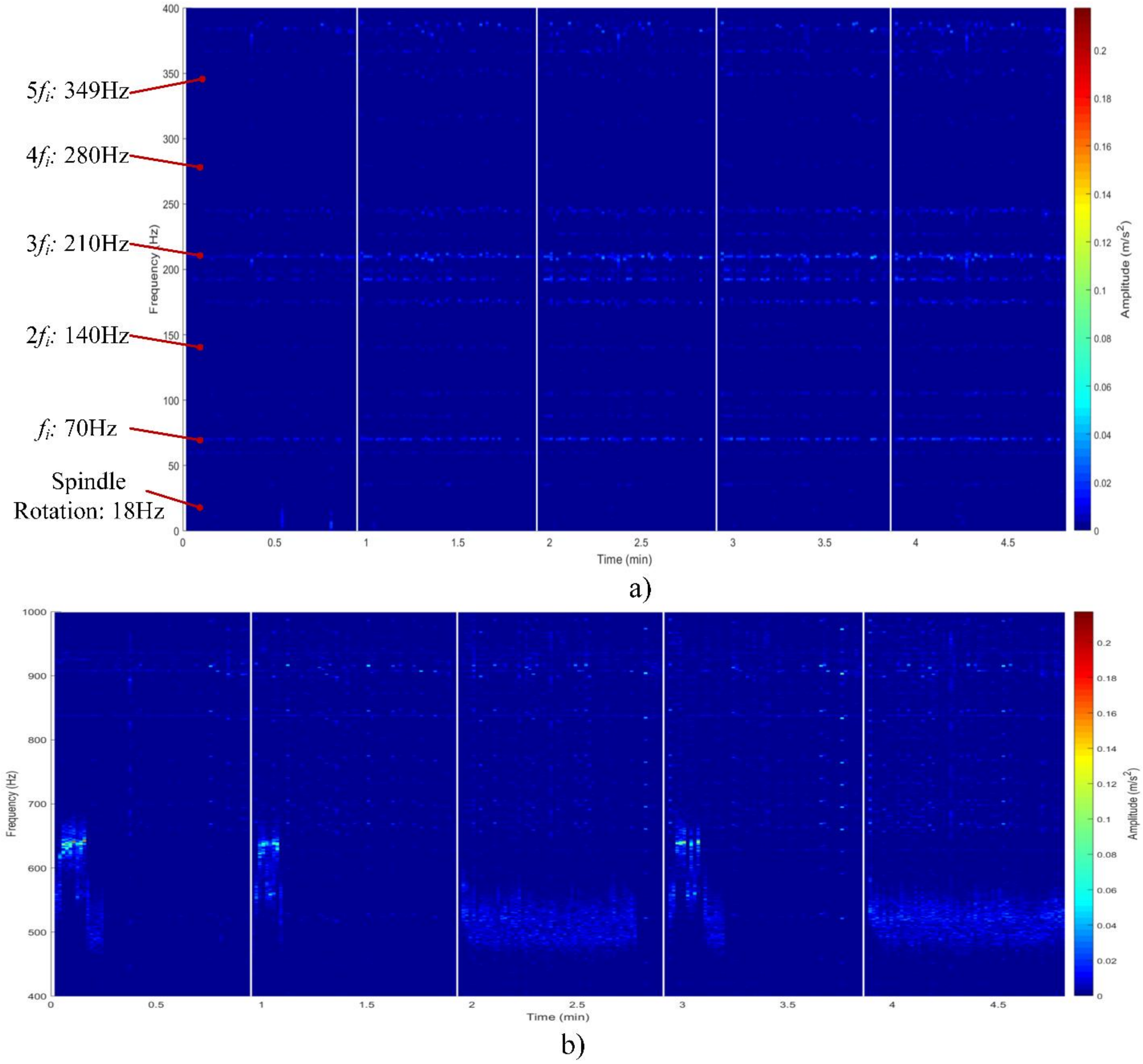


Figure 14 Time-frequency analysis for all the workpiece

From Figure 8, we can see that, for all workpieces, most values are comprised between an amplitude of 0 and 6 m/s^2; with some peaks during the machining process of each workpiece. However, we cannot clearly see that there was either a degradation or an improvement regarding the machining process in a part-to-part point of view. Also, when we look at the average values per part in Figure 9, we cannot clearly state that there is a direct relationship between the actual measurement and the evolution of the RMS, $Peak$ or PTP.

We could conclude the same thing regarding the evolution of the K and CF through time with the results shown in Figure 10 and Figure 11. However, it is interesting to also use these results to do the diagnostic of the machine health state. Based on these results, we could conclude that the machine is in a good health state. From Thomas [20], a kurtosis value around a value of 3 means a random signal, hence a machine in good health where no spike or impact were recorded. Values higher than that would start to indicate that impacts are being recorded. This is also supported by the values of the crest factors which are low and near the value indicating a good condition ($CF = 1.41$). We also tried to look at the fundamental frequencies related to the bearings installed in the spindle. However, since the machine and its component are relatively new, the amplitudes related to the typical fault (FTF, BPFI, etc.) do not stand out. This could indicate that they are in good condition and that their signal is lost in the noise of the machine during the machining process. A diagnostic when the machine is not cutting could probably help us identify with better accuracy these frequencies.

From Figure 12, we can see that the variation in terms of amplitude seems to increase between the workpiece 1 and workpiece 5. This would seems to be consistent with the claim of Proteau et al. [9] that the energy required to keep a material removal rate is increasing with tool wear. When we look at Figure 13, we can see that the values are increasing with every workpiece; for the first and third harmonics (f_i and $3f_i$). The 2nd, 4th and 5th harmonics ($2f_i$, $4f_i$ and $5f_i$) seems to have a low amplitude throughout the data we collected.

Moreover, when we look at the scale of the amplitude of the data, we can see that they are pretty low. This is somewhat counter intuitive to our belief. We believed that because the Inconel 625 is a very hard and difficult material to work with, we would have seen very high amplitudes due to the force required to remove the material. It was not possible to get the exact depth of cut used in this operation, therefore, maybe the engineers responsible of this product at APN used a very low depth of cut parameter in order to create less friction between the material and the cutting tool in order to facilitate the machining process.

When we look at the spectrogram shown in Figure 14, we can quickly see that the overall frequencies' amplitudes are consistent with our previous claim; it is low across most frequencies. We can still detect some frequencies of interest such as the spindle rotation (1047 RPM = 17.45 Hz), the cutting tool frequency (with 4 flutes: 69.8 Hz) and its harmonics (139.6 Hz, 209.4 Hz, 279.2 Hz and 349 Hz).

Aside from these specific frequencies, this low amplitude claim seems to hold true except for some spontaneous peak between 600 and 700 Hz. In fact, if we look at the graph in b), we see a phenomenon where we have not yet found the source. No video recording was made during the acquisition process. This kind of data would surely help us to correlate such phenomenon with actual events during the machining process. Additionally, this phenomenon is not consistent across all workpiece. Cutting parameter and overall machining strategy were not changed between the workpieces, hence we would have expected a similar pattern for each workpiece. However, we can denote two patterns: one related with workpiece 1, 2 and 4 and the other with the workpiece 3 and 5. The peaks in amplitude are also related to the first pattern for workpiece 1, 2 and 4.

In a diagnostic point of view, our conclusion related to the kurtosis and crest factor values seems to hold in the frequency domain since we do not seem to detect traces of impact during the utilization of the equipment.

The objective of this paper was to demonstrate that we can now have access to data allowing us to describe and diagnose a machining process and its cutting operations as well as making a step toward being able to do the prognostic of the overall quality of a workpiece. With the results shown in this section we can conclude that the propose data acquisition architecture enable us now to adequately contextualize, in real-time and automatically, signals acquired through sensing devices. However, when we look at the final objective of our project; that is the prognostic of the quality of a workpiece in term of the respect of its GD&T requirements, we have not seen a clear linear relationship or pattern between the cutting operation vibration signal and the evolution of the actual measurement of the specification #15 neither in the time and frequency domain. Throughout this article, we have been looking at one operation influencing the specification #15; in fact, there is a total of 11 cutting operations influencing this specific requirement. In other words, we conjuncture that a clear linear relationship cannot be establish between only one operation and the evolution of a specific requirement. On the contrary, it is maybe the "sum" or sequence of all these operations that could influence the conformity of a workpiece specific requirements. In other words, all the variations across all these operations could explain the evolution of a specification. Consequently, we believe that only through a machine learning approach we could be able to predict the quality of a workpiece. Our strategy to apply such approach to this research project still hold to this point.

6 Conclusion

To conclude this article, we wanted to showcase our data acquisition system architecture and demonstrate that we can now adequately contextualize a vibration signal to better do the diagnostic of a cutting operation to, in the end, facilitate the development of a prognostic methodology for the quality of a workpiece. We believe that we have successfully achieve these objectives by showing multiple results related to the cutting operation *OP_510*. However, we have not yet been able to showcase a linear relationship between the vibration signal of this operation and the evolution of the quality of the workpiece. The use of a machine learning approach could probably help us achieve this objective. Further work in order to close the gap between our current status and our final objective to predict the quality of a workpiece will include adding sensors to the GROB CNC machine: a tri-axial accelerometer, an acoustic emission sensor, current and voltage sensors to the motor of the spindle as well as the ones of the three main working axis of the CNC machine and try to apply cyclostationarity analysis based on the work of Lamraoui et al. [21]. We will also expand our system capacity to have it works in a more autonomous way and we will finally use and apply multiple machine learning approaches to perform sensors fusion and the actual prediction of a workpiece quality.

Acknowledgments

The authors would like to thank APN Inc.: their owners, Jean and Yves and all their employees related to this project for their support and time in helping our research team. Special thanks to Alain, Jonathan, Michel, David, Daniel, Raphael, Maxim and Mathieu for the many questions they answered and their help to make this project happen. We would also like to acknowledge the financial support of the *Fonds de Recherche du Québec – Nature et Technologies* (FRQNT) to this project through Grant #257668.

References

1. Kohler D, Weisz J-D (2016) Industrie 4.0 *Les défis de la transformation numérique du modèle industriel allemand*. France
2. Wang D (2011) *Robust Data-Driven Modeling Approach for Real-Time Final Product Quality Prediction in Batch Process Operation*. IEEE Transactions on Industrial Informatics 7 (2):371-377
3. Benardos PG, Vosniakos GC (2003) *Predicting surface roughness in machining: a review*. International Journal of Machine Tools and Manufacture 43 (8):833-844
4. Balamurugamohanraj G, Vijaiyendiran K, Mohanaraman P, Sugumaran V (2016) *Prediction of Surface Roughness Based on Machining Condition and Tool Condition in Boring Stainless Steel-304*. Indian Journal of Science and Technology 9 (47):6
5. Vogl GW, Weiss BA, Helu M (2016) *A review of diagnostic and prognostic capabilities and best practices for manufacturing*. Journal of Intelligent Manufacturing:1-17
6. Wang K-S (2013) *Towards zero-defect manufacturing (ZDM)—a data mining approach*. Advances in Manufacturing 1 (1):62-74
7. Peng Y, Dong M, Zuo MJ (2010) *Current status of machine prognostics in condition-based maintenance: a review*. The International Journal of Advanced Manufacturing Technology 50 (1):297-313
8. Lee J, Jin C, Bagheri B (2017) *Cyber physical systems for predictive production systems*. Production Engineering 11 (2):155-165
9. Proteau A, Tahan A, Thomas M (2019) *Specific cutting energy: a physical measurement for representing tool wear*. The International Journal of Advanced Manufacturing Technology:1-10
10. Balan GC, Epureanu A (2008) *The monitoring of the turning tool wear process using an artificial neural network. Part 1: The experimental set-up and experimental results*. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222 (10):1241-1252
11. Balan GC, Epureanu A (2008) *The monitoring of the turning tool wear process using an artificial neural network. Part 2: The data processing and the use of artificial neural network on monitoring of the tool wear*. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 222 (10):1253-1262
12. Aghazadeh F, Tahan A, Thomas M (2018) *Tool Condition Monitoring Using Spectral Subtraction Algorithm and Artificial Intelligence Methods in Milling Process*. International Journal of Mechanical Engineering and Robotics Research 7 (1):30-34
13. Waqar T, Demetgul M (2016) *Thermal analysis MLP neural network based fault diagnosis on worm gears*. Measurement 86:56-66
14. Aydin O, Guldamlasioglu S *Using LSTM networks to predict engine condition on large scale data processing framework*. In: 2017 4th International Conference on Electrical and Electronic Engineering (ICEEE), 8-10 April 2017 2017. pp 281-285

15. ASME (2009) *Dimensioning and Tolerancing; Engineering Drawing and related documentation practices*. American Society of Mechanical Engineers, New York, NY
16. Lei Y, Li N, Guo L, Li N, Yan T, Lin J (2018) *Machinery health prognostics: A systematic review from data acquisition to RUL prediction*. Mechanical Systems and Signal Processing 104:799-834
17. Elattar HM, Elminir HK, Riad A (2016) *Prognostics: a literature review*. Complex & Intelligent Systems 2 (2):125-154
18. Abellan-Nebot JV, Romero Subirón F (2010) *A review of machining monitoring systems based on artificial intelligence process models*. The International Journal of Advanced Manufacturing Technology 47 (1):237-257
19. Debongnie J-F (2006) *Usinage*. Editions du CEFAL,
20. Thomas M (2011) *Fiabilité, maintenance prédictive et vibration des machines*. Presses de l'Université du Québec, Montréal
21. Lamraoui M, Thomas M, El Badaoui M, Girardin F *Cyclostationarity analysis of instantaneous angular speeds for monitoring chatter in high speed milling*. In: IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society, Montreal, Qc, Canada, 25-28 Oct. 2012 2012. pp 3868-3873

Fault prognosis of planetary gearbox using acoustic emission and genetic algorithm: a case study

Félix Leaman[1], Cristián Molina Vicuña[2], Ralph Baltes[1] and Elisabeth Clausen[1]

[1]Institute for Advanced Mining Technologies, RWTH Aachen University
Wüllnerstr. 2, D-52062 Aachen, Germany
{fleaman, rbaltes, eclausen}@amt.rwth-aachen.de
[2]Laboratorio de Vibraciones Mecánicas, Departamento de Ingeniería Mecánica,
Universidad de Concepción, Edmundo Larenas s/n, Concepción, Chile
crimolin@udec.cl

Abstract
One of the most important aspects of machine fault prognosis is the selection of sensors and features to represent the degradation process of a faulty component. Several approaches in the literature have used features extracted from vibration signals to estimate the future condition based on time series forecasting. Another technology that has been used increasingly for this task are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. On gearboxes some studies have shown that the AE technology can be used effectively for fault diagnosis, but its use for fault prognosis is still a relative new field of research that offers encouraging opportunities. One downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature, which may lead to difficulties during the forecasting of an AE-based feature. Thus, in this study a novel feature based on a relative counting of the AE bursts is proposed and tested with data from a planetary gearbox with a ring gear fault. The proposed feature reduces the influence of the temperature on the generation of AE when it is compared to the counting based on a fixed amplitude threshold. Therefore, it can then be more suitable for fault prognosis than traditional AE counting. In this case study the forecasting of the proposed feature is carried out using an artificial neural network (ANN), whose hyperparameters were selected using a genetic algorithm. The results are promising and constitute a basis for further research.

1 Introduction

Machine diagnosis is defined as the examination of symptoms to detect and determine the nature of faults or failures, while machine prognosis is defined as the evaluation of the actual condition of a machine, forecast of the future condition and estimate the remaining useful lifetime (RUL) before a catastrophic failure based on the symptoms of faults [1, 2]. As it is well established for machine diagnosis, the type of fault that is going to be examined must be correctly represented by the monitoring data. A feature is a specific representation of the data in a lower dimension, which can be easier to relate to a specific fault. For instance, a misalignment can be examined using the vibration data and a feature related to the amplitude of the rotational speed harmonics can represent it in a better way than the RMS value. For prognosis purposes the same principle applies, but rather than only represent a fault the features should also be capable to represent the degradation process of the faulty component [3].

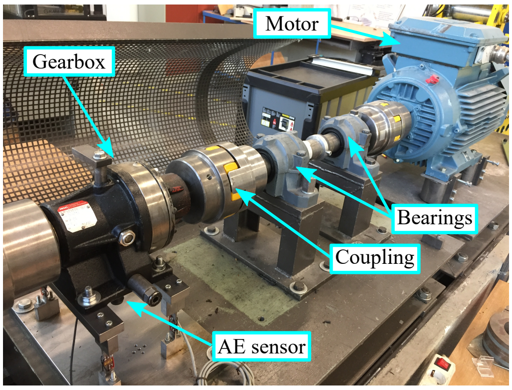
Prognosis models can be categorized into experience-based models, physical-based models or data-driven models. Experienced-based models use component failure history to predict the future condition. Physical-based models are based on physical principles of the failure process, which are modelled by mathematical equations. Data-driven models use historical monitoring data to generate a model capable of predicting the future condition of the machine. Within data-driven models artificial neural networks (ANN), hidden Markov and support vector machine models have been used for feature forecasting [2, 4, 5].

Condition monitoring data from vibration sensors have been extensively used for diagnosis purposes and to a lesser extent for prognosis purposes. Another type of sensor that has been used increasingly for these tasks are the acoustic emission (AE) sensors, which have frequency measurement ranges much higher than vibration sensors. AE is defined as elastic stress waves generated by rapid release of strain energy due to changes in the internal structure of a material [6]. Many sources of AE are damage-related such as: crack initiation and growth, crack opening and closure, dislocation movement and others [7]. One important feature to represent the AE data consists of the rate of generated AE bursts. This is adressed by many studies that have for example correlated the rate of AE bursts with crack propagation [8, 9].

In the case of gearboxes the AE technology has been effectively used for fault diagnosis [10, 11]. For the case of prognosis it is a relatively new field of research that offers encouraging opportunities [12]. It has been reported that during cyclic loading of a single tooth, the amplitude and energy of the AE bursts increase intermittently because of the accumulated damage and crack propagation [13]. Thus, it makes sense to evaluate the rate of AE bursts as a possible feature suitable for fault prognosis in gearboxes. However, one downside of the application of the AE technology in gearboxes is the strong dependence of the AE on the oil temperature [14], which may lead to difficulties during the forecasting of an AE-based feature. Therefore, in this study a novel feature based on a relative counting of the AE bursts that minimizes the effect of temperature is proposed and tested with data from a planetary gearbox with a ring gear fault. Then, a data-driven model is generated with an ANN and the progression of the feature is predicted.

2 Experiment

The test bench consisted of a planetary gearbox connected to an electric motor on its input side (high speed) and to a generator on its output side (low speed). The gearbox had the following number of teeth: $Z_s = 18$ for the sun gear, $Z_r = 72$ for the ring gear and $Z_p = 26$ for each of the planet gears. The number of planet gears was $N_p = 3$. The outside diameter of the ring gear was 144 mm. The speed reduction ratio was 1:5. A localized fault was produced in one tooth of the ring gear using a grinding tool. An AE sensor VS375-M was fixed under the gearbox case using a magnetic holder. The surface under the sensor was previously polished and copper paste was applied between the sensor and the surface to improve the acoustic coupling. The test bench and the fault can be seen in Figure 1.



(a)

(b)

Figure 1: Planetary gearbox test bench (a) and ring gear fault (b)

AE signals were measured using an input rotational speed of 1300 r/min and a load of 112 Nm in the output shaft. Thus, the sun gear rotating frequency was $f_s = 21.67$ Hz and the carrier rotating frequency was $f_c = 4.33$ Hz. For all measurements a sampling frequency of 1 MHz and acquisition time of 5 s for the AE signals were used. The AE signals were measured continuously in eight measurement campaigns (MCs), of which seven consisted of approximately 75 min of continuous operation and one of approximately 120 min. During this eight MCs neither planetary gearbox nor sensor were disassembled. Aditionally in each MC the same instrumentation for the measurement of the AE signals was employed. Summing the MCs, the planetary gearbox cumulated approximately 10.5 h of intermitent operation. After the 8^{th} MC the planetary gearbox was disassembled for inspection, which revealed no visual growth of the fault. Despite this observation, damage progression at a microscopic or subsurface level could not be discarded.

The external temperature of the planetary gearbox case was used to represent the temperature of the oil. It was measured every two minutes for each MC. The result is shown in Figure 2, where the end of each MC is indicated with a dashed line. As expected the temperature increases during each MC. The rate of temperature increase is higher at the beginning of each MC and is then progressively reduced.

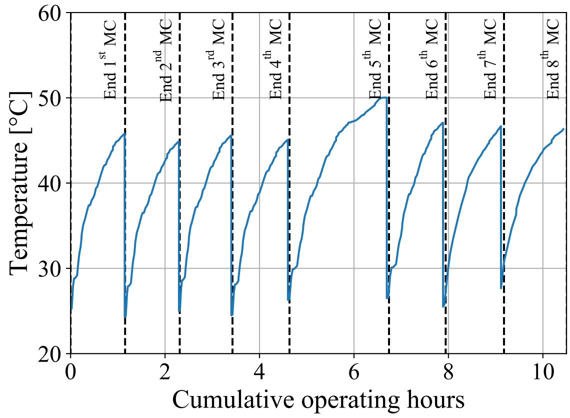


Figure 2: Planetary gearbox temperature during the tests.

As discussed in the previous section, the temperature affects the generation of AE in gearboxes. Therefore, the detection of AE bursts is expected to be affected by this phenomena. In the next section this topic is adressed.

3 Counting of AE Bursts

3.1 Fixed counting of AE bursts

The traditional approach to detect AE bursts consists of the definition of a fixed amplitude threshold T_a. When the amplitude of the signal exceeds this value an AE burst is detected (see Figure 3). The end of the AE burst is determined when after a hit definition time (HDT) there are no more threshold crossings. The beginning of the next AE burst can only be detected when a hit lockout time (HLT) has elapsed. The value for the threshold T_a is established empirically after inspection of the AE signals.

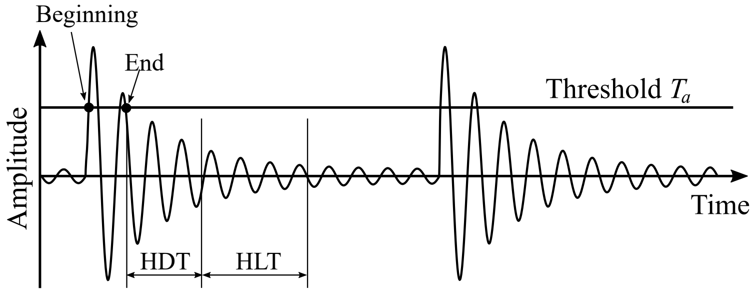


Figure 3: AE burst detection with fixed amplitude threshold

For this case study a threshold $T_a = 0.8$ mV was selected together with a HDT and a HLT of 300 μs each. The number of detected AE bursts is counted every 5 seconds in each signal, so the obtained result expresses the rate of detected AE bursts every 5 seconds. In order to minimize small fluctuations and concentrate the analysis on the global trend, a moving average filter of width 60 s was applied to the rate of AE bursts. Figure 4 shows the result of this analysis with fixed counting approach. It is observed that the rate of AE bursts rises at the beginning of each MC and then decreases as the MC continues. Although the same behavior occurs in each MC, the maximal rates of AE bursts have different values for different MCs. Moreover, the maximal values reached do not take place necessarily as the accummulated operating hours increase. However, when only the last parts of each MC are analyzed, it is observed that the rate of AE bursts increases as the hours of operation cumulate.

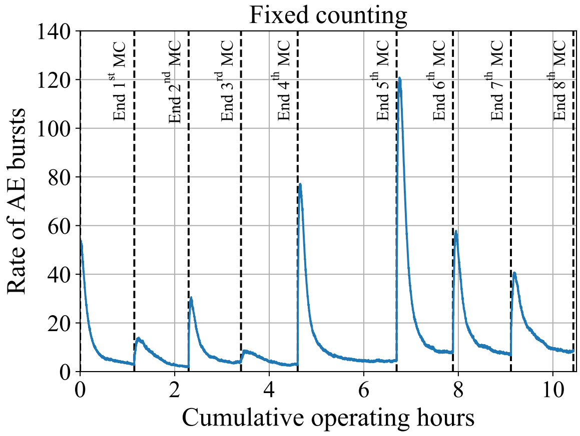


Figure 4: Rate of AE bursts during tests calculated with a fixed counting

The aforementioned observations make the fixed counting approach inappropriate for prognosis purposes in this study case, since the obtained rate of AE bursts has very low monotonicity. Therefore, it is required to explore other approaches as follows in the next subsection.

3.2 Relative counting of AE bursts

In this subsection an approach to make a relative counting of the AE bursts is proposed. The relative counting consists of taking into account only the AE bursts that should take place when the fault of the ring gear interacts with each of the planet gears. Knowing the theoretical temporal positions of the bursts originated by the interaction between fault and planet gears within a signal requires the use of an encoder that provides angular position. However, this requirement can be avoided by considering only the spacing between the detected bursts instead of their absolute position. Thus, the relative counting will be maximum if in the signal there are detected bursts equally spaced at the inverse of the ring gear fault frequency f_{fr}. This corresponds to the frequency at which the planet gears interact with the fault and corresponds to the rotational frequency of the planet carrier f_c times the number of planets N_p as expressed in equation 1:

$$f_{fr} = f_c \cdot N_p \tag{1}$$

Mathematically the relative counting can be achieved by taking the maximal cross-correlation between a binary signal $\mathring{x}(t)$ that represents where the bursts are actually detected, and another binary signal $\mathring{z}(t)$ that represents where the AE bursts should take place, if they are always originated by the interaction between the damaged ring gear tooth and the planet gears.

Figure 5 shows schematically the path of each planet gear across the fault and the originated AE bursts. Some of them overpass the threshold T_a and thus are detected. Accordingly, the binary signal $\mathring{x}(t)$ is constructed with its high value at the position of the detected bursts and its low value everywhere else. On the other hand, the binary signal $\mathring{z}(t)$ is constructed with its high values equally spaced at the inverse of the ring gear fault frequency f_{fr}. Thanks to the cross-correlation it is not necessary that the high values of the binary signal $\mathring{z}(t)$ are in phase with the path of the planet gears across the fault. By taking the maximal cross-correlation the phase difference can be corrected. Therefore, only the spacing between the high values of the binary signal $\mathring{z}(t)$ are considered in its construction.

The proposed relative counting gives values between zero and one. A value one means that all expected bursts have a maximal amplitude higher than the threshold. Analogously, a value zero means that none of the expected bursts have an amplitude higher than the threshold. This is schematically illustrated in Figure 6. Here, the binary signal $\mathring{x}_1(t)$ has a maximal cross-correlation of 1 with the binary signal $\mathring{z}(t)$, since they can match

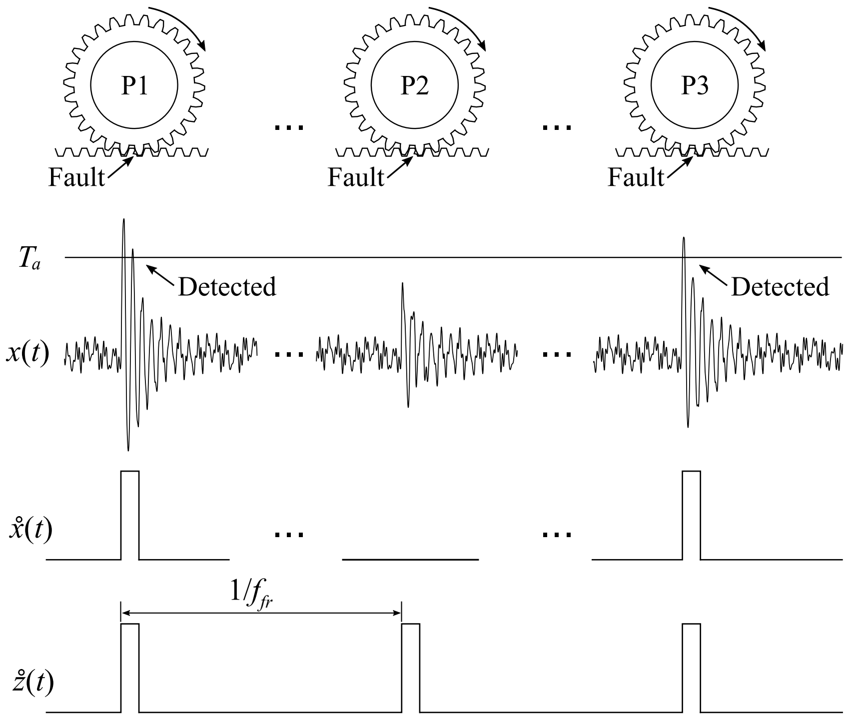


Figure 5: Binary signals representing actual $\mathring{x}(t)$ and hypothetic $\mathring{z}(t)$ time location of AE bursts

in all their high values. The binary signal $\mathring{x}_2(t)$ has a maximal cross-correlation of 0.5 with $\mathring{z}(t)$, since they can match in two out of four high values. The binary signal $\mathring{x}_3(t)$ has a maximal cross-correlation of 0.25 with $\mathring{z}(t)$, since they only can match in one out of four high values. Finally, the binary signal $\mathring{x}_4(t)$ has a maximal cross-correlation of 0 with $\mathring{z}(t)$ since it does not have any high values. Notice that this figure points out the aforementioned observation that $\mathring{z}(t)$ does not have to be in phase with the signals $\mathring{x}(t)$.

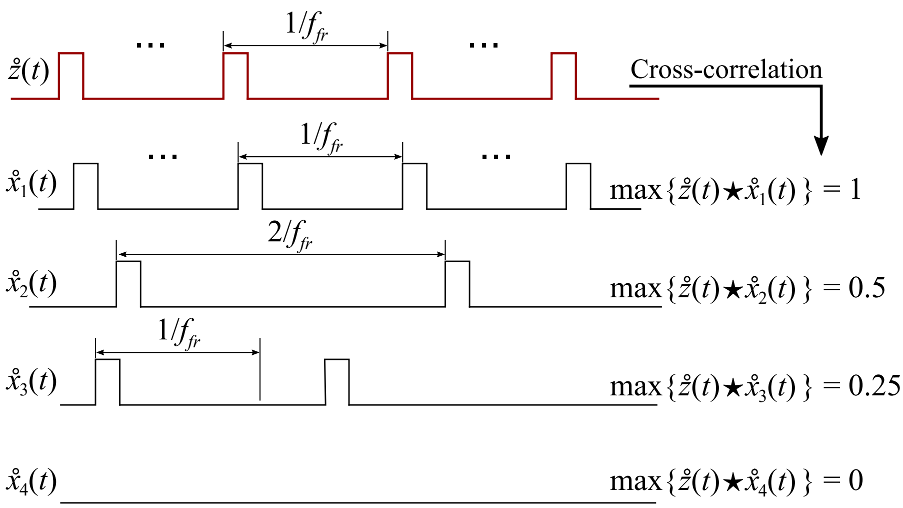


Figure 6: Relative counting of AE bursts based on cross-correlation with the signal $\mathring{z}(t)$

For the analysis of the case study the same values for T_a, HDT and HLT used for the fixed counting are now used for the relative counting. In the same way, a moving average filter of width 60 s is also employed to eliminate too small fluctuations in the rate of AE bursts. Figure 7 shows the result of this analysis with the relative counting approach. It is observed that the rate of AE bursts (now represented by the maximal cross-correlation) is much more monotonous than for the fixed counting approach. The large fluctuations due to the beginning of each MC are also considerably reduced.

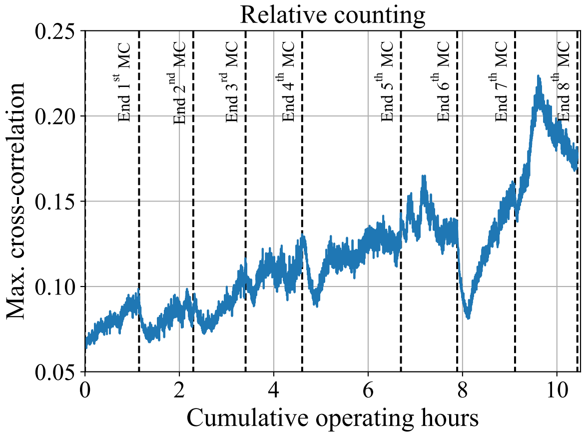


Figure 7: Maximal cross-correlation during tests calculated with a relative counting of AE bursts

In account of this observations the rate of AE bursts calculated with the relative counting approach is much more suitable to be uses as feature for failure prognosis. The forecasting of this feature is addressed in the following section.

4 Prognosis

4.1 Time series forecasting with an ANN

There are multiple possible architectures to build an ANN suited for forecasting a time series. One of the most straightforward architectures consists of using the observations x_0 to x_n from a time series $\{x\}$ as input features to predict the observation x_{n+1}. In the following prediction step (PS) the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} and so on. Another possible architecture consists of using the observations x_0 to x_n to predict the observations x_{n+1} to x_{n+m}. In the following PS the ANN uses the observations x_1 to x_{n+1} to predict the observation x_{n+2} to x_{n+m+1} and so on. Thus, although m observations are predicted in each PS, only the first of them is used as input feature to make the prediction in the next PS. One variation of such an architecture consists of using only the last predicted observation rather than the first one to make new predictions. By using these types of architectures a higher stability in the forecasting is expected. The training and test of such ANN are illustrated schematically in Figure 8. It shows how the training examples (TE) are constructed from a time series considering values $n = 4$ and $m = 3$. Each of the TE is used to iteratively adjust the weights of the ANN using the backpropagation algorithm. After the ANN has already been trained with all the TE, it is now capable to forecast the time series in successive PS and its performance can be evaluated with the test data.

During the building of the ANN the dataset is divided into two groups: training and test. Within the training group a datasubset is left aside to validate the results of the training process. One key objective of the training process is to find the most suited hyperparameters for the ANN. They include: number of nodes in input and output layers (n and m, respectively), number of nodes per hidden layer, regularization term α, number of mini-batches and number of epochs. More information regarding the hyperparameters can be found in [15]. Initially the most suited values for the hyperparameters are selected using grid search. This consists of selecting possible values for each hyperparameter and calculating which combination of hyperparameters achieves the best performance overall. For each combination of hyperparameters an ANN is trained without including the validation data. Then, the forecasting performance of the ANN is calculated in the data used for its training and then in the validation data. This division is carried out with the aim of minimizing the bias and variance errors of the ANN model as much as possible. A performance calculated only in the data used for training can cause higher variance since the model could over fit the training data. On the other hand, a performance calculated only in the validation data can cause higher bias since the model could under fit the training data. In this study a global performance is calculated as the average between the training and validation performances.

The performance of the model is measured with the use of the mean squared error (MSE) between prediction

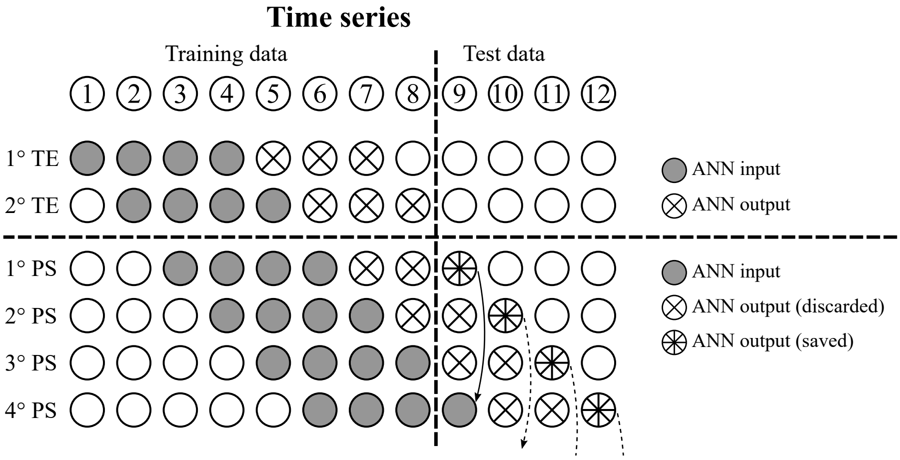


Figure 8: Training and prediction scheme of the ANN for time series forecasting

and real data. The MSE is calculated as follows in equation 2:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{2}$$

, where x_i represents the i-th observation of the time series $\{x\}$ and $\hat{x}_i$ represents the i-th observation of the predicted time series $\{\hat{x}\}$.

Due to computational restrictions the grid search can only consider a limited number of values for the hyperparameters. One way to increase the efficiency of the search is to set an initial group of values for the hyperparameters and determine the best combinations. After that, the search continues but considering only variations of the already determined best combinations. This can be achieved by using a genetic algorithm, which is explained in the following section.

4.2 Genetic algorithm

The genetic algorithms are computational procedures to find a sufficiently good solution to an optimization problem. They are inspired by the mechanism of natural selection and emulate biological processes such as mutation, crossover and selection. In this study a genetic algorithm is employed to guide consecutive grid searches and find the best possible combination of hyperparameters for the ANN model. First, an initial population of hyperparameters is evaluated using a fitness function. The best combination of hyperparameters are called *parents* and are used to generate *children* through crossover and mutation techniques. The crossover combines information from parents to generate the children, while the mutation consists of random alteration of the information traspased to the children. In this study, the information corresponds to the hyperparameters for the ANN model. The parents together with the children constitute a new *generation*. This new generation is once again evaluated with the fitness function to create new parents and new children in an iterative process. The procedure is repeated until a stoppage criterion is met. When this is accomplished, the best solution in the last generation is selected as the best solution of the whole iterative process. Two possible stoppage criteria are: a fixed number of generations to iterate or a fixed number of consecutive generations for which the best fitness score does not improve more than a pre-defined percentage.

Figure 9 shows schematically how the genetic algorithm works. Let us consider four hyperparameters to optimize: α, β, γ and δ. The initial population contains five possible combinations of hyperparameters, which are evaluated with the fitness function. The best two combinations are selected as parents and are used to generate three children. The employed crossover corresponds to a single-point crossover, which randomly selects at which single point (star marker) the information from the parents is divided. Then, in the mutation some information of the children is randomly altered. One important aspect is that the percentage of mutation must not be so high as to completely vanish the result of the crossover.

In this study single-point crossover, maximum 10% of mutation and a total of 5 generations are considered.

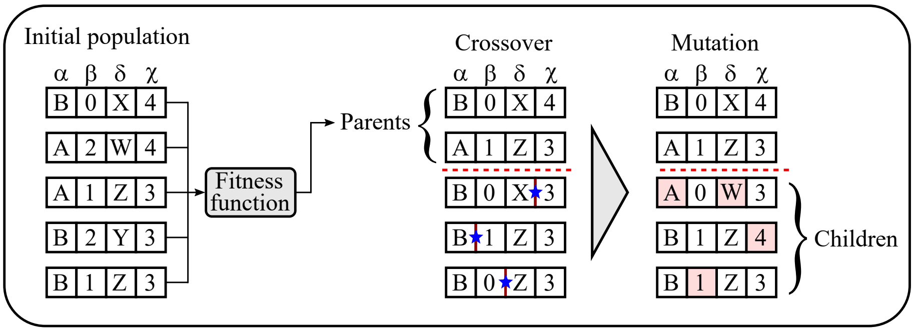


Figure 9: Schema of the genetic algorithm for hyperparameters optimization

The employed fitness function corresponds to the calculation of a fitness score (FS) that is the inverse of the MSE as it is shown in equation 3. Thus, higher values of the FS correspond to best performance of the model.

$$FS = \frac{1}{MSE} \tag{3}$$

4.3 Results

Table 1 shows the hyperparameters considered for the initial population. The values for the number of nodes in the input layer n and in the output layer m expressed as percentages refer to a fraction of the data used to train the ANN models. Two configurations for the hidden layers are initially tested: one hidden layer with a number of nodes equal to the mean value between n and m, and two hidden layers with number of nodes equal to $0.75(n+m)$ for the first layer and $0.25(n+m)$ for the second layer.

Hyperparameter	Values
n	50%; 32%; 16%
m	13%; 6%; One
Nodes per layer	$0.5(n+m)$; $0.75(n+m)$ - $0.25(n+m)$
α	1.e-1; 1.e-3
Mini-batches	32; 128
Epochs	Until 100

Table 1: Initial hyperparameters considered to generate ANN models

The total of 7512 observations of the relative counting of AE bursts was divided into 80% for training and 20% for test. For the process of hyperparameters estimation 25% of the training data was left aside for validation. Table 2 shows the fitness scores for the initial population and 5th generation during the search for the most suited hyperparameters. The $FS_{training}$ is highly improved from the initial population until the 5th generation, while the $FS_{validation}$ only decreases marginally. As a consequence, the overall $FS_{average}$ is highly increased. Table 3 shows the corresponding best values for the hyperparameters for the initial population and 5th generation that achieve the aforementioned fitness scores. It is observed that the number of nodes in each layer tends to increase. So does the regularization term α, while the number of epochs decreases. The number of mini-batches remains approximately constant.

Best score	For initial population	For 5th generation
$FS_{training}$	6.12e4	1.83e5
$FS_{validation}$	2.25e3	2.00e3
$FS_{average}$	3.17e4	9.25e4

Table 2: Fitness scores of the best models from the initial population and 5th generation in validation phase

Figure 10 shows the result of the forecasting for the maximal cross-correlation with the best hyperparameters of the initial population in (a) and for the 5th generation in (b). It is observed that for the initial population

Hyperparameter	Best values for initial population	Best values for 5th generation
n	50%	63.481%
m	13%	15.053%
Nodes per layer	1836 - 1002	1858 - 1014
α	1.0e-1	1.3085e-1
Mini-batches	32	33
Epochs	64	21

Table 3: Best combination of hyperparameters for the initial population and 5th generation

the prediction (test) deviates considerably from the real data. However, for the 5th generation the prediction is very close to the real data, which confirms the improvement produced by the genetic algorithm. This is also confirmed by the calculation of the fitness scores in Table 4. The FS calculated on the training data indicates how good the model fits the training data. Thus, it can be only considered as an estimation of the performance of the ANN model. The FS calculated on the test data indicates how good the model can predict unseen data. Both scores should be as high as possible and have a as low as possible difference between them. It is also expected that $FS_{training}$ is higher than FS_{test}, since the model should perform better on data that it has already seen. For the initial population both fitness scores $FS_{training}$ and FS_{test} have low values. Moreover, FS_{test} is higher than $FS_{training}$. For the 5th generation this deviation is corrected and both fitness scores reach much higher values.

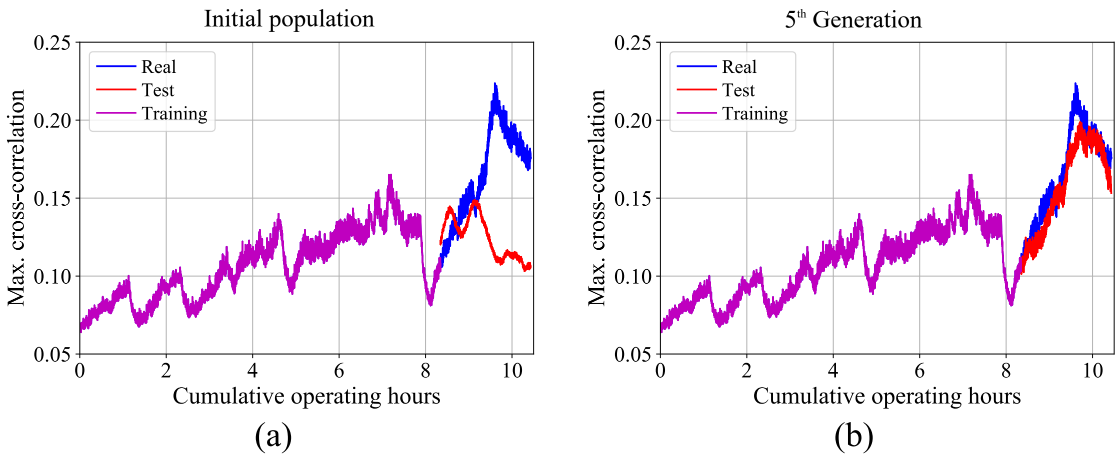


Figure 10: Forecasting of the maximal cross-correlation for the initial population (a) and 5th generation (b)

Best score	For initial population	For 5th generation
$FS_{training}$	9.39e1	3.64e4
FS_{test}	3.06e2	6.37e3

Table 4: Model fitness scores from the initial population and 5th generation in the testing phase

5 Conclusions

A novel approach for counting of AE bursts in gearboxes was presented and compared with the traditional approach with measurements from a planetary gearbox. In conditions where the gearbox accummulated several hours of intermittent operation with a localized fault in the ring gear, the proposed approach was able to provide a rate of AE bursts represented by a cross-correlation value suitable for prognosis purposes. The forecasting was carried out with an ANN, whose performance was enhanced through a genetic algorithm for the selection of its hyperparameters. With basis on the obtained results the authors conclude that:

- In presence of a localized fault in the ring gear of a planetary gearbox, the rate of detected AE bursts increases as the hours of operation cumulate.

- For intermittent operation the approach to detect AE bursts using a fixed amplitude threshold provides a rate of AE bursts highly affected by the temperature, which is not suitable for prognosis purposes.
- In comparison the proposed approach for a relative counting of the AE bursts provides a rate of AE bursts with higher monotonicity, and thus it is more suitable for prognosis.
- The rate of AE bursts, in this case represented by a cross-correlation value, can be effectively predicted with the use of an ANN for time series forecasting if its hyperparameters are correctly selected.
- The proposed way for selection of hyperparameters was an initial grid search followed by a genetic algorithm for hyperparameters optimization, which was capable to provide good results.

References

[1] W. Caesarendra, G. Niu, B. Yang *Expert Systems with Applications Machine condition prognosis based on sequential Monte Carlo method.*, Expert Systems with Applications, 37, 2010, pp. 2412-2420

[2] Y. Peng, M. Dong *A prognosis method using age-dependent hidden semi-Markov model for equipment health prediction*, Mechanical Systems and Signal Processing, 25, 2011, pp. 237-252

[3] X. Zhang, J. Kang, J.S. Zhao *Features for Fault Diagnosis and Prognosis of Gearbox*, Chemical Engineering Transactions, 33, 2013, pp. 1027-1032

[4] W.Q. Wang, M.F. Golnaraghi, F. Ismail, *Prognosis of machine health condition using neuro-fuzzy systems*, Mechanical Systems and Signal Processing, 18, 2004, pp. 813-831.

[5] J. Kang, X. Zhang, J. Zhao, et al., *Gearbox fault prognosis based on CHMM and SVM*, In: International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, 2012.

[6] R.K. Miller, P. McIntire *Non-destructive testing handbook: Volume 5 acoustic emission testing*, Publisher: American Society for Non-destructive Testing, 1987.

[7] M. Huang, L. Jiang, P. Liaw, et al. *Using acoustic emission in fatigue and fracture materials research.*, JOM-e: Research Summary, 50(11), 1998, pp. 1-12.

[8] C. Barile, C. Casavola, G. Pappalettera *Fatigue Damage Monitoring by Means of Acoustic Emission and Thermography in Ti Grade 5 Specimens*, Procedia Engineering, 114, 2015, pp. 487-492.

[9] P. Mazal, F. Vlasic, V. Koula *Use of Acoustic Emission Method for Identification of Fatigue Micro-cracks Creation*, Procedia Engineering, 133, 2015, pp. 379-388.

[10] A.B. Novoa, C.M. Vicuna *New aspects concerning the generation of acoustic emissions in spur gears, the influence of operating conditions and gear defects in planetary gearboxes*, Insight, 58(1), 2016, pp. 18-27.

[11] D.S. Gu, J.G. Kim, Y.S. An, et al. *Detection of faults in gearboxes using acoustic emission signal*, Journal of Mechanical Science and Technology, 25(5), 2011, pp. 1279-1286.

[12] D. Mba *Prognostic Opportunities Offered by Acoustic Emission for Monitoring Bearings and Gearboxes.*, In: 12th International Congress on Sound and Vibration, 2005.

[13] A. Singh, D.R. Houser, S. Vijayakar *Detecting Gear Tooth Breakage Using Acoustic Emission: a Feasibility and Sensor Placement Study*, Journal of Mechanical Design, 121, 1999, pp. 587-593.

[14] C.M. Vicuna *Effects of operating conditions on the Acoustic Emissions (AE) from planetary gearboxes*, Applied Acoustics 77, 2014, pp. 150-158.

[15] I. Loshchilov, F. Hutter *Online Batch Selection for Faster Training of Neural Networks*, In: International Conference on Learning Representations (ICLR), 2016, Workshop Track.

Rotating Machine Diagnosis using Acoustic Imaging and Artificial Intelligence
SURVISHNO Conference 2019

Abdelhakim Darraz[1,2], Jérôme Antoni[3], Pierre Mollon[2], Gabriel Kirié[2], Jean-Hugh Thomas[1], Anthony Larcher[4]

[1] Université du Mans/ENSIM, LAUM, UMR CNRS 6613, F-72085 Le Mans, France
[2] VALEO Equipements Electriques Moteur, F-38070 Saint-Quentin-Fallavier, France
[3] Univ Lyon, INSA Lyon, Laboratoire Vibration Acoustique, F-69621 Villeurbanne, France
[4] Université du Mans/ENSIM, LIUM, EA 4023, F-72085 Le Mans, France

Corresponding author: abdelhakim.darraz@valeo.com

Abstract

Mass production of quality equipment in the automotive industry requires controls throughout the production line. These controls are done through monitoring and validation tools for both production and finished products. The use of signal processing methods, applied to acoustic and vibratory recordings collected during the operating cycle, aims to ensure that they are in good working order, to maintain them and to guarantee the quality of the service provided by a manufacturer to its customers. However, sometimes the techniques used do not reach the expected performance, which of course depends on the defect to be recognized but also on the conditions under which the measurements were made. This paper introduces a preliminary study on machine diagnosis by combining both signal processing methods and artificial intelligence. This work is dedicated to develop a system which allows us to measure several high frequency channels and to transmit them to a computer via USB interface.

1 Introduction:

In a production environment many parts are unfairly detected as defective when monitoring is based on indicators from the literature. The causes of these errors are often related to the not conducive noisy environment to such a diagnosis by records sensitive to disturbance. Moreover, from one production site to another, it is not possible to apply the same default detection thresholds because of a different environment involving a variation of the structures and of the product frequency responses. Therefore, today it remains difficult to do a relevant diagnosis in a noisy environment and particularly on non-stationary signals. The aim of the study is to improve this diagnosis by first using a microphone antenna and then operating an artificial intelligence process on a database acquired on production benches. The microphone array leads to the provision of a spatial map of the acoustic field generated by the monitored system. An acoustic imaging approach allows the addition of a new spatial dimension in the data representation. The preliminary study presented consists in differentiating several states of the system to be monitored from the simultaneous exploitation of information expressed in the time-frequency-space domain.

1.1 Rotating Machine

The system considered for the study is a starter. A starter is an electric motor used to rotate a thermal engine until the self-combustion in the engine takes over. It is composed by several subassemblies such as a planetary gear reducer, a drive shaft, an armature permitting under the effect of a magnetic field the rotation of the motor and a current transmission subassembly allowing the current to flow to the armature by brushes.

Each subset radiates its own acoustic signal which, by interactions, contributes to the overall signal emitted by this motor. These interactions associated to the resonance phenomena in the transient operating phases of the machine make the diagnosis more difficult. Indeed, this diagnosis allows to focus the acoustic measurement on the rotating machine by a beamforming process while freeing of the disturbing sources coming from other directions. First results of comparisons will be presented.

There are three states of the electric motor to recognize:

- Healthy state
- Armature imbalance
- Gear defect

Figure 1: Electric motor

1.2 Classification

Automatic classification is a branch of artificial intelligence. Artificial intelligence, commonly known as AI, is defined as the set of theories and techniques used to make machines capable of simulating human intelligence [1]. This field was born in the middle of the 20th century with the development of computer science and the ambition to create machines with the ability to think in a similar way in their functioning to the human mind in terms of perception, understanding and even in taking decision.

Automatic classification is used in all pattern recognition systems. A pattern in the broad sense is an object of very varied nature. It can be a bar code, a face, a fingerprint and more generally a digital data suite that will constitute the signature of the belonging of an object to a family.

Indeed, these systems allow the algorithmic characterization of objects and consist of assigning an object to a class or category based on prior learning.

Learning is the process of constructing a general model based on particular observations of the real world in order to predict a behaviour or a decision in front of new unseen data. The second definition of learning, taking more the sense of training, consists of improving the performance of the model in a progressive way by being confronting to the exercise of an activity.

This idea of improving the accuracy of models through training is easy to perceive for humans since the ability of a person to perform a certain task is often judged by his experience in the field.

2 Experimentation

2.1 Simple sensor analysis

2.1.1 Machine learning

Supervised Machine Learning requires expert know-how in the intended field application. Indeed, it is a question of being able to label the samples correctly on the one hand but also to define relevant indicators to characterize samples regarding the classification to be carried out in the sense that these indicators must be representative for class distinction.

Here the chosen features are indicators coming from signal processing such as:

- RMS value
- Peak to peak
- Kurtosis
- Partial levels around kinematic frequencies
- Global level
- Energy in one-third octave band

In order not to bias the classification for the algorithms used in this part, a step of standardization of these different parameters is necessary. This consists in subtracting a value of a parameter by the average of the values of this parameter and then by dividing the obtained value by the standard deviation of the parameter. For most classification algorithms it is also important to use the same number of samples per class, i.e. to have a balanced dataset. Indeed, Bayesian algorithms introduce the probability of belonging to a class in the calculation of the conditional probability that a sample belongs to a class knowing its characteristics.

In this part we are primarily interested in a diagnosis of the electric motor based on a single microphone recording. The setup consists on the analysis of a signal coming from a single omnidirectional microphone located at 50 centimeters from the electric motor. The dataset is composed by 25 starters per class.

	Bayesian	Neural Network	Decision Tree	Support Vector Machine	Random Forest
True Healthy	24	21	21	21	25
False Healthy	2	5	2	7	1
True Imbalance	25	23	25	21	25
False Imbalance	0	1	1	2	0
True Gear	23	22	22	22	24
False Gear	1	3	4	2	0
# Error	3	9	7	11	1
Performance	96%	88%	90.67%	85.33%	**98.67%**

Table 1: Classification results

The classification mode used for the results obtained above is the so-called cross-validation mode. Cross-validation consists in dividing the population into N groups of constant size. Then, N-1 groups are randomly selected and used to learn and create a model while the Nth group is used as a test population. This step is repeated N times so that all samples are used for learning and testing in order to create N models.

The global performance is an average of the N classification results created. Cross-validation thus makes it possible to obtain a faithful and global performance of the tested algorithm but has the disadvantage of non-negligible cost in terms of computing time.

Table 1 indicates a very good fault recognition performance for all algorithms, particularly for Bayesian and Random Forest, which respectively achieved 96% and 98.67% of recognition. These results confirm the interest of using such methods for machine diagnosis. However, these measurements were carried out in a healthy environment without external disturbance.

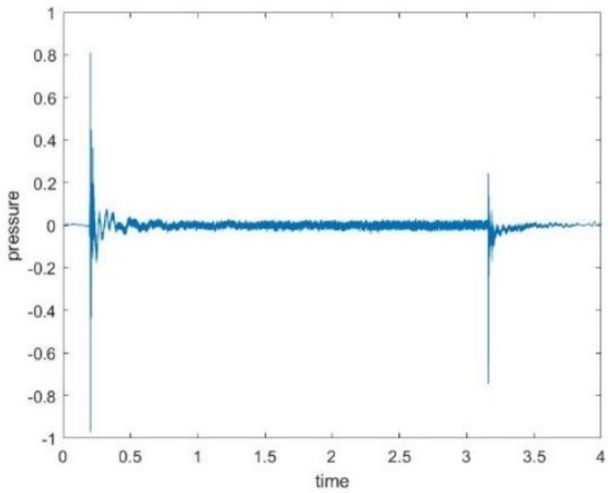


Figure 2: Acoustic pressure signal of the electric motor in unload condition

In the first case, the Mel Frequency Cepstrum Coefficients (MFCC) are extracted from the time signal. This extraction of the coefficients is developed using the Fast Fourier Transform (FFT) and the Discrete Cosine Transform (DCT) on the Mel scale which is a perceptual parametric model. These are the most used criteria in Automatic Speech Recognition (ASR) systems.

First, the signal is split into N windows of a few milliseconds with overlapping (usually in ASR the length of the window is 25 ms and the overlap is 10 ms). A Hamming window is then applied to the signal to limit the spectral distortion (appearance of parasitic high frequencies) related to the overlap and the disturbance at the beginning and the end of the window before passing to the frequency domain via the FFT.

A conversion of the frequency scale f to the Mel scale is performed according to

$$mel(f) = 2595 * \log(1 + \frac{f}{700}). \quad (1)$$

A triangular response filter bank with variable frequency bandwidth is applied to simulate the response of the human ear in the best possible way. This band variation represents the capacity of the human being to be able to easily distinguish two near frequencies at low frequencies than at high frequencies. For each triangular filter, the sum of the energies is calculated so we get as many coefficients as filters.

Finally, we convert the logarithmic spectrum of Mel obtained to the time domain with the DCT, and then we usually keep the first 12 coefficients for each of the windows.

The coefficients are calculated according to

$$C_k = \sum_{i=1}^{N} \log(E_i) * \cos[\frac{\pi k}{N}(i - 0.5)] \quad (2)$$

- N: number of filters.
- E_i: energy calculated with the ith filter.
- C_k: kth Mel Frequency Ceptsrum Coefficient.

Short windows are used for the transient operation phases of the motor while longer windows are used on the stationary phase (this phase is not used in its entire duration; an average and a standard deviation are calculated to take into account any variation of the remaining signal). In addition, to add more information and thus improve the signal recognition, the differential coefficients Δ and acceleration ΔΔ are implemented. These coefficients, calculated directly from the MFCC coefficients taking respectively the derivative of the first and the second order, make it possible to consider the dynamics of the signal.

According to [2], the addition of these differential coefficients and accelerations increases the recognition by about 20%. At least 702 features were extracted from each time signal. Here the dataset is reduced to 15 starters per class.

	Bayesian	Decision Tree	Support Vector Machine	Random Forest
True Healthy	13	11	11	14
False Healthy	7	7	4	3
True Imbalance	14	14	15	14
False Imbalance	0	2	2	0
True Gear	9	9	10	13
False Gear	2	2	3	1
# Error	9	11	9	4
Performance	80%	75.56%	80%	**91.11%**

Table 2: Classification results using MFCCs

Table 2 shows the performance achieved by algorithms. Random Forest remains more efficient than the other algorithms with a recognition performance of 91.11%. However, the overall performance achieved with this approach is less than the performance obtained previously in Table 1.

2.1.2 Deep Learning

The aim here is not to extract previously mentioned indicators, but rather to provide complete acoustic data to the algorithms. Spectrogram as input was studied.

The classification consists in identifying the class of each starter by recognizing the spectrogram of the acquired signal. For this study, an image processing approach based on a pre-learnt Convolutional Neural Network (CNN) AlexNet is used. AlexNet was developed at Toronto university [3] for the ImageNet LSVRC-2010 contest, a competition for which it is proposed to classify 1.2 million images into 1000 different classes. The authors achieved a winning top-5 test error rate of 15.3% for the ILSVRC-2012 contest.

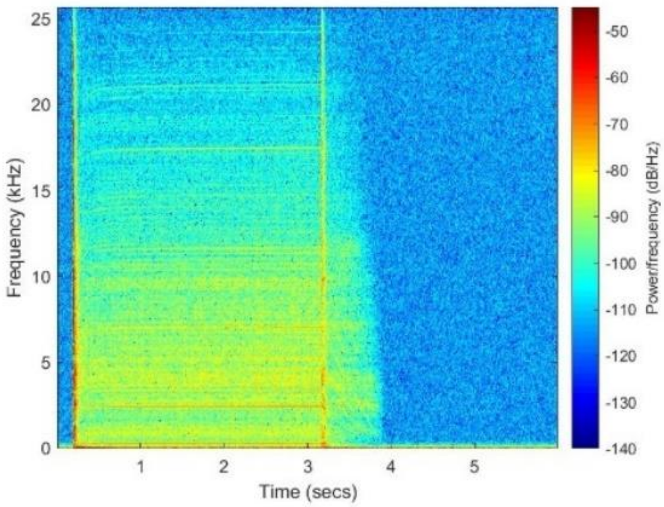


Figure 3: Acoustic signal spectrogram of the electric motor in unload condition

Transfer learning consists in using a trained Neural Network to extract features, which achieved very good performance in recognition on a specific field with thousands of samples for the learning phase, and to transfer the knowledge learnt in that field to another field. This is what we have done here, AlexNet has been trained on 1 million images and performs very well on object recognition in images.

This allows us to extract relevant features for image recognition even if we only have a low amount of data. A modification of the final layers and a training phase are necessary in order to adapt the classification to our field by changing the number of outputs and the weights of each neuron.

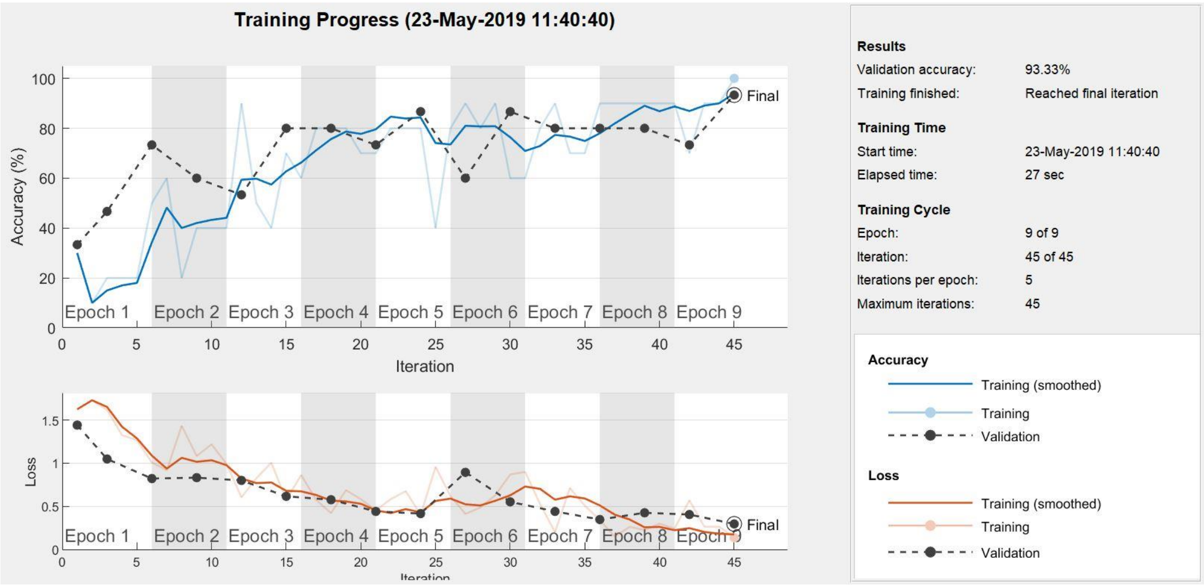


Figure 4: Acoustic pressure by time of the electric motor in unload condition

We achieved a good validation accuracy of 93.33% with this method compared to results obtained by MFCCs. However, it is possible that the model overfits the data because of a large number of parameters (60 million features) compared to the number of samples. In statistics, overfitting corresponds to a model or an analysis which fits perfectly with a dataset. In machine learning, overfitting is one phenomenon to avoid because these models perform very well on the data used to create the model but cannot generalize on new data which means low performances.

2.2 Acoustic localization

Unfortunately, it is not possible to obtain such a recognition rate in a production environment. Indeed, the random ambient noise generates ambiguity perturbing the recognition of defects. This is why we have chosen to introduce acoustical imaging to our work in order to locate the sound sources in a space, focus the measurement towards a privileged direction and thus filter the signal emitted by the electric motor from the ambient noise.

An acoustic antenna represents a distribution of microphones according to a specific geometry and number of sensors. The resulting geometry formed by the repartition of these microphones as well as their number condition the performance of the system, which is characterized by the frequency range of use, the resolution and the capacity of localization. For this project, we made the choice of digital MEMS (Micro electro-Mechanical Systems).

This technology introduced for the first time in 1967 [4] corresponds to miniaturized sensors or actuators which couple several physical principles including mechanics and electronics. The miniaturization allowed by this technology responds to a growing need related to congestion, sensitivity, mass production and to the complexification of systems by allowing the addition of new functions in a non-intrusive way.

The method used to represent the acoustic field is a Beamforming-based method.

2.2.1 Acquisition System

The acquisition system is composed of MEMS microphones. The working of these systems remains faithful to those of conventional microphones because in a general way the physical laws governing the different domains are unchanged. A MEMS consists of a fixed substrate (a semiconductor generally made of silicon) and a moving part.

In the case of the MEMS microphone, the moving part is represented by a membrane. A conductor measures the impedance variation induced by the deformation of the membrane and this allows to obtain a minimum size while not skimping on the performance of the sensor.

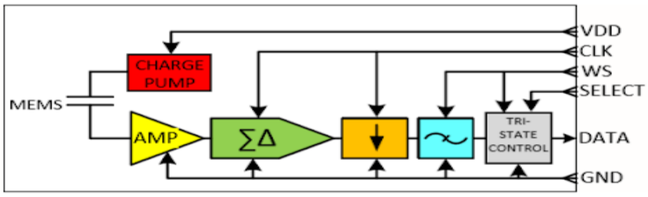


Figure 5: Block diagram of the I2S digital mic

- The charge pump (in red) allows the supply voltage to be raised to a level necessary to polarize the transducer.

- The MEMS is responsible for converting the measured pressure into a voltage.

- The amplifier (in yellow) stores the voltage from the MEMS and amplifies the signal.

- The Sigma Delta converter (in green) converts the analog signal in memory into a pulse-density modulation (PDM) signal on 1-bit resolution.

- The decimator (in orange) converts the PDM signal into a 24-bit pulse coded modulation (PCM) signal by down-sampling with a factor of 64.

- A low pass filter (in blue) removes the remaining high frequency components.

- The three-state mode (in gray) makes it possible to associate two MEMS on the same I2S line for a stereo recording.

Pre-assembled I2S digital MEMS microphones with welds on PCB plate already made were chosen for convenience because the size of the microphone involves a high precision in the welding process.

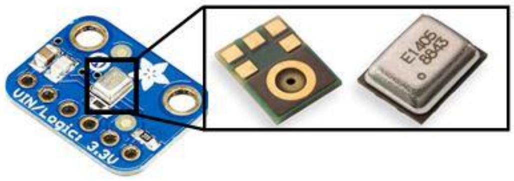


Figure 6: I2S MEMS microphone

An I2S output format MEMS implies that the microphone integrates a large part of the acquisition chain and a MiniDSP acquisition board is used to measure up to 8 synchronous I2S channels by board.

2.2.2 Beamforming

Beamforming [5] is a signal processing tool used in the field of antennas for the directional signal transmission or reception. The main principle of this method is to combine the elements of a sensor array in such a way that in a particular direction signals interfere constructively whereas in other directions the interferences are destructives. It is a question of measuring and applying a delay on the signals acquired by the microphones.

Beamforming is used with radio or sound waves and has many applications in radar, sonar, seismology and acoustics. It allows a representation of a wave field by an estimate of the directions of arrival.

Let an antenna formed by 10 microphones (red stars in FIG. 6), a sinusoidal source (green circle in FIG. 6) with a frequency of 2500 Hz and a white noise (blue circle in FIG. 6). The sources are both placed 0.20 meters from the plane of the antenna and spaced from each other by 0.05 meters.

Through a Beamforming algorithm, we are able to distinguish these two sources.

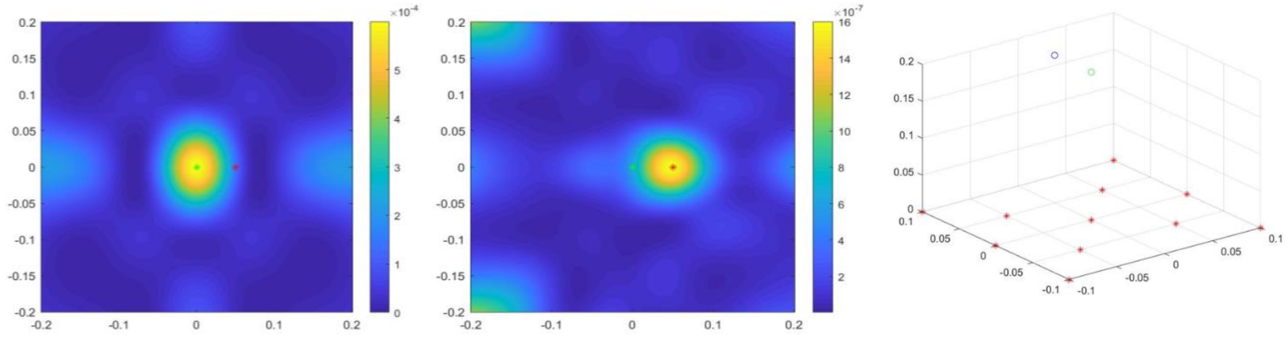


Figure 7: Measurement of the pressure field at 2500 Hz on the left, at a different frequency in the middle

Using an antenna makes the diagnosis less sensitive to disturbance and thus more reliable. We can from this observation reconstruct the acoustic signal coming only from the direction of the source in order to apply the diagnosis to a non-noisy signal. Therefore, we hope to find at least the results obtained in Table 1. An improvement in these results will then be considered from a more advanced imaging diagnosis.

Conclusion and further works

The results introduced above show that the use of classification algorithm is relevant to diagnose electric motors. However, even if these results seem good it is not possible to consider an implementation of these tools in a production bench because 1.33% of False Positive misclassification, achieved with Random Forest in 2.1.1, on a production of 5000 starters per day first means that 67 electric motors are misclassified and in this particular case 67 electric motors are in a bad condition but classified as healthy could be sent to customers.

We will first focus the work on collecting data with the microphone array in a production environment and then improve the recognition by using classification algorithms developed on time signals and spectrograms. By using algorithms near from our application field and adding spatial informations with Beamforming, we hope increase the recognition tasks.

References

[1] C. Bishop, Pattern Recognition and Machine Learning, New York, Springer, 2006.

[2] Hanson, Brian, and Ted Applebaum. "Robust speaker-independent word recognition using static, dynamic and acceleration features: Experiments with Lombard and noisy speech." Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on. IEEE, 1990.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. "ImageNet Classification with Deep Convolutional Neural Networks", Pages 1097-1105 NIPS'12, 2012.

[4] H.C.Nathanson, et al., "The Resonant Gate Transistor", IEEE Trans. Electron Devices, March 1967, vol. 14, 1967.

[5] J.J. Christensen, J. Hald, “Beamforming”, Brüel & Kjaer Technical Review N°1, 1-48, 2004.

[6] J.-H. Thomas, Diagnostic d’un système non stationnaire à partir d’une approche de reconnaissance des formes floue. Application au diagnostic de cliquetis moteur, Traitement du Signal, 16 (3), 203-216, 1999.

[7] M. Mein, Surveillance acoustique de moteurs diesel, thèse de doctorat de l’Université du Maine, Le Mans, 1997.

Macroscopic-Microscopic Attention in LSTM Networks based on fusion Features for prediction of bearing remaining life

QIN Yi[1,2] XIANG SHENG[1,2] HAORAO YAN[1,2]

[1] College of Mechanical Engineering, Chongqing University, Chongqing 400044, China
qy_808@aliyun.com
a1121623518@1636.com
[2] State Key Laboratory of Mechanical Transmission, Chongqing University, Chongqing 400044, China
18334729055@163.com

Abstract

In the mechanical transmission system, the bearing is one of the most widely used transmission components. The failure of the bearing will cause serious accident and huge economic loss. Therefore, the remaining life prediction of the bearing is greatly important. In order to predict the remaining life of the bearing, a prediction method combining macro-micro attention, long-short-term memory neural network and isometric mapping is proposed. First, some typical time-domain and frequency-domain characteristics of the vibration signal are calculated .respectively, such as the the mean frequency, the absolute mean value, the standard deviation, the RMS and so on. Then, the principal component of these characteristics is extracted by the isometric mapping method. The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

1 Introduction

Bearing is widely used in mechanical equipment and is one of the most universally used mechanical parts[1]. Under the complex working condition and environment, bearings are easily subject to failures, which may result in the catastrophe of the machine running and even threaten the personal safety [2]. Bearing remaining life prediction is beneficial to determine the equipment maintenance time reasonably, improve the production efficiency, reduce the accident rate, and prevent the sudden accidents, which is significant for engineering production [3].Recently, with the rapid developments in sensing, signal processing and artificial intelligence technology, Prognostics and Health Management (PHM) technologies based on data (experience) have gradually become the mainstream solution either in fault diagnosis or remaining useful life（RUL）estimation, instead of physics-based methods which can be expensive and tedious to develop[4]. Neural network is one of the most advanced models for the study of sequence classification and prediction in data driven methods. Zhao et al. [5] introduced the application of deep learning in machine health monitoring system, mainly from the perspective of automatic encoder (AE) and its variants, restricted boltzmann machine and its variants, including deep belief network (DBN) and deep boltzmann machine (DBM), convolutional neural network (CNN) and recursive neural network (RNN). Qin et al. [6] proposed a model for fault diagnosis of wind turbines' gearboxes based on deep belief networks (DBNs) with improved logistic sigmoid units via extracting impulsive features. Gebraeel et al. [7] used the amplitude of the bearing vibration decay signal as the criterion to assess the life, and used the BP neural network to predict the life.

Based on particular design, RNN is suitable for processing timing-related information [8]. However, RNN also has its own drawbacks, e.g. the excessive recursion time is indirectly equivalent to increasing the depth of neural network and the training time; the vanishing gradient problem usually [9]. In order to solve these problems, long short-term memory (LSTM) was proposed by Hochreiter and Schmidhuber in 1997 [10]. It can avoid the long-term dependence problem of RNN, thus it has been widely used. Yuan et al. [11] investigated three RNN models including vanilla RNN, LSTM and GRU models for fault diagnosis and prognostics of aero engine. They found that these advanced RNN model based on LSTM and GRU performed better than the convetional RNN via a number of experiments. Elsheikh et al. [12] proposed the bidirectional handshake network which solved the problem that bidirectional LSTM could not be well utilized in the prediction field, and proved the superiority of the method in the life experiment of turbine engines.

Attention mechanism can be regarded as a kind of contribution screening of information which improves the efficiency of neural network by selecting key information for processing. Attention-based recurrent networks have been successfully applied to a wide variety of tasks, such as handwriting synthesis[13], machine translation[14], image caption generation[15] and visual object classification[16]. In the prediction field, attention-based LSTM is getting more and more attention. Ran et al. [17] substituted a tree structure with attention mechanism for the unfolding way of standard LSTM to construct the depth of LSTM and modeling long-term dependence for travel time prediction. Fernando et al. [18] combined two kinds of attention and used LSTM for human trajectory prediction and abnormal event detection. Filtering key information can reduce computing resources, but it can also cause some degree of information loss. Differential treatment of input data according to the screening of attention mechanism can not only reflect the focus to important information, but also retain useful information as far as possible.The importance of fusional characteristic information is filtered via a proposed macro-micro attention mechanism, so that the input weight of neural network data and recursive data can reach multi-level real-time amplification. With the new long short-term memory neural network, the characteristic of the bearing vibration signal can be predicted based on the known fusional characteristic. The experimental results show that the method can predict the remaining life of the bearing well and has higher prediction accuracy than the conventional LSTMs.

2 Related work

2.1 LSTM

The LSTM neural network is a special RNN neural network which also consists of an input layer, a hidden layer, and an output layer. The difference lies in using an LSTM structure that includes the input gate, the output gate, the forget gate, and the memory cell as the hidden layer, as shown in Figure. 1.

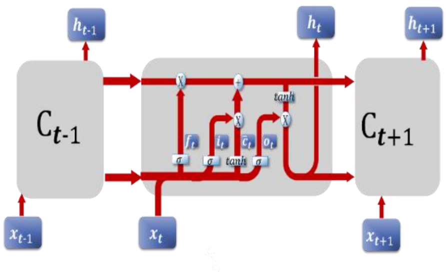


Figure. 1 Hidden layer structure of LSTM neural network

The forget gate is used to determine whether to keep the historical information stored in the current memory cell. If the door is opened, the historical information stored in the current memory cell is retained, otherwise the historical information is forgotten. The input gate is used to determine whether to allow the input layer information to enter the current memory cell. The open door allows the input layer signal to enter, and the closed door does not allow. The output gate is used to determine whether to output the current input layer signal to the next layer, the open door allows signal output and the closed door does not allow.

An LSTM network computes a mapping from an input sequence $\mathbf{x} = (x_1, x_2, \ldots x_n)$ to an output sequence $\mathbf{y} = (y_1, y_2, \ldots y_n)$ by calculating the network unit activations using the following equations iteratively from t = 1 to T:

$$
\begin{aligned}
\mathrm{i}_t &= \sigma\left(\mathrm{w}_{ix}\mathrm{x}_t + \mathrm{w}_{im}\mathrm{h}_{t-1} + \mathrm{b}_i\right) \\
\mathrm{f}_t &= \sigma\left(\mathrm{w}_{fx}\mathrm{x}_t + \mathrm{w}_{fm}\mathrm{h}_{t-1} + \mathrm{b}_f\right) \\
\mathrm{c}_t &= \mathrm{f}_t \quad \mathrm{c}_{t-1} + \mathrm{i}_t \quad \tanh\left(\mathrm{w}_{cx}\mathrm{x}_t + \mathrm{w}_{cm}\mathrm{h}_{t-1} + \mathrm{b}_c\right) \\
\mathrm{o}_t &= \sigma\left(\mathrm{w}_{ox}\mathrm{x}_t + \mathrm{w}_{om}\mathrm{h}_{t-1} + \mathrm{b}_o\right) \\
\mathrm{h}_t &= \mathrm{o}_t \quad \tanh\left(\mathrm{c}_t\right)
\end{aligned}
\tag{1}
$$

where, i is the input gate, o is the output gate, f is the forgetting gate, c includes cell activation vectors, and h is the memory cell outputs. w represents the weight matrix (for example, the weight w_{ix} matrix representing the input x to the input gate), and b represents the threshold (b_i which is the threshold of the input gate). σ is the sigmoid activation function, $\tanh$ is the tanh activation function, represents dot product.

3 Macroscopic-Microscopic Attention in LSTM

The structure of MMALSTM neural network is in Figure. 2. The number of input cells and the number of output cells in MMALSTM are respectively set as 60 and 1, and the learning rate is set as 0.05. The number of hidden layer cells is set as 17 in this study. The initialization method of neural network employs the standard initialization. And the recurrent neural network based on MMA is deduced as follows.

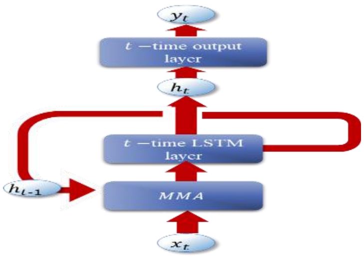


Figure. 2 structure of MMALSTM neural network

Firstly, we deal with the data matrix and calculate its macro and micro attention coefficient by the macro -micro attention mechanism. The input information $\mathbf{X}_t = \left[\mathbf{x}_1 \quad \mathbf{x}_2 \quad \dots \quad \mathbf{x}_t\right]$ at time t is defined as $\mathbf{x}_t = \left[x_{t,1} \quad x_{t,2} \quad \dots \quad x_{t,n}\right]$ and the recurrent information at time t-1 is defined as $\mathbf{h}_{t-1} = \left[h_{t-1,1} \quad h_{t-1,2} \quad \dots \quad h_{t-1,m}\right]$. Macro attention mechanism is to process the data in the whole time interval by attention mechanism. Micro attention mechanism is to process the data $\mathbf{x}_t = \left[x_{t,1} \quad x_{t,2} \quad \dots \quad x_{t,n}\right]$ and $\mathbf{h}_{t-1} = \left[h_{t-1,1} \quad h_{t-1,2} \quad \dots \quad h_{t-1,m}\right]$ in each time instant by attention mechanism. Thus the operation of attention mechanism on data in the whole time dimension and each time dimension is called macro-micro attention mechanism. In this paper, the macro -micro attention mechanism of input matrix and the micro attention mechanism of recurrent matrix are processed (The recurrent data in the whole time dimension is not known before input into the network, so it cannot be processed at the macro level). In the training process, $\bar{x}_{T+1}$ is set as query vector q_M at the macro level, $x_{t+1,n}$ is set as query vector q_m at the micro level. In the prediction process, $\bar{x}_T$, $x_{t,n}$ represents q_M and q_m separately. And then the attention coefficients are calculated by:

$$
\chi_i = \frac{\exp\left(S\left(\bar{x}_i, \mathrm{q}\right)\right)}{\sum_{j=1}^{T}\exp\left(S\left(\bar{x}_j, \mathrm{q}\right)\right)}
\tag{2}
$$

$$
\alpha_i = \frac{\exp\left(s\left(x_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n}\exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m}\exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)}
\tag{3}
$$

$$
\lambda_i = \frac{\exp\left(s\left(h_{t,i}, \mathrm{q}\right)\right)}{\sum_{j=1}^{n}\exp\left(s\left(x_{t,j}, \mathrm{q}\right)\right) + \sum_{p=1}^{m}\exp\left(s\left(h_{t-1,p}, \mathrm{q}\right)\right)}
\tag{4}
$$

where α_i , λ_i are micro attention coefficients corresponding to input data and recurrent data respectively, χ_i is macro attention coefficients corresponding to the whole input data, $\bar{x}_i$ represents the mean value of $\mathbf{x}_i$ .And

$$S(\bar{x}_i, q) = \frac{\bar{x}_i q}{\sqrt{t}} \tag{5}$$

$$s(x_i, q) = \frac{x_{t,i}^T q}{\sqrt{n+m}} \tag{6}$$

$$s(h_i, q) = \frac{h_{t-1,i}^T q}{\sqrt{n+m}} \tag{7}$$

Secondly, according to the macro -micro attention coefficient and Eq. (4), the weight of input data is amplified in real time at multiple levels. And the weight of recursive data is amplified in real time based on the micro attention coefficient.

$$\begin{aligned}
\mathbf{w}'_{ix} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ix} \\
\mathbf{w}'_{ox} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{ox} \\
\mathbf{w}'_{fx} &= (1+\chi_i)\times(1+\alpha_i)\mathbf{w}_{fx} \\
\mathbf{w}'_{ih} &= (1+\lambda_i)\mathbf{w}_{ih} \\
\mathbf{w}'_{oh} &= (1+\lambda_i)\mathbf{w}_{oh} \\
\mathbf{w}'_{fh} &= (1+\lambda_i)\mathbf{w}_{fh}
\end{aligned} \tag{8}$$

From the above, the flow chart of MMA, MA, ma are depicted in the follow figures. Figure. 3 represents the flowchart of MMA, Figure. 4 (a)represents the flowchart of MA, Figure. 5(b) represents the flowchart of ma.

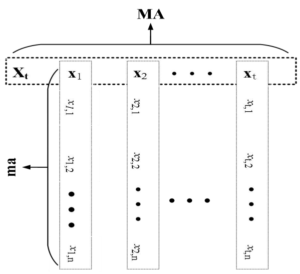
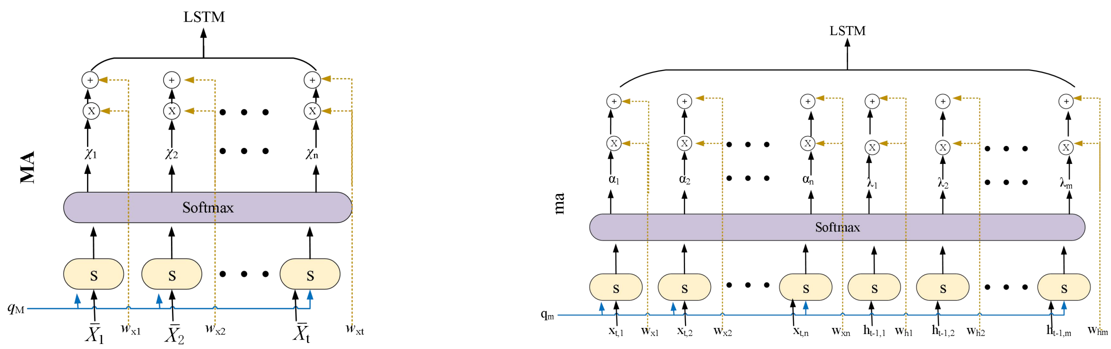


Figure. 3 The flowchart of MMA



(a)

(b)

Figure. 4 The flowchart of MA and ma.

With the weight amplified based on MMA, a new variant of LSTM is proposed, which is named as MMALSTM. Via Eq. (1), we can derive the calculation formula of MMALSTM as follows:

$$
\begin{aligned}
&\mathbf{i}_t = \sigma\left(\mathbf{w}'_{ix}\mathbf{x}_t + \mathbf{w}'_{ih}h_{t-1} + \mathbf{b}_i\right) \\
&\mathbf{f}_t = \sigma\left(\mathbf{w}'_{fx}\mathbf{x}_t + \mathbf{w}'_{fh}h_{t-1} + \mathbf{b}_f\right) \\
&\mathbf{c}_t = \mathbf{f}_t \quad \mathbf{c}_{t-1} + \mathbf{i}_t \quad \tanh\left(\mathbf{w}_{cx}\mathbf{x}_t + \mathbf{w}_{ch}\mathbf{r}_{t-1} + \mathbf{b}_c\right) \\
&\mathbf{o}_t = \sigma\left(\mathbf{w}'_{ox}\mathbf{x}_t + \mathbf{w}'_{oh}h_{t-1} + \mathbf{b}_o\right) \\
&\mathbf{h}_t = \mathbf{o}_t \quad \tanh\left(\mathbf{c}_t\right) \\
&\mathbf{y}_t = \mathrm{g}\left(\mathbf{w}_{yh}h_t + \mathbf{b}_y\right)
\end{aligned}
\tag{9}
$$

where g is the linear activation function.

4 Prediction method based on MMALSTM

In summary, the steps of the proposed method for predicting the bearing remaining life based on the combination of the isometric mapping algorithm [19] and MMALSTM neural network are given below:

1. During the bearing life cycle, collect the bearing vibration samples according to the interval Ts between two adjacent samples. Suppose that the number of sample is n .

2. Calculate the 5 kinds of time-frequency characteristics （time-domain :mean absolute difference，standard deviation and root-mean-square, frequency-domain: mean of frequency distribution, envelope spectrum: mean of frequency distribution）of these vibration samples separately, then the eigenvalue matrix X of the size $n \times 5$ can be obtained.

3. Select the eigenvalue matrix $X1$ of the $n1$ samples from X as the training matrix.

4. The training matrix $X1$ and original matrix X are respectively processed by the ISOMAP algorithm and Savitzky-Golay filtering, and the obtained eigenvectors with the largest eigenvalues after filtering $V1 = \left(v1_1, v1_2, \ldots, v1_{n1}\right)^T$ and $V = \left(v_1, v_2, \ldots, v_n\right)^T$ are used as their principal components respectively. The matrix X and the vector V are only used to verify the validity of this method, which don’t participate in neural network training and prediction.

5. Since the size of the matrix X is larger than that of the matrix $X1$ and both the sums of V and $V1$ are equal to zero according to the characteristic of ISOMAP algorithm, the two vectors may have different starting value, even though their trends are same. Therefore, it is necessary to unify them. By minimizing

$$
E = \sum_{i=1}^{n1} \left(v1_i - av_i - b\right)^2 \tag{10}
$$

a and b can be computed, and then all the elements of V are unified through the following equation

$$
v_i' = av_i + b \tag{11}
$$

6. Linearly normalize the vector $V1$ to obtain the normalized vector $W = \left(w_1, w_2, \ldots, w_{n1}\right)^T$.

7. Reconstruct matrix U :

$$
U = \begin{bmatrix} w_1 & w_2 & \cdots & w_{n1-p} \\ w_2 & w_3 & \cdots & w_{n1-p+1} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p+1} & w_{p+2} & \cdots & w_{n1} \end{bmatrix} = \begin{bmatrix} \mathbf{u_1} \\ \mathbf{u_2} \\ \vdots \\ \mathbf{u_{p+1}} \end{bmatrix} \tag{12}
$$

where p is the cell number of input layer and

$$
\mathbf{u}_i = \left[w_i \quad w_{i+1} \cdots \quad w_{n1-p+i-1}\right] \tag{13}
$$

8. Set the first p vectors of the matrix U as the input of the MMALSTM neural network and the last vectors as the output respectively, then train the MMALSTM neural network.

9. With the trained MMALSTM neural network, the mapping function f for prediction is defined. By inputting the last p vectors of the matrix U into the trained MMALSTM, the output $\overline{u}_{p+2}$ at the first prediction time (FPT) can be calculated as

$$
\overline{\mathbf{u}}_{p+2} = f\left(\mathbf{u}_2, \mathbf{u}_3 \cdots, \mathbf{u}_{p+1}\right) \tag{14}
$$

Then U is updated by

$$U = [\mathbf{u_1} \quad \mathbf{u_2} \quad \cdots \quad \mathbf{u_{p+1}} \quad \bar{\mathbf{u}}_{p+2}]^T \tag{15}$$

10. Repeating step 9 several times, we can obtain the predicted data series $\bar{u}_i$.

$$\begin{aligned}
&\bar{\mathbf{u}}_{p+3} = f\left(\mathbf{u}_3, \mathbf{u}_4 \cdots, \mathbf{u}_{p+1}, \bar{\mathbf{u}}_{p+2}\right) \\
&\bar{\mathbf{u}}_{p+4} = f\left(\mathbf{u}_4, \mathbf{u}_5 \cdots, \mathbf{u}_{p+1}, \bar{\mathbf{u}}_{p+2}, \bar{\mathbf{u}}_{p+3}\right) \\
&\bar{\mathbf{u}}_{p+5} = f\left(\mathbf{u}_5, \mathbf{u}_6 \cdots, \mathbf{u}_{p+1}, \bar{\mathbf{u}}_{p+2}, \bar{\mathbf{u}}_{p+3}, \bar{\mathbf{u}}_{p+4}\right) \\
&\qquad\qquad \vdots \\
&\bar{\mathbf{u}}_i = f\left(\mathbf{u}_{i-p}, \mathbf{u}_{i-p+1} \cdots, \mathbf{u}_{p+1}, \bar{\mathbf{u}}_{p+2} \cdots \bar{\mathbf{u}}_{p+(i-p-2)}, \bar{\mathbf{u}}_{p+(i-p-1)}\right)
\end{aligned} \tag{16}$$

When the predicted characteristic data $\bar{w}_{n1-p+i-1}$ ($\bar{w}_{n1-p+i-1} \in \bar{\mathbf{u}}_i$) exceeds a preset threshold, the bearing remaining life can be calculated by $(i-n1)\times T_s$. Then we anti-normalize the predicted data series and get $\bar{v}_i$ ($i = n1+1, n1+2, \cdots, n$), which can be compared with the actual vector $V' = \left(v'_{n1+1}, v'_{p+2}, \ldots, v'_n\right)^T$ to verify the validity of this method.

5 Experimental signal analysis

The experimental data come from PRONOSTIA in the IEEE PHM 2012 Data Challenge [20]. The platform mainly contains three major parts: a rotatory part, a degradation generation part, and a signal acquisition part. To accelerate the degradation of bearing, radial load force is applied with a controllable shaft speed. Two accelerated sensors perpendicular to each other are installed on the key position for the test bearing. The sampling frequency is 25 600 Hz. Each sample has a duration of 0.1 s, which means each sample has 2560 points. The record interval is 10 s. The test is ceased once the amplitude of the collected signal surpasses a certain level to prevent damage. Our proposed method is applied on the data set bearing1-1, which means that our method currently only considers the operation condition with constant speed and load. The run-to-failure bearing1-1 contains 2803 samples. Moreover, many samples were acquired under the steady stage, and their characteristics were almost the same, which were of little significance to use these data for prediction. Therefore, only the last 1750 samples were used for bearing life prediction.

The characteristic matrix of 1650 experimental samples is used as the training matrix to predict the 100 characteristic points corresponding to the next 100 time instants. Via the proposed method, the training, predictive and actual curves are illustrated in Figure. 7. As can be seen from Figure. 5, the trend of the predicted curve is not close to actual curves but both have down trend. It follows that the predicted value only exceed the threshold when the MMALSTM neural network predict a large number of points. In such case, the predicted sampling points are 1736, which is 14×10s different from the actual life. In this study, to assess the performance of the proposed approach, an error index is defined as

$$Er = \frac{Rul - \overline{Rul}}{Rul} \times 100\% \tag{17}$$

where Rul denotes the actual remaining useful life, and Rul denotes the predictive remaining useful life. It is easy to note that the prediction error Er is 9.3%.

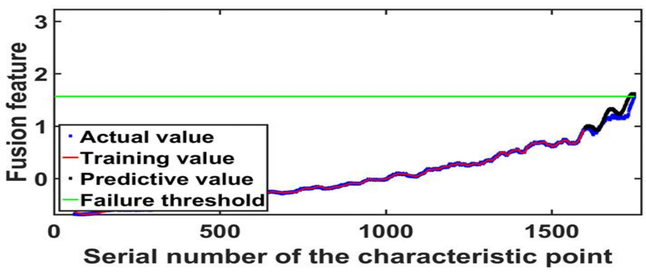


Figure. 5 Failure threshold, training curve, predicted curve and actual curve for 370 experimental samples of real bearing.

Then we add the number of known characteristic points to test the predictive power of the model, each group was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .6.

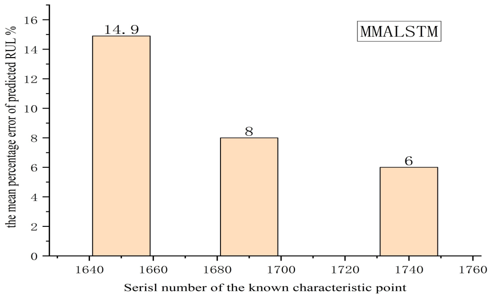


Figure. 6 The obtained prediction errors under the various numbers of the known characteristic point of data set .

Next, given 1690 characteristic points, MMALSTM, LSTM, deep LSTM (DLSTM) and LSTM with a projection (LSTMP) are used for comparison under the same conditions, each method was tested for 5 times, and the mean percentage error of predicted RUL was calculated, as shown in the Figure .7.

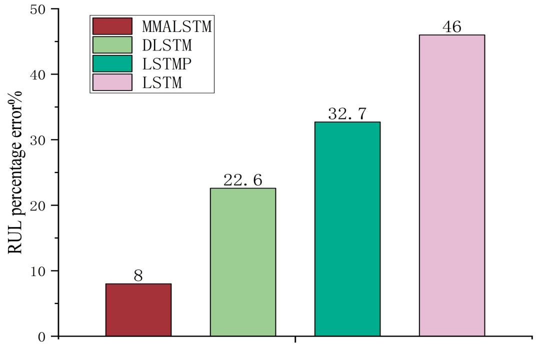


Figure. 10 Comparison of bearing remaining useful life prediction results obtained by the four types of LSTMs.

From the experimental results, it can be concluded that MMALSTM has not higher accuracy, while the four types of LSTMs do not have the satisfactory performance when fewer samples are used. And the closer the prediction point is to the failure point, the more accurate the prediction of bearing remaining life will be. Therefore, the proposed method is more suitable for bearing life prediction with enough samples.

6 Conclusion

Modern manufacturing requires high reliability and high efficiency, which makes bearing health analysis and residual life prediction an increasingly important research field. Advances in networked manufacturing and AI-oriented big data analysis provide new opportunities for bearing health analysis and residual life prediction. In this paper, isometric mapping algorithm is used to fuse 3 time-domain features,1 frequency-domain features and 1 envelope spectrum features of bearing vibration signal into a new feature. Due to the information characteristics of life data, the input weight of neural network and the weight of recurrent layer are all multi-amplified in real time by partial processing based on MMA. We use the improved long and short term neural network(MMALSTM) to study the bearing degradation model and predict the residual service life. The performance of the network model is verified by the data obtained from the bearing lifetime experiment and compared with LSTM, DLSTM and LSTMP. Numerical experiment results show that the proposed method not only has a higher accuracy of predicted. This will not only have certain reference significance and help for the online detection of bearing residual life, but also have great significance for the determination of equipment maintenance time and the prevention of unexpected accidents in engineering production.

In the future we will plan to study the performance of the proposed depth structure model in the field of bearing residual life prediction. Furthermore, the prediction of the residual life of bearings under variable working conditions by the deep learning model is also worth studying and working on

references

[1]Yi Qin, Jingqiang Zou, Baoping Tang, Yi Wang, Haizhou Chen. Transient feature extraction by the improved orthogonal matching pursuit and K-SVD algorithm with adaptive transient dictionary, IEEE Transactions on Industrial Informatics, 2019, DOI 10.1109/TII.2019.2909305

[2]Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," IEEE Transactions on Instrumentation and Measurement, vol. 66, no. 7, pp. 1693-1702, 2017.

[3]L. Song, H. Wang, and P. Chen, "Vibration-based intelligent fault diagnosis for roller bearings in low-speed rotating machinery," IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 8, pp. 1887-1899, 2018.

[4] X. Han, Y. F. Zhu. Prediction of Residual Fatigue Life of In-service Gears[J]. Journal of Wanxi University, 2003, 19(2): 85-87.

[5]R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," Mechanical Systems and Signal Processing, vol. 115, pp. 213-237, 2019.

[6]Y. Qin, X. Wang, and J. Zou, "The optimized deep belief networks with improved logistic Sigmoid units and their application in fault diagnosis for planetary gearboxes of wind turbines," IEEE Transactions on Industrial Electronics, vol. 66, no. 5, pp. 3814-3824, 2019.

[7]Gebraeel N, Lawley M, Liu R, et al. Residual life predictions from vibration-based degradation signals: a neural network approach[J]. IEEE Transactions on Industrial Electronics, 2004, 51(3):694-700.

[8]Graves A, Mohamed A R, Hinton G. Speech Recognition with Deep Recurrent Neural Networks[J]. 2013, 38(2003):6645-6649.

[9]Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 2002, 5(2):157-166.

[10]Hochreiter S, Schmidhuber J. Long short-term memory.[J]. Neural Computation, 1997, 9(8):1735-1780.

[11]M. Yuan, Y. Wu, and L. Lin, "Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network," in Aircraft Utility Systems (AUS), IEEE International Conference on, 2016, pp. 135-140: IEEE.

[12] A. Elsheikh, S. Yacout, M.-S. Ouali, Bidirectional handshaking LSTM for remaining useful life prediction, Neurocomputing, 323 (2019) 148-156.

[12]K. B. Charbonneau and O. Shouno, "Neural Trajectory Analysis of Recurrent Neural Network In Handwriting Synthesis," arXiv preprint arXiv:1804.04890, 2018.

[14]H. Choi, K. Cho, and Y. Bengio, "Fine-grained attention mechanism for neural machine translation," Neurocomputing, vol. 284, pp. 171-176, 2018.

[15]P. Anderson et al., "Bottom-up and top-down attention for image captioning and visual question answering," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 6077-6086.

[16]L. Wu, Y. Wang, X. Li, and J. Gao, "Deep attention-based spatially recursive networks for fine-grained visual recognition," IEEE transactions on cybernetics, no. 99, pp. 1-12, 2018.

[17]X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-Based Method with Attention Mechanism for Travel Time Prediction," Sensors, vol. 19, no. 4, p. 861, 2019.

[18]T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft+ hardwired attention: An lstm framework for human trajectory prediction and abnormal event detection," Neural networks, vol. 108, pp. 466-478, 2018.

[19]A. Najafi, A. Joudaki, and E. Fatemizadeh, "Nonlinear dimensionality reduction via path-based isometric mapping," IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 7, pp. 1452-1464, 2016.

[20]P. Nectoux et al., "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in IEEE International Conference on Prognostics and Health Management, PHM'12., 2012, pp. 1-8: IEEE Catalog Number: CPF12PHM-CDR.

Milling Diagnosis using Machine Learning Approaches

Dominique KNITTEL[1,2], Hamid MAKICH[1], Mohammed NOUARI[1]

[1]LEM3 (Laboratoire d'étude des microstructures et de mécanique des matériaux), site INSIC, 27 rue d'Hellieule, 88100 Saint Dié des Vosges, France
[2]University of Strasbourg, Faculty of Physics and Engineering, 3 - 5 rue de l'Université, 67000 Strasbourg, France
Corresponding author : knittel@unistra.fr

Abstract

The manufacturing industries more and more pay close attention to artificial intelligence (AI). For example, smart monitoring and diagnosis, real time evaluation and optimization of the whole production and raw materials management can be improved by using machine learning and big data tools.
In this work, a smart milling diagnosis has been developed for composite sandwich structures based on honeycomb core. The use of such material has grown considerably in recent years, especially in the aeronautic, aerospace, sporting and automotive industries. But the precise milling of such material presents many difficulties.
The objective of this work is to develop a data-driven industrial surface quality diagnosis for the milling of honeycomb material, by using supervised machine learning methods. In this approach cutting forces are online measured in order to predict the resulting surface flatness.
The developed diagnosis tool can also be applied to the milling of other materials (metal, polymer, …).

1 Introduction

The manufacturing industries more and more pay close attention to artificial intelligence (AI). For example, smart monitoring and diagnosis, real time evaluation and optimization of the whole production and raw materials management can be improved by using machine learning and big data tools [1]. An accurate milling process implies a high quality of the obtained material surface (roughness, flatness) [2]. With the involvement of AI-based algorithms, milling process is expected to be more accurate during complex operations.

T. Mikołajczyk *et al*. developed an Artificial Neuronal Network (ANN) for tool-life prediction in machining with a high level of accuracy, especially in the range of high wear levels, which meets the industrial requirements [3].

D. Pimenov *et al*. evaluated and predicted the surface's roughness through artificial intelligence algorithms (random forest, standard Multilayer perceptron) [4]: in their investigation the obtained performance depends on the parameters contained in the dataset.

M. Correa *et al.* compared the performances of Bayesian networks (BN) and artificial neural networks for quality detection in a machining process [5]. Even ANN models are often used to predict surface quality in machining processes, they preferred BNs for their significant representation capability and for the fast model building.

The work of C. Zhang *et al*. [21] focused on monitoring the condition and life of the cutting tool in dry milling environment. From de-noised vibration signal they extract some relevant features such as the root mean square, the skewness and the kurtosis in both time and time-frequency domain. Based on Neuro-Fuzzy Network (NFN), they implemented a tool wear prediction model which performs the best, with the smallest Mean Squared Error (MSE) and Mean Absolute Percentage Error (MAPE) compared with Back Propagation Neural Network (BPNN) and Radial Basis Function Network (RBFN) algorithms.

Z. Rui *et al*. [24] implemented a hybrid approach combining handcrafted feature design with automatic feature learning for machine health monitoring: local feature-based gated recurrent unit (LFGRU) networks. By comparison with some other methods such as the Support Vector Machine (SVM), the k-nearest neighbor (kNN), they verified the effectiveness and robustness of the proposed LFGRU model for tool wear prediction.

D. Wu *et al*. [25] have worked on cloud-based machine learning for tool wear prediction in milling. The research was about the development of a novel approach for machinery prognostics using a cloud-based random forest algorithm. Their experimental result have shown that despite the fact that random Forests give the best accuracy for large dataset, parallel random forest algorithm has the best ratio training time/accuracy. Future more, they will predict tool wear with other machine learning algorithms such as support vector machines as well as to make a comparison with their actual algorithms.

For machining result prediction, similar algorithms could be used but the recurrent problem is how to increase the accuracy of those algorithms. K. Javed. *et al.* [26] have worked on an enabling health monitoring approach based on vibration data for accurate prognostics. They have shown that prognostic efficiency is closely related to the extracted features and by the same way proposed a method for enabling features that can lead to simple and accurate prognostics.

K. Durmus [27], by using neuronal networks, worked on the prediction and the control of surface roughness in CNC lathe using artificial neural network. His study has concluded that artificial neural network (ANN) can produce an accurate relationship between cutting parameters and surface roughness. Based on the ANN training model, he could find the best machining parameters for obtaining a desired surface roughness.

By also using neuronal artificial neural network M. Azlan [28] has developed a surface roughness prediction models for end milling machining, in the logic to find the best ANN network structure for surface roughness prediction.

Another approach consists to measure and analyze the drive power (for example by current measuring) [31], which is not applicable in our experiment. In this paper few artificial intelligence methods are tested: random forest (RF), standard Multilayer perceptrons (MLP), Regression Trees, and radial-based functions.

In our work, a smart milling diagnosis has been developed for composite sandwich structures based on honey-comb core. The use of such material has grown considerably in recent years, especially in the aeronautic, aerospace, sporting and automotive industries. Recent development projects for Airbus A380 or Boeing 787 confirm the in-creased use of the honeycomb material. But the precise milling of such material presents many difficulties.

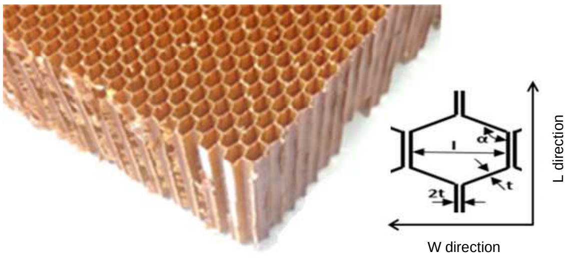
The objective of this work is to develop a data-driven industrial surface quality diagnosis for the milling of honeycomb material, by using supervised machine learning methods. Therefore, cutting forces are online measured in order to predict the resulting surface flatness.

2 Description of the Experiments

2.1 Workpiece material and tools

The workpiece material studied in this investigation is Nomex® honeycomb cores with thin cell walls. It is produced from aramid fiber dipped in phenolic resin (Fig. 1).

Honeycomb designation: **A10-72-5OX**



Density [Kg/m3]	Cell size l [mm]	Wall size t [mm]	Angle α [°]
72	5	0.08	100

Figure 1: *Nomex*® honeycomb cores and the main geometrical characteristics

The honeycomb cores consist of continuous corrugated ribbons of thin foil bonded together in the longitudinal direction. The aim of such a process is to create a structure allowing lightness and stiffness together thanks to the hexagonal geometry of formed cells. Figure 1 illustrates the geometric characteristics of the honeycomb core. The use of honeycomb material in sandwich composite is limited by the fragility of each wall of the honeycomb, which influences the quality of obtained surfaces after machining [7, 8, 9].

The Nomex® honeycomb machining presents several defects related to its composite nature (uncut fiber, tearing of the walls), the cutting conditions and to the alveolar geometry of the structure which causes vibration on the different components of the cutting effort [10].

It is clear that the use of ordinary cutting tools and also the mechanical and geometrical characteristics of honeycomb cores will have a crucial effect on machinability and on the quality of the resulting surface [11]. In fact, ordinary cutting tools for machining honey-comb core produce generally tearing of fibers and delamination of cell structures. Subsequently, these cause a reduction of bond strength between the skin and the honey-comb core, and thus a weaker joint for composite sandwich structures.

In our study, the used milling cutter is provided from our industry partner, the EVATEC Tools Company. As shown in figure 2, the used EVATEC tool is a combined specific tool with two parts designed to surfacing/dressing machining operation. The first part is a cutter body made of high speed steel with 16 mm in diameter and having ten helixes with chip breaker. This tool part is designated by Hogger. The second part is a circular cut-ting blade made of tungsten carbide with a diameter of 18.3 mm and having a rake angle of 22° and a flank angle of 2.5°. These two parts are mechanically linked to each other with a clamping screw.

Figure 2: Milling cutter used for Nomex® honeycomb core ‘‘CZ10’’.

2.2 Milling experiments

All experimental milling tests illustrated in this paper were carried out on a three-axis vertical machining center Realmeca® RV-8.

For assessing the performance of the machining process of Nomex® honeycomb core we monitored and measured the cutting forces generated during cutting, by using the Kistler dynamometer model 9129AA. The Kistler table is mounted below the Nomex sample in order to measure the three components of the machining force as shown in figure 3. During the measurements, the x-axis of the dynamometer is aligned with the feed direction of the milling machine and the longitudinal direction of the workpiece (parallel to core ribbons and the direction of honeycomb double wall). The three orthogonal components of machining force (Fx, Fy and Fz) were measured according to figure 3 using the Kistler table.

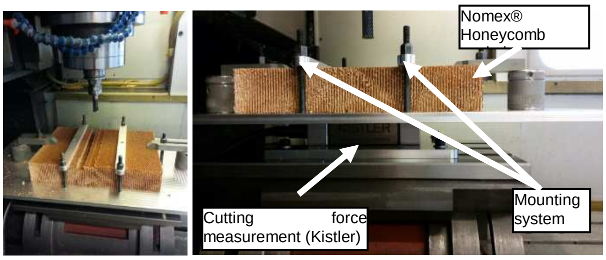


Figure 3: Experimental test setup

The milling experiment conditions are summarized in table 1. Four different speeds (high and low speeds) and four feed values were selected.

Spindle speed (rpm)	2 000	10 000	15 000	23 000
Feed rate (mm/min)	150	1 000	1 500	3 000

Table 1 : Milling experiment conditions

Two main modes of surface damage are observed (Figure 4): uncut aramid fibers along the machined surface and tearing of the walls. The appearance of the uncut fibers is a machining defect specific to the composite material which depends on the type of the fibers and their orientation. The tearing of Nomex® paper, linked to the cellular appearance of the honeycomb structure, occurs under the shear loading effect [28, 29].

Uncut fibers are observed on Figure 4 -a and -c. It is well known that the surface quality is of high importance for the use of the Nomex® honeycomb in sandwich materials. The machining defects cause a reduction of bond strength between the skin and the honeycomb core, and thus a weaker joint for composite sandwich structures.

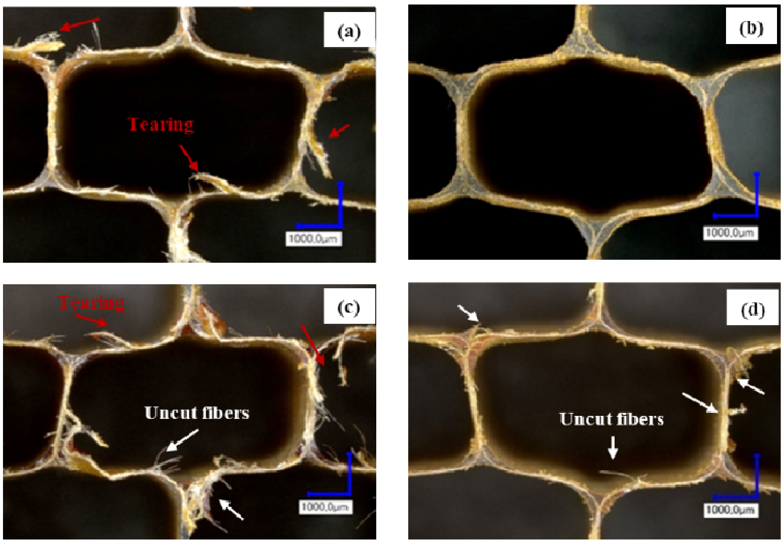


Figure 4: Obtained honeycomb machining surfaces.
The case (b) represents the best milling result

2.3 Measured signals

Many milling experiences have been made in our study. For example, figure 5 shows the milling forces measured for honeycomb at 2000 rpm spindle speed and 3000 mm/min feed rate.

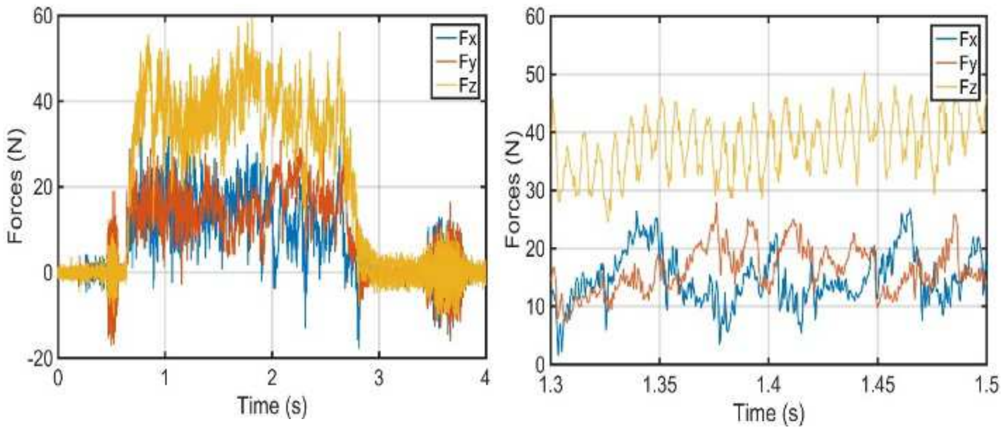


Figure 5: Milling force measurements for 2000 rpm spindle speed and 3000 mm/min feed rate:
(a) during all process; (b) during 0.2s (zoom)

Cutting forces are in the order of a few Newtons, they do not exceed 60 Newtons. Generally, the force in vertical direction (Fz) is quite small, thus, it is advised that to keeping vertical forces small in milling composite due to the delamination issue. In our case, the vertical cutting force component is greater than other forces components which can be attributed to the mechanical properties of the honeycomb structure where the honeycomb structure is characterized by a better out-of-plane compression behavior than its tensile and shear strength. The evolution of cutting forces shows significant oscillations. These oscillations are caused by vacuum in the cells of the honeycomb and the angle between the cutting direction and the honeycomb cell wall direction.

Figure 6 shows the obtained evolution of the surface quality (flatness) for various combinations of cutting conditions (spindle speed and feed rate). The defect of shape is higher for low speeds. Thus, for high feed rates that exceed the 1500 mm/min, the unevenness exceeds 500 µm which characterizes the severe tearing of the honeycomb walls.

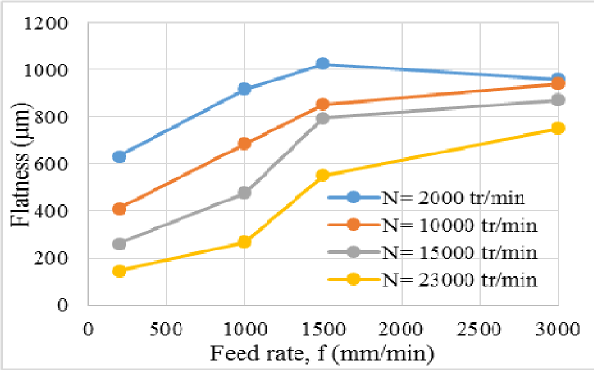


Figure 6: Effect of cutting parameters on surface flatness

Given the low level of cutting forces, the quality of the obtained machined surface allows to establish criteria for determining the machinability of the honeycomb structures. The appearance of the uncut fibers is a machining defect specific to the composite material which depends on the type of the fibers and their orientation. The tearing of Nomex® paper, linked to the cellular appearance of the honeycomb structure, occurs under the effect of shear loading [5, 12].

Alternatively a surface response [30] could have been built in order to predict the milling surface quality. But close milling parameters (such as splindle speed, feed rate, depth of cut) can lead to different results, depending on the material, the quality of the machining tool, etc.

Therefore, in our approach supervised machine learning techniques (with labeled measurements for the model training) are used. These tools need the construction of features associated with the measurements.

3 Milling diagnosis using machine learning techniques

Machine learning techniques can be separated mainly in two categories [17, 23]:

-Unsupervised approaches: based only on input data (data are unlabeled). The goal is to find groups and structures in the data set, in order to classify new observations (measurements) into the different groups.

- Supervised approaches: based on input and output data (Now the data are labelled.

The raw data (measurements) are firstly filtered, with low pass filters in order to eliminate high frequency noises, and labeled (“obtained signals for good surface quality”, “obtained signals for bad surface quality”). Then the features are calculated offline or online.

All the experiments are then split into two groups: 75% for the machine learning model training, 25% for the obtained model evaluation also called test phase in the literature (another percentage can be chosen, for example 60% - 40%, depending on the number of experiments). This can be made randomly, but the ratio “good surface quality” and “bad surface quality” must be kept in each group.

3.1 Features calculation

The features are calculated in the time domain and the frequency domain [6, 13] from the raw signal represented on figure 7, in steady state behaviour. In fact, transient zones (that means when the cutting tool entries or exits the honeycomb core) are not taken into account.

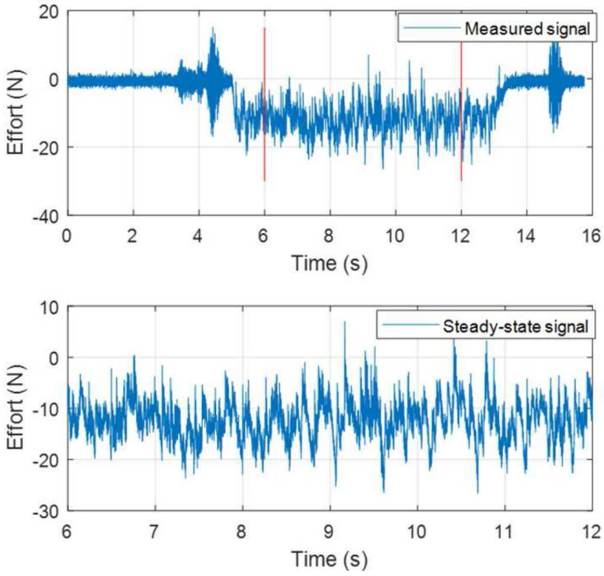


Figure 7: Measured milling force in time domain:
(a) total data plot, (b) signal during steady-state phase

After a first data processing (low pass filtering), firstly 19 features are calculated in the time domain for the measured milling force signal (for example : maximum, minimum, amplitude range, median value, maximum of the absolute value of the signal, interquartile range, average value of the signal, energy of the signal, Skewness, Kurtosis, Shannon entropy, …)

Secondly another 19 features are calculated in frequency domain in a similar way for the measured milling force signal. Therefore, the Fast Fourier transform (FFT) of the signal has been calculated.

All the calculated features (in time and frequency domains) are normalized and stored in a table whose lines and columns respectively represent the experimental number (also called instance) and the associated feature values. The description of the used features are indicated in [32]

The reduction of the features is then be made by using PCA (Principal Component Analysis) [18].

3.2 Labeled data

From the evaluation of the effect of the cutting parameters on surface flatness result, we defined two classes of surface quality applied to the output data of each observation (see table 2)

Label	Flatness (µm)	Qualitative value
'A‘	0 – 600	Best surface quality
‘B'	600 – …	Worst surface quality

Table 2: Label table for the experimental observations

3.3 Applied supervised learning algorithms

In this work, several classification algorithms have been implemented in the Matlab software environment [20, 21]: k-nearest neighbor (kNN), Decision trees (DT), Support Vector Machine (SVM). The different machine learning algorithms (with their adapted tuning parameters) are applied to the normalized labeled training data set (75% of the total experiments). The obtained trained models are then tested on the labeled test data set (25% of the total experiments). The objective is to find again the labels of the test data set: table 3 shows the obtained accuracy result of each algorithm.

Algorithms	Accuracy
KNN	83.4%
KNN k=2	81.3%
Weighted KNN k=2	83.4%
Chebychev KNN k=2	87.5%
Tree	99%
Pruned tree	66.67%
Linear SVM	83.4%
Gaussian SVM	66.67%

Table 3: Prediction error for the normalized data set

We used some news experimental data set in order to evaluate the performance of the trained model. The goal is to predict online (during milling) the surface quality. The results are presented here for the trained model by using the linear SVM classifier algorithm:

	Predicted class		
Actual class	A	B	
A	TP = 83%	FN = 17%	100%
B	FP = 0%	TN = 100%	100%

(TP: true positive rate; FN: false negative rate; FP: false positive rate; TN: true negative rate)

Table 4: Performance of the prediction using SVM classifier

The class B was the best predicted class. Despite the fact that linear SVM algorithm lost in performance for data set with large predictors (i.e. large number of features), it has been an accurate algorithm with a good prediction rate and the lowest training time.

4 Conclusion

The milling's performance is qualified by evaluating the roughness or the flatness of the resulted surface. In this work, different supervised machine learning algorithms have been implemented and compared. To do this, features were firstly calculated from measured milling forces and then each Artificial Intelligence (AI) based model has been trained by the labeled set of features. From the prediction results, SVM algorithm seems to be a good efficient diagnosis algorithm in this application of honeycomb material milling. The developed diagnosis approach can also be applied to the milling of other material.

References

[1] Beskri A., Mejri H., Mehdi K. and Rigal J.F., *Systèmes de surveillance automatique en usinage: Moyens et méthodes* (Automatic monitoring systems in machining: tools and methods), French Mechanics Congress 2013.

[2] IFPM-Formation, *Usinage: Tournage Fraisage* (IFPM courses : Machining: Milling Turning), September 2015

[3] Mikołajczyk T., Nowicki K., Bustillo A. and Pimenov D., *Predicting tool life in turning operations using neural networks and image processing*, Mechanical Systems and Signal Processing, 104: 503-513, 2018.

[4] Pimenov D. Bustillo A. and Mikołajczyk T., *Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth*, Journal of Intelligent Manufacturing, 29(5): 1045-1061, 2018.

[5] Correa M. Bielza C. and Pamies-Teixeira J., *Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process*, International journal of Expert Systems with Applications, 36: 7270-7279, 2009.

[6] A.I.H. Committee, ASM Handbook Volume 16: Machining, ASM International, 1989.

[7] J. Kindinger, Lightweight structural cores, ASM Handbook Met. Composites, 21: 2001.

[8] D. Gay, Matériaux composites (Composite materials), Hermes, 2015.

[9] R. Carl, *Three-dimensional honeycomb core machining apparatus and method*, US Pat. App. 13/707,670. 1, 2012.

[10] Jaafar M., Atlati S., Makich H., Nouari M., Moufki A. and Julliere B., *A 3D FE modeling of machining process of Nomex® honeycomb core: influence of the cell structure behaviour and specific tool geometry*, Procedia CIRP, 58: 505-510, 2017.

[11] Mendoza M. et al., *Development of a new milling cutter for aluminum honeycomb*, Int. J. Mach. Tool Des. Res. 23: 81-91, 1983.

[12] Tchoutouo H., *Adhesiveless honeycomb sandwich structure with carbon graphite prepreg for primary structural application: a comparative study to the use of adhesive film*, master thesis, Wichita State University, May 2012.

[13] Rion J. Leterrier Y. and Manson E., *Prediction of the adhesive fillet size for skin to honeycomb core bonding in ultra-light sandwich structures*, Compos. Part A, 39: 1547-1555, 2008.

[14] Agusmian P. Ooijevaar T. and Kilundu B., *Automated bearing fault diagnostics with cost effective vibration sensor*, WCEAM/VETOMAC, Conference, 2017

[15] Madhusudana CK, Budati S., Gangadhar N., Kumar H. and Narendranath S., *Fault diagnosis studies of face milling cutter using machine learning approach*, Journal of Low Frequency Noise, Vibration and Active Control, 35(2): 128-138, 2016

[16] Gopal S. and Kishore K., *Normalization: A Preprocessing stage*, CSE & IT department, VSSUT, Burla, India 2015.

[17] The MathWorks Inc., *Mastering Machine Learning A Step-by-Step Guide with MATLAB*, Matlab Ebook section 1 to 4, ©2018

[18] Vidal R., Ma Y., Sastry S.S., *Generalized Principal Component Analysis*, Springer 2016.

[19] Yang F., Yun Z., Haiyu Q., Dequn L., Huamin Z. and Jürgen L., *Analysis of feature extracting ability for cutting state monitoring using deep belief networks*, Conference on Modelling of Machining Operations (CIRP), 31: 29-34, 2015.

[20] Chen G. Wei X., Yan R. and Yuqing Z., *Numerical control machine tool fault diagnosis using hybrid stationary subspace analysis and least squares support vector machine with a single sensor*, MDPI conference, 7: 346-358, 2017

[21] Zhang C., Yao X., Zhang J. and Jin H., *Tool Condition Monitoring and Remaining Useful Life Prognostic Based on aWireless Sensor in Dry Milling Operations*, MDPI conference, 16: 795-815, 2016.

[22] Kubat M., *An introduction to machine learning*, Second Edition, Springer, 2017

[23] Shalev-Shwartz S. and Ben-David S., *Understanding Machine Learning: From Theory to Algorithms*, Published by Cambridge University Press, 2014.

[24] Rui Z., Dongzhe W., Ruqiang Y., Kezhi M., Fei S. and Jinjiang Wang, *Machine Health Monitoring Using Local Feature-Based Gated Recurrent Unit Networks*, IEEE Transactions on Industrial Electronics, 65(2): p. 1539-1548, February 2018.

[25] Wu D., Jennings C., Terpenny J. and Kumara S., *Cloud-Based Machine Learning for Predictive Analytics: Tool Wear Prediction in Milling*, IEEE International Conference on Big Data, p. 2062-2069, 2016.

[26] Javed K., Gouriveau R., Zerhounic N., and Nectoux P., *Enabling Health Monitoring Approach Based on Vibration Data for Accurate Prognostics*, IEEE Transactions on Industrial Electronics, 62(1): 647-656, January 2015.

[27] Durmus Karayel, *Prediction and control of surface roughness in CNC lathe using artificial neural network*, journal of materials processing technology 209: 3125–3137, 2009.

[28] Azlan Mohd Zain, Habibollah Haron and Safian Sharif, *Prediction of surface roughness in the end milling machining using Artificial Neural Network*, Expert Systems with Applications 37: 1755–1768, 2010.

[29] Codjo L., Jaafar M., Makich H., Knittel D. and Nouari M., *Milling diagnosis using machine learning techniques toward Industry 4.0*, 29th International Workshop on Principles of Diagnosis DX'18, Warsaw, August 2018.

[30] Myers R., Montgomery D., Anderson-Cook C, "Response surface methodology", Wiley 2016

[31] D. Yu. Pimenov, A. Bustillo, T. Mikolajczyk, *Artificial intelligence for automatic prediction of required surface roughness by monitoring wear on face mill teeth*, Journal of Intelligent Manufacturing, June 2018, Volume 29, Issue 5, pp 1045–1061

[32] Codjo L., Jaafar M., Makich H., Knittel D. and Nouari M., *Honeycomb Core Milling Diagnosis using Machine Learning in the Industry 4.0 Framework*, IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), Torino Italy, September 2018

Passive control of vibrations

Experimental identification of the corrective effect of a non-circular pulley: application to timing belt drive dynamics

Sébastien PASSOS[1,2], Lionel MANIN[2], Didier REMOND[2], Olivier SAUVAGE[1], Laurent ROTA[1] and Etienne BESNIER[2]

[1]Groupe PSA, Scientific and Future Technologies Department, OpenLab Vibro-Acoustic-Tribology@Lyon, F-78943 Vélizy-Villacoublay, France

[2]Univ Lyon, INSA-Lyon, LaMCoS, CNRS UMR5259, F-69621 Villeurbanne, France

sebastien.passos@mpsa.com
sebastien.passos@insa-lyon.fr

Abstract

The work presented in this paper aims at showing experimentally how an oval pulley can generate a corrective effect on a timing belt drive subjected to a 2nd order periodic excitation (fluctuating load torque). For that purpose, a simplified timing belt drive of a 4-cylinder car engine is reproduced on a test stand. The oval pulley is mounted on the crankshaft axis. Experiments are conducted for different phasing angles of the oval pulley at several driving speeds. Pulley angular vibrations and span tension fluctuations are monitored. The results obtained from these experiments are analyzed in the angular and angular frequency domains. The results are compared with experiments performed on a reference case (equivalent circular crankshaft pulley). The study focuses on the effect of the oval pulley phasing angle on the amplitude of the 2nd order harmonics governing the angular response of the transmission.

1 Introduction

In an automotive engine, the Timing Belt Drive (TBD) is a key component in charge of synchronizing the valve train system and the crankshaft (figure 1). This synchronization constitutes one of the essential functions of the engine. It ensures the correct timing of the valves opening and closing with respect to the piston movement.

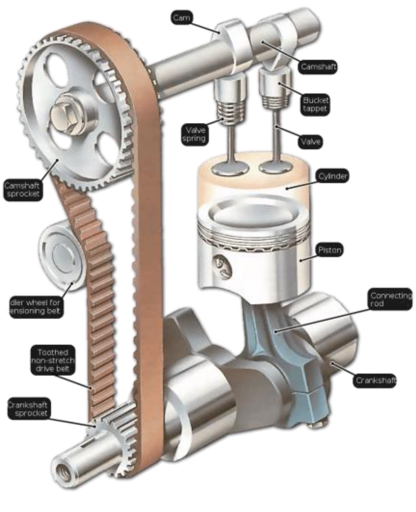


Figure 1: Principle of the valve train system of a car engine with a timing belt drive

In running conditions, TBD are exposed to a very harsh vibratory environment. Various excitation sources such as crankshaft acyclism and fluctuating load torques applied to driven pulleys, generate vibratory phenomena that may affect the TBD dynamic performances and life. In particular, two phenomena require careful monitoring and need to be minimized:

- Rotational vibrations of the transmission axes.
- Tension force fluctuations in the belt spans.

If mishandled, these two phenomena may induce increased component fatigue, power losses, noise and in the most extreme case can cause desynchronization of the TBD that could result in engine failure (piston-valve clash).

In response to ever-stricter requirements for engine efficiency and reliability, car manufacturers now commonly use TBD comprising innovative pulleys with Non-Circular (NC) pitch profile. When rotating, a NC pulley causes periodic elongations of its adjacent belt spans. Hence, it behaves like an exciter that may generate a corrective rotational excitation able to counteract the other excitation sources acting on the TBD. It is now known that for optimal profile shape and phasing in the transmission, the use of a NC pulley can improve considerably the vibratory performances of a transmission [1,2,3,4]. Nevertheless, determining the optimal design parameters of a non-circular pulley remains hard to accomplish. To achieve this, it is important to clearly understand and identify the impact of such pulleys on the dynamic behaviour of TBD.

The literature on this subject is relatively poor. Most works concern kinematics and quasi-static analyses as reported in [5,6]. In recent papers, Zhu et al. [2] and Passos et al. [3,4] propose numerical models developed to predict the dynamic behaviour of a transmission comprising NC pulleys. Using these models, the authors perform numerical studies that show how an oval crankshaft pulley can significantly reduce the rotational vibrations of a camshaft pulley in a four-cylinder engine (2nd order periodic camshaft load torque). The models are based on a discrete approach (also called 0D/1D approach) similar to that implemented by Hwang et al. [7] for poly-V belt transmissions comprising circular pulleys only. In his model, Hwang considers the belt spans as linear spring-damper elements connected to the pulleys represented by rotational inertias. In [8], Passos et al. conducted an experimental investigation on a basic transmission comprising two pulleys. Two configurations of the basic transmission were studied: one with an oval pulley mounted on the driving axis and the other with an equivalent circular pulley. The experiments were done with no excitation source (constant driving speed and constant load torque applied to the driven pulley). Instantaneous angular speed and acceleration of the driven pulley, transmission error and transmitted torques obtained for the two configurations of the transmission were analyzed and compared. The comparison provided a rigorous description of the proper effect of the oval pulley on the rotational dynamics of the transmission.

The present work aims at completing the studies presented in [3,4,8]. It shows experimentally how an oval pulley can generate a corrective effect able to counteract an excitation source such as a 2nd order periodic (H2) load torque. For that purpose, the oval pulley is mounted on the driving axis of a transmission whose architecture is similar to that of a 4-cylinder car engine timing belt drive. Usually, in this type of transmissions the camshaft pulley is subjected to a fluctuating load torque having a 2nd order periodicity. The test stand is presented in section 2. Experiments are conducted for different phasing angles of the oval pulley and at several driving speeds. Angular vibrations and span tension fluctuations are monitored. The results obtained from these experiments are analyzed and compared with those of a reference case (equivalent circular crankshaft pulley) in section 3.

2 Test stand

2.1 Studied transmission

The studied transmission is represented in figure 3 and its geometrical characteristics are given in table 1. The transmission has been designed in order to reproduce a simplified TBD of a 4-cylinder car engine. It comprises four pulleys: one oval crankshaft pulley (driving pulley), one camshaft pulley and one idler pulley on each side of the transmission. All the pulleys are circular except the crankshaft pulley that has an elliptical pitch profile. Such a pulley has two design parameters: its eccentricity e and its phasing angle ϕ_0 corresponding to the initial orientation angle between the oval pulley major axis and the y-axis (figure 3). In the present study, the eccentricity is imposed (marketed component) and the phasing angle is freely adjustable.

$$\phi_0 = \theta_{CS}(t=0) \tag{1}$$

Three levels of stationary crankshaft speeds are considered ($\omega_{CS} \in \{600 ; 1500 ; 3000 \text{ rpm}\}$). The camshaft pulley is subjected to a fluctuating load torque C_{CAM} having a 2nd order periodicity with respect to the crankshaft rotation θ_{CS}. The measured load torque is plotted versus the crankshaft angle for the three different driving speeds in figure 2 (a, c, e). The angular frequency content of the torque is shown in figure 2 (b, d, f). Whatever the driving speed, the torque is governed by a dominant 2nd order harmonic and secondary even harmonics of 4th, 6th and 8th orders. The torque is also modulated by harmonics of 0.5th and 1st orders due to geometry faults and misalignments affecting slightly the measurement. When the crankshaft runs at 3000 rpm, order 9.5 appears due to dynamic effects occurring in the mechanical device that generates the load torque (see section 2.2).

	Coordinates of the pulley centers		**Pitch profiles of the pulleys**
	X (cm)	**Y (cm)**	
Crankshaft	0	0	Ellipse: – Major axis $a = R_{CS} + e$ – Minor axis $b = R_{CS} - e$ with $\begin{cases} R_{CS} = 3.19\ cm \\ e = 0.07\ cm \end{cases}$
Idler (Tight side)	17.3	-2.8	Circle of radius $R_{IT} = 3\ cm$
Camshaft	39.3	0	Circle of radius $R_{CAM} = 2R_{CS}$
Idler (Slack side)	16.9	4.7	Circle of radius $R_{IT} = R_{IS}$

Table 1: Geometrical characteristics of the transmission

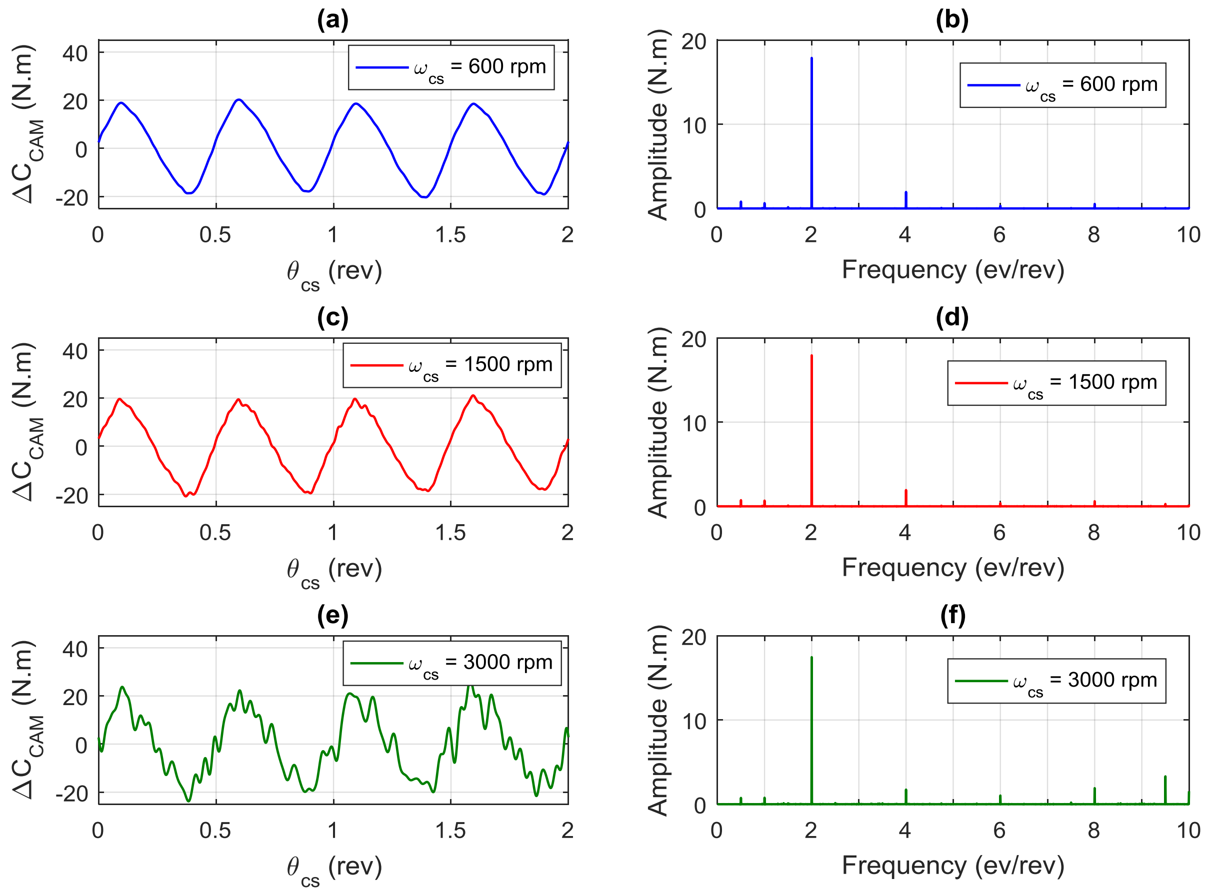


Figure 2: Variations of camshaft load torque in the angular (a, c, e) and angular frequency (b, d, f) domains

2.2 Global architecture

The global architecture of the test stand is illustrated in figures 3 and 4. It comprises two rotary shafts (one driving and one driven shaft) and two supporting parts for idler pulleys. The distances between the pulleys are freely adjustable, enabling a custom setting of the transmission geometry.

The crankshaft pulley is mounted on the driving axis coupled to a speed-controlled electric motor. The driven axis that supports the camshaft pulley of the studied transmission is coupled to the camshaft of a cylinder head fully equipped. It enables applying a real camshaft load torque to the transmission (see figure 2). This load torque has already been discussed in section 2.1.

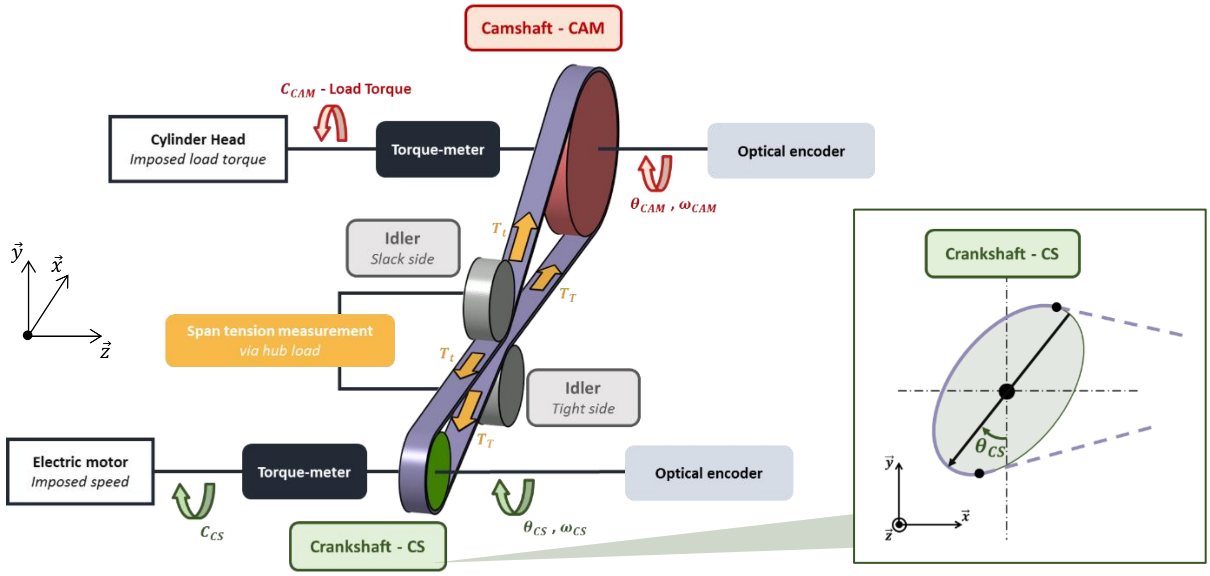
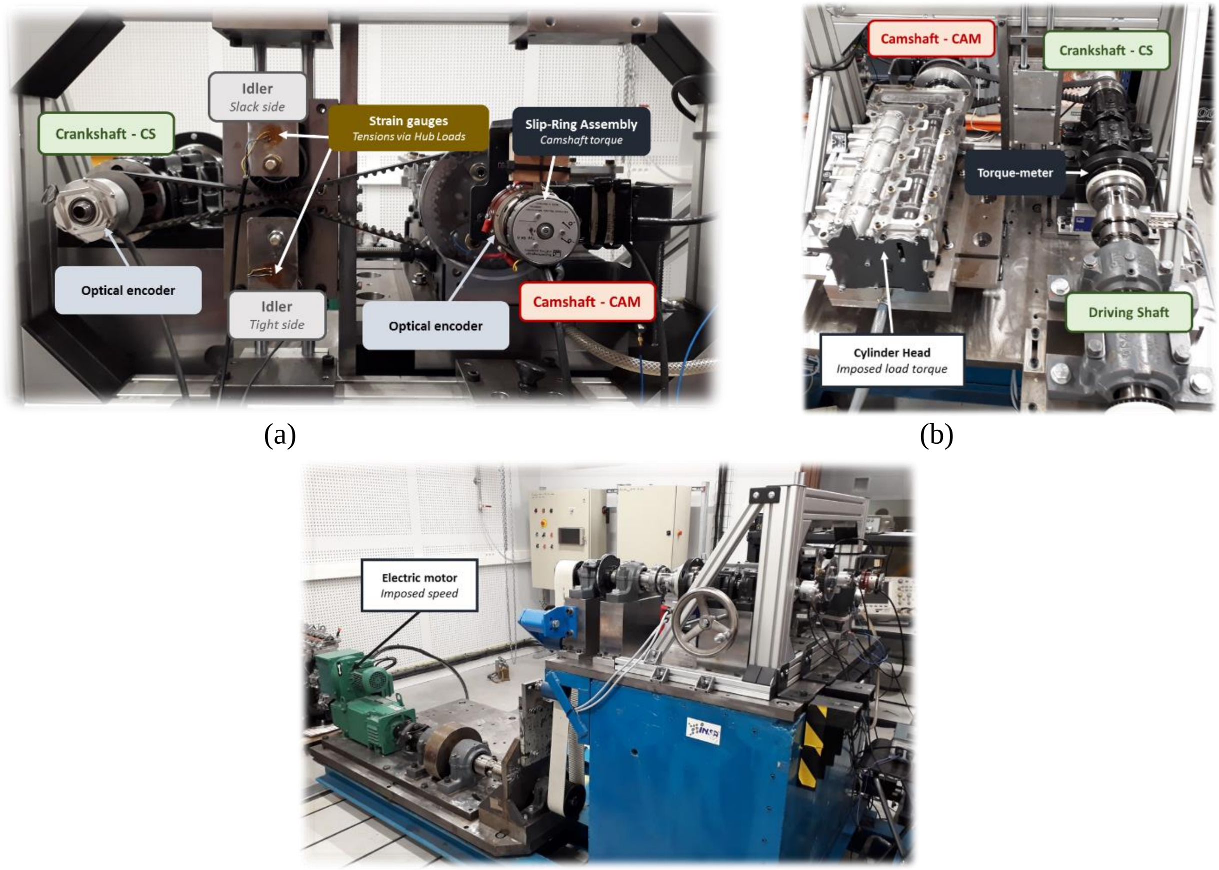


Figure 3: Scheme of the global test stand architecture



(a) (b)

(c)

Figure 4: Face (a), top (b) and global (c) views of the test stand

2.3 Measurement devices

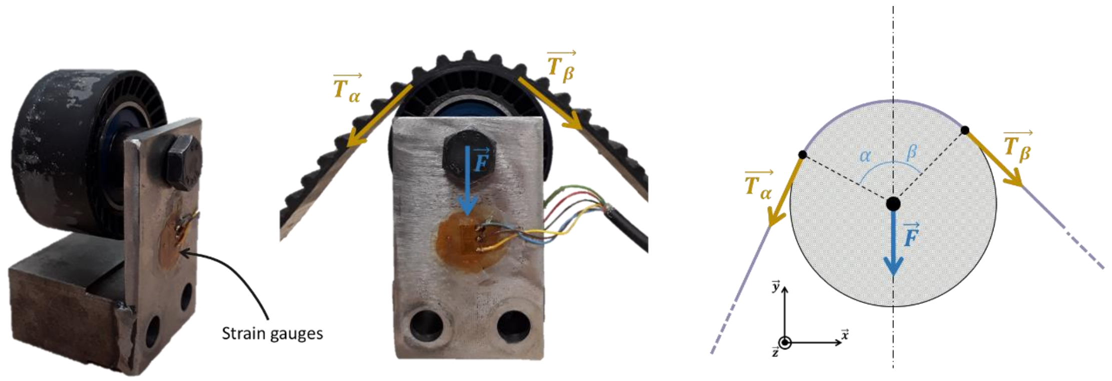
The measurement system is very similar to that used for the works presented in [8]. The system comprises usual devices employed for experimental investigations on gear transmission error and belt transmissions [9,10,11]. It enables measuring all the quantities depicting the rotational dynamics of the transmission.

Optical encoders with a resolution of 2500 pulses per revolution are mounted on the driving and driven axes for measuring their rotation angles and speeds. Torques transmitted by the driving and the driven axes are monitored. On the driving axis, a torque meter with a range of 200 N.m is placed before the driving pulley. On the driven axis, the camshaft is equipped with strain gauges for measuring the resistive load torque. The electrical connection for gauges is provided by means of a slip-ring assembly. The supporting parts used for idler pulleys are equipped with strain gauges so that to measure y-axis component of the hub load applied onto the idler axis. The belt span tension forces can be deduced from the hub load as illustrated in figure 5. On both sides of an idler, belt tensions can be considered equal ($T_\alpha = T_\beta = T$) and then:

$$T = \frac{F}{\sin\alpha + \sin\beta} \tag{2}$$

where F is the y-axis component of the hub load, T is the tension force in both belt spans and α and β are the respective orientation angles of the spans.

The data acquisition system is custom made within a N. I. PXI frame including counter boards for the optical encoders (pulse timing method [9]) and classical data acquisition boards for the other sensors.



(a) (b)

Figure 5: Idler supporting part used to measure the span tensions (a) and principle diagram (b)

3 Experimental analysis

The result analyses are performed by observing the following quantities in the angular and the angular frequency domains:

- Instantaneous angular speed variations for the camshaft pulley ($\Delta\omega_{CAM}$).
- Tension force variations in the tight and the slack sides (ΔT_T and ΔT_t)

These quantities are first studied for a reference case for which the crankshaft pulley is circular (section 3.1). Then, the transmission equipped with the oval crankshaft pulley is considered (section 3.2). A comparison study is conducted showing how the oval pulley affects the levels of angular vibrations and tension force fluctuations depending on its phasing angle and the driving speed.

3.1 Reference case: circular crankshaft pulley

The fluctuations and the frequency content of the camshaft angular speed are shown in figure 6. For all driving speeds, the variations are periodic and dominated by a 2nd order harmonic due to the dominant 2nd order harmonic governing the load torque. The amplitude of the H2 harmonic fluctuation is about 4.5 rpm when the driving speed is 600 rpm and rises to approximately 15 rpm when the crankshaft speed reaches 1500 and 3000 rpm. One can see also that the frequency spectrum contains secondary harmonics mainly due to the other camshaft torque harmonics (even orders and 9.5th order for ω_s = 3000 rpm) but also to some disturbing excitations such as resonances of the test stand and 0.5th and 1st order modulations due to misalignment and geometry faults on the crankshaft and camshaft pulleys.

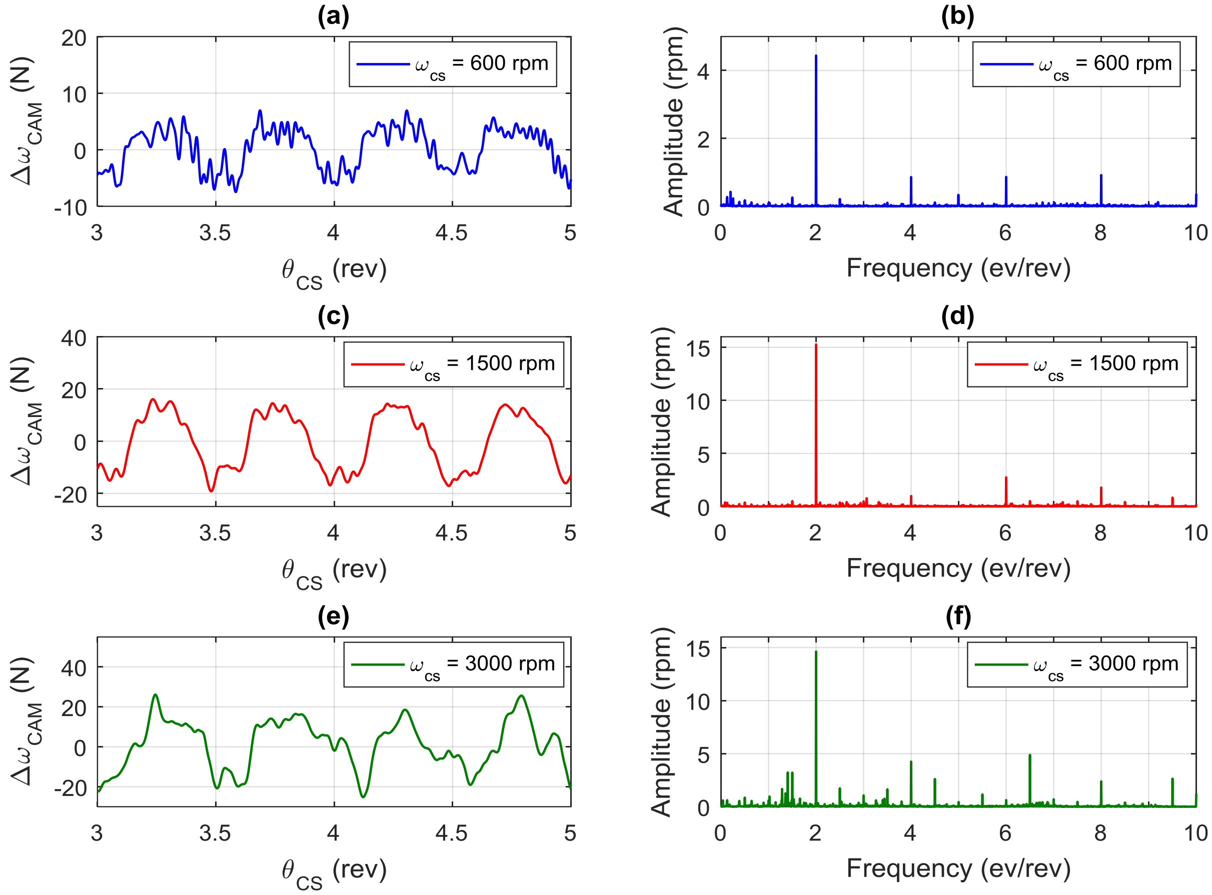


Figure 6: Camshaft angular speed variations in the angular (a, c, e) and angular frequency (b, d, f) domains

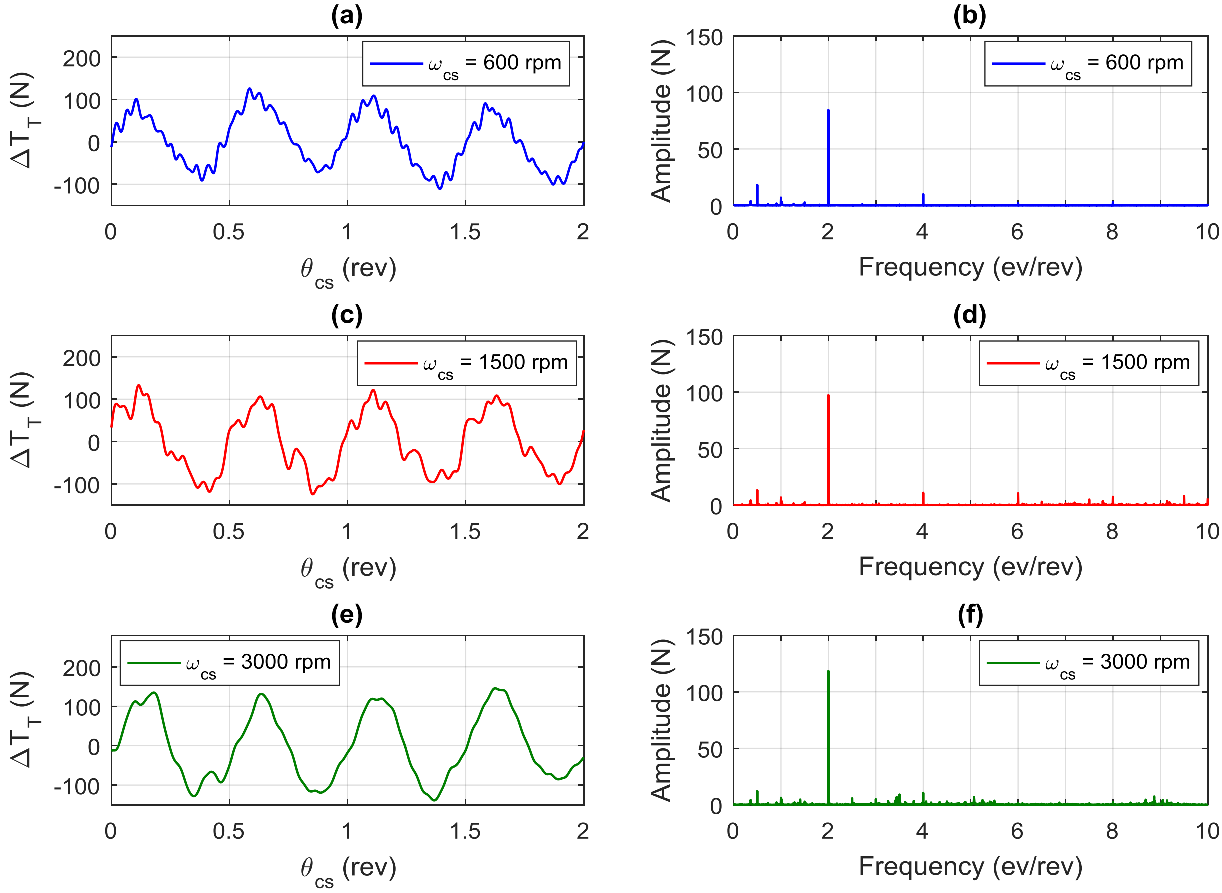


Figure 7: Tight span tension variations in the angular (a, c, e) and angular frequency (b, d, f) domains

Figure 7 and 8 show the variations and the frequency content of the belt tensions in the tight and slack spans of the transmission. The amplitude of the dominant H2 harmonic slightly increases with the driving speed. The H2 amplitude is of course always higher in the tight span than in the slack span. Moreover, one can note that the tension fluctuations in the tight and slack spans are out-of-phase, which is another common result for belt transmission subjected to fluctuating load torque.

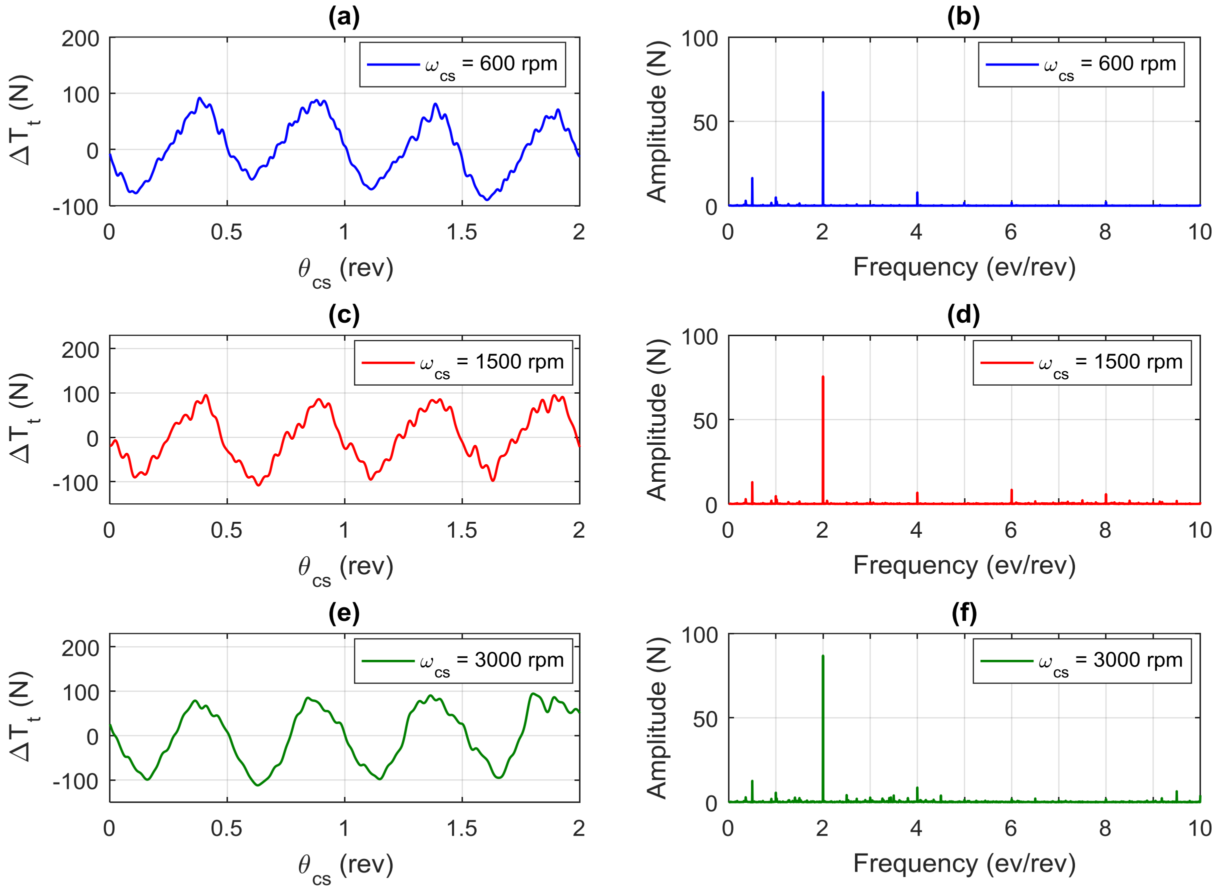


Figure 8: Slack span tension variations in the angular (a, c, e) and angular frequency (b, d, f) domains

3.2 Oval crankshaft pulley: study of the corrective effect

Results obtained for the reference case (section 3.1) have shown that the variations of the camshaft speed and the belt span tensions caused by the camshaft load torque are strongly dominated by a 2^{nd} order harmonic (H2) for all the driving speeds. Thus, in the following, the analyses focus only on the amplitude of H2 harmonics. A parametric study has been performed in order to show how the oval pulley phasing angle impacts the H2 amplitude. Considering that an oval pulley is symmetric along its major axis, the experiments are only performed for a phasing angle ranging between 0 and 180°.

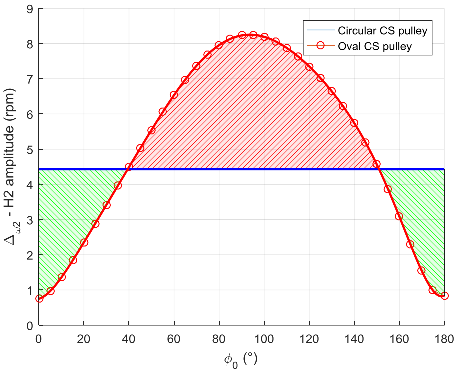
Figures 9, 10 and 11 show respectively the results obtained for the three driving speeds 600 ; 1500 and 3000 rpm. Each figure comprises three graphs on which the amplitudes of the H2 harmonics that govern the camshaft speed and tension forces in the tight and slack spans are respectively plotted versus the phasing angle of the crankshaft oval pulley. On these graphs, the H2 amplitude evolution with the phasing angle is represented with a red solid line marked with circles. Horizontal blue solid lines represent the amplitude of the H2 harmonics in the reference case (circular crankshaft pulley). Green and red hatched areas correspond to phasing angle ranges for which the oval pulley respectively involves a reduction or an increase of the H2 harmonic.

When the crankshaft runs at 600 rpm (figure 9), a speed fluctuation reduction can be obtained for a phasing angle range [0,40°]∪[150,180°] with a maximum reduction of nearly 82% for a phasing angle around 0°. In contrast, for a phasing angle in the range]40,150°[the amplitude is higher than in the reference case with a critical increase of 75% for a phasing angle equal to 90°. H2 amplitude of the tension forces in tight and slack sides are significantly reduced for non-overlapping phasing angle ranges, respectively [97,177°] and [0,82°]. Amplitudes of tension force variations are less impacted by the phasing angle of the oval pulley. For both spans, the maximum reduction and increase ratios are the same, respectively about 24 and 26%.

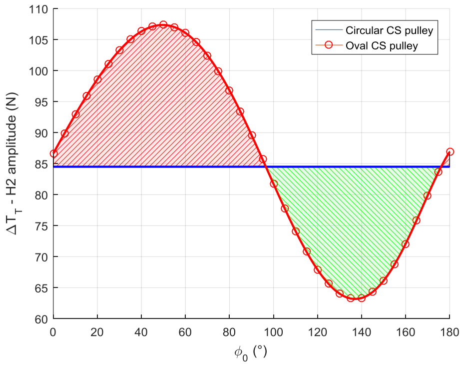
When the crankshaft runs at 1500 rpm (figure 10), the phasing angle ranges corresponding to an increase and a reduction of fluctuations are quite the same as for a driving speed equal to 600 rpm. The reduction and increase ratios are of the same order of magnitude and are obtained for very similar optimal and critical phasing angle values.

For a driving speed of 3000 rpm (figure 11), the angle ranges and the corresponding reduction and increase ratios differ. The speed variations are reduced when the phasing angle belongs to the range [0,30°]∪[170,180°]. The optimal and critical phasing angles are respectively 10° and 95° with a maximum reduction and increase ratios of 60% and 166% respectively. The tension force variations in the tight span are reduced for a phasing angle in the range [0,22°]∪[137°,180°]. The maximum reduction is about 50% when the phasing angle is around 10° and there is a maximum increase of nearly 67% for an angle of 95°. The tension force fluctuations induced in the slack span are reduced for a phasing angle comprised in the range [0,45°]∪[158,180°] with a minimum amplitude (-30%) and a maximum amplitude (+39%) for phasing angles of 10° and 105° respectively. It can be seen that if the phasing angle belongs to the range [0,22°]∪[170°,180°], the fluctuations are reduced for the camshaft speed and the tension forces in both spans simultaneously. In particular, for a phasing angle equal to 10° all the fluctuations are significantly reduced.

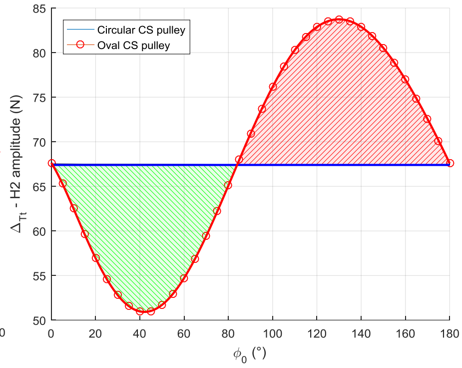
In the light of these results, one can finally note that it is possible to reach a compromise with a phasing angle around 0°. Whatever the driving speed, such a value ensures a strong reduction of the camshaft speed variations and limits the tensions variations in both spans.



(a)

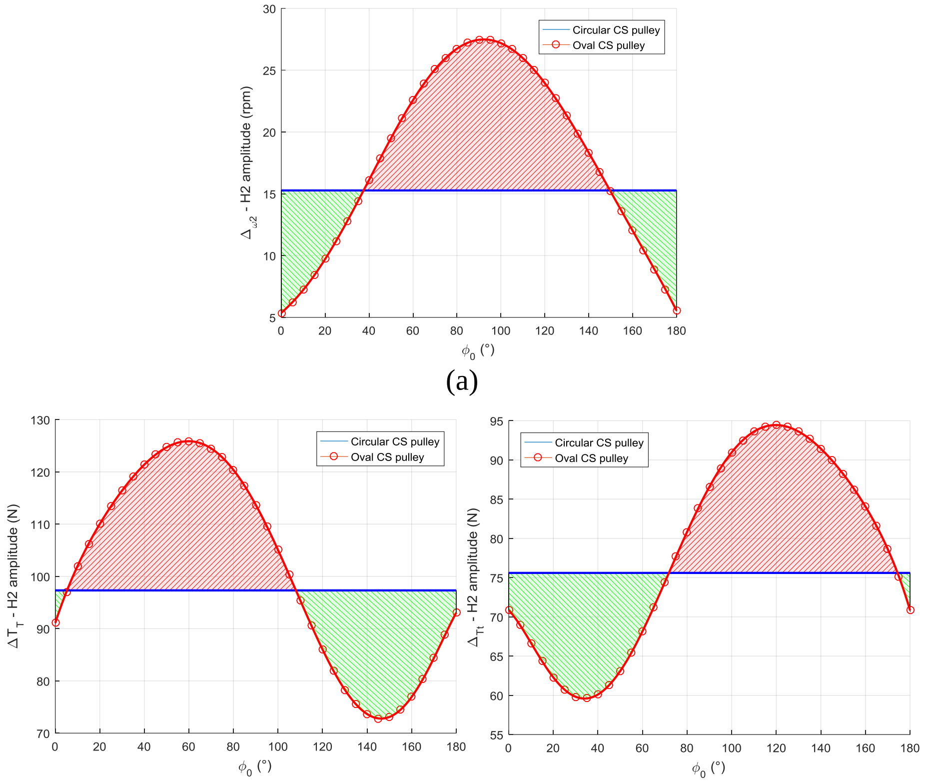


(b)



(c)

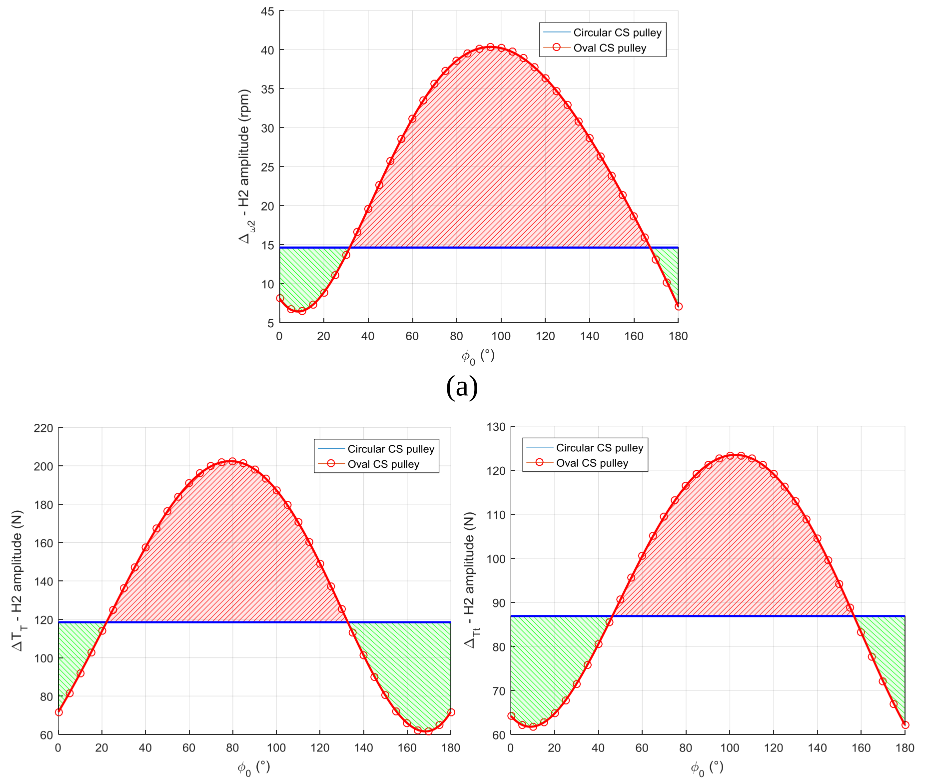
Figure 9: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at <u>600 rpm</u>



(a)

(b) (c)

Figure 10: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at <u>1500 rpm</u>



(a)

(b) (c)

Figure 11: H2 harmonic amplitude of the camshaft speed (a) and tension forces in the tight (b) and slack (c) spans when the crankshaft runs at <u>3000 rpm</u>

4 Conclusion

The experimental results presented in this paper demonstrate to what extent, for a well-chosen phasing angle, an oval pulley has a corrective effect enabling a reduction of the H2 fluctuations affecting the camshaft angular speed and the span tension forces. These results also show how much, for wrong phasing angles, the pulley can degrade the dynamic behaviour of the transmission inducing a strong increase of speed and tension fluctuations. This is why transmissions comprising NC pulleys must be designed carefully.

In addition, one can note that the impact of an oval pulley depends on the driving speed, which make the design of the transmissions more difficult. As already discussed in previous works [3,4], it is not so easy to reduce significantly and simultaneously speed and tension force fluctuations in belt spans using a NC pulley only.

For completing the present work, it could be meaningful to extend the researches to more complex systems. Thus, future works could involve transmissions subjected to other excitation sources such as acyclism and/or transmissions comprising a dynamic tensioner. It could be interesting as well to study the case of a NC pulley having a different profile shape adapted to the correction of harmonics of other order. Also, the effect of NC pulleys on TBD acoustic radiation still remains to investigate (impact on belt meshing noise, span transverse vibrations, …).

References

[1] GAJEWSKI, W., *Synchronous drive apparatus and methods,* U.S. Patent No 7.044.875, Litens Automotive, 2006.

[2] ZHU, H., HU, Y., ZHU, W. D., *Dynamic response of a front end accessory drive system and parameter optimization for vibration reduction via genetic algorithm*, Journal of Vibration and Control, Volume 24, Issue 11, pp. 2201-2220, 2018.

[3] PASSOS, S., MANIN, L., SAUVAGE, O., ROTA, L., DELATTRE, B., REMOND, D., *Investigation on the timing belt drive with a non-circular pulley*, Conference Proceedings - International Gear Conference, Lyon – Villeurbanne, France, Volume 2, pp. 1186-1198, 2018.

[4] PASSOS, S., MANIN, L., SAUVAGE, O., ROTA, L., REMOND, D., *Rotational dynamics of timing belt drives comprising non-circular pulleys,* SIA International Conference and Exhibition on Simulation, Paris – Saclay, France, 2019.

[5] ZHENG, E., JIA, F., SHA, H., WANG, S., *Non-circular belt transmission design of mechanical press,* Mechanism and Machine Theory, Volume 57, pp. 126-138, 2012.

[6] ENDO, G., YAMADA, H., YAJIMA, A., OGATA, M., HIROSE, S., *A passive weight compensation mechanism with a non-circular pulley and a spring,* IEEE International Conference on Robotics and Automation (ICRA), Anchorage, AK, pp. 3843-3848, 2010.

[7] HWANG, S. J., PERKINS, N. C., ULSOY, A. G., MECKSTROTH, R. J., *Rotational response and slip prediction of serpentine belt drive systems*, J. Vib. Acoust. , Volume 116, Issue 1, pp. 71-78, 1994.

[8] PASSOS S., MANIN, L., REMOND, D, SAUVAGE, O., ROTA, L., BESNIER, E., *Experimental investigation on timing belt drives dynamics with non-circular pulley,* ISMA International Conference on Noise and Vibration, Leuven, Belgium, 2018.

[9] REMOND, D., *Practical performances of high-speed measurement of gear transmission error or torsional vibrations with optical encoders,* Measurement Science and Technology, IOP Publishing, pp. 347-353, 1998.

[10] MAKITA, K., KAGOTANI, M., UEDA, H., KOYAMA, T., *Influence of idler on error transmission in synchronous belt drives (under transmission force),* ASME J. Mech. Des., Volume 125, Issue 2, pp. 404-410, 2003.

[11] MANIN, L., MICHON, G., REMOND, D., DUFOUR, R., *From transmission error measurement to pulley-belt slip determination in serpentine belt drives: influence of tensioner and belt characteristics*, Mechanism and Machine Theory, Volume 44, Issue 4, pp. 813-821, 2009.

Robust optimization of nonlinear energy sinks used for dynamic instabilities mitigation of an uncertain friction system

Cherif SNOUN, Baptiste BERGEOT, Sébastien BERGER

INSA Centre Val de Loire, Université d'Orléans,Université de Tours, LaMé EA 7494,
3 Rue de la Chocolaterie, CS 23410, 41034 Blois Cedex, France

Abstract

In this paper, robust optimization tool is proposed for nonlinear energy sinks used for the mitigation of friction-induced vibrations due to mode coupling instability in braking systems. The study is based on a mechanical system which is composed of two NESs coupled to the well-known two-degrees-of-freedom Hultèn's model. In such an unstable system coupled with NES, it is usual to observe a discontinuity in the steady-state amplitude profiles which separates the parameters space into two parts which contain mitigated and unmitigated regimes respectively. We developed a methodology based on Multi-Element generalized Polynomial Chaos to identify this discontinuity which allows us to determine the Propensity of the system to undergo a Harmless Steady-State Regime (PHSSR). The objective of this work is, therefore, to maximize the value of PHSSR to obtain an optimal design of the NESs. For that, several stochastic optimization problems are presented taking into account the dispersion of the uncertain parameters.

Keywords: Brake squeal noise, Friction-induced vibrations, Nonlinear energy Sink, Uncertainties, Robust optimization

1 Introduction

Dynamic instability also named friction-induced vibrations is one of the important problem that the friction system can confront. This phenomenon can appear by the generation of Limit Cycle Oscillation (LCO) induced by dry friction. They are explained, in most studies of self-excited systems, by the coupling of the tangential and normal modes [1, 2, 3]. As a way to model these dynamic instabilities related to friction, the well-known two-degrees-of-freedom Hultèn's model [4, 5] has been widely used and it is also considered in this paper.

The NES is a nonlinear spring mass damper with a strong cubic stiffness. It can adapt itself to the main system that it is attached to without being tuned to a specific frequency. Its operation is based on the concept of Targeted Energy Transfer (TET) which has become an important passive control technique for reducing or eliminating unwanted vibrations [6, 7].

In this paper, two ungrounded NESs are used in order to mitigate or eliminate mode coupling instability in the Hultèn's model. In [8], the authors classified the steady-state response regimes in two main regimes related to the dispersion of some uncertain parameters: the first are the mitigated regimes and the second are the unmitigated regimes. As usual in the context instability mitigation by means of NESs, the LCO amplitude profile presents a discontinuity between these regimes which makes the NES potentially very sensitive to uncertainty. For that, with a low computational cost, the aim is to maximize the performance of the NES (i.e. minimize the region parameter of unmitigated regimes), by optimizing its design parameters taking into account uncertain parameters.

NES has been extensively optimized in a deterministic context (see e.g. [9]), but there is very little work performing optimization of NES under uncertainties in general [10] and for mitigation of self-excited vibrations in particular [11]. In the latter, the authors have developed a methodology of optimization under uncertainties of a NES attached to a two-degree-of-freedom airfoil. In the present work, a robust optimization tool based on the Multi-Element generalized Polynomial Chaos (ME-gPC) is developed.

The article is constructed as follows: in Section 2, the two degrees-of-freedom Hultèn's model coupled to two NESs is presented. In Section 3, the steady-state of the system as well as the discontinued profile are introduced. The optimization formulation under uncertainties is formulated in Section 4. Section 5 describes

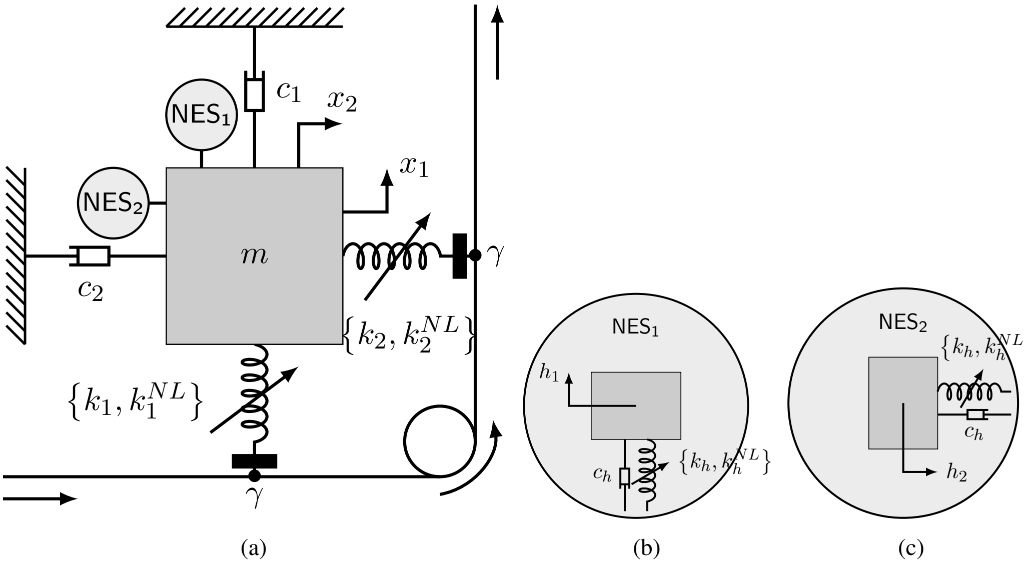


Figure 1: (a) Mechanical system with NESs. (b) Zoom on the NES$_1$. (c) Zoom on the NES$_2$.

the polynomial chaos theory and the used stochastic optimization algorithm. The results of the optimization methods are presented in Section 6. Finally, conclusions are given in Section 7.

2 The mechanical system

The mechanical system considered in this work is composed by the two degrees-of-freedom (DOF) Hultèn's model [4, 5], which represents the primary system, coupled to two identical NESs with masses m_h, damping coefficients c_h and cubic stiffnesses k_h^{NL}. The NESs are attached on the primary system in an ungrounded configuration as shown in Fig. 1.

The equations which describe the mechanical system are given by :

$$\frac{d^2x_1}{dt^2} + \eta_1\omega_1\frac{dx_1}{dt} + \omega_1^2 x_1 - \gamma\omega_2^2 x_2 + \varphi_1 x_1^3 - \gamma\varphi_2 x_2^3 + \mu\omega_1\left(\frac{dx_1}{dt} - \frac{dh_1}{dt}\right) + \xi_h(x_1 - h_1) + \varphi_h(x_1 - h_1)^3 = 0 \tag{1a}$$

$$\varepsilon\frac{d^2h_1}{dt^2} + \mu\omega_1\left(\frac{dh_1}{dt} - \frac{dx_1}{dt}\right) + \xi_h(h_1 - x_1) + \varphi_h(h_1 - x_1)^3 = 0 \tag{1b}$$

$$\frac{d^2x_2}{dt^2} + \eta_2\omega_2\frac{dx_2}{dt} + \omega_2^2 x_2 + \gamma\omega_1^2 x_1 + \gamma\varphi_1 x_1^3 + \varphi_2 x_2^3 + \mu\omega_1\left(\frac{dx_2}{dt} - \frac{dh_2}{dt}\right) + \xi_h(x_2 - h_2)^+ \varphi_h(x_2 - h_2)^3 = 0 \tag{1c}$$

$$\varepsilon\frac{d^2h_2}{dt^2} + \mu\omega_1\left(\frac{dh_2}{dt} - \frac{dx_2}{dt}\right) + \xi_h(h_2 - x_2) + \varphi_h(h_2 - x_2)^3 = 0, \tag{1d}$$

where $h_1(t)$ and $h_2(t)$ (respectively $x_1(t)$ and $x_2(t)$) represent the NESs displacements (respectively the displacements of the primary system), $\eta_i = c_i/\sqrt{mk_i}$, $\omega_i = \sqrt{k_i/m}$, $\varphi_i = k_i^{NL}/m$ (with $i = 1,2$), $\varepsilon = m_h/m$ assuming $0.01 < \varepsilon < 0.1$, $\xi_h = k_h/m$, $\mu = c_h/\sqrt{mk_1}$ and $\varphi_h = k_h^{NL}/m$.

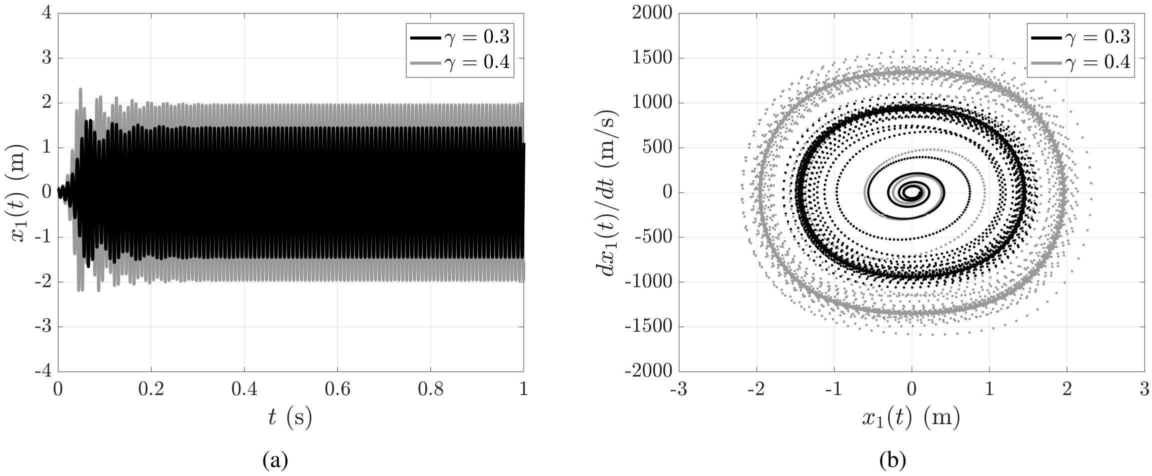


Figure 2: Numerical integration in the direction x_1 of the Hult'en's model without NESs for two values of the friction coefficient: (a) Times series; (b) Representation in the phase plane.

3 Preliminary results

3.1 Vibratory levels and possible steady-state regimes

A example of the numerical simulation of the system without NESs is plotted in Fig. 2 for two different values of the friction coefficient: $\gamma = 0.3$ and $\gamma = 0.4$. The other parameters are

$$\begin{array}{cccc} & \omega_1 = 2\pi 100, \quad \omega_2 = 2\pi 85, & & \\ \eta_1 = 0.02, & \eta_2 = 0.06, & \varphi_1 = 10^5, & \varphi_2 = 0, \\ \varepsilon = 0.05, & \xi_h = 0.001, & \mu = 0.02, & \varphi_h = 1.4 \cdot 10^5. \end{array} \tag{2}$$

Moreover, the initial conditions are small perturbation of the trivial equilibrium position: $x_1(0) = x_2(0) = 0$, $\dot{x}_1(0) = \dot{x}_2(0) = 10$. When $\gamma = 0.4$, Limit Cycle Oscillations (LCO) are observed. We focus our analysis on the capacity of the NESs to suppress or mitigate these LCOs.

Four main types of steady-state regimes can be generated when two NESs is attached on the: complete suppression of the instability, mitigation through Periodic Response (PR), mitigation through Strongly Modulated Response (SMR) or no mitigation.

An illustration of these regimes is shown in Fig. 3 which plots the displacements $x_1(t)$ with respect to times with and without the NESs. Hereafter, mitigated regimes referred to complete suppression, PR and SMR.

3.2 The objective function

We define the amplitude A_1^{wNES} of the variables x_1 of the coupled system Eq. (1) and within a steady-state regime as

$$A_1^{wNES} = \frac{\max\left[x_1^{SSR}(t)\right] - \min\left[x_1^{SSR}(t)\right]}{2}, \tag{3}$$

where $x_1^{SSR}(t)$ is the times series of the variables x_1 obtained from the numerical integration of the coupled system Eq. (1) within the steady-state regime[1].

The amplitudes A_1^{wNES} and A_1^{woNES} (i.e. the amplitude of the system without NESs) are plotted as a function of the friction coefficient γ in Fig. 4 for the set of parameters Eq. (2).

The figure highlights a jump (or discontinuity) in the amplitude profile A_1^{wNES}. This discontinuity corresponds, when γ increases, to the transition from SMR to no suppression regimes and separates mitigated regimes and unmitigated regimes.The value of γ at the jump is called mitigation limit (with respect to γ) and denoted γ_{ml}.

[1]The calculation was performed on 4 seconds in order to be sure that the steady-state has been reached.

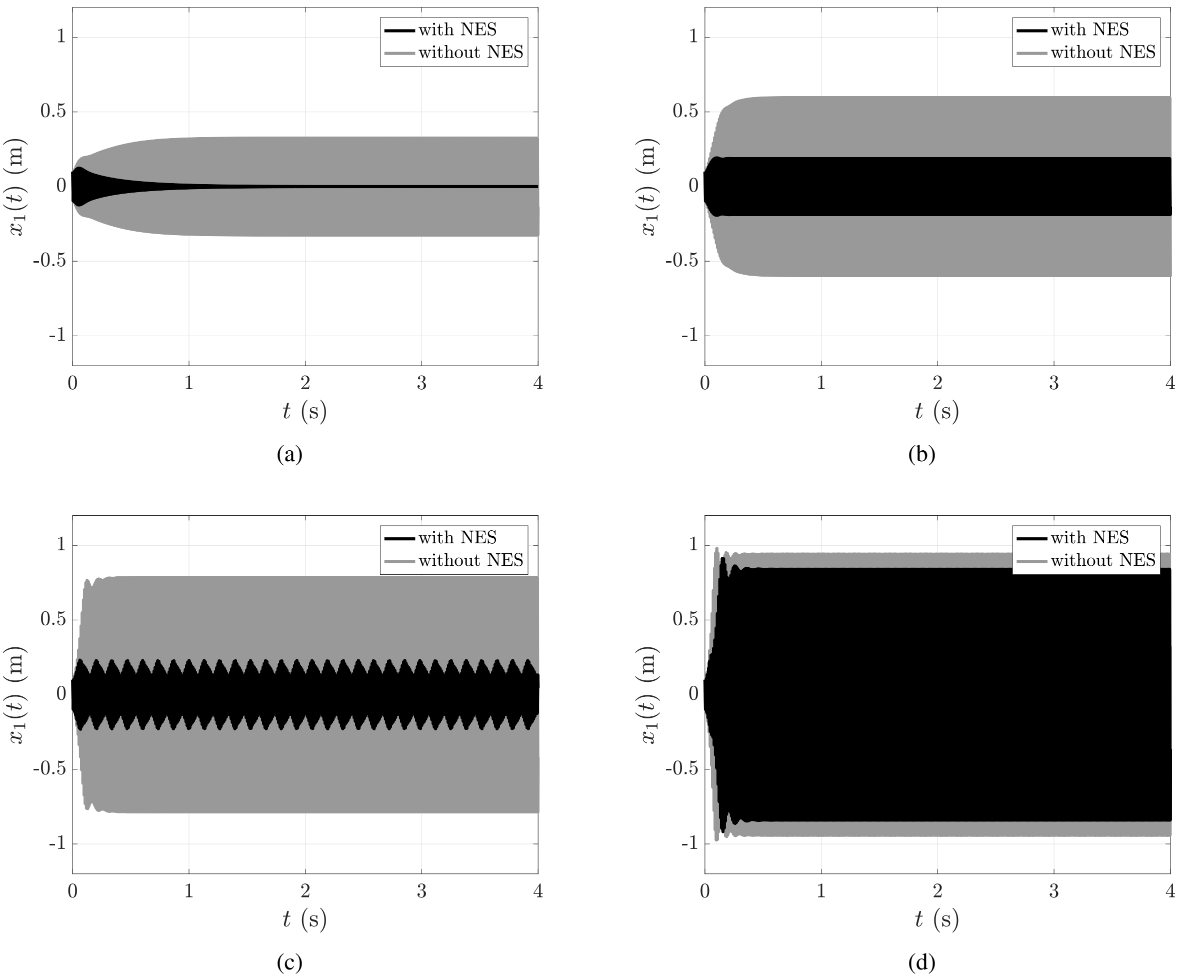


Figure 3: Comparison between time serie $x_1(t)$ resulting from the numerical integration of the braking system with and without NESs. (a) Complete suppression, $\gamma = 0.16$; (b) Mitigation: PR, $\gamma = 0.18$; (c) Mitigation: SMR, $\gamma = 0.2$; (d) No suppression, $\gamma = 0.22$. The set of parameters Eq. (2) is used.

Then the Propensity of the system to undergo a Harmless Steady-State Regime (PHSSR) of the oscillation by the NESs is estimated. We compute a set of S_{total} samples of the uncertain parameters within the considered uncertain space and following their distribution law. Then, the PHSSR is defined as follows

$$\text{PHSSR} = \frac{S_{\text{HSSR}}}{S_{\text{total}}} \times 100 \tag{4}$$

where S_{HSSR} is the number of samples within the region of uncertain parameters space in which the LCO of the system is mitigated or the system is stable.

The PHSSR represents the objective function of the study.

4 Optimization Formulation under uncertainties

The sensitivity of the friction coefficient is such that the steady state of the mechanical system is discontinuous and presents a jump. This jump induces areas in which the efficiency of the NES is either high or low. In order to obtain a robust design of NES which is insensitive to the dispersion of the uncertain parameters, a stochastic optimization problem is considred as follow:

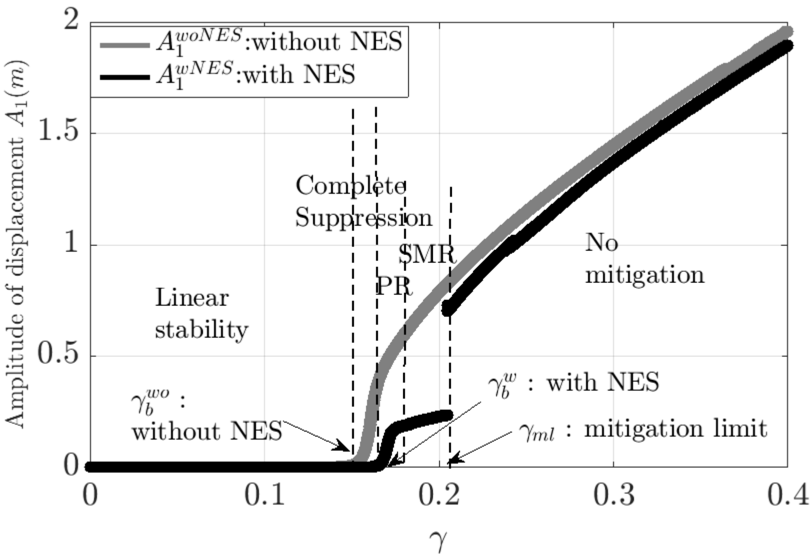


Figure 4: Amplitudes A_1^{wNES} and A_1^{woNES} as a function of the friction coefficient γ.

$$\begin{aligned} &\text{Maximize} \quad && PHSSR\ (A_1^{wNES}(x_d, x_u)) \\ &\text{Subject to} \quad && x_d^{(min)} \leqslant x_d \leqslant x_d^{(max)} \end{aligned} \tag{5a}$$

(5b)

where

- x_d represents the design variables of the NES with lower bounds $x_d^{(min)}$ and upper bounds $x_d^{(max)}$.
- x_u represent the uncertain parameters of the primary system.

5 Optimization Algorithm

In order to solve an optimization problem under uncertainties, the determination of the LCO of the mechanical system need to use a stochastic method to take into account the uncertain parameters. In this work, the Multi-Element generalized Polynomial Chaos is used as an alternative approach to reduce the computational cost of the traditional methods (Monte Carlo or deterministic simulations).

5.1 Polynomial Chaos theory

5.1.1 generalized Polynomial Chaos

The gPC theory [12, 13] allows to express a random process $X(\xi)$ called also the Quantity of Interest (QoI) with a truncated orthogonal polynomial function series such as

$$X(\xi) \approx \sum_{j=0}^{N_p} \bar{x}_j \phi_j(\xi), \tag{6}$$

where:

- $\bar{x}_j$ are the gPC coefficients of the stochastic process $X(\xi)$,
- $\phi_j(\xi)$ are orthogonal polynomial functions,
- $N_p = \frac{(p+r)!}{p!r!} - 1$ where p is the order of the gPC and r is the number of the uncertain parameters,
- $\xi(\xi_1, ..., \xi_r)$ is a vector of r independent random variables within$[-1, 1]^r$.

5.1.2 Multi-Element Generalized Polynomial Chaos (ME-gPC)

The ME-gPC consists to split ξ into a collection of m non-intersecting elements and to approximate the stochastic process $X(\xi)$ using the gPC in each element [14]. It is given by

$$X(\xi) \approx \sum_{k=1}^{m} X_k(\bar{\xi}^k) J_k, \tag{7}$$

where

- $X_k(\bar{\xi}^k) = \sum_{j=1}^{N_p} \bar{x}_{k,j} \phi_j(\bar{\xi}^k)$ where $\bar{\xi}^k$ is an uniform random variables corresponding to the k^{th} element,
- J_k is the element size.

We denote by $\sigma^2_{p,k}$ the variance of the QoI estimated in the k^{th} element directly from the gPC coefficients as:

$$\sigma^2_{p,k} = \frac{1}{2^r} \sum_{j=1}^{N_p} \bar{x}^2_{k,j} \langle \phi_j^2 \rangle. \tag{8}$$

5.2 Proposed algorithm

The ME-gPC metamodel would not be efficient in its initial configuration due to the presence of discontinuities. In fact, the goal is not to obtain an accurate representation of the response of the system, but to be able to locate the discontinuity. Figure 5 shows the different steps of the ME-gPC based method to detect the jump. The method consists to split the stochastic parameter space into two equal non-intersecting elements and to evaluate in each element the amplitude of displacement of the system. Then, the verification of the presence or not of discontinuity is addressed according to the comparison of the variance $\sigma^2_{p,k}$ to a chosen threshold θ_1. The process stops either if the element size reaches a threshold

$$J_{kmin} = \theta_2 \, J_{k0}, \tag{9}$$

where J_{k0} represents the size of the initial element or if the number N' of numerical simulations required with the ME-gPC based method is smaller than the typical number N of simulations required with the reference method. The mitigation limit is, therefore approximated by the upper bound of the last found element and then the robust evaluation of the PHSSR is calculated.

6 Results

In this work, the friction coefficient γ is considred as the uncertain parameter (x_u) and the damping rate of the NES μ as the design variables (x_d).

6.1 Reference study

In this section, the reference is evaluated with the deterministic model (1). In the first case, 50 deterministic samples of the damping rate of the NES ($\mu \in [0, 0.04]$) are used. For each sample of μ, 1 000 samples of the friction coefficient ($\gamma \in [0, 0.4]$) are used in order to determine the mitigation limit $\gamma_m l$. The number of simulations is therefore equal to 50 000 (50 $\times$ 1 000). Fig. 6(a) shows the variation of the mitigation limit as a function of μ. Because of the presence of a plateau, we realize that the number of samples of γ is not sufficient to determine the optimal design value of μ. For that, 10 added deterministic samples of the damping rate of the NES ($\mu \in [0.02, 0.03]$) with now 10 000 samples of the friction coefficient γ to determine the mitigation limit. The total number of simulations required is now equal to 150 000 (50 000 + 10 $\times$ 10 000). Fig. 6(b) shows the new variation of the mitigation limit as a function of μ. In this case, the maximum (i.e. optimal) value of γ_{ml} is 0.2004 which gives a maximum value of PHSSR equals to 50.1%. The corresponding optimal design value for the damping rate is $\mu^{\text{Ref}}_{\text{opt}} = 0.024$.

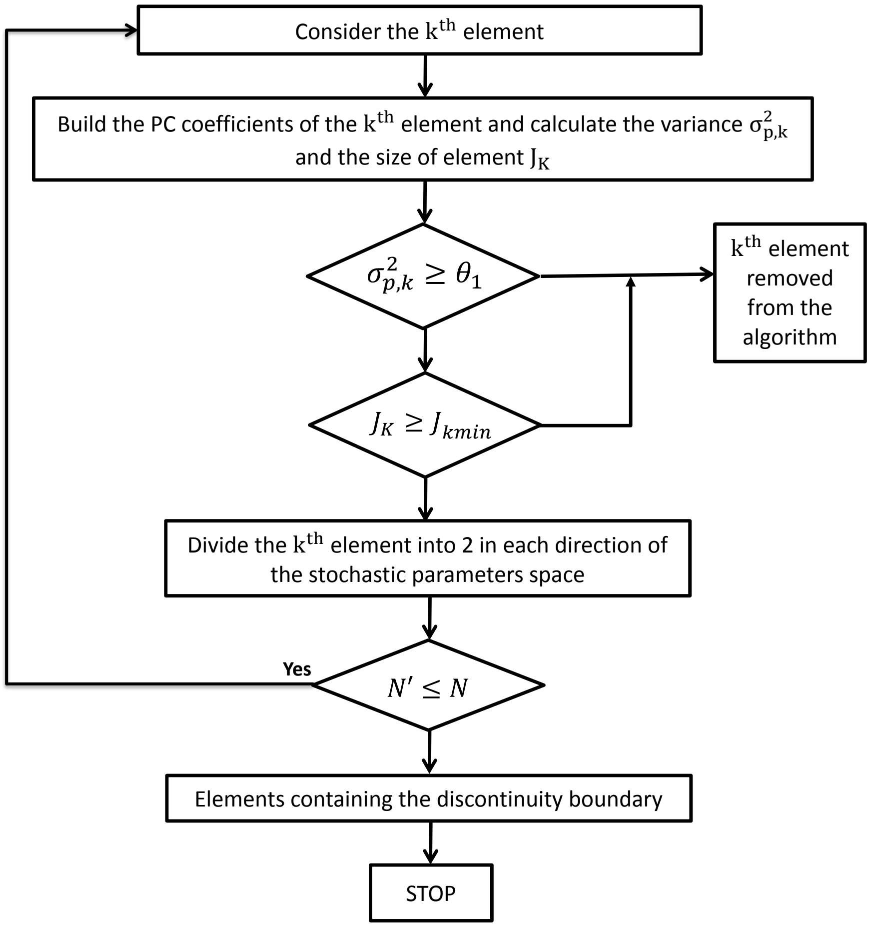


Figure 5: Algorithm of the ME-gPC based method to identify the mitigation limit.

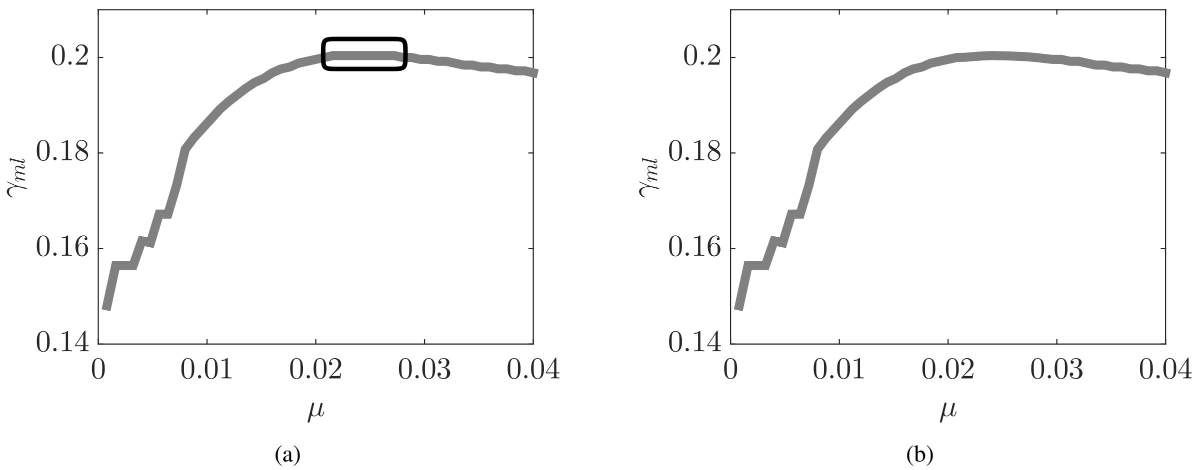


Figure 6: Variation of the mitigation limit as a function of the damping ratio of the NES (μ) by the deterministic method. (a) 50 samples of $\mu \in [0, 0.04]$; (b) 10 added samples of $\mu \in [0.02, 0.03]$.

6.2 Optimization using the ME-gPC based method

In this section, the algorithm presented in Fig. 5 is applied to estimate A_1^{wNES} (the considered QoI) and then the mitigation limit and the PHSSR. The gPC order is $p = 1$ and the threshold $\theta_1 = 2.10^{-3}\%$. Fig. 7 shows an example, with a given value of μ, of the algorithm described in Section 5.2. As described above, the mitigation limit is approximated by the upper bound of the last found element. Now, 50 deterministic samples of the damping rate of the NES ($\mu \in [0, 0.04]$) are used. The samples of the friction coefficient γ are built

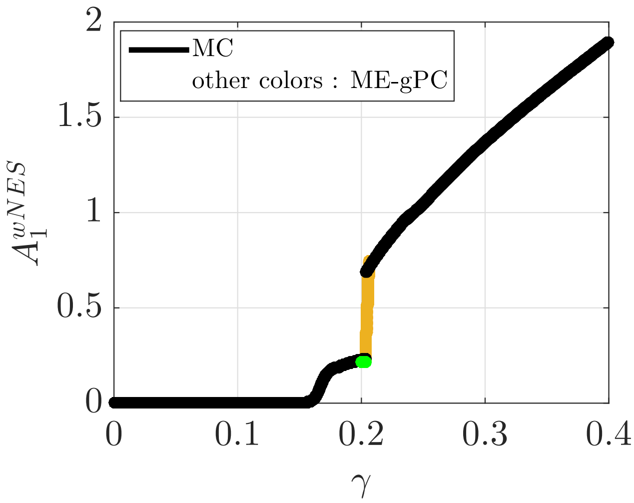


Figure 7: Last step of the ME-gPC based method to identify of the mitigation limit (γ_{ml}).

using the Latin Hypercube Samples (LHS) method. Fig. 8(a) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 1\%$. In this case, the number of simulations is equal to 6547. The optimal value of μ is in the interval $I_{\text{opt},1} = [0.016, 0.032]$. In order to reduce the size of I_{opt} the threshold θ_2 and the mitigation limit is determined within $I_{\text{opt},1}$. Fig. 8(b) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 0.1\%$. The added number of simulations is equal to 2766 and the optimal value of μ is in the interval $I_{\text{opt},2} = [0.019, 0.026]$. Fig. 8(c) shows the variation of the mitigation limit as a function of μ with $\theta_2 = 0.05\%$ used within $I_{\text{opt},2}$. In this case, the added number of simulations is equal to 1395 and the optimal value of μ is now in the interval $I_{\text{opt},2} = [0.0208, 0.0248] = 0.0228 \pm 8\%$. This interval has a midpoint equal to 0.0228 which gives a relative error equal to 5% compared to the reference. The total number of simulations is equal to 10 708 which represents 14 times less than the reference.

7 Conclusion

In this paper, the optimization of two NESs attached to the two degrees-of-freedom (DOF) Hultèn's model is presented. It accounts the discontinuity presented in the steady-state profiles using a stochastic optimization method. The results show the efficiency of the ME-gPC based method to detect the discontinuity and therefore the so-called mitigation limit. This allows us to provide an optimal design of the NESs with a low computational cost compare to that the deterministic optimization reference method.

References

[1] J.T. Oden and J.A.C. Martins. Models and computational methods for dynamic friction phenomena. *Computer Methods in Applied Mechanics and Engineering*, 52(1):527–634, 1985.

[2] Guillaume Fritz, Jean-Jacques Sinou, Jean-Marc Duffal, and Louis Jézéquel. Investigation of the relationship between damping and mode-coupling patterns in case of brake squeal. *Journal of Sound and Vibration*, 307(3):591–609, 2007.

[3] B. Hervé, J.-J. Sinou, H. Mahé, and L. Jézéquel. Analysis of squeal noise and mode coupling instabilities including damping and gyroscopic effects. *European Journal of Mechanics - A/Solids*, 27(2):141–160, mar 2008.

[4] J. Hultén. Friction phenomena related to drum brake squeal instabilities. In *ASME Design Engineering Technical Conferences, Sacramento, CA*, 1997.

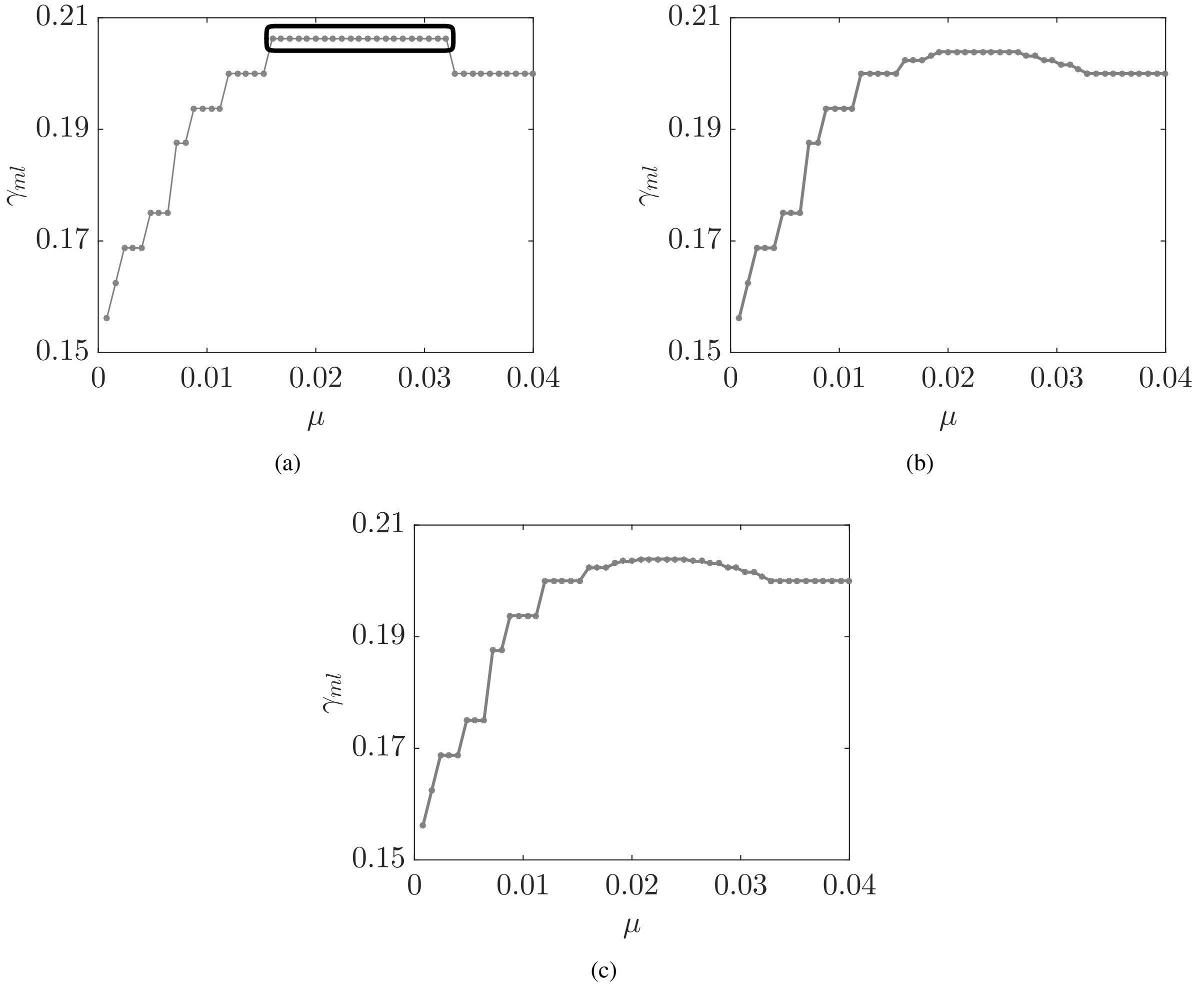


Figure 8: Variation of the mitigation limit as a function of the damping ratio of the NES using the ME-gPC based method. (a) $\theta_2 = 1\%$; (b) $\theta_2 = 0.1\%$; (c) $\theta_2 = 0.05\%$.

[5] J. Hultén. Brake squeal - a self-exciting mechanism with constant friction. In *SAE Truck and Bus Meeting, Detroit, Mi, USA*, 1993.

[6] A.F. Vakakis and O.V. Gendelman. Energy pumping in nonlinear mechanical oscillators: Part II - Resonance capture. *Journal of Applied Mechanics*, 68:42–48, 2001.

[7] A. F. Vakatis, O. V. Gendelman, L. A. Bergman, D. M. McFarland, G. Kerschen, and Y. S. Lee. *Nonlinear Targeted Energy Transfer in Mechanical and Structural Systems*. Springer-Verlag, Berlin, New York, 2008.

[8] B Bergeot, S Berger, and S Bellizzi. Mode coupling instability mitigation in friction systems by means of nonlinear energy sinks : numerical highlighting and local stability analysis. *Journal of Vibration and Control*, 24(15):3487–3511, 2017.

[9] Tuan Anh Nguyen and Stéphane Pernot. Design criteria for optimally tuned nonlinear energy sinks—part 1: transient regime. *Nonlinear Dynamics*, 69(1-2):1–19, nov 2011.

[10] Ethan Boroson, Samy Missoum, Pierre-Olivier Mattei, and Christophe Vergez. Optimization under uncertainty of parallel nonlinear energy sinks. *Journal of Sound and Vibration*, 394:451 – 464, 2017.

[11] Bharath Pidaparthi and Samy Missoum. Stochastic optimization of nonlinear energy sinks for the mitigation of limit cycle oscillations. *AIAA Journal*, pages 1–11, jan 2019.

[12] Roger G. Ghanem and Pol D. Spanos. Stochastic finite elements: A spectral approach. 1991.

[13] R.H. Cameron and W.T. Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, 48(2):385–392, 1947.

[14] Xiaoliang Wan and George Em Karniadakis. An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. *Journal of Computational Physics*, 209(2):617 – 642, 2005.

Energy exchange between a nonlinear absorber and a pendulum under parametric excitation

Gabriel HUREL[1], Alireza TURE SAVADKOOHI[1], Claude-Henri LAMARQUE[1]

[1]Univ. Lyon, ENTPE, LTDS UMR CNRS 5513, F-69518 Vaulx-en-Velin, France
gabriel.hurel@entpe.fr

Abstract

The studied system is a planar pendulum coupled with a nonlinear absorber and parametrically excited at its basis. The dynamical equations are treated with a multiple scale method. At fast time scale, a slow invariant manifold represents the asymptotic behavior. At slow time scale, the equilibrium points and their stability are investigated. Several phase portraits complete the analysis of the dynamical behavior of the system. Finally, numerical examples are given to confirm analytic predictions.

1 Introduction

Vibrations may be problematic in mechanical systems. They can provoke abnormal wear, noise or discomfort especially in case of transportation. Some devices have been designed in order to control these vibrations. Frahm [1] proposed a tuned mass damper i.e. a spring mass device coupled with the main system and able to reduce the energy of one mode. Later on, Roberson [2] showed that a nonlinear behavior of the control device can be more efficient. Since then, several nonlinear absorbers have been designed such as nonlinear tuned vibration absorber (NTVA) [3] or the nonlinear energy sink (NES) [4, 5]. The latter is purely nonlinear i.e. there is no linear term in the restoring force function. Whereas the tuned mass damper is efficient only for one mode, the NES can be used on a wider range of frequency.

Here, the system is a pendulum subject to a parametric excitation corresponding to the vertical displacement of its rotation axis. It corresponds to many industrial system, in particular to a rope-way vehicle excited by the movement of the cable. Matsuhisa et al. [6] designed several linear and nonlinear tuned mass damper in order to control the oscillations of a pendulum. Song [7] analyzed a parametrically excited pendulum used as a nonlinear absorber with an harmonic balance method. Hurel et al. [8, 9] studied a NES coupled to a two-dof pendulum excited by a generalized force with a multiple scale method. Here, the dynamical behavior of the system is also analyses with a multiple scale method.

In the section 2, the system is presented and dynamic equations are written. These equations are analyzed at two different scales of time in section 3. Then, in section 4, two numerical examples are given to illustrate analytic developments. Finally, the paper is concluded in section 5.

2 Description of the studied system

2.1 Main system

The main system is a pendulum in the plan $(\vec{e}_x,\ \vec{e}_y)$ attached at the point P by a hinged joint characterized by a viscous damping coefficient C_φ as seen on Fig. 1. Its mass, moment of inertia and center of mass are noted respectively M, J and G. The length L is the distance between the points P and G. The pendulum rotates around the point P with an angle φ. A gravitational field of magnitude g and direction $-\vec{e}_y$ exists.

2.2 Nonlinear absorber

In order to control the oscillations of the pendulum, a nonlinear absorber is coupled to the main system at a distance a from the point P. The mass m of the absorber is very small compared to the mass of the main system.

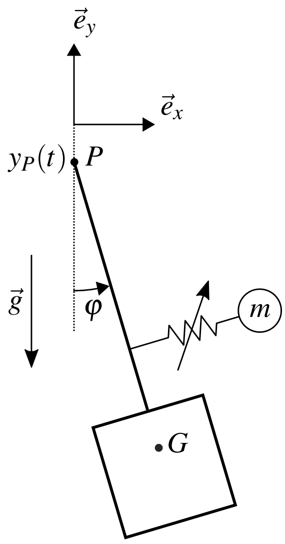


Figure 1 – Parametrically excited pendulum coupled with a nonlinear energy sink.

The ratio of mass is called ε:

$$\varepsilon = \frac{m}{M} \ll 1 \tag{1}$$

The nonlinear force function of the absorber reads:

$$s(u) = Ku^3 + C_u \dot{u} \tag{2}$$

where u is the relative displacement between m and the attached point with the main system and C_u is a viscous damping coefficient.

2.3 Parametric excitation

The main system is subject to a parametric excitation: an imposed vertical displacement of the point P called $y_P(t)$. We assume the displacement small (order of ε) and periodic with a frequency Ω. It can be written as Fourier series:

$$y_P(t) = \varepsilon \sum_{n \in \mathbb{Z}} y_n e^{in\Omega t} \tag{3}$$

where i is the complex number such as $i^2 = -1$.

2.4 Dynamical equations

The coordinates of the center of mass G and the mass of the absorber m read:

$$\begin{cases} x_G = L\sin(\varphi) \\ y_G = y_P - L\cos(\varphi) \end{cases}, \qquad \begin{cases} x_m = a\sin(\varphi) + u\cos(\varphi) \\ y_m = y_P - a\cos(\varphi) + u\sin(\varphi) \end{cases} \tag{4}$$

The kinetic $\mathscr{K}$ and potential $\mathscr{U}$ energies of the system became:

$$\mathscr{K} = \frac{1}{2}J\dot{\varphi}^2 + \frac{1}{2}M\left(\dot{x}_G^2 + \dot{y}_G^2\right) + \frac{1}{2}m\left(\dot{x}_m^2 + \dot{y}_m^2\right) \tag{5}$$

$$\mathscr{U} = Mgy_G + mgy_m + \frac{1}{4}Ku^4 \tag{6}$$

The non-conservative internal forces of the system are:

$$F_\varphi = C_\varphi \dot{\varphi} \tag{7}$$

$$F_u = C_u \dot{u} \tag{8}$$

We deduct from Eqs. 4, 5, 6, 7 and 8 with the Lagrange equations, the dynamic equations of the system:

$$\begin{cases} \left[L + j + \varepsilon(a^2 + u^2)\right]\ddot{\varphi} + \varepsilon a\ddot{u} + \varepsilon c_\varphi \dot{\varphi} + 2\varepsilon\dot{\varphi}\dot{u}u + \left[L\sin(\varphi) + \varepsilon(a\sin(\varphi) + u\cos(\varphi))\right](g + \ddot{y}_P) = 0 \\ \varepsilon\left[a\ddot{\varphi} + \ddot{u} - \dot{\varphi}^2 u + c_u\dot{u} + (g + \ddot{y}_P)\sin(\varphi)\right] + ku^3 = 0 \end{cases} \tag{9}$$

where $j = \dfrac{J}{M}$, $c_\varphi = \dfrac{C_\varphi}{M}$, $c_u = \dfrac{C_u}{m}$ and $k = \dfrac{K}{M}$. The natural frequency ω_0 of the main system alone at small angle reads:

$$\omega_0 = \sqrt{\frac{Lg}{j + L^2}} \tag{10}$$

3 Asymptotic behavior

We use a multiple scale method to understand the behavior of the system at several scales of time. To this end, the time t is broken down in several scales τ_n, thanks to the small parameter ε:

$$\tau_n = \varepsilon^n t,\ n \in \mathbb{Z} \tag{11}$$

The derivative operator can be redefined:

$$\frac{\mathrm{d}}{\mathrm{d}t} = \sum_{n\in\mathbb{Z}} \varepsilon^n \frac{\partial}{\partial \tau_n} \tag{12}$$

We assume the angle φ and the displacement u are small. A change of scale can be performed:

$$\varphi = \sqrt{\varepsilon}\overline{\varphi} \tag{13}$$

$$u = \sqrt{\varepsilon}\overline{u} \tag{14}$$

Then the complex variables of Manevitch [10] are introduced:

$$\Phi \mathrm{e}^{i\Omega t} = \dot{\overline{\varphi}} + i\Omega\overline{\varphi} \tag{15}$$

$$U \mathrm{e}^{i\Omega t} = \dot{\overline{u}} + i\Omega\overline{u} \tag{16}$$

In the following development, we keep only the first harmonics thanks to a Galerkin method. This is carried out for an arbitrary function of the system $h(\tau_0, \tau_1, \tau_2, ...)$ via:

$$H = \frac{\Omega}{2\pi}\int_0^{\frac{2\pi}{\Omega}} h(\tau_0, \tau_1, \tau_2, ...)\mathrm{e}^{-i\Omega\tau_0}\mathrm{d}\tau_0 \tag{17}$$

We assume the frequency of the first harmonic of the excitation Ω is closed to ω_0:

$$\Omega - \omega_0 = \sigma\varepsilon \tag{18}$$

3.1 Slow time scale τ_0

At fast time scale τ_0, the Eqs. 9 of the system yield to:

$$\frac{\partial \Phi}{\partial \tau_0} = 0 \tag{19}$$

$$\frac{\partial U}{\partial \tau_0} + i\frac{a\omega_0^2 - g}{2\omega_0}\Phi + \frac{i\omega_0 + c_u}{2}U - i\frac{3k}{8\omega_0^3}|U|U^2 = 0 \tag{20}$$

We conclude from the Eq. 19 that the amplitude Φ is independent of fast time τ_0. We are looking for the asymptotic state of the system at fast time: $\tau_0 \longrightarrow \infty$ and $\frac{\partial U}{\partial \tau_0} = 0$. By writing the complex variables in the polar form $\Phi = N_\varphi \mathrm{e}^{i\delta_\varphi}$ and $U = N_u\mathrm{e}^{i\delta_u}$, the Eq. 20 gives:

$$(a\omega_0^2 - g)^2 N_\varphi^2 = \left(\frac{3k}{4\omega_0^2}N_u^3 - \omega_0^2 N_u\right)^2 + c_u^2\omega_0^2 N_u^2 \tag{21}$$

The Eq. 21 describes the slow invariant manifold of the system (SIM). It is showed on Fig. 2 with the following parameters: $k = 0.15\,\mathrm{m^{-2}\,s^{-2}}$, $j = 10\,\mathrm{m}$, $L = 1\,\mathrm{m}$, $a = 1\,\mathrm{m}$, $c_u = 0.1\,\mathrm{s^{-1}}$ and $g = 9.81\,\mathrm{m\,s^{-2}}$. By following the method described by Ture Savadkoohi et al. [11] we find the following singular points:

$$N_{u1,2} = \frac{2\omega_0^{3/2}\sqrt{2\omega_0 \pm \sqrt{\omega_0^2 - 3c_u^2}}}{3\sqrt{k}} \tag{22}$$

The zone of the SIM between these singular points is unstable.

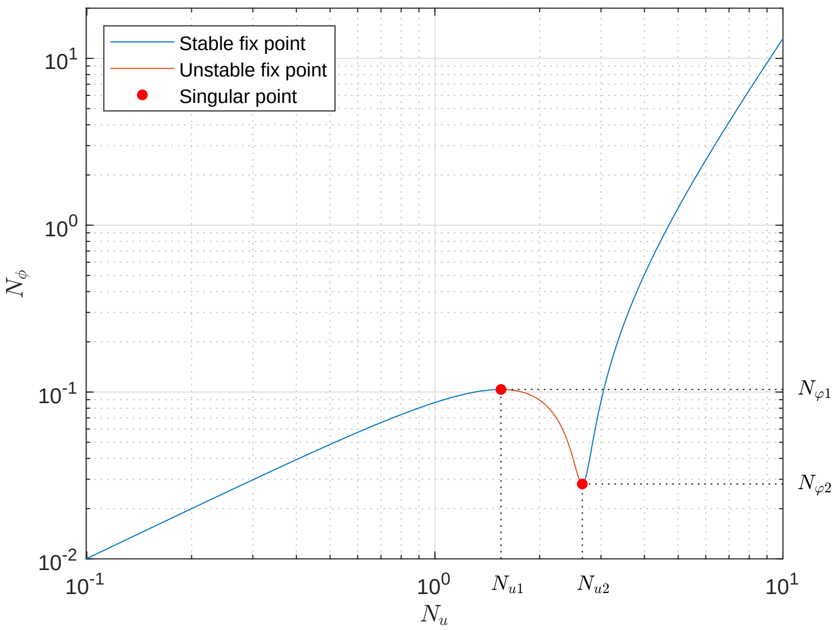


Figure 2 – Slow Invariant Manifold of the system with stable and unstable zones and singular points.

3.2 Fast time scale τ_1

We study now the system at fast time scale around the SIM. The analysis of Eqs. 9 gives:

$$Lg\frac{\partial \Phi}{\partial \tau_1} + \left(i\frac{2\sigma Lg + a\omega_0(a\omega_0^2 - g)}{2} + \frac{c_\varphi \omega_0^2}{2}\right)\Phi + i\frac{\omega_0(a\omega_0^2 - g)}{2}U - 2iy_2 L\omega_0^3 \Phi^\star + i\frac{Lg\,|\Phi|^2\,\Phi}{16\omega_0} = 0 \tag{23}$$

The complex variable S can be expressed as a function of U thanks to the Eq. 20 of the SIM:

$$\Phi = \frac{U}{a\omega_0^2 - g}\left(\frac{3k\,|U|^2}{4\omega_0^2} - \omega_0^2 + 4ic_u\omega_0\right) \tag{24}$$

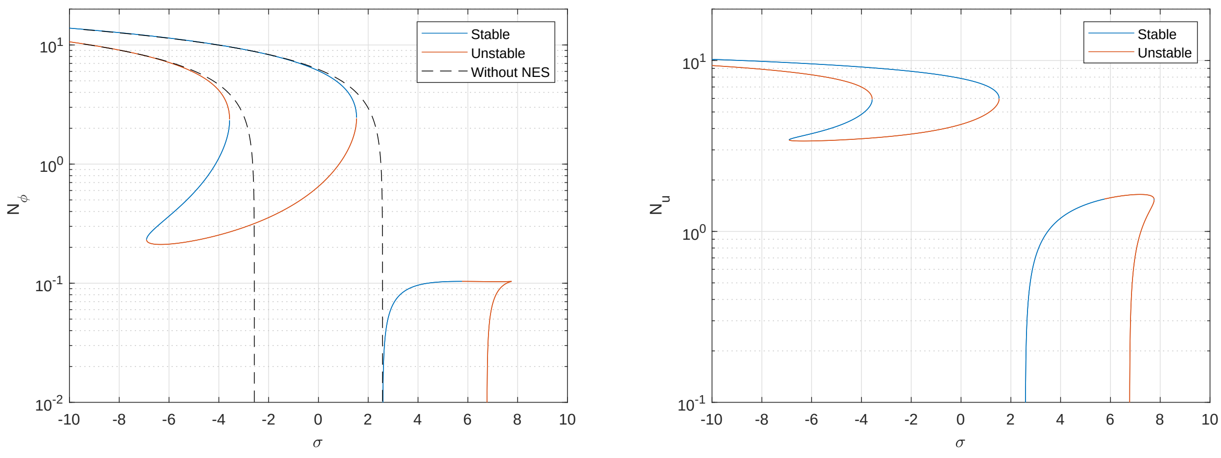
To find he equilibrium points, we consider no variation of Φ at fast time i.e. $\frac{\partial \Phi}{\partial \tau_1} = 0$. By replacing Eq. 24 in Eq. 23 and by taking the norm, we obtain a polynomial of degree 9 in N_u^2. The solutions are obtained numerically and represented on Fig. 3 as a function of σ with $c_\varphi = 50\,\mathrm{m\,s^{-1}}$ and $y_2 = 20\,\mathrm{m}$.

The stability of the equilibrium points is determined by a perturbation method with Eqs. 23 and 24:

$$\delta_u \to \delta_u + \Delta\delta_u, \qquad N_u \to N_u + \Delta N_u \tag{25}$$

After linearisation, we can write:

$$\mathbf{A}\begin{bmatrix}\dfrac{\partial \Delta\delta_u}{\partial \tau_1} \\ \dfrac{\partial \Delta N_u}{\partial \tau_1}\end{bmatrix} = \mathbf{B}\begin{bmatrix}\Delta\delta_u \\ \Delta N_u\end{bmatrix} \tag{26}$$



(a) N_φ

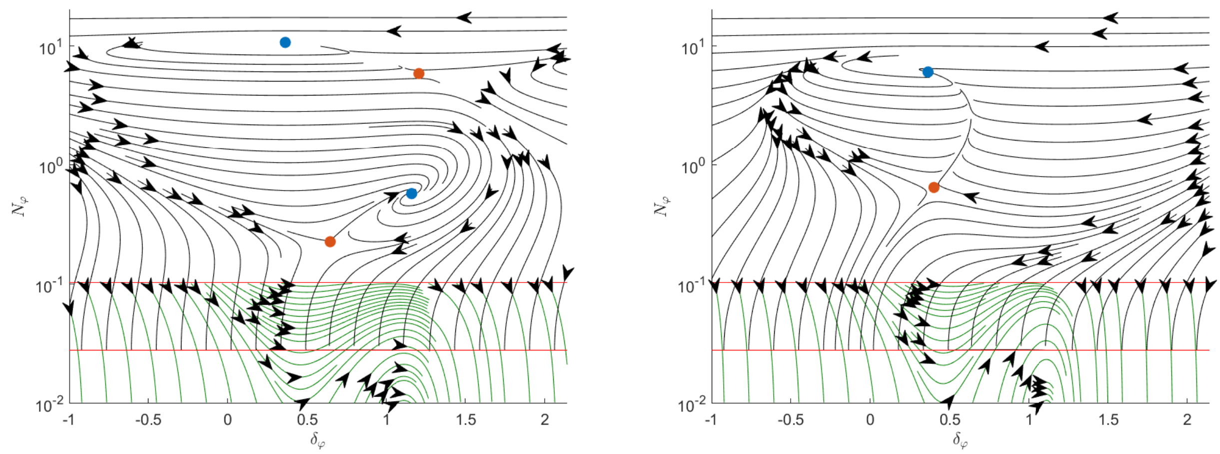
(b) N_u

Figure 3 – Equilibrium point of the system with NES compared to the equilibrium points of the system without NES (dashed line) as a function of σ.

where $\mathbf{A}$ and $\mathbf{B}$ are matrices. The stability depends on the signs of the eigenvalues of $\mathbf{A}^{-1}\mathbf{B}$. On the Fig. 3, the stability of the equilibrium points is represented with blue (stable) and red (unstable) colors. Note that $N_u = N_\varphi = 0$ is also an equilibrium point for every σ but is not represented because of the logarithmic scale of the graph. Its stability depends on δ_u.

3.3 Phase portrait

The knowledge of the equilibrium points and their stability is not enough to predict the behavior of the system. To complete the analysis, phase portraits are computed and plotted on Fig. 4 for two values of σ. In each case, for a starting point of the system with $N_\varphi(0) < 0.2\,\mathrm{s}^{-1}$, the system will evolute to stay below the second singular point i.e. $\exists t_1, \forall t > t_1\; N_\varphi(t) \leq N_{\varphi 1}$. This fact is illustrated with numerical examples in the next section.



(a) $\sigma = -5\,\mathrm{Hz}$

(b) $\sigma = 0\,\mathrm{Hz}$

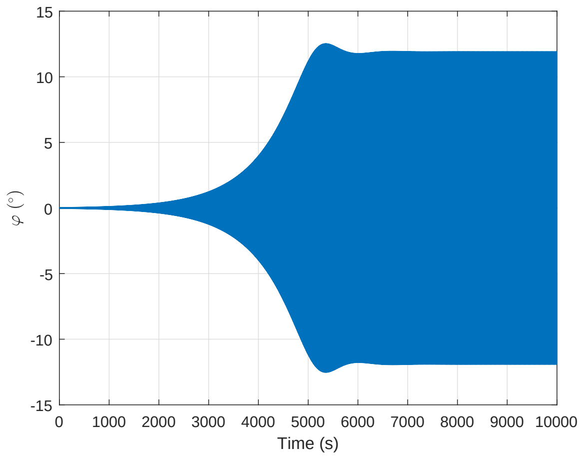
Figure 4 – Phase portrait of the system for two values of σ. A blue point is a stable equilibrium point whereas a red point is an unstable equilibrium point. Red lines correspond to singular points $N_{\varphi 1}$ and $N_{\varphi 2}$. The first stable zone of the SIM ($N_u < N_{u1}$) is represented by green curves and the second stable zone ($N_u > N_{u2}$) by black curves. The equilibrium point $N_\varphi = 0$ is not visible because of the logarithmic scale.

4 Numerical simulations

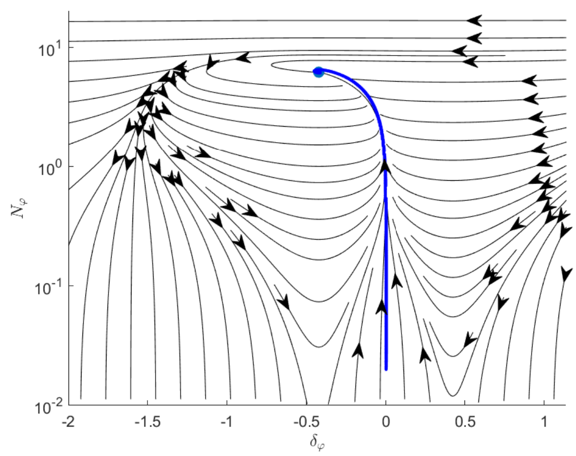
In order to illustrate the efficiency of the NES, two cases are shown with $\sigma = 0\,\text{Hz}$:

- case 1: the pendulum without NES with a very low initial amplitude $N_\varphi(0) = 0.1\,\text{s}^{-1}$ but with a particular phase $\delta_\varphi(0) = 0\,\text{rad}$.
- case 2: the pendulum with NES with a relatively high initial amplitude $N_\varphi(0) = 0.4\,\text{s}^{-1}$ and an arbitrary phase $\delta_\varphi(0) = 1\,\text{rad}$.

The results are shown on figure 5 and 6 with $\varepsilon = 10^{-2}$. First, we note a good agreement between numerical calculations and analytic phase portraits. In the first case, despite the initial amplitude is low, the system moves toward an equilibrium point with high amplitude ($N_\varphi = 0.8$). In the second case, the system goes to the equilibrium point with zero amplitude while the initial condition was higher than in the previous case. This is true for any initial phase angle $\delta_\varphi(0)$.

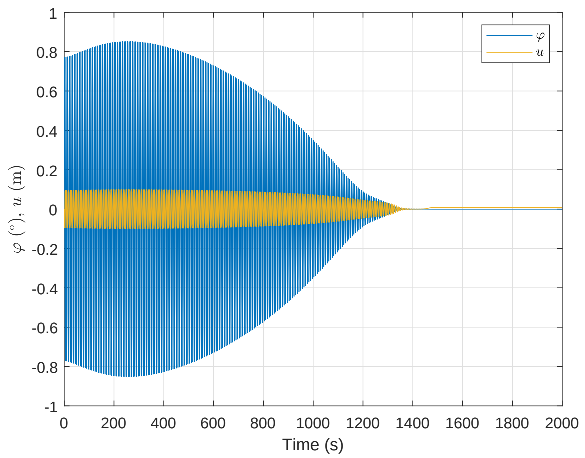


(a) φ as a function of time.

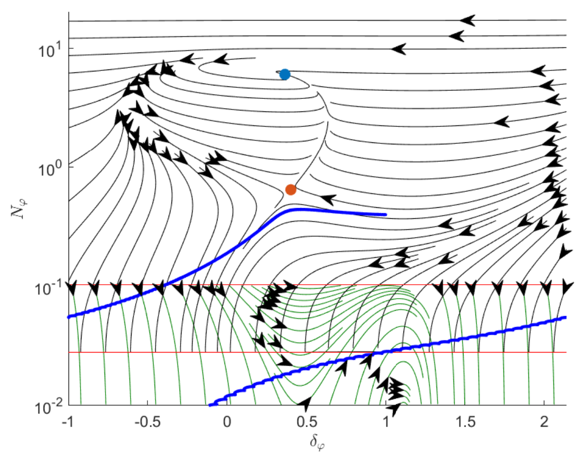


(b) Phase portrait and numerical result.

Figure 5 – Case 1: numerical result without NES, $N_\varphi(0) = 0.1\,\text{s}^{-1}$, $\delta_\varphi(0) = 0\,\text{rad}$



(a) φ and u as functions of time.



(b) Phase portrait and numerical result.

Figure 6 – Case 2: numerical result with NES, $N_\varphi(0) = 0.4\,\text{s}^{-1}$, $\delta_\varphi(0) = 1\,\text{rad}$

5 Conclusion

After a presentation of a system of a pendulum coupled with a nonlinear energy sink under parametric excitation, the dynamic equations are written thanks to the Lagrange equations. The analysis with a multiple scale method at fast time shows a slow invariant manifold of the system. At the next order, the equations give the equilibrium points and their stability. They are traced as a function of the frequency of excitation. To better predict the behavior of the system, phase portraits are drawn at different values of frequency. They show that the nonlinear energy sink can help the system to stay below a value instead of going to an equilibrium point with very high amplitude. This result is illustrated by two numerical examples. The first one shows that even if the initial amplitude of the angle of the pendulum without absorber is very low, the system can reach very high amplitude. As shown in the second example, this does not happen with a nonlinear absorber.

Acknowledgements

The authors would like to thank “La Région Auvergne-Rhône-Alpes” for supporting this work in the frame of the CALIPSO project 17 010971 01 - 15 713.

References

[1] H. Frahm, “Device for damping vibrations of bodies.,” Apr. 1911. US Patent 989,958.

[2] R. E. Roberson, “Synthesis of a nonlinear dynamic vibration absorber,” *Journal of the Franklin Institute*, vol. 254, pp. 205–220, Sept. 1952.

[3] G. Habib and G. Kerschen, “A principle of similarity for nonlinear vibration absorbers,” *Physica D: Nonlinear Phenomena*, vol. 332, pp. 1–8, Oct. 2016.

[4] O. Gendelman, L. I. Manevitch, A. F. Vakakis, and R. M’Closkey, “Energy Pumping in Nonlinear Mechanical Oscillators: Part I-Dynamics of the Underlying Hamiltonian Systems,” *Journal of Applied Mechanics*, vol. 68, no. 1, p. 34, 2001.

[5] A. F. Vakakis and O. Gendelman, “Energy Pumping in Nonlinear Mechanical Oscillators: Part II. Resonance Capture,” *Journal of Applied Mechanics*, vol. 68, no. 1, p. 42, 2001.

[6] H. Matsuhisa and M. Yasuda, “Dynamic vibration absorber for pendulum type structure,” Oct. 1995. US Patent 989,958.

[7] Y. Song, H. Sato, Y. Iwata, and T. Komatsuzaki, “The response of a dynamic vibration absorber system with a parametrically excited pendulum,” *Journal of Sound and Vibration*, vol. 259, pp. 747–759, Jan. 2003.

[8] G. Hurel, A. Ture Savadkoohi, and C.-H. Lamarque, “Nonlinear vibratory energy exchanges between a two degrees-of-freedom pendulum and a nonlinear absorber.,” *Journal of Engineering Mechanics*, 2019. (in press).

[9] G. Hurel, A. Ture Savadkoohi, and C.-H. Lamarque, “Passive control of a two degrees-of-freedom pendulum by a non-smooth absorber,” *Nonlinear Dynamics*, 2019. (in press).

[10] L. I. Manevitch, “The Description of Localized Normal Modes in a Chain of Nonlinear Coupled Oscillators Using Complex Variables,” *Nonlinear Dynamics*, vol. 25, pp. 95–109, July 2001.

[11] A. Ture Savadkoohi, C.-H. Lamarque, M. Weiss, B. Vaurigaud, and S. Charlemagne, “Analysis of the 1:1 resonant energy exchanges between coupled oscillators with rheologies,” *Nonlinear Dynamics*, vol. 86, pp. 2145–2159, Dec. 2016.

Smart Structures

Hybrid crankshaft control for the reduction of torsional vibrations and rotational irregularities

Guillaume PAILLOT[1], Simon CHESNÉ[1], Didier RÉMOND[1]

[1]Univ Lyon, INSA-Lyon, CNRS UMR5259, LaMCoS, F-69621, France
guillaume.paillot@insa-lyon.fr

Abstract

Rotative systems do not have a constant revolution speed. The problem is even more significant in internal combustion engines, where the crankshaft is submitted to a torque that is far from being constant over a revolution period. This matter causes unacceptable noises in the the gearbox and fatigues the shaft, and it can be even further amplified by the shaft dynamics. This well-known problem is already tackled by numerous passive systems, and some active devices are being introduced to further enhance their capabilities. However, these ones are complex, not always fail-safe, need a control unit and consume high levels of energy.
In this paper, the authors introduce a hybrid self-fed damper for the so-called rotational irregularities, based on a tuned massed damper controled through an electromagnetic coupling by the irregular behaviour itself. For a better description, the whole behaviour of this non-stationary process is governed with an angular approach.

1 Introduction

In an internal combustion engine, the impulse initiated by the combustion of gas within the cylinders is able to put the crankshaft into a rotating movement, thus transmitting a torque to the wheels through the drivetrain. However, the movement of the piston is not uniform, and the transmitted torque is far from being constant at a steady engine regime [1]. Indeed, the important pressure variations on the piston as well as the influence of the alternating mass inertia create a periodic oscillation around an average value in torque. This creates a similar oscillation in angular acceleration, speed, and position, which is commonly referred to as rotational irregularities.
This oscillation tends to generate unacceptable noises, such as rattling noise in the gearbox, and fatigues the drivetrain parts, shortening the operating lifetime. For these reasons, such irregularities need to be reduced, which is traditionally done by a flywheel located between the crankshaft and the gearbox, while a viscous Lanchester damper on the front end reduces the torsional magnitude through energy dissipation. However, due to the downsizing of engines for fuel-consumption reasons, the simple flywheel is not enough anymore for acceptable noise levels.
This rotational irregularity phenomenon can be further amplified when the critical frequency of the crankshaft is a multiple of the current engine average speed. Here again, downsizing is a problem as a crankshaft length reduction increases the critical frequency, enabling a larger spectrum of engine speeds to excite the resonance. The presence of the viscous damper ensures that the crankshaft does not immediately break, but at the cost of a dissipated power up to 1kW for heavy duty vehicles.

For those reasons, there is a continuous research for more efficient solutions tackling the two presented issues : mitigation of the torsional levels in service and especially at the critical frequencies, and reduction of the dissipated energy in the damper. A large amount of passive solutions have been introduced over the two last decades in order to decrease the magnitude of torsional motions, typical examples being the Double (or Triple) Mass Flywheel ([2],[3]), variable inertia flywheels ([4],[5]), planetary gear systems ([6]), or Centrifugal Pendulum Vibrations Absorbers ([7],[8],[9]). These latter rely on the creation of an extra degree of freedom, enabling an "escape way" during resonance. However, its behaviour is still dependent on the steady engine regime, complicating a tuning process[10].

Some active concepts have also been proposed ([11],[12],[13],[14]), as the control possibilities they offer greatly enhance the capabilities of such dampers. Their limitation is the need for a power source.
Regarding the second issue, a promising way would be to collect the otherwise dissipated energy, and to store it so that it can be used by other devices in the vehicle. Several possibilities exist for the conversion, based on piezoelectricity - R-EHVTA [15] or pendulum harvester [16] - or on electromagnetic coupling, with some concepts developped over the last years, such as [17] or [18].

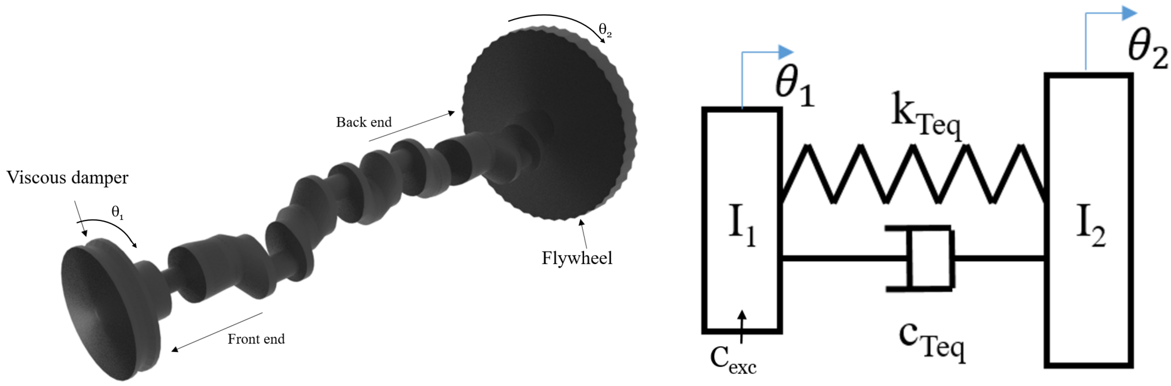
In this paper, the authors introduce a concept for a new self-fed rotational hybrid damper, combining an electromagnetic harvester with a hybrid Tuned Mass Damper (hTMD), that has already proved its capabilities for translational motions applications [19]. It is based on the generation of a counteracting torque by a set of permanent magnets and coils, similar to what happens in active dampers, coupled with the principle of the TMD, efficient against resonance issues. The TMD embeds a permanent magnet that moves in a coil. Both movements will change the local magnetic fields and generate an induced current in a common circuit, acting on each other.
In order to tackle non-stationary operating conditions, the present model involves angular and time approaches to describe the cyclic excitations and the frequency resonances.

2 Model building, important concepts and assumptions

The present concept can be applied to any rotating shaft submitted to rotational irregularities, no matter their profile. However in this paper we consider a 6-cylinder crankshaft, each with a cubic capacity of 1,3L.

2.1 Mechanical model

The crankshaft is first reduced to a 2 DoF-in-torsion part, as only the first torsional mode is of interest in this academic first application. As described in Fig.1, a regime-dependant cyclic excitation C_{exc} is introduced on the front degree of freedom. This excitation is the sum of the the torque produced by the explosion in the cylinders and of inertial effects. It is further described in the next section.
The measures are conducted on both ends, the second being the output of the shaft, connected to the gearbox. On this one, a resistive torque C_c assumed constant is applied, so as to balance the system and enable a steady speed. The actual interest is indeed the difference between both angles $\theta_2 - \theta_1$ for a constant engine regime. Numerical values of the introduced parameters are to be found in Table.1.



(a) Full crankshaft (b) Reduced model

Figure 1 – Model reduction

This model is governed by the system of equations :

$$\begin{cases} I_1\ddot{\theta}_1 + c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2) = C_{exc} \\ I_2\ddot{\theta}_2 + c_{Teq}(\dot{\theta}_2 - \dot{\theta}_1) + k_{Teq}(\theta_2 - \theta_1) = -C_c \end{cases} \quad (1)$$

Parameter	Value
I_1	0.23 kg.m^2
I_2	1.65 kg.m^2
k_{Teq}	524 114 Nm/rad
c_{Teq}	100 Nms/rad

Table 1 – Parameters of the 2-DoF model

In a perfect engine, the natural frequency is excited only by the orders that are multiples of the number of cylinders, N_{cyl}, devided by two. To that extent, only the orders multiple of 3 are here taken into account. According to Table 2, as our engine speed range is from 600RPM to 2600RPM and the natural frequency is 256,8 Hz, we can restrict the analysis up to the 24th order : higher orders will not excite the natural frequency .

Order	Engine steady regime
3	-
6	2569 RPM
9	1713 RPM
12	1285 RPM
15	1028 RPM
18	857 RPM
21	734 RPM
24	642 RPM

Table 2 – The investigated orders and the corresponding engine regime at resonance

2.2 Excitation torque profile and non-stationary conditions

In a 4-stroke internal combustion engine, a complete period is accomplished in two revolutions, each stroke corresponding to half a revolution. Among them, only the firing stroke produces a useful torque, with a non-constant force over the revolution. However, the presence of the crank also causes it to be highly dependant on the instantaneous position α, as it is depicted in Fig.2. Noting λ the ratio of the rod length L_B to the crank radius R, F_{bg} the projection on the rod of the gas-induced force and d its lever arm, a geometric projection of $C_g = -F_{bg}.d$ brings :

$$C_g(\alpha) = -F_g(\alpha) \cdot R \cdot \sin(\alpha) \cdot \left(1 + \frac{\cos(\alpha)}{\sqrt{\lambda^2 - \sin(\alpha)^2}}\right) \tag{2}$$

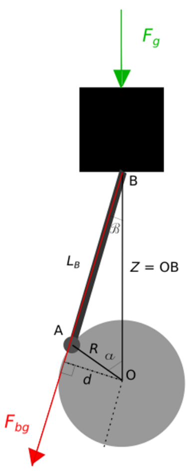


Figure 2 – Crankshaft, piston and rod

The force F_g, created by the explosion in the piston, is also dependant on the engine steady regime. In this expression, a constant part is the engine torque, and the oscillations around it are a first source of rotational irregularities.

The second source originates from the motions of the masses around and above the axis : the sum of the forces created by their inertias is not 0. Here again, there is a strong dependence on the instantaneous angular position, and also on the instantaneous angular speed. We note m_{alt} the mass considered in a translational motion when the shaft rotates. Assumed that the inertias do not change with speed and position :

$$\begin{cases} F_i = -m_{alt}.\dfrac{d^2Z}{dt^2} \\ Z = R\Big(\cos(\alpha) + \sqrt{\lambda^2 - \sin(\alpha)^2}\Big) \\ C_i = -F_i.R.\Big(cos(\alpha) + \sqrt{\lambda^2 - sin(\alpha)^2}\Big)\dfrac{sin\alpha}{\sqrt{\lambda^2 - sin(\alpha)^2}} \end{cases} \tag{3}$$

Which provides :

$$C_i = -m_{alt}.R^2\left(\left(sin(\alpha) + \frac{sin(2\alpha)}{2\sqrt{\lambda^2 - sin(\alpha)^2}}\right)\ddot{\alpha} + \left(cos(\alpha) + \frac{\lambda^2 cos(2\alpha) + sin(\alpha)^4}{(\lambda^2 - sin(\alpha)^2)^{3/2}}\right)\dot{\alpha}^2\right)sin(\alpha)\left(1 + \frac{cos(\alpha)}{\sqrt{\lambda^2 - sin(\alpha)^2}}\right) \tag{4}$$

The final excitation torque C_{exc} is the sum of the inertial torque and the oscillating gas torque. Its process is non-stationnary, though the simplifying stationary assumption $\alpha = \omega.t$ with a constant ω is often made. However, the instantaneous speed does not remain constant, since the input torque itself is harmonic : the angular acceleration is not 0, and its magnitude is directly linked to C_{exc}. In order to prevent any hidden behaviour to disappear in the analysis, and as the angular position θ_1 appears as the natural variable for such a rotating system, we use this variable instead of time for the resolution. The counterpart here is the introduction in the equations of a non-linear behaviour when the variable is swapped, as there is now a division by the instantaneous angular speed : $\frac{dX}{d\theta_1} = \frac{1}{\dot{\theta}_1}\frac{dX}{dt}$.

2.3 TMD model

First addition to the previous 2-DoF model is the mass damper, tuned on its resonance frequency according to Den Hartog's principles [20]. It consists in a magnet clamped on a mass, linked to the rotating shaft with a spring set at a certain distance R_{TMD} of the rotation axis. The resulting torsional stiffness k_{TMD} depends on the tuning. Around this magnet, a coil is set, clamped to the shaft, in such a way that the TMD can translate within.

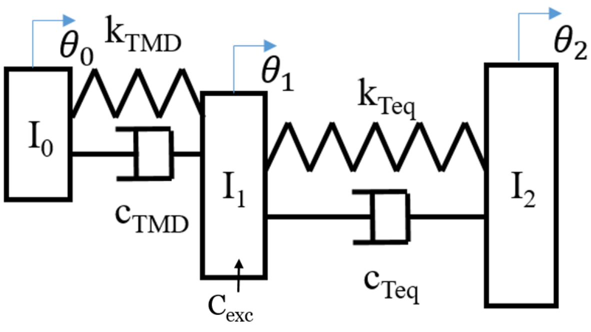


Figure 3 – Reduced model of crankshaft including a TMD in front position

However, as we want to control the torsional behaviour, determined by the difference of the angles, both degrees of freedom are potential candidates to host the TMD set, but the two will not act the same.
The front end has the greatest amplitude during vibration due to its lower inertia, and the lower the inertia, the lower the efficient modal inertia I_{eq}. At a given R_{TMD}, a lower mass for the TMD is then required, as the ratio of its inertia I_0 to I_{eq} is the key parameter of the tuning. To that extent, better performance can be achieved for

Parameter	Value
I_0	0.013 kg.m^2
μ	0.05
k_{TMD}	29 617 Nm/rad
c_{TMD}	7.36 Nms/rad
R_{TMD}	0.2 m
m_{TMD}	0.325 kg

Table 3 – Parameters of the 3-DoF model

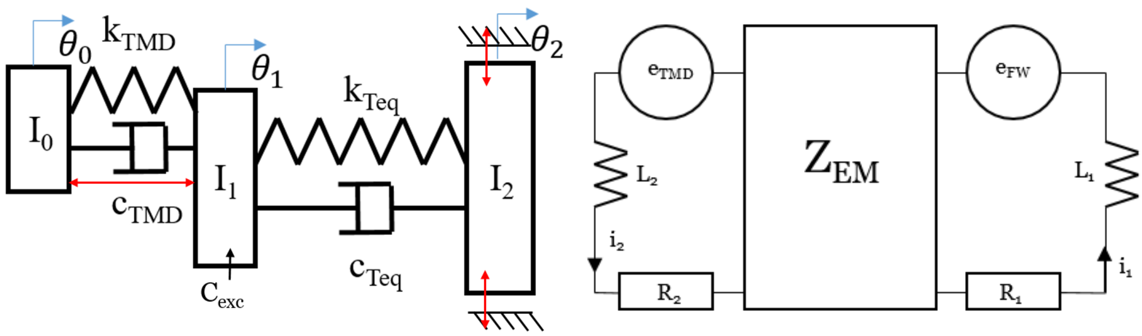
a TMD on the front end and in the rest of the paper, this integration possibility is chosen, described in Fig.3. Its absolute angle around the rotation axis is parameterized by θ_0.
To bring what happens to the TMD in a more practical frame, let us define a relative variable for the new degree of freedom : $x = R_{TMD}(\theta_0 - \theta_1)$. This variable translates the torsional motion of the TMD around the axis of rotation in regard to the disc to which it is clamped. The values for all introduced parameters can be found in Table 3. The inertia ratio μ is chosen to 0.05, as this value allows a realistic mass and a realistic distance to the shaft to be chosen.

2.4 Electromagnetic model

The electromagnetic model actually consists in two parts. One has already been introduced in subsection 2.3, and describes the interaction of the magnet on the TMD and its surrounding coil. Indeed, any motion of the TMD creates an induced current in the coil, and conversely, any current circulating in the coil creates a magnetic field interacting with the magnet.

The second part is located around the flywheel. It consists in a set of coils and another of magnets, one being embedded in the flywheel and rotating with the shaft, the other being fixed in the vehicle, all around the flywheel. Again, moving a magnet close to a coil will create a current, itself generating a magnetic field that will slow down the motion, according to Lenz's law, and eventually damp the rotational irregularities. We note ϕ the magnetic flux.

The red arrows in Fig.4a show the different electromagnetic interactions that occur. The two parts are linked to another through an impedance Z_{EM}, as to be seen in Fig.4b. What this impedance covers depends on the interaction scenario.



(a) Reduced model with EM interactions (b) Electromagnetic circuit

Figure 4 – Electromagnetic model

With the electromotive forces e_{FW} at the flywheel and e_{TMD} at the TMD, defined as:

$$\begin{cases} e_{FW} = -\dfrac{d\phi}{dt} \\ e_{TMD} = -T_2\dot{x} \end{cases} \tag{5}$$

and under the assumption that $Z_{EM} = 0$, the total model can thus be described with a system of five equations, three mechanical equations and two electrical ones :

$$\begin{cases} I_0\ddot{x} + (1+\frac{I_0}{I_1})(c_{TMD}\dot{x} + k_{TMD}x) - \frac{I_0}{I_1}R_{TMD}(c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2)) = -\frac{I_0}{I_1}R_{TMD}.C_{exc} + R_{TMD}^2 T_2 i_2 (1+\frac{I_0}{I_1}) \\ I_1\ddot{\theta}_1 + c_{Teq}(\dot{\theta}_1 - \dot{\theta}_2) + k_{Teq}(\theta_1 - \theta_2) - \frac{c_{TMD}}{R_{TMD}}\dot{x} - \frac{k_{TMD}}{R_{TMD}}x = C_{exc} - R_{TMD}^2 T_2 i_2 \\ I_2\ddot{\theta}_2 + c_{Teq}(\dot{\theta}_2 - \dot{\theta}_1) + k_{Teq}(\theta_2 - \theta_1) = -C_c + T_1.R_{ext} i_1 \\ L_1\frac{di_1}{dt} + R_1 i_1 = -pT_1\dot{\theta}_2 \\ L_2\frac{di_2}{dt} + R_2 i_2 = -T_2\dot{x} \end{cases} \tag{6}$$

Where T_2 is the coupling factor depicting the electromagnetic interaction between the TMD and the surrounding coil [21], p is a scalar and T_1 is a coupling function linked to $\frac{d\phi}{dt}$. Both are further described in the next subsection.
In the case where the TMD is actually located on the flywheel, or depending on Z_{EM}, these equations can be slightly different.

2.5 The T_1 function and p parameter

In subsection 2.2, we have stated that an angular approach is better suited to describe the rotation of the shaft and the irregularities that occur. Moreover, Faraday's law states that the electromotive force is proportional to the time derivative of the magnetic flux, which itself is then a function of the instantaneous angular position, here measured at the flywheel. Changing the variable t to θ_2 provides $e_{FW} = -\dot{\theta}_2\frac{d\phi}{d\theta_2}$.
On the other hand, the countering torque C_{count} created at the flywheel by the current also depends on the location of the various coils and magnets. It is able to mitigate the rotational irregularities when it and the excitation torque C_{ext} have the same main harmonics, but with a 180° phase shift. Assuming this torque and $\frac{d\phi}{d\theta_2}$ have a same sine shape involving the same harmonics, the introduced T_1 function aims to bridge the gap between them. To that extent, the torque exerted when a number n of coils and a number m of magnets interact during the revolution can be described using the expression :

$$T_1 = \sum_{i=1}^{n}\sum_{j=1}^{m} K_{i,j} m_j \mu_0 \frac{N_i}{L_{sol,i}} sin(\beta_{i,j}) \tag{7}$$

and the relationship :

$$C_{count} = T_1.R_{ext} i_1 \tag{8}$$

In (7), N_i is the number of turns in solenoid i, $L_{sol,i}$ the length of the same solenoid, m_j the magnetic moment of magnet j, and K is a matrix describing the coupling coefficient between coil i and magnet j. $\beta_{i,j}$ is a local variable standing for the angular position θ_2 when a magnet interacts with a coil.
With this description, we also have the relationship :

$$e_{FW} = -p\dot{\theta}_2 T_1 \tag{9}$$

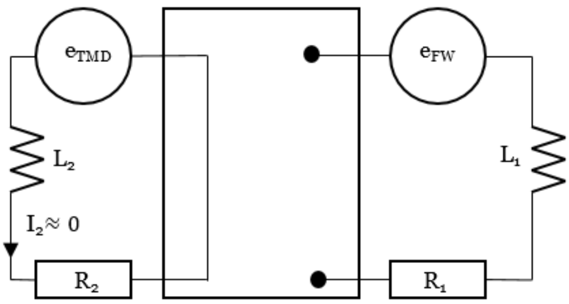
the description of the harmonics being included in T_1, except for their amplitude. To that extent, in order to keep realistic values for the electromotive force e_{FW}, we introduce an additional parameter, p, which purpose is to adjust the magnitude of the harmonics.

3 Model results

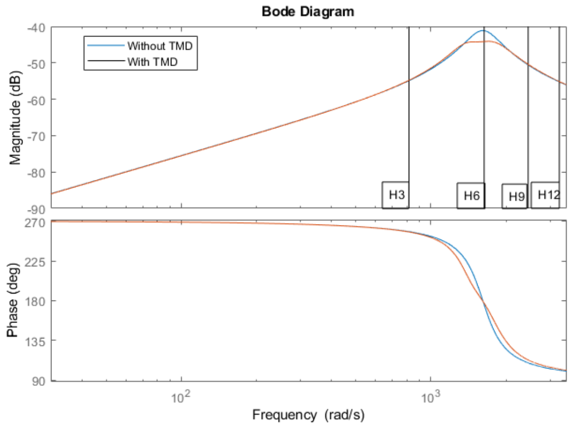
3.1 Efficiency of the TMD

To display the efficiency of the concept in our model, let us first consider a decoupled electromagnetic circuit, as modeled in Fig.5a. With a low value for T_2, the created current is very low, and the natural motion of the TMD is almost not impeded. Regarding the electromagnetic setup around the flywheel, there is also little impact on the overall rotation as the countering torque C_{count} is four orders of magnitude lower than the excitation torque C_{exc}. Such configuration is equivalent to actually no circuit at all, and enables to measure the effect of a passive TMD.

As we can see on the Bode diagram in Fig.5b, the TMD flattens the peak at the natural frequency of the torsional mode. The tuning parameters used are those detailed in Table.3. This enables a reduction of 15% the amplitude in torsion, as displayed in Fig.6a for the response at 2600RPM. At this regime, the 6^{th} order excites a frequency which is very close to the natural frequency. At 2200RPM, situation displayed in Fig.6b, the frequencies excited are further from the natural frequency, and the damping is smaller, even though still present.

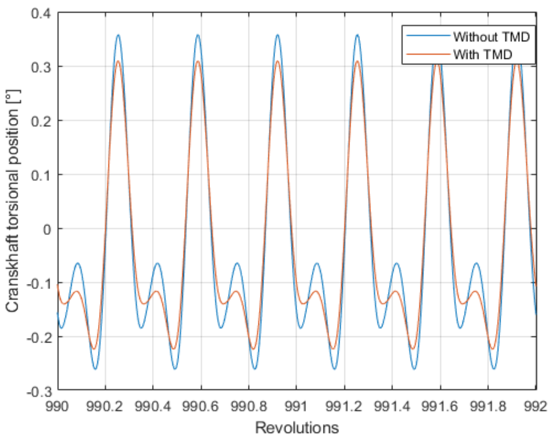


(a) Decoupled electromagnetic circuit

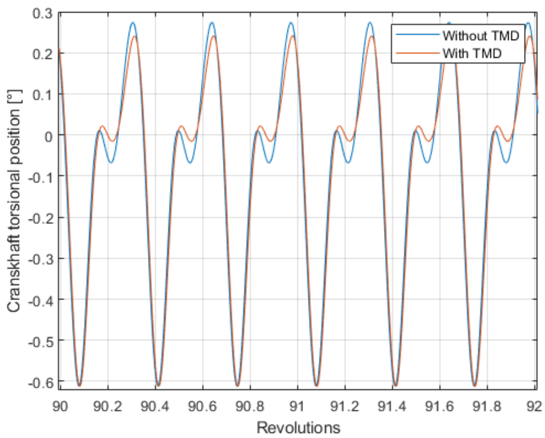


(b) Bode diagram with the harmonics at 2600RPM

Figure 5 – Passive TMD scenario



(a) At 2600RPM



(b) At 2200RPM

Figure 6 – Typical magnitude reductions

3.2 Coupling effect in the hybrid damper

Now let us consider a direct hookup between the two constituents of the electromagnetic circuit, a situation depicted in Fig.7. The equations in (6) must then be changed, as it is now imposed $i_1 = i_2$. The parameters used are listed in Table.4. In this situation, we still have $Z_{EM} = 0$

Parameter	Value
L_1	555mH
L_2	185mH
R_1	15.51Ω
R_2	5.17Ω
T_2	0.8
p	0.25
$T_1(\theta)$	$0.01.sin(6\theta)$

Table 4 – Parameters for the coupled model

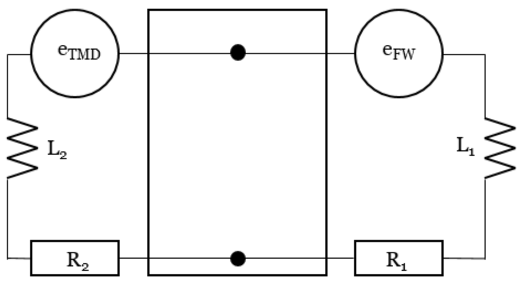
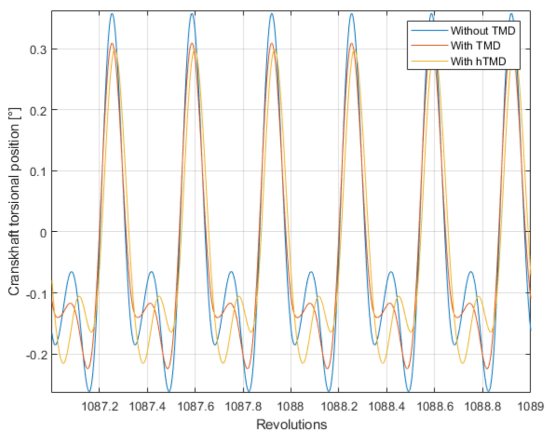
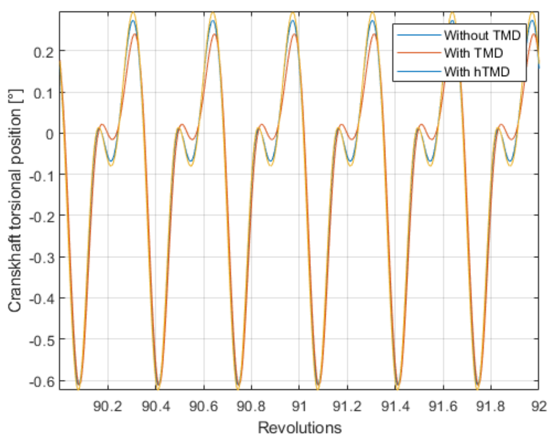


Figure 7 – Direct hookup electromagnetic circuit



(a) At 2600 RPM



(b) At 2200 RPM

Figure 8 – Comparison with the previous case

As we can see in Fig.8a, at 2600RPM in that particular situation, the damping is slightly better, even though the values considered are not tremendous. However, this results must be handled carefully, since no phase effect has yet been taken into account in T_1, effects that could make the hTMD ineffective or even worsen the situation, which is what appears at 2200RPM in Fig.8b. Indeed, the hookup is equivalent to a modification of c_{TMD} and - to a lesser extent - of k_{TMD}. A modification of the damping factor can thus be advantageous or detrimental, and what happens is highly dependent on the involved harmonics and phases, and on the hookup configuration. To that extent, it appears necessary to further investigate on the effect of each harmonic's phase and on the configurations for Z_{EM}, in order to properly damp the torsional mode in all cases.

Discussion - Conclusions

In this paper, a new concept for a hybrid mass damper against rotational irregularities has been introduced and described. The self-fed property avoids any exterior source of power, but also complicates the tuning of the system due to the instantanous angular position and the steady regime dependence. The interaction between the two parts of the electromagnetic circuit is not yet fully understood, whereas this is a major requirement to have them work together properly. Indeed, it has been seen in 3.2 that a direct hookup can provide unexpected behaviour if the phases are not carefully handled.

Further work will also be devoted to a check of the assumptions that are made. In particular, it is necessary to prove the possibility to approximate the $\frac{d\phi}{d\theta_2}$ function to a sine with given harmonics. There is also no guarantee that any harmonic signal can be approximated by at least one particular coil-and-magnet spatial layout. The tayloring is quite straightforward when only one or two harmonics are involved, though here 8 different harmonics are being investigated. However a truncation to the fundamental or to a predominant harmonic could be sufficient to efficiently damp the rotational irregularities. In all cases, further work on the damping possibilities of such a hTMD is expected, in order to present its capabilities in comparison to a fully passive TMD.

The hookup of the full electromagnetic circuit in a real engine also matters. It has been stated in 2.3 that the TMD is more efficient when mounted on the lower inertia, however the rest of the electromagnetic circuit is necessarily mounted on the flywheel, which is the other end of the crankshaft. In this configuration, it is nessary to embody the wires along the shaft in the rotating referential. The choice is left to equip the flywheel with the TMD instead, in another configuration that provides a better integration rate at the cost of a lower performance.

Acknowledgements

This work has been supported by the chair ‘Solutions for the future of Road Freight Transport’, jointly created by INSA Lyon LaMCoS and Volvo Group.

References

[1] J.-L. Ligier and E. Baron, *Acyclisme et vibrations: applications aux moteurs thermiques et aux transmissions. 1, Notions de base*. Paris, France: Ed. Technip, 2002, ISBN: 978-2-7108-0820-6.

[2] U. Schaper, O. Sawodny, T. Mahl, and U. Blessing, “Modeling and torque estimation of an automotive dual mass flywheel”, in *2009 American Control Conference*, St. Louis, MO, USA: IEEE, 2009, pp. 1207–1212, ISBN: 978-1-4244-4523-3. DOI: 10.1109/ACC.2009.5160136.

[3] L. Wramner, V. Berbyuk, and H. Johansson, “Vibration dynamics in non-linear dual mass flywheels for heavy-duty trucks”, in *Proceedings of ISMA2018 International Conference on Noise and Vibration Engineering*, 2018, pp. 1935–1947.

[4] X. Dong, “Magneto-rheological variable inertia flywheel”, *Smart Materials and Structures*, vol. 27, no. 11, Jul. 2018.

[5] L. L. Schumacher, “Controllable inertia flywheel”, U.S. Patent 4995282A, Feb. 26, 1991.

[6] A. Suryanarayana, “Engine dynamics and torsion vibration reduction”, Master Thesis, Chalmers, 2015, 52 pp.

[7] D. E. Newland, “Developments in the design of centrifugal pendulum vibration absorbers”, presented at the ICSV 24, 2017, p. 8.

[8] A. Kooy, Ed., *Isolation is the key*, Solving the Powertrain Puzzle: 10th Schaeffler Symposium April 3/4, 2014, Springer Vieweg, 2014, ISBN: 978-3-658-06194-4.

[9] M. Pfabe and C. Woernle, “Reduction of periodic torsional vibration using centrifugal pendulum vibration absorbers”, *PAMM*, vol. 9, no. 1, pp. 285–286, 2009, ISSN: 1617-7061. DOI: 10.1002/pamm.200910116.

[10] K. Nishimura, T. Ikeda, and Y. Harata, “Localization phenomena in torsional rotating shaft systems with multiple centrifugal pendulum vibration absorbers”, *Nonlinear Dynamics*, vol. 83, no. 3, pp. 1705–1726, Feb. 1, 2016, ISSN: 1573-269X. DOI: 10.1007/s11071-015-2441-2.

[11] J. Pfleghaar and B. Lohmann, “The electrical dual mass flywheel -an efficient active damping system”, *IFAC Proceedings Volumes*, 7th IFAC Symposium on Advances in Automotive Control, vol. 46, no. 21, pp. 483–488, Jan. 1, 2013, ISSN: 1474-6670. DOI: 10.3182/20130904-4-JP-2042.00046.

[12] S. Schwunk, “Method for vibration damping of a drive train by means of an electric machine”, U.S. Patent 20170334448A1, Nov. 23, 2017.

[13] A. Gruendl, B. Hoffmann, U. Masberg, T. Pels, and K.-P. Zeyen, “System zur aktiven dämpfung von drehungleichförmigkeiten der kurbelwelle eines verbrennungsmotors oder einer damit gekoppelten welle”, European pat. 0847485B1, Jun. 23, 1999.

[14] F. Knopf and M. Steidl, “Verfahren zum aktiven isolieren eines antriebsstrangs von drehschwingungen einer welle einer maschine, insbesondere einer kurbelwelle einer hubkolbenmaschine, und eine entsprechende anordnung zur durchführung des verfahrens”, pat. WO2014118246A1, Aug. 7, 2014.

[15] F. Infante, W. Kaal, S. Perfetto, and S. Herold, “Design development of rotational energy harvesting vibration absorber (r-EHTVA)”, Sep. 10, 2018. DOI: 10.1115/SMASIS2018-7902.

[16] L. Gu and C. Livermore, “Pendulum-driven passive self-tuning energy harvester for rotating applications”, *Applied Physics Letters*, p. 4,

[17] S. Troebst, “Dämpfungssystem mit Energierückgewinnung für ein Kraftfahrzeug und Antriebsvorrichtung mit einem derartigen Dämpfungssystem”, European pat. 2910813A2, Aug. 26, 2015.

[18] B. Gunn, S. Theodossiades, and S. J. Rothberg, “A nonlinear concept of electromagnetic energy harvester for rotational applications”, *Journal of Vibration and Acoustics*, Nov. 19, 2018, ISSN: 1048-9002. DOI: 10.1115/1.4042040.

[19] S. Chesne and C. Colette, “A simple hybridization of active and passive mass dampers”, presented at the ISMA2016, KU Leuven, Sep. 2016.

[20] J. P. Den Hartog, *Mechanical vibrations*, 4th. New York, NY : McGraw-Hill, 1956, 446 pp.

[21] A. Preumont and B. Mokrani, *Electromagnetic and piezoelectric transducers*.

Semi-active Torsional Vibrations Control of a Rotor Using a Smart ER Dynamic Absorber

Yulan Sun and Marc Thomas

Department of Mechanical Engineering, ETS,
1100, Notre-Dame West Street, Montreal, Quebec, (H3C 1K3), CANADA
marc.thomas@etsmtl.ca

Abstract
Torsional rotor vibration is always very difficult to control since the implementation of a control system is not an easy task while the machine is rotating. Excessive torsional vibration can lead to failures of many mechanical components. A common method for controlling vibrations involves the use of dynamic absorbers. Due to their variable properties, smart materials can be used to increase the frequency range in order to control vibration. This article is concerned with the application of Electrorheological Fluids (ERF) to the reduction of torsional vibrations of a rotor by controlling damping and stiffness of a rotational dynamic absorber. A cylindrical type of electrorheological (ER) torsional absorber is designed and manufactured according to the required damping force level and the critical velocity of the rotor system. This paper presents torsional vibration control performance of a smart ER dynamic absorber using a bang-bang (ON-OFF) control strategy. The experimental results very closely approximate the simulation results. These results show that the ER dynamic absorber exhibits very good performances in terms of reducing the torsional vibration of rotor system.

Key words: Electrorheological fluids, torsional vibrations, dynamic absorber, Bang-Bang control, smart material.

1. Introduction

Torsional rotor vibration is always very difficult to control since the implementation of a control system is not an easy task while the machine is rotating. Excessive torsional vibration can lead to failures of many mechanical components. Furthermore, in many industrial applications, torsional vibration problems may not be apparent until a failure occurs. For example, it is the case when the machine contains a gearbox which causes the torsional vibration to cross-couple into lateral vibrations [1]. A common method to control vibrations involves the use of dynamic absorbers [2]. However, when the physical parameters of dynamic absorbers are constant, their frequency range efficiency is tight and thus not really suitable for variable speeds systems. To remedy this, a damping could be added to the absorber, but this would lead to a loss of efficiency at the main frequency and increases the transmissibility on a large frequency range. On the other hand, due to their variable properties, smart materials may be used to increase the frequency range in order to control vibration. Electrorheological (ER) fluids are attractive materials that undergo very fast reversible changes in their rheological properties (mainly their yield stress and effective viscosity) upon the application of electric fields [3]. This fluid may potentially be applied in many industrial areas [4]. They may be used to control vibrations [5], to control valves [6] or rotor squeeze film dampers [7]. ER fluid may also be applied to control active suspensions [8] or seismic vibrations [9], through the application of high electric fields. Previous studies conducted by the authors have allowed the development of a new very efficient Electro-Rheological (ER) fluid [10]. In this paper, we propose the design of a new dynamic torsional absorber by using ER fluid for a rotor system. The smart absorber is controlled by varying the electric field applied to the fluid.

2. ER smart dynamic torsional absorber

2.1 Model of ER fluid

In general, most ER fluid exhibit Bingham rheological behaviors and shear-thinning. In previous studies, the authors developed a model call the Quasi-Bingham model to explain this complex phenomenon (Sun *et al*., 2009, 2010), which is:

$$\begin{aligned}
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} \quad or \quad \tau = \eta_\infty \dot{\gamma} && when \quad E = 0 \\
&\tau = \tau_{yQB} + \eta_0 \dot{\gamma} && when \quad E \neq 0 \quad \dot{\gamma} \leq \dot{\gamma}_1 \\
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} + (\eta_0 - \eta_\infty)\frac{\dot{\gamma}}{1+(t\dot{\gamma})^n} && when \quad E \neq 0 \quad \dot{\gamma}_1 \leq \dot{\gamma} \leq \dot{\gamma}_2 \\
&\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} && when \quad E \neq 0 \quad \dot{\gamma} \geq \dot{\gamma}_2
\end{aligned} \tag{1}$$

where τ is the shear stress (Pa); $\dot{\gamma}$ is the shear rate; τ_{yQB} is the elastic yield stress (Pa) which is dependent on the strain and on the applied electric field E.

According to the dielectric loss model of Hao *et al*. [11], the yield stress τ_{yQB} is a function of E^2; η_0 is the zero shear viscosity; it is defined as the value at a very low shear rate and is a function of E; η_∞ is the infinite shear viscosity; it is defined as the value at very high shear rate; the parameter "n" is known as the Cross Rate Constant [12]. It is a dimensionless factor and is a measurement of the degree of dependence of viscosity on shear rate in the shear-thinning region; "t" is known as the Cross Time variable and has a dimension of time. The reciprocal, 1/t, gives us a critical shear rate that is a useful indicator of the onset shear rate for shear thinning; $\dot{\gamma}_1, \dot{\gamma}_2$ are two critical shear rates. We can obtain the values by using the two equations:

$$1+\left(t\dot{\gamma}_1\right)^n \approx 1; \quad \dot{\gamma}_2 \gg 1/t \tag{2}$$

This model can explain very well the rheological behaviors. An empirical equation (equation (3) of stresses was developed for a mixture called ETSERF40-20 as an example of the application of the Quasi-Bingham model. This equation is very useful for modeling a control system with the ER fluid [13].

$$\tau = \tau_{yQB} + \eta_\infty \dot{\gamma} + (\eta_0 - \eta_\infty)\frac{\dot{\gamma}}{1+(t\dot{\gamma})^n} \tag{3}$$

$$\tau \quad = 27.04\times10^{-9} E^2 + 0.218\dot{\gamma} + 29.4\times10^{-3} E \frac{\dot{\gamma}}{1+t\dot{\gamma}} \tag{4}$$

The electric field E is expressed in kV/mm.
Equations 3 and 4 may be replaced by Equation (5) accordingly with the following relationships:

$$\tau \quad = \alpha E^2 + \eta_\infty \dot{\gamma} + \beta E \frac{\dot{\gamma}}{1 + t\dot{\gamma}} \tag{5}$$

$$\tau_{yQB} = \alpha E^2 ; \eta_0 = \beta E + \eta_\infty \tag{6}$$

The α , β and η_∞ are intrinsic values of the ER fluid to be experimentallv determined. The field-dependent yield stresses of these ER fluids were experimentally obtained by $\tau_{yQB} = 27.04 \times 10^{-3} E^2$ Pa. The viscosity dynamic was experimentally obtained by $\eta_0 = 29.4E + 0.218$ for fluid ETSERF40-20. This rheological model allows for the exploration of the suitability of ER fluids to control the torsional vibrations of rotors through simulations.

2.2 ER torsional absorber design

A cylindrical type of ER torsional absorber is designed and manufactured according to the required damping force level and a critical velocity of a rotor system. Figure 1 describes the prototype of the proposed ER torsional absorber, which consists in an outer cylinder for the rotor and an inner cylinder for the absorber with the ER fluid enclosed between both cylinders. The ER fluid is composed of diatomite (called ETSERF40-20) mixed into silicone oil and is used to study the effect of ER fluids on the dynamic absorber [10] , [13].

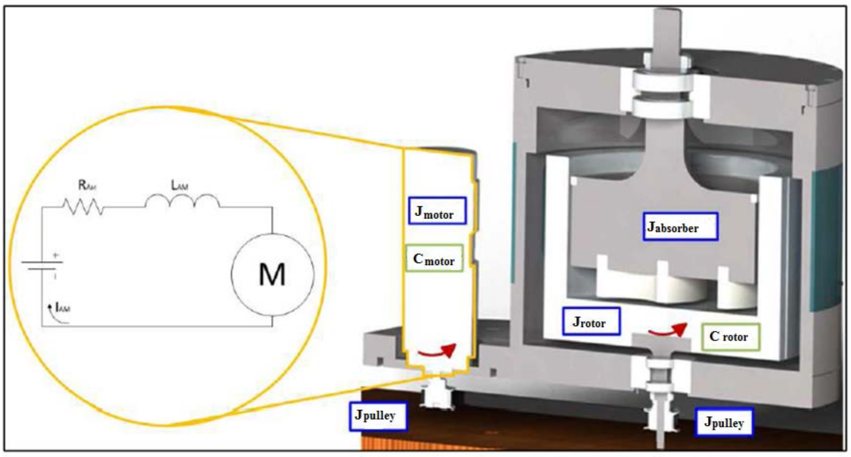


Figure 1. Overview of the torsional dynamic absorber

The positive (+) voltage is connected to inner cylinder and the negative voltage (-) connected to the outer cylinder. The proposed ER absorber is detailed accordingly with the following design parameters; electrode length (H)= 62.5mm and electrode gap (e)= 4.4mm. In the absence of electric field, the ER absorber produces a damping force only due to the fluid resistance when rotating. This damping (called C_2) of ER fluid can be calculated from equation (7) [13]. It is inherent, and is not influenced by the electric field.

$$C_2 = \frac{2\pi R^3 H \eta_\infty}{e} \tag{7}$$

R is the internal radius of outer cylinder, η_∞ is the viscosity at infinite velocity that is intrinsic value of the ER fluid to be experimentally determined. Applied to the considered fluid ETSERF40-20, we found C_2 = 0.0092 N.m.s/rad.

If a certain level of the electric field is applied to the ER absorber, the ER absorber produces an additional damping force. This damping force can be continuously tuned by controlling the intensity of the electric field. This system is then called semi-active.

2.3 Mechanical model of the smart system

2.3.1 Dynamic mode of a rotor system with the ER absorber

The vibration of the initial system may be described as a single degree-of-freedom (SDOF) damped rotating system, composed of the rotational inertia J_1 = 0.116 kg.m^2, the torsional stiffness K_1 = 801.51 N.m/rad, and the torsional damping C_1 =2.192 N.m.s/rad. The theoretical natural frequency of this primary system is 13.23 Hz with a damping rate of 11.37%. It is the amplitude close to this frequency that must be controlled.

This vibration control method in this system consists in adding another SDOF system, which is the torsional dynamic absorber. The smart ER dynamic absorber is composed of a rotational inertia J_2 = 0.0265kgm^2 driven by the ER fluid, which can thus be modeled by a torsional spring K_2 , a linear viscous dashpot C_2 and a viscous dashpot C_{ER} . The resulting mechanical system is a two DOF system which is illustrated in Figure 2.

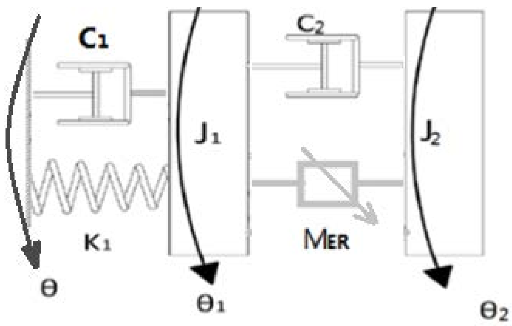


Figure 2. ER torsional dynamic absorber

The torque M_{ER} which is due to the effect of the ER fluids is composed of a nonlinear torsional stiffness torque and a nonlinear viscous torque, which depend on the electric field. It is only active when the electric field is applied. It can be expressed as:

$$M_{ER} = C_{ER}(\dot{\theta}_1 - \dot{\theta}_2) + K_{ER}(\theta_1 - \theta_2) \tag{8}$$

The total torque of the proposed ER absorber can be obtained as $M_F = C_2\dot{\theta} + M_{ER}$, where C_2 is the effective damping due to the fluid viscosity, θ is the excitation angular displacement, and M_{ER} is the field-dependent damping torque which is tunable as a function of the applied electric field.

The rotor dynamic system could be described as follows:

$$\begin{aligned} J_1\ddot{\theta}_1 &= -M_{ER} - C_1(\dot{\theta}_1 - \dot{\theta}) - K_1(\theta_1 - \theta) - C_2(\dot{\theta}_1 - \dot{\theta}_2) \\ J_2\ddot{\theta}_2 &= -C_2(\dot{\theta}_2 - \dot{\theta}_1) + M_{ER} \end{aligned} \tag{9}$$

2.3.2 Modeling of the system by state space formulation

Equation 9 can be expressed as the following equations:

$$
\begin{aligned}
&J_1\ddot{\theta}_1 + K_1\theta_1 + C_1\dot{\theta}_1 + C_2(\dot{\theta}_1 - \dot{\theta}_2) = K_1\theta + C_1\dot{\theta} - M_{ER}\\
&J_2\ddot{\theta}_2 + C_2(\dot{\theta}_2 - \dot{\theta}_1) = M_{ER}\\
&\ddot{\theta}_1 = \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - C_1\dot{\theta}_1 - C_2(\dot{\theta}_1 - \dot{\theta}_2))\\
&\ddot{\theta}_2 = \frac{1}{J_2}(M_{ER} - C_2(\dot{\theta}_2 - \dot{\theta}_1))
\end{aligned}
\tag{10}
$$

By modifying the variables as follows:

$$
\begin{aligned}
&x_1 = \theta_1 - h_0\theta\,;\\
&x_2 = \dot{x}_1 - h_1\theta = \dot{\theta}_1 - h_0\dot{\theta} - h_1\theta\\
&x_3 = \theta_2\,;\\
&x_4 = \dot{x}_3 = \dot{\theta}_2
\end{aligned}
\tag{11}
$$

We can write:

$$\theta_1 = x_1 + h_0\theta\,;\ \dot{x}_1 = x_2 + h_1\theta \tag{12}$$

We obtain:

$$\dot{x}_1 = x_2 + h_1\theta \tag{13}$$

$$\dot{x}_2 = \ddot{\theta}_1 - h_0\ddot{\theta} - h_1\dot{\theta} = \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - C_1\dot{\theta}_1 - C_2(\dot{\theta}_1 - \dot{\theta}_2)) - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}(K_1\theta + C_1\dot{\theta} - M_{ER} - K_1\theta_1 - (C_1 + C_2)\dot{\theta}_1 + C_2\dot{\theta}_2) - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}\left\{K_1\theta + C_1\dot{\theta} - M_{ER} - K_1(x_1 + h_0\theta) - (C_1 + C_2)(x_2 + h_0\dot{\theta} + h_1\theta) + C_2x_4\right\} - h_0\ddot{\theta} - h_1\dot{\theta}$$

$$= \frac{1}{J_1}\left\{K_1 - K_1h_0 - (C_1 + C_2)h_1\right\}\theta + \left\{\frac{1}{J_1}\left[C_1 - (C_1 + C_2)h_0\right] - h_1\right\}\dot{\theta} - h_0\ddot{\theta}$$

$$+ \frac{1}{J_1}\left\{-M_{ER} - K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\right\}$$

Take $h_0 = 0$ and $\frac{1}{J_1}\left[C_1 - (C_1 + C_2)h_0\right] - h_1 = 0$, we can obtain $h_1 = \frac{C_1}{J_1}$

We have:

$$
\begin{aligned}
&\dot{x}_1 = x_2 + h_1\theta = x_2 + \frac{C_1}{J_1}\theta\\
&\dot{x}_2 = \frac{1}{J_1}\left\{K_1 - K_1h_0 - (C_1 + C_2)h_1\right\}\theta + \left\{\frac{1}{J_1}\left[C_1 - (C_1 + C_2)h_0\right] - h_1\right\}\dot{\theta} - h_0\ddot{\theta}\\
&+ \frac{1}{J_1}\left\{-M_{ER} - K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\right\}\\
&= \frac{1}{J_1}\left\{K_1 - (C_1 + C_2)\frac{C_1}{J_1}\right\}\theta - \frac{1}{J_1}M_{ER} + \frac{1}{J_1}\left\{-K_1x_1 - (C_1 + C_2)x_2 + C_2x_4\right\}\\
&\dot{x}_3 = x_4
\end{aligned}
\tag{14}
$$

$$\dot{x}_4 = \frac{C_1 C_2}{J_1 J_2}\theta + \frac{1}{J_2}M_{ER} + \frac{1}{J_2}\{C_2 x_2 - C_2 x_4\}$$

By regrouping these equations under a matrix form, we obtain:

$$\begin{pmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{K_1}{J_1} & -\frac{(C_1+C_2)}{J_1} & 0 & \frac{C_2}{J_1} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{C_2}{J_2} & 0 & -\frac{C_2}{J_2} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} + \begin{pmatrix} \frac{C_1}{J_1} & 0 \\ \frac{1}{J_1}\left\{K_1 - (C_1+C_2)\frac{C_1}{J_1}\right\} & -\frac{1}{J_1} \\ 0 & 0 \\ \frac{C_1 C_2}{J_1 J_2} & \frac{1}{J_2} \end{pmatrix} \begin{pmatrix} \theta \\ M_{ER} \end{pmatrix} \tag{15}$$

$$x = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix}^T$$

$$\dot{x} = \begin{pmatrix} \dot{x}_1 & \dot{x}_2 & \dot{x}_3 & \dot{x}_4 \end{pmatrix}^T$$

$$u = \begin{pmatrix} \theta & M_{ER} \end{pmatrix}^T$$

By calling these matrices as:

$$y = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \tag{16}$$

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\frac{K_1}{J_1} & -\frac{(C_1+C_2)}{J_1} & 0 & \frac{C_2}{J_1} \\ 0 & 0 & 0 & 1 \\ 0 & \frac{C_2}{J_2} & 0 & -\frac{C_2}{J_2} \end{pmatrix}, \quad B = \begin{pmatrix} \frac{C_1}{J_1} & 0 \\ \frac{1}{J_1}\left\{K_1 - (C_1+C_2)\frac{C_1}{J_1}\right\} & -\frac{1}{J_1} \\ 0 & 0 \\ \frac{C_1 C_2}{J_1 J_2} & \frac{1}{J_2} \end{pmatrix}$$

$$C = \begin{pmatrix} 0 & 1 & 0 & 0 \end{pmatrix}$$

We obtain the classic state space formulation:

$$\dot{x} = Ax + Bu$$
$$y = Cx$$

We consider a simple and linear case, in order to establish an ON-OFF control. The block diagram is shown in Fig. 3.

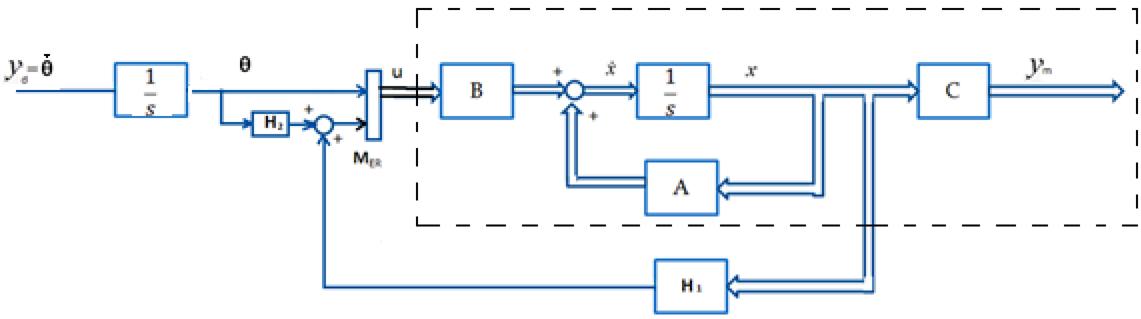


Figure 3. Block diagram du system

By defining

$$H_1 = \left(K_{ER} \quad C_{ER} \quad -K_{ER} \quad -C_{ER} \right); \quad H_2 = \frac{C_{ER} C_1}{J_1} \quad ; \tag{17}$$

The controllable torque may be expressed as:

$$M_{ER} = K_{ER}(x_1 - x_3) + C_{ER}(x_2 - x_4 + \frac{C_1}{J_1}\theta)$$
$$M_{ER} = K_{ER} x_1 - K_{ER} x_3 + C_{ER} x_2 - C_{ER} x_4 + \frac{C_{ER} C_1}{J_1}\theta = H_1 x + H_2 \theta \tag{18}$$

H_1 *and* H_2 can be controlled by the electrical field by checking the difference of the angular velocity between $y_d = \dot{\theta}$ and $y_m = \dot{\theta}_m$. If the difference of the absolute value of is greater than zero, we must apply the electric field; otherwise we set the electric field to zero.

3. Numerical simulation

The control system with an Dynamic absorber (SERDA) has been numerically simulated by using Simulink (figure 4), in order to observe the reaction of the system, we use a linear chirp signal which is sine wave whose frequency varies linearly with time (figure 5), the amplitude is 1. We select E = 0 and 1 kv/mm in this simulation, the results is showed in the figure 6. We can see that amplitude of the response varies with the frequency; the amplitude is very small when the frequency is down 13Hz or more than 13Hz, but it arrives the maximum value, when the frequency is around 13Hz without electric fields and with the electric field, the amplitude decreases to about half of the original amplitude when electric fields applied is 1 kv/mm.

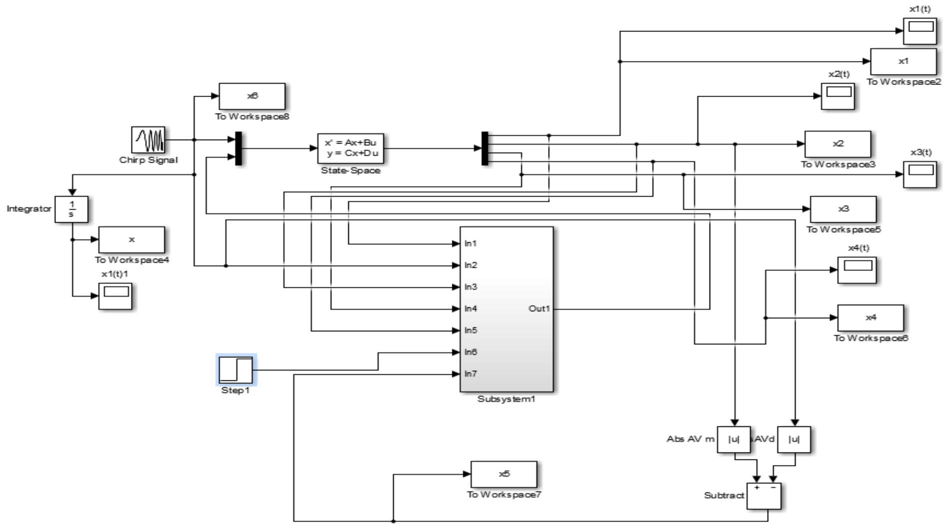


Figure 4. Block Simulink

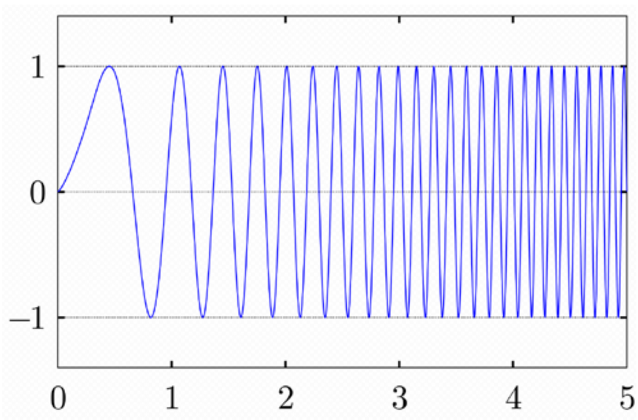


Figure 5. Linear chirp waveform

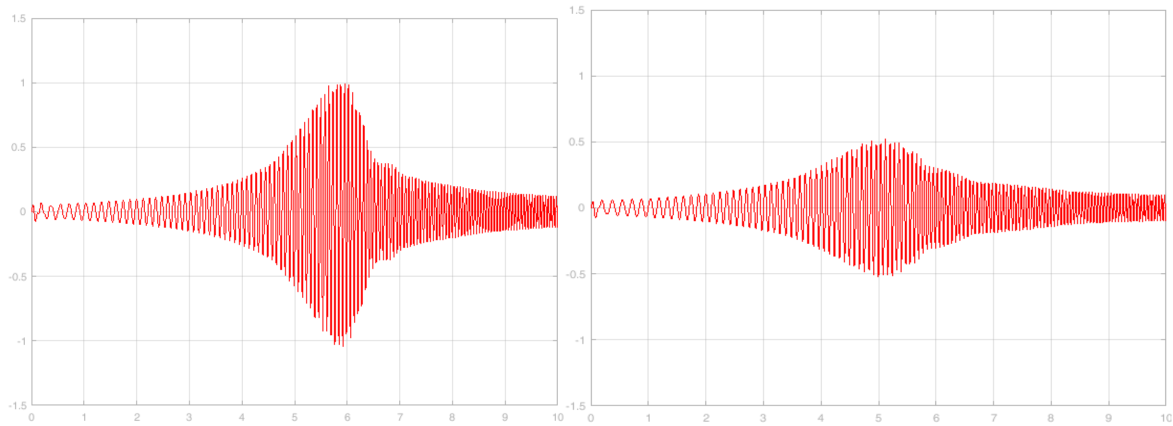


Figure 6.Reponse of system under a chirp signal obtained by simulation (left figure: E=0kV/mm; right figure : E-1kV/mm)

4. Experimental verification

In this system, the two cylinders J_1 and J_2 serve as electrodes for the ER effect in this study. The oscillation caused by the torsional vibration is measured by a triaxial accelerometer placed in the radial position as shown in Figure.6.

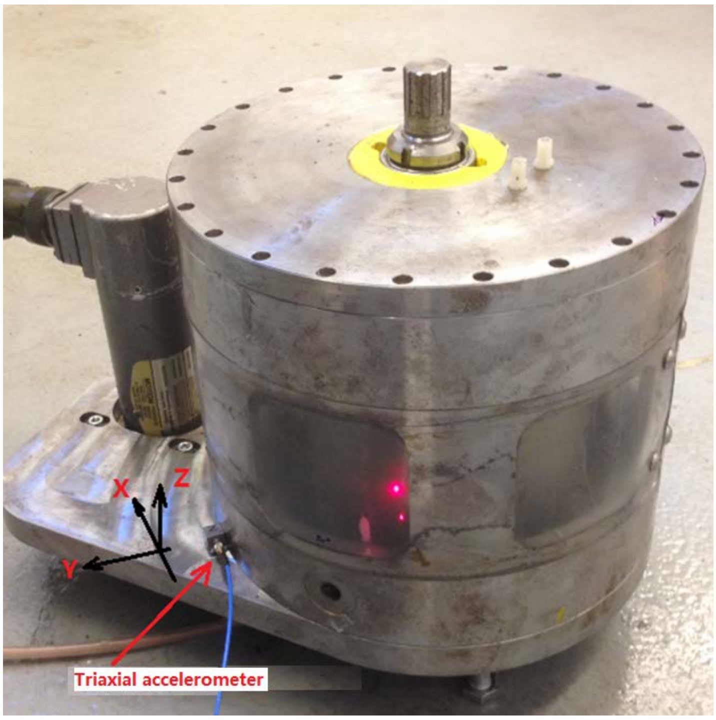


Figure 7. Electrorheological torsional absorber

We clearly see that vibration control performance in figures 8, 9 and 10, especially at 13 Hz, since the natural torsional frequency of the system is designed for this value as 13.24Hz, the critical angular velocity is 794 RPM. The two peaks as shown in Figure 8 are two natural frequencies, one is of flexible vibration, and the other is of vibration torsional. In fact, this absorber is designed only to attenuate the torsional vibration to determine the effects of the fluid ER in torsional vibration. Clearly, the torsional vibrations are attenuated in the range of 12.5 to 15Hz, but the peak which is at 12Hz does not change anything, it shows that this absorber is effective for torsional vibration, but not for flexible vibration. It can be noticed that the results of simulation are very comparable with the experimental results.

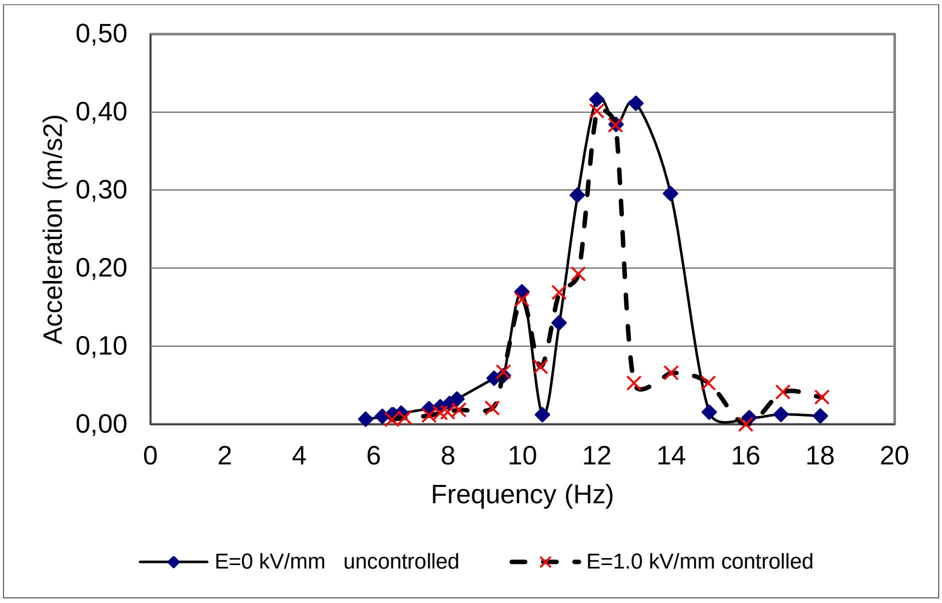


Figure 8.Response acceleration of system under different electrical field

The measurement results in time and frequency domain are shown in figure 9 and 10 for different applied electric fields. The amplitude of sinusoidal vibration decreases when the applied field increases, but not for the high electric field as E = 1.36 kv / mm since there is an optimal value of electric fields that applies on ER fluid to attenuate the torsional vibrations when a simple On-Off control strategy is chosen to realize this control system.

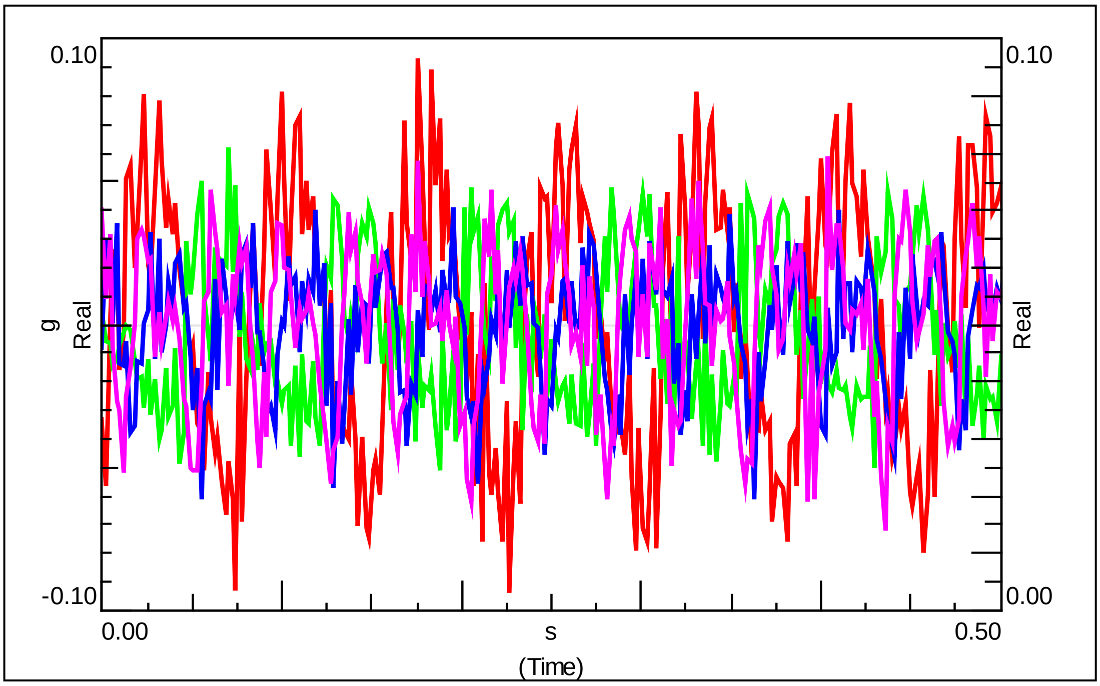


Figure 9.Temporal response of system when the rotor rotates in 780 RPM under different electrical field (red: E=0kV/mm; green: E=0.68 kV/mm; blue: E=1kV/mm; pink: E=1.36kV/mm)

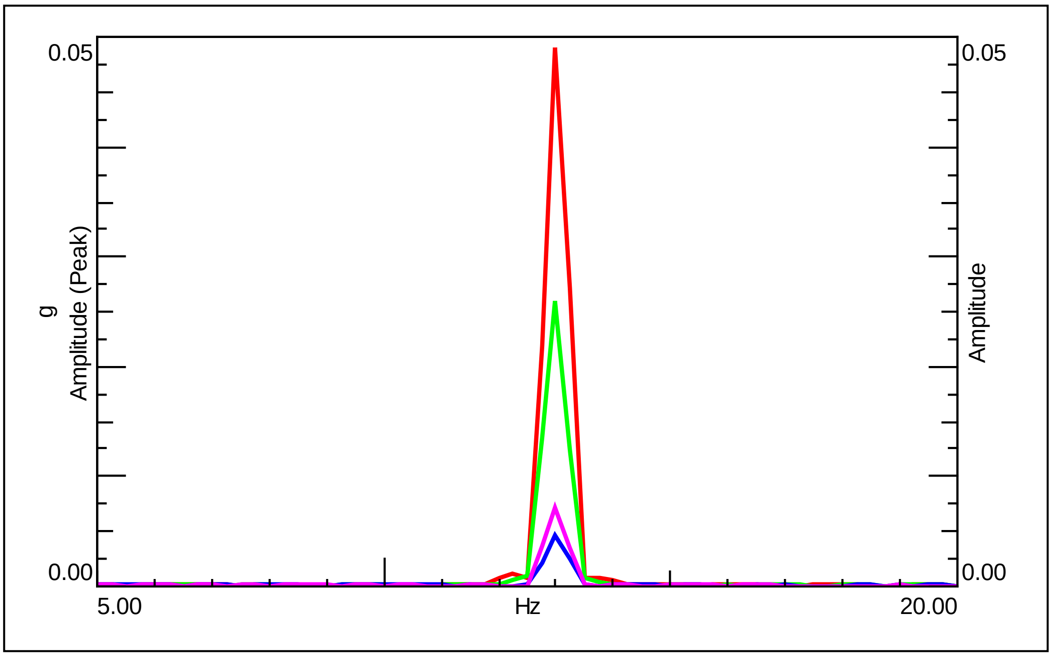


Figure 10. Frequency response of system when the rotor rotates in 780 RPM under different electrical field (red: E=0kV/mm; green: E=0.68 kV/mm; blue: E=1kV/mm; pink: E=1.36kV/mm)

5. Conclusion

This article is concerned with the application of Electrorheological Fluids (ERF) to the reduction of torsional vibrations of a rotor by controlling damping and stiffness of a rotational dynamic absorber. A cylindrical type of electrorheological (ER) torsional absorber was designed and manufactured according to the required damping force level and the critical velocity of the rotor system. This paper presents torsional vibration control performance of a smart ER dynamic absorber using a bang-bang (ON-OFF) control strategy. The absorber efficiency is measured, and the results show that the ER dynamic absorber exhibits very good performances in terms of reducing the torsional vibration of rotor system. The experimental results very closely approximate the simulation results. By use of controlled damping one can reduce (minimize) unwanted torsional vibration not only in theory but also in practice, this study shown that ER fluid are useful smart materials to realize a controllable system.

References

[1] Troy; Hill, Charles L. (2009). *Prevention Of Torsional Vibration Problems In Reciprocating Machinery*. Texas A&M University. Turbomachinery Laboratories.

[2] Den Hartog J.P. (2008), *Mechanical vibrations*, Crastre Press, ISBN 1443725366, 496 p.

[3] Winslow W.M., 1949. *Induced vibration of suspensions.* Journal of applied physics. **20**(12): p. 1137-1140.

[4] Block H. and Kelly J.P., 1988. *Review Article Electro-rheology*. Journal of Physics D: Applied Physics. **21**: p. 1661-1667.

[5] Stanway R., Sprostonz J.L., and El-Wahedz A.K., 1996. *Applications of electro-rheological fluids in vibration control: a survey*. Smart Materials and Structures. 5: p. 464-482.

[6] Abed A., Bouzidane A., Thomas M. and Zahloul H., December 2016, *Performance characteristics of a three-pad hydrostatic squeeze film damper compensated with new electrorheological valve restrictors*. Journal of Engineering Tribology (vols 208-210) 1994-1996.

[7] Bouzidane A. and Thomas M., July 2013. *Nonlinear Dynamic Analysis of a Flexible Shaft Supported by Smart Hydrostatic Squeeze Film Dampers*, Journal of Tribology (ASME), 135(3), 031701-1, 031701-9.

[8] Choi S. B., Y. M. Han, and all, 2007. *Field Test on Vibration Control of Vehicle Suspension System Featuring ER Shock Absorbers*, Electrorheological Fluids and Magnetorhelogical Suspensions, pp. 496-503 (2007)

[9] Symans M. D., M. C. Constantinou, 1999. *Semi-active control systems for seismic protection of structures: a state-of-the-art review*, Engineering Structures 21 (1999) 469–487

[10] Sun Y., Thomas M. and Masounave J., 2004. *Investigation of Dynamic Behavior of Electrorheological Fluids by Experimental Design*, Proceedings of the 7th Cansmart International workshop on Smart Materials and structures, Montreal, pp 153-162.

[11] Hao T., A. Kawai, and F. Ikazaki, 2000. *The Yield Stress Equation for the Electrorheological Fluids*. Langmuir, 16: pp. 3058-3066.

[12] Cross, M.M., 1965. *Rheology of Non Newtonian Fluids: A New Flow Equation for Pseudoplastic Systems*. Journal Colloid Science. 20 pp.417-437.

[13] Sun Y., Thomas M. and Masounave J., (2010). *A quasi-Bingham model for predicting electrorheological fluid behaviour*, Multidiscipline Modeling in Materials and Structures. **6** (1), 141-165.

Shunted piezoelectrical flextensionnal suspension for vibration insulation

K. BILLON [1], N. MONTCOUDIOL [1], A. AUBRY [2], R. PASCUAL [2], F. MOSCA [2], F. JEAN [2], C. PEZERAT[3], C. BRACAULT [3], S. CHESNÉ [1]

[1]Univ Lyon, INSA-Lyon, CNRS UMR5259, LaMCoS, F-69621, France
simon.chesne@insa-lyon.fr

[2]PYTHEAS Technology, 100 Impasse des Houllières, ZA Le Pontet, F-13590, France

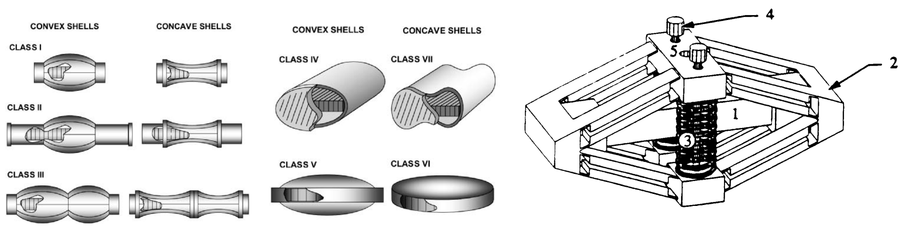
[3]Laboratoire d'Acoustique de l'Université du Mans - UMR CNRS 6613, F-72085, France

Abstract

The objective of the PyDAMP project is to develop a hybrid mechanical suspension to reduce the vibrations transmission on a wide frequency band. The undesired vibrations are generated by small electric motors (few kilograms). A suspension with piezoelectric pillar developped by PYTHEAS Technology is compared to a conventional viscoelastic suspension in terms of performances in the audible frequency range. The principle and design of the piezoelectrical suspension are approached through an electromechanical model and a finite element model. The electromechanical coupling of the transducer allows the introduction of mechanical damping and electric damping with different shunts based on resistor and negative capacitance.

1 The piezoelectrical suspension: principle and conception

The concept of suspension is inspired by a Class IV flextensionnel transducer [1, 2, 3] (figures 1). Flextensional transducers are a class of mechanical amplifiers composed of an active part, usually piezoelectric (bars, discs, rings), or magnetostrictive, and a shell that radiates in the surrounding fluid[4, 5, 6].



(a) (b)

Figure 1: a) Sketch of various classes of flextensional transducers [1, 2]. b) Sketch of a Class IV flextensional piezoelectric transducer. [3]

A finite element study has been achieved to ensure the validity of the concept in terms of maximum admissible Von Mises stress, maximum displacement and modes shapes. Figure 2 shows the CAD view of the piezoelectric suspension.

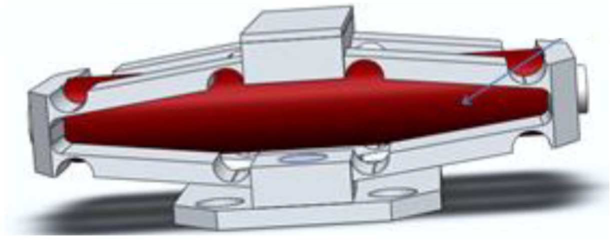


Figure 2: CAD view of the piezoelectric suspension.

2 Electrical model and simulations

An electromechanical model of the piezoelectrical suspension has been developed. Mechanical elements are converted in electrical components and an equivalent electrical circuit can be found. The simulation and the shunt optimisation are facilitated with only one physic, taking into account the whole dynamic behaviour of the piezoeletrical suspension. Figure 3 shows the schema of the piezoelectrical suspension.

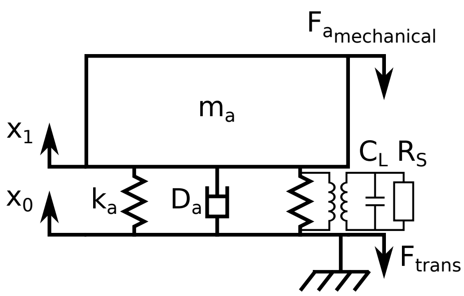


Figure 3: Schema of the piezolelectrical suspension.

The equivalent electrical circuit of the piezoelectrical suspension with mechanical excitation and resistor shunt is shown on figure 4.

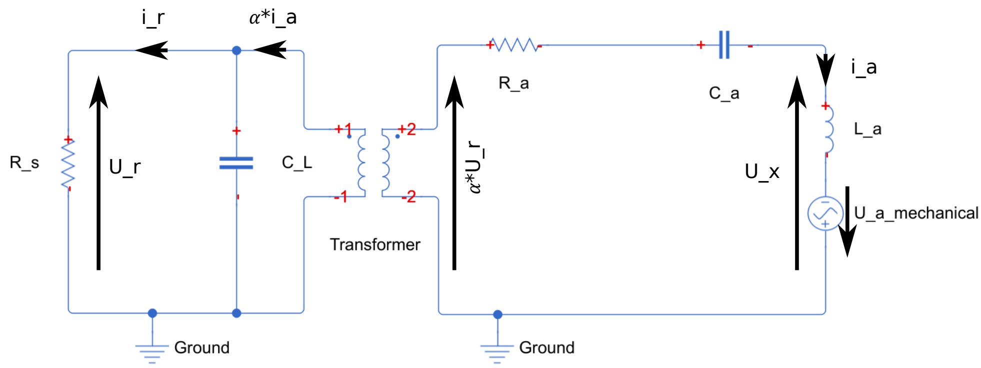


Figure 4: Equivalent electrical circuit of the piezoelectrical suspension.

Figure 5 shows the displacement transmissibility functions in open-circuited, in short-circuited conditions, with resistor shunt and with negative capacitance shunt. The damping (ξ) in open-circuited, in short-circuited conditions is similar, equal to 0.3%. As expected, the surtension is reduced by the resistor shunt, the damping (ξ) is equal to 1.6%. The negative capacitance increases the performances of the suspension to reach 2.9% of damping.

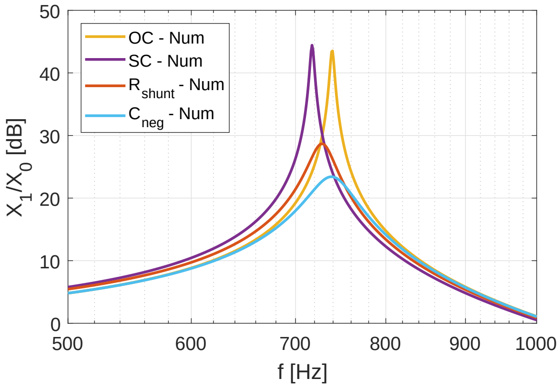
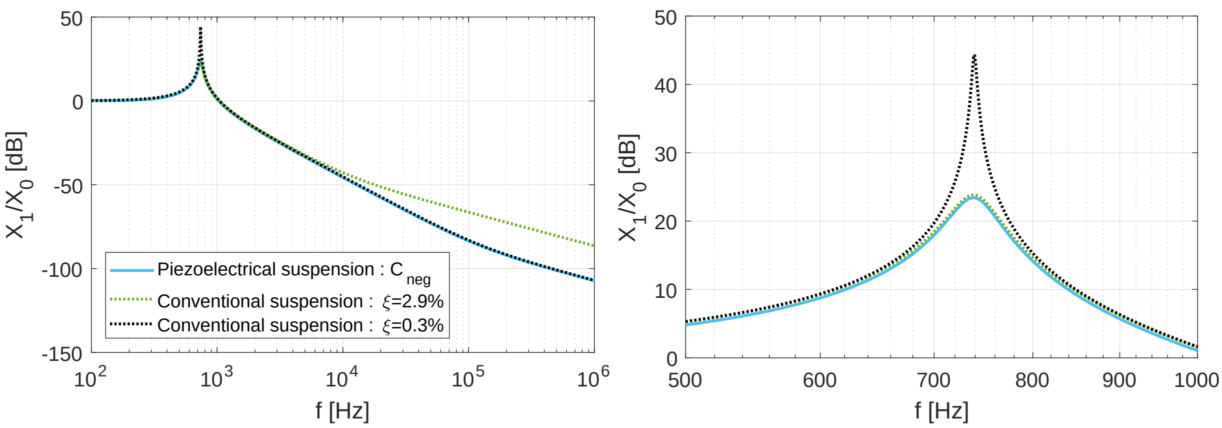


Figure 5: Transmissibility functions.

3 Comparison with a conventional viscoelastic suspension

A comparison with a conventional viscoelastic suspension is done. The conventional viscoelastic suspension is a classical spring (k_a), mass (m_a), damper (D_a). Two damping values are tested and compared with the piezoelectrical suspension namely $\xi = 0.2\%$ and 2.9%. Figures 6 show the displacement transmissibility functions with negative capacitance shunt (blue curves) compared to conventional viscoelastic suspension with damping equal to $\xi = 2.9\%$ (green curves) and 0.2% (black curves). Figures 6 show the comparison with equivalent damping. Differences can be observe in high frequencies from 10 kHz where the slope moves from $-40\ dB/dec$ to $-20\ dB/dec$ for the conventional viscoelastic suspension. For the piezoelectric suspension, the change appears a decade later. For audible perturbation between 10 kHz and 20 kHz, in this case the difference is between 2 dB and 6 dB. Figures 6 show the comparison where the slope moves from $-40\ dB/dec$ to $-20\ dB/dec$ at the same frequency. Differences can be observe around 700 Hz, for the conventionnal suspension, the maximum transmission level is twice as high compared to the piezoelectrical suspension.



(a) (b)

Figure 6: a) Comparison between the transmissibility functions of a conventional viscoelastic suspension and the piezoelectrical suspension. b) Zoom.

4 Experimental setup

Figure 7 shows the experimental setup. Two accelerometers are used to obtain the transmissibility functions. The mechanical excitation is provided by a shaker. The piezoelectrical suspension is glued on the shaker. Different electrical conditions can be applied on the stack. Experimental results are under post-processing.

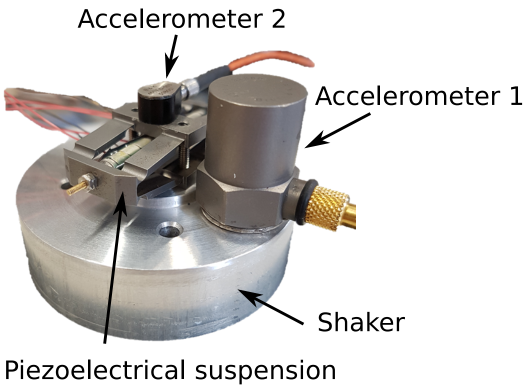


Figure 7: Experimental setup

5 Conclusion

A suspension with piezoelectric pillar developed by PYTHEAS Technology has been studied. An electromechanical model and a finite element model allowed the conception and the design of the suspension. A resistor shunt optimisation has been performed and good results were observed in order to reduce the surtension in the displacement transmissibility. Performances have been increased using a negative capacitance shunt. A comparison with conventional suspension has been conducted. For the piezoelectric suspension, no static problems are observed and there is no creep behaviour. The slope modification from $-40\ dB/dec$ to $-20\ dB/dec$ appears a decade later and is not sensitive to damping. It could be important for audible applications.

References

[1] J. Butler and C. Sherman, *Transducers and arrays for underwater sound*, Springer, 2016.

[2] D. Jones and N. Abboud, "Performance analysis of a low-frequency barrel-stave flextensional projector," in *Proceeding of ONR Transducer Materials and Transducers Workshop*, pp. 1–4, 1996.

[3] N. Hagood, J. Aldrich, and A. V. Flotow, *Design of passive piezoelectric damping for space structures*, National Aeronautics and Space Administration, Langley Research Center, 1994.

[4] B. Damiri, *Modèle analytique pour l'analyse du transducteur flextensionnel de classe IV dans l'air*. PhD thesis, Lille 1, 1996.

[5] G. Brigham and L. Royster, "Present status in the design of flextensional underwater acoustic transducers," *The Journal of the Acoustical Society of America* **46**(1A), pp. 92–92, 1969.

[6] L. Royster, "Predicting the effective dynamic coupling coefficient of the flextensional transducer (preliminary design considerations)," *The Journal of the Acoustical Society of America* **38**(5), pp. 879–880, 1965.

Exploring periodicity and dispersion diagrams in muffler design

Vinícius D. Lima[1], Belisário N. Huallpa[2] and José R. F. Arruda[1]

[1]University of Campinas, Rua Mendeleyev, 200, Campinas, SP, Brazil
vinidiaslima@gmail.com, arruda@fem.unicamp.br
[2]Universidade Federal de Lavras, Lavras, Aquenta Sol, MG, Brazil
belisario.nina@deg.ufla.br

Abstract

Periodic waveguides may be analyzed using dispersion diagrams, which plot the wavenumbers as functions of frequency. Imaginary wavenumbers mean propagation is not possible and, therefore, normal modes cannot build up. Muffler design has traditionally explored periodicity, but usually not using dispersion analysis. In this work, we show how to model one-dimensional acoustic waveguides with plane wave assumption using spectral elements (SE), how to obtain dispersion diagrams and, using semi-infinite elements, transmission loss from an SE model. The technological challenge consists in opening band gaps at low frequencies with a limited size muffler, and SE models are handy for low cost parameter optimization. For arbitrary shapes, this work uses scaled SE models or, alternatively, a state-space formulation recently developed by the authors. Additive manufacturing is an enabling technology for the implementation of the designed mufflers. In this work, we show experimental results for simple periodic mufflers built using 3D printing. The proposed simulation methodology is simple and can be used for quick design of 3D-printed polymer mufflers.

1 Introduction

Acoustic mufflers are usually characterized by their insertion loss (IL) or transmission loss (TL) [1]. The latter is a more absolute characterization, as it does not depend upon the acoustic impedance of the system connected to the muffler. Mufflers can be reactive, active of hybrid depending on the sound attenuation mechanism. Reactive mufflers do not need to dissipate energy and attenuate using acoustic impedance discontinuities and acoustic resonators that cause destructive interference. Active mufflers attenuate propagating sound by energy dissipation. Mufflers may combine these two mechanisms. The design of reactive sound mufflers usually start from a chosen geometry combining Helmholtz resonators, quarter length tubes and expansion chambers and adjusting their parameters to maximize the TL at desired frequency ranges.

In recent years, inspired by the scientific advances in photonics [2], researchers started using wave dispersion diagrams of periodic elastic systems to investigate the existence of frequency stop bands created by destructive interference (Bragg scattering) and local resonance due to resonators. This is exactly the effect desired in reactive mufflers. Therefore, in this work, we investigate the use of dispersion diagrams to design periodic reactive mufflers.

Dispersion diagrams are usually obtained using the Plane Wave Expansion (PWE) method [3]. Otherwise, given a transfer matrix of a periodic cell, the Floquet-Bloch theorem can be used to obtain the wavenumbers as a function of frequency. In this work we derive a spectral acoustic element to model the linear acoustics of ducts using the plane wave assumption. It is straightforward to derive a semi-infinite acoustic duct element. Using finite and semi-infinite spectral elements a model can be easily built to compute the TL and simulate the forced responses.

We first introduce the acoustic spectral elements, show how to assemble these elements, how to reduce the global matrix and transform it into a transfer matrix, and, finally, how to compute the dispersion diagram. Then we show how to assemble a duct system consisting of source, uniform duct, periodic muffler and semi-infinite duct that allows computing the TL. We show that with a small number of periodic cells a high TL is achieved. Finally, experimental results obtained in an impedance tube for a 3D-printed periodic muffler are shown.

2 Analytical Formulation

2.1 Acoustic spectral element

The derivation of the acoustic duct spectral element is analogous to the spectral element for elementary rods [4]. Assuming plane wave propagation, the spectral element for an acoustic waveguide starts from the one-dimensional non-dissipative wave equation [5]

$$\frac{\partial^2 p}{\partial x^2} - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \tag{1}$$

where p is the acoustic pressure at a position x along the waveguide and c is the speed of sound. In the frequency domain, Eq. 1 becomes an ordinary differential equation

$$\frac{\partial^2 \hat{p}}{\partial x^2} + \frac{\omega^2}{c^2}\hat{p} = 0 \tag{2}$$

Equation 2 has a wave-based solution form given by:

$$\hat{p} = C_1 e^{-ikx} + C_2 e^{-ik(L-x)} \tag{3}$$

where k is the wavenumber. For low sound levels, the acoustic pressure and the particle velocity u are related by the linear Euler's equation [5]

$$\rho\frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x} \tag{4}$$

where ρ is the mass density of the fluid. Replacing the solution given by Eq. 3 into Eq. 4 and rearranging in terms of the volume acceleration $\hat{v} = S\partial u/\partial t$ yields

$$\hat{v} = \frac{S}{\rho c}\left(-C_1 e^{-ikx} + C_2 e^{-ik(L-x)}\right) \tag{5}$$

where S is the cross-sectional area.

Considering a straight acoustic waveguide with length L, constants C_1 and C_2 can be determined for given boundary conditions at $x = 0$ and $x = L$. Finding the solution for these constants yields the following matrix system [4] [6]:

$$\begin{bmatrix} \hat{V}_l \\ \hat{V}_r \end{bmatrix} = \frac{ikS}{\rho(1-e^{-2ikL})}\begin{bmatrix} 1+e^{-2ikL} & -2e^{-ikL} \\ -2e^{-ikL} & 1+e^{-2ikL} \end{bmatrix}\begin{bmatrix} \hat{p}_l \\ \hat{p}_r \end{bmatrix} = \begin{bmatrix} D_{ll} & D_{lr} \\ D_{rl} & D_{rr} \end{bmatrix}\begin{bmatrix} \hat{p}_l \\ \hat{p}_r \end{bmatrix} = \mathbf{D}\begin{bmatrix} \hat{p}_l \\ \hat{p}_r \end{bmatrix} \tag{6}$$

where $\mathbf{D}$ is the spectral element matrix for the acoustic waveguide.

2.2 Dispersion diagram

An acoustic tube with periodically varying cross section ca be called a phononic crystal [2]. The band structure of a phononic crystal may be represented by a dispersion diagram, i.e., a plot of the wavenumber versus frequency (or vice-versa). A transfer matrix relates the acoustic pressure and the volume acceleration at the left and right ends of a periodic cell as

$$\begin{bmatrix} \hat{p}_r \\ -\hat{V}_r \end{bmatrix} = \mathbf{T}\begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} = \begin{bmatrix} -D_{ll}D_{lr}^{-1} & -D_{lr}^{-1} \\ D_{rl} - D_{ll}D_{rr}D_{lr}^{-1} & -D_{rr}D_{lr}^{-1} \end{bmatrix}\begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} \tag{7}$$

The transfer matrix of a periodic system with n cells can be obtained from the matrices of each cell $\mathbf{T_1}, \mathbf{T_2}, ..., \mathbf{T_{n-1}}, \mathbf{T_n}$ as

$$\mathbf{T} = \mathbf{T_n}\mathbf{T_{n-1}}...\mathbf{T_2}\mathbf{T_1} \tag{8}$$

The Floquet-Bloch theorem states the relation between the state vector at the right and left boundaries of the unit cell as

$$\begin{bmatrix} \hat{p}_r \\ -\hat{V}_r \end{bmatrix} = e^{-ikL} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} \tag{9}$$

Equations 7 and 9 yield the following eigenproblem:

$$\mathbf{T} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} = e^{-ikL} \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} = \lambda \begin{bmatrix} \hat{p}_l \\ \hat{V}_l \end{bmatrix} \tag{10}$$

The solution of the eigenproblem in Eq. 10 gives the normalized wavenumber kL for each angular frequency ω using the relation

$$kL = i\ln(\lambda) \tag{11}$$

2.3 Transmission loss

In this section we recall the two-load method for computing the TL with an impedance tube. The loudspeaker is positioned at the left extremity, and measurements are performed with two different terminations at the right end. The sample is positioned at the central position, with two microphones measuring the pressures at each side. With the positions of the microphones and the pressures, the amplitudes of the transmitted and reflected waves can be determined by

$$A = i\sqrt{G_{rr}} \frac{H_{1r}e^{ikx_2} - H_{2r}e^{ikx_1}}{2\sin k(x_1 - x_2)} \tag{12a}$$

$$B = i\sqrt{G_{rr}} \frac{H_{2r}e^{-ikx_1} - H_{1r}e^{-ikx_2}}{2\sin k(x_1 - x_2)} \tag{12b}$$

$$C = i\sqrt{G_{rr}} \frac{H_{3r}e^{ikx_4} - H_{4r}e^{ikx_3}}{2\sin k(x_3 - x_4)} \tag{12c}$$

$$D = i\sqrt{G_{rr}} \frac{H_{4r}e^{-ikx_3} - H_{3r}e^{-ikx_4}}{2\sin k(x_3 - x_4)} \tag{12d}$$

where G_{rr} is the autospectrum of the loudspeaker signal (there is no need to use the actual loudspeaker velocity, the voltage input signal is sufficient) and H_{nr} and x_n are the frequency response functions (FRF) for input r and output pressure measured with microphones $n = 1,2,3,4$. The pressures and particle velocities at the sample extremities are given by

$$p_0 = A + B \tag{13a}$$

$$u_0 = \frac{(A-B)}{\rho c} \tag{13b}$$

$$p_d = Ce^{-jkd} + De^{-jkd} \tag{13c}$$

$$u_d = \frac{Ce^{-jkd} - De^{-jkd}}{\rho c} \tag{13d}$$

Denoting with subscripts a and b the measurements with two different tube terminations (e.g. open and closed), the transfer matrix of the sample can be determined by

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \frac{1}{p_{da}v_{db} - p_{db}v_{da}} \begin{bmatrix} p_{0a}v_{db} - p_{0b}v_{da} & p_{0b}p_{da} - p_{0a}p_{db} \\ v_{0a}v_{db} - v_{0b}v_{da} & p_{da}v_{0b} - p_{db}v_{0a} \end{bmatrix} \tag{14}$$

The transmission loss of the sample is obtained by:

$$TL = 20\log_{10}\left|\frac{1}{2}\left(T_{11} + \frac{T_{12}}{\rho c} + \rho c T_{21} + T_{22}\right)\right| \tag{15}$$

In the cases where the sample is symmetric, another two conditions can be set, and only one load is necessary to determine the transfer matrix. These conditions are the reciprocity and symmetry given by

$$T_{11} = T_{22} \tag{16a}$$

$$T_{11}T_{22} - T_{12}T_{21} = 1 \tag{16b}$$

With the conditions of Eqs. 16 the transfer matrix is determined by:

$$\begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \frac{1}{p_0 v_d + p_d v_0} \begin{bmatrix} p_d v_d + p_0 v_0 & p_0^2 - p_d^2 \\ v_0^2 - v_d^2 & p_d v_d + p_0 v_0 \end{bmatrix} \tag{17}$$

where subscripts a and b are no longer necessary.

2.4 Insertion loss

The insertion loss is the difference between the acoustic power transmitted with and without the muffler, denoted by the subscripts 2 and 1, respectively

$$IL = L_{W_1} - L_{W_2} = 10 \log \left(\frac{W_1}{W_2} \right) \tag{18}$$

If the temperature gradient is negligible and the source pressure is constant [1] it becomes simply

$$IL = 20 \log \left| \frac{p_1}{p_2} \right| \tag{19}$$

where p_1 is the pressure at the outlet without the muffler and p_2 with the muffler.

3 Numerical results

Using the spectral element previously derived, the periodic muffler can be modeled by assembling the element matrices by imposing continuity at the element connections. Two connected elements are assembled as

$$\mathbf{D} = \begin{bmatrix} D_{ll}^{(1)} & D_{lr}^{(1)} & 0 \\ D_{rl}^{(1)} & D_{rr}^{(1)} + D_{ll}^{(2)} & D_{lr}^{(2)} \\ 0 & D_{rl}^{(2)} & D_{rr}^{(2)} \end{bmatrix} \tag{20}$$

After assembling all the elements composing the periodic cell, the resulting matrix can be condensed to the nodes at the two ends, resulting in a 2 x 2 matrix

$$\bar{\mathbf{D}} = \begin{bmatrix} D_{11} - D_{21} D_{22}^{-1} D_{12} & -D_{23} D_{22}^{-1} D_{12} \\ -D_{21} D_{22}^{-1} D_{32} & D_{33} - D_{32} D_{22}^{-1} D_{23} \end{bmatrix} \tag{21}$$

which can be transformed into a transfer matrix using Eq. 6 , in terms of $\bar{\mathbf{D}}$

$$\mathbf{T} = \begin{bmatrix} -\bar{D}_{12}^{-1} \bar{D}_{11} & \bar{D}_{12}^{-1} \\ -\bar{D}_{21} + \bar{D}_{22} \bar{D}_{12}^{-1} \bar{D}_{11} & -\bar{D}_{12}^{-1} \bar{D}_{22} \end{bmatrix} \tag{22}$$

From the transfer matrix of the periodic cell the dispersion relation is computed as previously described. It is shown in Fig. 2. Furthermore, using Eq. 15, the TL can be computed, as shown in Fig. 1. To compute the FRFs for a volume acceleration input at one end and pressure outputs at the two ends matrix $\bar{\mathbf{D}}$ is inverted and multiplied by vector $\hat{\mathbf{V}} = [1\ 0]^T$.

4 Experimental results

The two loads used in the impedance tube were a rigid and a nearly anechoic termination, implemented using absorptive foam in the open tube end. Figure 4 shows the experimental setup.

The sample is a periodic muffler built using 3D printing (polyamide and selective laser sintering). Figure 5 shows the sample manufactured for the experiment.

Measurements were made with a commercial impedance tube with four-microphones. The loudspeaker was driven with a Gaussian noise signal approximately white (flat spectrum) in the frequency range from 0 Hz to 5272 Hz. A single roving microphone was used to avoid the phase calibration required in the four-microphone measurement [7].

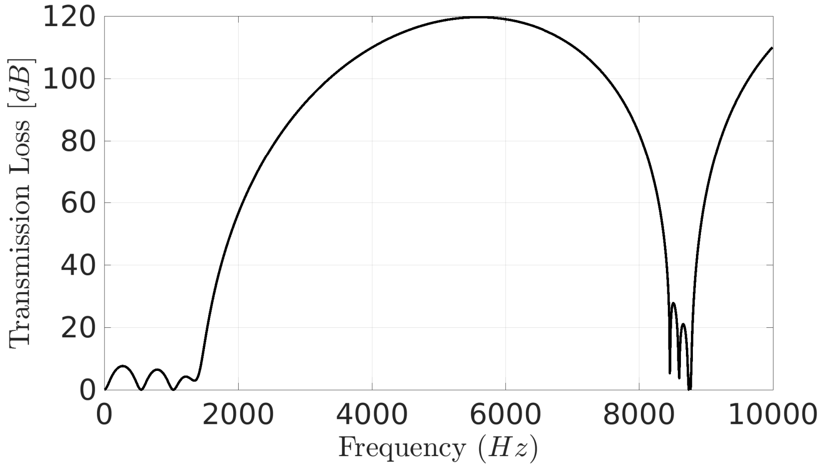


Figure 1: Simulated TL.

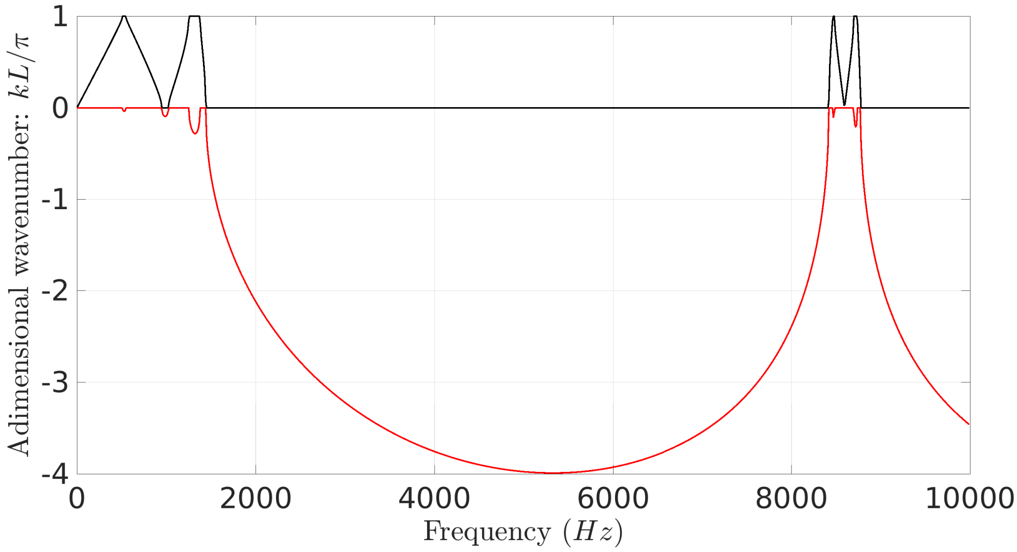


Figure 2: Simulated dispersion diagram.

Figure 6 shows a comparison between simulated and experimental results for the TL. The one-load method, allowed by the symmetric muffler geometry in our case, yielded better results. Experimental background noise did not allow to obtain TL values above approximately 20 dB.

Figure 7 shown a comparison between the predicted and measured IL. The measurements are less noisy, but again background noise did not allow obtaining IL values larger than 20 dB. Measurements inside an anechoic chamber are under way to try to get measurements with lower background noise in the case of the open tube.

5 Conclusions

We have shown how to use dispersion diagrams to predict band gaps in periodic mufflers, which result in large transmission losses. Using acoustic spectral elements it is straightforward to obtain the muffler transfer matrix, from which the dispersion diagram and TL can be computed. For arbitrary shapes, the periodic cell can be computed by discretizing the cell with uniform spectral elements. Otherwise, as shown by the authors [8], a state-space formulation and Riccati-type equation for the acoustic impedance can be used. Optimized mufflers can be constructed using 3D printing. Different optimization strategies are under investigation by the authors. The tools shown in this work set a framework for periodic muffler design optimization using dispersion relations.

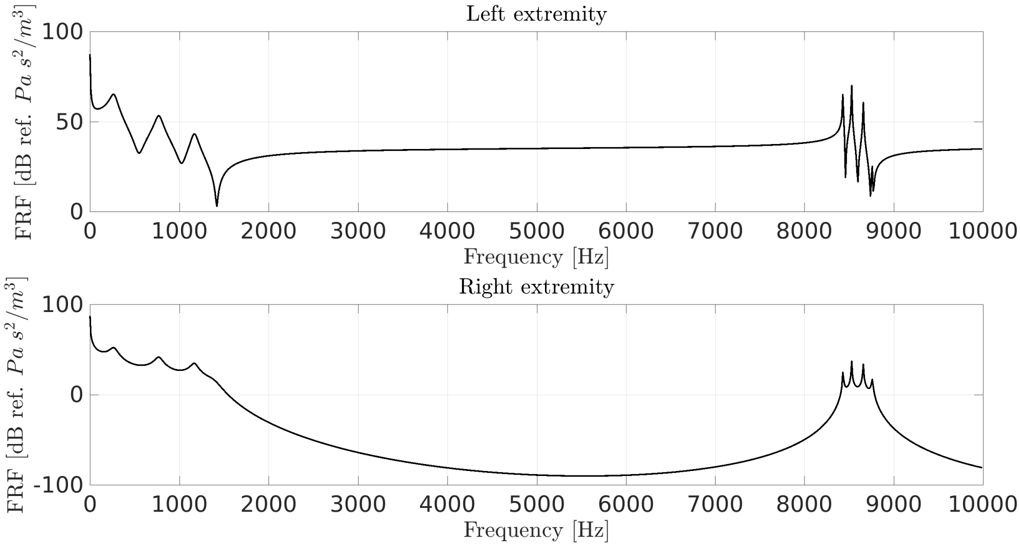


Figure 3: Simulated FRFs.

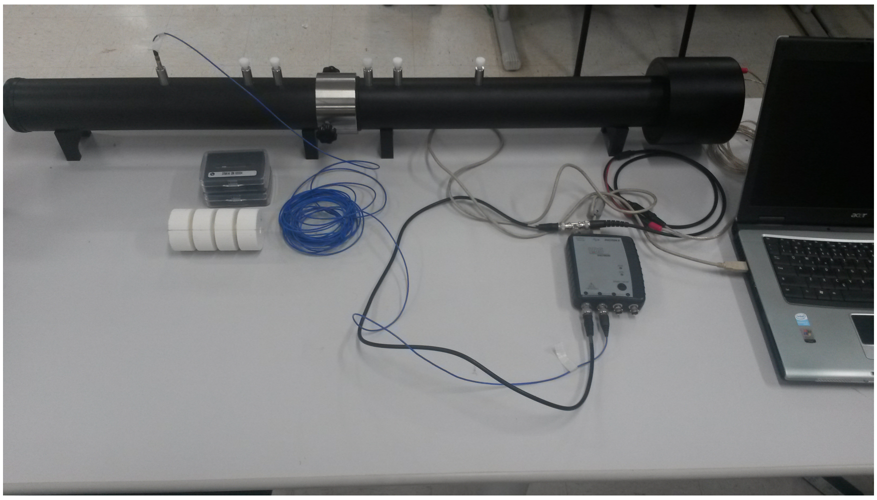


Figure 4: Experimental setup.

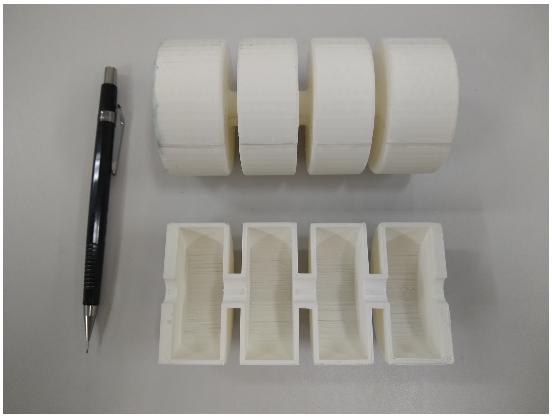


Figure 5: Manufactured muffler.

Acknowledgements

The authors are thankful to Coordenação de Aperfeiçoamento de Pessoal de Nível Superiorl (CAPES Finance Code 001) and the São Paulo Research Foundation (FAPESP 2018/15894-0) for the financial support.

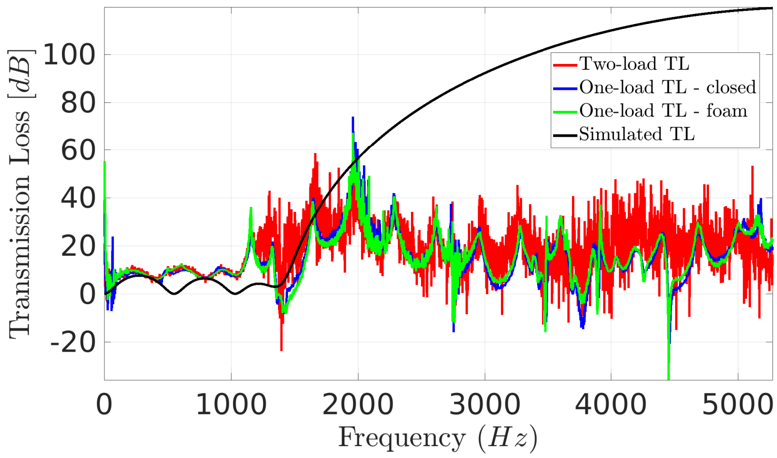


Figure 6: Simulated and experimental transmission loss.

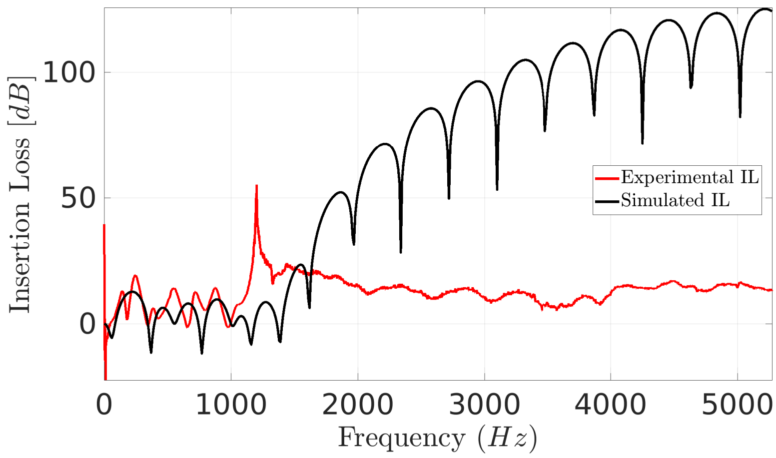


Figure 7: Simulated and experimental insertion loss.

References

[1] M.L. Munjal, *Acoustic of ducts and mufflers with application to exhaust and ventilation system design*, John Wiley and Sons, Canada, 1987.

[2] M.I. Hussein, M.J. Leamy and M. Ruzzene, *Dynamics of phononic materials and structures: Historical origins, recent progress, and future outlook*, Applied Mechanics Reviews, American Society of Mechanical Engineers, 2014, V. 66, n. 4, pp.1-4.

[3] M.M. Sigalas and E.N. Economou, *Elastic and acoustic wave band structure*, Journal of Sound and Vibration, 1992, V. 158, n. 2, pp. 337-382.

[4] J.F. Doyle, *Wave Propagation in Structures: Spectral Analysis Using Fast Discrete Fourier Transforms*, Mechanical Engineering Series, Springer New York, USA, 1997.

[5] L.E. Kinsler, A.R. Frey, A.B. Coppens and J.V. Sanders, *Fundamentals of acoustics*, John Wiley and Sons, USA, 2000.

[6] M. I. N. Rosa, V. D. Lima, J. M. C. Santos and J. R. F. Arruda, *Numerical and Experimental Investigation of Interface Modes in Periodic Acoustic Waveguides*, Proceedings of the International Conference on Structural Engineering Dynamics, Viana do Castelo, Portugal, 24-26 June 2019, 10 pages.

[7] B.H. Song and J.S. Bolton, *A transfer-matrix approach for estimating the characteristic impedance and wave numbers of limp and rigid porous materials*, Journal of the Acoustical Society of America, 2000, V. 107, n. 3, pp.1131-1152,

[8] G.F.C.A. Assis, D. Beli, E.J.P. Miranda Jr., J.F. Camino, J.M.C. dos Santos and J.R.F. Arruda, *Wave Propagation in Structures: Spectral Analysis Using Fast Discrete Fourier Transforms*, International Journal of Mechanical Sciences (submitted), 2019.