Applied Condition Monitoring

Ahmed Hammami ·
Philippus Stephanus Heyns ·
Stephan Schmidt · Fakher Chaari ·
Mohamed Slim Abbes ·
Mohamed Haddar *Editors*

Modelling and Simulation of Complex Systems for Sustainable Energy Efficiency

Contributions to the First International Workshop on Modelling and Simulation of Complex Systems for Sustainable Energy Efficiency, MOSCOSSEE'2021, February 25–26, 2021

Springer

Applied Condition Monitoring

Volume 20

The book series Applied Condition Monitoring publishes the latest research and developments in the field of condition monitoring, with a special focus on industrial applications. It covers both theoretical and experimental approaches, as well as a range of monitoring conditioning techniques and new trends and challenges in the field. Topics of interest include, but are not limited to: vibration measurement and analysis; infrared thermography; oil analysis and tribology; acoustic emissions and ultrasonics; and motor current analysis. Books published in the series deal with root cause analysis, failure and degradation scenarios, proactive and predictive techniques, and many other aspects related to condition monitoring. Applications concern different industrial sectors: automotive engineering, power engineering, civil engineering, geoengineering, bioengineering, etc. The series publishes monographs, edited books, and selected conference proceedings, as well as textbooks for advanced students.

** Indexing: Indexed by SCOPUS, WTI Frankfurt eG, SCImago

More information about this series at http://www.springer.com/series/13418

Ahmed Hammami · Philippus Stephanus Heyns ·
Stephan Schmidt · Fakher Chaari ·
Mohamed Slim Abbes · Mohamed Haddar
Editors

Modelling and Simulation of Complex Systems for Sustainable Energy Efficiency

Contributions to the First International Workshop on Modelling and Simulation of Complex Systems for Sustainable Energy Efficiency, MOSCOSSEE'2021, February 25–26, 2021

Springer

Editors

Ahmed Hammami
National School of Engineers of Gabes
Gabes, Tunisia

Stephan Schmidt
Centre for Asset Integrity Management,
Department of Mechanical
and Aeronautical Engineering
University of Pretoria
Pretoria, South Africa

Mohamed Slim Abbes
National School of Engineers of Sfax
Sfax, Tunisia

Philippus Stephanus Heyns
Centre for Asset Integrity Management,
Department of Mechanical
and Aeronautical Engineering
University of Pretoria
Pretoria, South Africa

Fakher Chaari
National School of Engineers of Sfax
Sfax, Tunisia

Mohamed Haddar
National School of Engineers of Sfax
Sfax, Tunisia

ISSN 2363-698X ISSN 2363-6998 (electronic)
Applied Condition Monitoring
ISBN 978-3-030-85583-3 ISBN 978-3-030-85584-0 (eBook)
https://doi.org/10.1007/978-3-030-85584-0

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The first "International Workshop on MOdelling and Simulation of COmplex Systems for Sustainable Energy Efficiency" MOSCOSSEE'2021 was organized by the LAboratory of Mechanics, Modelling and Production (LA2MP) from University of Sfax, Tunisia, and the Department of Mechanical and Aeronautical Engineering at the University of Pretoria, on 25 and 26 February 2021. This workshop is in the framework of a Tunisian South-African research project entitled "Design, Modelling and Diagnostic of Wind Turbines for Sustainable Energy Efficiency." It was scheduled online due to the COVID-19 pandemic. The MOSCOSSEE'2021 workshop comprised high-level contributions in the fields of complex systems for sustainable energy efficiency in order to promote communication and collaboration between participants. Three plenary sessions were presented by eminent scientists who kindly agreed to share their knowledge in the workshop field. The organizers of the conference were honored by their participation with very interesting keynotes, namely

- Prof. José ANTUNES, Applied Dynamics Laboratory (ADL), Superior Technical Institute Lisbon, Portugal.
- Prof. Abdelkhalek ELHAMI, Mechanical Engineering Department, National Institute of Applied Sciences in Rouen (INSA de Rouen), France.
- Prof. Mohamed Amine BEN SOUF, National School of Engineers of Sfax, Tunisia.

This book contains 24 chapters selected from the presented papers by eminent scientists which were rigorously peer reviewed. During the 2 days of the workshop, oral communications discussed several topics such as

- Sustainable energy efficiency,
- Vibrations of complex systems,
- Structural and machine dynamics,
- Fault diagnosis and prognosis,
- Nonlinear dynamics,
- Vibration field measurements,

- Material behavior in dynamics,
- etc.

The editors are grateful to all participants from Tunisia, South Africa, Poland, Portugal and France, as well as the reviewers of the chapters. We acknowledge the financial support of the Ministry of Higher Education and Scientific Research in Tunisia and the National Research Foundation in South Africa under the Tunisia South Africa Agreement for Cooperation in Science and Technology. We would also like to thank Springer for their support of the MOSCOSSEE'2021 workshop.

February 2021

Ahmed Hammami
Philippus Stephanus Heyns
Stephan Schmidt
Fakher Chaari
Mohamed Slim Abbes
Mohamed Haddar

Contents

Numerical Assessment of the Structural Performance of a Segmented Wind Turbine Blade

Majdi Yangui(✉), Salma Thabti, Slim Bouaziz, and Mohamed Haddar

Mechanics, Modelling and Production Laboratory (LA2MP), Mechanic Department,
National School of Engineers of Sfax, University of Sfax, BP 1173, 3038 Sfax, Tunisia
mohamed.haddar@enis.rnu.tn

Abstract. Segmented wind turbine blade (SWTB) development remains a major challenge for constructors so as to reduce blade transport and manufacturing costs. The blade structural properties must be examined in the design stage to enhance their mechanical behavior and fatigue life. This paper presents a numerical investigation of a SWTB prototype. Teeth inserted in holes at the interfaces of segments, were designed to avoid relative displacements between the segments assembled along a spar. Modal and fatigue analysis were established using ANSYS Workbench software to evaluate the structural performance of the investigated wind turbine blade (WTB). This work covers the impact of the assembly force and the rotation velocity effect on the blade fatigue life. Previous findings of an experimental study, of the SWTB at rest, were considered to validate the blade finite element model. To assess the used spar location, along the blade segments, the edgewise and flapwise deflections of the blade under assembly force effects were analysed. This study reveals the significant impact of the exerted assembly force on the SWTB fatigue life versus the rotation velocity effects. Interestingly, the obtained results indicate that a segments assembly force must be respected in the blade assembling to ensure the optimum service life.

Keywords: Fatigue analysis · Modal analysis · Segmented wind turbine blade · Finite element modeling

1 Introduction

In recent years, the concerns about the global warming consequences, caused essentially by the excessive fossil energy production, has made renewable energy development more and more indispensable for a sustainable future. In this context, wind energy is treated as one of the most profitable clean energy sources. Actually, power generation efficiency and cost represent the primary factors which govern wind turbine development. Thus, maintenance and manufacturing cost reduction remains a primary need. Undoubtedly, the wind turbine blades are the most crucial parts, in terms of performance and cost, of the wind power system. For this reason, many works have analysed the rotor blades fatigue and vibration, aiming to extend their service life. Shokrieh and Rafiee (2006) studied

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 1–7, 2022.
https://doi.org/10.1007/978-3-030-85584-0_1

the fatigue phenomena of a 23 m wind turbine blade (WTB) manufactured by the Vestas Company to predict its lifetime. By adopting a stochastic approach, the service life of the selected blade has been predicted to be limited to 18.66 years. Jensen et al. (2006) tested to failure, a composite WTB, under flapwise loading. The structure displacements were registered throughout the loading history. The experimental measurements and the numerical simulation results were processed to determine the location of the initial failure.

Recently, because of the intricacy of manufacturing and transport processes of the long WTBs, the blade fragmentation was proposed to solve such issues (Abdulaziz et al. 2018). Yangui et al. (2020) developed a numerical model of a SWTB using shell elements. To update the used material properties in the numerical model from the natural frequencies identified experimentally, an iterative technique was followed based on the substructuring technique. Dutton et al. (2001) tested a segmented form of a 13.4 m blade with a connecting tube to investigate the durability of the advanced fragmentation method. Static load tests, in the edgewise and flapwise directions were performed and repeated after a five million cycle fatigue test in the flapwise direction. The blade inspection shows that no damage was occurring in the segment interfaces and connections. Static and fatigue analysis of a SWBT were performed by Bhat et al. (2015) to evaluate its structural performance. The determined numerical results, of the non-segmented and the segmented blade, indicate that the effect of the fragmentation on the entire structural performance is minimal. Nevertheless, the outcomes of the load applied to assemble the segments were neglected. Yangui et al. (2019) performed an experimental analysis to inspect the effects of the assembly force adjusted by a nut on the WTB dynamical behavior. The determined experimental results, using the Eigen-system Realization Algorithm (ERA) modal identification method, showed the notable influence of the assembly force change on the blade eigenfrequencies versus the effects resulting from the blade rotation. Nevertheless, the impacts of the applied force on the blade shape and lifetime were not addressed.

In the present paper, an attempt to address this issue has been made by investigating the displacements and the fatigue life of a SWTB taking into account the mounting force of the segments. Based on previous experimental modal identification, the blade numerical model developed using Ansys Workbench software was validated. Static and fatigue analysis were performed to assess the effects resulting from the assembly force and the rotation velocity on the SWTB structural performance.

2 Blade Finite Element Modeling

A SWTB model, consisting of 5 segments assembled along a spar, was designed as seen in Fig. 1.

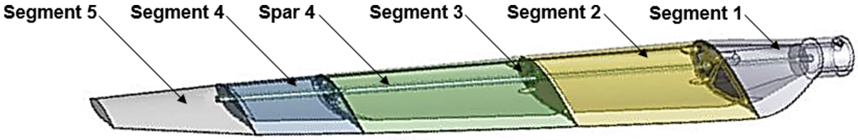


Fig. 1. Segmented blade CAD model

The full length of the designed blade is 500 mm and the segment skin thickness is about 3 mm. Regarding to the assembly, a spar with a length of 420 mm and a diameter of 4 mm was used. The material properties of the SWTB components are presented in the following Table 1.

Table 1. Material properties of the blade components.

Parameters	Material	Density (kg/m^3)	Poisson's ratio	Elastic modulus (GPa)
Blade segments	PC-ABS	1070	0.3879	2.25
Spar	Steel	7850	0.3	210

Using ANSYS Workbench finite element software, solid elements with three degrees of freedom per node and a free mesh were adopted to model the blade structure. Considering the blade segment's complex shape, the tetrahedral finite element was used. For the spar, the mesh was simplified by adopting the quadratic element as shown in Fig. 2.

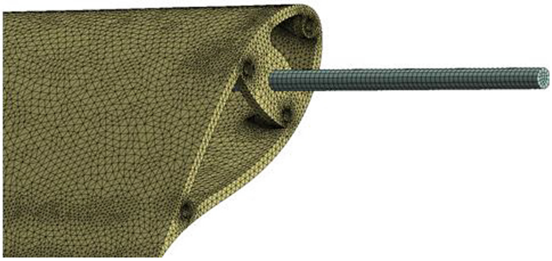


Fig. 2. Segmented blade mesh

To simulate the contact between the different parts of the blade, a frictional contact was defined between the segments and the spar and at the interfaces of segments. The contact between the tip and the root segments of the blade and the spar was bonded.

To validate the established numerical model, modal analysis was carried out without applying the assembly force. The assembled blade structure was clamped at its root. To optimise the mesh size, a convergence study was conducted for various mesh sizes. Accordingly, 979306 nodes and 613326 elements were generated. The natural frequencies, obtained numerically and those reached experimentally by Yangui et al. (2019), are given in Table 2.

Table 2. Segmented blade natural frequencies.

	Present work	Yangui et al. (2019)	Error %
1st natural frequency	17.9	17.4	2.87
2nd natural frequency	23.5	24.8	5.24
3rd natural frequency	80.9	85.7	5.60

An acceptable agreement is found, between the simulated and experimental findings, where the maximum error is about 5.6%. Thus, the introduced numerical model can be reliably employed to evaluate the fragmented blade structural performance.

3 Structural Performance Assessment

The WTB efficiency depends essentially on the blade shape. Thus, the segments mounting force and the spar location must be primarily treated during the blade design to avoid structural distortion. Figure 3 shows the blade tip edgewise and flapwise displacements for different assembly loads. In this section, only the static assembly load of the blade segments was exerted.

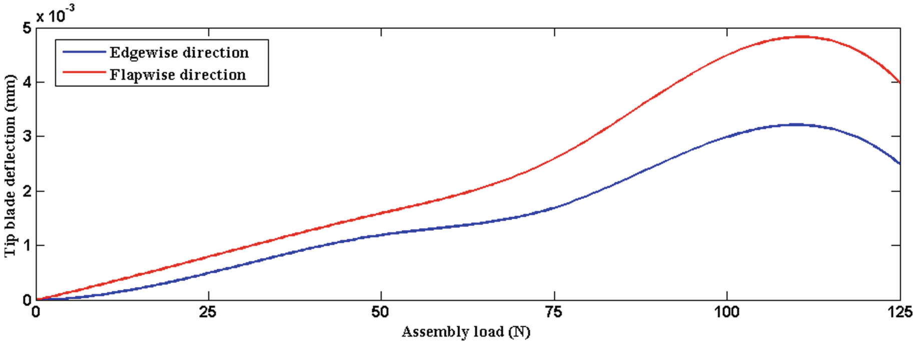


Fig. 3. Edgewise and flapwise displacements as a function of the assembly load.

The displacement amplitudes proved the negligible influence of the assembly force on the blade shape, where, up to a significant assembly effort of 125 N, the maximum deflection does not exceed 0.005 mm. Therefore, the spar location along the blade is well designed.

To assess the impact of the assembly force on the lifetime of the WTB structure, fatigue analysis was performed using Ansys Workbench fatigue module. Stress life type analysis was adopted based on Stress-Cycle (S-N) curves of the segments and spar materials. A bending load, in the flapwise direction, equal to 5 N was applied on the third blade segment as illustrated in Fig. 4. Based on the results obtained from the static load analysis, the fatigue analysis was performed where the zero-based constant amplitude loading type was assumed.

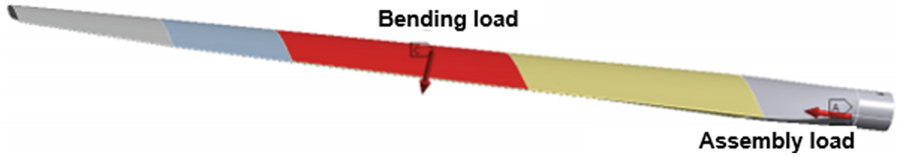


Fig. 4. Blade bending and assembly loads.

Figure 5 represents the SWTB lifetime change as a function of the applied assembly load.

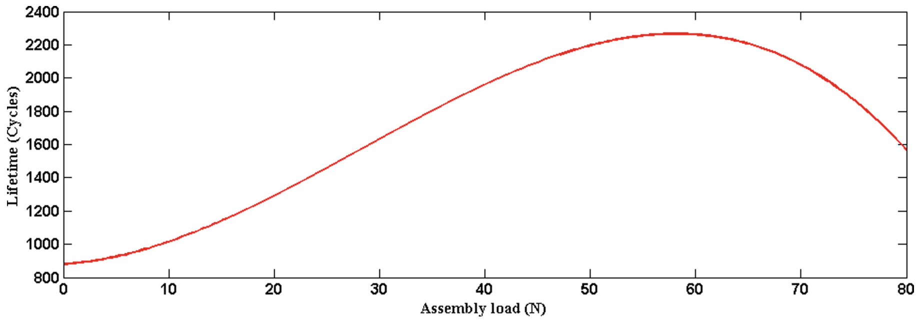


Fig. 5. Segments assembly load impact on the blade fatigue life.

It is observed that the lifetime of the blade is significantly dependent on the assembly load. The maximum fatigue life of the blades, equal to 2273 cycles, is obtained by applying an assembly load equal to 59 N. For an assembly load higher than 60 N, the blade cycles of fatigue life start to decrease.

To analyse the rotation velocity results on the blade fatigue life, a constant assembly load equal to 59 N was applied. The rotational velocity direction was performed as seen in Fig. 6.

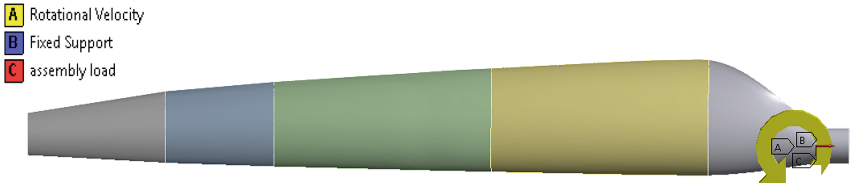


Fig. 6. Rotating blade model

Based on a static analysis, the blade fatigue life investigation was performed assuming a ratio loading type with a tolerance of 10%. Thus, the alternating load, provoked by the blade spinning velocity, was limited to 10% of the static load amplitude. The prediction of the rotating blade fatigue life is presented in Fig. 7.

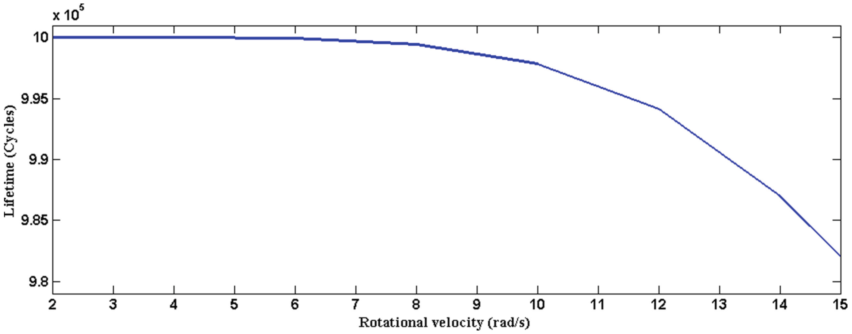


Fig. 7. Rotation velocity effects on the blade fatigue life.

Clearly, the increase of the WTB rotation velocity from 2 rad/s to 15 rad/s engenders a negligible loss of the blade fatigue life. Thus, the spar is well localized along blade segments in a way that it did not generate an important bending force component in the centrifugal load produced by the blade rotation.

4 Conclusion

In the present study, numerical analysis were performed to evaluate the structural performance of a SWTB. The numerical model of the blade was validated by reference to experimental analysis. The blade deflections, at rest, under the assembly load were determined the prove the satisfactorily location of the spar along the blade segments. Static and fatigue analysis were also performed to inspect the impacts of the segments assembly force and the rotation velocity on the blade fatigue life. Findings show that an assembly load must be respected during the WTB mounting to ensure the maximum service life, and thus, the sustainability of the wind turbine system. Again, the designed spar location was perfectly assessed by the negligible effects of the rotation velocity on the blade lifetime. In this work, only the flapwise bending fatigue was treated. Therefore, this research can be extended through the development of a dual-axis, edgewise and flapwise, fatigue testing.

References

Abdulaziz, A.H., Elsabbagh, A., Elnady, T.: Testing and validation of a novel segmented wind turbine blade. J. Test. Eval. **47**(5), 3719–3739 (2018)

Bhat, C., Noronha, D.J., Saldanha, F.A.: Structural performance evaluation of modularized wind turbine blade through finite element simulation. J. Energy Econ. Dev. **1**(1), 79 (2015)

Dutton, A.G., Kildegaard, C., Dobbe, T.: Design, structural testing, and cost effectiveness of sectional wind turbine blades. Technical Report JOR3970167 (2001)

Jensen, F.M., Falzon, B.G., Ankersen, J., Stang, H.: Structural testing and numerical simulation of a 34 m composite wind turbine blade. Compos. Struct. **76**(1–2), 52–61 (2006)

Shokrieh, M.M., Rafiee, R.: Simulation of fatigue failure in a full composite wind turbine blade. Compos. Struct. **74**(3), 332–342 (2006)

Yangui, M., Bouaziz, S., Taktak, M., Debut, V., Antunes, J., Haddar, M.: Numerical and experimental analysis of a segmented wind turbine blade under assembling load effects. J. Theor. Appl. Mech. **57**(1), 85–97 (2019)

Yangui, M., Bouaziz, S., Taktak, M., Haddar, M.: Experimental updating of a segmented wind turbine blade numerical model using the substructure method. J. Strain Anal. Eng. Des. **56**(2), 67–75 (2020)

Identifiability Considerations for Rotating Machine Fault Diagnosis and Prognosis

Stephan Schmidt^(✉), P. Stephan Heyns, and Daniel N. Wilke

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
stephan.schmidt@up.ac.za

Abstract. It is important to develop reliable fault diagnosis and prognosis methods for critical mechanical assets such as wind turbines. Reliable fault diagnosis and prognosis methods ensure that the damage is detected early, the damage modes are accurately characterised, and the correct remaining life is inferred. This enables the appropriate maintenance decisions to be made and can decrease the risk of unexpected breakdowns. Identifiability is an important criterion for the development of new fault diagnosis and prognosis methods. Therefore, in this work, we present the identifiability problem for fault diagnosis and prognosis on academic examples and we place a specific emphasis on gearbox applications. This chapter provides an overview of the concepts and is intended for neophytes to experienced researchers and practitioners. Hence, the examples are purposefully simple. We specifically highlight the importance of sensor positioning and also discuss the influence of varying operating conditions on the diagnosis and prognosis steps. Thereafter, we present the fundamental steps in the fault diagnosis and prognosis process and highlight the associated challenges with identifiability. We also propose potential solutions for these challenges. Lastly, we propose requirements for the different phases of the fault diagnosis and prognosis steps, which could be beneficial when developing new methods.

Keywords: Gearbox · Diagnosis · Prognosis · Identifiability · Bearing · Gears · Signal processing

1 Introduction

Condition-Based Maintenance (CBM), which comprises diagnosis and prognosis tasks, is important for expensive machine assets such as wind turbines [1]. Wind turbines, for example, are prone to gearbox, brake and rotor blade failures that result in long downtimes [1]. In CBM, condition monitoring data such as temperature, vibration, and pressure measurements are acquired from the machine and subsequently used to infer the health of the machine (i.e. diagnosis) and to determine the remaining useful life of the machine (i.e. prognosis) [1]. Vibration-based methods are most commonly used to monitor the condition of rotating machine components such as bearings and gears [2,3]

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 8–20, 2022.
https://doi.org/10.1007/978-3-030-85584-0_2

Usually, the damaged assets result in subtle changes in the condition monitoring data that are difficult to detect [2]. This makes processing the data into more meaningful representations essential. Several signal processing and learning-based[1] condition monitoring methods have been developed (e.g. Ref [4,5]) to address the challenges posed by the weak signal components-of-interest. Here, extraneous signal components, time-varying operating conditions, and non-Gaussian noise mask the components-of-interest and impede diagnosis [3,6].

Ultimately, the diagnosis and prognosis steps are used to inform maintenance decisions and therefore the diagnosis and prognosis methods should be carefully designed. In the diagnosis step, the damaged component and damage severity are inferred from the condition monitoring data and used for prognosis algorithms to predict the remaining useful life. Many data-driven prognosis approaches are proposed in the literature [7]. However, the methods require representative historical fault data from the different damage modes, which are difficult to obtain in practice [8]. Hence, inferring the relevant fault information is critical.

Therefore, in this work, we highlight that identifiability is essential for developing fault diagnosis and prognosis methods. Firstly, we present the identifiability problem for rotating machine condition monitoring in Sect. 2, whereafter we highlight the different tasks in the fault diagnosis and prognosis process in Sect. 3. In Sect. 4, we propose requirements for the development of fault diagnosis and prognosis methods, whereafter we conclude the work in Sect. 5.

2 Identifiability for Condition Monitoring

2.1 The Condition Monitoring Problem

Figure 1(a) shows the condition monitoring problem for a gearbox with four bearings (B1, B2, B3, B4), a gear, a pinion and two events E1 and E2. In Fig. 1(b), the sources of the events are shown: The pinion is damaged, resulting in event E1, and the outer race of the bearing is damaged, resulting in event E2. In this example, two sensors are placed on the casing of the gearbox S1 and S2. Ultimately, we would like to infer the gearbox's condition from one (or both) sensor(s) measurements. In this example, actual measurements from different sensors on a gearbox are shown to emphasise the influence of the sensor positions.

The excitations due to the damage should travel between the source (i.e. the events) and the sensors [9], with possible transmission paths highlighted in Fig. 1(c). The time domain signals can be decomposed as follows:

$$x_1(t) = h_{11}(t) \otimes e_1(t) + h_{12}(t) \otimes e_2(t) + n_1(t), \tag{1}$$

$$x_2(t) = h_{21}(t) \otimes e_1(t) + h_{22}(t) \otimes e_2(t) + n_2(t), \tag{2}$$

---

[1] We refer to statistical learning, machine learning and deep learning methods as learning-based methods.

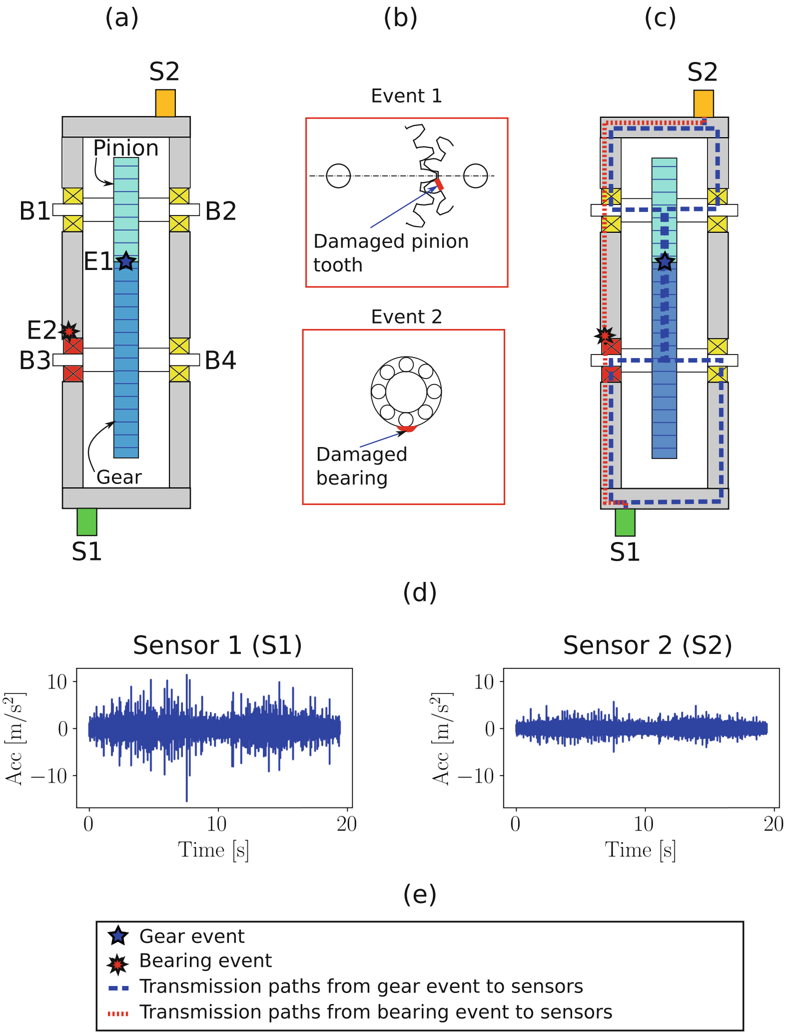


Fig. 1. The condition monitoring problem: (a) The gearbox, two events inside the gearbox E1 and E2, four bearings B1, B2, B3, B4 and two sensors, S1 and S2, are shown. (b) The two damage components causing the events are shown. (c) Some of the transmission paths between the events and the sensors are shown. (d) Example signals are shown for the two sensors. These were acquired from two sensors on the test-rig described in Ref. [3]. (e) The legend for the events and the transmission paths.

where $x_i(t)$ and $n_i(t)$ are the vibration signal and the extraneous components[2] in the ith sensor signal, respectively. The jth event, attributed to the excitation at the source of the damage (e.g. the rolling element's interaction with a spall), is denoted e_j. The impulse response function that captures the transmission path between event j and sensor i is denoted by h_{ij}. The measured vibration signal is discrete and denoted $\boldsymbol{x} = [x(0), x(\Delta t), x(2 \cdot \Delta t), \ldots]$, where Δt is the sampling period. By using the convolution theorem, it is possible to decompose the signals as follows:

$$\begin{bmatrix} X_1(f) \\ X_2(f) \end{bmatrix} = \begin{bmatrix} H_{11}(f) & H_{12}(f) \\ H_{21}(f) & H_{22}(f) \end{bmatrix} \begin{bmatrix} E_1(f) \\ E_2(f) \end{bmatrix} + \begin{bmatrix} N_1(f) \\ N_2(f) \end{bmatrix}, \tag{3}$$

where H_{ij}, X_i, E_j, N_i are the Fourier transforms of h_{ij}, x_i, e_j and n_i respectively. Changes in the signal X_i can potentially be influenced by changes in the frequency response function H_{i1} and H_{i2}; the events E_1 and E_2; and/or the noise components N_i. Therefore, it is important to carefully interpret the changes in the raw signal and its statistics. Ultimately, we desire to extract or enhance the event information E_j to perform diagnosis and prognosis.

2.2 Identifiability Problem

Identifiability and observability are closely related; observability indicates whether we can infer the latent state of the system from the system's response, while identifiability indicates whether we can infer the system's parameters from its response [10]. In condition monitoring, we typically would like to infer the parameters of the damage (e.g. size of the damage) and therefore we present this discussion in an identifiability context. We will use Eq. (3) to present the identifiability problem for condition monitoring, without accounting for estimation errors (e.g. the influence of finite length signals). We can write Eq. (3) in matrix form

$$\boldsymbol{X}(f) = \boldsymbol{H}(f) \cdot \boldsymbol{E}(f) + \boldsymbol{N}(f), \tag{4}$$

and leverage elementary linear algebra theory to gain insight into the identifiability of the events. The following four interesting cases are considered here for highlighting the identifiability problem:

$$\boldsymbol{H}_1 = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \quad \boldsymbol{H}_2 = \begin{bmatrix} a & c \\ c & b \end{bmatrix}, \quad \boldsymbol{H}_3 = \begin{bmatrix} a & a \\ a & a \end{bmatrix}, \quad \boldsymbol{H}_4 = \begin{bmatrix} a & 0 \\ 0 & 0 \end{bmatrix}. \tag{5}$$

where $a \in \mathbb{C}$, $b \in \mathbb{C}$, $c \in \mathbb{C}$. If the frequency response function matrix $\boldsymbol{H} = \boldsymbol{H}_1$, the events will feature independently in the sensors. Therefore, different sensors can be used to characterise specific events. If $\boldsymbol{H} = \boldsymbol{H}_2$ and assuming $\det(\boldsymbol{H}) \neq 0$, the matrix is invertible and therefore we obtain two signals with independent event information using

$$\boldsymbol{H}^{-1}(f) \cdot \boldsymbol{X}(f) = \boldsymbol{E}(f) + \boldsymbol{H}^{-1}(f) \cdot \boldsymbol{N}(f), \tag{6}$$

---

[2] Extraneous components refer to signal components attributed to physical mechanisms (e.g. healthy gear mesh components), environmental conditions and noise that are not related to the fault components-of-interest.

where $\boldsymbol{H}^{-1}(f) \cdot \boldsymbol{X}(f)$ is the spectrum of the processed signal. Note, that unless we know $\boldsymbol{N}(f)$, we cannot identify the events E_j, but only scaled and translated versions of the events, i.e. $\alpha \cdot E_j + \beta$. This is often sufficient for most condition monitoring tasks. If $\boldsymbol{H} = \boldsymbol{H}_3$, the rows are linearly dependent, which means that the sensors contain duplicate information (e.g. the sensors are symmetric for the structure and the excitations) and $\boldsymbol{H}_3$ is not invertible. This means that we cannot recover the original events from the response. Lastly, if $\boldsymbol{H} = \boldsymbol{H}_4$, it is only possible to identify the information concerning event 1; event 2 is not identifiable irrespective of the signal processing algorithm used. These four academic examples highlight the importance of sensor positioning to obtain well-conditioned matrices.

Many condition monitoring algorithms (e.g. [2,3,11]) only utilise the information from a single sensor to infer the machine's condition. In some classes of condition monitoring algorithms (e.g. [12,13]), we aim to design filters $P_{ij}(f)$ that can be applied to the data of a single sensor $X_j(f)$ so that the jth event is identifiable from the processed signals. The spectrum of these processed signals can be decomposed as follows:

$$\begin{bmatrix} P_{i1}(f) \\ P_{i2}(f) \end{bmatrix} X_i(f) = \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix} \cdot \begin{bmatrix} E_1(f) \\ E_2(f) \end{bmatrix} + \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \tag{7}$$

where $P_{ij}(f) \cdot X_i(f)$ is the spectrum of the processed signal aiming to extract event E_j, α_i is a scaling term and β_i is an offset term. The scaled events E_j are identifiable if we can design a filter

$$P_{ij}(f) = k \cdot H_{ij}(f), \tag{8}$$

where $k \neq 0$ and $H_{i1}(f) \cdot H_{i2}(f) = 0$, i.e. the frequency response functions of the two events are independent. This is possible if the two events manifest in two separate narrow frequency bands. Targeted informative frequency band identification methods (e.g. [11,12]) and targeted blind deconvolution algorithms [13] aim to find the optimal filters $P_{ij}(f)$ to extract the events-of-interest.

In conclusion, the measurement signal (and its processed signal) is influenced by the [9]:

- transmission path, which means that it is influenced by the sensor location and the damaged component's location. Planetary gearboxes have time-varying transmission paths between the planet gears and the sensors [14].
- excitation characteristics (e.g. if there is a bearing crack or a bearing spall in the inner race or outer race).
- the extraneous components (e.g. dominant healthy gear mesh components, impulsive environmental noise).

All these effects can impede identifiability of the events and therefore impede the fault diagnosis and prognosis tasks discussed in the next section.

3 Fault Diagnosis and Prognosis

3.1 Overview of Diagnosis and Prognosis Steps

The diagnosis and prognosis problems are shown in Fig. 2 using the available information from a selected sensor. In the condition inference problem, we infer the actual condition (i.e. the damaged component, the damage mode and the size of the damage) from the processed condition monitoring data. The estimated condition is subsequently used with an appropriate degradation model (e.g. Paris' law) to estimate the system's remaining useful life. The reliability of the remaining useful life estimation process depends on our ability to diagnose the machine.

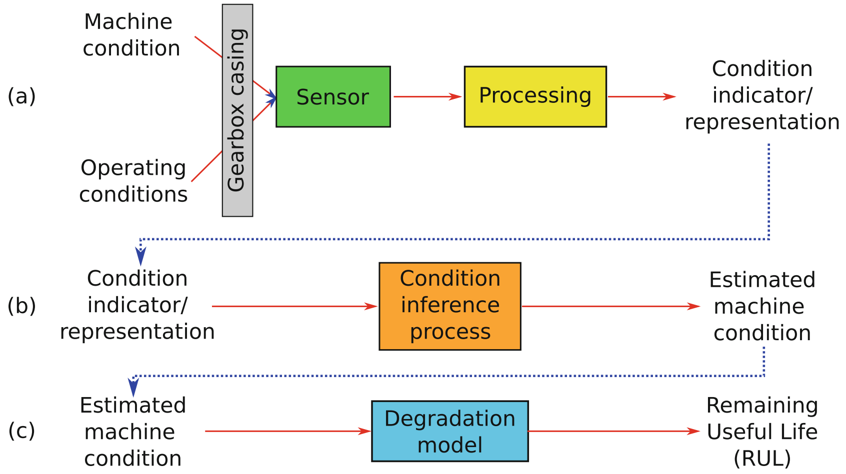


Fig. 2. The conventional diagnosis and prognosis problem. (a) Condition monitoring and data processing; (b) Diagnosis; (c) Prognosis.

In summary, the following steps need to be followed to perform diagnosis (Steps 1–4) and prognosis (Step 5):

1. Damage detection (Presented in Sect. 3.1)
2. Damage component identification (Presented in Sect. 3.2)
3. Damage mode identification (Presented in Sect. 3.3)
4. Fault severity quantification (Presented in Sect. 3.4)
5. Remaining useful life estimation (Presented in Sect. 3.5)

3.2 Damage Detection

In the damage detection phase, changes in the condition monitoring data due to the deteriorating machine are detected and flagged by comparing condition

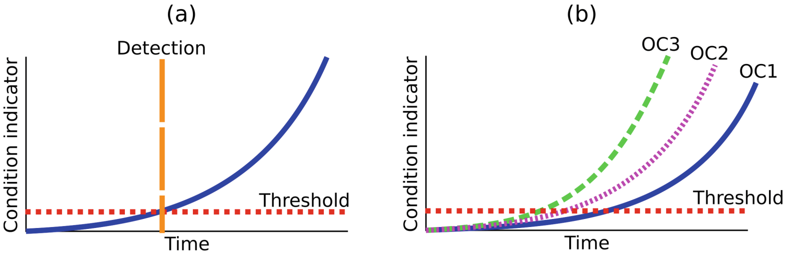


Fig. 3. (a) The generic fault detection problem is shown for a single threshold and fault detection point. (b) The influence of operating conditions on fault detection. Abbreviations: OC - Operating Condition.

indicators against thresholds as shown in Fig. 3(a). For example, the threshold can be determined using the statistical methodology proposed by Antoni and Borghesani [4]. In learning-based methods, the detection threshold is usually determined from the reference density obtained from healthy historical data [5]. However, time-varying operating conditions could influence the condition indicator and could therefore influence the time of detection (i.e. our ability to identify the damage) and increase the false positive rate (i.e. detecting changes due to operating condition as opposed to changes in machine condition). This is highlighted in Fig. 3(b). This problem can be alleviated by using the appropriate pre-processing of the signal (e.g. [15]) or post-processing of the condition indicators (e.g. [16]). For example, in Ref. [15], the amplitude modulation caused by the changes in the operating conditions is estimated and attenuated, while retaining the damage information. In contrast, Zimroz et al. [16] first obtain the relationship between a condition indicator and the wind turbine's power using a regression model . Thereafter, changes in this model are used for condition monitoring. It is expected that the sensitivity of the condition indicator to the operating conditions (e.g. power) is dependent on the condition of the machine and can therefore be used for damage detection.

3.3 Damage Component Identification

The damage component identification (also referred to as damage localisation) problem is shown in Fig. 4 for the example gearbox in Fig. 1. As shown in Fig. 4(a), the measured signal $\boldsymbol{x}$ can be generated by multiple potential events (e.g. a single damage mode or multiple simultaneous damage modes) and contains extraneous components $\boldsymbol{n} = [n(0), n(\Delta t), \ldots,]$. The damaged component is identifiable if we can process the signals $\boldsymbol{x}$ to extract the scaled events $\alpha \cdot e_i + \beta$. This is shown in Eq. (7) with the spectral representation. Methods to identify the damaged component include:

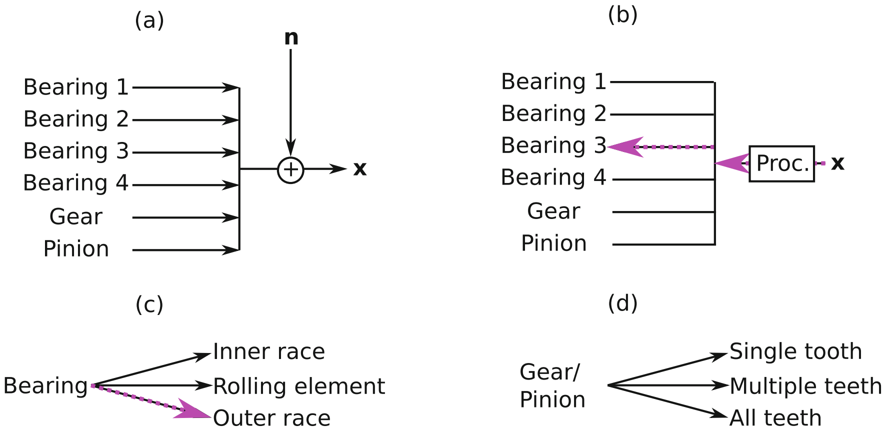


Fig. 4. The damage component identification problem: (a) The signal generation process for the gearbox in Fig. 1; (b) The damage subcomponent identification problem with bearing 3 identified using a Processing step; (c) Some of the potential bearing damage cases; (d) Some of the potential gear damage cases.

- The synchronous average can be used to extract synchronous deterministic components. If there are extraneous synchronous components, these components would contaminate the synchronous average and impede the damaged component's identifiability.
- The synchronous average of the squared envelope does not facilitate identifiability as it is sensitive to non-synchronous changes in the signal [3]. The synchronous median of the squared envelope and the synchronous geometric average of the squared envelope are much better suited as they are robust to non-synchronous impulsive components [3].
- Cyclostationary analysis tools such as the squared envelope spectrum, the spectral correlation and the improved envelope spectrum can be used to determine the characteristics of the component-of-interest [2]. Cyclostationary analysis methods for time-varying speed conditions are proposed in Refs. [17] and cyclostationary analysis methods for impulsive noise conditions are proposed in Ref. [6,18].
- In discrepancy analysis, the anomalies' localised behaviour is used to determine the source of the anomalous behaviour [19].
- In contrast to blind condition indicators, targeted condition indicators make it possible to focus on specific cyclic orders and can be used to target specific mechanical components [11]. The RMS for example is sensitive to any energy changes in the data and cannot be used to identify the damaged component and would therefore not be a reliable estimate of the fault severity.

To be able to perform damage identification, the rotational speed of a reference shaft and the system's kinematics (e.g. the gear mesh frequencies, the ball-pass outer race component of the bearing) need to be known [19]. The damaged component identifiability is also impeded if two components have the same statistical behaviour (e.g. both are first-order cyclostationary) and have very similar characteristic frequencies, e.g. if the same bearings are located on the same shaft. However, using multiple sensors might separate the contributions of the different sources if $\boldsymbol{H}$ in Eq. (4) is invertible.

3.4 Damage Mode Identification

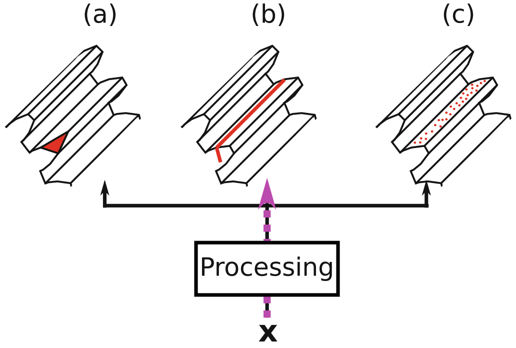


Fig. 5. The damage mode identifiability problem is illustrated when gear damage is present and only three damage modes are considered. The measured signal is processed to either identify (a) a tooth chip, (b) a tooth crack, or (c) pitting, with the tooth crack identified in this example.

In the damage mode identification phase, the mode of degradation (e.g. crack propagation in the gear tooth, pitting formation and spalling generation) is identified. The damage identification problem is shown in Fig. 5 for the case where there is gear damage in the gearbox. The degradation mode influences the remaining useful life of the system and therefore it is important to identify this for prognosis. Feng et al. [20] could distinguish between changes in the signals due to abrasive wear and pitting and therefore the damage mode is identifiable with their proposed procedure.

3.5 Fault Severity Quantification

In the fault severity quantification phase, the damage's size is estimated (e.g. characteristic crack length, material loss due to wear). This is important as the remaining useful life depends on the size of the damage [21] and this information needs to be known when using the appropriate degradation model. The ICS2 indicator, which measures the degree of second-order cyclostationarity, has a

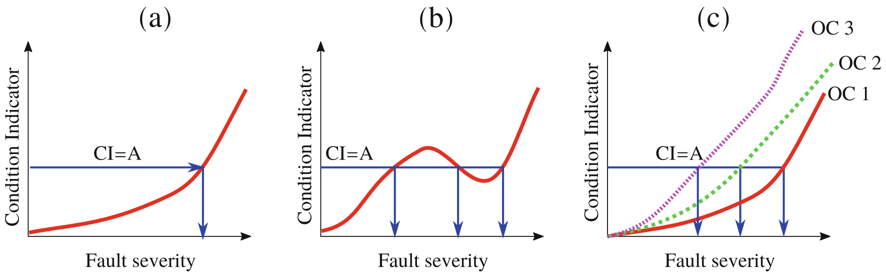


Fig. 6. The typical behaviour of condition indicators are shown. (a) The fault severity is identifiable for this specific case. (b) A condition indicator where the fault severity is not identifiable. (c) The fault severity is non-identifiable when a condition indicator is used that is sensitive to varying operating conditions. This behaviour is shown for three Operating Conditions (OC).

good correlation with the severity of the abrasive wear and pitting [20]. Three cases are shown in Fig. 6. These problems are often addressed in classification-based fault diagnosis, where the raw or processed data are mapped to a damage mode and/or a fault severity. The problem with this is that the degradation is continuous, not discrete, and often we do not have sufficient historical fault data to train the models [8]. Physics-based models can potentially aid with this task, but model calibration becomes an important consideration.

3.6 Remaining Useful Life (RUL) Estimation

The remaining useful life estimation process is shown in Fig. 7(a) for the ideal case and in Fig. 7(b) for the practical scenario. The larger the variance of the

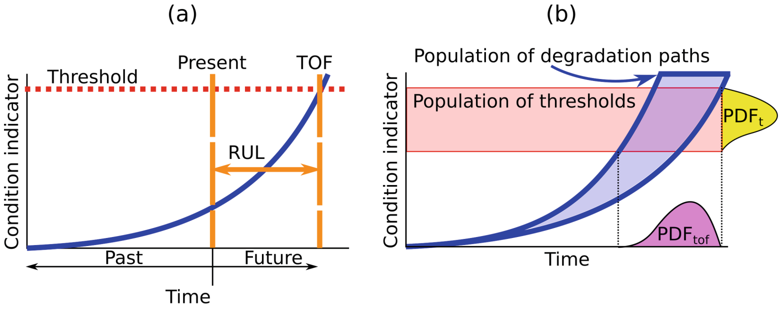


Fig. 7. The Remaining Useful Life (RUL) estimation process: (a) The conventional RUL estimation process is shown for a known threshold, a known degradation path and a known Time-Of-Failure (TOF). (b) The practical prognosis problem is illustrated as a population of failure thresholds, with the Probability Density Function (PDF) denoted by PDF_t, and a population of degradation paths. This makes us uncertain of the actual TOF, which is described by PDF_{tof}.

potential degradation path and the appropriate threshold (i.e. the more uncertain we are), the more difficult it becomes to identify the remaining useful life. Hence, it is important to perform the diagnosis phase properly to ensure that the remaining useful life estimates are reliable. Data-driven prognosis methods estimate the remaining useful life from the raw or processed signals directly using the available historical data [7,22]. However, a representative historical fault dataset is difficult to obtain; there are many potential failure modes and sufficient representative samples of each failure mode are required. Therefore, purely data-driven prognostics methods are difficult to implement in practice, which highlights the need for hybrid prognosis strategies [23].

4 Proposed Requirements for Fault Diagnosis and Prognosis Methods

In this work, the following requirements are proposed for the different phases of fault diagnosis and prognosis:

- Sensor placement: The sensors should be placed to ensure that the matrix $\boldsymbol{H}$ in Eq. (4) is as well-conditioned as possible. This means that if more sensors are added to the monitored machine, it should be ensured that the new sensor is as informative as possible (e.g. that the damage detection accuracy is increased [24]).
- Damage detection: Condition indicators should only change when the machine degrades and should be robust to changes in operating conditions. Robustness against varying operating conditions is possible by pre-processing the signals (e.g. [15]) or processing the condition indicators (e.g. [16]).
- Damage component identification: Condition indicators should only be sensitive to the presence of a subset of possible damage events (e.g. different bearing damage modes) to ensure that the condition indicators can be properly interpreted. A method to perform this task is presented in Ref. [3].
- Damage mode identification: Condition indicators should provide a unique representation for the different damage modes to ensure the damage mode is identifiable. The appropriate degradation model can subsequently be selected or inferred.
- Remaining useful life estimation: The RUL estimate should only be dependent on the component-of-interest and should not be influenced by changes in the condition of other machine components or changes in the machine's operating conditions. This can be performed by using a robust condition indicator.

The damage mode identification problem is especially challenging because the damage often manifests as weak components in the signal and there might be subtle differences between the different damage modes. However, Eq. (4) highlights that by using multiple well-placed sensors and by having some prior knowledge about the transmission paths (e.g. from a physics-based model) it could be possible to have a better view of the events, which could aid in identifying the damage mode and the severity of the damage.

5 Conclusions

In this work, we presented the identifiability problem for rotating machine fault diagnosis and prognosis and highlighted its significance for the different fault diagnosis and prognosis tasks. The importance of sensor location and the importance of using the appropriate signal processing methods are emphasised. We also proposed requirements for the development and implementation of fault diagnosis and prognosis methods. We believe these requirements can help to develop reliable condition-based maintenance methods for machinery such as wind turbines and will be investigated in more detail in future work.

References

1. Salameh, J.P., Cauet, S., Etien, E., Sakout, A., Rambault, L.: Gearbox condition monitoring in wind turbines: a review. Mech. Sys. Sig. Process. **111**, 251–264 (2018)
2. Randall, R.B., Antoni, J.: Rolling element bearing diagnostics—a tutorial. Mechan. Syst. Sig. Process. **25**(2), 485–520 (2011)
3. Schmidt, S., Zimroz, R., Heyns, P.S.: Enhancing gearbox vibration signals under time-varying operating conditions by combining a whitening procedure and a synchronous processing method. Tech. Syst. Sig. Process. **156**, 107668 (2021)
4. Antoni, J., Borghesani, P.: A statistical methodology for the design of condition indicators. Mech. Syst. Sig. Process. **114**, 290–327 (2019)
5. Booyse, W., Wilke, D.N., Heyns, S.: Deep digital twins for detection, diagnostics and prognostics. Mech. Syst. Sig. Process. **140**, 106612 (2020)
6. Kruczek, P., Zimroz, R., Antoni, J., Wyłomańska, A.: Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech. Syst. Sig. Process. **159**, 107737 (2021)
7. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J.: Machinery health prognostics: a systematic review from data acquisition to rul prediction. Mech. Syst. Sig. Process. **104**, 799–834 (2018)
8. Schmidt, S., Heyns, P.S.: An open set recognition methodology utilising discrepancy analysis for gear diagnostics under varying operating conditions. Mech. Sys. Sig. Process. **119**, 1–22 (2019)
9. Lin, J., Zhao, M.: A review and strategy for the diagnosis of speed-varying machinery. In: 2014 International Conference on Prognostics and Health Management, pp. 1–9. IEEE (2014)
10. Villaverde, A.F.: Observability and structural identifiability of nonlinear biological systems. Complexity **2019** (2019)
11. Smith, W.A., Borghesani, P., Ni, Q., Wang, K., Peng, Z.: Optimal demodulation-band selection for envelope-based diagnostics: a comparative study of traditional and novel tools. Mech.l Syst. Sig. Process. **134**, 106303 (2019)
12. Schmidt, S., Heyns, P.S., Gryllias, K.C.: An informative frequency band identification framework for gearbox fault diagnosis under time-varying operating conditions. Mech. Syst. Sig. Process. **158**, 107771 (2021)
13. Buzzoni, M., Antoni, J., d'Elia, G.: Blind deconvolution based on cyclostationarity maximization and its application to fault identification. J. Sound Vib. **432**, 569–601 (2018)
14. Lei, Y., Lin, J., Zuo, M.J., He, Z.: Condition monitoring and fault diagnosis of planetary gearboxes: a review. Measurement **48**, 292–305 (2014)

15. Schmidt, S., Heyns, P.S.: Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring. Measurement **149**, 106964 (2020)
16. Zimroz, R., Bartelmus, W., Barszcz, T., Urbanek, J.: Diagnostics of bearings in presence of strong operating conditions non-stationarity—a procedure of load-dependent features processing with application to wind turbine bearings. Mech. Syst. Sig. Process. **46**(1), 16–27 (2014)
17. Abboud, D., Baudin, S., Antoni, J., Rémond, D., Eltabach, M., Sauvage, O.: The spectral analysis of cyclo-non-stationary signals. Mech. Syst. Sig. Process. **75**, 280–300 (2016)
18. Borghesani, P., Antoni, J.: Cs2 analysis in presence of non-gaussian background noise-effect on traditional estimators and resilience of log-envelope indicators. Mech. Syst. Sig. Process. **90**, 378–398 (2017)
19. Heyns, T., Heyns, P.S., De Villiers, J.P.: Combining synchronous averaging with a gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox. Mech. Syst. Sig. Process. **32**, 200–215 (2012)
20. Feng, K., Smith, W.A., Borghesani, P., Randall, R.B., Peng, Z.: Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution. Mech. Syst. Sig. Process. **150**, 107258 (2020)
21. Zhao, F., Tian, Z., Bechhoefer, E., Zeng, Y.: An integrated prognostics method under time-varying operating conditions. IEEE Trans. Reliab. **64**(2), 673–686 (2015)
22. An, D., Kim, N.H., Choi, J.H.: Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliab. Eng. Syst. Saf. **133**, 223–236 (2015)
23. Liao, L., Köttig, F.: Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Trans. Reliab. **63**(1), 191–207 (2014)
24. Zhao, Z., Xu, Q., Jia, M.: Sensor network optimization of gearbox based on dependence matrix and improved discrete shuffled frog leaping algorithm. Nat. Comput. **15**(4), 653–664 (2016)

Comparative Study Between Experimental and Theoretical Frictional Power Losses of a Geared System

Maroua Hammami[1,2](✉), Olfa Ksentini[1,2], Nabih Feki[1,3], Mohamed Slim Abbes[1],
and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modelling and Manufacturing (LA2MP), University of Sfax,
National Engineering School of Sfax, BP 1173, 3038 Sfax, Tunisia
slim.abbes@enis.tn, Mohamed.haddar@enis.rnu.tn
[2] University of Gabes, Higher Institute of Industrial Systems of Gabes (ISSIG), 6029 Gabes,
Tunisia
{MarouaHammami123,olfa.ksountini}@issig.u-gabes.tn
[3] University of Sousse, Higher Institute of Applied Science and Technology of Sousse,
ISSATSo, 4003 Sousse, Tunisia
nabih.feki@issatso.u-sousse.tn

Abstract. In this study, a frictional dynamic gear model is propounded for a lubricated FZG C40 spur gear pairs. The proposed nonlinear dynamic model accounts the actual time variable gear mesh stiffness with considering the frictional effects of meshing gear teeth. This developed model aims to predict with high accuracy the gear power losses and to highlight the influence of the friction coefficient. Two listed cases were considered through introducing a constant friction coefficient and a local coefficient of friction in the developed dynamic model. The influence of the constant and local coefficient of friction on the frictional spur gears power losses has been evaluated. All key parameters like axle gear oil dynamic and kinematic viscosity, contact conditions between meshing gear teeth through equivalent radius of curvature, slide-to-roll ratio, maximum Hertzian pressure and tooth profile deviations on which the local friction coefficient depends are introduced in the model formulation. The simulation results were obtained through solving the motion equations of the generalized translational torsional coupled dynamic system by an iterative implicit scheme of Newmark. The results of simulation were compared to experimental results investigated on an FZG test rig under several operating conditions of load and speed.

Keywords: Frictional dynamic model · Coefficient of friction · Tooth profile deviations · FZG test rig

1 Introduction

The industry provides worldwide efforts to optimise mechanical systems aiming to decrease the energetic efficiency of equipment and save resources (Touret et al. 2020).

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 21–30, 2022.
https://doi.org/10.1007/978-3-030-85584-0_3

These systems suffer in their operation from power losses which occur in the main mechanical components and especially in the gear transmissions. For that, it is crucial to study the gear efficiency inside gearbox under dynamic conditions which can be useful to design reliable and efficient gearboxes.

In the reviewed studies, several dynamic models were developed. Their focus was the transmission errors, the dynamic overloads, the critical speed and mainly the vibration and noise behavior (Velex and Cahouet 2000, Wang et al. 2003, Diab et al. 2006, Tounsi et al. 2016). New details are introduced in gears dynamic models like the varying mesh stiffness, frictional effects and more degrees of freedom. The friction is a critical factor in the gear power loss estimations since it affects the gear life of the geared system. It influences not only vibration and noise behavior of the mechanical system but also its efficiency. For that, it is interesting to estimate the frictional power losses considering real operating conditions like dynamic effects, variable coefficient of friction under a range of load and speed.

The work aims to exploit the proposed dynamic model for C 40 spur gears to assess the dynamic and frictional effects in gears power loss at different operating conditions using two distinct coefficients of friction formulations (average and local COF). Power losses are investigated as they are proportional to the friction coefficient in the gear contact. A local friction coefficient is inserted in the mathematical formulation based on EHL approach (Xu and Kahraman 2007) and an average COF is determined through validation and calibration of quasi-static model based on experimental results (Hammami et al. 2018). A comparison between the predicted gear power losses in dynamic steady conditions using two friction coefficient approaches with experimental results performed using an FZG test rig are exposed and investigated.

2 Geared System Model

The developed geared system model is described in Fig. 1. The total model comprises four nodes with three degrees-of-freedom (DOFs) per node corresponding to 2 translational (v_j, w_j) and 1 torsional (θ_j) displacements. This dynamic model is composed of four main elements: (1) a motor and a load imposing the input torque (C_m) at the first node of the gear shaft and the resisting load (C_r) at the last node of the pinion shaft, respectively, (2) the input shaft 1 and the output shaft 2 are modelled by two-node Timoshenko beam elements with circular cross-sections, (3) the pinion gear pair simulated as two rigid cylinders related with nonlinear mesh stiffness ($K_m(t)$) and a dampener (C_m), (4) four rolling bearings supporting shafts introduced as lumped damping (C_{xj}, C_{yj}) and stiffness (K_{xj}, K_{yj}) components.

To evaluate the spur gear frictional dynamic behavior, an equation of motion of the generalized translational torsional coupled dynamic system was developed from Lagrange principle considering frictional effects (average and local coefficient of friction (μ). This motion equation does not consider gyroscopic and centrifugal effects. Tear surface roughness variation is introduced in the model by the tooth profile errors ($e(M_i)$) through the local coefficient of friction.

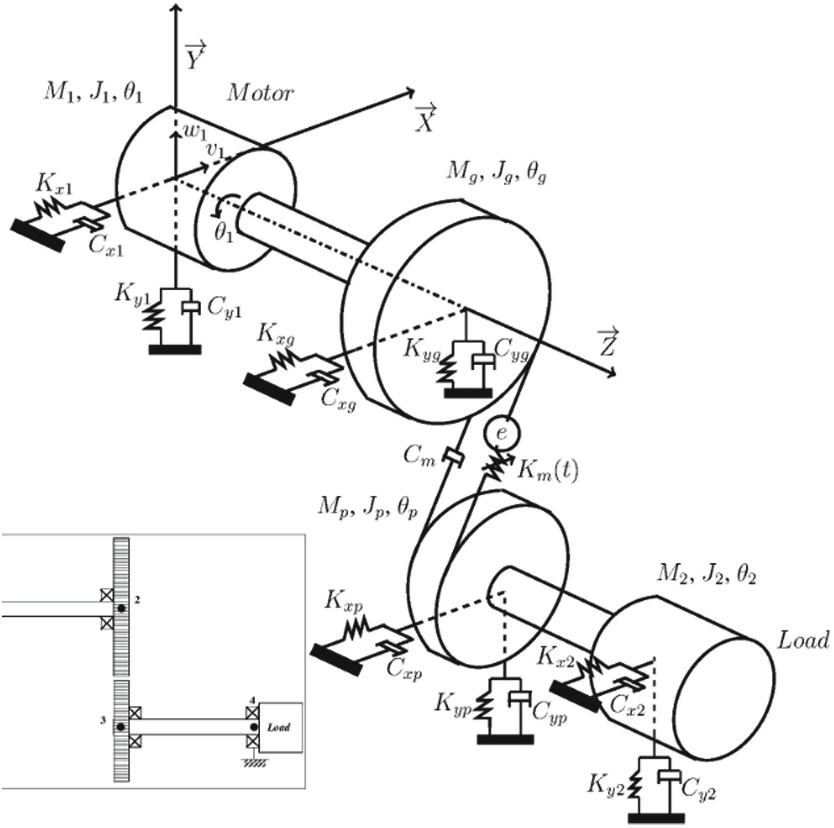


Fig. 1. 3D dynamic gear model (12 DOFs) (Hammami et al. 2019b).

After executing the differentiations in Lagrange equation the differential equation of the system is obtained:

$$M\ddot{X} + C\dot{X} + \big(K(t, X) + K_{\mu}(t, X, \mu)\big)X = F_1(t, X, \mu) + F_2(t, X, e(M)) \qquad (1)$$

where $X = \{v_1, w_1, \theta_1, v_g, w_g, \theta_g, v_p, w_p, \theta_p, v_2, w_2, \theta_g\}$ is the DOF's vector; M, C, $K(t, X)$ and $K_{\mu}(t, X, \mu)$ are the total mass, proportional damping, time-dependent nonlinear stiffness and frictional stiffness matrices, respectively; $F_1(t, X, \mu)$ is the input and resisting torques vector depending on internal frictional excitation and $F_2(t, X, e(M))$ is a tooth shape deviations excitation vector.

The variable tooth profile error along the action contact line and the theoretical contact line is considered similar for all teeth (see Fig. 2).

The corresponding tooth profile error defined as the difference between the theoretical and the real tooth profiles is induced at the point M_i using an estimated roughness profile for C 40 spur gear as shown in Fig. 3.

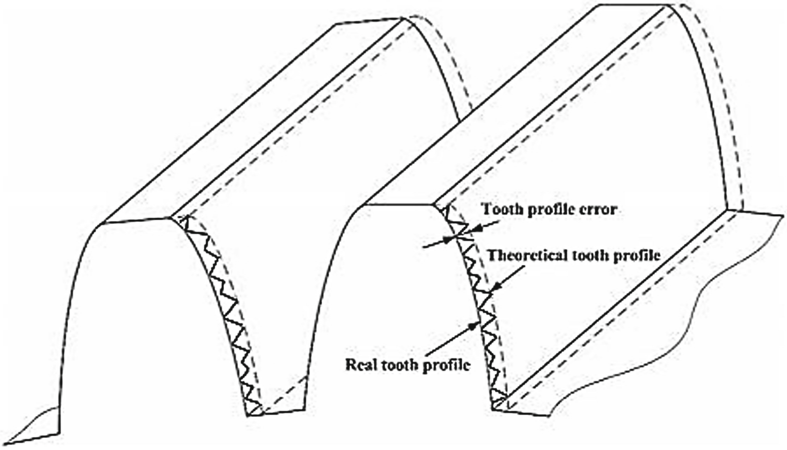


Fig. 2. Estimated tooth profile error (Feki et al. 2020)

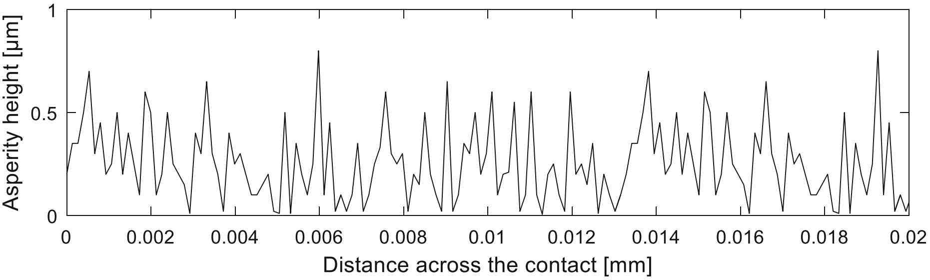


Fig. 3. C40 pinion root mean square roughness (Rq) in radial direction.

The local coefficient of friction (µ) formulation based on EHL model that experimentally validated by Xu (Xu 2005) is inserted in the gear dynamic simulations using the following equation:

$$\mu = e^{f(SR, P_h, \nu_0, S)} P_h^{b_2} |SR|^{b_3} V_r^{b_6} \mu_0^{b_7} R^{b_8} \tag{2}$$

$$f(SR, P_h, \nu_0, S) = b_1 + b_4 |SR| P_h log_{10}(\nu_0) + b_5 e^{-|SR| P_k log_{10}(\nu_0)} + b_9 e^S \tag{3}$$

where $SR, P_h, \nu_0, S, V_r, \mu_0, R$ are respectively the slide to roll ratio, maximum Hertzian pressure, oil dynamic viscosity, surface roughness ($e(M_i)$), rolling velocity, oil kinematic viscosity and equivalent radius of curvature. More details about Xu formulation can be found in reference (Xu 2005). The coefficients used in this formulation from b_1 to b_9 in Eqs. (2) and (3) are given as follow $b_{i:1..9} = \{-8.916465; 1.03303; 1.036077; -0.354068; 2.812084; -0.100601; 0.752755; -0.390958; 0.620305\}$.

3 Experimental set-up

In this study, a back-to-back FZG machine was used. It is composed of slave and test gearboxes in a locked mechanical loop (see Fig. 4). The test gearbox contains type C standard gears (characterized in Table 1). Detailed Gear description can be found in Ref. (Hammami et al. 2018). The gear set is splash lubricated with commercial 75W90-A oil whose properties are presented in Table 2.

ETH DRDL torque meter is used to measure the resistive torque generated in the system. Type K thermocouples are placed in the oil sump and outside of the gearbox to measure the operating temperature in several locations. Both the slave and the test gearbox will suffer from frictional gears power loss during experimental tests. Measuring friction in the contact of only test gearbox is challenging, especially in a closed mechanical loop such as the FZG test rig.

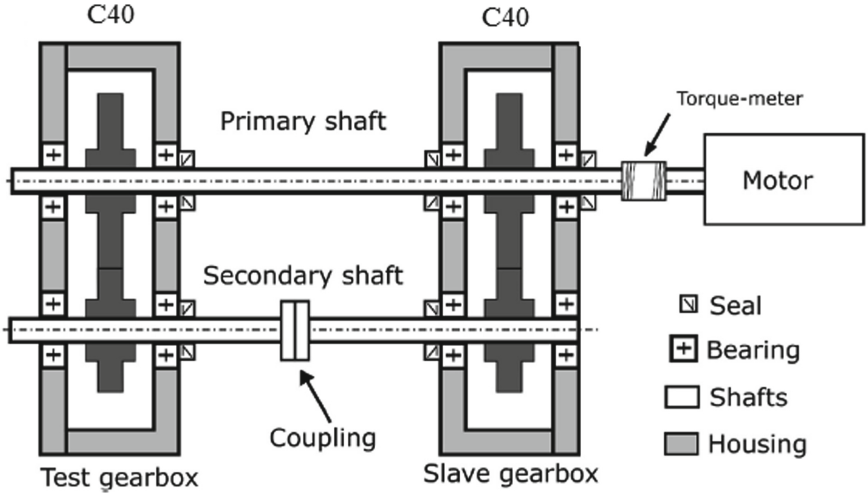


Fig. 4. Cross section representation FZG back-to-back gear test rig

During the test, the friction evolution in the gear contact will be measured based on the total torque loss values. As the rotational speed is controlled, power losses are immediately extracted from torque measurements. To study the correlation between friction and the variations of power losses, an experimental C40 gears power loss lubricated with 75W90-A oil and an average coefficient of friction were determined and validated in author previous work (Hammami et al. 2018). Experimental conditions are detailed in Table 3. The selected load stages K5, K8 and K11 were characterized with a constant torque on the test pinion. The constant torque is obtained through using calibrated weight of the FZG machine (Hammami et al. 2018).

Table 1. FZG type C40 gear definition

Type C gear set				
Parmeters	Symbol	Pinion value	Wheel value	Unit
Tooth width	B	40		[mm]
Number of teeth	Zi	16	24	[–]
Module	M	4.5		[mm]
Operating center distance	A	91.5		[mm]
Pressure angle	A	20		[deg]
Contact ratio	ε_α	1.44		[–]
Surface roughness	R_q	0.5		[µm]

Table 2. Oil properties (Hammami et al. 2019a)

Oil properties		
Reference	75W90-A	Unit
Density	870	[kg/m^3]
Kinematic viscosity@70	36.7	[cSt]
Kinematic viscosity@100	16.37	[cSt]

Table 3. Operating conditions during experimental tests

Load stage	K5	K8	K11
Pinion torque [Nm]	70	172	319
Oil temperature [°C]	80		
Wheel rotating speed [rpm]			
250	500	1000	1750

4 Experimental and theoretical results

To validate the dynamic proposed models, several measurements were performed on the FZG test rig.

The simulation results shown in this work are presented under stabilized operating conditions. As an example, the total power loss for C40 gear at 1000 rpm under K8 load stage considering local COF presented some variations at the beginning as transient regime and became steady from a definite time value (see Fig. 5).

The analytical dynamic model results are presented and quantified under only a stable regime. Before, the power loss predictions, an important key parameter should be quantified which the local coefficient of friction is considering the tooth profile error.

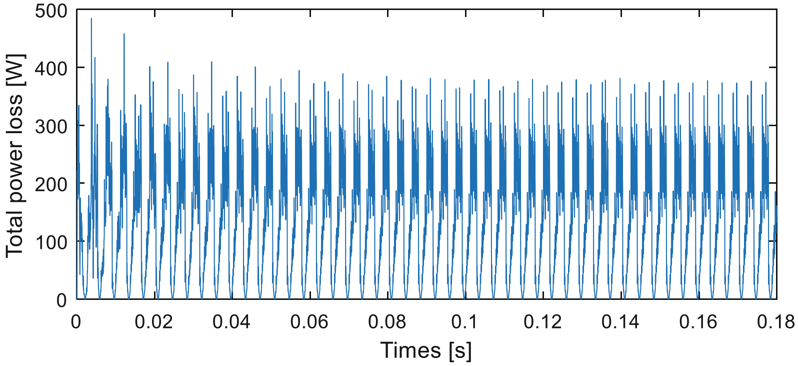
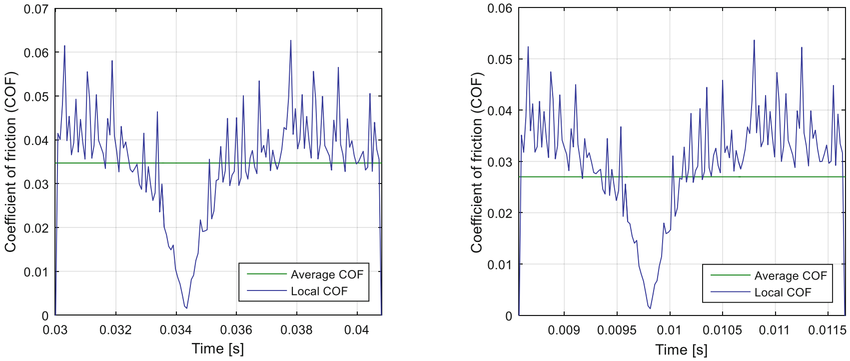


Fig. 5. Total frictional power loss considering local coefficient of friction for C40 at K8–1000 rpm

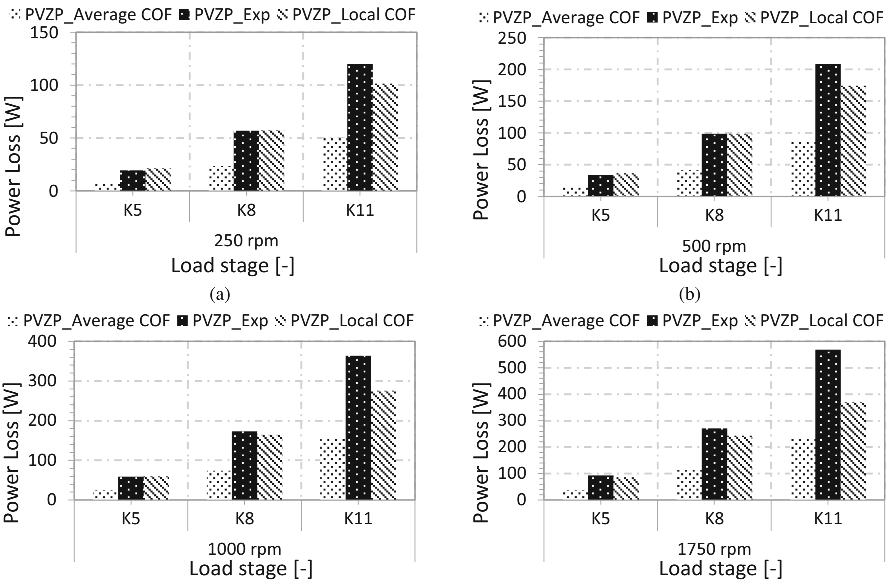
The experimental (average COF found in (Hammami et al. 2019b)) and simulated (local COF) coefficient of friction behavior along the path of contact for C 40 spur gear is shown in Fig. 6.



(a) (b)

Fig. 6. Average and local coefficients of friction for the C40 gear geometry: (a) COF at K8/500 rpm, (b) COF at K8/1750 rpm.

These results are performed under K8 load stage and at the following rotational speed: (a) 500 rpm, (b) 1750 rpm. Under the same load, the COF values slightly decrease with increasing rotational speed. The shape and the amplitudes are significantly different between the constant and local COF which can be explained with the no-time dependency of the average COF. The local COF is expected to evolve with surface roughness. It can be observed that the predicted μ values are influenced by the dynamic effects. The local COF is close to zero in the pitch point where no relative sliding velocity and show lager oscillations when the contact takes place near the edge where the contact pressure is high.



(a) (b) (c) (d)

Fig. 7. Average of the RMS of the experimental gears power loss vs. dynamic model simulation of RMS gears power loss considering average COF and local COF for C40 spur gears: (a) RMS gears power losses at 250 rpm, (b) RMS gears power losses at 500 rpm, (c) RMS gears power losses at 1000 rpm and (d) RMS gears power losses at 1750 rpm

Since the friction coefficient is quantified, it is possible to calculate the frictional meshing gear power loss which can be directly calculated from the gear dynamic model using the following equation (Hammami et al. 2018):

$$P_{VZP}(M_i) = F_N(M_i).\mu(M_i).V_{12}(M_i)$$

where $\mu(M_i)$ represents the average or the local friction coefficient, $V_{12}(M_i)$ is the instantaneous sliding velocities and $F_N(M_i)$ is the dynamic normal load.

The comparison between the simulated gear power losses using local and average COF and the experimental results are presented in Fig. 7.

In this figure, the RMS of the power losses under three FZG load stages (K5, K8 and K11) and four rotational speed (250, 500, 1000 and 1750 rpm) at stable conditions are exposed.

From K5 ($C_m = 69, 98\ Nm$) to K11 ($C_m = 319, 15\ Nm$), the RMS power losses increases due to the increase of µ values which are directly related to the pressure of the contact surfaces.

The presented results show high values when the rotational speed increases from 250 rpm to 1750 rpm because of the increased sliding velocities.

A relevant error is observed over 57% between the RMS power loss using average COF and the experimental results proving the limited prediction of this dynamic model (see Table 4). A constant COF should be used carefully since it was limited to certain system parameters.

However, promising results are expected since a local coefficient of friction was developed for large ranges of parameters study like gears operating conditions, surface roughness and contact surfaces. It is noticed that the introduction of the local COF in the dynamic model to account all relevant factors mainly the tooth profile errors increases impressively the prediction of the developed dynamic model.

The RMS power loss using local COF presents closer results to experimental values with low errors (see Table 4). A significant reduction between error 1 (PVZP_average COF/PVZP Exp) and error 2 (PVZP_Local COF/PVZP Exp) is observed. Since the gear power loss is an important key to evaluate the system mechanical efficiency, the frictional gears dynamic model with the highest prediction accuracy should be used.

Table 4. Errors between experiments and simulations of an RMS friction power loss using average and local COF for C40 spur gears for all operating conditions.

Rotational speed [rpm]	Load stage [–]	Error 1 [%] (PVZP_average COF/PVZP Exp)	Error 2 [%] (PVZP_Local COF/PVZP Exp)
250	K5	58.04	9.67
	K8	58.15	0.4
	K11	58.16	15.33
500	K5	57.88	7.84
	K8	57.91	0.19
	K11	57.92	16.39
1000	K5	57.33	0.45
	K8	57.32	4.77
	K11	57.47	24.23
1750	K5	58.39	6.9
	K8	58.51	10.21
	K11	58.51	35.1

5 Conclusion

A dynamic gear model for C40 spur gear was developed to evaluate the total power loss considering frictional effects through a constant and local friction coefficient. The friction coefficient shows an important role in the prediction of an accurate meshing spur gear power loss. To validate the developed dynamic model, several experimental results were used. A clear difference between simulations with local and constant COF is revealed. This observation supports the time COF dependency affecting the Hertzian pressure and the tooth profile deviations. Furthermore, a good correlation is presented between the predicted frictional gears power loss using EHL formulation with experimental results.

References

1. Touret, T., Changenet, C., Ville, F., Cavoret, J.: Experimental investigation on the effect of micropitting on friction-Part 2: Analysis of power losses evolution on a geared system. Tribol. Int. **153**, 106551 (2020). https://doi.org/10.1016/j.triboint.2020.106551
2. Velex, P., Cahouet, V.: Experimental and numerical investigations on the influence of tooth friction in spur and helical gear dynamics. J. Mech. Des. **122**, 515–522 (2000). https://doi.org/10.1115/1.1320821
3. Wang, J., Li, R., Peng, X.: Survey of nonlinear vibration of gear transmission systems. Appl. Mech. Rev. **56**, 309–329 (2003). https://doi.org/10.1115/1.1555660
4. Diab, Y., Ville, F., Velex, P.: Prediction of power losses due to tooth friction in gears. Tribol. Trans. **49**, 260–270 (2006). https://doi.org/10.1080/05698190600614874
5. Tounsi, M., Beyaoui, M., Abboudi, K., Feki, N., Walha, L., Haddar, M.: Influence of uncertainty in aerodynamic performance on the dynamic response of a two stage gear system. J. Theor. Appl. Mech. **54**(2), 601–612 (2016). https://doi.org/10.15632/jtam-pl.54.2.601
6. Xu, H., Kahraman, A.: Prediction of friction-related power losses of hypoid gear pairs. Proc. Inst. Mech. Eng. K: J. Multi-body Dyn. **221**(3), 387–400 (2007). https://doi.org/10.1243/14644193JMBD48
7. Hammami, M., Fernandes, C.M., Martins, R., Abbes, M.S., Haddar, M., Seabra, J.: Torque loss in FZG-A10 gears lubricated with axle oils. Tribol. Int. **131**, 112–127 (2018). https://doi.org/10.1016/j.triboint.2018.10.017
8. Hammami, M., Abbes, M.S., Martins, R., Seabra, J.H.O., Haddar, M.: Losses in thrust ball bearings lubricated with axle gear oils. In: Fakhfakh, T., Karra, C., Bouaziz, S., Chaari, F., Haddar, M. (eds.) ICAV 2018. ACM, vol. 13, pp. 23–36. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-94616-0_3
9. Hammami, M., Feki, N., Ksentini, O., Hentati, T., Abbes, M.S., Haddar, M.: Dynamic effects on spur gear pairs power loss lubricated with axle gear oils. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **234**(5), 1069–1084 (2019). https://doi.org/10.1177/0954406219888236
10. Feki, N., Hammami, M., Ksentini, O., Abbes, M.S., Haddar, M.: Frictional dynamic model predictions of FZG-A10 spur gear pairs considering profile errors. Proc. Inst. Mech. Eng. J: J. Eng. Tribol. **235**(7), 1390–1404 (2020)
11. Xu, H.: Development of a generalized mechanical efficiency prediction methodology for gear pairs. Doctoral dissertation, The Ohio State University (2005)

Early Damage Detection in Planetary Gear Transmission in Down-Time Regime

Ayoub Mbarek[1,2,3](✉), Ahmed Hammami[1], Alfonso Fernández Del Rincón[2],
Fakher Chaari[1], Fernando Viadero Rueda[2], and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modelling and Production (LA2MP),
National School of Engineers of Sfax, BP1173, 3038 Sfax, Tunisia
`mohamed.haddar@enis.rnu.tn`
[2] Department of Structural and Mechanical Engineering,
Faculty of Industrial and Telecommunications Engineering, University of Cantabria,
Avda de Los Castros S/N 39005, Santander, Spain
`{alfonso.fernandez,fernando.viadero}@unican.es`
[3] LMEE, Univ Evry, Université Paris-Saclay, 91020 Evry Cedex, France

Abstract. Planetary Gear systems (PGs) are widely used in many mechanical engineering applications such as industrial robots, mining machinery and aircraft transmission. During their fitness they can be affected by several kinds of damage for example crack and spalling that are the most systematic malfunction. As those systems are frequently provoked in non-stationary condition defined as variable loading and variable speed regimes, their diagnosis under these regimes using the vibration response presents a huge challenge for research and investigation. In this work, the vibration response of two-stage planetary gears in presence of spalling faults are investigated during the down-time excitation regime. The system is already preloaded, and a ramp of speed is applied on the system using a frequency inverter and the sensors are mounted on the ring gear components. A series of measurements tests are carried out using an experimental test bench where a spalling defect is introduced to one of the planet components. The paper examines the damages detection using time domain, frequency domain and order domain analysis approaches. In the end, the obtained results that are processed with angular order domain analysis based on the angle domain order tracking method shows an interesting result compared to others method.

Keywords: Two stages planetary gear · Spalling defect · Down-time regime · Vibration signal

1 Introduction

Planetary gear transmissions are extensively used in many industrial and mechanical engineering applications like CNC machines, car transmissions and wind turbine gearboxes. During their service, they are confronting several running conditions. During the last few decades, diagnosis and condition monitoring of planetary gear systems

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 31–37, 2022.
https://doi.org/10.1007/978-3-030-85584-0_4

become a challenge for research and industrial manufacturing with the aim of reducing maintenance cost. Many research papers were published in which the authors focus on modelling of defects and, the impact of defects on the system using numerical or experimental data signal processing tools. Also, some authors focus on prognostics [1]. Concerning the modelling of spalling defects, Saxena et al. [2] studied the influence of rectangular spalling shape and its location on meshing function. Other researchers agreed that the presence of tooth spalling failure can be considered by a decrease in amplitude on meshing functions [3, 4]. The influences of gear failure on dynamic behaviour are studied by many researchers. In Ma and Chen [5], the time domain and spectrum analysis are considered to study the influence of spalling failure on the dynamic response of one stage gear transmission.

Moreover, many studies addressed gear fault diagnostic is addressed [6, 7]. First, the focus was either on extracting the frequency [7, 8]. In a second context, features of the frequency domain were used to determine the failure gear [9]. A third, the wavelet transformation methods were developed for failure detection [10–12]. In addition, in Hong and Dhupia [13], a novel approach for gear failure detection based on.

A novel approach for gear failure detection was based on combining the kurtosis and the fast-dynamic time warping method is proposed [13]. Other analsis trategies were using the Hilbert Huang transform to characterize the gear faults in cases of variable speed condition. In this context, an experimental study focused on the dynamic behaviour of a planetary gear system in a downtime regime in the presence of a spalling defect is analysed using the order tracking method [14].

2 Experimental Test Bench

To study the impact of tooth failure on the vibration response of planetary gear transmission, a spalling damage on one of the teeth of planet gear was artificially manufactured using an electro erosion machine. Figure 1 displays the test bench where the tests are achieved. This bench was used for several research purposes by Mbarek et al. [15] and it is well described in [16, 17] (Fig. 2).

Fig. 1. Experimental test bench

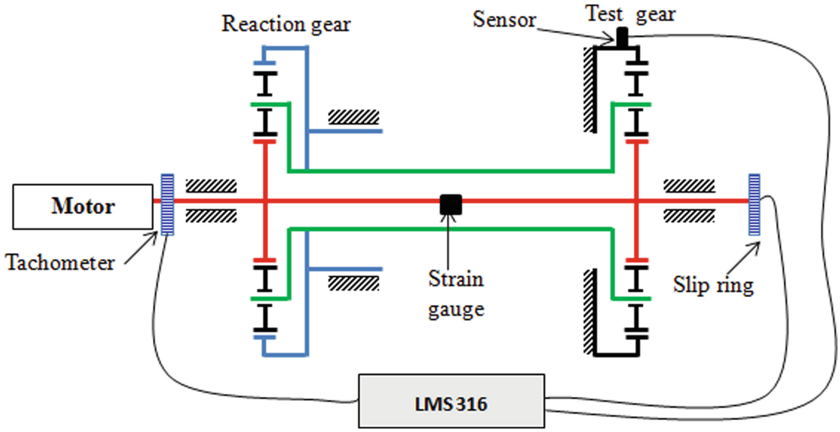


Fig. 2. Experimental test bench

Table 1. Basic dimension planetary gear

	Sun	*Planet (3)*	*Ring*	*Carrier*
Number of teeth	16	24	65	–
Module m]	0.004	0.004	0.004	–
Base diameters m]	0.06	0.09	0.24	0.1
Tip diameters m]	0.06	0.10	0.25	0.1
Mass kg]	0.4	1.2	28.1	3.6
Moment of inertia kg·m^2]	$35 \cdot 10^{-6}$	$20 \cdot 10^{-6}$	$69 \cdot 10^{-6}$	$21 \cdot 10^{-6}$

3 Results

Measurement tests are carried out when the system is exited in a downtime regime, during this regime, a ramp of speed is imposed on the motor according to the speed profile given in [6]. Figure 3 shows that the vibration amplitude decreases with the time that is explained by the fact that the period of the gear mesh stiffness increase as the speed decrease.

Results presented in Fig. 3 shows that time analysis cannot give a piece of interesting information about the faulty components. Therefore, a time-frequency analysis based on STFT is used to determine the features of the sapling defect as illustrated in Fig. 4.

The presence of an oblique segment that shows the evolution of the gear meshes frequency in time. In addition, the system can be excited by critical resonances frequency presented by the vertical lines like the lines presented on the frequencies 152 Hz, 346 Hz.

The STFT presentation is limited; there is no indication about the damaged tooth component. For this reason, an angular approach is adopted. Figure 5 shows the order map of the acceleration of the test ring; a vertical line shows the meshing order which is surrounded by their harmonic.

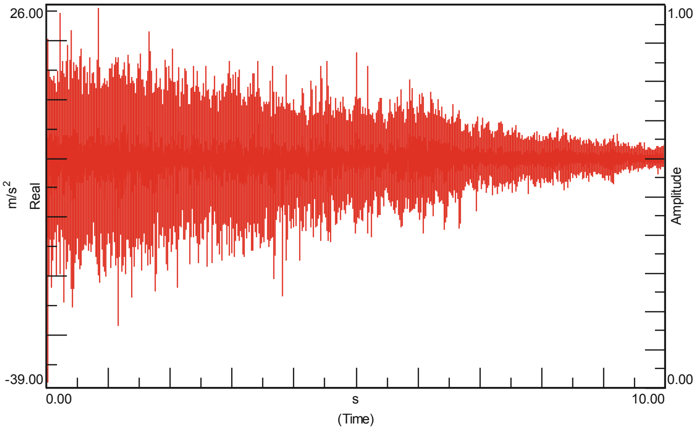


Fig. 3. Time response of the acceleration measured on the test ring

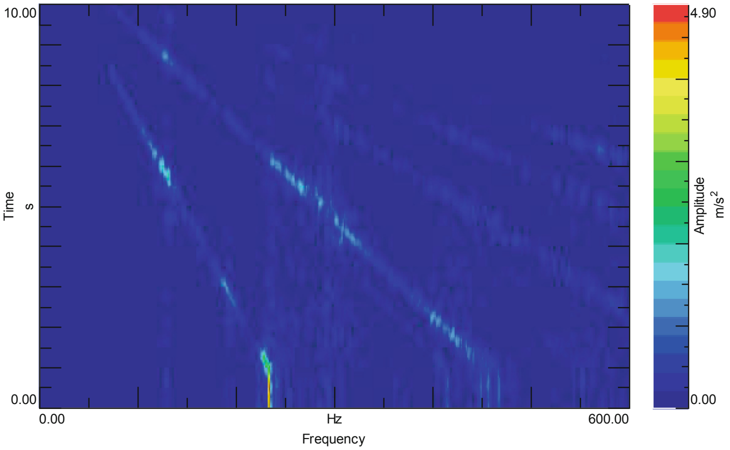


Fig. 4. STFT of the acceleration measured of the test ring

Respecting the number of teeth of each component displayed in Table 1 and the formula given by Mbarek et al. [14], the order associated with the planet is Op 0,4 and the mesh order O_m. To well visualize the order associated with the spalling defects, a caption around an order bandwidth [0-10] is established as displayed in Fig. 6.

The orders related to the faulty planet are presented as mentioned by the black cursor (Fig. 6). This fact is explained by that the damaged tooth enters in contact with the other component; the fault is reflected by a reduction of the meshing function. In this case, the reduction is presented in the ring planet and the sun planet mesh function which influences directly on the accelerations. Also, the meshing order is well observed around the 9th order as shown in Fig. 6. In fact, order domain presentation shows a good capability to detect faults in non-stationary condition, based on gear order features; this

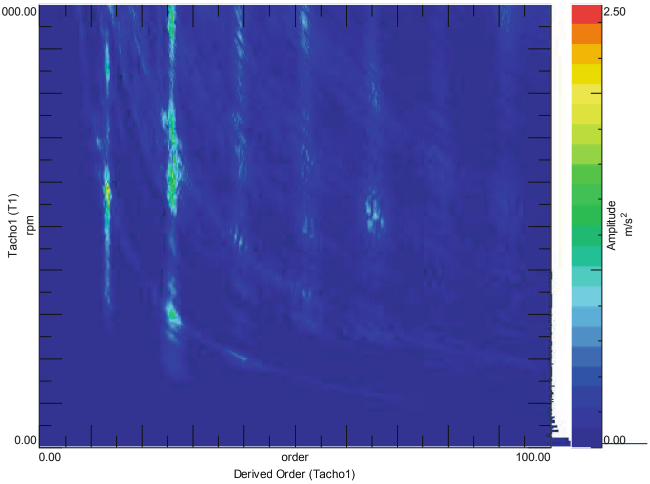


Fig. 5. Order map of the acceleration measured on the test ring

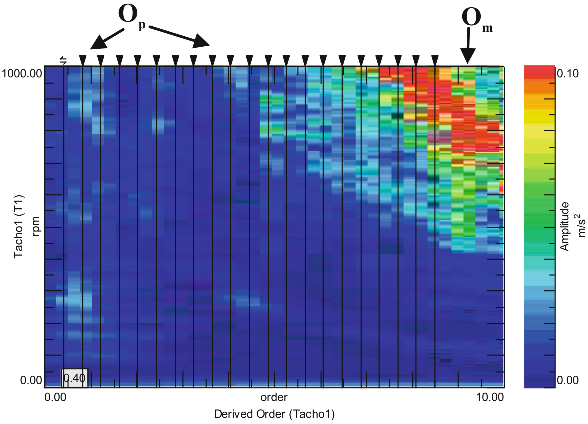


Fig. 6. Zoom of the order map

method gives a piece of information about the faulty components compared to the time and the spectrum response.

4 Conclusion

In this work, an experimental study on the dynamic behaviour of planetary gear transmission was investigated. A series of measurements were carried out using an experimental test bench composed of two-stage PGs mounted back-to-back in presence of a spalling defect. This failure was introduced to one of the three planets and the system was driven during the down-time regime. In fact, the recorded time response signal of the test ring is presented and analysed via different method. First, the time response proved that the signal is characterized by higher vibration amplitude due to the defect. It is worth noting that this analysis approach is limited because it is not informative enough when it consernes fault component. For this reason, frequency-time analysis based on STFT presentation is adopted in the next step, this technique reveals the evolution of the gear mesh frequency in the time-frequency domain while there is no clear indication of the defect. In addition, an order tracking method based on the angular method was used and adopted using the signal of the tachometer. Based on the formula related to each component, order map presentation indicates the defect's location as well as the meshing order. Additional work need to be realized in order to model and include defect in a munerical model. Also, the Order tracking method will be modelled to processed with the theoretical signal.

Acknowledgements. "The authors would like to acknowledge Project DPI2017-85390-P funded by the Spanish Ministry of Economy, Industry, and Competitiveness for supporting this research."

References

1. Farhat, M.H., Chiementin, X., Chaari, F., Bolaers, F., Haddar, M.: Digital twin-driven machine learning: ball bearings fault severity classification. Measure. Sci. Technol. **32**(4), 044006 (2021)
2. Saxena, A., Parey, A., Chouksey, M.: Time varying mesh stiffness calculation of spur gear pair considering sliding friction and spalling defects. Eng. Fail. Anal. **70**, 200–211 (2016)
3. Chaari, F., Fakhfakh, T., Haddar, M.: Dynamic analysis of a planetary gear failure caused by tooth pitting and cracking. J. Fail. Anal. Prev. **6**(2), 73–78 (2006)
4. Maatar, M.: Contribution a I'analyse du comportement dynamique de reducteurs a engrenages avec ecarts de forme et defauts de montage. Doctoral dissertation, These de Doctoral. INS A Lyon (1995)
5. Cheng, Z., Hu, N.: Quantitative damage detection for planetary gear sets based on physical models. Chinese J. Mech. Eng. **25**(1), 190–196 (2012)
6. Zhang, J., Dhupia, J.S., Gajanayake, C.J.: Stator current analysis from electrical machines using resonance residual technique to detect faults in planetary gearboxes. IEEE Trans. Industr. Electron. **62**(9), 5709–5721 (2015)
7. Kia, S.H., Henao, H., Capolino, G.A.: Fault index statistical study for gear fault detection using stator current space vector analysis. IEEE Trans. Ind. Appl. **52**(6), 4781–4788 (2016)
8. Glowacz, A., Glowacz, W.: Vibration-based fault diagnosis of commutator motor. Shock Vibration (2018)
9. Dhamande, L.S., Chaudhari, M.B.: Compound gear-bearing fault feature extraction using statistical features based on time-frequency method. Measurement **125**, 63–77 (2018)
10. Wang, C., Gan, M., Zhu, C.: A supervised sparsity-based wavelet feature for bearing fault diagnosis. J. Intell. Manuf. **30**(1), 229–239 (2016). https://doi.org/10.1007/s10845-016-1243-9
11. Wang, C., Gan, M., Zhu, C.: Fault feature extraction of rolling element bearings based on wavelet packet transform and sparse representation theory. J. Intell. Manuf. **29**(4), 937–951 (2015). https://doi.org/10.1007/s10845-015-1153-2
12. Hong, L., Dhupia, J.S.: A time domain approach to diagnose gearbox fault based on measured vibration signals. J. Sound Vib. **333**(7), 2164–2180 (2014)

13. Wu, T.Y., Chen, J.C., Wang, C.C.: Characterization of gear faults in variable rotating speed using Hilbert-Huang transform and instantaneous dimensionless frequency normalization. Mech. Syst. Signal Process. **30**, 103–122 (2012)
14. Mbarek, A., Hammami, A., Del Rincon, A.F., Chaari, F., Rueda, F.V., Haddar, M.: Effect of load and meshing stiffness variation on modal properties of planetary gear. Appl. Acoust. **147**, 32–43 (2019)
15. Mbarek, A., et al.: Comparison of experimental and operational modal analysis on a back to back planetary gear. Mech. Mach. Theory **124**, 226–247 (2018)
16. Mbarek, A., Hammami, A., Del Rincon, A.F., Chaari, F., Rueda, F. V., Haddar, M.: Effect of gravity of carrier on the dynamic behavior of planetary gears. In: International Conference Design and Modeling of Mechanical Systems (pp. 975–983). Springer, Cham (2017)
17. Mbarek, A. et al.: Dynamic behavior of back to back planetary gear in presence of pitting defects. In: International Conference on Acoustics and Vibration (pp. 16–22). Springer, Cham, March 2018

The Role of Untangled Latent Spaces in Unsupervised Learning Applied to Condition-Based Maintenance

Daniel N. Wilke(✉), P. Stephan Heyns, and Stephan Schmidt

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
`nico.wilke@up.ac.za`
`https://www.up.ac.za/centre-for-asset-integrity-management`

Abstract. Advances in signal processing are complemented by advances in machine and deep learning and vice versa. In general, machine and deep learning are employed as discriminative models within a supervised setting. Progress in unsupervised generative modelling allows for generative models to be employed in discriminative (discrete classes) and deviation (continuous deviation from a baseline) tasks. This only requires the samples to be chronological. Discriminative and deviation analysis is usually based on the reconstruction loss. However, this is limited as it offers only a single scalar from which inference can be made. Generative models do, however, learn a latent representation of the data from which additional scalars can be derived. Whether these derived scalars are informative depends on the quality of the latent representations. Most learning algorithms derive a latent representation that efficiently explains the variance in the data, which can be informative when the property of interest is well explained by variance. Alternatively, a lesser known class of learning algorithms aim to learn a latent representation that aims to identify sources in the data. Hence, given the same data, an infinite number of latent representations are possible, of which only a fraction are informative. We consider three classes of latent spaces that are stochastic, entangled and untangled. Furthermore, we highlight the importance of untangled latent spaces to obtain informative signals for condition monitoring.

Keywords: Signal processing · Generative modelling · Unsupervised learning · Latent variable · Untangled latent spaces

1 The Intersection Between Signal Processing and Learning Models

Signal processing in condition-based maintenance (CBM) is primarily concerned with extracting and analysing informative features from raw time-series data that enable diagnosis and prognosis to prevent asset failure and downtime, as shown in Fig. 1. Given sensor data, $\mathbf{x}(t)$, two questions are raised:

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 38–49, 2022.
https://doi.org/10.1007/978-3-030-85584-0_5

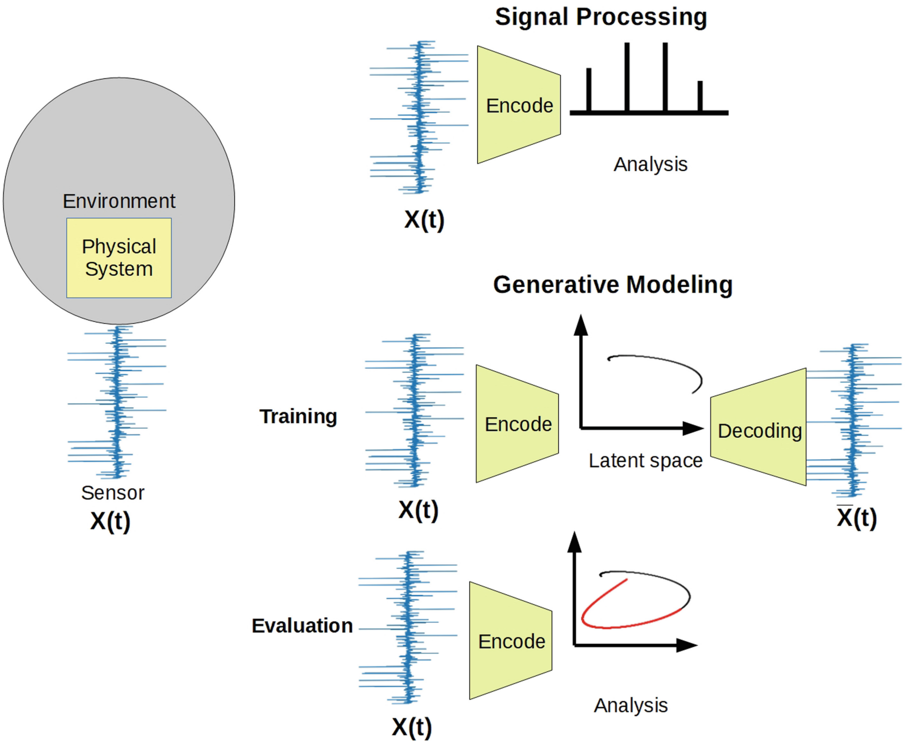


Fig. 1. Signal processing viewed as the encoding of useful or informative information from raw sensed signal data that the analyst can analyse. Signal processing is contrasted against the typical autoencoding training of generative models.

1. What is the information content of the signal, i.e. what of the physical system can be identified from the signal, and
2. Is the identifiable information in the signal captured such that the informativeness is maximised?

In signal processing, the analyst aims to isolate identifiable information from $\mathbf{x}(t)$ for analysis. The main aim is to extract useful information that informs an assets' condition. This task of separating identifiable information or extracting relevant information for the condition monitoring problem can be seen as encoding, followed by the encoded signal analysis.

The analyst's encoding task is equivalent to finding a transformation or representation that enhances specific information and attenuates extraneous signal components. This makes it possible to process the raw time-series signal into a more meaningful representation. Examples include the wavelet transform and the spectral coherence of the signal [1,2].

Alternatively, a time-series signal can be viewed as high dimensional data, i.e. as an n-dimensional vector, $\mathbf{x}(t) \in \mathbb{R}^n$, where n corresponds to the number of data points in the signal. The analyst is now tasked with finding a lower-dimensional representation or features that is/are informative. A single n-dimensional observation $\mathbf{x}(t)$ does not in itself enable the analyst to find a lower-dimensional representation. However, several strategies exist and have been employed to recast a single n-dimensional observation into multiple correlations/observations of high dimensional data. The auto-correlation matrix recasts $\mathbf{x}(t)$ into $\mathbf{X}_{AC}(t) \in \mathbb{R}^{n \times n}$ [3], while through a sliding window approach $\mathbf{X}(t) \in \mathbb{R}^{m \times l}$ recasts $\mathbf{x}(t)$ into m observations of window length l [4–6]. This approach is also used to obtain the Hankel matrix. Given some $\mathbf{X}(t)$, the aim is to find a lower-dimensional representation that enhances specific information and suppresses components unrelated to the component of interest.

In signal processing, irrespective of whether the data is being transformed or projected, the analyst is essentially conducting the following operations: finding informative transformations (e.g. [7]) and features or condition indicators (e.g. [8]), whereafter they are analysed for damage. This requires knowledge and mastering of signal processing principles, experience and extensive knowledge of the physical mechanisms that generate the measured signals.

In artificial intelligence, statistical learning [9], machine learning [10] and deep learning [11] are sub-domains that focus on developing generative models. Incidentally, generative modelling is the task of transforming high-dimensional data to a lower-dimensional representation or latent space through encoding, as shown in Fig. 1. Samples in this latent space can be resonstructed back to the high-dimensional space through decoding. This encoding-decoding process is used in training where the error between the original signal $\mathbf{x}(t)$ and the reconstructed signal $\bar{\mathbf{x}}(t)$ is minimised. Hence, the task of encoding is common between generative modelling and signal processing. An additional intersection being the analysis of the latent variables to inform on the condition of an asset.

Therefore, it is critical to find latent spaces that are informative and ensure that the latent space components correspond to distinct physical processes. In CBM under the effects of time-varying environmental and operational conditions (EOCs), this implies that time-varying components can be isolated and identified as independent latent representations [12,13]. The challenge in generative modelling is to encode informative latent spaces. Hence, by applying generative modelling with untangled latent spaces to CBM, the aim is to automate signal processing tasks that require extensive domain knowledge.

2 Process of Lower-Dimensional Representations

The process of finding lower-dimensional representations entails two steps, as shown in Fig. 2. Firstly, by finding an informative, low dimensional coordinate system, and secondly, by projecting the high dimensional data onto the low dimensional coordinate system. Depending on the generative modelling, the coordinate system can be linear (see Fig. 2), i.e. the coordinate system forms a basis that a set of vectors can describe. Some transformation function can describe a nonlinear or curvilinear coordinate system (see Fig. 2).

Statistical learning specifically aims to alleviate the analyst from finding features to merely supervising the task of finding features, a.k.a. feature engineering. In particular, statistical learning restricts itself to finding a linear coordinate system, i.e. linear basis to represent the latent space. Machine learning has similar aims, but the latent space is described by nonlinear coordinate systems to which the raw signals can be transformed. This enables the analyst to focus on analysing the projected signals instead of first finding informative features. Deep learning has more ambitious goals. It aims to automate the processing, feature engineering, and analysis of raw time-series signals, i.e., automate the entire signal processing value chain.

Hence the task of statistical, machine and deep learning is to find a coordinate system to represent the latent space onto which a time-series signal can be projected for analysis. This is achieved by defining an appropriate optimisation problem. Here, the analyst needs to decide a proper optimisation model, e.g. primal problem formulation or dual problem formulation, to solve a formulated min-max, maximisation or minimisation problem [14]. The choice of the loss function, norm (e.g. l_1-norm and l_2-norm), regulariser (e.g. l_2-norm weight norm penalty (ridge regression), or l_1-norm weight penalty (LASSO for sparsity)) and constraints (e.g. orthogonality of coordinates) allows the analyst to craft and target specific characteristics in the data that are assumed to be informative. This problem can be cast within a statistical language by assuming a distribution for the noise (e.g. Laplacian or Gaussian, which are related to the l_1-norm and l_2-norm, respectively) and prior distribution (e.g. Laplacian or Gaussian, which are associated with the l_1-norm weight norm penalty and l_2-norm weight penalty). The process of solving a formulated optimisation problem is known as training, see Fig. 1.

To summarise, data science, particularly generative modelling, requires the formulation of an optimisation that can be solved to find informative coordinate systems onto which high-dimensional data can be projected.

3 Supervised, Semi-supervised and Unsupervised Learning

The availability and type of data dictate which learning strategies are available for the analyst. In the context of anomaly or fault detection, supervised learning requires signals that were taken when a physical asset was known to be in a healthy and damaged state. This is referred to as labelled data. In turn, unsupervised learning only requires physical asset data without knowing whether it was taken while the physical asset was healthy or damaged. However, we distinguish between data where the samples are chronologically ordered for unsupervised learning, i.e. measured samples are time-stamped, or not, i.e. each sample is chronological but no time-stamps were recorded for the samples. Lastly, semi-supervised learning relates the learning with predominantly unsupervised data and access to few supervised samples. The most prevalent data in CBM is unsupervised data with chronological samples without prior knowledge of a machine's

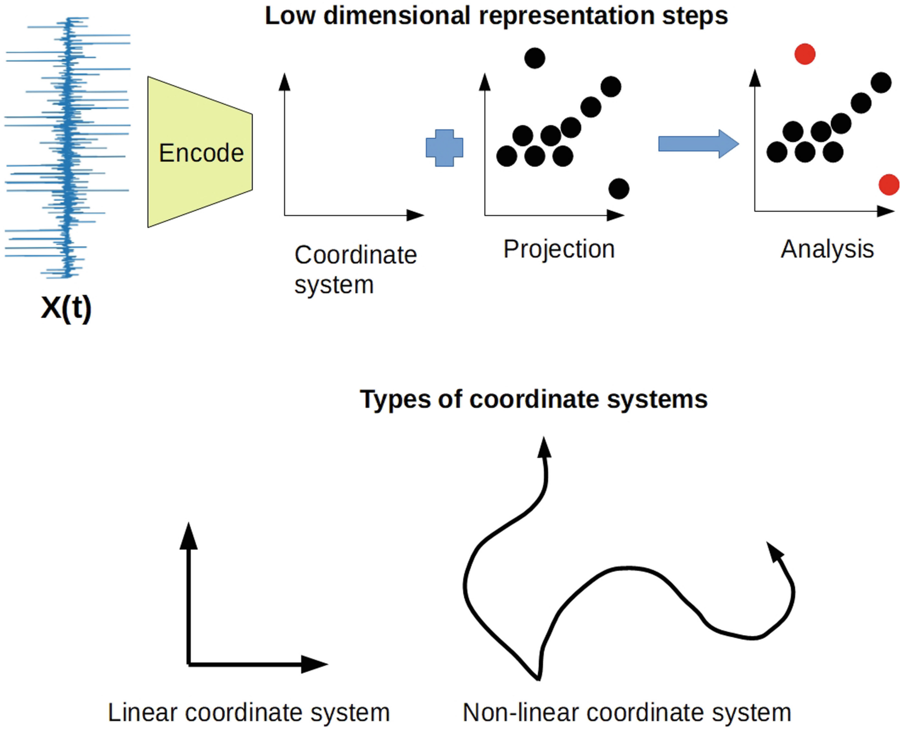


Fig. 2. Encoding a raw signal viewed as finding a coordinate system onto which the data is projected for analysis.

initial state. Data is selected to represent a reference state on which to train a model [13]. This reference state is often referred to as “healthy” when the initial data in the chronological data flow was used to train on [13]. Hence, unsupervised learning with chronologically sorted data is an ideal learning framework to train statistical, machine and deep learning models, which we will explore in more detail.

4 Unsupervised Generative Learning with Chronologically Sorted Data

Unsupervised generative learning can be achieved primarily by one of two learning principles, namely,

1. auto-associative learning [15], or
2. generative adversarial learning [16],

thereby enabling the training of generative models in an unsupervised fashion.

4.1 Auto-associative Learning

Auto-associative learning [15] or autoencoding [17] is pervasive in statistical, machine and deep learning and follows the process of encoding and decoding outlined in Fig. 1. In essence, autoencoding aims to reconstruct the input signal by encoding the signal in a lower-dimensional coordinate system. It reconstructs the signal back to the same dimensionality as the input signal through decoding. This lower-dimensional encoding is often referred to as the bottleneck, an essential mechanism to find an informative lower-dimensional signal. The aim is to learn to regenerate the samples $\mathbf{X}$ without memorizing the samples.

This is exemplified by considering m, l-dimensional samples arranged in matrix $\mathbf{X} \in \mathbb{R}^{m \times l}$. Given that we have centred the data $\mathbf{X}$, which we denote $\mathbf{C}$, we aim to find a coordinate system described by a $l \times k$ matrix $\mathbf{W}$. Here, $\mathbf{CW}$ encodes the data $\mathbf{C}$ into a k-dimensional coordinate system $\mathbf{W}$, whereas $(\mathbf{CWW}^{\mathrm{T}})$, decodes the projected data $\mathbf{CW}$ back to the original l-dimensional space.

Given $k = l$ with $\mathbf{W} = \mathbf{I}$ then encoding results in $\mathbf{CI} = \mathbf{C}$, and decoding gives $\mathbf{CI}^{\mathrm{T}} = \mathbf{C}$ perfect reconstruction. However, using $\mathbf{W} = \mathbf{I}$ is equivalent to memorising the data instead of learning useful information about the data, again highlighting the importance of a lower-dimensional latent space $k \ll l$.

Learning algorithms can be constructed to optimally reconstruct signals. For this case, the aim is to maximise the variance explained in a dataset by finding an optimal lower-dimensional coordinate system to reconstruct signals. Strategies that focus on reconstruction include principal component analysis (PCA), autoencoders (AE) and variational autoencoders (VAE) [15,18,19]. Alternatively, learning algorithms can be constructed to find an informative lower-dimensional coordinate system [20–24], or an informative latent space. Latent focussed approaches include independent component analysis (ICA) and beta-variational autoencoders (β-VAEs) [5,25].

Autoencoding with a Reconstruction Focus: Principal component analysis (PCA) [19] and singular value decomposition (SVD) [26] encapsulate the foundation of autoencoding that aims to maximise the variance explained or aimed at reconstructing the signal as efficiently as possible.

Given $\mathbf{C}$, we aim to find a new orthogonal coordinate system described by a $l \times k$ matrix $\mathbf{W}$. With $k \ll l$ to describe the data, we can achieve it by minimising the following constrained optimisation problem:

$$\mathbf{W}^{*} = \arg \min_{\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}} \left((\mathbf{CW})\mathbf{W}^{\mathrm{T}} - \mathbf{C}\right) : \left((\mathbf{CW})\mathbf{W}^{\mathrm{T}} - \mathbf{C}\right). \quad (1)$$

$\mathbf{CW}$ encodes the data $\mathbf{C}$ into a k-dimensional coordinate system $\mathbf{W}$, whereas $\mathbf{W}^{\mathrm{T}}(\mathbf{CW})\mathbf{W}^{\mathrm{T}}$ decodes the projected data $\mathbf{CW}$ back to the original l-dimensional space. Since $\mathbf{C} \in \mathbb{R}^{m \times l}$, we require a double contraction, denoted: to reduce the expression to a scalar. Autoencoding (AE) [15], Variational Autoencoding (VAE) [27] and Singular Spectrum Analysis (SSA) [28] aim to find a coordinate system that is efficient at explaining the variance in the signal, which in some cases can be informative if the aspect of interest manifests as the variance. It is untangled if it is the sole aspect to do so, however, this is seldom the case.

Autoencoding with a Latent Space Focus: Independent component analysis (ICA) [20,29,30] complements autoencoding that aims to have an informative latent space or aims to find an informative decomposition. ICA terminology often refers to the components of such latent spaces as the sources [31]. The application of ICA in prognostic maintenance of renewable energy systems is well known, but the connection to untangling is not that well established [32].

Machine learning and deep learning offer improved abilities to untangle latent spaces over linear models [33], but remain largely unexplored for CBM applications. Most time-series studies are focussed on human speech and animal acoustics [34].

Given $\mathbf{C}$, we aim to find an informative new coordinate system $l \times k$ matrix $\mathbf{W}$, such that the projected data

$$\mathbf{S} \approx \mathbf{CW} \tag{2}$$

is maximally statistically independent and non-Gaussian, according to some measure of non-Gaussianity. The measures of non-Gaussianity define the characteristics of the expected sources, such as kurtosis and negentropy [20,30,35,36]. Damaged signals often manifest as non-Gaussian sources in the measured data [21,24,37], making ICA useful for fault diagnosis. This is also aligned with the developments of the signal processing community, where different measures of non-Gaussianity are used to identify damaged machine components [21,24,37,38].

Note, the reconstruction of the given data is given by

$$\mathbf{SW}^{-1} \approx \mathbf{C}, \tag{3}$$

but was obtained by finding statistically independent latent variables, which implies that $\mathbf{S}$ is uncorrelated

$$\mathbf{SS}^{\mathrm{T}} = \mathbf{I}. \tag{4}$$

Hence, ICA aims to find a statistically independent latent space during training in contrast to the reconstruction-based focused models. Each latent component represents an isolated physical process or source contributing to the measured signal. This directly implies a latent space that is untangled and informative as the measured signal is decomposed into independent components that can be readily analysed and interpreted.

In machine learning and deep learning, an extension of variational autoencoders (VAE) towards an untangled latent space is β–VAE [25].

4.2 Generative Adversarial Learning

The premise of adversarial training is to transform the unsupervised generative modelling problem into a supervised classification problem. This is achieved by introducing two sub-models in the adversarial learning framework: a generator and a discriminator model. The generator model generates new samples *a priori* from a chosen latent space, while the discriminator classifies between actual

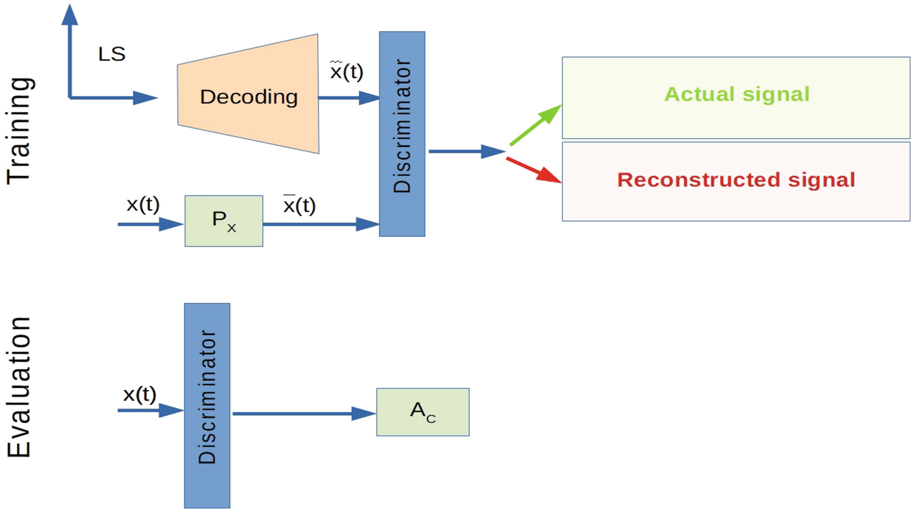


Fig. 3. Adversarial training turns an unsupervised learning problem into supervised learning by introducing a discriminator between actual $\mathbf{x}(t)$ and the reconstructed $\tilde{\mathbf{x}}(t)$ signals. Training constructs a decoding network and discriminator network. Evaluation then uses the discriminator to evaluate whether a signal $\mathbf{x}(t)$ is from the training set data or not by analysing, A_c, the discriminator signal.

signals and generated signals. In an adversarial framework, the two models are trained together.

In a basic adversarial framework, the latent space is chosen to be stochastic without any structure enforced. The implication is that the latent space is uninformative. However, the adversarial framework makes an additional measure available in the form of the discriminator.

A generative adversarial network (GAN) discriminator can be informative in condition monitoring applications [13]. However, an informative latent space would supplement adversarial learning approaches. Ref. [12] extended on the work of Ref. [13] with latent space conditioning semi-supervised learning for CBM applications. Improvements to construct informative latent spaces for GANs include adversarial latent autoencoders [39].

5 Latent Representations

Recall, the latent representations given the same data are not all equivalent. Let us consider concrete manifestations of latent representations using a foundational example. Since faults or environmental and operating conditions can manifest in several variations from a nominal signal, e.g. variations in amplitude, frequency, phase, and offset, we consider a simple sine wave signal with amplitude magnitude variation (between 1 (black) and 5 (white)) as shown in Fig. 4. A low-frequency signal over 10 s is purposefully constructed for clarity.

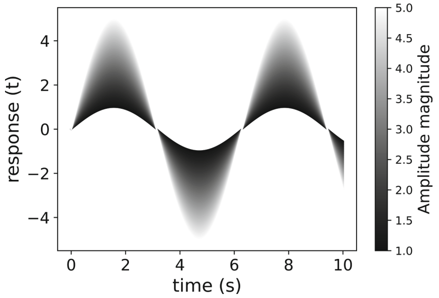
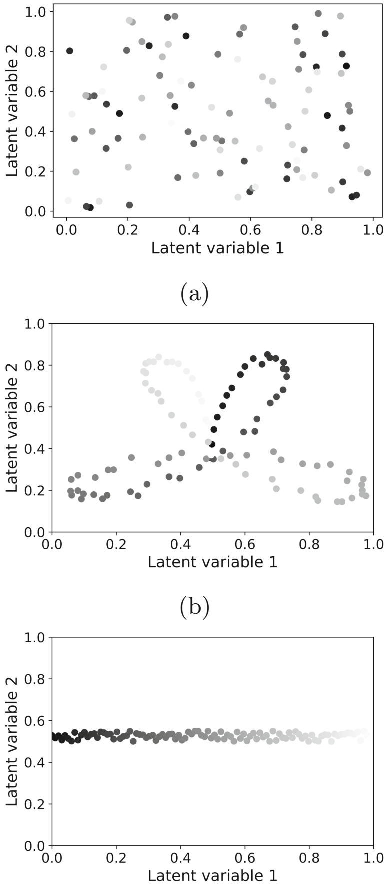


Fig. 4. Foundational example of amplitude magnitude change in a sine wave signal from 1 (black) to 5 (white).



(a)

(b)

(c)

Fig. 5. Latent representations for the amplitude variation from 1 (black) to 5 (white) of the sine wave given as (a) stochastic, (b) entangled and (c) untangled.

Let us focus on finding latent representations highlighting the variation in amplitude magnitude (given by the colour variation of the data points from black (amplitude magnitude of 1) to white (amplitude magnitude of 5)) for the sine wave. We depict a two-dimensional latent space representation for a stochastic latent space in Fig. 5(a), entangled latent space in Fig. 5(b) and untangled latent space in Fig. 5(c).

Given, we obtain a stochastic latent space signal, as the amplitude varies in the sine wave. It is clear that the stochastic latent space signal does not inform the amplitude variation, i.e. the latent space is uninformative. Suppose we would instead obtain an entangled latent space signal, as the amplitude varies in the sine wave. The entangled latent space signal does contain some information of the amplitude variation. In this case, some analysis may be required to interpret the amplitude variation from this entangled latent signal. Lastly, the untangled latent space signal informs the amplitude variation independent of any other variation in the latent signal, resulting in an informative latent signal requiring no additional or minimalistic processing for analysis and interpretation.

Given a condition monitoring problem, where damage is variance or amplitude driven (e.g. modulation due to bearing impacts), we may find similar performance between reconstruction focussed (e.g. SSA, AE or VAE) and latent focussed autoencoding (e.g. ICA or β–VAE) as the variance is a good independence measure for damage in this example. Here, we will obtain an entangled or untangled latent space depending on the additional variations in the signal. However, should damage manifest weakly in the signal's variance, we may find that untangled latent spaces are restricted to latent focussed autoencoding. An untangled latent space is also critical in CBM under the effects of time-varying environmental and operational conditions (EOCs). It allows for informative time-varying components to be identified and isolated [12,13], or uninformative components to be identified and suppressed.

6 Conclusions

This study explored the role of untangled latent spaces obtained with semi-supervised or unsupervised learning. An infinite number of latent spaces exist, of which only a fraction are informative. Significant effort is required to obtain an untangled and informative latent space. However, the additional latent signals available for analysis given an untangled and informative latent space make this endeavour all worthwhile. In some cases, the variance may be a good proxy for statistical independence, given that the fault of interest manifests in the variance of a sensed signal. In these cases, a reconstruction-focused learning strategy may result in a partially untangled latent space that is informative. However, should the fault of interest not manifest in the signal variance or under time-varying operating conditions, then a latent focussed learning strategy is imperative to obtain an untangled and informative latent space.

References

1. Kruczek, P., Zimroz, R., Antoni, J., Wyłomańska, A.: Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech. Syst. Sig. Process. **159**, 107737 (2021)
2. Al-Badour, F., Sunar, M., Cheded, L.: Vibration analysis of rotating machinery using time-frequency analysis and wavelet techniques. Mech. Syst. Sig. Process. **25**(6), 2083–2101 (2011)
3. Cadzow, J.A., Baseghi, B., Hsu, T.: Singular-value decomposition approach to time series modelling. IEE Proc. F Commun. Radar Sig. Process. **130**(3), 202–210 (1983)
4. Bozzo, E., Carniel, R., Fasino, D.: Relationship between Singular Spectrum Analysis and Fourier analysis: theory and application to the monitoring of volcanic activity. Comput. Math. Appl. **60**(3), 812–820 (2010)
5. He, Q., Feng, Z., Kong, F.: Detection of signal transients using independent component analysis and its application in gearbox condition monitoring. Mech. Syst. Sig. Process. **21**(5), 2056–2071 (2007)
6. Bekiroglu, K., Tekeoglu, A., Andriamanalimanana, B., Sengupta, S., Chiang, C.-F., Novillo, J.: Hankel-based unsupervised anomaly detection. In: 2020 American Control Conference (2020)
7. Buzzoni, M., Antoni, J., d'Elia, G.: Blind deconvolution based on cyclostationarity maximization and its application to fault identification. J. Sound Vibr. **432**, 569–601 (2018)
8. Antoni, J., Borghesani, P.: A statistical methodology for the design of condition indicators. Mech. Syst. Sig. Process. **114**, 290–327 (2019)
9. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics, Springer, New York (2001)
10. Aggarwal, C.C.: Outlier Analysis. Springer, New York (2013)
11. Goodfellow, I.J., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016)
12. Baggeröhr, S., Booyse, W., Heyns, P., Wilke, D.: Novel bearing fault detection using generative adversarial networks. In: Condition Monitoring and Diagnostic Engineering Management (2018)
13. Booyse, W., Wilke, D.N., Heyns, S.: Deep digital twins for detection, diagnostics and prognostics. Mech. Syst. Sig. Process. **140**, 106612 (2020)
14. Snyman, J.A., Wilke, D.N.: Practical Mathematical Optimization. Springer, Heidelberg (2018)
15. Kramer, M.A.: Nonlinear principal component analysis using autoassociative neural networks. AIChE J. **37**(2), 233–243 (1991)
16. Goodfellow, I.J., et al.: Generative adversarial nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems, NIPS 2014, vol. 2, pp. 2672–2680, MIT Press, Cambridge (2014)
17. Makhzani, A., Frey, B.: k-sparse autoencoders. In: International Conference on Learning Representations (ICLR) (2014)
18. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3/4), 321–377 (1936)
19. Golyandina, N.: Particularities and commonalities of singular spectrum analysis as a method of time series analysis and signal processing (2020)
20. Hyvärinen, A., Oja, E.: Independent component analysis: algorithms and applications. Neural Netw. **13**(4–5), 411–430 (2000)

21. Hou, S., Wentzell, P.D.: Fast and simple methods for the optimization of kurtosis used as a projection pursuit index. Anal. Chim. Acta **704**(1–2), 1–15 (2011)
22. He, Q., Du, R., Kong, F.: Phase space feature based on independent component analysis for machine health diagnosis. J. Vibr. Acoust. Trans. ASME **134**(2), 1–11 (2012)
23. Debals, O., Lathauwer, L.D.: Stochastic and deterministic tensorization for blind signal separation. In: Lecture Notes in Computer Science, vol. 9237, no. 1, pp. 3–13 (2015)
24. Qian, Y., Yan, R.: Gearbox fault diagnosis in a wind turbine using single sensor based blind source separation. J. Sens. **2016**, 1–14 (2016)
25. Burgess, C.P., Higgins, I., Pal, A., Matthey, L., Watters, N., Desjardins, G., Lerchner, A.: Understanding disentangling in β-VAE. arXiv (2018)
26. Jiang, Y., Tang, B., Qin, Y., Liu, W.: Feature extraction method of wind turbine based on adaptive Morlet wavelet and SVD. Renew. Energy **36**(8), 2146–2153 (2011)
27. Kingma, D.P., Welling, M.: An introduction to variational autoencoders. Found. Trends® Mach. Learn. **12**(4), 307–392 (2019)
28. Sulandari, W., Subanar, Lee, M.H., Rodrigues, P.C.: Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. Energy **190**, 116408 (2020)
29. Herault, J., Jutten, C., Ans, B.: Detection de grandeurs primitives dans un message composite par une architeture de calcul neuromimetique en apprentissage non supervise. In: Colloque sur le traitement du signal et des images, pp. 1017–1022 (1985)
30. Herault, J., Jutten, C.: Space or time adaptive signal processing by neural network models. In: AIP Conference Proceedings, vol. 151, no. 1, pp. 206–211 (1986)
31. Li, Z., Yan, X., Wang, X., Peng, Z.: Detection of gear cracks in a complex gearbox of wind turbines using supervised bounded component analysis of vibration signals collected from multi-channel sensors. J. Sound Vibr. **371**, 406–433 (2016)
32. Afridi, Y.S., Ahmad, K., Hassan, L.: Artificial intelligence based prognostic maintenance of renewable energy systems: a review of techniques, challenges, and future research directions (2021)
33. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**(8), 1798–1828 (2013)
34. Sainburg, T., Thielk, M., Gentner, T.Q.: Latent space visualization, characterization, and generation of diverse vocal communication signals. bioRxiv (2019)
35. Hyvrinen, A., Oja, E.: Independent component analysis: a tutorial. Technical report, TUH (1999)
36. Hyvärinen, A.: Independent component analysis: recent advances. Philos. Trans. Roy. Soc. A Mathe. Phys. Eng. Sci. **371**(1984), 20110534 (2013)
37. Cheng, W., Jia, Z., Chen, X., Gao, L.: Convolutive blind source separation in frequency domain with kurtosis maximization by modified conjugate gradient. Mech. Syst. Sig. Process. **134**, 106331 (2019)
38. Peña, D., Prieto, F.J., Viladomat, J.: Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure. J. Multivar. Anal. **101**(9), 1995–2007 (2010)
39. Pidhorskyi, S., Adjeroh, D.A., Doretto, G.: Adversarial latent autoencoders. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) (2020)

Modal Analysis of the Differential Bevel Gear with Uncertainties

Wassim Lafi[1](✉), Fathi Djmal[1], Ali Akrout[1,2], Lassad Walha[1],
and Mohamed Haddar[1]

[1] Mechanics, Modeling and Production Laboratory (LA2MP),
National Engineering School of Sfax, University of Sfax, Sfax, Tunisia
`mohamed.haddar@enis.rnu.tn`
[2] National Engineering School of Tunis, University of Tunis El Manar, Tunis, Tunisia

Abstract. The application of a differential mechanism is widespread in automotive applications. Its importance stems from the fact that a vehicle cannot take a curve safely without the presence of the differential. Deterministic dynamic analysis can be misleading due to the presence of uncertainties in the system. Such study can lead to catastrophic events since the natural frequencies are sensitive to errors in measurement, assembly lines or parameter modellings. In this case, resonance can occur, which may jeopardize the main purpose of using the system. Thus, it is fundamental to take into account the uncertainties in the system. In this chapter, the modal analysis of the differential system is investigated with the presence of uncertain parameters. The polynomial chaos expansion is used to represent the uncertain parameters. The inverse power method (IPM) is utilized to solve the stochastic eigenvalue problem. Besides, the effects of the uncertain parameter on the natural frequencies and the modal properties are scrutinized. A perfect match is found between the results determined by the IPM and that of Monte Carlo method. This study highlights the main dynamic features of the system, and it can provide important insight into the design and vibration control of the differential system.

Keywords: Automotive differential · Modal analysis · Straight bevel gear · Inverse power method · Natural frequencies

1 Introduction

The presence of a differential mechanism in automotive applications is inevitable due to the fact that it allows a vehicle to take a curve without posing the axle under excessive stress. Its vitality comes at the expense of toughness in analyzing the dynamic properties of the system due to its unique configuration. Such system has two planet gears that possess two angular velocities about two-nonparallel axes. To make the matter worse, some physical parameters of the system cannot be attributed to deterministic values. The prediction of the modal

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 50–64, 2022.
https://doi.org/10.1007/978-3-030-85584-0_6

properties of the mechanical system with deterministic parameters is straightforward. Although this kind of study could help us to determine the unique features of the dynamic behavior of the system, it gives us only partial knowledge of the dynamic behavior in question because we don't have any clue about how the system will behave if the deterministic parameters are assigned to different values. Thus, uncertainties in some parameters should be taken into account in order to determine the dynamic properties of the system that accurately reflect reality.

In contrast to the parallel gear systems [1–7], there is scarcity of the scientific papers that have tackled the dynamic properties of the mechanical system made up by a set of bevel gears.

Driss et al. [8] scrutinized the dynamic properties of the two-stage straight bevel gear system with and without defects. In their studies, the Lagrange-Euler formulation has been used to derive the dynamic equations. Accurate descriptions of the geometric features of the straight bevel gear have been used to describe the variation of the contact point along the spherical involute.

Lafi et al. [9] have analyzed the same system, but the mesh stiffness for the straight bevel gear has been determined by means of the slice theory and potential energy method. Besides, the mesh stiffness function has been assumed as time-varying uncertain parameter. The non-probabilistic interval method has been used to investigate the effects of the uncertain time-varying mesh stiffness in the dynamic properties of the system.

Morselli et al. [10] proposed the reduced and detailed dynamic models to simulate four types of differential. The aim of the detailed model is to simulate and to scrutinize the internal phenomena that can significantly modify the differential dynamics. On the other hand, the purpose of the reduced model is tailored only to highlight the essential features of the differential. The authors used the reduced model to compare the four differentials dynamically.

Shi et al. [11] have conducted design analysis on a differential bevel gear unit under light torque condition and investigated its NVH features. Wang and Liu [12] have scrutinized the differential system to investigate its strength and capability under load. The authors used dynamic software LS-Dyna to investigate gear strength. Safarov et al. [13] have conducted a study on the way of adjusting the satellites and gears. To do so, the authors used the universal method.

Bond graph modeling technique [14–17] has been used many times in scientific articles in order to derive the equations of motion for differential.

Lafi et al. [18] have investigated the dynamic properties of the simple configuration of the differential system. The analytical expression of the mesh phasing relation has been derived. The dynamic equations have been derived by means of the newton-Euler formulation. The assembly error of the planet has been incorporated in the mechanism to investigate its effect on the dynamic response.

The aim of this chapter is twofold. The first is to derive the dynamic modeling of the such system by means of Lagrange-Euler method. The second is to use the inverse stochastic method to scrutinize the stochastic eigenvalue problems and to investigate the natural frequencies of the system with the presence of various uncertain parameters.

2 Dynamic Model of the Differential Mechanism

The differential system studied in this chapter is composed of a set of straight bevel gears. The kinematic representation of the system is depicted in Fig. 1.

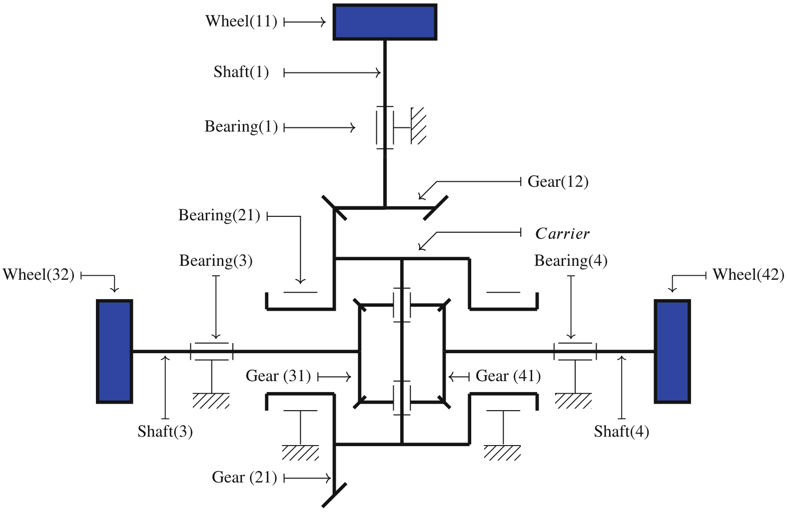


Fig. 1. Kinematic representation of the differential mechanism

To facilitate the comprehension of the mechanism, the system has been dismantled in four blocks. The wheel (11), the gear (12) and the shaft (1) linking them together make up the first block. The second block is composed of the carrier to which two planets are attached and the gear (21). The block (3) and (4) have identical components in order to make sure that even torque must be distributed equally between the two side wheels (denoted by the wheel (32) and the wheel (42)). Each of them consists of a wheel, a gear and a shaft that make them attached.

The lumped parameter of the system is represented in Fig. 2.

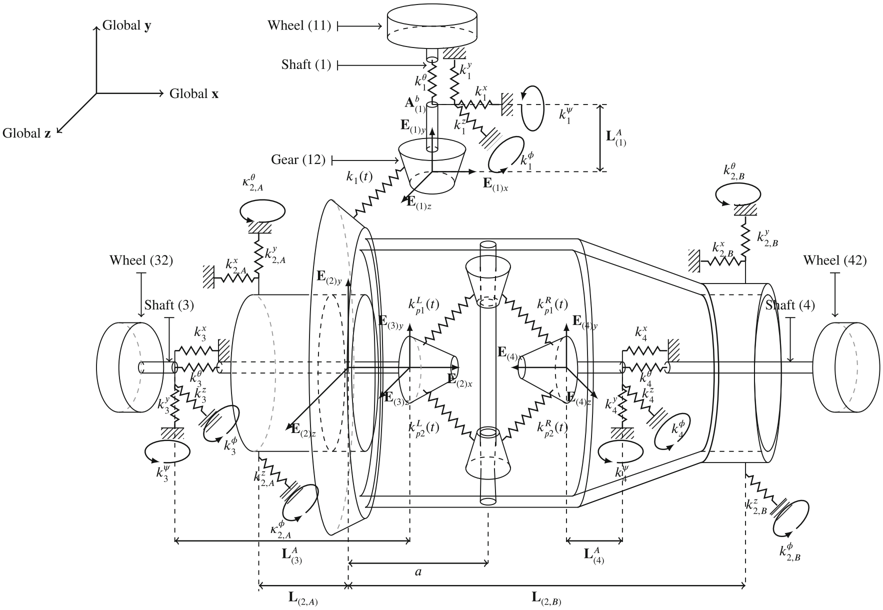


Fig. 2. Lumped parameter of the differential model

Due to the presence of various deformed elements in the system, the system possesses 39 degrees of the freedom. The bearings supporting various blocks in the system have been represented by the three springs placed along three directions. For each gear meshing process in the system, the elastic liaison that represents the contact between the gear teeth is represented by the spring acting along the line of action and perpendicular to the active surfaces of the gear teeth. The spring in question behaves as mesh stiffness that tackles the variation of the number of the gear pair in contact. The mesh stiffness for the straight bevel in this chapter has been calculated by dint of the potential energy method. The proportional damping method is used in this chapter so there is no need to depict the damping symbol in Fig. 2. The trajectory of a vehicle can have significant impact on the dynamic behavior of the system. For the reason of conciseness, the analysis of the system will be conducted for the case in which the vehicle is taking a curve.

3 Dynamic Equations of the System

The Lagrange-Euler formulation has been used to derive the dynamic equations governing the motion of the mechanism. The dynamic equation in matrix from is written as follows:

$$\mathbf{M}\mathbf{q} + \mathbf{C}\mathbf{q} + \left(\mathbf{K}\left(t\right) + \mathbf{K}_s\right)\mathbf{q} = \mathbf{F}_{ext} \tag{1}$$

Where: the generalized coordinates of the differential mechanism are listed in vector $\mathbf{q}$ that can be written:

$$\mathbf{q} = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4] \tag{2}$$

q_j represents the generalized coordinates collecting the degrees of the freedom for block (j).

$\mathbf{M}$ denotes the mass matrix that collects all masses of the components of the system. The time-varying mesh matrix is denoted by $\mathbf{K}(t)$ while the bearing stiffnesses and the torsional stiffnesses of the shafts are listed in the matrix $\mathbf{K}_s$. The damping matrix is denoted by $\mathbf{C}$.

4 Modal Analysis of the Differential System with Uncertainties

Eigenvalue analysis is a fundamental step in the dynamic analysis of a system. Its importance stems from the fact that natural frequencies of the system can be revealed, which can enlighten the engineers about the critical speeds that should be avoided. Presently, the eigenvalues and eigenvectors can be computed easily for deterministic problems. On the other hand, in various cases, physical features are not deterministic. Uncertainties can be incorporated in the mass, geometries, stiffness and other features in the structures. Such uncertainties may induce deleterious issues such as resonance occurrence. That is ascribed to the fact that the impact of the uncertainties is generally unclear. Safety assessment of structures is one of the main reasons to analyze the modal analysis with uncertainty. Once the uncertainties are introduced in the physical structures of the system, it will propagate to the output of the system through its mathematical formulation. The computation of these random results is generally denoted as stochastic analysis. To do so, the analysis in question computes the statistical feature of the stochastic system outputs in the form of their mean values and standard deviations.

4.1 Representation of the Random Variables

In this paper, the bearing stiffnesses, mesh stiffnesses and masses of the components making the system are considered as random variables. Their expressions can be written as:

$$\begin{aligned} k_i(\theta) &= k_{0,i} + \sigma_i^b \theta. \\ k_i(t,\theta) &= k_{mean,i}(t) + \sigma_i(t)\,\theta. \\ m_i(\theta) &= m_{0,i} + \sigma_i^m \theta. \end{aligned} \tag{3}$$

where: $k_{0,i}, k_{mean,i}(t)$ and $m_{0,i}$ stand for the mean value of the uncertain bearing stiffness, the uncertain mesh stiffness and the uncertain mass respectively. $\sigma_i^b, \sigma_i(t)$ and σ_i^m are the standard deviations of the uncertain bearing stiffness, the uncertain mesh stiffness and the uncertain mass respectively. Subscript i indicates the index of the uncertain parameter in the system. The uncertain parameters in Eq. 3 are written in the function of a set of random variables $\theta = \{\theta_i\}_{i=1}^{i=k}$.

4.2 Response Surface by Polynomial Chaos Expansion

In this chapter, the approximation of the stochastic response is achieved by Hermite polynomial function. Therefore, a random response surface $Y\left(\theta\right)$ is generally written in the function of a random variable as can be seen in the following expression:

$$Y\left(\theta\right)=\sum_{i=0}^{P}y_i\psi_i\left(\theta\right). \tag{4}$$

$\psi_i\left(\theta\right)$ stands for ith basis function modeled by Hermite polynomial of the random variables (θ). y_i denotes the deterministic coefficient that must be determined in order to compute the response surface. It is important to remember that hermite polynomials are a set of polynomials that two different polynomials are orthogonal to each other with respect to $W\left(\theta\right)$ as can be seen in the following expression:

$$\int_D \psi_i\left(\theta\right)\psi_j\left(\theta\right)W\left(\theta\right)d\theta=\left\langle\psi_i^2\left(\theta\right)\right\rangle\delta_{ij}. \tag{5}$$

Where D stands for the domain of standard normal random variable $(-\infty<\theta_i<+\infty)$, and δ_{ij} denotes the Kronecker-Delta property.

If Hermite polynomials are utilized as basis functions, the expression of weight function $W\left(\theta\right)$ which stands for the probability density function of the standard normal distribution [19] is written as:

$$W\left(\theta\right)=\frac{1}{\sqrt{\left(2\pi\right)^n}}e^{-\frac{1}{2}\theta^T\theta} \tag{6}$$

The general expression of hermite polynomial is written as:

$$\psi_n\left(\theta\right)=\left(-1\right)^n e^{\left\{\frac{1}{2}\theta^T\theta\right\}}\frac{\partial^n e^{\left\{\frac{1}{2}\theta^T\theta\right\}}}{\partial\theta_1...\partial\theta_n} \tag{7}$$

These orthogonal basis functions must satisfy the following equation:

$$\left\langle\psi_m\left(\theta\right)\right\rangle=\int_D\psi_m\left(\theta\right)W\left(\theta\right)d\theta=0\quad for\ m=1,2,.....,p \tag{8}$$

The expression of the standard deviation and the mean of the stochastic response Eq. 4 can be written as [19]:

$$\mu_Y=E\left[Y\right]=\left\langle Y\right\rangle=y_0 \tag{9}$$

$$\sigma_Y=\sqrt{\mathrm{var}\left[Y\right]}=\sqrt{\left\langle\left(Y-E\left[Y\right]\right)^2\right\rangle}=\sqrt{\sum_{i=1}^{p}y_i^2\left\langle\psi_i^2\left(\theta\right)\right\rangle} \tag{10}$$

After representing the input random variable and showing important parameters used to model the stochastic system, we move on now to present the stochastic eigenvalue problem.

4.3 Stochastic Eigenvalue Problem Formulation

The general eigenvalue problem is given by:

$$\mathbf{Ku} = \lambda\mathbf{Mu} \tag{11}$$

$\mathbf{K}$ stands for a deterministic stiffness matrix, and $\mathbf{M}$ denotes the deterministic mass matrix of the gear system. λ stands for an eigenvalue of the system while u denotes the corresponding eigenvector. In order to simplify the analysis of the system, the previous Eq. 11 can be written as:

$$\mathbf{Au} = \lambda\mathbf{u} \tag{12}$$

Incorporating the uncertainties in the system can alter the expression of the eigenvalue problem. Its new expression is defined as:

$$\bar{\mathbf{A}}\left(\theta\right)\bar{\mathbf{u}}\left(\theta\right) = \bar{\lambda}\left(\theta\right)\bar{\mathbf{u}}\left(\theta\right). \tag{13}$$

The new expressions of the matrices, the eigenvalue and the eigenvectors are written in the function of their deterministic values along with Hermite orthogonal polynomials and random variables.

$$\bar{\mathbf{A}}\left(\theta\right) = \sum_{i=0}^{n_1-1} \mathbf{A}_i \psi_i\left(\theta\right), \tag{14}$$

$$\bar{\mathbf{u}}\left(\theta\right) = \sum_{i=0}^{n_1-1} \mathbf{u}_i \psi_i\left(\theta\right), \tag{15}$$

$$\bar{\lambda}\left(\theta\right) = \sum_{i=0}^{n_1-1} \lambda_i \psi_i\left(\theta\right). \tag{16}$$

Superscript $\bar{()}$ denotes the expression of the uncertain parameter in the function of hermite polynomials and random variables.

The parameter n_1 can be determined as follows:

$$n_1 = \frac{(N+k)!}{k!N!} \tag{17}$$

Where: k stands for the number of random variables while N denotes the order of polynomial chaos expansions.

Plugging the above equations into the generalized stochastic eigenvalue problems, it gives the following expressions:

$$\sum_{i=0}^{n_1-1} \mathbf{A}_i \psi_i\left(\theta\right) \left(\sum_{j=0}^{n_1-1} \mathbf{u}_j \psi_j\left(\theta\right)\right) = \sum_{i=0}^{n_1-1} \lambda_i \psi_i\left(\theta\right) \left(\sum_{j=0}^{n_1-1} \mathbf{u}_j \psi_j\left(\theta\right)\right) \tag{18}$$

Then, $\psi_t\left(\theta\right) W\left(\theta\right) d\theta$ should be multiplied in both sides. After simple simplification, the updated form of the above equation is written as:

$$\sum_{i=0}^{n_1-1} \sum_{j=0}^{n_1-1} \mathbf{A}_i \mathbf{u}_j \left\langle \psi_i\left(\theta\right) \psi_j\left(\theta\right) \psi_t\left(\theta\right)\right\rangle = \sum_{i=0}^{n_1-1} \sum_{j=0}^{n_1-1} \lambda^{'} \mathbf{u}_j \left\langle \psi_i\left(\theta\right) \psi_j\left(\theta\right) \psi_t\left(\theta\right)\right\rangle \tag{19}$$

$$where \qquad t = 0, ..., n_1 - 1.$$

The corresponding eigenvectors should be normalized so that:

$$\mathbf{u}_j^T\left(\theta\right) \mathbf{u}_j\left(\theta\right) = 1. \tag{20}$$

The matrix form of the above equations is written as :

$$\begin{bmatrix} \mathbf{A}_{00}^{''} & \cdots & \mathbf{A}_{0(n_1-1)}^{''} \\ \vdots & \ddots & \vdots \\ \mathbf{A}_{(n_1-1)0}^{''} & \cdots & \mathbf{A}_{(n_1-1)(n_1-1)}^{''} \end{bmatrix} \begin{bmatrix} \mathbf{u}_0^{'} \\ \vdots \\ \mathbf{u}_{(n_1-1)}^{'} \end{bmatrix} = \sum_{i=0}^{(n_1-1)} \lambda_i^{'} \begin{bmatrix} c_{i00}^{''}\mathbf{I} & \cdots & c_{i0(n_1-1)}^{''}\mathbf{I} \\ \vdots & \ddots & \vdots \\ c_{i(n_1-1)0}^{''}\mathbf{I} & \cdots & c_{i(n_1-1)(n_1-1)}^{''}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{u}_0^{'} \\ \vdots \\ \mathbf{u}_{(n_1-1)}^{'} \end{bmatrix}. \tag{21}$$

where:

$$\begin{aligned} \mathbf{A}_{tj}^{''} &= \sum_{i=0}^{n_1-1} \mathbf{A}_i^{'} \left\langle \psi_i\left(\theta\right) \psi_j\left(\theta\right) \psi_t\left(\theta\right)\right\rangle \\ c_{ijt} &= \left\langle \psi_i\left(\theta\right) \psi_j\left(\theta\right) \psi_t\left(\theta\right)\right\rangle \end{aligned} \tag{22}$$

The stochastic eigenvectors should be normalized such as:

$$\bar{\mathbf{u}}_j^T\left(\theta\right) \bar{\mathbf{u}}_j\left(\theta\right) = 1 \tag{23}$$

Same as Eq. 19, multiplying both sides by $\psi_t\left(\theta\right) W\left(\theta\right) d\theta$, and integrating the above equation from $-\infty$ to $+\infty$ yields the following expression.

$$\sum_{i=0}^{n_1-1} \sum_{k=0}^{n_1-1} \mathbf{u}_{j,i}^T \mathbf{u}_{j,k} \left\langle \psi_i\left(\theta\right) \psi_k\left(\theta\right) \psi_t\left(\theta\right)\right\rangle - \left\langle \psi_t\left(\theta\right)\right\rangle = 0 \quad t = 0, 1, 2...... \left(n_1 - 1\right). \tag{24}$$

The above Eqs. 19 and 24 can be solved by means of Newton-Raphson method. However, it poses difficulty in the sense that how to solve it. The difficulty stems from the fact that nonlinear equations are very elusive to solve especially about predicting the initial guess of the nonlinear equations.

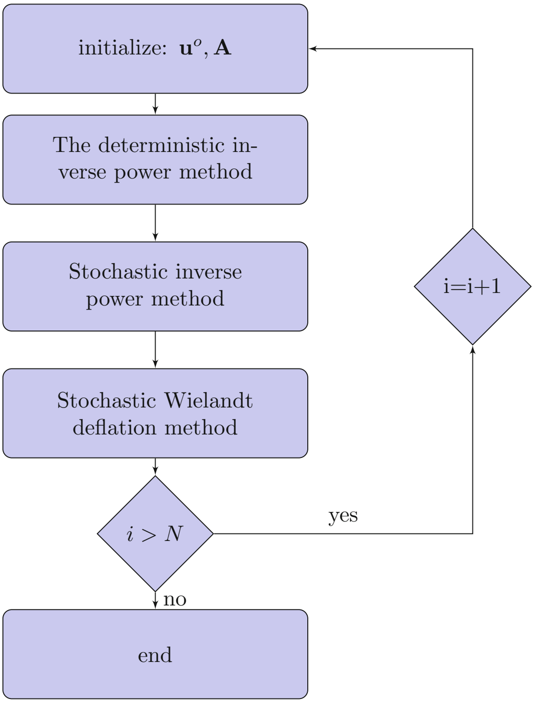


Fig. 3. The algorithm of the stochastic inverse power method and the wielandt deflation method.

Alternatively, the stochastic inverse algorithm along with Wielandt deflation method will be used to solve the nonlinear equations. By means of that algorithm, the initial guess can be bypassed. What follows is the account of the proposed algorithm used to solve stochastic eigenvalue problem.

4.4 Stochastic Method to Solve Stochastic Eigenvalue Problem

As aforementioned, the inverse power method along with Wielandt deflection method can be used in order to determine the stochastic eigenvalues without resorting to predict the initial solution.

Figure 3 shows the diagram of the algorithm to solve stochastic eigenvalue problem.

To do so, two sub-algorithms should be used simultaneously. The first one is the stochastic inverse method. Its role is to determine the minimum random eigenvalue as shown in Fig. 4.

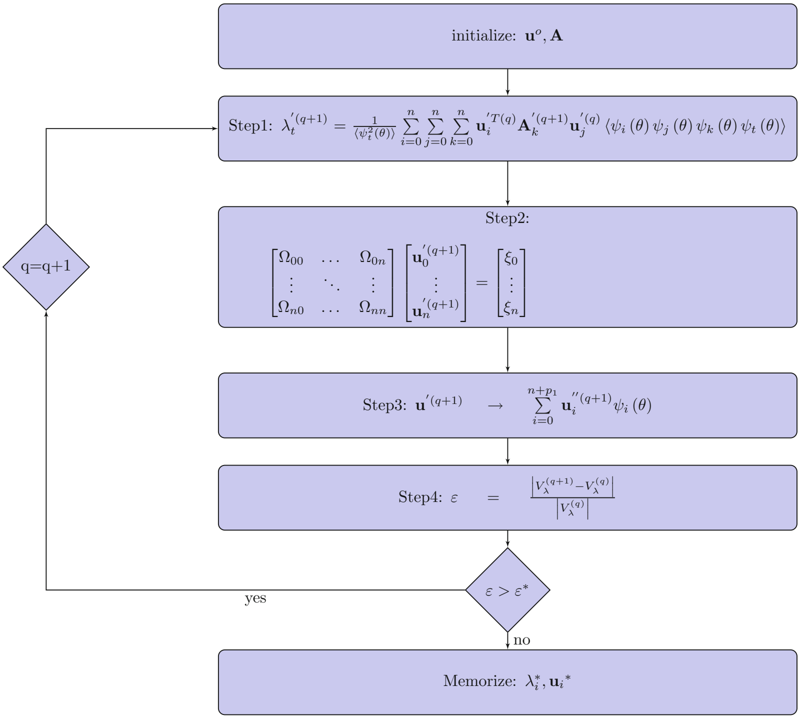


Fig. 4. The algorithm of the stochastic inverse power method.

In order to determine the other values, the dimension of the matrix should be reduced, and the stochastic inverse power should be re-used once again to determine the minimum eigenvalue of the new matrix as shown in Fig. 5.

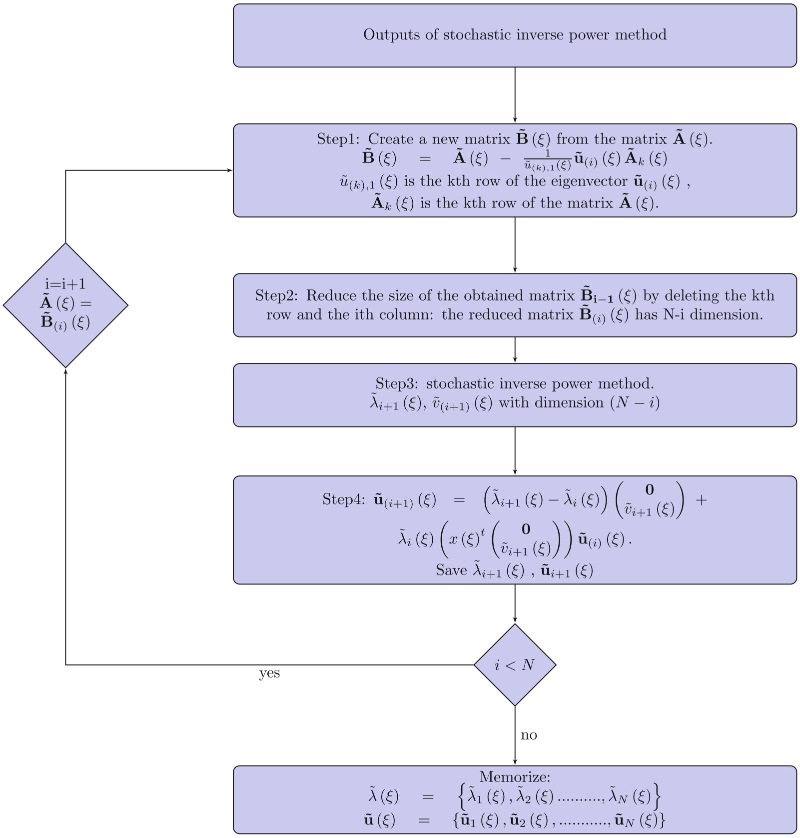


Fig. 5. The algorithm of the stochastic Wielandt deflation method.

The algorithm will work like that until reach the final eigenvalue. The algorithm is thoroughly described in [20].

5 Numerical Results

In this section, the modal properties of the differential bevel gear with uncertainties will be scrutinized. The eigenvalue problem will be solved by means of the stochastic inverse method. It is important to analyze the eigenvalue of the system in hopes that the system will bypass resonance from the occurrence.

Table 1. Essential parameters of the differential mechanism

Parameters	Block (1)	Block (2)	Planet(i)	Side block
Rotational stiffness of the shaft (j) k^{θ} ($\mathrm{N\,m\,rad^{-1}} \times 10^5$)	3	–	–	3
Axial stiffness of bearing (j) k_j^p ($\mathrm{N\,m^{-1}} \times 10^8$) $p \in \{x, y, z\}$	8	–	–	8
Number of the teeth of gear (ji)	$Z_{12} = 42$	$Z_{21} = 52$	$Z_{pi} = 18$	$Z_{31} = 22$
Tooth width (m) $\times 10^{-3}$	$W = 40$	$W = 40$	$W = 20$	$W = 20$
Mass of the block (j) (kg)	$m_1 = 21.072$	$m_2 = 24.8160$	$m_{p1} = 2.8$	$m_3 = 18.4$
Rotational inertia of block (j) ($\mathrm{kgm^2}$)	$I_{12} = 7.2e-2$	$I_{21} = 0.1693$	$I_{p1} = 0.01$	$I_{31} = 0.024$
External torque (N m)	$T_{11} = 100$	–	–	$T_{31} = -61.9048$

Table 1 lists all essential parameters of the system. When the uncertainty is incorporated in the differential system, it will propagate to the outputs of the system. Thus, the eigenvalues of the system will be regarded as random parameters. As mentioned in the previous sections, the uncertainty has been incorporated into the bearing stiffnesses, mesh stiffnesses and the masses of the components making up the system whose expressions are written in Eq. 3.

In this paper, the standard deviation of each uncertain variable is assumed to be equal to 10% of the mean value of the uncertain parameter.

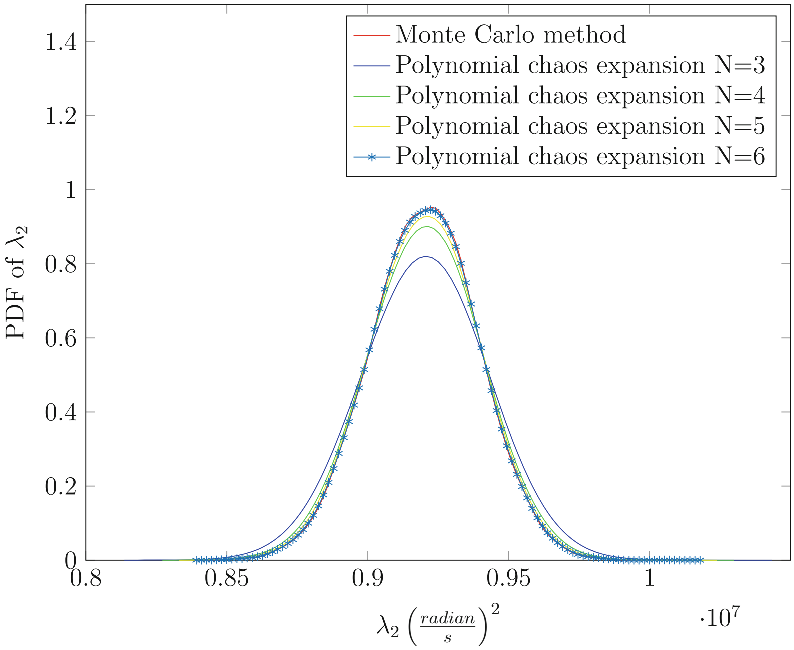


Fig. 6. Comparison between PDF of λ_2 calculated by Polynomial chaos expansion and that of λ_2 calculated by Monte Carlo method

The uncertain outputs are represented by their probability density functions and their mean values. For this reason, the stochastic eigenvalues and their corresponding eigenvectors will be depicted by means of their probability distribution functions. Their mean values will be represented with respect to the variation of the random variable. To this end, two method will be used.

The first is the polynomial chaos method, and the second is the well-known Monte Carlo method.

The latter method is used in this paper as reference method to test the accuracy of the polynomial chaos expansion method used to handle the gear systems with random variables. Monte Carlo Simulation is regarded as a statistical method applied in dynamic modeling of a gear system that cannot be solved easily due to the interference of a random variable. The simulation is based on the repeated random samples to yield numerical results. It can be used to assimilate the influence of uncertainty and randomness in gear models.

Figure 6 presents the probability distribution function of the eigenvalue λ_2 that is determined by two methods.

It is safe to conclude from the data depicted in Fig. 6 that there is a perfect match between the results derived from the proposed method and that of Monte Carlo method.

The perfect match can only occur when the order of polynomial is higher enough. Thus, the higher the order of polynomial is, the more accurate the result is.

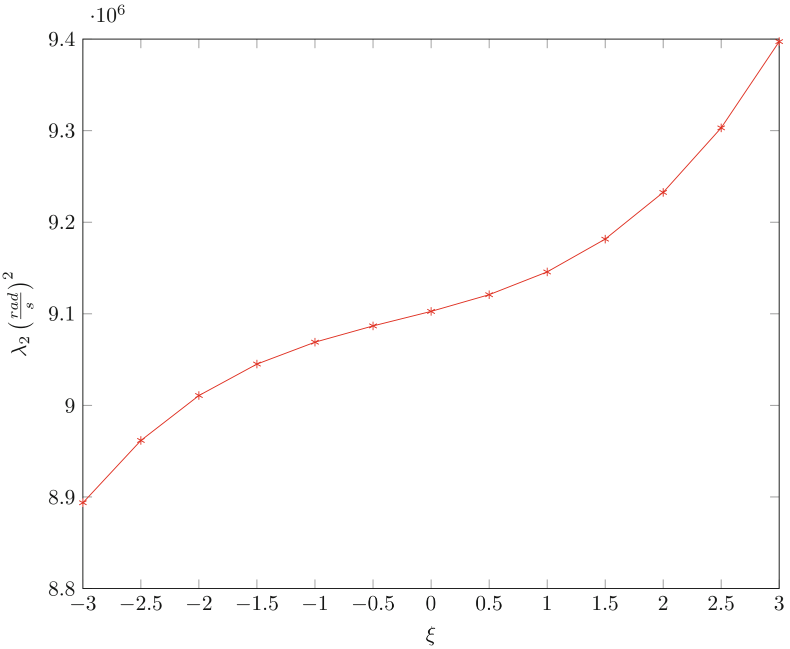


Fig. 7. The variation of the eigenvalue λ_2 with respect to the random variable.

The analytical expression of the stochastic eigenvalue λ_2 can be written as:

$$\lambda_2(\xi) = 9.105e + 06\psi_0(\xi) + 2.06e + 04\psi_1(\xi) + 1.19e + 03\psi_2(\xi) + 7.123e + 02\psi_3(\xi) \tag{25}$$

The variation of the eigenvalue λ_2 with respect to the random variable ξ is represented in Fig. 7. This study can be used to analyze the sensitivity of the eigenvalue λ_2 with respect to the uncertain parameters.

6 Conclusion

The stochastic eigenvalue problem of the differential system has been scrutinized in this chapter. The uncertainty has been incorporated in the mean values of the mesh stiffness, the bearing stiffnesses used in this system and masses of various components making up the system. We have found that the eigenvalue problem has become difficult to be solved by conventional method. Thus, we have proposed the inverse power method to achieve this aim. First of all, we have determined the probability density function of the second eigenvalue of the system. The obtained result has been compared with the result derived by Monte Carlo method and we have found a perfect match between them, which makes the proposed the algorithm adequate to solve stochastic eigenvalue of the gear system. However, the order of polynomial chaos expansion can play significant role in increasing the accuracy of the analysis. We have found out also that the eigenvalue of the system is very sensitive to the random value, which sheds the light on the importance of analyzing the dynamic properties of the system with taking into account the uncertainties.

References

1. Chaari, F.: Contribution a l'etude du Comportement Dynamique Des Trains Epicycloidaux a denture doite en presence de defauts. Ph.D. dissertation, l'Ecole Nationale d'Ingénieurs de Sfax (2005)
2. Guo, Y.: Analytical study on compound planetary gear dynamics. Ph.D. dissertation, The Ohio State University (2011)
3. Lim, T.C., Cheng, Y.: A Theoretical Study of the Effect of Pinion Offset on the Dynamics of Hypoid Geared Rotor System. J. Mech. Des. **121**(4), 594–601 (1999)
4. Sondkar, P., Kahraman, A.: A dynamic model of a double-helical planetary gear set. Mech. Mach. Theory **70**, 157–174 (2013)
5. Kim, W., Lee, J., Chung, J.: Dynamic analysis for a planetary gear with time-varying pressure angles and contact ratios. J. Sound Vibr. **331**(4), 883–901 (2012)
6. Mbarek, A., Del Rincon, A., Hammami, A., Iglesias, M., Chaari, F.: Dynamic behavior of back to back planetary gear in presence of pitting defects. In: Advances in Acoustics and Vibration II, ICAV 2018. Applied Condition Monitoring, vol. 13. Springer, Cham (2018)
7. Yousfi, N., Akrout, A., Zghal, B., Walha, L., Haddar, M.: Rayleigh damping coefficients identification using the wavelet transform on two stage gear system. In: Advances in Materials, Mechanics and Manufacturing. Springer (2020)

8. Yassine, D., Ahmed, H., Walha, L., Haddar, M.: Effects of gear mesh fluctuation and defaults on the dynamic behavior of two-stage straight bevel system. Mech. Mach. Theory **82**, 71–86 (2014)
9. Lafi, W., Djemal, F., Tounsi, D., Akrout, A., Walha, L., Haddar, M.: Non-probabilistic interval process method for analyzing two-stage straight bevel gear system with uncertain time-varying parameters. Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci. 1–17 (2020)
10. Morselli, R., Zanasi, R., Sandoni, G.: Detailed and reduced dynamic models of passive and active limited-slip car differentials. Math. Comput. Model. Dyn. Syst. **12**(4), 347–362 (2006)
11. Shi, Z., Chen, J., Kolivand, M., Sun, Z., Kopp, G., Peng, Y.: Design optimization of differential bevel gear for NVH improvement. SAE Technical Paper. SAE International (2019)
12. Wang, P., Liu, K.: Analysis and design of differential bevel gear. In: 2018 IEEE International Conference on Advanced Manufacturing (ICAM), pp. 243–245 (2018)
13. Safarov, D.T., Kondrashov, A.G.: Methods of quality control manufacturing gears of the differential satellites. In: Radionov, A.A., Kravchenko, O.A., Guzeev, V.I., Rozhdestvenskiy, Y.V. (eds.) Proceedings of the 5th International Conference on Industrial Engineering (ICIE 2019), pp. 249–259. Springer, Cham (2019)
14. Deur, J., Ivanović, V., Hancock, M., Assadian, F.: Modeling and analysis of active differential dynamics. J. Dyn. Syst. Measur. Control **132**(6), 061501 (2010)
15. Hrovat, D., Tobler, W.E.: Bond graph modeling of automotive power trains. J. Franklin Inst. **328**(5–6), 623–662 (1991)
16. Hsu, C.-H.: A graph notation for the kinematic analysis of differential gear trains. J. Franklin Inst. **329**(5), 859–867 (1992)
17. Karnopp, D.: An approach to derivative causality in bond graph models of mechanical systems. J. Franklin Inst. **329**(1), 65–75 (1992)
18. Lafi, W., Djemal, F., Tounsi, D., Akrout, A., Walha, L., Haddar, M.: Dynamic modelling of differential bevel gear system in the presence of a defect. Mech. Mach. Theory **139**, 81–108 (2019)
19. Xiu, D., Karniadakis, G.E.: The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J. Sci. Comput. **24**(2), 619–644 (2002)
20. Lafi W.: Robustness and modal analysis of the differential system with uncertainties. Dissertation, National Engineering school of Tunis (2020)

Effect of Non-linear Suspension on the Recognition of the Road Disturbance

Dorra Ben Hassen[1](✉), Mariem Miladi[1], Mohamed Slim Abbes[1],
S. Caglar Baslamisli[2], Fakher Chaari[1], and Mohamed Haddar[1]

[1] Mechanics, Modeling and Production Laboratory,
National Engineering School of Sfax (ENIS), BP 1173, 3038 Sfax, Tunisia
`Mohamed.haddar@enis.rnu.tn`

[2] Department of Mechanical Engineering, Hacettepe University, 06800 Beytepe, Ankara, Turkey

Abstract. The road excitation is one of the major forces which act on the vehicle and affect the passenger's comfort so it constitutes a crucial field of interest when suspension systems are designed. Hence, identifying this type of excitation remains very benefiting since it contributes to study the dynamic behavior of the vehicle and to apply a controller law in order to ensure the passenger comfort. Direct recognition techniques (longitudinal profile analyser or laser sensors…) of the road profile are costly, thus, it is necessary to find other methods such as numeric ones to recover the road disturbance. In this paper, the inverse problem theory is employed to pick out the road profile disturbance applied to vehicle. This proposed technique, known as the Independent Component Analysis (ICA), can recreate initial excitation sources by using physically measurable signals named observed signals of the system under study. These signals are obtained numerically in this study by using the Newmark approach. Starting from these dynamic responses, the ICA algorithm is applied to a non-linear vehicle model to identify the road excitations. The performance of this technique is studied using some criteria which are the relative error and the MAC number. The obtained results show a good relevance between the original signals and the estimated ones.

Keywords: Non-linear vehicle model · Road excitation · Inverse problem

1 Introduction

The effect of the road excitation on the suspension's performance has been the center of the attention of various scientific papers. One of them is: Hunt (Hunt 1991) has scrutinized the dynamic response of a vehicle that is subjected to the random excitations. Furthermore, E. Duni (E. Duni et al. 2003) have investigated the dynamic behavior of a full vehicle model submitted to different types of road excitations through the use of finite element method. So, in order to analyze the dynamic response of a vehicle under real condition, the road profile excitation should be identified accurately in order to its effect on the ride quality and passenger's comfort (Yan 2012). Kim (Kim et al. 2002) have measured the road roughness directly by means of visual inspections. The Monte

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 65–74, 2022.
https://doi.org/10.1007/978-3-030-85584-0_7

Carlo technique (Harris et al. 2010) has been used to estimate the road profile. Fauriat (Fauriat et al. 2016) proposed the ‘Augmented Kalman filter’ to derive the road profile excitation. Mariem (Mariem Miladi et al. 2019) showed that ICA technique has higher efficiency than other techniques in road identification. The main aim of this paper is to use the ICA to recover the excitation of the road profile for a non-linear full vehicle model. This technique was commonly used to evaluate the excitation force in many studies (Dhief et al. 2016, B. Hassen et al. 2017, Taktak et al. 2012). Its main advantages are that it is simple to be implemented and has a feature of a real time identification process. This paper is organized as follows: the first section shows the two axle vehicle model along with its mathematical formulation. Then the applied method, the ICA, is modeled. In the third section, the obtained results are showed and finally the efficiency of the method is confirmed by means of some performance criteria.

2 Two Axle Vehicle Model

The figure (Fig. 1) depicts the full dynamic model of the car studied in Meywerk (2015).

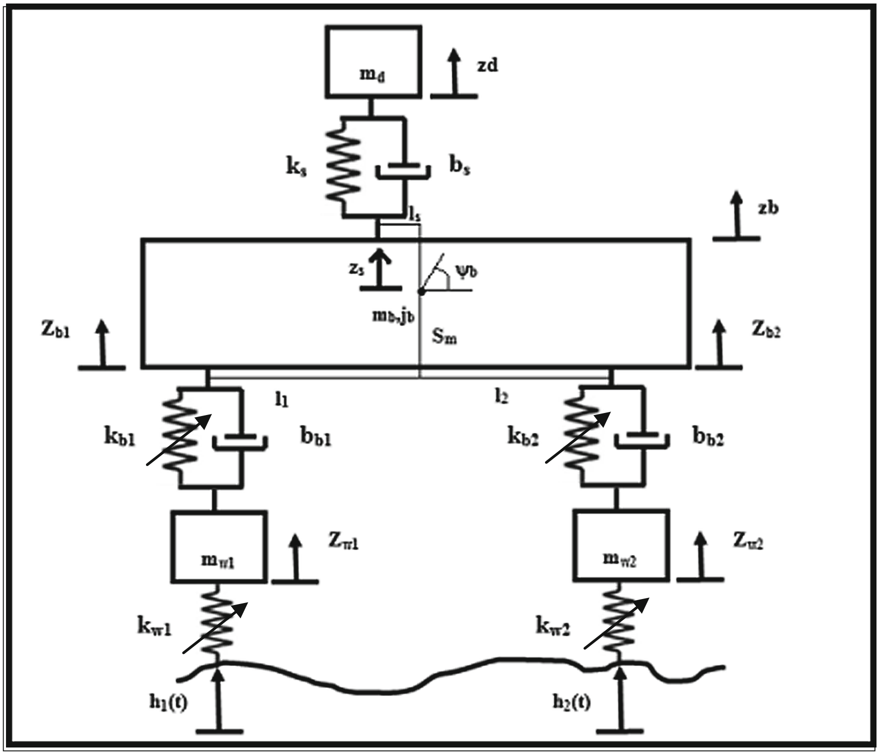


Fig. 1. Two axle vehicle model

This model consists of four masses. In order for this kind of modeling to yield a good insight of the actual vibrations of a real vehicle, the following assumptions should be included:

- The road disturbance is evenly applied to the right and the left wheel. The vehicle is considered to have symmetrical inertia features (Meywerk 2015).
- The excitations applied to the rear wheels $h_2(t)$ are taken identically to those on the front wheels $h_1(t)$, with a short delay as it is mentioned in Fig. 2.

This model is described as follows:

- m_{w1} and m_{w2} stand for the masses of the wheels. They are linked to the road by means of two non-linear springs denoted as k_{w1} and k_{w2}. The deflections of the two masses m_{w1} and m_{w2} are denoted as z_{w1} and z_{w2} respectively.
- z_{b1} and z_{b2} represent the vertical displacement of the suspension systems. These suspensions are composed by non-linear stiffness's k_{b1} and k_{b2} in parallel with the two dampers are denoted b_{b1} and b_{b2}.
- z_b is symbolized as the center of gravity's displacement
- φ_b denotes the pitch angle.
- z_d stands for the vertical displacement of the driver's seat.

Taking into account the assumptions described above, zb_1 and zb_2 are expressed as follows:

$$z_{b1} = z_b - l_1 \varphi_b \tag{1}$$

$$z_{b2} = z_b + l_2 \varphi_b \tag{2}$$

And the coordinate z_s is expressed as:

$$z_s = z_b - l_s \varphi_b \tag{3}$$

- The non-linear Spring's excitations are expressed as follows:

$$F_{b1} = k_{b1}\,\Delta l + \beta_1 k_{b1}\,\Delta l^2 + \beta_2 k_{b1}\,\Delta l^3 \tag{4}$$

and

$$F_{b2} = k_{b2}\,\Delta l_1 + \beta_1 k_{b2}\,\Delta l_1^2 + \beta_2 k_{b2}\,\Delta l_1^3 \tag{5}$$

Where:

- Δl stands for the deflection between z_{b1} and z_{w1} in Eq. (4) and Δl_1 is written in Eq. (5) as the deflection between z_{b2} and z_{w2}.
- $\beta_1, \ \beta_2$ represent two non-linear constants (Li et al. 2011) as $\beta_1 = 0.1 \ \textit{and} \ \beta_2 = 0.4$

For the tire, it is designed as a non-linear spring k_2. The expression of the non-linear tire stiffness is taken from Li et al. (2011) as:

$$F_{w1} = k_{w1}\,\Delta l_3 + \beta_3 k_{w1} \Delta l_3^2 \tag{6}$$

And

$$F_{w2} = k_{w2}\,\Delta l_4 + \beta_3 k_{w2} \Delta l_4^2 \tag{7}$$

Where:

- Δl_3 is the difference between kw_1 and $h_1(t)$ in Eq. (6) and Δl_4 represents the difference between kw_2 and $h_2(t)$ in Eq. (7).
- β_3 is the non-linear coefficient of the tire. Its value is taken from Li et al. (2011):

$$\beta_3 = 0.01 \tag{8}$$

So the equations of motion of the half vehicle model can be written in a matrix form as: (Ref CMSM DORRA°)

$$[\mathrm{M}]\left\{\ddot{\mathrm{X}}\right\} + [\mathrm{C}]\left\{\dot{\mathrm{X}}\right\} + [\mathrm{K}]\{\mathrm{X}\} + \{\mathrm{F_{NL}}\} = \{\mathrm{F}\} \tag{9}$$

Where $\{\mathrm{X}\}$, $\left\{\dot{\mathrm{X}}\right\}$ and $\left\{\ddot{\mathrm{X}}\right\}$ are respectively displacement, velocity and acceleration vectors. [M] is the mass matrix [C] is the damping matrix and [K] depicts the stiffness matrices of the studied system. {F} is the excitation force vector due to road disturbance. $\{\mathrm{F_{NL}}\}$ is the non-linear spring force vector.

$$M = \begin{bmatrix} m_d & 0 & 0 & 0 & 0 \\ 0 & m_b & 0 & 0 & 0 \\ 0 & 0 & J_b & 0 & 0 \\ 0 & 0 & 0 & m_{w1} & 0 \\ 0 & 0 & 0 & 0 & m_{w2} \end{bmatrix} \quad K = \begin{bmatrix} k_s & -k_s & k_s l_s & 0 & 0 \\ -k_s & k_s + k_{b_1} + k_{b_2} & -k_s l_s - k_{b_1} l_1 + k_{b_2} l_2 & -k_{b1} & -k_{b2} \\ k_s l_s & -k_s l_s - k_{b_1} l_1 + k_{b_2} l_2 & k_s l_s^2 + k_{b_1} l_1^2 + k_{b_2} l_2^2 & k_{b1} l_1 & -k_{b2} l_2 \\ 0 & -k_{b_1} & k_{b_1} l_1 & k_{b_1} + k_{w1} & 0 \\ 0 & -k_{b_2} & -k_{b_2} l_2 & 0 & k_{b_2} + k_{w2} \end{bmatrix}$$

$$C = \begin{bmatrix} b_s & -b_s & b_s l_s & 0 & 0 \\ -b_s & b_s + b_{b_1} + b_{b_2} & -b_s l_s - b_{b_1} l_1 + b_{b_2} l_2 & -b_{b1} & -b_{b2} \\ b_s l_s & -b_s l_s - b_{b_1} l_1 + b_{b_2} l_2 & b_s l_s^2 + b_{b_1} l_1^2 + b_{b_2} l_2^2 & b_{b1} l_1 & -b_{b2} l_2 \\ 0 & -b_{b_1} & b_{b_1} l_1 & b_{b_1} & 0 \\ 0 & -b_{b_2} & -b_{b_2} l_2 & 0 & b_{b_2} \end{bmatrix}$$

$$F = \begin{bmatrix} 0 \\ 0 \\ 0 \\ k_{w1} h_1 \\ k_{w2} h_2 \end{bmatrix}$$

The implicit Newmark's technique coupled with Newton Raphson Method has been used to resolve the non-linear dynamic equations governing the motion of the system.

A residue is computed. At each iteration k this residue has the following expression:

$$R_{i+1}^{k} = \left[\overline{K}\right]\{X\}_{i+1}^{k} + F_{NL}^{k} - \left\{\overline{F}\right\}_{i+1} \tag{10}$$

With:

$$\left[\overline{K}\right] = [K] + a_0[M] + a_1[C] \tag{11}$$

And:

$$\begin{aligned}\left\{\overline{F}\right\}_{i+1} = {} & \{F\} + [M]\left(a_0\{X_i\} + a_2\left\{\dot{X}_i\right\} + a_3\left\{\ddot{X}_i\right\}\right) \\ & + [C]\left(a_1\{X_i\} + a_4\left\{\dot{X}_i\right\} + a_5\left\{\ddot{X}_i\right\}\right)\end{aligned} \tag{12}$$

Where a_i ($i = 0..5$) are the Newmark's constants.

If the residue is not acceptable i.e. $R > \varepsilon$, a correction should be made for the displacement vector as following:

$$\{\Delta X\} = \left(\frac{\partial R}{\partial X}\bigg|\begin{matrix}k\\ i+1\end{matrix}\right)^{-1}\left(-R_{i+1}^{k}\right) \tag{13}$$

So that the displacement will be:

$$\{X\}_{i+1}^{k+1} = \{X\}_{i+1}^{k} + \{\Delta X\} \tag{14}$$

The values of the model parameters are presented in the following Table 1:

Table 1. Parameters of the studied vehicle model

Parameters	Variable value	Variable unit
Mass of the chassis	$mb = 960$	[Kg]
Mass of the tires	$m_{w1} = m_{w2} = 36$	[Kg]
Suspension stiffness	$k_{b1} = k_{b2} = 16000$	[N/m]
Tire stiffness	$K_{w1} = k_{w2} = 10^5$	[N/m]
Suspension damping	$bb_1 = bb_2 = 100$	[Ns/m]
Driver's mass	$m_d = 90$	[Kg]
Moment of inertia	$Jb = 500$	[Kg/m^2]
Driver seat's rigidity	$k_s = 2000$	[N/m]
Driver seat's damping	$b_s = 10$	[Ns/m]
l_1	$l_1 = 1.8$	[m]
l_2	$l_2 = 0.8$	[m]

For the road profile excitation, a random road profile has been included in the first wheel. For the second the same excitation with a short delay is taken (Fig. 2). This profile is modeled according to ISO 8608 (ISO 8608) (Table 2) which classifies the profiles to different classes based on the power spectral density (PSD) (Yan 2012).

Table 2. Road profile classification

Road Class	Degree of roughness Gd(n_0) (10^{-6} m^3)		
	Lower limit	Geometric mean	Upper limit
Road A	–	16	32
Road B	32	64	128
Road C	128	256	512
Road D	512	1024	2048
Road E	2048	4096	8192

The Integral White Noise method is employed to create the road roughness. It considers that the road roughness is the issue of a filtered white noise defined by Eq. 15:

$$\dot{q}(t) = 2\pi n_0\, w_1(t)\sqrt{G_d(n_0)v} \tag{15}$$

Where: $w_1(t)$ stands for the Gaussian white noise with a variance equal to 1, q(t) denotes the road roughness while v represents the vehicle velocity. The applied road in this paper is a profile of type A as mentioned in Fig. 2.

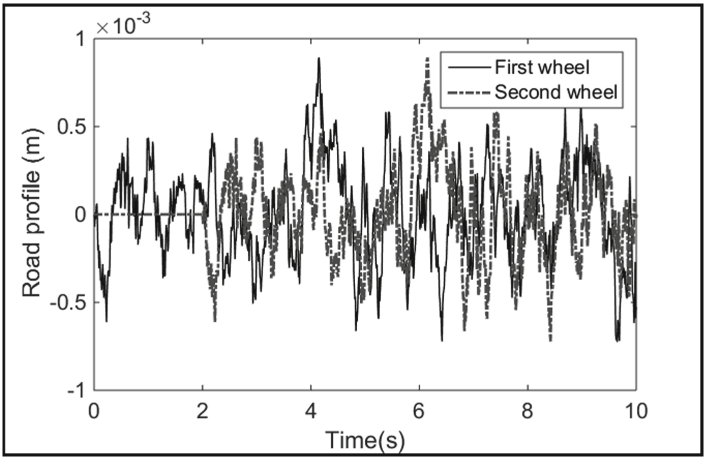


Fig. 2. Road profile

3 Identification Technique: ICA

This algorithm aims to decompose a random signal X in statistically independent components (Dhief et al. 2016). This random signal is expressed by the following equation (Hassen et al. 2017)

$$\mathrm{X(t)} \ = [\mathrm{A}]\{\mathrm{S}\} \tag{16}$$

Where:

A: The mixing matrix.

S: The source signals' vector.

The ICA has to estimate A and S based only on the cognition of the vector X. This estimation requires some assumptions and some pretreatments.

So a matrix W must be searched where the estimated signal $\bar{\mathrm{y}}(\mathrm{t})$ are defined by the following equation

$$\bar{\mathrm{y}}(\mathrm{t}) \ = [\mathrm{W}]\{\mathrm{S}\} \tag{17}$$

To find the matrix W the independence criterion are used in the sense of the maximization of non gaussianity defined by the kurtosis (Eq. 18)

$$\{\mathrm{Y}\} = [\mathrm{W}]^{\mathrm{H}}\{\mathrm{X}\} \tag{18}$$

Where $(.)^{\mathrm{H}}$ denotes the conjugate-transpose operator To find the matrix W the independence criterion are used in the sense of the maximization of non gaussianity defined by the kurtosis (Eq. 19) defined by Zarzoso and Comen (ISO 8608) article apac6344 as the normalized fourth-order marginal cumulate defined by the following equation in order to guarantee a non-Gaussianity distribution.

$$\mathrm{K(ka)} = \frac{\mathrm{E}\left\{|\mathrm{y}|^4\right\} - 2\mathrm{E}^2\left\{\left|\mathrm{y}^2\right|\right\} - \left|\mathrm{E}\left\{\mathrm{y}^2\right\}\right|^2}{\mathrm{E}^2\left\{\left|\mathrm{y}^2\right|\right\}} \tag{19}$$

Where E is is the orthogonal matrix of eigenvectors of $\mathrm{E}\left\{\mathrm{X}\,\mathrm{X}^{\mathrm{T}}\right\}$.

For more details about this method, the reader can refer to these references (Hassen et al. 2017; Comon 1994).

4 Numerical Results

The dynamic responses which are the deflections of the two suspensions system (Fig. 3) are used as the observed signals for the ICA algorithm. Based on the knowledge of these signals (Noted X1 which is equal to $\mathrm{z_b - z_{w1}}$ and X2 which is equal to $\mathrm{z_b - z_{w2}}$), the ICA aims to identify the road profile excitations.

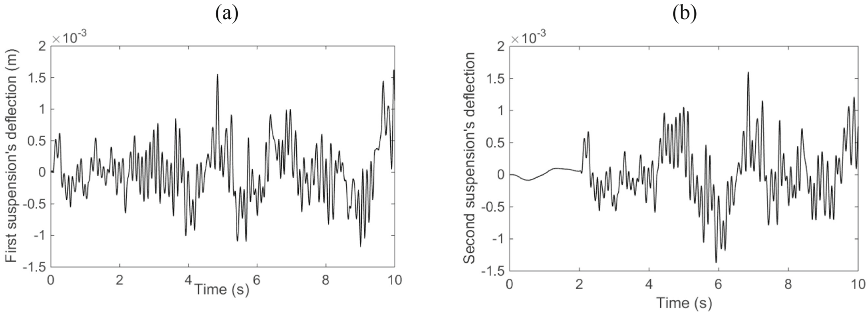


Fig. 3. Observed signals (a) X1 (b) X2

After applying the ICA, The results of the road profile estimation are presented by the following Fig. 4:

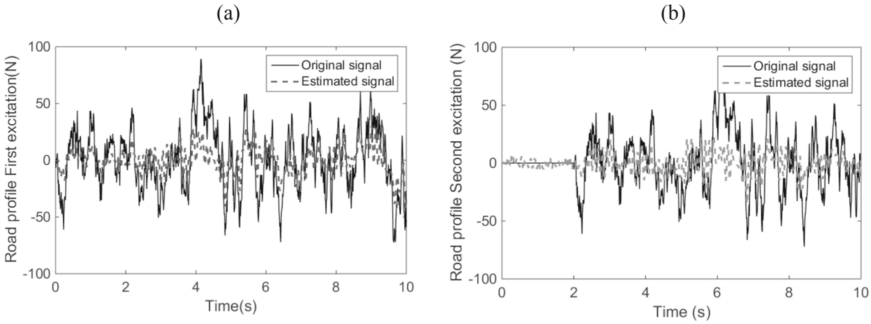


Fig. 4. Estimation of the road profile excitation (a) excitation1 (b) excitation2

Based on these results, it can be said that the ICA can identify the original signals. The small delay and perturbation are due to the effect of the non-linearity. To study the efficiency of this method, the Mac number and the relative error are computed according to the following equations. If the MAC number has a value close to zero, then the obtained results are not compliant and if it has a value close to 1, then the results are compliant. Table 3 resumes the obtained results:

$$E_r = 100 * \frac{y_i - \overline{y}_i}{y_i} \tag{20}$$

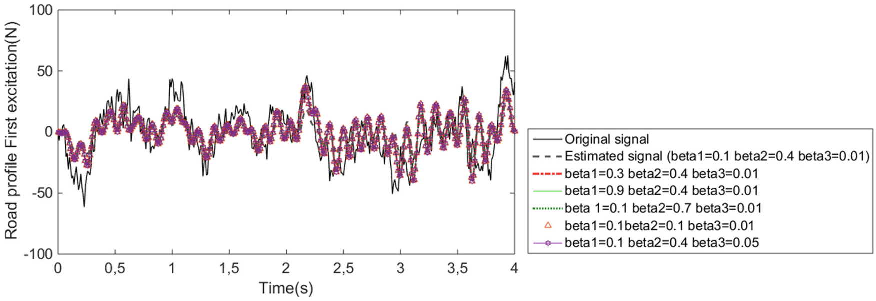
$$\mathrm{MAC_i} = \frac{(y_i^{\mathrm{T}}\overline{y}_i)^2}{(y_i^{\mathrm{T}}\overline{y}_i)(\overline{y}_i^{\mathrm{T}}\overline{y}_i)} \tag{21}$$

Where yi is the original signal and $\overline{y}_i$ is the estimated one.

Table 3. Performance criteria

	MAC	Relative error E_r (%)
Profile 1	0.65	4.712
Profile 2	0.71	3.85

It is seen that the Mac value is near to one for the two studied signals and this shows that the estimated profile is conformed to the original one. Moreover, the values of the relative error confirm these results.



5 Conclusion

The proposed method ICA gives a good estimation of the road profile excitation even with the use of the non-linear parameters. The strength of the ICA technique is that it is applicable with no need to specific road instruments and it is inexpensive.

This method based on the inverse problem, can be used over thousands of kilometers as a real time estimation which is rapid enough.

This identification process will help us to choose the adequate controller law in future work in order to ameliorate the passenger comfort.

References

Hunt, H.E.M.: Stochastic modelling of traffic-induced ground vibration. J. Sound Vib. **144**, 53–70 (1991)

Duni, E. et al.: Numerical simulation of full vehicle dynamic behaviour based on the interaction between ABAQUS/Standard and explicit codes. In: Abaqus Users' Conference, June Munich (2003)

Yan, Z.: Road condition predicting with Kalman. Filter for magneto-rheological. Damper in suspension system, Thesis, Department of Applied Signal Processing, Blekinge Institute of Technology, July 2012

Kim, H.J., Yang, H.S., Park, Y.P.: Improving the vehicle performance with active suspension using road-sensing algorithm. Comput. Struct. **80**, 1569–1577 (2002)

Harris, N.K., González, A., OBrien, E.J., McGetrick, P.: Characterisation of pavement profile heights using accelerometer readings and a combinatorial optimisation technique. J. Sound. Vibr. **329**, 497–508 (2010)

Fauriat, W., Mattrand, C., Gayton, N., Beakou, A., Cembrzynski, T.: Estimation of road profile variability from measured vehicle responses. Vehicle Syst. Dyn. **54**(5), 585–605 (2016)

Chaabane, M.M., et al.: Road profile identification using estimation techniques: comparison between independent component analysis and Kalman filter. J. Theor. Appl. Mech. **57** (2019)

Dhief, R., Taktak, M., Tounsi, D., Akrout, A., Haddar, M.: Application of the independent components analysis in the reconstruction of acoustic sources in duct systems. Arabian J. Sci. Eng., 1–10 (2016)

Hassen, D.B., Miladi, M., Abbes, M.S., Baslamisli, S.C., Chaari, F., Haddar, M.: Application of the operational modal analysis using the independent component analysis for a quarter car vehicle model. In: Fakhfakh, T., et al. (eds.) Advances in Acoustics and Vibration. Applied Condition Monitoring, vol. 5, pp. 125–133. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-41459-1_12

Taktak, M., Tounsi, D., Akrout, A., Abbès, M.S., Haddar, M.: One stage spur gear transmission crankcase diagnosis.using the independent components method. Int. J. Vehicle Noise Vibr. **8**(4), 387–400 (2012)

Meywerk, M.: Vehicle Dynamics. John Wiley & Sons (2015)

Li, S., Lu, Y., Li, H.: Effects of .parameters on dynamics of a nonlinear vehicle-road coupled system. JCP **6**(12), 2656–2661 (2011)

ISO, 8608.: Mechanical vibration. Road surface profiles. Reporting of measured data. ISO 8608:1995 (1995)

Hassen, D.B., et al.: Road profile estimation using the dynamic responses of the full vehicle model. Appl. Acoustics (2017)

Comon, P.: Independent component analysis, a new concept? Signal Process. **36**, 287–314 (1994)

Learning-Based Methods for Vibration-Based Condition Monitoring

Ryan Balshaw(✉), P. Stephan Heyns, Daniel N. Wilke, and Stephan Schmidt

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
`https://www.up.ac.za/centre-for-asset-integrity-management`

Abstract. In this chapter, the fundamentals of learning-based techniques are introduced in a condition monitoring context. The objective is to provide a general overview of learning-based techniques and detail how these techniques are used in vibration-based condition monitoring. This chapter introduces learning-based techniques and introduces the notable distinctions between the supervised and unsupervised learning frameworks. The distinctions between the learning-based techniques are highlighted as the implications of these distinctions are crucial for industrial applicability. Specifically, the assumption of a learning-based methodology of access to a labelled historical fault dataset is often infeasible in industrial applications. Popular methodologies such as supervised learning, semi-supervised learning and transfer learning all require access to a labelled fault dataset. Unsupervised learning techniques offer the opportunity to circumvent the labelled fault data requirements and provide useful condition monitoring insights for condition inference. Under suitable data evaluation strategies, where the preservation of the temporal structure present in time-series data is key, unsupervised learning offers health indicators that can be investigated for fault detection and isolation purposes. Finally, the complementary nature of signal processing and learning-based approaches in an unsupervised learning setting is discussed.

Keywords: Supervised learning · Semi-supervised learning · Unsupervised learning · Latent variable models · Learning-based condition monitoring · Temporal latent variable analysis

1 Introduction to Vibration-Based Condition Monitoring

Condition monitoring (CM) is a research field dedicated to improving and maximising industrial asset productivity. The generalised CM research objective is to develop techniques that can accurately infer a physical asset's state to optimise maintenance decision-making tasks [1]. To initialise this inference process, data is required that captures aspects of the asset's state. This data can be obtained through various sensors, of which vibration data is the most prevalent [2]. In vibration data, faults typically manifest in the signal covariates, and CM

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 75–86, 2022.
https://doi.org/10.1007/978-3-030-85584-0_8

techniques aim to use these covariates for condition inference tasks. However, monitoring these covariates in complex applications, such as wind turbine gearboxes, is non-trivial [2]. In this chapter, focus is given to vibration-based CM techniques. This chapter provides an overview of the current learning-based literature and gives a general overview of how learning-based techniques are applied in vibration-based CM.

There are two main research categories in learning-based CM: *supervised learning* and *unsupervised learning*. There are two general classes of CM problems that are typically addressed: categorical problems and continuous problems. The main difference between the research categories is the assumption of access to a set of labels $\mathbf{y}$, which may be fault categories or continuous variables such as the remaining useful life (RUL) in CM applications. However, as Schmidt and Heyns [3] detailed, there is often a mismatch between the CM problem and how it is addressed. For example, categorical strategies are often used to solve continuous problems. [3] critically discussed the efficacy of *supervised learning* techniques. The assumption that data labels can only originate from the class labels in a historical fault dataset is highlighted as a clear drawback in *supervised learning* tasks. Furthermore, the natural transition between healthy or faulty states cannot be modelled using conventional *supervised learning* techniques. Typical *supervised learning* techniques develop distinct boundaries between each of the historical dataset fault labels. To further develop the CM discussion in this chapter, a basic set of requirements must be detailed for CM techniques, whereby it should be possible to:

- Estimate or infer the damage severity or level of damage.
- Identify the damaged component from the vibration data.
- Estimate the time until the asset is no longer useful.
- Deal with steady and time-varying operating conditions.
- Deal with non-damage related impulsiveness in vibration data due to harsh operating environments.

These requirements are simple enough to isolate but are by no means trivial in nature, as each is an investigation category [4]. To detect damage or estimate the level of damage severity is referred to as the *diagnostics problem*. This approach typically requires a health indicator (HI) to indicate deviance in the state of an asset. To estimate the time until critical asset failure, otherwise referred to as the RUL, is known as the *prognostics problem*. The *diagnostics problem* is often a precursor to the *prognostics problem*, as some estimate of the damage severity is required to infer the RUL [5]. The reliability requirement for operating and environmental conditions (OECs) exists as sensitivity to OECs can significantly hinder the CM technique's practical applicability [6,7].

Signal processing and learning-based techniques are two dominant fields in CM, aiming to address the requirements above [1,8]. Both of these approaches entail *i*) methods, *ii*) training, and *iii*) evaluation, as shown in Fig. 1. However, the two approaches differ in focus, with signal processing techniques predominantly focusing on the methods used and how time-series data is evaluated. The training component in signal processing techniques is defined through expert knowledge obtained through years of experience. Learning-based techniques focus on the methods used and the training procedures required to maximise method performance.

Industrial assets are complex, and extensive knowledge and experience are required to interpret vibration data when using signal processing-based methods. Learning-based methods could reduce some of these requirements and have much potential for vibration-based CM.

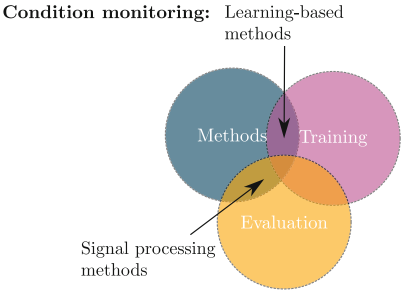


Fig. 1. A basic relationship model for CM methodologies. Signal processing research is predominantly focused on the methods used and the evaluation procedure applied to time-series data. In contrast, learning-based research is predominantly focused on the methods used and the training procedure used.

2 Introduction to Learning-Based Techniques

In this section, an introduction to learning-based techniques is given. Note that the umbrella term "learning-based techniques" includes statistical learning, machine learning and deep learning as sub-concepts of the category [9,10]. It is essential to note the differences between these sub-concepts as there is an underlying philosophical difference. Statistical learning uses a predefined hypothesis and domain knowledge to formulate linear models that facilitate automatic decision making. Machine learning increases the flexibility of the model by introducing non-linear models to maximise model performance. In contrast, deep learning attempts to reduce the requirement for domain knowledge by using powerful parametric models to learn important features automatically. Learning-based techniques, as a type of data-driven technique, are sometimes referred to as digital twins [11]. A digital twin broadly refers to a virtual model that encapsulates and represents a physical system. This virtual model is updated over time in an online setting as more data becomes available from the physical system.

In learning-based applications, the two common approaches of interest are *supervised learning* and *unsupervised learning* [9,12]. These techniques both require access to a set of data $\mathbf{x}$, but how this data is handled differs. *Supervised learning* is a framework for predicting a target label $\mathbf{y}$ for a given sample $\mathbf{x}$ from the data distribution $p(\mathbf{x})$ [13]. *Unsupervised learning* is a framework where the common objective is to perform density estimation to capture the data distribution $p(\mathbf{x})$ [9]. From a probabilistic perspective, these approaches are fundamentally different.

2.1 Supervised Learning

In the *supervised learning* framework, the distribution of interest is the conditional distribution $p(\mathbf{y}|\mathbf{x})$. *Supervised learning* exists as two sub-frameworks, a discriminative framework that aims to model $p(\mathbf{y}|\mathbf{x})$ directly, or a generative framework that aims to model $p(\mathbf{y}, \mathbf{x})$, and then evaluate $p(\mathbf{y}|\mathbf{x})$ using Bayes' theorem. Although there are apparent differences between these sub-frameworks, the end goal is still to evaluate $p(\mathbf{y}|\mathbf{x})$. The target variable $\mathbf{y}$ is either a discrete or continuous variable, with the variable type dictated by the application to categorical or continuous problems, respectively.

To use the discriminative *supervised learning* framework in practice, it is common to assume that a parametric discriminative function $f_\theta(\mathbf{x})$, with parameters θ, parametrises the conditional distribution $p(\mathbf{y}|\mathbf{x})$. This requires one to make explicit choices for the form of this distribution [9]. This function is then optimised using a historical dataset $\mathbb{D} = \{\mathbf{X}, \mathbf{Y}\}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]^T$, $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n]^T$ and N is the dataset size. This optimisation occurs by maximising the conditional likelihood

$$\tilde{\theta} = \arg\max_\theta p(\mathbf{Y}|\mathbf{X}, \theta) \overset{\text{i.i.d}}{=} \arg\max_\theta \prod_{n=1}^{N} p(\mathbf{y}_n|\mathbf{x}_n, \theta), \tag{1}$$

where i.i.d stands for the independently and identically distributed random variables. It is also common to maximise the log-likelihood to avoid numerical underflow problems and simplify gradient computations [9].

2.2 Unsupervised Learning

In the *unsupervised learning* framework, the distribution to be captured is $p(\mathbf{x})$ through a density estimation process. To facilitate this process, latent variable model (LVM) methods are a cornerstone. LVMs utilise a latent variable $\mathbf{z}$ to define the generative distribution $p(\mathbf{x}|\mathbf{z})$, which implies that observed data $\mathbf{x}$ is conditioned on some unobserved latent variable $\mathbf{z}$ [9]. The data distribution is then obtained through the marginalisation of the joint distribution $p(\mathbf{x}, \mathbf{z})$ with respect to $\mathbf{z}$

$$p(\mathbf{x}) = \int_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}) d\mathbf{z} = \int_{\mathbf{z}} p(\mathbf{x}|\mathbf{z}) p(\mathbf{z}) d\mathbf{z}. \tag{2}$$

For sample generation, it is common to assume that it is simple to sample from the prior $p(\mathbf{z})$. Once a latent sample $\mathbf{z} \sim p(\mathbf{z})$ is obtained, data samples can be obtained by sampling from the generative distribution $\mathbf{x} \sim p(\mathbf{x}|\mathbf{z})$. Finally, the posterior distribution $p(\mathbf{z}|\mathbf{x})$ can also be evaluated using Bayes' theorem.

Two fundamental learning frameworks commonly employed in LVMs are: *i*) auto-associative learning, where Principal Component Analysis (PCA) [9,14] and Variational Auto-Encoders (VAEs) [15] are standard methods, and *ii*) generative adversarial learning, where Generative Adversarial Networks (GANs) are the defacto standard [16]. The differences in these learning frameworks are illustrated in Fig. 4. As shown in the model training step in Fig. 4, the principle behind auto-associative methods is to use an encoding function f_ϕ to take data $\mathbf{x} \in \mathbb{R}^{data}$ and map it to a latent space $\mathbf{z} \in \mathbb{R}^{latent}$, and then reconstruct this data using a decoding function g_θ. The latent space is a manifold embedded in the original data space, often assumed to be in an under-complete representation ($\mathbb{R}^{latent} < \mathbb{R}^{data}$) that captures the intrinsic features and important information in the observed data [9]. This auto-encoding process is commonly optimised using an L_p norm error objective function, where the L_1 or squared L_2 norm are two typical choices [8,17–19]. It is also common to regularise the encoding function f_ϕ to conform to the prior $p(\mathbf{z})$ [15]. Generative adversarial learning uses two competing players in a non-cooperative game. A discriminative function determines whether observed data is real or fake, and a generative function tries to generate realistic samples to fool the discriminative function [16]. The non-cooperative game allows each player to improve until the discriminative function cannot distinguish between real and generated samples.

3 Learning-Based Techniques in Condition Monitoring

This section discusses and identifies the differences between learning-based frameworks in vibration-based CM. In Fig. 2, the differentiation between *supervised* and *unsupervised learning* is highlighted. To use learning-based approaches in CM, the *supervised learning* framework aims to perform fault classification or RUL estimation [5,18]. In the *unsupervised learning* framework, the aim is to estimate the likelihood of unobserved data from a reference distribution. It is common to assume that this reference distribution captures an asset in a healthy state [11].

3.1 Time-Series Data Processing

To use learning-based techniques in vibration data applications, typically, the first step is data collection. Time-series data, however, is sampled at a sampling rate F_s for a length of time t, and this gives signals of high dimensionality $\mathbb{R}^{t \times F_s}$. Due to computational limitations, it is often infeasible to use an entire signal as an input to a model. Fortunately, several processing techniques exist that alter the form of the data presented to the model to obtain a data representation $\mathbf{x}$ that satisfies computational limitations.

Manual Feature Selection: Expert knowledge can be used to extract features from time-series data using three domains: *i*) the time domain, *ii*) the frequency domain, and *iii*) the time-frequency domain. This includes, but is not

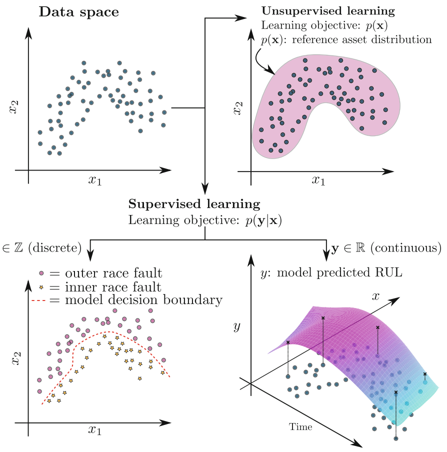


Fig. 2. The relationship between *supervised learning* and *unsupervised learning*, given a set of data $\mathbf{x}$, where the former predicts the corresponding labels $\mathbf{y}$ while the latter does not. Note that the reference asset distribution $p(\mathbf{x})$ is a distribution of the available condition monitoring data for a reference asset state, and the model category decision boundary or RUL can be obtained from both the discriminative and generative *supervised learning* frameworks.

limited to: *i*) mean, median, maximum, minimum, root-mean-square, peak-to-peak ratio, variance, skewness or kurtosis features *ii*) mean frequency, root variance frequency, spectral skewness or spectral kurtosis features, and *iii*) wavelet energy or energy entropy features [8,18]. These scalar features are combined to represent $\mathbf{x}$.

Raw Time-Series Transformations: To negate the requirement for expert knowledge during the processing stage, the raw time-series data can be used directly. Several time-series transformations can be applied to time-series data. These transformations aim to make time-series data accessible to the model by reducing the dimensionality of time-series signals. Typically, time-series data is transformed into a data matrix $\mathbf{H}$ by arranging time-series data into a matrix using a model window length L_w and with a window overlap controlled by a shifting parameter L_{sft} [11,20,21]. This process is given as

$$\mathbf{H}[\mathbf{x}_i(n)] = \begin{bmatrix} \mathbf{x}_i(0) & \mathbf{x}_i(1) & \mathbf{x}_i(2) & \cdots & \mathbf{x}_i(L_w) \\ \mathbf{x}_i(L_{sft}) & \mathbf{x}_i(L_{sft}+1) & \mathbf{x}_i(L_{sft}+2) & \cdots & \mathbf{x}_i(L_w + L_{sft}) \\ \vdots & & \ddots & & \vdots \end{bmatrix}, \quad (3)$$

where $\mathbf{H} \in \mathbb{R}^{N_{obs} \times L_w}$ and N_{obs} is the number of windows extracted from a single signal. The rows of $\mathbf{H}$ are then used as model input vectors.

Alternative Transformations: The data matrix transformation step is often implicit in some alternative transformations. Suppose the Fourier transform is used on each signal segment in the rows of $\mathbf{H}$, where it is common to use the magnitude of the complex Fourier coefficients. In that case, it can be decided to either use the local or global Fourier coefficient magnitude content as input to the model. For the former case, the model input data is the signal segment frequency content, while for the latter, the input data is a Spectrogram [8,18]. It is important to note that the local Fourier case produces vector representations, while the global Fourier case, the Spectrogram, produces matrix representations.

Feature Scaling: After the raw time-series data is transformed, it is expected to perform feature scaling to ensure that the learning-based model is not biased to the magnitude of the features [9]. Standard feature scaling techniques include min-max scaling, mean normalisation, standardisation and unit vector scaling [22].

3.2 Approaches in Supervised Learning

In the *supervised learning* framework, fault classification is an important research task as it tackles point one of the requirements mentioned in Sect. 1. Many *supervised learning techniques* exist in the literature to evaluate $p(\mathbf{y}|\mathbf{x})$, such as linear discriminant analysis, k-nearest neighbours, support vector machines or classification neural networks [9,18]. However, these techniques are dependent on how the fault category information is reflected in $\mathbf{x}$. During the initial stages of *supervised learning* research, manual feature extraction practices, as detailed in Sect. 3.1, are prevalent [18,23]. This manual feature extraction process relies on extensive domain knowledge. Naturally, it reduces the information available for the condition inference task, as explicit decisions are made into which features benefit performance [18].

Deep Learning-Based Diagnosis. Although a sufficient starting point, the feature engineering process is subject to issues such as a lack of generic feature importance, a result of CM problem diversity, and requires expert knowledge to discern which features are informative to a given problem. The development of powerful computational resources naturally leads to the *deep learning* era, aiming to utilise powerful network architectures to perform automatic feature extraction. This process, generally, includes a feature extraction component and a classification component, and is commonly referred to as a *semi-supervised learning* framework, as shown in Fig. 3 [12,19,24–27]. The feature extraction component is often fundamentally based on auto-associative methods as a dimensionality reduction technique [18,19]. In this framework, as shown in Fig. 3, the training and evaluation procedure followed is different. The latent representation of the processed time-series data is fed to the discriminative or generative *supervised learning* model to predict the asset's categorical fault state [8,18].

3.3 The Data Labelling Problem

An immediate drawback of the *supervised learning* framework is the requirement for fault labels [3,18]. Industrial assets typically operate for long periods in a

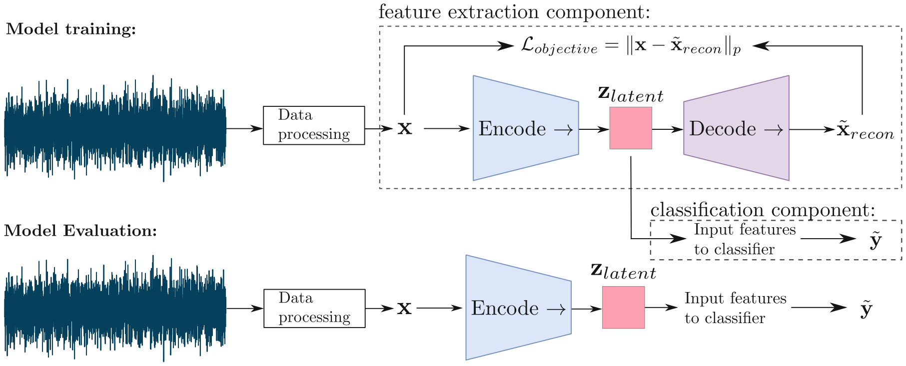


Fig. 3. A generalised *semi-supervised learning* framework with automatic feature extraction. Note that $\tilde{\mathbf{x}}_{recon}$ is the model reconstruction of the input $\mathbf{x}$. There is a clear difference between model training and evaluation. During evaluation, the decoding step is often disregarded, and the predicted labels $\tilde{\mathbf{y}}$ are the focus.

healthy state, and there are often significant financial costs required to obtain fault labels [18]. This is referred to in this work as the *data labelling problem*.

Transfer learning tackles the *data labelling problem* by reusing the diagnosis task knowledge from one asset for another asset [18]. This transfer can occur between identical machines or between different machines. A fundamental assumption to transfer learning techniques is asset distribution overlap. It is common to reduce the asset distribution discrepancy using some shared information space, such that a shared classifier can be used. It is clear that transfer learning still requires access to fault labels.

3.4 Metrics and Applications of Latent Variable Models

However, if the objective is to completely negate the requirement for fault labels, *supervised learning* methods cannot be applied. *Unsupervised learning* methods, such as LVMs, can capture an asset's probabilistic representation in a healthy condition, whereby this distribution can be used to measure any deviations in the asset state. LVMs introduce HIs that can be used to track the departure from the healthy asset state. In Fig. 4, the training and evaluation stages of LVMs are shown. During the model training stage, raw time-series transformations, as detailed in Sect. 3.1, are used to process healthy time-series data into a data matrix with $L_{sft} > 1$. The model parameters are then optimised. During the evaluation stage, HI measures are obtained from the model. The HI measures are then either compared to the healthy HI distribution or evaluated to preserve the temporal characteristics of time-series data.

A common LVM HI is the reconstruction negative log-likelihood, which measures the error between a sample $\mathbf{x}$ and its reconstruction $\tilde{\mathbf{x}}_{recon}$ [28]. If a GAN is used, the discriminator can be used to measure whether the observed data is from the healthy distribution, as shown in the work of Booyse et al. [11]. These metrics focus on the information and covariates in the data $\mathbf{x}$. However, the latent manifold is also a representation of the healthy asset state, and thus it is expected that damage is equivalently detectable in this space. For the latent manifold to be informative, the latent space must be disentangled; otherwise, the presence of damage may be hidden in the latent manifold dynamics [20]. A disentangled manifold allows the model to separate and identify non-damage related changes, such as the operating conditions, from damage, which is crucial for condition inference [6]. Baggeröhr [20] used a latent adversarial framework with a Wasserstein GAN critic [29] to produce a HI on the latent space. In the work of Balshaw [21], a set of model agnostic latent HIs were proposed that showed that damage is directly detectable in the latent manifold.

Non-Preserved Temporal Analysis. In a conventional application of LVMs in CM, it is common to process time-series data into a data matrix with $L_{sft} > 1$. In the model evaluation step, the data used for model training and validation is used to obtain time-independent HI measures and construct a healthy reference HI distribution. The HI measures are time-independent as the segment samples are obtained randomly from the data, and thus, time is not preserved. From this, any new data is processed in the same manner, fed through the model and the new HI measures are evaluated against the reference healthy distribution, as shown in Fig. 4. Akin to hypothesis testing, HIs that lie in low healthy density likelihood regions indicate that the asset has deviated in some way from the reference asset state [11,20]. It is important to note that this deviation cannot always be directly linked to damage. For example, if the asset is to operate in an operating condition cycle that the model has not seen before, it would be expected that the HI could indicate this deviation.

Preserved Temporal Analysis. In the work of Balshaw [21], it was shown that LVMs could be used to isolate and identify the cause of damage. This development arose through the realisation that the model evaluation step is incomplete for applications involving time-series data. It was shown that by preserving the temporal structure in time-series data during model evaluation, by processing data into $\mathbf{H}$ with $L_{sft} = 1$, the temporal characteristics of time-series data are transferred to the HIs to produce discrepancy signals, a term commonly used in discrepancy analysis literature [3,30]. This difference is highlighted in the HI evaluation step in Fig. 4. These HI signals can then be analysed to not only show that faults are detectable using LVMs, but it is also possible for an analyst to isolate the root cause of the fault in localised fault applications using signal processing techniques. This unites signal processing and *unsupervised learning* in a complementary manner, whereby signal processing enriches the LVM driven CM inference procedure. This result highlighted that the evaluation method is carefully considered when using time-series data, thereby bringing all three facets from Fig. 1 to learning-based CM.

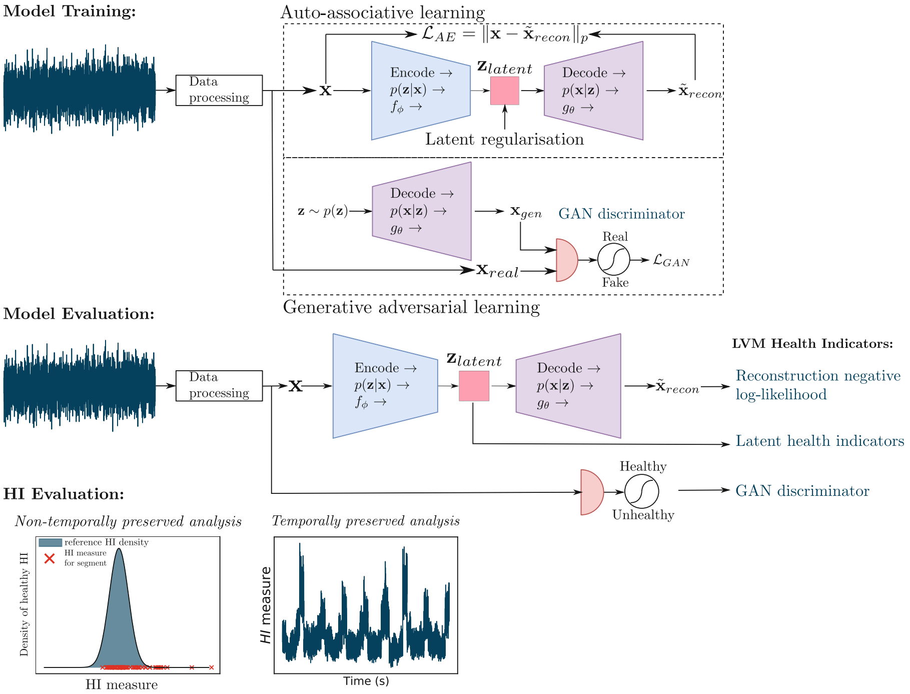


Fig. 4. A general overview of LVM learning's training and evaluation stages for CM applications. Note that auto-associative learning learns both f_ϕ and g_θ while GANs only learn g_θ. Auto-associative methods typically assume that the posterior and generative distributions are represented by Gaussian distributions [9,15]. However, GANs make implicit distribution assumptions, which increases the flexibility of g_θ as it can capture more complex distributions [16].

4 Conclusion

Learning-based CM is an exciting and rapidly evolving field of research for CM. However, a firm grasp on the underlying concepts of the techniques used is a necessity for those who would like to meaningfully contribute to this field. In this chapter, the application of learning-based techniques has been discussed for vibration-based CM. The formulations of *supervised learning* and *unsupervised learning* were discussed from a fundamental perspective to identify the two fields underlying conceptual components for CM applications under the *data labelling problem*.

For learning-based research, the importance of industrial applicability cannot be understated. The dependence on fault labels for *supervised learning* techniques cannot be realistically scaled for the industry. LVMs overcome the *data labelling problem*, detect damage, and can uncover the root cause of a fault in localised fault applications. This is achieved by preserving the temporal structure during the model evaluation stage, which is an essential deviation from conventional analysis methods. LVM-based CM research is an emerging field, and there are still numerous CM requirements that need to be investigated. The research will improve the performance, reliability and interpretability of LVMs for CM, which could enable mainstream adoption in industry.

References

1. Jardine, A.K., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. **20**(7), 1483–1510 (2006)
2. Randall, R.B.: 1. In: Vibration-based Condition Monitoring: Industrial, Aerospace and Automotive Applications, pp. 1–23. Wiley, Chichester (2011)
3. Schmidt, S., Heyns, P.S.: An open set recognition methodology utilising discrepancy analysis for gear diagnostics under varying operating conditions. Mech. Syst. Signal Process. **119**, 1–22 (2019)
4. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D.: Prognostics and health management design for rotary machinery systems - reviews, methodology and applications. Mech. Syst. Signal Process. **42**(1–2), 314–334 (2014)
5. Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J.: Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. **104**, 799–834 (2018)
6. Schmidt, S., Heyns, P.S.: Normalisation of the amplitude modulation caused by time-varying operating conditions for condition monitoring. Meas. J. Int. Meas. Confeder. **149**, 106964 (2020)
7. Mahgoun, H., Chaari, F., Felkaoui, A., Haddar, M.: L-kurtosis and improved complete ensemble EMD in early fault detection under variable load and speed. In: Fakhfakh, T., Karra, C., Bouaziz, S., Chaari, F., Haddar, M. (eds.) Advances in Acoustics and Vibration II, vol. 13, pp. 3–15. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-94616-0_1
8. Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., Gao, R.X.: Deep learning and its applications to machine health monitoring. Mech. Syst. Signal Process. **115**, 213–237 (2019)
9. Bishop, C.: Pattern Recognition and Machine Learning. Springer, New York (2006)
10. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. Springer Series in Statistics. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
11. Booyse, W., Wilke, D.N., Heyns, P.S.: Deep digital twins for detection, diagnostics and prognostics. Mech. Syst. Signal Process. **140**, 106612 (2020)
12. Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning (2016)
13. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
14. Tipping, M.E., Bishop, C.M.: Probabilistic principal component analysis. J. R. Stat. Soc. Ser. B Stat Methodol. **61**(3), 611–622 (1999)
15. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv:1312.6114 (2013)
16. Goodfellow, I.J., et al.: Generative adversarial networks. arXiv:1406.2661 (2014)
17. Khan, S., Yairi, T.: A review on the application of deep learning in system health management. Mech. Syst. Signal Process. **107**, 241–265 (2018)

18. Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K.: Applications of machine learning to machine fault diagnosis: a review and roadmap. Mech. Syst. Signal Process. **138**, 106587 (2020)
19. Shao, H., Jiang, H., Lin, Y., Li, X.: A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders. Mech. Syst. Signal Process. **102**, 278–297 (2018)
20. Baggeröhr, S.: A deep learning approach towards diagnostics of bearings operating under non-stationary conditions. Masters thesis, University of Pretoria (2019)
21. Balshaw, R.: Latent analysis of unsupervised latent variable models in fault diagnostics of rotating machinery under stationary and time-varying operating conditions. Masters thesis, University of Pretoria (2020)
22. Han, J., Pei, J., Kamber, M.: Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science (2011)
23. Liu, R., Yang, B., Zio, E., Chen, X.: Artificial intelligence for fault diagnosis of rotating machinery: a review. Mech. Syst. Signal Process. **108**, 33–47 (2018)
24. Bull, L.A., Worden, K., Dervilis, N.: Towards semi-supervised and probabilistic classification in structural health monitoring. Mech. Syst. Signal Process. **140**, 106653 (2020)
25. Mao, W., Feng, W., Liu, Y., Zhang, D., Liang, X.: A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. Mech. Syst. Signal Process. **150**, 107233 (2021)
26. Wu, X., Zhang, Y., Cheng, C., Peng, Z.: A hybrid classification autoencoder for semi-supervised fault diagnosis in rotating machinery. Mech. Syst. Signal Process. **149**, 107327 (2021)
27. Xu, X., Cao, D., Zhou, Y., Gao, J.: Application of neural network algorithm in fault diagnosis of mechanical intelligence. Mech. Syst. Signal Process. **141**, 106625 (2020)
28. An, J., Cho, S.: Variational autoencoder based anomaly detection using reconstruction probability. Special Lecture IE **2**, 1–18 (2015)
29. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223 (2017)
30. Schmidt, S., Heyns, P.S., Gryllias, K.C.: A discrepancy analysis methodology for rolling element bearing diagnostics under variable speed conditions. Mech. Syst. Signal Process. **116**, 40–61 (2019)

Parametric Study for PCM-Based Heat Sinks: A Numerical Investigation

Bessem Debich[1,2(✉)], Ahmed Yaich[3], Wajih Gafsi[1], Abdelkhalak El Hami[2], Lassaad Walha[1], and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modeling and Manufacturing (LA2MP), Mechanical Engineering Department, National School of Engineers of Sfax, Sfax, Tunisia
{bessem.debich,wajih.gafsi}@enis.tn,
walhalassaad@yahoo.fr,mohamed.haddar@enis.rnu.tn
[2] Laboratory of Mechanics of Normandy (LMN), National Institute of Applied Sciences of Rouen, Rouen, France
{bessem.debich,abdelkhalak.elhami}@insa-rouen.fr
[3] Institut Clément Ader (ICA), CNRS/INSA/ISAE/Mines Albi/UPS, Université de Toulouse, Toulouse, France
yaich@insa-toulouse.fr

Abstract. Studying thermal management of mechatronic components presents a great challenge for researchers in several domains in the last few decades. In most cases, classical cooling systems may not be sufficient. To this end, phase change material-based heat sinks are well recommended for a passive cooling problem. In this paper, we present an efficient thermal energy management of a PCM-based round pin-finned heat sink that leads to provide its optimal configuration. The numerical results compare relatively well with that of the experimental. A detailed parametric study is based on the effect of several uncertain parameters which are in relation with boundary conditions, geometry and materials. This study leads to determine that salt hydrate has the ability to store more thermal energy of the presented cooling system comparing to n-Eicosane and paraffin wax. Furthermore, a heat sink with a 2 mm pin diameter presents the optimal heat sink geometry for the charging phase under a uniform heat flux about 2800 W/m^2. Also, increasing the input heat power leads to decrease the latent heating period and then decreasing the ability to store a high amount of thermal energy. Finally, the parametric study leads to determine an optimal configuration with an efficient thermal energy behavior. Thermal performance of each configuration is determined for the charging phase of the electronic device.

Keywords: Thermal energy storage · PCM-based heat sink · Finite element analysis · Heat flux · Latent heating phase

1 Introduction

Recently, thermal energy storage presents the most important factor for researchers in several domains, particularly in the field of mechatronics such

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 87–95, 2022.
https://doi.org/10.1007/978-3-030-85584-0_9

as smartphones, laptops automobile and even planes. Due to the development of new technologies, devices packages become even smaller. In fact, an overheating of electronic components may lead to reduce their lifetime or even their failure. For this reason, a good thermal energy management is extremely needed to avoid its deterioration.

In this context, classical cooling methods are not sufficient. Therefore, many methods of passive cooling by heat dissipation are developed and some of them are based on Phase Change Materials (PCM) [1–3]. In fact, it has been confirmed that PCMs are more efficient particularly with components that operate intermittently [4]. Recently, PCM-based heat sink is recommended to increase the performance of the cooling system as well as to extend its lifetime. In fact, PCMs are characterized by their high latent heat of fusion, high specific heat and very small variation in volume during the phase change. It can be noted also that PCMs have the ability to store a large quantity of thermal energy benefiting from its thermophysical properties and then to release it back later. An experimental investigation is studied by Arshad et al. [1] to ensure and improve the functionality and reliability of the installed structures. In this study, a parametric analysis is proposed to ameliorate the thermal behavior of PCM-based heat sink by changing the diameter of pin-fin as well as PCM volume fraction at different heat fluxes. In [5,6], a Reliability-Based Design Optimization (RBDO) studies are developed in order to determine an efficient and optimal model of a PCM-based heat sink cooling system where 2D and 3D Finite Element Analysis (FEA) are established respectively.

In this paper, a parametric analysis is proposed to determine an optimal model of PCM-based round pin-fin heat sink. This study is verified and validated experimentally [1]. It consists in comparing the thermal behavior of the proposed PCM-based heat sink for charging phase when changing: input power level, heat sink geometry, PCM thermo-physical properties as well as its volume fraction. Finally, an optimal and efficient configuration is then proposed and it presents an efficient thermal energy management.

2 Studied Cooling System Model: Governing Equations

The thermal energy generated by the heat source is transmitted to all heat sink surfaces. In fact, PCMs have the ability to absorb and store thermal energy during the charging phase until its total melting. For the discharging phase, the stored thermal energy will be dissipated in the ambient air by convection and radiation.

- **Energy conservation equation**

PCM thermophysical properties are supposed independent of temperature. It can be assumed that only conduction equation is considered, as follow:

$$\frac{\partial}{\partial t}\left(\rho C_p T\right) = \frac{\partial}{\partial x}\left(\lambda \frac{\partial T}{\partial x}\right) + \frac{\partial}{\partial y}\left(\lambda \frac{\partial T}{\partial y}\right) + \frac{\partial}{\partial z}\left(\lambda \frac{\partial T}{\partial z}\right) + S_h. \tag{1}$$

Note that C_p, ρ and λ are the specific heat, the density and the thermal conductivity, respectively.

The latent heat storage can be defined by the energy source term S_h due to melting and it can be written as follow:

$$S_h = -\frac{\partial}{\partial t}(\rho \Delta H) \tag{2}$$

The sum of the latent heat ΔH and specific enthalpy h_s presents the total enthalpy of PCM:

$$H = \Delta H + h_s. \tag{3}$$

And the specific enthalpy h_s is mathematically defined by:

$$h_s = h_{s,ref} + \int_{T_{ref}}^{T} C_p dT. \tag{4}$$

In addition, the latent heat ΔH can be defined by:

$$\Delta H = \beta L_f \qquad \text{where} \qquad \beta = \begin{cases} 0 & \text{if } T \leq T_{sol} \\ \frac{T - T_{sol}}{T_{mel} - T_{sol}} & \text{if } T_{sol} < T < T_{mel} \\ 1 & \text{if } T \geq T_{mel} \end{cases} \tag{5}$$

Note that, β and L_f present the liquid fraction and the latent heat of fusion of PCM.

- **Mass conservation**

$$\frac{\partial(\rho u)}{\partial x} + \frac{\partial(\rho v)}{\partial y} + \frac{\partial(\rho w)}{\partial z} = 0 \tag{6}$$

Note that u, v and w are the velocity components in x, y and z directions, respectively.

- **Momentum conservation**

$$\rho\left(\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} + \frac{\partial(uv)}{\partial y} + \frac{\partial(uw)}{\partial z}\right) = -\frac{\partial P}{\partial x} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) + S.u \tag{7}$$

$$\rho\left(\frac{\partial v}{\partial t} + \frac{\partial(uv)}{\partial x} + \frac{\partial v^2}{\partial y} + \frac{\partial(vw)}{\partial z}\right) = -\frac{\partial P}{\partial y} + \mu\left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2}\right) + S.v \tag{8}$$

$$\begin{aligned} \rho\left(\frac{\partial w}{\partial t} + \frac{\partial(uw)}{\partial x} + \frac{\partial(vw)v^2}{\partial y} + \frac{\partial w^2}{\partial z}\right) = -\frac{\partial P}{\partial z} + \mu\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}\right) \\ + S.w + \rho\, g\alpha_w\,(T - T_s) \end{aligned} \tag{9}$$

3 Numerical Results: Physical Description of the Problem

A detailed 3D design is presented in Fig. 1 of the studied PCM-based round pin-fin heat sink. It is composed by 11 × 12 pins heat sink matrix with an overall dimension of 114 × 114 × 25 mm^3. It is made from aluminum which is characterized by its high thermal conductivity and its lower density.

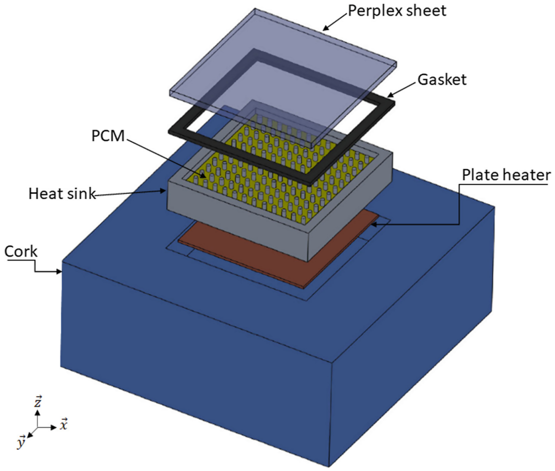


Fig. 1. PCM-based heat sink model: a 3-D assembly.

To minimize thermal losses, an insulator is applied to exterior heat sink walls, except the superior area. The top surface of the heat sink is covered using perspex sheet and silicon gasket for monitoring control PCM phase changes during the simulation. Note that these parts are not taken into account in the numerical model. A three-dimensional numerical model is considered to compare trends detected in the experimental study [1]. A constant heat flux is applied from the heat source to the heat sink bottom.

Table 1. Thermo-physical properties of each material

Material	Th. conductivity (W/mK)	Specific heat (kJ/kgK)	Latent heat (kJ/kg)	Solidification point (°C)	Melting point (°C)	Density (kg/m^3)
Aluminum	201	0.9	-	-	606.4	2700
Rubber pad	0.043	1.23	-	-	-	2500
Paraffin Wax	0.212 (s)	2.8	173.6	56	58	880 (s)
	0.167 (l)					790 (l)

The thermophysical properties of each material is presented in Table 1. To study the round pin-fin heat sink behavior fulfilled with PCM, a Finite Element (FE) computing software ANSYS is proposed. A uniform heat flux ($Q = 2800$ W/m^2) is applied at the bottom of the heat sink.

The temperature-time profile is presented in Fig. 2 for a 3 mm pin-fin heat sink diameter. Note that black and blue curves correspond to numerical and experimental results respectively, throughout the charging phase under $Q = 2800$ W/m^2.

For the charging phase, three zones can be distinguished: solid, liquid and latent heating region.

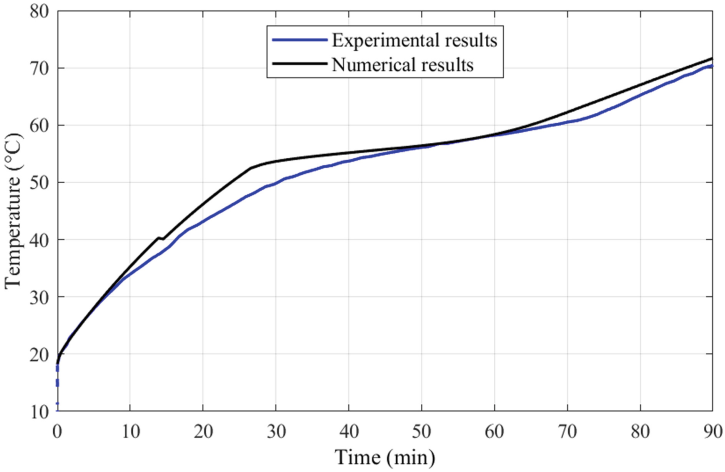


Fig. 2. Temperature-time profile for both experimental and numerical results $Q = 2800$ W/m^2.

It can be seen that numerical results compare relatively well with the experimental data found by Arshad et al. [1]. Therefore, the parametric analysis can be then proceeded.

4 Parametric Analysis

4.1 PCM Selection

For the PCM-based heat sink cooling technology, the melting temperature is considered the most essential factor for PCM thermophysical properties. The thermophysical properties of each studied PCM material are presented in Table 2.

Figure 3 presents the temperature-time evolution measured at the bottom of the heat sink under a uniform heat flux $Q = 2800$ W/m^2. In this figure, thermal energy performance is studied for n-Eicosane, paraffin wax and salt hydrate presented respectively by the black, red and blue curves.

Table 2. PCMs thermo-physical properties

Material	Th. conductivity (W/mK)	Specific heat (kJ/kgK)	Latent heat (kJ/kg)	Solidification point (°C)	Melting point (°C)	Density (kg/m^3)
Paraffin Wax	0.212 (s)	2.8	173.6	56	58	880 (s)
	0.167 (l)					790 (l)
n-Eicosane	0.39 (s)	1.9 (s)	237.4	35.5	36.5	810 (s)
	0.157 (l)	2.2 (l)				770 (l)
Salt hydrate	0.6	2	200	27	32	1500

It can be considered that salt hydrate is one of the best PCM material comparing with other PCMs. In fact, salt hydrate has the ability to store more thermal energy than other PCM benefiting from higher thermal conductivity as well as its higher density.

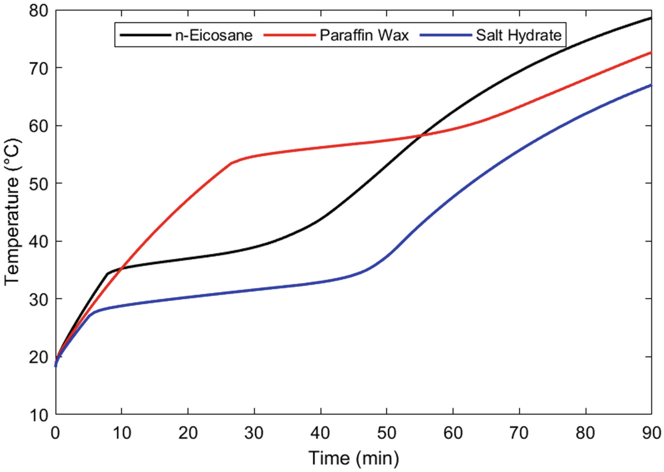


Fig. 3. Temperature-time profile for the studied PCMs.

At t = 90 min, it can be clearly seen that the maximum reached temperature for salt hydrate is about 67 °C and it presents the lower temperature comparing with other ones. Then, salt hydrate will be taken as PCM for the following sections.

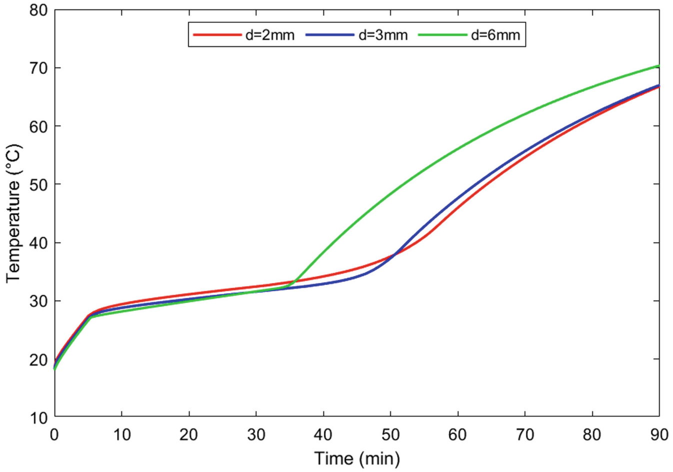


Fig. 4. Temperature-time evolution for each pin diameter variation.

4.2 Pin Diameter Variation

The configuration of the studied heat sink is a 11 × 12 round pin-finned matrix. Figure 4 shows the temperature-time profile for each pin-fin diameter. Noting that red, blue and green curves present the temperature evolution for $d = 2$ mm, $d = 3$ mm and $d = 6$ mm pin-fin diameter.

This study demonstrates that the configuration with a pin diameter of 2 mm presents the best thermal energy behavior in term of thermal energy storage as well as maximum reached temperature.

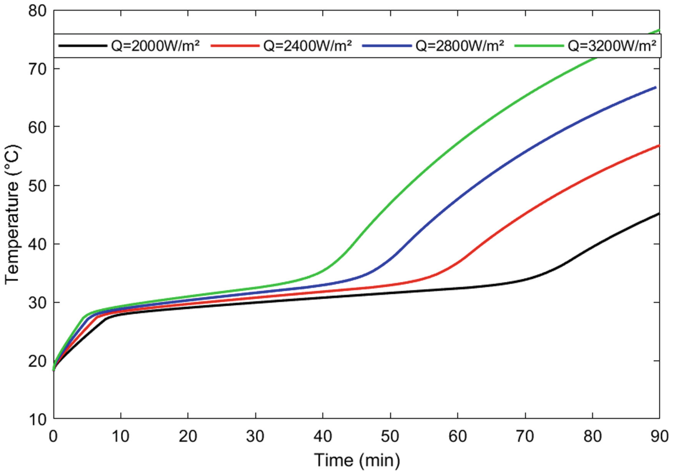


Fig. 5. Temperature-time evolution for different heat fluxes load.

4.3 Power Level Variation

The temperature-time profile history is presented in Fig. 5 for different heat loads: 2000, 2400, 2800 and 3200 W/m^2. It can be noted that salt hydrate is used as PCM. Referring to Fig. 4, it can be clearly seen that the latent heating phase, which it the period where PCM changes from the solid state to its melting, decreases when the heating power increases. For $Q = 2000$ W/m^2, PCM reaches its melting phase at $t = 75$ min but it ends within $t = 40$ min for a heat power about $Q = 3200$ W/m^2.

5 Adopted Parameters for the Studied Cooling System

The parametric analysis leads to investigate the thermal performance as well as to determine an optimal design which has the ability to give a minimal reached temperature for the charging phase. Several parameters are studied to give the best thermal energy management. In this study, a PCM-based 11×12 round pin-fin heat sink matrix is carried out. A parametric study shows that salt hydrate has the ability to store a high thermal energy comparing with n-Eicosane as well as paraffin wax. In addition, a heat sink with a 2 mm pin-fin diameter presents the optimal heat sink geometry.

6 Conclusion

In this study, a finite element simulation of a cooling system is carried out to quantify the thermal energy behavior of a PCM-based round pin-fin heat sink fulfilled with PCM. This study presents the impact of parameters (PCM selection, pin diameter and input heat flux) on thermal energy performance of the studied model. In fact, salt hydrate is considered as the best PCM which has the ability to store more thermal energy during the charging phase comparing to other PCMs (n-Eicosane and Paraffin Wax). Increasing the heat power level has the same effect on the maximum temperature measured at the end of charging phase. Otherwise, the latent heating phase, which is the period where PCM changes from the solid state to its melting, decreases when the heat power increases. Numerical results confirm the benefits of finite element simulations for passive cooling applications especially for PCM-based heat sink case. This study provides an optimal design and it presents an efficient thermal energy management. An experimental study is required to ameliorate the optimal obtained configuration and to determine all studied physical parameters of the optimal model.

References

1. Arshad, A., Ali, H.M., Khushnood, S., Jabbal, M.: Experimental investigation of PCM based round pin-fin heat sinks for thermal management of electronics: effect of pin-fin diameter. Int. J. Heat Mass Transf. **117**, 861–872 (2018). https://doi.org/10.1016/j.ijheatmasstransfer.2017.10.008

2. Debich, B., Yaich, A., Elhami, A., Gafsi, W., Walha, L., Haddar, M.: Coupling PCM-based Heat Sinks finite elements model for mechatronic devices with Design Optimization procedure. In: 2020 IEEE 6th International Conference on Optimization and Applications (ICOA), pp. 1–4. IEEE, April 2020
3. Kumar, A., Kothari, R., Sahu, S.K., Kundalwal, S.I., Paulraj, M.P.: Numerical investigation of cross plate fin heat sink integrated with phase change material for cooling application of portable electronic devices. Int. J. Energy Res. **45**, 8666–8683 (2021). https://doi.org/10.1002/er.6404
4. Baby, R., Balaji, C.: Thermal performance of a PCM heat sink under different heat loads: an experimental study. Int. J. Thermal Sci. **79**, 240–249 (2014). https://doi.org/10.1016/j.ijthermalsci.2013.12.018
5. Debich, B., El Hami, A., Yaich, A., Gafsi, W., Walha, L., Haddar, M.: Design optimization of PCM-based finned heat sinks for mechatronic components: a numerical investigation and parametric study. J. Energy Storage **32**, 101960 (2020). https://doi.org/10.1016/j.est.2020.101960
6. Debich, B., El Hami, A., Yaich, A., Gafsi, W., Walha, L., Haddar, M.: An efficient reliability-based design optimization study for PCM-based heat-sink used for cooling electronic devices. In: Mechanics of Advanced Materials and Structures, pp. 1–13 (2020). https://doi.org/10.1080/15376494.2020.1836291

Modeling and Simulation of a Swarm Robot Application Using MBSE Method and Multi-agent Technology: Monitoring Oil Spills

Khalil Aloui[1,2](✉), Moncef Hammadi[1], Amir Guizani[2], Thierry Soriano[1],
and Mohamed Haddar[2]

[1] QUARTZ Lab EA7393, SUPMECA, 3 rue Fernand Hainaut, 93400 Saint-Ouen, France
`{moncef.hammadi,thierry.soriano}@supmeca.fr`
[2] LA2MP, University of Sfax, Ecole National D'Ingénieurs, 3038 Sfax, Tunisia
`mohamed.haddar@enis.rnu.tn`

Abstract. Swarm Robotics is a new approach to the coordination of a large number of robots inspired by nature. This approach aims to design collective behaviors for many robots. Several researchers have tried to develop structured design methods, but unfortunately, these methods are still limited. Today, swarm robotics are used in many fields that include agriculture, medicine, industrial, etc. One of the most important fields that require swarm robots is surveillance. In this chapter, we present in the first section some methods of designing swarm robot systems by identifying swarm engineering based on a model (MBSE) and multi-agent simulation. Then, we study the energy problem of these robotic systems and the solution proposed by the researchers. In the second section, we present an application for detecting oil in the sea and cleaning it using swarm robots. We will model this application using the MBSE method with the SysML language. We will use the different diagrams of SysML to specify the system requirements and model the functions offered by the system. Finally, we will simulate the models on a multi-agent tool to identify the functional and structural architecture of the system. Throughout this approach, we check the transition from one step to another to ensure the consistency and continuity of the method.

Keywords: Swarm robotic · SysML language · MBSE method · Multi-agent technology · Sustainable energy

1 Introduction

1.1 Swarm Robotics Criteria

Swarm robotics is a new approach to coordinate a group of robots inspired by social insects (Şahin 2005). It aims to design a scalable, robust and flexible collective behaviors for the coordination of a large group of robots. This approach is often inspired from social animals such as bees, ants, fish and birds (Dorigo et al. 2007). Indeed, swarm

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 96–105, 2022.
https://doi.org/10.1007/978-3-030-85584-0_10

robotics tends to develop a robotics system that are similar to behaviors of social animals. In particular, swarm robotics systems are meant to be robust, scalable and flexible (Camazine et al. 2003). Scalability can be defined as the ability to generate modularized solutions to different tasks that work well with different group sizes. In swarm robotics, scalability is promoted by local detection and communication. Robustness can be defined as the ability to operate despite failures in individuals. In swarm robotics, robustness is the redundancy and the absence of a leader. Flexibility can be defined as the ability to operate under a wide range of different environments and tasks. The simplicity of behaviors and mechanisms represent flexibility in swarm robotics.

1.2 Swarm Robotics Advantages

A swarm robotic system is characterized by its members' autonomy. These robots must be homogeneous. In some cases, there may be different types of robots in the swarm, but these groups should not be too large. In addition, the number of robots in the swarm must be large. These robots must have distributed coordination. Ultimately, robots must collaborate to achieve or improve the performance of a primary task that cannot be solved by a single robot since the detection range of a group of robots is wider than that of a single robot (Navarro et al. 2013). Groups of robots can make decomposable tasks more efficient when they use parallelism. Moreover, a single robot is unable to operate in different places simultaneously while a group of robots can do that. Another advantage is fault tolerance (i.e. thanks to the redundancy of the system, the task remains accomplished despite the failure of a single robot within a group). There are also several drawbacks of swarm robot systems, such as uncertainty about other robots' intentions (i.e., coordination requires knowing what other robots are doing. If this is not clear, robots can compete instead of cooperating). Due to collisions, robots can interfere with each other in a group. Finally, the system's overall cost is very high since the use of more than one robot can increase the economic cost (Arkin et al. 1998).

1.3 Swarm Robotics Applications

Several real-world issues require the application of swarm robots, such as monitoring an environmental accident. A swarm robot system can instantly monitor and detect hazardous events, such as a chemical spill due to its distributed detection capability. A swarm robot system can focus on one problem. First, the swarm identifies the nature of the problem. It mobilizes its members towards the problem's source. Second, the swarm forms a patch that blocks the leak using its self-assembly characteristic (Şahin 2005).

Depending on the given mission, the swarm robot systems can increase or decrease their number of individual robots. For example, during an oil spill, a swarm robot system will be built to contain an initial spill in the area. We will study this case in the rest of this paper.

2 Design Methods

The design of the swarm robot system means the planning and development of a swarm system from the initial specifications and requirements. Unfortunately, there is no precise and structured methodology for designing desired collective behavior through individual behaviors. The main ingredient in developing these types of systems is always the designer's intuition (Ammar et al. 2018).

Traditionally, many designers have used the code-and-fix approach to develop robot swarms. It is a bottom-up method means that the developer creates and improves the individual behaviors of the robots until the desired collective behavior is achieved. Kazadi et al. (2007) applied the swarm engineering methodology called the "Hamiltonian method" to the problem of artificial physics. This methodology is used to create swarms of predefined global properties. They apply this methodology to the problem of artificial physics which locally creates hexagonal grids of agents. Brambilla et al. (2014) presented a new top-down design method for swarms of robots based on prescriptive modeling and model verification called "Property Driven Design". This method consists of four phases: first, the designer specifies the requirements of the robot swarm. Second, the designer creates a normative model of the swarm. He uses model verification to verify that this normative model satisfies the desired properties; third, the designer simulates the model and implement the simulated version to validate the developed model; fourth, the designer implements the desired robot swarm to validate the previous steps. Aloui et al. (2020) improved the "Property Driven Design" design method. They used the SysML language for the specification and modeling of the Swarm robotic system. They used SysML state machine diagrams to describe robot behaviors. Then, the robot behavior models described with SysML are implemented with a multi-agent technique (Guizani et al. 2017).

2.1 Model-Based Systems Engineering MBSE

In general, the model is a simplified version of a concept, relation or phenomenon to facilitate understanding of a structure or system. In the field of swarm robotics, it is possible to use the MBSE method (Model-Based Systems Engineering) to model swarm robot systems to understand behavioral analysis, performance analysis, requirements traceability, system architecture, simulation, testing, etc. (Mhenni et al. 2014). Model-based systems engineering (MBSE) is a formalized methodology that is used to support the requirements, design, analysis, verification, and validation associated with the development of complex systems. A simplified definition of MBSE is provided by Mellor et al. (2003) as "…is simply the notion that we can construct a model of a system that we can transform into the real thing". The main advantage of a model-based approach is to determine the scope of the problem. That is, the developed robot swarm best meets expectations when the properties are clear and complete with the models. The detailed description of the requirements by the models reduces the risk of developing "the wrong robot swarm". To ensure model-based swarm engineering, the swarm developer can use the graphical modeling language SysML. SysML is a visual modeling language that supports the specification, design, analysis, verification, and validation of complex systems. SysML is used with different methodologies including object orientation, structured analysis, and others (Hause et al. 2006). This language includes diagrams that can be used to specify system requirements, system behavior, system structure, and parametric relationships.

2.2 Multi-Agent Simulation MAS

Recently, several researchers have started to use multi-agent systems in different fields, such as automobile design (Guizani et al. 2015), power electronics (Hammadi et al. 2011) and aerospace design. A multi-agent system is a field of artificial intelligence (AI) (Guizani et al. 2014) are focused on the agent and his autonomy. They defined the agent as a computer system located in its environment and able to act autonomously in order to achieve its design goals. The study of swarm robotic systems has been considered today as a study of multi-agent systems. Developers of swarm robot systems use realistic simulators to test and accelerate the development of new design methods. They use simulators to model the interactions between robots and the interactions of robots with their environment (Şahin et al. 2008). Luke designed a multi-agent simulation toolkit called MASON to present a many multi-agent simulation tasks ranging from machine learning and swarm robotics to environments of social complexity (Luke et al. 2005). Anylogic is another multi-agent simulation tool developed by The AnyLogic Company. It is a graphical modeling language that makes it easy to extend the simulation model with Java code (Borshchev et al. 2002).

2.3 Sustainable Energy Efficiency of Swarm Robots

Each time the groups of robots in the swarm recharge to their maximum level. Swarm robot systems do not estimate the energy needed for their next task. They require a method to minimize the overall energy cost in a swarm and simultaneously maximize the performance of the swarm. However, in the current literature, only a few articles consider energy efficiency in a swarm. Labella et al. (2004) have developed an adaptation method which manages the number of robots foraging in the environment. Indeed, they efficiently use energy when needed, but they measure only the time needed to complete the mission without measuring the amount of energy already allocated. Liu et al. (2006) focused on energy efficiency. They modify the foraging time for each robot. Stirling et al. (2010) have worked on flying swarms in indoor environments. They proposed a new approach to find energy-efficient search algorithms. They noticed that the progressive launch of the robots increases the search time and reduces the total energy consumption and the collision rate. Chen et al. (2019) propose a method that improves the efficiency of energy management methods. This method uses upper and lower battery very rapid thresholds.

In the next section, we will model a swarm robot application for oil spill monitoring using the different SysML diagrams. Then, we will simulate this application on a multi-agent tool.

3 Case Study: Oil Spill Monitoring

An oil spill in the areas has both immediate and long-term effects on the environment. The fauna of these areas can be directly exposed: either by swimming in the oil, by inhaling its toxic fumes, or by swallowing food affected by the oil. This exposure to petroleum can cause serious health problems and even lead to death. Thus, researchers are still trying to find solutions to clean up the affected environment (Şahin 2005). We use the different SysML diagrams to model this application.

3.1 Requirement Specification

The requirements diagram shown in Fig. 1 specifies the application requirements for detecting environmental pollutants and cleaning them. Detecting environmental pollutants requires exploring the entire environment by deploying all the robots that detect and clean these pollutants. Cleaning should be quick and efficient using autonomous robots.

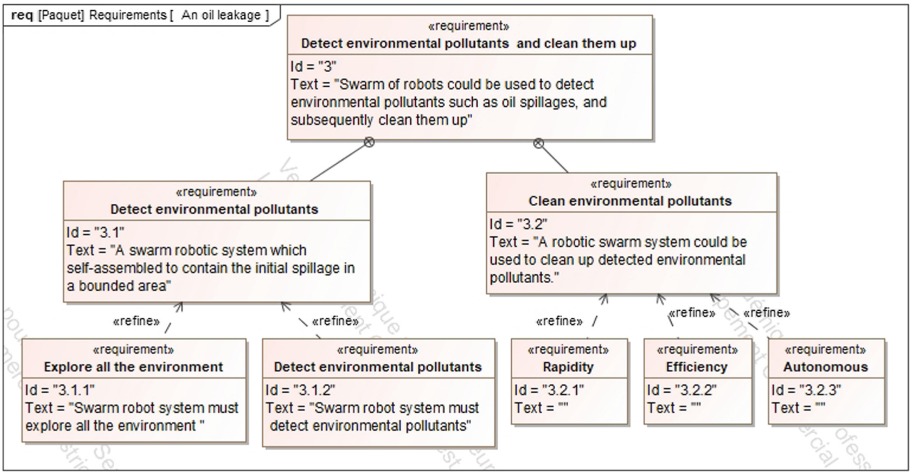


Fig. 1. Requirements diagram

3.2 Functional Architecture

In this step, we describe the functional scenario using the different collective behaviors of swarm robots. Figure 2 shows this scenario. First, members of the swarm begin to explore the entire region. Then they detect environmental pollutants. Finally, they regroup in these areas, form groups and clean up.

To model this case study, some collective behaviors of swarm robots are used:

- **Collective exploration**: The swarm of robots cooperatively navigates the environment in order to explore and monitor it. First, the swarm is initialized with random positions and constant speeds. If the robots locate spilled oils, they send positive information to the control center. If they detect obstacles or other robots within their reach, they avoid the collision. Otherwise, they send negative information to the control center indicating that there is no oil leak in that area. Subsequently, they update their speeds.
- **Collective localization:** Allows the robots in the swarm to find their position and orientation in relation to each other by establishing a local coordinate system throughout the swarm.
- **Aggregation**: Individual robots gather spatially in a specific region of the environment. This behavior is used to ensure that the robots are assembled in the area to be cleaned.

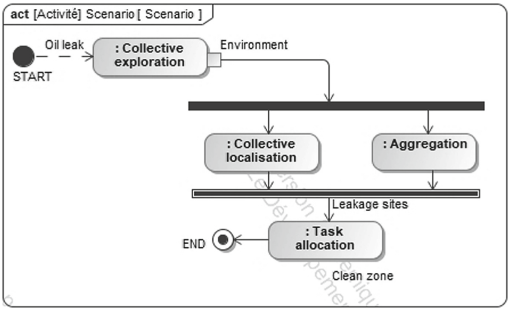


Fig. 2. Descriptive model for cleaning an oil leak

- **Task allocation:** Assignment of tasks dynamically assigned emerging tasks to individual robots in the swarm. Its objective is to maximize the performance of the entire swarm system. In this case study, there are 4 groups of robots distributed throughout the area to completely clean the contaminated area.

After identifying the functional model of the system, we simulate the model using AnyLogic.

3.3 Swarm Simulation

In this step, AnyLogic is used as a multi-agent software tool to implement the swarm simulation model. In this case study, we chose 100 robots to do this mission. As shown in Fig. 3, we have divided the region into four zones according to the degree of contamination: zone A and D are the two most polluted zones. So, the swarm of robots must be directed towards them more. Zone B is the moderately polluted zone and finally, Zone E is the smallest contaminated zone. Figure 3 gives the initial number of robots in each zone.

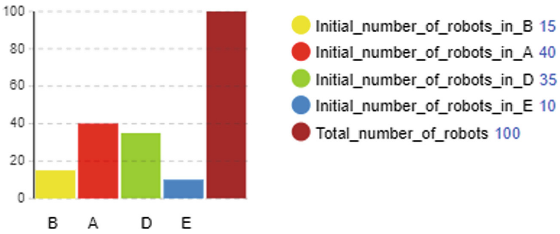
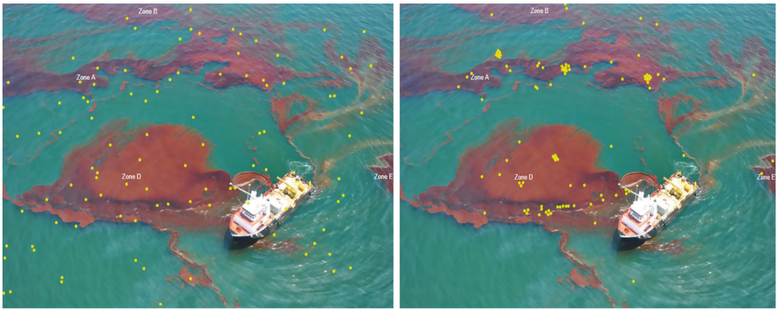


Fig. 3. Initial distribution of robots in zones

Each time, the swarm of robots must return to the boat to dispose of the captured pollutants. The initial speed of each robot is 20 km/h. In fact, we have the right to direct the swarm by adjusting the probability values of the robots passing from the boat to different areas. The exploration time for the whole area is 1 h, the service time in each area is 50 min, and the time to dispose of the pollutants in the boat is 30 min. In addition, the robots must form swarms of 10 robots in each area. At first, robots begin to explore the entire region to identify contaminated areas. They then regroup in these areas, form groups and do the cleaning (Fig. 4).



(a) **(b)**

Fig. 4. Simulation of an oil spill cleanup in the ocean (a): Robots explore the entire region. (b) Robots clean the different areas

Figure 5 represents the indicators of system performance: The red curve represents the number of robots that clean area A. This number is always greater than 15 robots. Likewise, the green curve represents the number of robots that clean the area D. This number is always greater than 10 robots. This means that the characteristics specified initially have been validated (i.e., the robotic swarm is already directed towards the two most polluted regions). In addition, the robotic swarm treated the two less polluted areas B and E with the fewest robots. The yellow curve indicates the number of robots that return to the boat and throw the oil into the storage tanks. The most important main step is to contain the spill to as small an area as possible because the oil diffusion is very rapid. Each robot swarm was made up of 10 robots that scatter throughout the region and soak up the spilled oil.

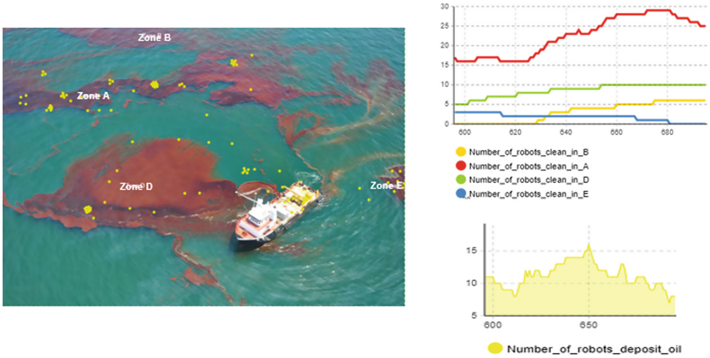


Fig. 5. System performance indicators

3.4 Structural architecture

In this section, the block definition diagram shown in Fig. 6 is used to define the various components of the system. The swarm consists of a group of individual autonomous robots. Several values are defined, such as the size of the swarm, the area covered by the robots, the initial position and the duration of the mission.

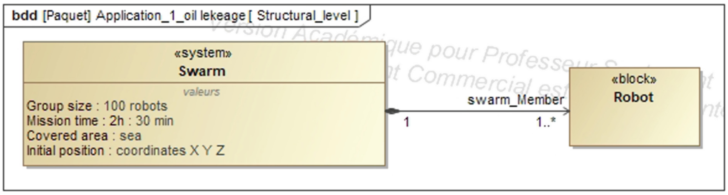


Fig. 6. The swarm architecture

In Fig. 7, the individual robots that form the swarm are described in the block definition diagram. Indeed, each robot consists of a movement system that ensures the robot swimming in the sea, a power system, and a communication and control system. At this level, the various sensors and electronic components are defined to design the system.

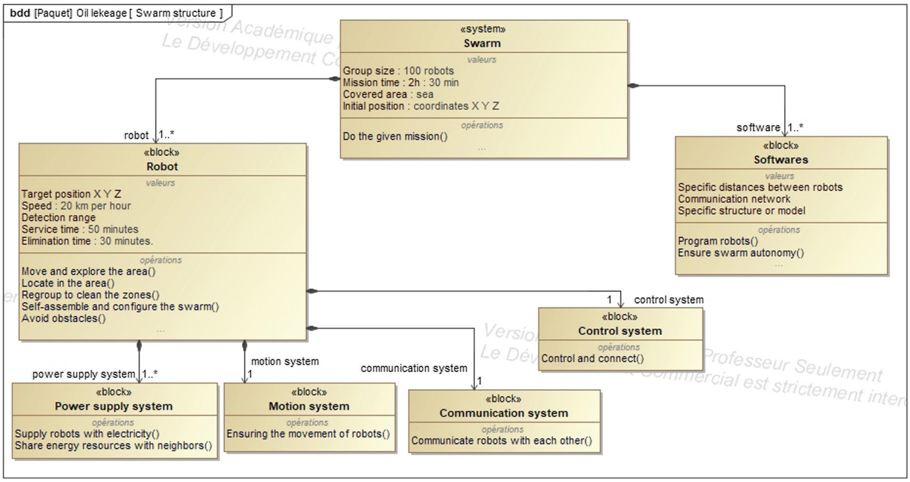
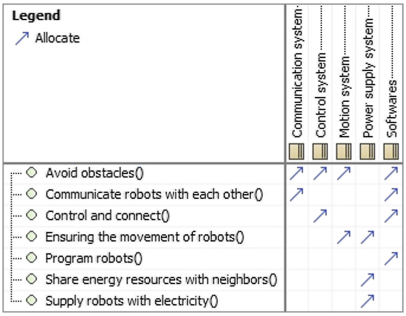


Fig. 7. Block definition diagram of a swarm of robots

Finally, it is necessary to verify that the components of the system described in the BDD correspond to the various functions of the system. The allocation matrix shown in Fig. 8 relates each component of the system to its functions.

Finally, the swarm developer chooses the implantation system (like the ROS platform) to implement this system on real robots. Researchers can realize the task of implementation in further research and next steps.



Legend
↗ Allocate

	Communication system	Control system	Motion system	Power supply system	Softwares
Avoid obstacles()	↗	↗	↗		↗
Communicate robots with each other()	↗				↗
Control and connect()		↗			↗
Ensuring the movement of robots()			↗	↗	
Program robots()					↗
Share energy resources with neighbors()				↗	
Supply robots with electricity()				↗	

Fig. 8. Allocation matrix component-function

4 Conclusion

Swarm robotics is a new approach to robot coordination inspired by social insects. This approach aims to design evolving, robust, and flexible collective behaviors for many robots. Unfortunately, the design methods for these systems are still limited. Several researchers have proposed approaches to simplify design techniques, but this field is still limited.

This chapter has described swarm robotics: their main definition, their areas of application, their design methods, and their sustainable energy efficiency. Then, we presented the modeling and simulation methods of multi-agent systems. We modeled an oil spill detection application using the MBSE method with SysML diagrams. Finally, we simulated this application on a multi-agent tool to determine the robot swarm's functional and structural architecture. The implementation of real robots and the production of prototypes can be the subject of future papers.

References

1. Aloui, K., Hammadi, M., Soriano, T., Guizani, A., Haddar, M.: On the continuity of the swarm robot design using MBSE method and simulation. In: 13th International Conference on Modelling, Optimization and Simulation (MOSIM 2020) (2020)
2. Ammar, R., Hammadi, M., Choley, Jean-Yves., Barkallah, M., Louati, J., Haddar, M.: The Design and Modeling of an Optimized Mechatronic System Using a Set Based Concurrent Engineering. In: Fakhfakh, T., Karra, C., Bouaziz, S., Chaari, F., Haddar, M. (eds.) ICAV 2018. ACM, vol. 13, pp. 111–120. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-94616-0_11
3. Arkin, R.C., et al.: Behavior-based robotics. MIT Press (1998)
4. Borshchev, A., Karpov, Y., Kharitonov, V.: Distributed simulation of hybrid systems with AnyLogic and HLA. Futur. Gener. Comput. Syst. **18**, 829–839 (2002)
5. Brambilla, M., Brutschy, A., Dorigo, M., Birattari, M.: Property-driven design for robot swarms: a design method based on prescriptive modeling and model checking. ACM Trans. Autonom. Adapt. Syst. **9**, 1–28 (2014)
6. Brambilla, M., Ferrante, E., Birattari, M., Dorigo, M.: Swarm robotics: a review from the swarm engineering perspective. Swarm Intell. **7**, 1–41 (2013)

7. Camazine, S., Deneubourg, J.-L., Franks, N.R., Sneyd, J., Bonabeau, E., Theraula, G.: Self-organization in biological systems. Princeton University Press (2003)
8. Chen, A., Harwell, J., Gini, M.: Maximizing energy battery efficiency in Swarm robotics. *arXiv preprint* arXiv:1906.01957 (2019)
9. Dorigo, M., Birattari, M., et al.: Swarm intelligence. Scholarpedia **2**, 1462 (2007)
10. Friedmann, M.: Simulation of autonomous robot teams with adaptable levels of abstraction. Ph.D. Dissertation, Technische Universität (2010)
11. Guizani, A., Hammadi, M., Choley, J.-Y., Soriano, T., Abbes, M.S., Haddar, M.: Agent-Based Approach for the Optimal Design of Mechatronic Systems. In: Haddar, M., et al. (eds.) Multiphysics Modelling and Simulation for Systems Design and Monitoring. ACM, vol. 2, pp. 189–198. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-14532-7_20
12. Guizani, A., Hammadi, M., Choley, J.-Y., Soriano, T., Abbes, M.S., Haddar, M.: Agent-based approach for collaborative distributed mechatronic design. In: 2014 10th France-Japan/8th Europe-Asia Congress on Mecatronics (MECATRONICS2014-Tokyo), pp. 156–161 (2014)
13. Guizani, A., et al.: Multi-agent approach based on a design process for the optimization of mechatronic systems. Mech. Ind. **18**, 507 (2017)
14. Hause, M., et al.: The SysML modelling language. In: Fifteenth European Systems Engineering Conference, vol. 9, pp. 1–12 (2006)
15. Hammadi, M., Choley, J.Y., Penas, O.: Layout optimization of power modules using a sequentially coupled approach. Int. J. Simul. Model. **10**(3), 122–132 (2011)
16. Kazadi, S., Lee, J.R., Lee, J.: Artificial physics, swarm engineering, and the Hamiltonian method. In: World Congress on Engineering and Computer Science, pp. 623–632 (2007)
17. Labella, T., Dorigo, M., Deneubourg, Jean-Louis.: Efficiency and task allocation in prey retrieval. In: Ijspeert, A.J., Murata, M., Wakamiya, N. (eds.) BioADIT 2004. LNCS, vol. 3141, pp. 274–289. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-27835-1_21
18. Liu, W., Winfield, A., Sa, J., Chen, J., Dou, L.: Strategies for energy optimisation in a swarm of foraging robots. In: Şahin, E., Spears, W.M., Winfield, A.F.T. (eds.) SR 2006. LNCS, vol. 4433, pp. 14–26. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-71541-2_2
19. Luke, S., Cioffi-Revilla, C., Panait, L., Sullivan, K., Balan, G.: Mason: a multiagent simulation environment. Simulation **81**, 517–527 (2005)
20. Mellor, S.J., Clark, T., Futagami, T.: Model-driven development: guest editors' introduction. IEEE Softw **20**(5), 14–18 (2003)
21. Mhenni, F., Choley, J.-Y., Penas, O., Plateaux, R., Hammadi, M.: A SysML-based methodology for mechatronic systems architectural design. Adv. Eng. Inform. **28**, 218–231 (2014)
22. Navarro, Iñaki., Matía, F.: An introduction to swarm robotics. ISRN Robot. **2013**, 1–10 (2013). https://doi.org/10.5402/2013/608164
23. Şahin, E.: Swarm robotics: From sources of inspiration to domains of application. In: Şahin, E., Spears, W.M. (eds.) SR 2004. LNCS, vol. 3342, pp. 10–20. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-30552-1_2
24. Şahin, E., Girgin, S., Bayindir, L., Turgut, A.: Swarm robotics. In: Blum, C., Merkle, D. (eds.) Swarm intelligence, pp. 87–100. Springer Berlin Heidelberg, Berlin, Heidelberg (2008). https://doi.org/10.1007/978-3-540-74089-6_3
25. Stirling, T., Wischmann, S., Floreano, D.: Energy-efficient indoor search by swarms of simulated flying robots without global information. Swarm Intell. **4**, 117–143 (2010)

Diagnosis and Prognosis of Mechanical Components Using Hybrid Methods

Brian Ellis(✉), P. Stephan Heyns, and Stephan Schmidt

Centre for Asset and Integrity Management, University of Pretoria, Pretoria, South Africa
brian.ellis@live.co.za, {stephan.heyns,stephan.schmidt}@up.ac.za

Abstract. Diagnosis and prognosis of mechanical components are important for critical rotating machinery found in the power generation, mining, and aviation industries. Data-driven diagnosis and prognosis methods have much potential; however, their performance is dependent on the quality of historical data . Usually only limited historical data are available for newly commissioned parts and for parts that do not go through a full degradation cycle before being replaced. Physics-based diagnosis and prognosis methods require assumptions of the underlying physics; the governing equations need to be derived and solved; and the model needs to be calibrated for the underlying system. Physics-based methods require extensive domain knowledge and could have modelling biases due to missing physics. Hybrid methods for diagnosis and prognosis of mechanical components have the potential for improving the accuracy and precision of remaining useful life (RUL) estimation when historical fault data are scarce. This is because hybrid methods combine data-driven and physics-based models to alleviate the shortcomings of the respective methods. For these reasons, hybrid methods are getting more attention in the condition monitoring community as a solution for diagnosis and prognosis tasks. Therefore, in this chapter, we present a review of the state-of-the-art implementations of physics-based, data-driven, and hybrid methods for diagnosis and prognosis. The methods are organised using a condition monitoring framework and contributions of various techniques are discussed. We identify gaps in the hybrid diagnosis and prognosis field that could be the focus of future research projects.

Keywords: Hybrid methods · Diagnosis · Prognosis · Remaining useful life

1 Introduction

Mechanical components such as bearings, gears or turbomachine blades are affected by different degradation mechanisms. These degradation mechanisms include, but are not limited to creep, wear, and fatigue crack growth (Cubillo et al. 2016). Remaining useful life (RUL) is defined as the time or number of cycles the damage in a component will remain within the specified limits set by the engineer (“ISO 13381-1” 2015). RUL is one of the most important aspects of predictive maintenance. In predictive maintenance strategies engineers need to know the condition of the component and the RUL of the component to make decisions that either (i) ensure safety against unplanned failure or (ii)

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 106–115, 2022.
https://doi.org/10.1007/978-3-030-85584-0_11

maximise the component's use before unnecessarily replacing the component (Saxena 2010). Condition monitoring is collection of techniques that use sensors to determine the damage of a component while the component is in operation. It forms an integral part of diagnosis and prognosis. Lee et al. (2018) reviewed different condition monitoring techniques applied to rotating mechanical components. The focus of this review was on vibration sensors and acoustic emission sensors. According to the review, vibration-based condition monitoring methods are the most popular for mechanical components (Lee et al. 2018).

Methods for estimating the condition and the RUL of mechanical components from condition monitoring data are often categorised into either (i) physics-based, or (ii) data-driven methods (An et al. 2015). Physics-based methods use first principles to model the damage in the component and to model the response of the system. In contrast, data-driven methods only rely on the available data to find relationships between the different correlated variables and do not account for the underlying physical mechanisms that generate the data. Both physics-based methods and data-driven methods have their limitations (Khan and Yairi 2018). Physics-based can contain model errors, while data-driven methods require historical fault data. Hybrid methods can improve existing methods by addressing the drawbacks of the other methods.

In this chapter, we organise the contribution of different authors into a generalised framework and highlight the applications of hybrid methods in this framework. The objective of this chapter is to present the state-of-the-art methods in diagnosis and prognosis of mechanical components. We identify areas that still need to be addressed, which are presented in Sect. 4 of this chapter.

2 Diagnosis and Prognosis Using Condition Monitoring Data

Figure 1 presents a summary of the condition monitoring process and the various parts that are required to estimate RUL. The most common vibration-based condition monitoring techniques are broadly categorised as (i) machine learning techniques or (ii) Fourier-based methods (Lee et al. 2018). The purpose of these methods is to relate the characteristics of the vibration of the system to the condition of the component.

Diagnosis refers to the identification, localisation, and quantification of the damage. Identification and localisation (otherwise known as anomaly detection) refer to detecting the damaged component, the degradation mechanism, and the exact location of the damage. Anomaly detection is usually performed in early stages of the component's condition monitoring process (Carden and Fanning 2004). Wang et al. (2016) demonstrated the use of spectral kurtosis as an anomaly detection method for determining faults in gear teeth from accelerometer data. Quantification refers to an estimation of the extent of the damage (e.g., the crack length in a steam turbine blade). An accurate estimate of the damage severity is essential for making the appropriate maintenance decisions. The accuracy of the estimate can be quantified by its bias and its variance. The estimate's variance is influenced by electrical, mechanical, and environmental noise sources. The bias is typically influenced by incorrect measurement models (i.e., the model that relates the underlying state to the measured data). Continuous condition monitoring applications enable the implementation of Bayesian filtering algorithms to update and improve the

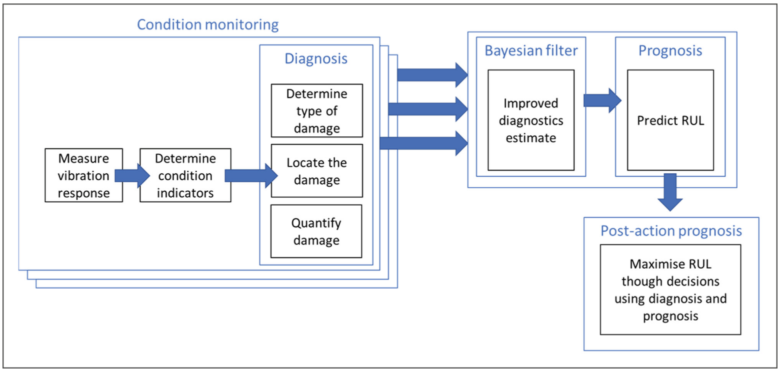


Fig. 1. Condition-based maintenance process from sensor data collection to RUL estimation of critical mechanical components

estimate of the damage severity over time (Jouin et al. 2016). The purpose of a Bayesian filtering algorithm is to improve the diagnosis using a known degradation model, which is critical for RUL prediction.

Prognosis, refers to the estimation of RUL. Sankararaman and Goebel (2014) emphasise the treatment of uncertainty when RUL is estimated and the ways that uncertainty is represented and interpreted from the models. Physics-based models, for instance, are deterministic in nature. Uncertainty can be introduced in these methods by implementing an ensemble of physics-based methods. Data-driven methods, on the other hand, can measure the RUL uncertainty only when enough representative data are collected. RUL estimates from physics-based methods and data-driven methods are limited by the models and data that capture the uncertainty of the component's degradation.

3 Hybrid Methods for Diagnosis and Prognosis of Mechanical Components

Lei et al. (2018) highlight that one of the largest drawbacks of data-driven methods is the availability of run-to-failure data for mechanical components. Mechanical systems are becoming more complex as requirements for flexible manufacturing become more prevalent. As a result, prognostics has reached a tipping point where insufficient data are available for training, particularly for newly commissioned components. Physics-based models, on the other hand, require data to validate their accuracy since they are limited by the modelling assumptions (An et al. 2015). Model inadequacy, where the model does not capture all of the physics of the system, could also be detrimental to the performance of physics-based methods. Liao and Köttig (2014) suggest the use of hybrid techniques for cases where data are scarce and simplified physics-based models are available. We distinguish between hybrid combination frameworks (that is, combinations of different model types for unique purpose in a condition monitoring framework); and hybrid fusion models (that is, different types of models that are combined for the same purpose). An example of a hybrid fusion model is illustrated by Coppe et al. (2012) whereby physics-based crack growth models (e.g. Paris' law) are combined with data-driven observations of the crack length to form a hybrid crack growth prediction model. Most of the models that Liao and Köttig (2014) propose are hybrid combination frameworks. An example of a hybrid combination framework is exhibited by Sanchez et al. (2016) for determining the RUL of wind turbine blades. Sanchez et al. (2016) proposes a physics-based method for determining the stiffness of a blade from the blades' vibration characteristics and a data-driven method for predicting the changes in the blades' stiffness due to crack growth at the root of the blade.

3.1 Diagnosis: Estimating Current Damage

Diagnosis requires condition indicators that unique identify the component and the damage mechanism present in the component (Lei et al. 2018). Incorrect identification may cause all other methods to fail and could result in unnecessary maintenance costs. Hence, anomaly detection methods often form part of the diagnosis procedure. Serafini et al. (2019) use simulated data to detect localised stiffness reduction of helicopter blades from strain gauge measurements.

Physics-based diagnosis models, model the relationship between the condition indicator and the damage using a first-principles approach. Zeng et al. (2018) demonstrated using finite element simulations that there is a comparable difference in the vibration characteristics of a compressor rotor blade with and without a crack. Elshamy et al. (2018) used the first three natural frequencies of a cantilever beam to uniquely identify the depth and location of cracks. Corrado et al. (2018) expanded on this idea to diagnose multiple cracks and the locations of these crack from only the mode shape. The methods only use finite element simulations to construct these models.

Data-driven diagnosis, on the other hand, apply machine learning techniques to quantify the damage mode and the extend of the damage. Here we refer to the data as collections of the condition indicators and the damage histories. Kaloop and Hu (2015) detected and localised faults in stayed-cable bridges from accelerometer data. Jia et al. (2016) and Zhang et al. (2018) demonstrated the use of deep neural networks and convolutional neural networks respectively for classifying types and severity of faults in bearings from raw accelerometer data. The accuracy of the data-driven diagnosis method surpasses the accuracy of most physics-based diagnostics models (Khan and Yairi 2018). This is because the performance of data-driven methods scale well with the size of the data (Bishop 2006). The more data that is available, the less likely a complex model will overfit the data.

There is however a gap in the literature for methods that diagnose the damage from condition indicators using hybrid fusion models. In a *hybrid fusion diagnosis* model, the physics-based diagnosis and data-driven diagnosis models are combined to improve the quantification of the damage during condition monitoring. That is, after determining the condition indicators from vibration sensors a diagnostics model should quantify the damage and estimate the uncertainty of this quantification. There is presently very limited literature on hybrid fusion diagnostics models. This may be attributed to Bayesian filtering techniques that are used to update the estimation of the current damage from known degradation models. As a result, most researchers deem it unnecessary to improve the quantification of the current damage twice. We discuss these Bayesian filtering techniques next.

3.2 Improved Diagnosis: Bayesian Inference Methods

It has become standard practice to use Bayesian filtering methods in condition monitoring methods (Corbetta et al. 2018a, b) . The purpose of Bayesian filtering is outlined by Jouin et al. (2016): A Bayesian filtering method is only used for improving state estimation and has no predictive capabilities. Since, models that describe the rate of degradation are often a function of damage in the component (Cubillo et al. 2016) it is important to improve the diagnostics obtained from regular condition monitoring to improve RUL estimation.

Let x_k denote the hidden health state estimate at the k^{th} condition monitoring step. The condition indicators at all condition monitoring steps are denoted $y_{1:k}$. A Bayesian filter typically consists of two steps. Firstly, predict the probability of the health state at the using the k^{th} condition monitoring step with

$$p(x_k|y_{1:k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|y_{k-1})dx_{k-1} \tag{1}$$

This is also called the Chapman-Kolmogorov equation. Secondly, update the probability distribution of the health state using all the condition monitoring measurements until step k using Bayes' rule

$$p(x_k|y_{1:k}) = \frac{p(y_k|x_k)p(x_k|y_{1:k-1})}{p(y_k|y_{1:k-1})} \tag{2}$$

The normalising constant is determined from

$$p(y_k|y_{1:k-1}) = \int p(y_k|x_k)p(x_k|y_{1:k-1})dx_k \tag{3}$$

Jouin et al. (2016) show that there are analytical and approximate solutions to these equations. Corbetta et al. (2018a, b) used particle filter-based methods for performing Bayesian filtering. Particle filters are approximate inference methods that solve the update and predict equation, using Monte Carlo methods. These particle filters can be degenerate under some modelling assumptions. Therefore, using non-additive process noise is advised. Valeti and Pakzad (2017) and Zaidan et al. (2015) applied particle filtering methods to cracks in wind turbine blades and aerospace gas turbine engines.

3.3 Prognosis: Estimating RUL

RUL is estimated from the posterior distribution of the Bayesian filter. The estimated damage at the last condition monitoring step is propagated using a degradation model. Cubillo et al. (2016) presented an extensive summary of physics-based degradation models for creep, wear, and fatigue crack growth. These models are mostly derived in the form of a differential equation

$$\frac{dx}{dt} = f(x, u(t)) \tag{4}$$

where $f(x, u)$ denotes the non-linear function that describes degradation rate from the currently estimated damage x, and the future operating condition $u(t)$. The Paris-Erdogan law is an example of this model where $x = a$ denotes the crack length and $u(t) = \Delta S(N)$ denotes the stress amplitude as a function of the number of loading cycles, N, instead of time, t. The equation

$$\frac{da}{dN} = C(\Delta K)^m = C\big(F(a)\Delta S(N)\sqrt{\pi a}\big)^m \tag{5}$$

predicts the crack growth rate from material parameters C and m, and the geometric factor $F(a)$. Keprate et al. (2017) presented a Gaussian processes regression surrogate model to quickly evaluating the stress intensity factor range ΔK, in offshore pipelines. Typically, finite element simulations or empirical formulas are used for determining the stress intensity factor range using a physics-based approach. Integrating the degradation model analytically or numerically will solve the degradation path (that is, the damage as a function of time). When the degradation path crosses the fault-specification-limit, the RUL is recorded.

Wang et al. (2020) present a collection of data-driven methods that do not directly measure the condition of the component. The methods estimate the RUL from the condition indicators directly using a relevance vector machine. Data-driven prognostics methods often use run-to-failure data to predict RUL directly from condition indicators without a diagnosis of the component. Khan and Yairi (2018) demonstrate the capabilities of data-driven methods that can perform classification and regression from very little understanding of the data. However, these models do not address uncertainty, which is a critical aspect for RUL estimation.

Hybrid-fusion-based prognostics methods have become popular since the introduction of damage propagation model parameters as part of the Bayesian filter step. Coppe et al. (2010, 2012) first introduced the concept by introducing Paris's law parameters as part of the hidden state to model cracks in a large plate. When introducing the model parameters as part of the hidden state, a posterior probability distribution of the model parameters can be inferred from regular diagnosis. Thus, a physics-based model is augmented with condition monitoring data to improve the model. Corbetta et al. (2018a, b) applied a similar technique to multi-degradation modes of fibre reinforced laminates with matrix cracks and delamination. Corbetta et al. (2018a, b), further termed these types of models as artificial dynamics models. (Chen et al. 2018; Saidi et al. 2018) applied the artificial dynamics approaches to attachment lugs and wind turbine bearings, respectively.

4 Future Aspects of Hybrid Methods

In Fig. 2 we present our hypothesis for the performance of hybrid techniques for RUL predictions. Performance metrics of RUL estimation refer to the accuracy, precision, consistency and robustness of estimating the true RUL of a mechanical component (Saxena et al. 2010); hence, we omit the scale of the dependent axis in Fig. 2. Physics-based models that do not rely on the data for model updating do not change in performance. A Bayesian approach to hybrid diagnosis and prognosis is expected to use physics-based models as a prior model and augment the model with data. Thus, it is expected that hybrid methods will have increased performance compared to physics-based models.

Data-driven methods, on the other hand, are also expected to increase with the accumulation of data. After a certain amount of data is collected data-driven methods will meet the performance of hybrid methods since the likelihood will be very localised and therefore most dominant in the posterior probability of the RUL (Khan and Yairi 2018).

Future applications of condition monitoring methods will require automated diagnosis and prognosis of critical assets. Even though data-driven models could theoretically enable automatic diagnosis and prognosis, mechanical assets typically do not have representative failure data available to properly train data-driven models. This means that the machines will operate on the left end of the graph in Fig. 2 and therefore hybrid methods will be essential for future diagnosis and prognosis tasks.

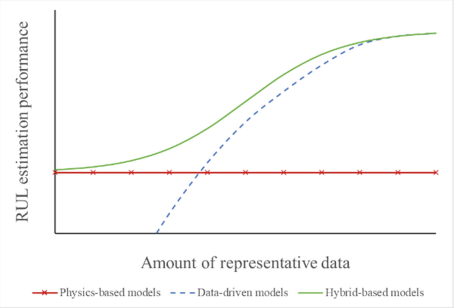


Fig. 2. A hypothesis of the performance of hybrid methods compared to data-driven and physics-based methods.

Corbetta et al. (2017) and Le et al. (2015) investigated hybrid prognostics frameworks for co-existing damage modes. Treatment of degradation modes on an individual basis has some severe consequences since the co-existing damage modes ‘fuel’ the degradation process and consequently the component may fail earlier than predicted. Cubillo et al. (2016) mentioned that fatigue crack growth and creep are stimulated by one-another. Therefore, future work in diagnosis should not only identify a single fault in a component but also identify when multiple failures occur simultaneously.

The future development of hybrid methods is, however, not limited to improvements of diagnosis and prognosis. Sikorska et al. (2011) presented a review of prognostics options for industry and listed many limitations of practically implementing these methods. Most of these problems have since then been addressed. However, possible actions that maximise the RUL of a component based on diagnosis and prognosis is still a developing field. Methods for optimising the RUL by proposing potential actions that could stop, remove, or slow down the failure rate of the component. Gao and Liu (2021) refer to these techniques as resilient control strategies and propose the use of RUL estimation in control systems to reduce the rate of crack growth in wind turbine blades.

5 Conclusion

In this chapter, we presented a review of hybrid methods using a condition monitoring framework. Hybrid methods are identified at the different steps of the frameworks. It is emphasised that the purpose of hybrid methods is to improve the estimation of the health state and the RUL. Two potential gaps for future research are identified namely (i) the effect that hybrid-fusion-based diagnosis has on the RUL estimation of a component and (ii) suggesting actions based on the diagnosis and prognosis to maximise the RUL of the component.

References

An, D., Kim, N.H., Choi, J.H.: Practical options for selecting data-driven or physics-based prognostics algorithms with reviews. Reliab. Eng. Syst. Saf. **133**, 223–236 (2015). https://doi.org/10.1016/j.ress.2014.09.014

Bishop, C.M.: Pattern Recognition and Machine Learning. Springer, New York (2006)

Carden, E.P., Fanning, P.: Vibration based condition monitoring: a review. Struct. Health Monit. **3**(4), 355–377 (2004)

Chen, J., Yuan, S., Qiu, L., Wang, H., Yang, W.: Online prognosis of fatigue crack propagation based on Gaussian weight-mixture proposal particle filter. Ultrasonics **82**, 134–144 (2018). https://doi.org/10.1016/j.ultras.2017.07.016

Coppe, A., Haftka, R.T., Kim, N.H., Yuan, F.G.: Uncertainty reduction of damage growth properties using structural health monitoring. J. Aircr. **47**(6), 2030–2038 (2010). https://doi.org/10.2514/1.C000279

Coppe, A., Pais, M.J., Haftka, R.T., Kim, N.H.: Using a simple crack growth model in predicting remaining useful life. J. Aircr. **49**(6), 1965–1973 (2012). https://doi.org/10.2514/1.C031808

Corbetta, M., Saxena, A., Giglio, M., Goebel, K.: An investigation of strain energy release rate models for real-time prognosis of fiber-reinforced laminates. Compos. Struct. **165**, 99–114 (2017). https://doi.org/10.1016/j.compstruct.2017.01.002

Corbetta, M., Sbarufatti, C., Giglio, M., Saxena, A., Goebel, K.: A Bayesian framework for fatigue life prediction of composite laminates under co-existing matrix cracks and delamination. Compos. Struct. **187**, 58–70 (2018a). https://doi.org/10.1016/j.compstruct.2017.12.035

Corbetta, M., Sbarufatti, C., Giglio, M., Todd, M.D.: Optimization of nonlinear, non-Gaussian Bayesian filtering for diagnosis and prognosis of monotonic degradation processes. Mech. Syst. Signal Process. **104**, 305–322 (2018b). https://doi.org/10.1016/j.ymssp.2017.11.012

Corrado, N., Durrande, N., Gherlone, M., Hensman, J., Mattone, M., Surace, C.: Single and multiple crack localization in beam-like structures using a Gaussian process regression approach. JVC/J. Vibr. Control **24**(18), 4160–4175 (2018). https://doi.org/10.1177/1077546317721418

Cubillo, A., Perinpanayagam, S., Esperon-Miguez, M.: A review of physics-based models in prognostics: application to gears and bearings of rotating machinery. Adv. Mech. Eng. **8**(8), 1–21 (2016). https://doi.org/10.1177/1687814016664660

Elshamy, M., Crosby, W.A., Elhadary, M.: Crack detection of cantilever beam by natural frequency tracking using experimental and finite element analysis. Alex. Eng. J. **57**(4), 3755–3766 (2018). https://doi.org/10.1016/j.aej.2018.10.002

Gao, Z., Liu, X.: An overview on fault diagnosis, prognosis and resilient control for wind turbine systems. Processes **9**(2), 1–19 (2021). https://doi.org/10.3390/pr9020300

ISO 13381-1. Condition monitoring and diagnostics of machines — prognostics — Part 1: General Guidelines (2015). https://www.iso.org/standard/51436.html

Jia, F., Lei, Y., Lin, J., Zhou, X., Lu, N.: Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech. Syst. Signal Process. **72–73**, 303–315 (2016). https://doi.org/10.1016/j.ymssp.2015.10.025

Jouin, M., Gouriveau, R., Hissel, D., Péra, M.C., Zerhouni, N.: Particle filter-based prognostics: review, discussion and perspectives. Mech. Syst. Signal Process. **72**, 2–31 (2016). https://doi.org/10.1016/j.ymssp.2015.11.008

Kaloop, M.R., Hu, J.W.: Stayed-cable bridge damage detection and localization based on accelerometer health monitoring measurements. Shock. Vib. (2015). https://doi.org/10.1155/2015/102680

Keprate, A., Ratnayake, R.M.C., Sankararaman, S.: Adaptive Gaussian process regression as an alternative to FEM for prediction of stress intensity factor to assess fatigue degradation in offshore pipeline. Int. J. Press. Vessels Pip. **153**, 45–58 (2017). https://doi.org/10.1016/j.ijpvp.2017.05.010

Khan, S., Yairi, T.: A review on the application of deep learning in system health management. Mech. Syst. Signal Process. **107**, 241–265 (2018). https://doi.org/10.1016/j.ymssp.2017.11.024

Le, T.T., Chatelain, F., Bérenguer, C.: Hidden Markov models for diagnostics and prognostics of systems under multiple deterioration modes. In: Safety and Reliability: Methodology Applications - Proceedings of the European Safety and Reliability Conference, ESREL 2014, pp. 1197–1204 (2015)

Lee, G.-Y., et al.: Machine health management in smart factory: a review. J. Mech. Sci. Technol. **32**(3), 987–1009 (2018). https://doi.org/10.1007/s12206-018-0201-1

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., Lin, J.: Machinery health prognostics: a systematic review from data acquisition to RUL prediction. Mech. Syst. Signal Process. **104**, 799–834 (2018). https://doi.org/10.1016/j.ymssp.2017.11.016

Liao, L., Köttig, F.: Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Trans. Reliab. **63**(1), 191–207 (2014). https://doi.org/10.1109/TR.2014.2299152

Saidi, L., Ben Ali, J., Benbouzid, M., Bechhofer, E.: An integrated wind turbine failures prognostic approach implementing Kalman smoother with confidence bounds. Appl. Acoust. **138**(April), 199–208 (2018). https://doi.org/10.1016/j.apacoust.2018.04.005

Sanchez, H., Sankararaman, S., Escobet, T., Frost, S., Goebel, K.: Analysis of two modeling approaches for fatigue estimation and remaining useful life predictions of wind turbine blades. In: European Conference of The Prognostics and Health Management Society, pp. 1–11 (2016)

Sankararaman, S., Goebel, K.: Uncertainty in prognostics and health management: an overview. In: European Conference of the Prognostics and Health Management Society, pp. 1–11 (2014)

Saxena, A.: Conference slides: prognostics tutorial. In: Annual Conference of the PHM Society (PHM2010) (2010)

Saxena, A., Celaya, J., Saha, B., Saha, S., Goebel, K.: Metrics for offline evaluation of prognostic performance. Int. J. Prognostics Health Manage. **1**(1), 1–20 (2010)

Serafini, J., Bernardini, G., Porcelli, R., Masarati, P.: In-flight health monitoring of helicopter blades via differential analysis. Aerosp. Sci. Technol. **88**, 436–443 (2019). https://doi.org/10.1016/j.ast.2019.03.039

Sikorska, J.Z., Hodkiewicz, M., Ma, L.: Prognostic modelling options for remaining useful life estimation by industry. Mech. Syst. Signal Process. **25**(5), 1803–1836 (2011). https://doi.org/10.1016/j.ymssp.2010.11.018

Valeti, B., Pakzad, S.N.: Remaining useful life estimation of wind turbine blades under variable wind speed conditions using particle filters, pp. 1–10 (2017)

Wang, B., Member, S., Lei, Y., Li, N., Li, N.: A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. IEEE Trans. Reliab. **69**(1), 401–412 (2020)

Wang, Y., Xiang, J., Markert, R., Liang, M.: Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: a review with applications. Mech. Syst. Signal Process. **66–67**, 679–698 (2016). https://doi.org/10.1016/j.ymssp.2015.04.039

Zaidan, M.A., Harrison, R.F., Mills, A.R., Fleming, P.J.: Bayesian hierarchical models for aerospace gas turbine engine prognostics. Expert Syst. Appl. **42**(1), 539–553 (2015). https://doi.org/10.1016/j.eswa.2014.08.007

Zeng, J., Chen, K., Ma, H., Duan, T., Wen, B.: Vibration response analysis of a cracked rotating compressor blade during run-up process. Mech. Syst. Signal Process. **118**, 568–583 (2018). https://doi.org/10.1016/j.ymssp.2018.09.008

Zhang, W., Li, C., Peng, G., Chen, Y., Zhang, Z.: A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. Mech. Syst. Signal Process. **100**, 439–453 (2018). https://doi.org/10.1016/j.ymssp.2017.06.022

Time Domain Identification of Multi-stage Planetary Gearbox Characteristic Frequencies Using Piezoelectric Strain Sensor

O. Graja[1(✉)], K. Dziedziech[2], A. Jablonski[2], A. Ghorbel[1], F. Chaari[1],
T. Barszcz[2], and M. Haddar[1]

[1] Laboratory of Mechanics Modelling and Production,
National School of Engineers of Sfax, BP 1173, 3038 Sfax, Tunisia
[2] AGH University of Science and Technology, Krakow, Poland

Abstract. Nowadays, condition monitoring of gearboxes is based on analyzing vibration signal (mostly acceleration, velocity; fewer strain) measured from mechanical components or on analyzing acoustic emission and temperature data. Strain sensors are rarely used because of their high cost, mounting difficulties and their sensitivity harsh environment. They are based on strain gage technology which implements 'Wheatstone Bridge Conditioner' to be able to measure strain values. This bridge circuit, unfortunately, becomes non-linear out of the measurement range resulting in significant errors. To avoid those errors, a new technology of piezoelectric strain sensors, relatively new to condition monitoring, emerges: they have good precision since the accuracy is improved and provide good results especially when attempting to do measurements at extremes of the range. In this work, a piezoelectric strain sensor mounted on an industrial planetary gearbox (PG) in vertical direction is used to acquire time domain signals. This latter is introduced to focus on the importance of the new time domain representation of a gearbox dynamic behavior by a such sensor. Finally, an analysis of the RHM sensor time domain data is conducted to identify the PG characteristic frequencies.

Keywords: Planetary gearbox · Vibration signal · Piezoelectric · Strain

1 Introduction

Planetary gearbox are widely used either in automobiles and in industrial machinery in order to transmit the power from input shaft to output shaft with a high ratio in a compact space. Many researchers tried to characterize its dynamic behavior. It was a hard task because of the complexity of its design and motions of its parts, especially the motion of planets. They used sensors and microphones which acquire signal (vibration or acoustic) to be processed in order, for instance, to cancel the noise (masking signal) which makes analysis of the signal easier.

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 116–122, 2022.
https://doi.org/10.1007/978-3-030-85584-0_12

Noll et al. [1] applied a piezoelectric strain sensor (SG) on the external housing of the fixed ring gear. They experienced a piezoelectric sensor in AC- and DC-coupling. Results show that strain signals can be obtained even down to 0.002 Hz (low spin). The work of S. Kiddy et al. [2] shows that the surface strain is collected by the use of Fiber-Bragg-sensor mounted on the ring gear of a bell helicopter OH-58C gearbox. They compared signals for healthy and damaged cases and it was demonstrated that strain signal are useful to detect damages. Due to the proper reproducibility of the signal, there is no need to average the signal over multiple revolutions as done for signals issued from accelerometers. Zhang et al. [4] did an excessive use of strain gauges. They mounted 6 strain sensors along a shaft system test rig to identify the load on each bearing supporting the shaft. They developed a strain gauge method which can reduce the modeling error taking into account the best location of sensors where the strain is sensitive to the change of the bearing load. Oskoueian et al. [5] monitor a Mitsubishi 4 cylinders internal combustion engine block by taking measurement with strain sensor. After processing signals captured by strain sensors, they were able to localize in which cylinder the mechanical wear is. In another research focused on gear root bending stress analysis, Lisle et al. [6] used strain gauges method to validate a numerical finite element analysis (ANSYS) highlighted the advantageous of this method and to compare it with ISO 6336:2006 and AGMA 2101-D04. Yoon et al. [7] proposed a new methodology for the diagnosis of planetary gearbox faults. Considering that the amplitude modulation is less effecting strain signals, the proposed method is based essentially on processing strain signals. So, they used a piezoelectric strain sensor mounted on the external surface of a planetary gearbox ring gear. Results show that all faults made in the gearbox were identified and they were fruitful compared with whom conducted from acceleration analysis. As mentioned in literature, strain sensors (classical and piezoelectric) are used for many purposes, for instance to compare different signals acquired from the same sensor with different coupling [1], identification of health condition and fault diagnosis as investigated in [3] and [7]. This work is focused on identification of vibration sources in a multi-stage planetary gearbox in healthy case using a piezoelectric strain sensor by analyzing only the time data. This paper is organised as follows. Section 2 is dedicated to present the experimental setup. Section 3 analyses time domain signal issued from the piezoelectric strain sensor. Conclusion is drawn at the end of the paper.

2 Experimental Setup

Figure 1 presents the Test rig used to investigate the dynamic behavior of a three stages planetary gearbox (PG) given by its kinematic scheme in Fig. 2. The motor is operating at 1008 rpm. Passing through parallel gearbox (ratio = 1/2.91), the speed is reduced to 346,4 rpm and it is reduced again passing through the planetary gearbox to achieve 1,73 rpm at its output.

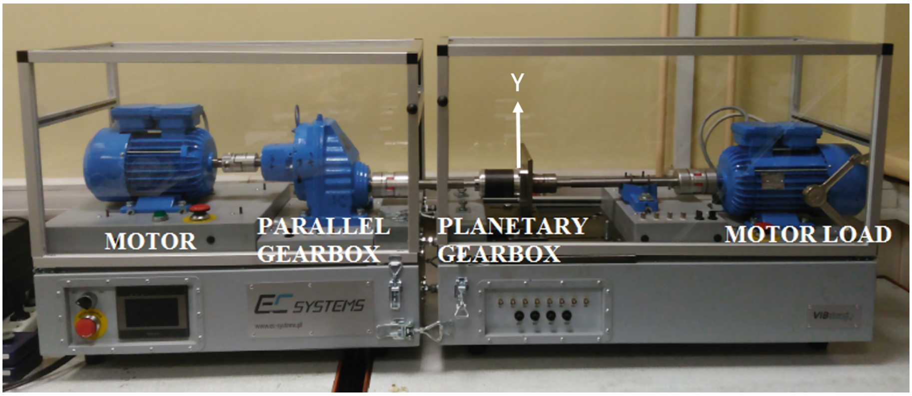


Fig. 1. Test rig

A quartz strain gage RHM-240A02 is attached to the PG to measure strain on Y direction as indicated in Fig. 1. It incorporates a built-in MOSFET microelectronic amplifier. This serves to convert the high impedance charge output into a low impedance voltage signal for analysis or recording. ICP quartz strain sensors, powered from a separate constant current source, operate over long ordinary coaxial or ribbon cable without signal degradation. The low impedance voltage signal is not affected by tribo-electric cable noise or environmental contaminants.

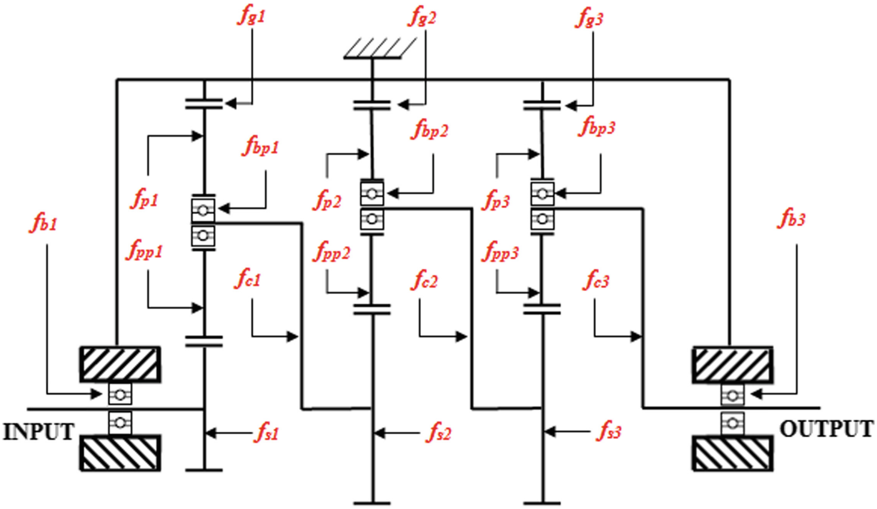


Fig. 2. Kinematic scheme of three stages planetary gearbox

Since the experimental conditions are given, the PG characteristic frequencies can be calculated. Those latter are given in Table 1. For more information about experimental setup, the reader can refer to the work of Zghal et al. [8].

Table 1. Characteristic frequencies of the three stages planetary gearbox when the motor was operating at 1008 rpm, $^{(*)}$:Parallel Gearbox

Fr(Hz)	Description	1^{st} stage	2^{ed} stage	3^{ed} stage
f_{si}	Sun gear frequency	5.77	0.72	0.15
f_{ci}	Carrier frequency	0.7223	0.15	0.03
f_{pi}	Planet frequency	1.68	0.39	0.08
f_{ppi}	Planet pass frequency	2.16	0.45	0.12
f_{gi}	Gear mesh frequency	60.67	12.13	2.43
f_{b1}	Sun bearing frequency	5.77	–	–
f_{bpi}	Planet bearing frequency	0.96	0.25	0.05
f_{b3}	Carrier bearing frequency	–	–	0.03
f_{in}	Pa.G$^{(*)}$ input frequency	16.78		
f_{out}	Pa.G output frequency	5.77		
f_{gpg}	Pa.G gearmesh frequency	386.4		

3 Time Domain Identification

In the previous section, all PG characteristic frequencies are calculated theoretically by using equations based on its kinematics. In this section, data collected from the Piezoelectric strain sensor RHM240A02 is presented. The chosen time signal is analyzed in order to identify the calculated frequencies in the previous section. This step will allow us checking the efficiency and robustness of the use of the strain sensor in describing PG vibration sources. The processing is simple and described as the following: we present the row signal than we do multiple zoom section to identify repetitive behavior. Figures 3 and 4 present all repetitive oscillations in the acquired strain signal.

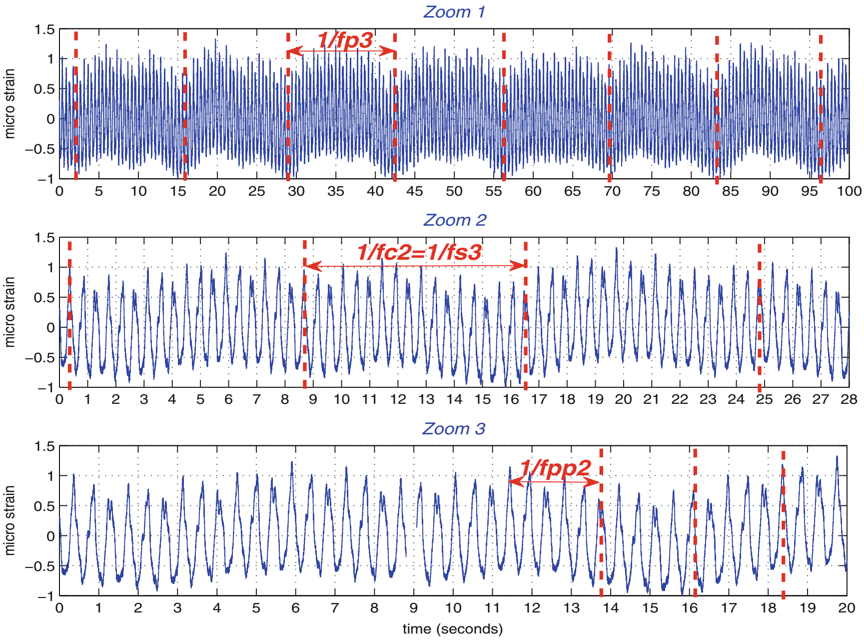


Fig. 3. First three zoom sections

Identified frequencies are summarized in Table 2.

The checked values are those identified previously. As shown, almost 40% of characteristic frequencies are identified from only the time domain representation. Hence, this piezoelectric strain sensor shows 40% of the planetary gearbox dynamic behaviour since the collected time signal reproduces with a good accuracy what was really happening inside the gearbox.

Table 2. Summary of the identified frequencies

Frequencies(Hz)	Description	1^{st} stage	2^{ed} stage	3^{ed} stage
f_{si}	Sun gear frequency	5.77	0.72 ✓	0.15 ✓
f_{ci}	Carrier frequency	0.72 ✓	0.15 ✓	0.03
f_{pi}	Planet frequency	1.68	0.39	0.08 ✓
f_{ppi}	Planet pass frequency	2.16	0.45 ✓	0.12
f_{gi}	Gear mesh frequency	60.64 ✓	12.13 ✓	2.43
f_{b1}	Sun bearing frequency	5.77	–	–
f_{bpi}	Planet bearing frequency	0.96	0.25	0.05
f_{b3}	Carrier bearing frequency	–	–	0.03

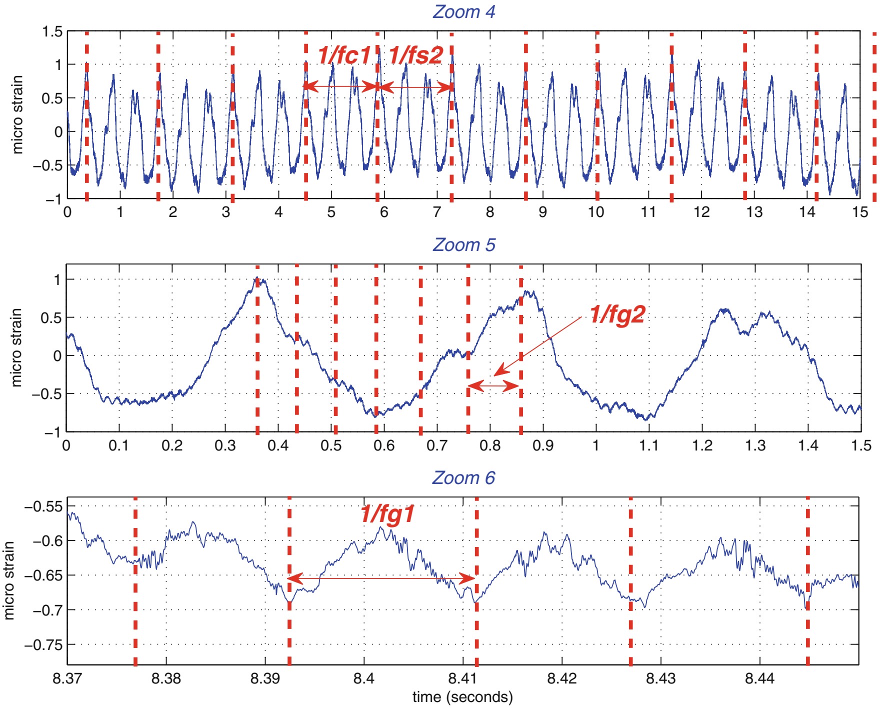


Fig. 4. Second three zoom sections

4 Conclusion

In this work, experimental measurement was done on a real industrial gearbox which is mounted on test rig. Strains of the housing surface of the ring gear were measured to identify characteristic frequencies of the PG in order to investigate its dynamic behaviour. The time signal collected using the piezoelectric strain sensor has a high SNR and an excellent reproducibility of periodic component of the signal especially in lower frequencies. In the way of future research, a supplementary analysis will be focused on the detection of gear faults using time signal acquired by using a strain gauge in order to investigate the capability of this sensor in damage cases.

Acknowledgement. This work is partially supported by NATIONAL SCHOOL OF ENGINEERS OF SFAX (ENIS)/Laboratory of Mechanics, Modelling and Production (LA2MP) and The National Centre of Research and Development (NCRD) in Poland under the research project no. PBS3/B6/21/2015.

References

1. Noll, M.-C., Godfrey, J.W., Schelenz, R., Jacobs, G.: Analysis of time-domain signals of piezoelectric strain sensors on slow spinning planetary gearboxes. Mech. Syst. Sig. Process. **72–73**, 727–744 (2016)
2. Kiddy, J.S., Samuel, P.D., Lewicki, D.G., LaBerge, K.E., Ehinger, R.T., Fetty, J.: Fiber Optic Strain Sensor for Planetary Gear Diagnostics, pp. 1–14. NASA, Cleveland (2011)
3. Lei, Y., Liu, Z., Wu, X., Li, N., Chen, W., Lin, J.: Health condition identification of multi-stage planetary gearboxes using a mRVM-based method. Mech. Syst. Sig. Process. **60–61**, 289–300 (2015)
4. Zhang, S.D., Yang, J.G., Li, Y.Y., Li, J.: Identification of bearing load by three section strain gauge method: theoretical and experimental research. Measurement **46**, 3968–3975 (2013)
5. Oskoueian, A., Nuawi, M.Z.: Internal combustion monitoring using strain gauge and analyzing with I-Kaz. In: 2012 2nd International Conference on Mechanical, Industrial and Manufacturing Engineering (2012)
6. Lisle, T.J., Shaw, B.A., Frazer, R.C.: External spur gear root bending stress: a comparison of ISO 6336:2006, AGMA 2101–D04, ANSYS finite element analysis and strain gauge techniques. Mech. Mach. Theor. **111**, 1–9 (2017)
7. Yoon, J., He, D., Van Hecke, B., Nostrand, T.J., Zhu, J., Bechhoefer, E.: Planetary gearbox fault diagnosis using a single piezoelectric strain sensor. In: Annual Conference of the Prognostics and Health Management Society (2014)
8. Zghal, B., Graja, O., Dziedziech, K., Chaari, F., Jablonski, A., Barszcz, T., Haddar, M.: A new modeling of planetary gear set to predict modulation phenomenon. Mech. Syst. Sig. Proces. **127**, 234–261 (2019)

Study of the Stiffness of a Polymer Pinion Tooth in a Polymer-Metallic Spur Gear System

Ala Eddin Chakroun[1,2](✉), Chaima Hammami[1], Ahmed Hammami[1], Ana De-Juan[1], Fakher Chaari[1], Alfonso Fernandez[2], Fernando Viadero[2], and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modelling and Production (LA2MP), National School of Engineers of Sfax, 1173, 3038 Sfax, Tunisia
`mohamed.haddar@enis.rnu.tn`
[2] Department of Structural and Mechanical Engineering, Faculty of Industrial and Telecommunications Engineering, University of Cantabria, Avda de Los Castros S/N, 39005 Santander, Spain
`{ana.dejuan,alfonso.fernandez,fernando.viadero}@unican.es`

Abstract. Over the years, the gear mesh stiffness is being studied extensively by researchers due to its ability to give a better understanding of a gear box dynamic behaviour. Many studies have used various methods to model the mesh stiffness of metallic spur gears such as square wave signal and finite element modelling. These models are based on combining the mesh stiffness pattern of each tooth involved in the gear meshing. This has proven to be effective in case of metallic gears. However, due to the viscoelastic behaviour of polymer, it is no longer possible to model the contact stiffness of a polymer pinion paired with a metal gear as a square-wave signal. In this study, the viscoelastic behaviour is taken into consideration. A Rheological model based on parallel Maxwell cells, called generalized Maxwell model, is used to model the contact behaviour of a polymer gear tooth. This model is ideal for steady-state harmonic excitation, which is the case of spur gears. The proposed approach is applied in the case of a pure nylon 6,6 gear paired with a metal pinion. It is then compared to the contact behaviour of a metallic gear tooth also paired with a metallic pinion.

Keywords: Gear · Stiffness · Polymer · Generalized Maxwell Model · Viscoelastic behaviour

1 Introduction

The modelling of the dynamic behaviour of gear pairs depends on many parameters. Hence the rigidity of the material used to manufacture the gears. In general, steel is considered to be a rigid body. For steel spur gears, square wave signals (Farhat et al. 2020) and finite element models (Rincon et al. 2013) are used to model the gear mesh stiffness. It is also possible to model the stiffness with a linear spring. This spring is a direct application of Hook's law. The materials modelled by this approach are called Hookean materials.

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 123–130, 2022.
https://doi.org/10.1007/978-3-030-85584-0_13

It is not possible to model non-Hookean materials with Hook's law. For this study, it is proposed to use polymer as pinion material and steel as gear material. The polymer is considered to be viscoelastic. Its viscoelastic behaviour can be modelled using rheological models. These models are considered as a combination of springs and dashpots connected in various configurations.

The Maxwell model is a combination of a spring and a dashpot in series. In this model, the stress applied to the elements is the same on each element but the strain is equal to the sum of the strains applied on the two elements. In this model, the stress tends to infinity when applying a constant strain. Also, the damper cannot retract itself after releasing the stress. For this reason, Maxwell's model can give a basic modelling of viscoelasticity but cannot model a real viscoelastic behaviour (Guedes 2019). The Kelvin-Voight model is a combination of a spring and a dashpot in parallel. In this model, the applied stress is the sum of the stresses in each element while the strain is the same for each element. In this model a stepwise strain excitation leads to infinite stress (Guedes 2019). Zener or standard solid model is a combination between Kelvin-Voight and Maxwell models. It is a spring connected in parallel with a Maxwell element. The relaxation is described as exponential behaviour which is not the case in experimental measurements. Also, the exponential results conduct rapid variations that cannot be observed experimentally.

Generalized models are also used. The best known are Generalized Maxwell Model (GMM) and generalized Kelvin model (GKM). GMM is a combination of Maxwell cells connected in parallel. In the study of Jrad et al. (2013), it is mentioned that depending on amplitude, frequency and preload, this model provides an accurate description and a good knowledge of the dynamic behaviour of viscoelastic components. It also can characterize stiffness and damping during steady-state harmonic excitation. Jrad et al. (2017) mentioned that for GMM, it is impossible to identify the creep when applying constant stress. GKM is a combination of Kelvin cells connected in series. It is mentioned that only GKM explains the results of tests under forced vibration. The latter statement is mentioned in the study of Hiltcher et al. (2006). The choice of the best model depends on the system to be modelled. In the case of gear teeth, for one cycle, a tooth enters loading phase for a meshing period. Then it has a second phase of no applied load. This harmonic excitation is what makes GMM the best choice to model the viscoelastic behaviour of the pinion teeth made with polymer.

In this study, GMM is used to model the viscoelastic behaviour of pure nylon 6,6. This polymer is chosen for a spur pinion coupled to a spur metal gear. It is proposed to study the behaviour of a tooth of the pinion for a cycle of engagement and disengagement with the gear. The stiffness of the polymer is negligible compared to that of the metal. Thus, it is considered that all deformations occur in the pinion. Strain and stiffness are determined from a constant and variable stress applied to the pinion tooth. For future work, gear mesh stiffness will be deduced from this work with performing an overlay on the stiffness of all the teeth of the pinion.

2 Generalized Maxwell Model (GMM)

Rheological models are made basically with linear springs and linear dashpots. The spring is used to model elastic behaviour. Its response to an instantaneous stress σ is an instantaneous strain ε. The relation is given in Eq. (1) where E is the Young modulus.

$$\sigma = \varepsilon E \tag{1}$$

The dashpot is used to model viscous behaviour. The relation between stress σ and strain ε is given in Eq. (2) where η is the coefficient of viscosity.

$$\sigma = \eta \frac{d\varepsilon}{dt} \tag{2}$$

Maxwell model is a combination of a spring and a dashpot connected in series. It is presented in Fig. 1. In this study, it is aimed to obtain the creep-recovery behaviour of the material. So, this response can be obtained by the following Eq. (3).

$$\varepsilon(t) = \frac{\sigma_0}{E} + \frac{\sigma_0}{\eta} t \tag{3}$$

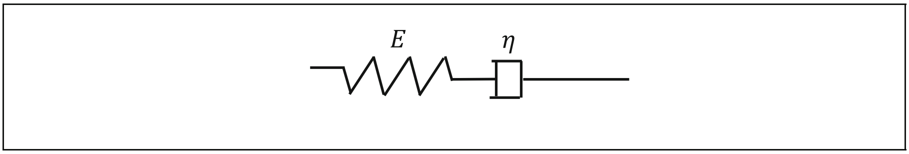


Fig. 1. Maxwell cell

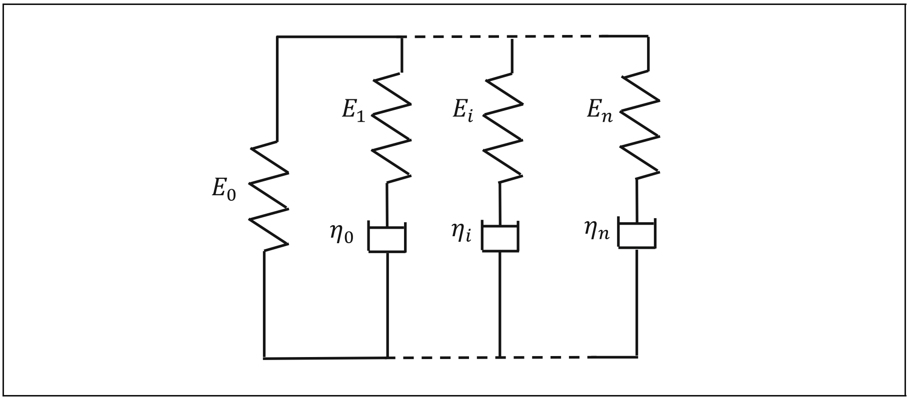


Fig. 2. Generalized Maxwell model.

The unlimited deformation resulted with Maxwell model under a constant stress over time do not give the real behaviour of a viscoelastic material. Moreover, the behaviour after releasing the stress cannot be modelled using this model. That is due to the fact that the spring can retract and return to its initial state. But there is no available load to retract the dashpot. In order to ameliorate these results, GMM are used instead of Maxwell model.

It consists of multiple Maxwell cells connected in parallel. A single spring can also be added in parallel to make it possible for the dashpots to retract after releasing the stress. The rheological presentation is given in Fig. 2 where $i \in [1n]$ and n is the number of Maxwell cells. The use of multiple springs and dashpots can give more accurate viscoelastic results. Yet, the complexity of the model is related to the number of elements used. The more elements used the more model parameters to identify.

The rheological formulation of the impedance of GMM is given in Eq. (4).

$$Z(\omega) = E_0 + \sum_{i=1}^{n} \frac{j\omega E_i \eta_i}{E_i + j\omega\eta_i} \tag{4}$$

Where ω is the angular frequency. E_0 is the elastic modulus at $\omega = 0$. E_i and η_i are the $\mathrm{i^{th}}$ elastic modulus and the $\mathrm{i^{th}}$ viscous modulus, respectively. The viscoelastic behaviour is then modelled using this equation. Multiple results can be found such as elasticity, viscous flow, creep with various retardation times and relaxation with various relaxation times (Guedes 2019).

3 Application of the GMM to Model the Polymer Gear Viscoelastic Behaviour

GMM is used in this study to model the stiffness of a tooth of a polymer pinion paired with a metal gear. The metal gear is considered as non-deformable solid. That was considered based on the fact that polymer will receive all the deformation because of its low stiffness in comparison with metal.

In the numerical simulation, Pole-Zero identification method was used to identify GMM parameters. It is an asymptote-based optimization algorithm useful to identify the parameters of transfer functions such as the Eq. (4) of GMM. The identification method is detailed in the study of Jrad et al. (2013). The mentioned study used a non-linear GMM whereas in this study linear model is used. The choice is based on the material used. In the study of Jrad et al., the non-linear spring is used to model the nonlinear dynamic behaviour of rubber. This is not the case in this study which considers polymer as material of the pinion. The only difference is the parallel spring E_0. It is considered non-linear in the study of Jrad et al. whereas in this study it is considered constant. Figure 3 shows the steps followed in order to identify the parameters of GMM. Initial values such as the elastic modulus E_0 at $\omega = 0$ are identified from a dynamic mechanical analysis (DMA) test. Using Pole-Zero equation depends on poles and zeroes.

The chosen polymer is pure nylon 6,6. The identification of the parameters of the GMM is based on a DMA of the said polymer. The results of the DMA test are extracted from the study of Almagableh et al. (2008). The numerical GMM showed a good corrolation with the results of the DMA test of the study of Almagableh et al.

After the determination of the impedance Z, its real part is used to model the creep while its imaginary part is used to model the recovery. The applied stress variation is determined from the load-sharing factor proposed in the study of (Raghuwanshi and Parey 2016). The load-sharing factor used in this study is shown in Fig. 4. The strain is then determined from Eq. (1) considering two results. The creep and the recovery are determined from the real part and the imaginary part of the impedance Z respectively.

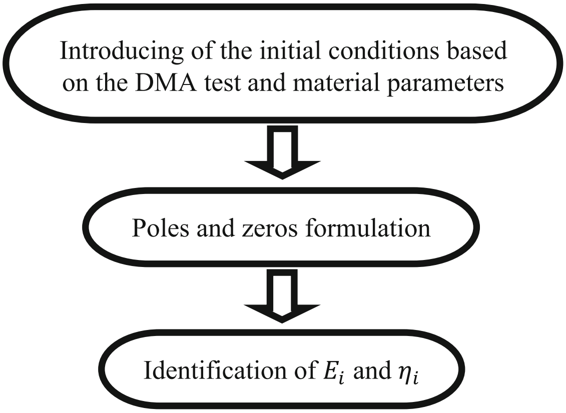


Fig. 3. Steps of the identification of the GMM parameters.

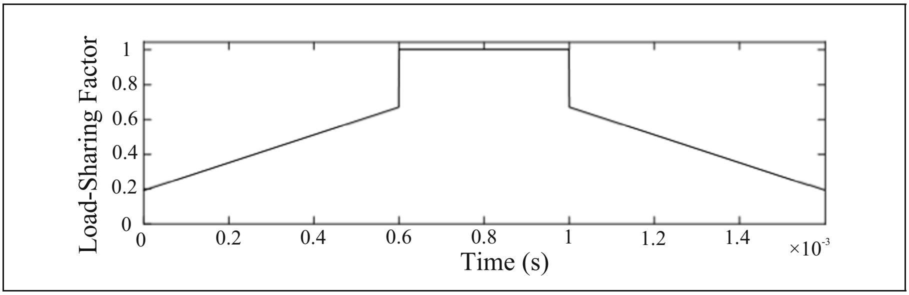


Fig. 4. Load-sharing factor of one tooth

4 Numerical Results

The numerical parameters of the studied gear pair are given in Table 1. The behaviour presented in Fig. 5 shows the creep-recovery of the tooth for a constant stress. It behaves differently in compared to metallic tooth behaviour that can be modelled as a square shape.

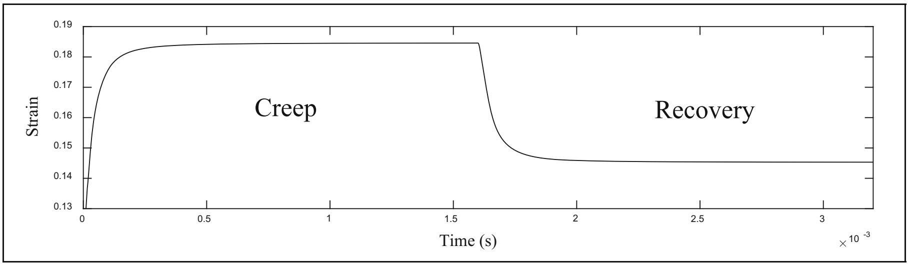


Fig. 5. Creep-recovery of the pinion tooth for a meshing period

Table 1. Parameters of the spur gear pair.

	Pinion	Gear
Teeth numbers	20	30
Material	Pure nylon 6,6	Steel
Base circle (mm)	18.8	28.2
Torque (N m)	1000	−1500
Rotation speed (rpm)	3000	2000
Module (mm)	2	
Pressure angle	20°	
Teeth width (mm)	23	
Contact ratio	$c = 1.6$	

It takes the tooth from the engagement to the disengagement a period of cT_e as shown in Fig. 6. The mesh period T_e is divided in two periods. The first period is equal to $(c - 1)T_e$ when two pair of teeth are engaged, where $c = 1.6$ is the contact ratio. The second period is when only one pair of teeth is engaged. It is equal to $(2 - c)T_e$. These details are mentioned in the study of Chaari et al. (2008). The shape of stiffness of one tooth shown in Fig. 6 is used to model steel spur gear pairs. In this study, it is modelled using GMM. The result of GMM given by Eq. (4) is used to measure the GMS. This measurement relates the material properties with the geometry of the teeth. The

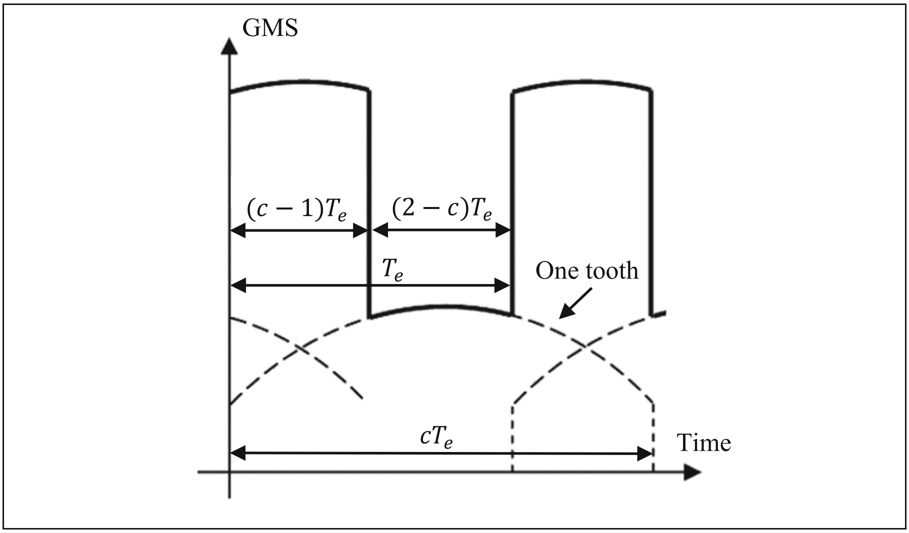


Fig. 6. Time varying gear mesh stiffness computation.

deflection occurs mainly due to the characteristics of the polymer. For this fact, it is the only deflection considered.

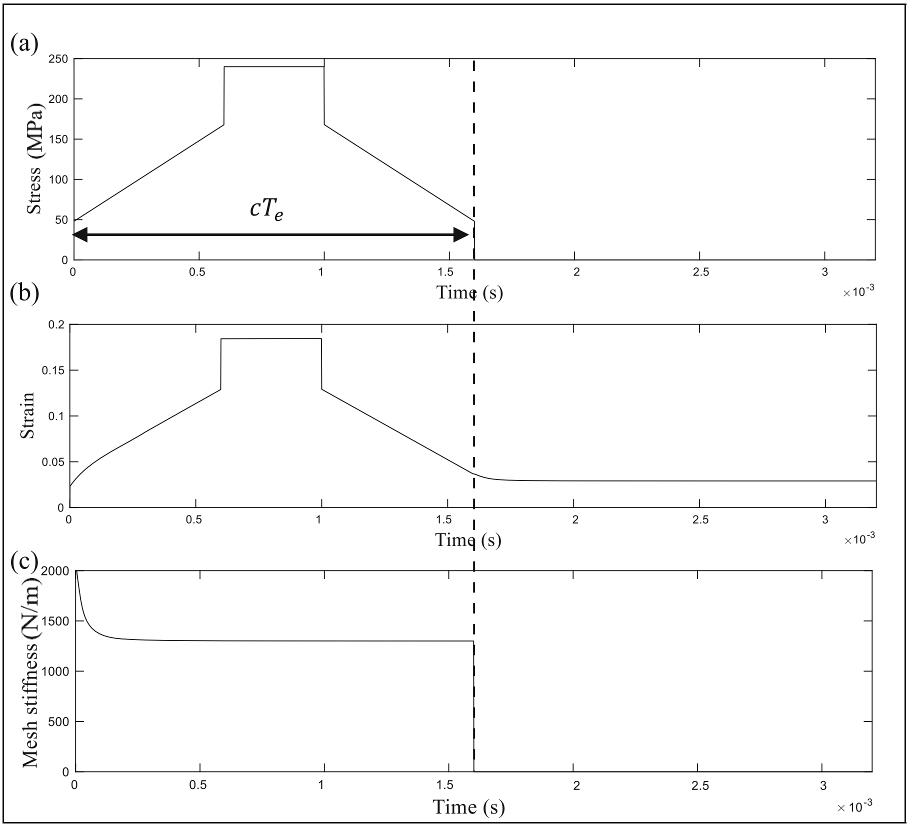


Fig. 7. (a) Stress, (b) strain and (c) stiffness of a pair of teeth in a meshing period with a variable stress.

To add more credibility the result, a variable stress applied to a tooth of a gear system that have parabolic shape as shown in Fig. 7(a). This variation is the result of having the stress shared by the teeth engaged in the meshing. This variation goes from the maximum value of the stress applied by the system on one tooth to the half of this value that is shared by a pair of teeth. The viscoelastic behaviour of the tooth shown in Fig. 7(b) is totally different from the one shown in Fig. 6. The stiffness shown in Fig. 7(c) is also different in comparison of metallic tooth that is shown in Fig. 6. It showed a delay in the beginning of the contact between teeth. This delay is the result of the viscous character of the polymer. This difference can have a big influence on the dynamic behaviour of the system.

This new approach can give a totally different understanding to the behaviour of polymer gears. Also, an overlay can be executed in order to deduce the gear mesh stiffness. This is the aim of a future work using the approach proposed in this study.

5 Conclusion

Polymer gears have a totally different behaviour in compared with metallic gears. The GMM is used and gives a realistic behaviour of the proposed polymer based on the DMA test carried out by the study of Almagableh et al. (2008). The proposed approach of modelling viscoelastic behaviour of nylon 6,6 with GMM gave a new understanding to dynamic behaviour of a polymer pinion tooth paired with a metallic gear tooth. This new understanding can lead to better investigations results such as gear mesh stiffness. It is known that investigating gear mesh stiffness can detect gear defects without the need to turn off the system. In future works, it is proposed to deduce the gear mesh stiffness of a nylon 6,6 paired with a metallic gear and compare it with metallic gear pair.

Acknowledgements. This work has been supported by project DPI2017-85390-P funded by the Spanish Ministry of Economy, Industry and Competitiveness.

References

1. Almagableh, A., Gupta, S., Mantena, P.R., Al-Ostaz, A.: Dynamic mechanical analysis of graphite platelet and nanoclay reinforced vinyl ester, and MWCNT Reinforced Nylon 6,6 Nanocomposites. In: International SAMPE Technical Conference (2008)
2. Chaari, F., Baccar, W., Abbes, M.S., Haddar, M.: Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. Eur. J. Mech. A/Solids **27**(4), 691–705 (2008)
3. Farhat, M.H., et al.: Numerical model of a single stage gearbox under variable regime. Mech. Des. Struct. Mach. (2020). https://doi.org/10.1080/15397734.2020.1863226
4. Fernandez Del Rincon, A.A., Del Rincon, F., et al.: A model for the study of meshing stiffness in spur gear transmissions. Mech. Mach. Theory **61**, 30–58 (2013). https://doi.org/10.1016/j.mechmachtheory.2012.10.008
5. Guedes, R.M.: Creep and Fatigue in Polymer Matrix Composites. Elsevier, Amsterdam (2019)
6. Hiltcher, Y., Guingand, M., de Vaujany, J.-P.: Load sharing of worm gear with a plastic wheel. J. Mech. Des. **129**(1), 23 (2006)
7. Jrad, H., et al.: Experimental characterization, modeling and parametric identification of the non linear dynamic behavior of viscoelastic components. Eur. J. Mech. A/Solids **42**, 176–187 (2013)
8. Jrad, H., et al.: Experimental and numerical investigation of energy dissipation in elastomeric rotational joint under harmonic loading . Mech. Time-Depend. Mater. **21**(2), 177–198 (2017). https://doi.org/10.1007/s11043-016-9325-9
9. Raghuwanshi, N.K., Parey, A.: Experimental measurement of gear mesh stiffness of cracked spur gear by strain gauge technique. Meas. J. Int. Meas. Confederation **86**, 266–275 (2016). https://doi.org/10.1016/j.measurement.2016.03.001

Generalized Prototype Bootstrapping for Nonlinear System Identification in an Accelerated Fatigue Testing Context

J. Crous(✉), S. Kok, D. N. Wilke, and P. S. Heyns

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria 0001, South Africa

Abstract. The Prototype Bootstrapping Method (PBM) is a sampling technique for nonlinear system identification specifically when applied to Accelerated Fatigue Testing (AFT). This application requires the number of samples taken to be limited as much as possible to preserve the integrity of the test. The work presents a generalization of this technique by removing system specific requirements from the algorithm. The generalized method is developed and tested in this paper. Various numerical experiments are conducted. These experiments show that Generalized PBM (GPBM) outperforms other sampling techniques, being 5.4 times more accurate. Through these experiments, the importance of sampling in this application is emphasized.

Keywords: Nonlinear system identification · Bootstrapping · Accelerated Fatigue Testing setting · Statistical learning

1 Introduction

In this work, we consider mechanical systems under realistic loadings for a short duration. Typically, field tests are conducted in hours while a system is designed to last years. During field testing, measurements to identify high fatigue incidents are taken. Accelerated fatigue testing then attempts to recreate these high fatigue incidents in a laboratory environment. Servo hydraulic actuators are used to excite a laboratory equivalent of the mechanical system. The challenge is to reproduce the high fatigue incidents in the laboratory system. The challenge arises due to the large number of measurements required to model the system accurately (as a highly nonlinear system) under severe operating conditions. A large number of measurements of the system compromises the integrity of the fatigue test. The focus of this work is to develop a technique that identifies a small number of measurements with which the system is identified. The identified system is then used to infer what input is required to recreate the target response in the system.

Accelerated fatigue testing (AFT) is performed as part of the design phase to identify the system's degradation patterns when in harsh operating conditions. This includes determining the system's longevity when exposed to a high fatigue environment for prolonged periods on time [1]. To ensure that the system's fatigue matches that of an

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 131–143, 2022.
https://doi.org/10.1007/978-3-030-85584-0_14

equivalent field operated specimen, the response of the system has to be reproduced as accurately as possible [1–4].

Since the system is modelled under severe operating conditions, nonlinear system identification is required. The identification of nonlinear systems has been extensively studied over the past few decades (see [5] and references), with particular emphasis on predicting the behavior of the system [6]. This is usually done in conjunction with sophisticated control strategies that are formulated to compensate for model errors in the predicted model [7].

Although well-established [5], the field of nonlinear system identification has experienced a change in direction due to the growing success of machine learning techniques [8] applied to control and system identification. In recent years a strong line of research in this field has been to successfully incorporate machine learning strategies into existing system identification strategies [9–15]. Continuing along this line of research, the authors have developed an alternative approach to nonlinear system identification. The proposed approach follows a well-established line of thought found in machine learning research where an unstructured regularized high dimensional regression technique is employed to model complex data [4, 8]. The approach views essential events in the time series as high dimensional features. A high dimensional feature map is then constructed to map the excitation signal to the response signal. This approach has been shown to outperform current techniques when applied to highly nonlinear systems.

Furthermore, it has been shown that greater performance gains are possible if the data presented to the regression algorithm is sampled correctly [4]. To this end, this paper presents a generalized sampling method explicitly designed for the AFT context: samples are drawn such that events closely resemble the target response. The system's dynamics are identified only to the extent required to describe the system's behavior when operating under the condition that generated the target response.

2 Sampling

The sampling process, forming part of the design process, is vitally important to a successful experimental design. Latin Hypercube Sampling (LHS) has gained popularity in Design and Analysis of Computer Experiments (DACE), where evaluations of the function to be computed are computationally expensive [16]. LHS has been successfully implemented in various DACE problems [17]. LHS divides the sample space into disjoint subsets of equal probability. This results in a space-filling sampling strategy where samples are distributed uniformly in the sample space.

DACE applies LHS to the entire sample space. In the current work, nonlinear system identification for AFT leads to early saturation of the residual: more samples do not significantly improve the predictions. Instead, PBM shrinks down the sample space to identified regions of interest. These regions are then sampled using LHS.

3 System Identification

The authors have developed an alternative approach to nonlinear system identification in previous work [4, 15]. This approach, which uses Event-Based Mapping (EBM), follows a well-established line of thought in machine learning literature where complex data can be represented by regularized linear models in high dimensional space. An EBM approach views the time series as a sequence of events. Each event is considered as high dimensional features of independent variables. An event consists of two features: an input feature and the corresponding response feature. EBM constructs a mapping from the feature that excites the system, the input feature, the corresponding response, and the output feature. Causality can be modelled implicitly by using multivariate methods to construct the mapping [18].

This is different from a differential equation type approach where the correlation between time steps is modelled explicitly. NARX models, and all their variants, assume a structure for the correlation between time steps [5]. Most time series modelling, including that done with machine learning techniques, make this basic assumption [6, 8]. One shortcoming of this approach is that when highly nonlinear systems are modelled, complex representations of the systems are required [5, 6, 10, 13, 19, 20].

Multivariate Principal Component Regression (MPCR) was identified as an efficient algorithm to construct the EBM. MPCR traces the change in the basis functions from the input to the output space using a set of rotation and scaling matrices. For a brief description of MPCR,[1] we recommend [21], a more in-depth discussion on the algorithm is presented in [18].

4 Simulations

When a system operated in the field is brought into a laboratory environment, the system is fixed to servo-hydraulic actuators and excited. To do this, parts of the system are often removed. Furthermore, the direction of excitation is chosen to be the dominant one. This implies that the system in the laboratory environment is not the same as the field operated one.

A half car and quarter car model are used to simulate this process: A half-car model is used to represent the field operated system where the target response is generated. The simpler quarter car model to represent the laboratory system where the number of degrees of freedom in the system has been reduced.

The nonlinearities in the half and quarter car models arise for the cubic stiffening springs and the quadratic dampers. Further details of the quarter and half car models and the parameters used for these experiments can be found in [18]. The displacements reported in this work for the half car are the vertical displacement of the center of mass and the sprung mass for the quarter car model.

To ensure that the models are operating in regions where the nonlinearities will be significant, a road profile of class G is generated using the ISO 8608 standard (Mechanical Vibration - Road surface profiles - Reporting measured data)[2] [22].

---

[1] MPCR is used to construct the event mapping in this work, however, any suitable mapping algorithm can used.

[2] The ISO 8608 standard provides a means to classify road profiles according to different damage levels.

5 Prototype-Based Bootstrapping

In this section, the prototype-based bootstrapping method for sampling is discussed. This is a high dimensional sampling technique due to the nature of the system identification technique used in this work, as outlined in Sect. 3. The subsequent sections describe the various stages of the algorithm.

5.1 Generating the Initial Signal

The bootstrapping algorithm is initiated by using a low-frequency signal with gradually increasing amplitude. This is used to estimate the magnitude of the input signal required. Having identified an initial region of interest an initial estimate, Y_{target}, of the response is computed by fitting MPCR to the low-frequency inputs, $X_{initial}$, and outputs, $Y_{initial}$.

When generating road profiles with ISO 8608, the road profile classes specify the lower and upper limits to the Spatial Power Spectral Density (SPSD) amplitudes as a function of position x for a specific road class. To generate the artificial road profiles, one uses a stochastic representation, a function of the SPSD of vertical displacements. This SPSD is obtained through a Fourier Transform of the auto-correlation function of the stochastic process describing the road profile [23]. In general, an input profile can be generated using this approach with the following equation

$$X_{in}(t) = \sum_{n=1}^{k} \left(\frac{1}{n}\right)^{m} R(t)\cos(\omega(n)t + \phi_i) \tag{1}$$

The function $R(t)$ is the ramp-up function, which starts small and ramps up to maximum amplitude. Finally, the exponent m changes the weighting of the frequency content to place more emphasis on low frequencies as m increases. In this work, m was selected to be 1.75, thus giving smooth low-frequency signals with some discontinuities. Figure 1 presents initial signals generated in this way and the corresponding response for the half-car model.

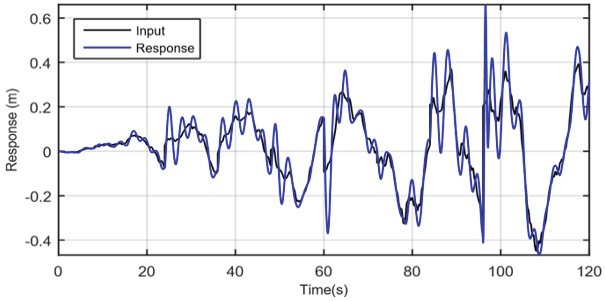


Fig. 1. Initial signal and the corresponding response. The response was generated using the half-car model discussed in Sect. 5.3.

This initial signal aims to find the magnitudes of the input signals and to this end crude low-frequency estimations are sought. Figure 2 shows the response to the estimated input signal and the target response. Note that the low-frequency content of the target response is already captured.

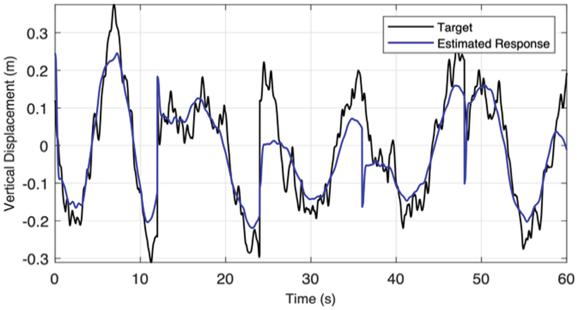


Fig. 2. Response generated from the initial estimate of the input with the target response for the half-car model driving on a road.

5.2 Finding Prototypes

The prototype bootstrapping problem is an unsupervised learning problem where feature prototypes represent a cluster of features. In this particular application, the features describe events that explain important behaviors of the system. Initially, these feature prototypes are centroids of the clusters and are low dimensional estimates of the required inputs. As more information is obtained, as discussed in Sect. 4.3, these estimations improve by shifting the centroid in the input space.

In this work, three different clustering algorithms were used: K-means, K-medoids and spectral clustering. All three of the spectral clustering methods presented by Von Luxburg in [25] are used in this work.

The configuration of the algorithm is based on the choice of clustering algorithm and the associated metric. When performing AFT, one is confronted with a new system for which the best configuration is unknown. Running tests to determine this is not feasible due to the fatigue life accrued while taking these measurements. A consistent and robust configuration is therefore preferred.

K-means clustering minimizes the square distance between points to find a centroid for a cluster. K-medoids, however, minimizes the sum of dissimilarities and chooses an existing data point at the center of a cluster to be the centroid. In general K-medoids are more robust to noise and outliers compared to K-means clustering [26]. In this application this attribute was found to be a disadvantage since the high dimensional space is large and the data is sparse, causing a resilience to outliers to diminish the overall performance of PBM.

Spectral clustering constructs a similarity graph for the data set. From this similarity graph a graph Laplacian is constructed. This process transforms the data into a sparse representation of similarities and pairwise distances of the data. The eigenvectors of the graph Laplacian are then used to represent the data which is then clustered using K-means clustering. Due to the sparse nature of the input and output spaces, the sparse characteristic of centroids found by spectral clustering could be beneficial. For a thorough tutorial on spectral clustering, we refer the interested reader to the text cited above by Von Luxburg [25].

Previous work showed that working in high dimensional space with fewer clusters performed better than a large number of clusters in a low dimensional space [4]. The feature lengths for generating the data were 1000 variables each.

5.3 Generating New Input Signals

In order to populate the input space, the centroids are perturbed at locations spaced equally far apart. At each of these locations, samples are drawn based on an LHS technique. This ensures initial perturbations are not clumped together around these locations but rather are spread out.

Since the target response is known, the error between the response of the predicted inputs and the target response is actively monitored. When the residuals of the target response saturate or increases, the number of perturbation locations is increased. As the algorithm progresses and the number of perturbation locations increases, the input signals become better approximations. The smoothness of the perturbation is determined by the interpolation method used. In this work, cubic splines were used for interpolating between the samples drawn.

The number of perturbation locations on the centroid determines the frequency content of the perturbation. Starting with a small number of locations results in a low-frequency input signal. As the number of locations is gradually increased, higher frequency content is added to the input signal. Figure 3 shows a low-frequency input signal resulting from a small number of perturbation locations. Figure 4 shows an input signal with low and high-frequency content produced by adding many perturbation locations.

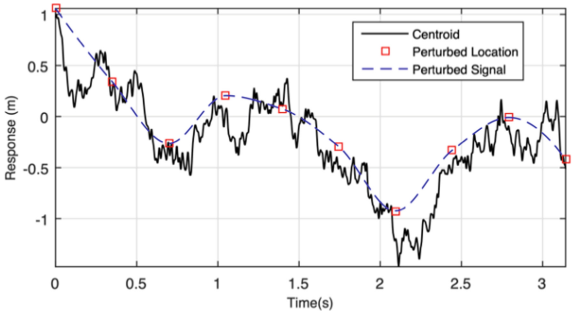


Fig. 3. A low dimensional perturbation of a road profile. The small number of locations at which the centroid is perturbed results in a smooth low-frequency signal.

When a small number of locations are perturbed on the input signal, a subset of variables of the feature vector is varied. Therefore, this is called a low dimensional perturbation as the centroid is varied based on a small subset of the variables. Alternatively, when a large number of variables are varied, it is termed a high dimensional perturbation. Using LHS at each location implies the variables of the feature vector is assumed to follow a uniform distribution. Since the amount of information available is scarce, this is a valid assumption. Future work will investigate a Bayesian update mechanism by which the posterior distributions of the variables can be estimated.

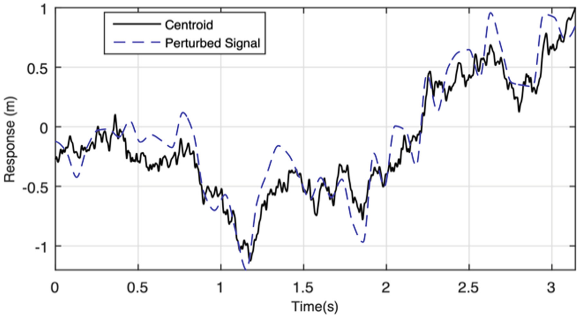


Fig. 4. A high dimensional perturbation of a road profile.

Once new samples are generated, the responses to these inputs are generated. The input and output feature spaces are updated with the new data, and a new centroid in the input space is found given the new information. Generally, a gradual shifting of the centroids is seen. This is controlled by the amplitude of the perturbations and the weighting assigned to new data versus old data.

The amplitude of the perturbations is determined dynamically based on a linearity assumption: we assume that the scale of the residuals can be translated to the perturbation amplitudes. First, the difference between the responses is determined, then this difference is scaled using the ratio between the input and response amplitudes. This scaled difference is then used to perturb the centroids.

The algorithm progresses by constructing a batch of input signals, generating the responses from these inputs and supplementing the dataset with the new data generated. The number of samples drawn from each cluster is determined by how many times a centroid is perturbed. This is done by measuring the information entropy of each cluster [27]. The information entropy of each cluster is then used to dynamically weight the number of samples drawn from each centroid. The higher the entropy of a cluster compared to the other clusters, the larger the number of samples drawn from it in that iteration. At each iteration, the entropy is computed, and the distribution of the samples is adjusted.

The process for populating the input space once the initial centroids have been computed is:

- Prototypes of the input events are determined using a clustering algorithm.
- The computed prototypes are perturbed based on the information entropy of the associated cluster.
- Generate the responses for the batch of input signals generated.
- Update the estimated input signal using MPCR.

6 Numerical Experiments

Section 6.1 presents a comparison between the performance of PBM and DACE for generating data to identify the half-car model, a highly nonlinear model. Section 6.2 uses both the half car and quarter car model to introduce model error into the system by attempting to reconstruct the response of the half-car model in the quarter car model. Finally, Sect. 6.3 discusses the effect of over or under regularization of the chosen regression algorithm.

6.1 Half-Car Model Excited by a Road

The half-car model is excited by a 200m long road profile while travelling at 60 km/h. The process of reconstructing the vertical response, $x(t)$, for the car's center of mass starts with an initial signal generated from Eq. (1). The half-car model's response to the generated input signal is shown in Fig. 5.

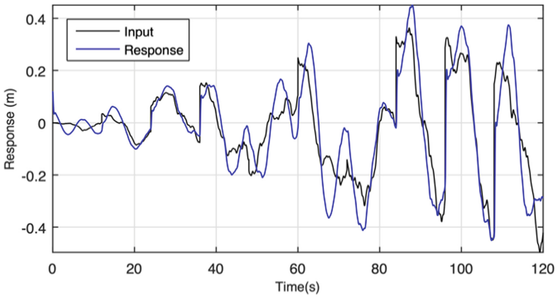


Fig. 5. Input signal and half car model's corresponding response to an initial signal generated using Eq. (1).

MPCR is used to map the responses to the input events. The number of samples generated during testing is restricted to 1 h. Batch samples are 1 min in duration, MPCR is fitted to each of these batches to update the centroid positions. The Mean Squared Error (MSE) of the response generated from the estimated inputs is reported in Table 1. Each of the simulations is repeated five times to quantify the variance in the output target error.

From Table 1, it is seen that the spectral clustering normalized according to Malik performed best. Thus, both the clustering algorithm and the associated metric had a significant influence on PBM's performance.

DACE is applied by first finding the subspace within which the target response was generated. Then, following the initial estimate, the identified subspace is populated using an LHS approach.

Figure 6 presents the response generated by the system, the estimated input using samples generated using DACE and PBM. For PBM, a spectral clustering algorithm normalized according to Malik, and a Euclidean distance metric was used. The MSE when using PBM is 5.33×10^{-4} compared to 4.2×10^{-3} when using DACE.

Table 1. MSE of the response of estimated inputs for clustering algorithms considered in this study

Metric	K-means	K-Medoids	Malik	Unnormalized (Normal)	Unnormalized (Mutual)	Ng (Normal)	Ng (Mutual)
City block	7.44×10^{-4}	6.43×10^{-4}	5.65×10^{-4}	5.94×10^{-4}	6.43×10^{-4}	6.75×10^{-4}	5.66×10^{-4}
Euclidean	6.30×10^{-4}	5.91×10^{-4}	5.13×10^{-4}	6.00×10^{-4}	6.40×10^{-4}	6.38×10^{-4}	6.17×10^{-4}
Cosine	6.64×10^{-4}	6.65×10^{-4}	6.48×10^{-4}	7.66×10^{-4}	5.59×10^{-4}	6.97×10^{-4}	6.96×10^{-4}
Correlation	5.84×10^{-4}	6.64×10^{-4}	5.07×10^{-4}	5.52×10^{-4}	5.48×10^{-4}	6.49×10^{-4}	6.65×10^{-4}

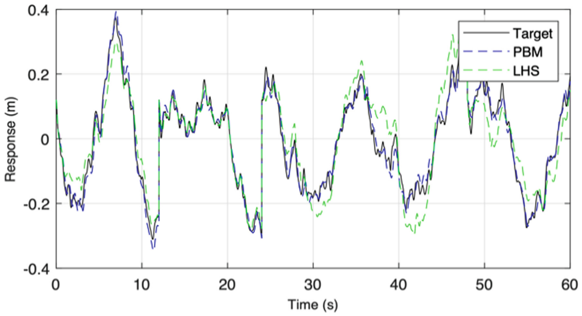


Fig. 6. Response generated from inputs estimated using the PBM and LHS, respectively.

One shortcoming of the DACE sampling strategy in this setting is the high variance of the MSE[3]. The results are therefore unreliable. For example, the 98% confidence interval for DACE in this experiment is $[1.2 \times 10^{-3}, 7.1 \times 10^{-3}]$. Furthermore, the MSE does not decrease consistently when drawing more samples using LHS and the high variance of the residuals persists throughout. This is due to the curse of dimensionality, showing that a space-filling sampling technique in a high dimensional space is not feasible when the number of samples is small compared to the dimensionality of the space.

6.2 Response Reconstruction When a Model Error is Present

When performing AFT model, an error is usually present since the field setup is not the same as the laboratory setup. Due to not all the degrees of freedom being excited or only part of the system being tested in the laboratory.

In this experiment, the half-car model's responses represent the field system. On the other hand, the quarter car model represents the laboratory system and introduces model

---

[3] High variance for DACE persists in the residual throughout the sampling process. Of particular concern however is high variance at the terminal residual, this being high implies that the accuracy of the final response would vary greatly from one simulation to the next.

error into the system identification problem. The displacement of the center of gravity of the half-car model is then to be reconstructed in the displacement of the sprung mass of the quarter car model.

Figure 7 shows the MSE for the input and output of PBM when reconstructing the half-car model's response in the quarter car model. The increase of the input signal MSE is due to the quarter car model's input signal being different from that of the half-car model. Figure 8 shows the response reconstructed in the system using PBM and DACE to estimate the target inputs.

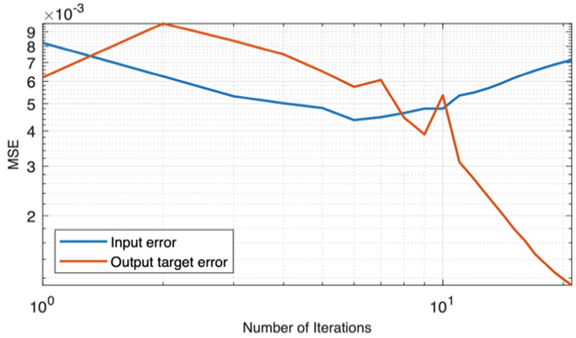


Fig. 7. MSE of the input and output target for reconstructing the half-car model response in the quarter model.

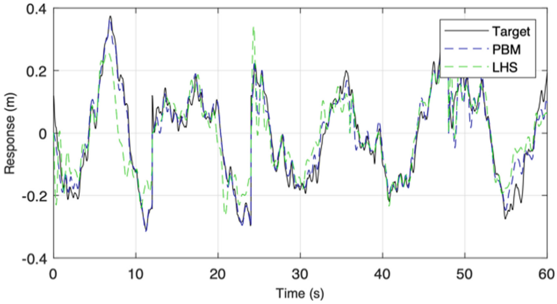


Fig. 8. Response generated from inputs estimated using the PBM and LHS respectively on the quarter car model when reconstructing half-car model's response.

The responses reconstructed when exciting the quarter car model through inputs generated using PBM and DACE had an MSE of 1.06×10^{-3} and 5.64×10^{-3} and 98% confidence intervals of $[9.78 \times 10^{-4}, 1.14 \times 10^{-3}]$ and $[4.45 \times 10^{-3}, 6.83 \times 10^{-3}]$ respectively. These results show that the model error resulted in a 72.2% and 85.9% decrease in performance when using the PBM and DACE.

In the AFT setting, PBM is shown to perform better than DACE: First, the accuracy of PBM is 81.2% higher than the DACE. Second, PBM has higher precision, as seen from the smaller confidence interval. Finally, the input residuals of PBM are seen to improve consistently as more samples were drawn, while for DACE, the residuals saturated early in the sampling process.

6.3 Importance of Regularization in PBM

In this section, a discussion on the importance of regularization when using PBM is presented. The number of components to use in MPCR is crucially linked to the performance of the algorithm. Although regularization is applied, the main item in this discussion is sufficient regularization. To this end, there are three cases to consider: insufficient, sufficient and over regularization. K-fold cross-validation is used in each of these cases. However, in the range of identified components, the minimum or maximum identified values are used in the over and under regularized cases.

In the first case, too little regularization gives rise to sudden jumps in the MSE. Figure 9 illustrates the MSE for estimating the target inputs at various degrees of regularization for the half-car model. This occurs when the variance of the fit is not constrained, which results in PBM proposing an input to the system that falls outside the known data set. The result of this is that shortly afterwards, the algorithm remarkably recovers. This is an artefact of the exploration aspect of PBM. If this property is managed, it improves overall performance. However, the main issue is that if left unattended, these jumps can occur at any time. This makes the final solution unpredictable.

The other extreme is where the regression algorithm is over regularized. PBM is too constrained and thus does not suggest inputs far from the inputs included in the data set. Due to this limitation, PBM does not yield the same improvement after each iteration, as seen with the under-regularized case. However, it is also noted that the MSE curves are smooth, and no sudden jumps are observed.

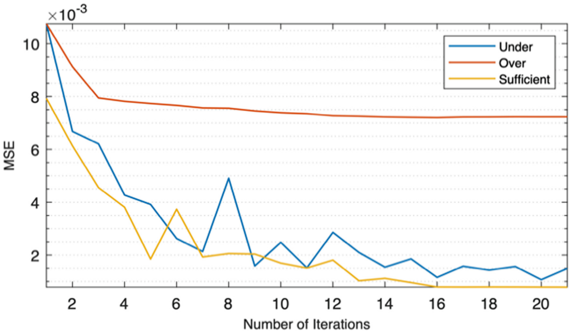


Fig. 9. MSE for the input error of half-car model when under, over and sufficiently regularized

In the case where sufficient regularization is applied at every iteration, PBM can improve. The optimal regularization lies between the two cases mentioned above. A hybrid strategy was adopted to obtain better results: PBM is under-regularized during the first half of the sampling process. This allows PBM to suggest inputs sufficiently far away from existing data to make improvements. After the initial exploration phase is over, PBM is over-regularized, ensuring that the residual stabilizes and the variance between runs is low.

7 Conclusion

In this work, PBM and DACE are compared in two important numerical experiments: The first is in reconstructing the input signal for a highly nonlinear system, the half-car model. The response reconstructed in the half-car model using PBM was an order of magnitude more accurate than DACE. This showed that PMB is more effective than DACE in reconstructing the response in a highly nonlinear system.

The second numerical experiment compared PBM and DACE when a model error is present: reconstructing the half-car model's response in the quarter car model. This is important for the AFT context, where the system is changed when moving for field tests to a laboratory environment. In this case, the target reconstructed using PBM was 5.3 times more accurate compared with DACE. Furthermore, the confidence intervals for PBM compared to DACE was 5.2 times smaller compared to DACE, indicating that PBM was more reliable in reconstructing an accurate input signal.

Finally, different clustering methods and distance metrics were compared for the PBM method. A Euclidian metric and Malik's version of spectral clustering showed the best performance. The clustering algorithm and metric affected performance with up to 19.8% and 24.1%, respectively.

References

1. Escobar, L.A., Meeker, W.Q.: A review of accelerated test models. Stat. Sci. **2006**, 552–577 (2006)
2. MIL-STD-810G. Environmental engineering considerations and laboratory tests, Department of Defence, 2008 (2008)
3. Petracconi, C., Ferreira, S., Palma, E.: Fatigue life simulation of a rear tow hook assembly of a passenger car. Eng. Fail. Anal. **17**(2), 455–463 (2010)
4. Chen, D.-G., Lio, Y., Ng, H. K. T., Tsai, T.-R. (eds.): Statistical Modeling for Degradation Data. IBSS, Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-5194-4
5. Kerschen, G., Worden, K., Vakakis, A.F., Golinval, J.-C.: Past, present and future of nonlinear system identification in structural dynamics. Mech. Syst. Signal Process. **20**(3), 505–592 (2006)
6. Billings, S.A.: Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains. Wiley, New York (2013)
7. Eksteen, J.J.A.: Advances in iterative learning control with application to structural dynamic response reconstruction. Ph.D. thesis, University of Pretoria (2014)
8. Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., Ljung, L.: Kernel methods in system identification, machine learning and function estimation: a survey. Automatica **50**(3), 657–682 (2014)
9. Tijani, I.B., Akmeliawati, R., Legowo, A., Budiyono, A.: Nonlinear identification of a small scale unmanned helicopter using optimized NARX network with multiobjective differential evolution. Eng. Appl. Artif. Intell. **33**(2014), 99–115 (2014)
10. Han, X., Xie, W.-F., Fu, Z., Luo, W.: Nonlinear systems identification using dynamic multi-time scale neural networks. Neurocomputing **74**(17), 3428–3439 (2011)
11. Wei, H.-L., Billings, S.A., Balikhin, M.A.: Wavelet based non-parametric NARX models for nonlinear input-output system identification. Int. J. Syst. Sci. **37**(15), 1089–1096 (2006)
12. Zhao, W., Chen, H.-F., Bai, E.-W., Li, K.: Kernel-based local order estimation of nonlinear nonparametric systems. Automatica **51**(2015), 243–254 (2015)

13. Spiridonakos, M., Chatzi, E.: Metamodeling of dynamic nonlinear structural systems through polynomial chaos NARX models. Comput. Struct. **157**(2015), 99–113 (2015)
14. Smolders, K., Volckaert, M., Swevers, J.: Tracking control of nonlinear lumped mechanical continuous-time systems: a model-based iterative learning approach. Mech. Syst. Signal Process. **22**(8), 1896–1916 (2008)
15. Crous, J., Kok, S., Wilke, D.N., Heyns, P.S.: An alternative approach to system identification. In: The Proceeding of the 10th South African Conference on Computational and Applied Mechanics (2016)
16. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. J. Global Optim. **13**(4), 455–492 (1998)
17. Liefvendahl, M., Stocki, R.: A study on algorithms for optimization of Latin hypercubes. J. Statist. Plan. Inference **136**(9), 3231–3247 (2006)
18. Crous, J.M.: A statistical learning approach to response reconstruction for accelerated destructive degradation testing. Ph.D. thesis, University of Pretoria (2019)
19. Deng, J.: Dynamic neural networks with hybrid structures for nonlinear system identification. Eng. Appl. Artif. Intell. **26**(1), 281–292 (2013)
20. Eksteen, J.J.A., Heyns, P.S.: An alternative update formula for nonlinear model-based iterative learning control. Inverse Probl. Sci. Eng. **24**(5), 860–888 (2016)
21. Sun, J.: A multivariate principal component regression analysis of NIR data. J. Chemom. **10**(1), 1–9 (1996)
22. I. Standard, 8608. Mechanical vibration – road surface profiles – reporting measured data (1995)
23. Agostinacchio, M., Ciampa, D., Olita, S.: The vibrations induced by surface irregularities in road pavements–a MATLAB$^{®}$ approach. Eur. Transp. Res. Rev. **6**(3), 267–275 (2014)
24. Moon, F.C.: Chaotic Vibrations: An Introduction for Applied Scientists and Engineers. Wiley-Interscience, New York (2004)
25. Von Luxburg, U.: A tutorial on spectral clustering. Statist. Comput. **17**(4), 395–416 (2007)
26. Hastie, T., Tibshirani, R., LFriedman, J.: The Elements of Statistical Learning. SSS, Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
27. Chen, C.-B., Wang, L.-Y.: Rough set-based clustering with refinement using Shannon's entropy theory. Comput. Math. Appl. **52**(10–11), 1563–1576 (2006)

Adaptive On-Line Estimation of Road Profile in Semi-active Suspension

Maroua Haddar[1](✉), Fathi Djmal[1], Riadh Chaari[1], S. Caglar Baslamisli[2],
Fakher Chaari[1], and Mohamed Haddar[1]

[1] Mechanics, Modeling and Production Laboratory (LA2MP), Mechanic Department,
National Engineering School of Sfax (ENIS), 1173, 3038 Sfax, Tunisia
`maroua.haddar@enis.tn`
[2] Department of Mechanical Engineering, Hacettepe University, Beytepe, 06800 Ankara, Turkey

Abstract. An on-line algebraic estimator for parameter identification is integrated a recent road profile estimator. The classical version of a road profile estimator that is based transmission characteristics suffered from parametric variation. Sprung mass is variable by changing the weight of passengers and baggage's. Furthermore, the damping of semi-active suspension does change rapidly during vehicle driving. The challenge is to identify all imperceptible modifications that can threaten the vehicle performances and to keep the credibility of road profile estimator online. A Differential algebra and operational calculus rules can be helpful to overcome the impact of variation in real time. The algebraic estimator has the fastest detection time and non-asymptotic behavior. In literature, the algebraic observers shown a good estimation of vehicle mass uncertainties. Additionally, this estimator requires a lower number of sensors and has a lower computational overhead. Measurements of vertical accelerations are only required to algebraically identify the sprung mass and damping coefficient of a quarter car model. Some numerical simulation results in time domain and frequency domain are provided. This new version of adaptive estimator can be integrated with the active controller in the future with easy implementation.

Keywords: Road profile · Algebraic estimator · Parametric identification · Sprung mass · Damping coefficient

1 Introduction

Several research methods have been proposed to identify road profiles and to get precise information about road service ability. Indeed, road profile excitation is classified as one of the main exogenous perturbation that acts on road vehicles' ride dynamics.

This chapter is focused on the proposed scheme based on the principle transmission characteristics of the system proposed by (Liu et al. 2020). The basic idea is to use vehicle dynamic responses in order to reconstruct the road profile with one sensor. The algorithm requires only the unspring mass, hence the whole process straightforward to tune and is not expensive to implement. This technique is can replace traditional methods

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 144–151, 2022.
https://doi.org/10.1007/978-3-030-85584-0_15

of estimations. However, for getting a good estimation, a new concept of algebraic parametric estimators will be integrated to previous road profile estimators. The main advantage of this technique, only one setting parameter can be manually adjusted to improve the quality of the estimated road profile. The identification method based on the algebraic parametrical technique was first introduced by (Fliess and Sira-Ramirez 2003).

Actually, the use of operational calculus rules in combination with differential algebra creates an effective methodology for the estimation of dynamic system parameters. Major contributions of the proposed technique have been made in fields such as intelligent controller design (Haddar et al. 2019). The choice of the differential-algebraic theory for estimation is based on characteristic features of finite-time algebraic estimators (non-asymptotic state estimation).In fact, the influence of the initial conditions is indeed removed as claimed by (Beltrán-Carbajal and Silva-Navarro 2013). This truly is an improvement over the classical observers, which need the right initial conditions. Unknown or incorrect initial conditions invariably entail slow convergence of recursive type of observers. In addition, the presence of integrals in the estimation procedure acts like a low pass filter, which naturally reduces the influence of noise and external perturbation and hence is good at estimating vehicle parameters from a noisy signal.

The organization of the paper is as follows: a simple car model and the description of the proposed scheme of road profile estimator are presented in Sect. 2. Section 3 describes the algebraic parametric estimator and its principle rules for implementation process. The effectiveness of the proposed algebraic estimator in enhancing the credibility of road profile estimator is illustrated in Sect. 4. Finally, the conclusion is given in Sect. 5.

2 Scheme of Road Profile Estimation

A model of a vehicle with two degrees of freedom is considered as the most basic model that could describe the automotive suspension (Fig. 1). It consists of an assumption based on considering that the total mass of vehicle is equally distributed among the four wheels. Only vertical movements are considered. Dampers or springs prevent the amplification of disturbances caused by the road profile while maintaining good road contact. The selected simplified model is helpful, for a first study, to validate the proposed estimator. The dynamic behavior of a quarter-car model with a semi-active suspension is described by:

$$m_s \ddot{x}_s = -d_s(\dot{x}_s - \dot{x}_u) - k_s(x_s - x_u) \tag{1}$$

$$m_u \ddot{x}_u = k_s(x_s - x_u) + d_s(\dot{x}_s - \dot{x}_u) - k_t(x_u - x_r) \tag{2}$$

where, x_s, x_u and x_r are the sprung mass displacement, unsprung mass displacement and road profile excitation, respectively. The chassis is represented by m_s, the wheel and the tire are represented by m_u. k_s is the suspension stiffness and k_t is the tire stiffness. The damper d_s in this case, called “controllable damper” and is variable.

The road excitation will be estimated from information’s given by the un-sprung mass acceleration and the transmission characteristics of the quarter car model:

$$T(s) = \ddot{x}_u(t) \tag{3}$$

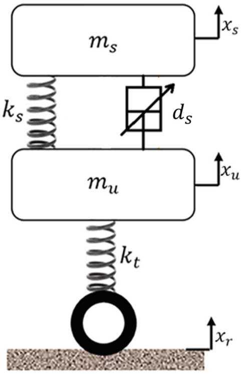


Fig. 1. Semi-active quarter car model

After a Laplace transformation of equation of motion and get rid of the sprung mass expression, we can get the following equation:

$$\left(m_s s^2 + d_s s + k_s + k_t\right) x_u(s) - \frac{(k_s + d_s s)^2}{m_s s^2 + d_s s + k_s} x_u(s) = k_t x_r(s) \tag{4}$$

The relationship between input signal and output signal of Eq. (4), allow us to write a transfer function between road excitation and un-sprung mass acceleration can be created:

$$\frac{x_r(s)}{T(s)} = \frac{m_s m_u s^4 + (m_s d_s + m_u d_s)s^3 + (m_s k_s + m_u k_s + m_s k_t)s^2 + k_t d_s s + k_t k_s}{m_s k_t s^4 + k_t d_s s^3 + k_t k_s s^2} \tag{5}$$

It is observable that the road excitation can be estimated directly by measuring only the vertical wheel acceleration.

3 Estimator of Vehicle Parameters

As we know, the sprung mass is variable and can affects the vehicle dynamics. Furthermore, the damping of semi-active suspension can change rapidly during vehicle driving. In order to estimate failures, we try to incorporate on-line the coefficient of the transfer function for reducing the sensitivity of proposed estimator.

A parametric algebraic estimator is proposed for obtaining the sprung mass and damping coefficient values. Starting by Laplace transformation of the first equation of motion (1):

$$m_s\left[s^2 X_s(s) - s x_s(0) - \dot{x}_s(0)\right] + d_s[sX_s(s) - x_s(0) - sX_u(s) + x_u(0)] = -k_s[X_s(s) - X_u(s)] \tag{6}$$

A double differentiation with respect to s is required for get rid of initial conditions:

$$m_s\left[2X_s + 4s\frac{dX_s}{ds} + s^2\frac{d^2X_s}{ds^2}\right] + d_s\left[2\frac{dX_s}{ds} + \frac{d^2X_s}{ds^2} - 2\frac{dX_u}{ds} - \frac{d^2X_u}{ds^2}\right] = -k_s\left[\frac{d^2X_s}{ds^2} - \frac{d^2X_u}{ds^2}\right] \tag{7}$$

No high power is allowed in order to attenuate the impact of noise:

$$m_s\left[2s^{-2}X_s + 4s^{-1}\frac{dX_s}{ds} + \frac{d^2X_s}{ds^2}\right] + d_s\left[2s^{-2}\frac{dX_s}{ds} + s^{-1}\frac{d^2X_s}{ds^2} - 2s^{-2}\frac{dX_u}{ds} - s^{-1}\frac{d^2X_u}{ds^2}\right]$$
$$= -k_s\left[s^{-2}\frac{d^2X_s}{ds^2} - s^{-2}\frac{d^2X_u}{ds^2}\right] \tag{8}$$

From Laplace domain to time domain, we can get the following relation based on iterated integrals and required only verticals displacements x_s and x_u:

$$m_s\left[2\iint x_s\,dt + 4\int t\,x_s\,dt + t\,x_s^2\right] + d_s\left[2\iint t\,x_s\,dt + \int t^2\,x_s\,dt - 2\iint t\,x_u\,dt - \int t^2\,x_u\,dt\right]$$
$$= -k_s\left[\iint t^2\,x_s\,dt - \iint t^2\,x_u\,dt\right] \tag{9}$$

The last Eq. (9), after some more integrations, leads to the linear system of equations:

$$A(t)\theta = B(t) \tag{10}$$

where $\theta = [m_s, d_s]^T$ denotes the parameter that should be identified online, $A(t)$ and $B(t)$ are 2×2 and 2×1 matrices respectively;

$A(t) = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B(t) = \begin{bmatrix} b_{11} \\ b_{12} \end{bmatrix}$.

whose components are time functions defined as

$$a_{11} = 2\iint x_s\,dt + 4\int t\,x_s\,dt + tx_s^2,\ a_{12} = 2\iint t\,x_s\,dt + \int t^2\,x_s\,dt - 2\iint t\,x_u\,dt - \int t^2\,x_u\,dt,$$

$$b_{11} = -k_s\left[\iint t^2\,x_s\,dt - \iint t^2\,x_u\,dt\right],\ a_{21} = \int a_{11},\ a_{22} = \int a_{12},\ b_{12} = \int b_{11}$$

By solving system (10) one obtains the parameter vector θ as

$$\theta = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \begin{bmatrix} b_{11} \\ b_{21} \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} \tag{11}$$

with

$$\Delta = a_{11}a_{22} - a_{12}a_{21} \tag{12}$$

$$\Delta_1 = a_{22}b_{11} - a_{12}b_{21} \tag{13}$$

$$\Delta_2 = a_{11}b_{21} - a_{21}b_{11} \tag{14}$$

For avoiding singularities, we propose the following algebraic identifier (Beltrán-Carbajal and Silva-Navarro 2013)

$$\widehat{m}_s = \frac{\iint \Delta_1}{\iint \Delta},\ \widehat{d}_s = \frac{\iint \Delta_2}{\iint \Delta} \tag{15}$$

4 Numerical Simulation

To check the performance of the road profile estimator in the presence of algebraic parametric identifier, some numerical simulations were performed on a 2-DOF quarter car model characterized by the parameters given in Table 1.

The verticle road exciataion is selected according to model presented by Múčka et al. (2020).

Table 1. Parameters of suspension model.

Parameters	Value
Body mass m_s	317.5–635 kg
Wheel mass m_u	45.5 kg
Damping coefficient d_s	1000–7000 N.s/m
Spring stiffness k_s	20000 N/m
Tire stiffness k_t	192000 N/m

Figure 2 and Fig. 3 show the comparison between the real road disturbance and estimated profile in the time domain and frequency domain, respectively. It is perceivable from the Fig. 2 that the road information obtained from transfer function of Eq. (5) gives a good tracking of the original input road (The relative error is equal to 2%). Furthermore the power spectra density (PSD) verifies the effectiveness of the proposed road disturbance estimator Fig. 3.

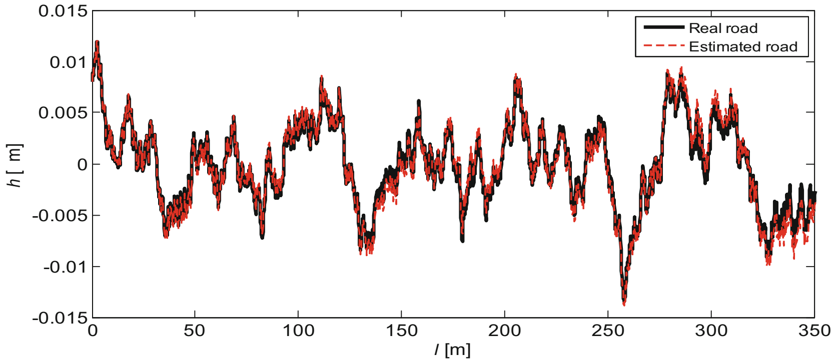


Fig. 2. Estimated road profile

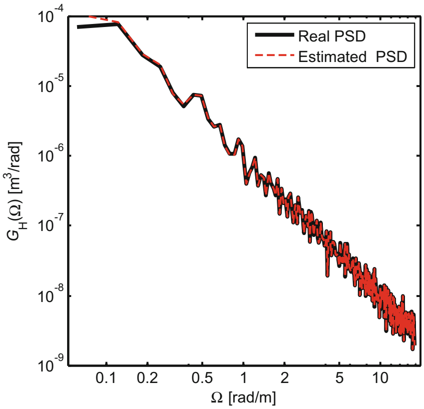


Fig. 3. PSD of estimated road profile

However, the vehicle parameters are able to be changed. The influence of variation is depicted in Fig. 4 and Fig. 5 under different cases:

Estimated 1: $m_s = 317.5\,kg,\ \ d_s = 7000\,N.s/m.$
Estimated 2: $m_s = 338\,kg,\ \ d_s = 1000\,N.s/m.$
Estimated 3: $m_s = 635\,kg,\ \ d_s = 7000\,N.s/m.$

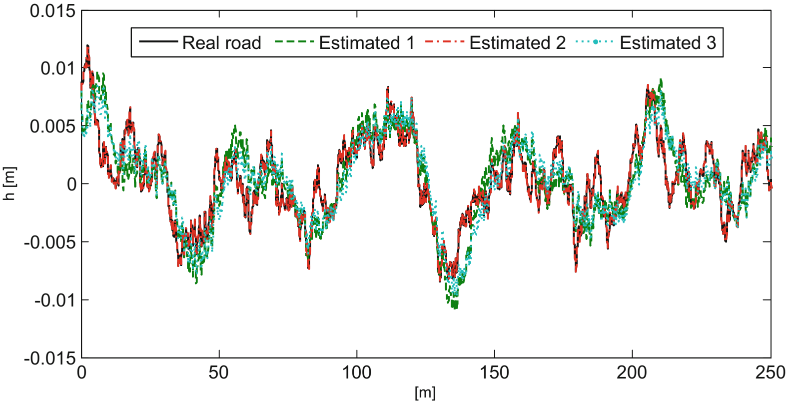


Fig. 4. Estimated road profile with parametric variation

Figure 4 and Fig. 5 shown the sensitivity of transfer function to quarter car model and specially to damping coefficient. The damping is varying in semi active suspension implementation. Sometimes, this change is very rapid. Therefore, it will be with a significant impact of damping variation on the estimate.

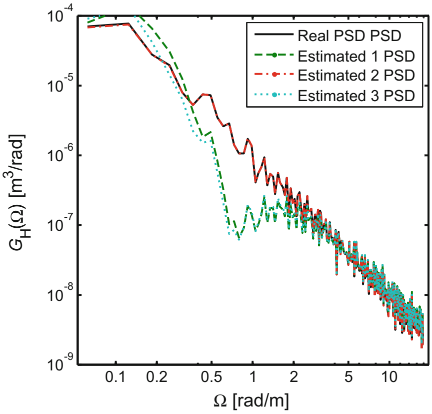


Fig. 5. PSD of estimated road profile with parametric variation

Then, the road disturbance estimator should be adjust the transfer function in real-time according to the variation of sprung mass and the damping coefficient, so as to obtain accurate estimation results.

Figure 5 shows the sprung mass and damping coefficient variations using the algebraic identifiers (15). Runge Kutta method with fixed small step time of 0.001s were used in the simulation implementation with Simulink (Haddar et al. 2017). Fast parameter estimation before t = 0.1 s.

In Fig. 6 the values of the estimated parameters were inserted in the transfer function online after 2 s in the case 3 (where $m_s = 635\,kg,\ d_s = 7000\,N.s/m$). Before 2 s, it is shown that the estimation results of the road disturbance in time-domain is deteriorated by the variation of vehicle parameters. However, the estimation results can still well follow the road input in the presence of algebraic identifier (Fig. 7).

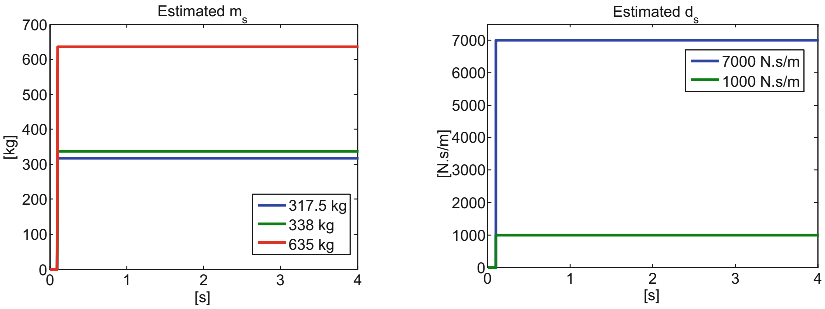


Fig. 6. Estimated vehicle parameters with algebraic estimator

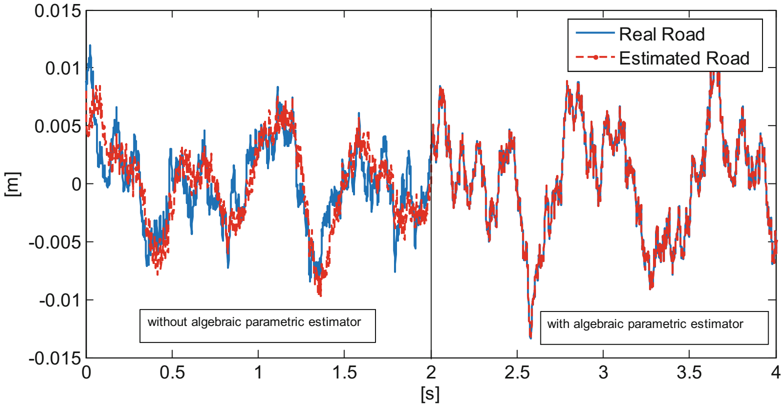


Fig. 7. Estimated road profile with and without algebraic estimator

5 Conclusion

In this chapter, the credibility of modified on-line estimation technique of road profile were tested by changing the sprung mass and damping coefficient The numerical results show that the algebraic estimator can well adjust the transfer function to the variation of model parameters. The proposed scheme can be useful in the case of all process with damping varying, the algebraic estimator always transmits the damping coefficient to the road profile estimator in real-time. In previous work, it is noted that in fact, the damping of the shock absorber cannot be obtained directly during vehicle driving. For exemple, in the case of magnetorheological (MR), only the control current of MR damper can be directly obtained. The corresponding calculation is required to obtain the damping coefficient. However, the proposed scheme is able to avoid this kind of problem and estimate directly the damping coefficient.

References

1. Liu, W., Wang, R., Ding, R., Meng, X., Yang, L.: On-line estimation of road profile in semi-active suspension based on unsprung mass acceleration. Mech. Syst. Signal Process. **135**, 106370 (2020). https://doi.org/10.1016/j.ymssp.2019.106370
2. Fliess, M., Sira-Ramirez, H.: An algebraic framework for linear identification. ESAIM: Control Optim. Calc. Var. **8**, 151–168 (2003)
3. Haddar, M., Chaari, R., Baslamisli, S.C., Chaari, F., Haddar, M.: Intelligent PD controller design for active suspension system based on robust model-free control strategy. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **233**(14), 4863–4880 (2019)
4. Beltrán-Carbajal, F., Silva-Navarro, G.: Adaptive-like vibration control in mechanical systems with unknown paramenters and signals. Asian J. Control **15**(6), 1613–1626 (2013)
5. Múčka, P., Stein, G.J., Tobolka, P.: Whole-body vibration and vertical road profile displacement power spectral density. Veh. Syst. Dyn. **58**(4), 630–656 (2020)
6. Haddar, M., Baslamisli, S. C., Chaari, F., Haddar, M.: On-line adaptive scaling parameter in active disturbance rejection controller. In: Felkaoui, A., Chaari, F., Haddar, M. (eds.) SIGPROMD'2017 2017. ACM, vol. 12, pp. 79–86. Springer, Cham (2019). https://doi.org/10.1007/978-3-319-96181-1_7

A Parametric Study to Investigate the Dynamic Behavior of Worm Gear with Plastic Wheel

Chaima Hammami[1](✉), Ala Eddin Chakroun[1,2], Ahmed Hammami[1], Ana De-Juan[2], Fakher Chaari[1], Alfonso Fernandez[2], Fernando Viadero[2], and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modelling and Production (LA2MP), National School of Engineers of Sfax, BP1173, 3038 Sfax, Tunisia

[2] Department of Structural and Mechanical Engineering, Faculty of Industrial and Telecommunications Engineering, University of Cantabria, Avda de los Castros s/n, 39005 Santander, Spain

Abstract. Plastic gears use is in an increasing growth. Plastic material has many interesting characteristics that cannot be found in metallic ones. Searching for more efficiency, combining advantages of metallic and plastic components of worm gears sets is thus employed. Most of the researches done for years had focused on the phenomena that occur in the contact area between the worm and the wheel. In this paper, a parametric approach is presented to, quickly and efficiently, investigate the vibration levels of worm gears including plastic wheel and metallic worm. A dynamic model is first described to establish equations of motion. The viscoelasticity of the worm gear's material is then taken into consideration. It is described by two parameters: The storage modulus describing the elastic behavior; and the loss factor describing the viscous one. The dependency of viscoelastic materials on frequency and temperature is considered. A parametrized linear viscoelastic problem is thus investigated here. Finally, numerical simulation based on modal analysis is established. The result is an estimation of the evolution of natural frequencies regarding the variation of the viscoelastic material's parameters. Consequently, it can give the range of frequencies that can reach a worm gear model including a specific plastic material.

Keywords: Worm gear · Dynamic · Viscoelastic · Parametric study · Modal analysis

1 Introduction

Worm gear set is widely used in several mechanical applications. It is used in power transmission to its highest reduction ratio and quieter level compared to other gears. To reduce cost manufacturing and deal with high sliding in the mesh gear, plastic material is introduced in worm gear (Tsukamoto 1995). Plastic material will also add dissipation that limits the vibration level and then enhance the general performance. It is also resistant to wear as it has low friction coefficient, and it is self-lubricated. Both components or just the wheel can be plastic. The worm gear is also characterized by its rather complex

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 152–159, 2022.
https://doi.org/10.1007/978-3-030-85584-0_16

geometry. This reflected on the number of publications that is not important as the case of spur gears for example.

Among the works that have been done on worm gears with plastic components, works described below here: The study of the quasi-static loaded behavior of a worm gear with a steel worm and a nylon wheel using Kelvin's model (Hiltcher et al. 2006). Kim et al. (2012) proposed a method to calculate the efficiency of worm gears with plastic wheel. This was performed through several steps including experiments to measure the nylon6 friction coefficient according to normal contact pressure. Koide et al. (2014) determined the effects of worm tooth forms on the strength of the plastic helical wheels by running fatigue tests. Liu et al. (2014) investigated about the mesh performance of steel worm with plastic involute cylindrical gear using finite element method. Jiaxing and Ilie (2014) proposed a calculation method to determine the gear shear strength with different geometry and material based on experimental test. Liang et al. (2019) studied a plastic worm against metallic wheel to estimate the contact stress of tooth. Yun et al. (2019) developed an approach to analyze the load distribution for plastic helical gear meshed with steel worm. Recently, Shi et al. (2021) proposed an inequal pitch theory to have more reliable design theory for plastic worm gears. This can improve the meshing area condition.

Through all the works that have been done, scientists showed interest specially on the phenomena that occur in the meshing area. They investigated about the load sharing and the heat generation that cause wear and impact the gear lifetime.

In this paper, the dynamic of the worm gear set is under investigation. a parametric study is established on worm gear model to focus on the viscoelastic behavior of plastic material used in it. Through this study, a simplified approach is developed. It describes the variation, in frequency domain, of the characteristics of viscoelastic material. It also illustrates its effect on the dynamic behavior of worm gear and consequently on vibration levels.

In the first section, the dynamic model of a worm gear where components are in perfect rotation motion around their axis, is described. Using Lagrange method, the equations of motion are established.

In the second section, a parametric approach describing viscoelasticity in the frequency domain is detailed. The parameter studied is the storage modulus of the plastic material. Small displacements and not preloaded configurations are taken into consideration.

In the third section, numerical simulation is presented. Through a modal analysis, the effect of the viscoelastic behavior on natural modes is illustrated.

2 Dynamic Model

The dynamic model used in this paper and shown in Fig. 1 is composed of a steel worm (w) and a wheel called also worm gear (g) in plastic material. One can assume in this study a perfect configuration, that can be enhanced in future works. The clearance between teeth is ignored. No error on the gear/ worm tooth profile is taken into consideration. The worm is single thread. There are no translational degrees of freedom. Only rotations around components axis are considered.

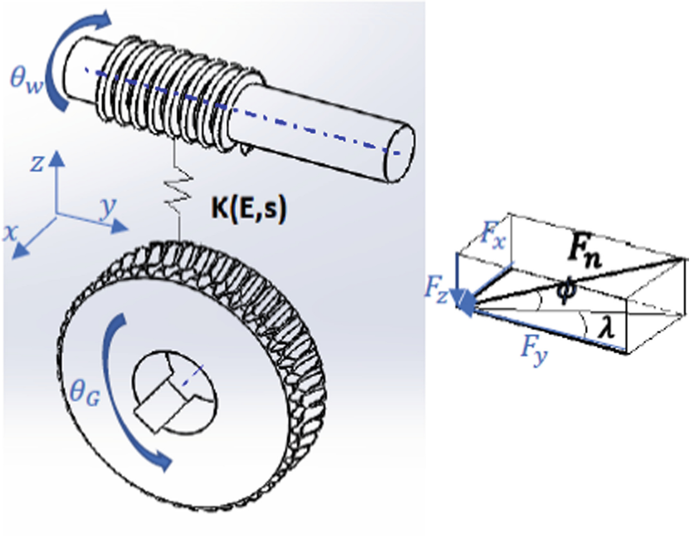


Fig. 1. Dynamic model of the worm gear and loads on the meshed teeth

Loads transmitted in the meshing gear between teeth are the normal load F_n and the friction load F_f. These loads depend on the contact points varying along the action line with rotation by time. The model under investigation here is a lumped one then $F_n = F_n(t)$. In this study, considering that $F_f = \mu F_n$ with μ the friction coefficient. Every force exerted on the worm can be decomposed into three orthogonal force components written as follows (Yeh and Wu 2009):

$$F_x = F_n cos\phi sin\lambda + F_f cos\lambda \tag{1}$$

$$F_y = F_n cos\phi cos\lambda - F_f sin\lambda \tag{2}$$

$$F_z = F_n sin\phi \tag{3}$$

Where, ϕ is the pressure angle and λ is the lead angle of the worm.

Two rotational equilibrium equations can be expressed for the model above for every component. For the worm in rotation θ_w around his axis, the equation of motion is written:

$$I_w\ddot{\theta}_w = -F_x R_w + T_w \tag{4}$$

Where, I_w is the inertia moment, R_w is the pitch radius and T_w is the input torque respectively of the worm. For the worm gear in rotation θ_g around his axis, the equation of motion is written:

$$I_g\ddot{\theta}_g = F_y R_g + T_g \tag{5}$$

Where, I_g is the inertia moment, R_g is the pitch radius and T_g is the output torque respectively of the worm gear.

The deflections $\delta_i (i = w, g)$ on the teeth of worm and worm gear in contact during the meshing period are calculated by:

$$\delta_w(t) = \frac{F_n(t)}{k_w(t)} \tag{6}$$

$$\delta_g(t) = \frac{F_n(t)}{k_g(t)} \tag{7}$$

Where, $k_i(t)$ is the stiffness of the component $i = w, g$. To simplify the writing of next equations, the dependence on time of different quantities will not be taken into consideration. The assumption of Eqs. (6) and (7) gives:

$$\delta_w + \delta_g = F_n \left(\frac{k_g + k_w}{k_g k_w} \right) \tag{8}$$

Teeth are modelled as springs. Each meshing gear pair is so composed of two springs joined in series. The total mesh stiffness k_{gw} is then equal to the stiffness of set of springs k_i joined in parallel. This equivalent stiffness is varying periodically in time and with the number of meshing gear pair.

$$k_{gw} = \frac{k_g + k_w}{k_g k_w} \tag{9}$$

To maintain the contact of teeth, the total displacement must be the same for each teeth pair in contact simultaneously. Profile error is considered negligible, so one can write (Chung and Shaw 2007):

$$l_w \theta_w - \frac{\delta_w}{\sin\left(\frac{\pi}{2} - \lambda\right)} = R_g \theta_g + \frac{\delta_g}{\sin\left(\frac{\pi}{2} - \lambda\right)} \tag{10}$$

Where l_w is the lead of the worm. This leads to:

$$l_w \theta_w - R_g \theta_g = \frac{\delta_w + \delta_g}{\sin\left(\frac{\pi}{2} - \lambda\right)} \tag{11}$$

Combining Eq. (8) with (11), one can obtain:

$$l_w \theta_w - R_g \theta_g = F_n \frac{k_{gw}}{\sin\left(\frac{\pi}{2} - \lambda\right)} \tag{12}$$

The normal load can then be expressed as follows:

$$F_n = \frac{\left(l_w \theta_w - R_g \theta_g\right) \sin\left(\frac{\pi}{2} - \lambda\right)}{k_{gw}} \tag{13}$$

Replacing the normal load F_n in Eq. (1) by its expression above, Eq. (4) can be written as:

$$I_w \ddot{\theta}_w = -\left(l_w \theta_w - R_g \theta_g\right) k_{1gw} + T_w \tag{14}$$

Where,

$$k_{1gw} = \frac{R_w}{k_{gw}}(cos\phi.sin\lambda + \mu cos\lambda)\sin\left(\frac{\pi}{2} - \lambda\right) \tag{15}$$

The same procedure applied for the gear's Eq. (5) leads to:

$$I_g\ddot{\theta}_g = \left(l_w\theta_w - R_g\theta_g\right)k_{2gw} + T_g \tag{16}$$

Where,

$$k_{2gw} = \frac{R_g}{k_{gw}}(cos\phi.cos\lambda - \mu.sin\lambda)\sin\left(\frac{\pi}{2} - \lambda\right) \tag{17}$$

The total equation of motion of the worm gear model can then be written in the matrix form as:

$$\begin{bmatrix} I_w & 0 \\ 0 & I_g \end{bmatrix}\begin{Bmatrix} \ddot{\theta}_w \\ \ddot{\theta}_g \end{Bmatrix} + \begin{bmatrix} l_wk_{1gw} & -R_gk_{1gw} \\ -l_wk_{2gw} & R_gk_{2gw} \end{bmatrix}\begin{Bmatrix} \theta_w \\ \theta_g \end{Bmatrix} = \begin{Bmatrix} T_w \\ T_g \end{Bmatrix} \tag{18}$$

Equation (18) has the form of:

$$[M]\{\ddot{q}(t)\} + [K(t)]\{q(t)\} = \{W\} \tag{19}$$

Where, $[M]$ is the total mass matrix, $[K(t)]$ is the time-varying total stiffness matrix, $\{q(t)\}$ is the generalized displacements vector and $\{W\}$ is the forces vector.

3 Modelling of viscoelasticity's Parameters in the Frequency Domain

The plastic material is characterized by the viscoelasticity that combines the advantage of elasticity and damping that limits the vibrations levels. To define viscoelasticity, the relation between stress and deformation must be defined. Many rheological models were developed to describe this relation. It can be represented by a set of springs and dashpots. The rheological model used here is the structural damping, composed of complex stiffness matrix $Z(s) = E(1 + i\eta)$, where E is the storage modulus describing elasticity and η is the loss factor of the material describing damping. The worm gear set is considered as a structure when components are in contact. The parameters of this mating are under investigation here.

In this study, the large deformation and prestressed components defining the nonlinear viscoelastic behavior are not taken into consideration. Only the linear behavior of viscoelastic materials is considered. Therefore, the dependence on history and temperature is treated. This dependence can in the frequency domain be represented as a complex modulus $E(s, T)$ which depends on frequency and temperature. For a specific value of a temperature, when assembling a model, one can group elements affected by a given modulus and rewrite the dynamic stiffness as a linear combination of fixed matrices with frequency dependent coefficients (Hammami et al. 2016).

In the frequency domain, the equation of motion (18) is written:

$$[Z(s)]\{q(s)\} = \left[Ms^2 + K(s)\right]\{q(s)\} = [b]\{u(s)\} \tag{20}$$

where q is the degree of freedom vector, the load is the product of the input shape matrix b and the input u(s), Z is the dynamic stiffness matrix combining the mass M and the frequency dependent complex stiffness K.

To investigate the dynamic behavior of the model described in the previous section, natural modes will be calculated as:

$$\{q(t)\} = Re\left\{\{\psi_j\}e^{\lambda_j t}\right\} \tag{21}$$

Where ψ_j and λ_j are the eigenvector and the eigenvalue of the j th mode, respectively. Complex modes are obtained as solutions of these equations:

$$[Z(\lambda_j)]\left\{\psi_{jR}\right\} = \{0\} \text{ and } \left\{\psi_{jL}\right\}^t [Z(\lambda_j)] = \{0\} \tag{22}$$

Where, ψ_{jR} is the right eigenvector and ψ_{jL} is the left one.

Dynamic stiffness $Z(s)$ is an analytical function of the complex plan. One can then write under the Laurent series shape around the poles of its inverse:

$$[Z(s)]^{-1} = \sum_j \frac{\left\{\psi_{jR}\right\}\left\{\psi_{jL}\right\}^t}{\alpha_j(s - \lambda_j)} \tag{23}$$

where, α_j is the normalization coefficient generally considered equal to one to simplify resolution.

In this paper, the goal is to establish a parametric study to simplify the dynamic behavior investigation of a model containing viscoelastic material. The parameters chosen here are the storage modulus E and the loss factor η depending on frequency. The stiffness of the gear can thus be written, around a reference point defined by a specific real modulus E_0, as:

$$k_g(s) = E(s)\frac{k_g(E_0)}{E_0} \tag{24}$$

By scanning a wide range of storage modulus E and loss factor η, a parametric representation of modes defined by the couple $\left\{\lambda_j(E), \{\psi_j(E)\}\right\}$ and dependent on the complex modulus $E(s) = E(1 + i\eta)$ is constructed.

4 Numerical Simulation

The parametric approach detailed above, is numerically simulated under MATLAB software. The worm gear set design' parameters used are detailed in Table 1.

To detect the effect of the variation of the storage modulus on the dynamic of the worm gear set, a range of storage modulus values in the interval $[10^3, 2.10^6]MPa$ is scanned around the reference point describing the Nylon 6 material. This interval is

Table 1. Model design parameters

	Worm	Worm gear
Module (mm)	3	
teeth number	1	20
Inertia moment (kg.mm2)	17.92	73.12
Angle of pressure	20°	
Worm’s lead angle	3°55′	
Normal pressure angle	14°33′	
Worm’s lead (mm)	9.4	
Pitch radius (mm)	8	30

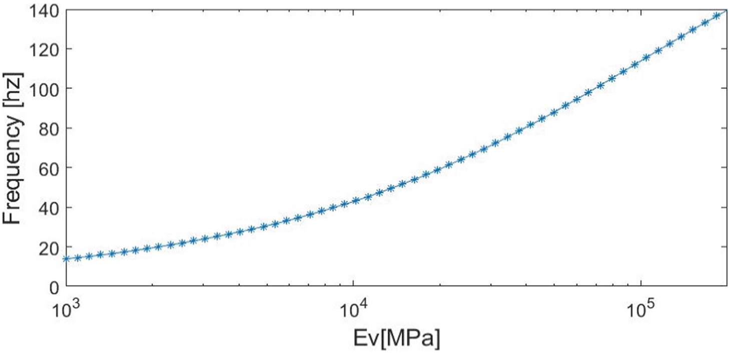


Fig. 2. Evolution of the first natural frequency with the variation of the storage modulus

chosen to cover the maximum of modulus values possibly reached with a plastic material by frequency variation. Natural modes are then calculated for a specific value of loss factor $\eta = 1$. The first natural frequency (after the rigid one), shown in Fig. 2, illustrates the limitation of vibration level with viscoelastic material with low modulus. It also gives an estimation of the range of frequencies that can be obtained with a specific material.

The loss factor is an important characteristic of the viscoelastic material. It depends on frequency as the storage modulus. Thus, to have a reliable representation of the viscoelastic behavior, the evolution of the dynamic behavior of the worm gear must be done against both the storage modulus and the loss factor variation. A map of isovalues, in a space storage modulus/loss factor, can be thus produced to have a complete information about modal damping and frequencies of the model.

5 Conclusion

A parametric approach to study the dynamic behavior of a worm gear set with plastic material is presented along this paper. The parameters of the study are the storage modulus and the loss factor of the worm gear’s viscoelastic material. These characteristics depends on frequency. Describing the dynamic equations of motion in the frequency domain was the first step. A modal analysis was then done through a numerical simulation to illustrate the effect of the variation of the storage modulus on the natural frequency of the model. The estimation of the frequencies range values that can be reached with the viscoelastic material is established. A more completed model including all the translation and rotation degrees of freedom can make a continuity of this work. Profile error and spalling or tooth breakage can also be taken into consideration.

References

Chung, M.Y., Shaw, D.: Parametric study of dynamics of worm and worm-gear set under suddenly applied rotating angle. J. Sound Vib. **304**, 246–262 (2007). https://doi.org/10.1016/j.jsv.2007.02.022

Hammami, C., Balmes, E., Guskov, M.: Numerical design and test on an assembled structure of a bolted joint with viscoelastic damping. Mech. Syst. Sig. Process. **70–71**, 714–724 (2016). https://doi.org/10.1016/j.ymssp.2015.06.031

Hiltcher, Y., Guingand, M., de Vaujany, J.-P.: Load sharing of worm gear with a plastic wheel. J. Mech. Des. **129**, 23–30 (2006). https://doi.org/10.1115/1.2359469

Jiaxing, Z., Ilie, K.: Static shear strength calculation of plastic helical gears mating with steel worm. Int. J. Precis. Eng. Manuf. **15**(2), 235–239 (2014). https://doi.org/10.1007/s12541-014-0330-0

Kim, S.H., Shin, M.C., Byun, J.W., et al.: Efficiency prediction of worm gear with plastic worm wheel. Int. J. Precis. Eng. Manuf. **13**, 167–174 (2012). https://doi.org/10.1007/s12541-012-0021-7

Koide, T., Ishida, Y., Ueda, A., et al.: Strength of plastic helical wheels meshed with various types of worms. In: International Gear Conference (2014). https://doi.org/10.1533/9781782421955.24

Liang, D., Li, C., Hua, C., Luo, T.: Experimental and numerical study of a plastic worm meshed with a steel helical gear. Trans. Can. Soc. Mech. Eng. **43**, 283–293 (2019). https://doi.org/10.1139/tcsme-2018-0016

Liu, Z., Huang, C., Hao, Y., Lin, C.: The mesh property of the steel involute cylindrical worm with a plastic involute helical gear. J. Mech. **30**, 185–192 (2014). https://doi.org/10.1017/jmech.2013.70

Shi, Z., Ren, J., Feng, Z., Li, J.: Key technology and experimental study of unequal pitches meshing between metal worm and plastic helical gears. Appl. Sci. **11** (2021). https://doi.org/10.3390/app11010333

Tsukamoto, N.: Argument on plastic gears for power transmission. JSME Int. J. Ser. C Dyn. Control Robot. Des. Manuf. **38**, 1–8 (1995). https://doi.org/10.1299/jsmec1993.38.1

Yeh, T.-J., Wu, F.-K.: Modeling and robust control of worm-gear driven systems. Simul. Model. Pract. Theory **17**, 767–777 (2009). https://doi.org/10.1016/j.simpat.2009.01.002

Yun, Y., Hu, H., Ta, J.: Load distribution analysis for the plastic helical gear meshing with steel worm. IOP Conf. Ser. Earth Environ. Sci. **237**, 032080 (2019). https://doi.org/10.1088/1755-1315/237/3/032080

A Review of Singular Spectral Analysis to Extract Components from Gearbox Data

Daniel N. Wilke(✉), Stephan Schmidt, and P. Stephan Heyns

Centre for Asset Integrity Management, Department of Mechanical
and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
{nico.wilke,stephan.schmidt,stephan.heyns}@up.ac.za

Abstract. Condition monitoring for rotating machines under time-varying environmental and operating conditions remains an important research problem for several industries, including wind turbines within the renewable energy sector; ship, train and freight transport within the supply-chain sector; crushing and grinding comminution within the mining sector. Proposed methods to solve this problem include synchronous statistics, the squared envelope spectrum, the order-frequency spectral coherence and the integrated squared spectral coherence. Singular Spectral Analysis (SSA) offers a non-parametric alternative to automatically identify potential components of interest. The components of interest are obtained from the resultant matrices, which are computed from grouping elementary matrices. Their contribution to signal reconstruction is achieved through diagonal averaging. SSA is fundamentally a reconstruction-focused linear latent variable model aiming at efficiently explaining the variance in the signal. Firstly, we expect SSA to be more informative when the component of interest manifests strongly in the signal's variance. Secondly, we only expect SSA to isolate the component of interest from the other components in the signal, if variance can isolate its contribution. However, the latter is rather unlikely. Although Singular Spectral Analysis (SSA) is a well-established technique, it has not been critically analysed for its ability to separate components for a damaged gearbox under time-varying operating conditions. This work shows that SSA can separate a damaged gearbox under time-varying operating conditions into various components. Furthermore, SSA is shown to decompose vibration signals from a damaged gearbox into various signal components that could aid fault diagnosis.

Keywords: Singular Spectral Analysis · Component extraction · Gearbox fault detection · Time-varying operating conditions

1 Introduction

Growing fleets of high-value assets within the renewable energy sector that include wind turbines [1], tidal turbines and paddles [2], geothermal power plants [3] and

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 160–172, 2022.
https://doi.org/10.1007/978-3-030-85584-0_17

biomass reactors [4] demand reliable fault diagnosis methods. Several methods have been developed to address some of the challenges in gearbox diagnostics and prognostics [5–7]. Fault diagnosis under time-varying and impulsive noise conditions [6,8,9] remains challenging. In particular, when damaged components manifest in low energy bandwidths, time-varying operating conditions and extraneous impulsive events dominate the raw signal and its processed representations. Data-driven methods for machinery fault diagnosis [10], have recently started to complement established signal processing approaches [6,11–13]. However, several established data-driven time series approaches from the 1980s include singular spectral analysis, a.k.a. singular system analysis (SSA) [14], and independent component analysis (ICA) [15] as part of a larger family of spectral methods. These methods aim to decompose a time-series signal into separate informative components. These components can then be further analysed for condition-based maintenance applications. SSA uses variance as a separability measure, whereas ICA uses statistical independence to identify sources.

Spectral methods comprise two main steps, namely, decomposition or encoding and reconstruction or decoding. Both approaches find a suitable lower-dimensional coordinate system to decompose a signal into components and reconstruct the original signal back during decomposition. Singular spectrum analysis (SSA) is a non-parametric spectral estimation approach that combines time-series analysis elements with multivariate statistics. The singular spectrum relates to the singular values of the singular value decomposition (SVD) of a covariance matrix that aids the decomposition of a time-series signal into components that can easily be interpreted. Since SSA relies on SVD, these components are:

1. extracted based on explaining maximum variance per coordinate, and
2. coordinates that are mutually orthogonal.

This is in contrast to independent component analysis (ICA) that extract signals based on:

1. non-Gaussianity of the components, and
2. components that are statistically independent.

As a result, SSA separates a signal into informative components using variance as an appropriate proxy to extract information that is independent and informative. In contrast, ICA extracts informative components if the degree of non-Gaussianity of the signal is a sensible proxy to extract independent and informative information. SSA does not attempt to explicitly untangle the latent space into independent and informative components or independent sources. SSA may only realise partially untangled latent spaces by using variance as a sensible proxy for independent sources. Although SSA has been applied in isolated cases to bearing defect detection [16], internal-combustion engine sound analysis [17], automobile gearbox vibration analysis [17] and waste-water treatment plants [18], its wide-spread adoption remains elusive. It can complement signal analysis techniques such as the Squared Envelope Spectrum (SES) [11] and the Synchronous Median of the Squared Envelope (SMSE) [13].

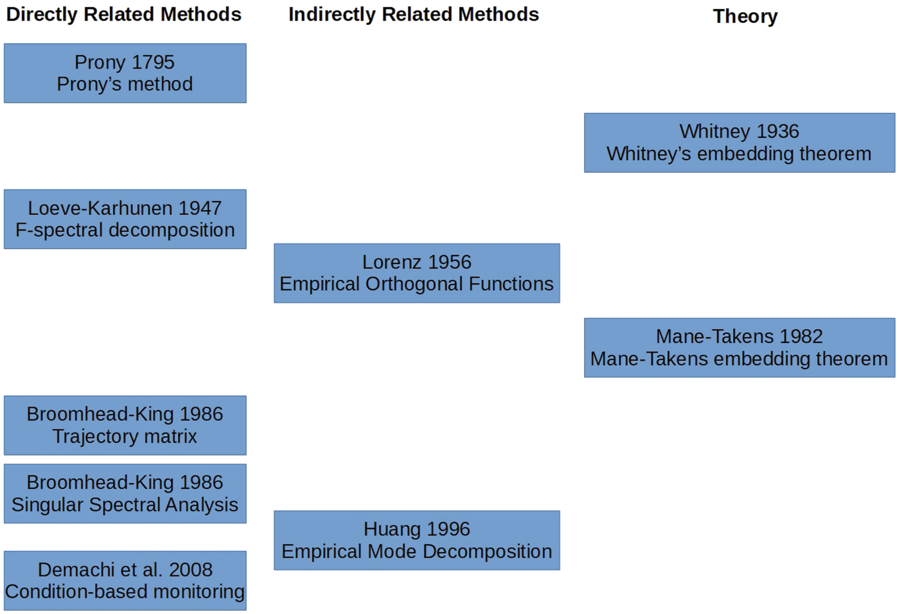


Fig. 1. A historical outline for SSA is shown from inception in 1795 with Prony's method to the first time-series analysis application.

Historically, SSA is a sub-space based method or latent variable model (LVM) based on the Whitney embedding theorem that states that any function from an n-dimensional manifold to an m-dimensional manifold can be approximated by a smooth embedding as long as $m > 2n$ [19]. For time-series data, this was extended into the delay embedding theorem of Mane-Takens after Pickard inspired the field with his paper entitled the *Geometry of time series*. A historical outline of SSA and related approaches are presented in Fig. 1, with SSA often referred to as the PCA for time series. As proposed by [14], SSA is principal component analysis (PCA) applied to the time domain, or also known as the Karhunen–Loeve decomposition. It is related to spectral decomposition, empirical orthogonal function (EOF) and empirical mode decomposition (EMD) [20–23]. SSA is an alternative to wavelet shrinkage to denoise mixtures of trends, transients and Gaussian noise that have only recently been employed in condition monitoring applications [24–27].

In Sect. 2, the SSA method is introduced. The process of applying SSA to an experimental signal from a damaged gearbox is studied in Sect. 3, followed by a conclusion of the study and recommendations for future work in Sect. 4.

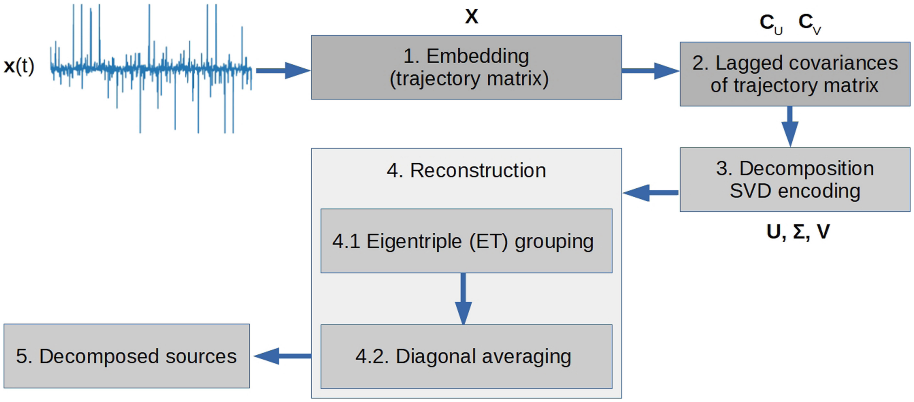


Fig. 2. SSA is shown for a single measured vibration signal that is 1) embedded, 2) lagged covariance matrices are constructed, 3) whereafter the decomposition step is performed. After that, 4) diagonal averaging and 5) the reconstruction of source components is performed.

2 Singular Spectral Analysis (SSA)

The general five steps of SSA, shown in Fig. 2, are listed:

1. embedding a discrete signal $\mathbf{x} = [x_1, \dots, x_N] \in \mathbb{R}^N$,
2. computing the lagged covariance matrices,
3. decomposition through a singular-value decomposition (SVD),
4. reconstruction of source components through
   (a) eigentriple (ET) grouping,
   (b) diagonal averaging, and
5. analysis of the decomposed sources and reconstructed signal $\tilde{\mathbf{x}} \in \mathbb{R}^N$.

These five steps are now discussed in detail.

2.1 Embedding

Given a signal $x(t)$ that is discretely sampled N times to obtain $\mathbf{x} = [x_1, \dots, x_N] \in \mathbb{R}^N$, which is subsequently embedded as a two-dimensional trajectory, lag or Hankel matrix $\mathbf{X} \in \mathbb{R}^{L \times K}$. K lagged vectors of dimension window length L are arranged as columns in the embedding matrix to form the L-trajectory matrix

$$\mathbf{X} = \begin{pmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & \cdots & x_N \end{pmatrix},$$

where the first column is the first L components of the original signal, the second column is the one-stepped lag vector of the first column etc. The columns are referred to as L-lagged vectors. Alternatively, each row represents a K-lagged vector of dimension K, where $K = N - L + 1$.

The embedding matrix is also known as the trajectory or Hankel matrix. Here, $2 < L < N/2$, should be chosen sufficiently large to reconstruct the minimum frequency of interest and to cover the maximum periodicity of interest [28].

2.2 Lagged Covariances

The lagged-covariance matrices of the embedding matrix are then calculated as $\mathbf{C}_U = \mathbf{X}\mathbf{X}^{\mathrm{T}} \in \mathbb{R}^{L \times L}$ and $\mathbf{C}_V = \mathbf{X}^{\mathrm{T}}\mathbf{X} \in \mathbb{R}^{K \times K}$.

2.3 Decomposition

The eigendecomposition of $\mathbf{C}_U = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^{\mathrm{T}}$ and $\mathbf{C}_V = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}$ reconstructs

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}^{\mathrm{T}}\mathbf{V}, \tag{1}$$

with the singular values ordered in the $L \times K$ diagonal matrix $\boldsymbol{\Sigma} = [\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L]$. Consequently, the SVD decomposes the embedding matrix $\mathbf{X}$ into a sum of rank-one matrices $\mathbf{X} = \sum_{i=1}^{L} \bar{\mathbf{X}}_i = \sum_{i=1}^{L} \sigma_i \mathbf{U}_i \mathbf{V}_i^{\mathrm{T}}$, where $\bar{\mathbf{X}}_i$ is the elementary matrix generated from the i-th eigentriple (ET) set, $(\sigma_i, \mathbf{U}_i, \mathbf{V}_i)$. The standard deviation $\bar{\sigma}$ contribution of the ith ET is given by:

$$\bar{\sigma}_i = \frac{\sqrt{\sigma_i}}{\sum_{i=1}^{L} \sqrt{\sigma_i}} \tag{2}$$

2.4 Reconstruction

SSA specifically focuses on grouping ETs to be informative. The elementary matrices are usually grouped based on singular and period spectra criteria [29].

ETs grouping is based on the separability measure of variance to untangle the original signal into the trend, weak fluctuations and noise. Sources of the signal are extracted based on their contributions towards explaining the variance of the original signal.

Instead of reconstructing the entire signal, SSA first divides the elementary matrices $\bar{\mathbf{X}}_i$ into M groups. In each group, the elementary matrices are added together to give a resultant matrix $\mathbf{R}_i \in \mathbb{R}^{L \times K}$. Hence, the embedded matrix can now be computed from $\mathbf{X} = \sum_{j=1}^{M} \mathbf{R}_j$.

Groups are selected based on information to detect weak separability. This is contained in the w-correlation matrix, which contains weighted cosines of angles between the reconstructed time-series components as entries. The w-correlation matrix is a standard approach to identify weak separability between elementary time-series components [30]. Strongly correlated elementary time-series components are usually grouped.

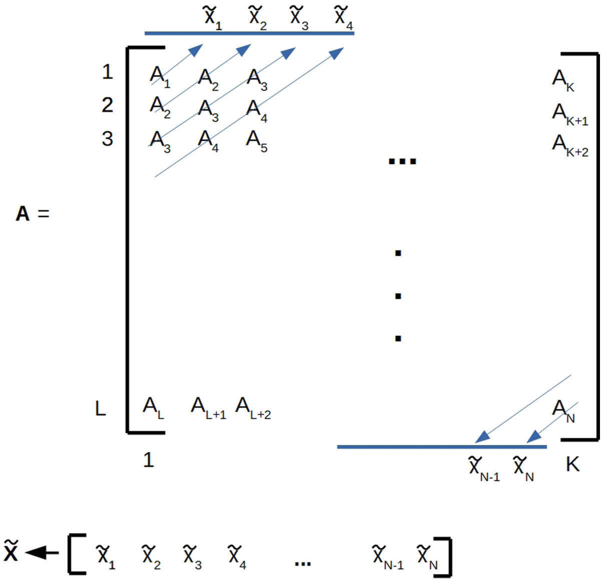


Fig. 3. The diagonal averaging depicted to reconstruct contributions of some L-trajectory matrix **A** to the signal **x**.

Diagonal Averaging (DA) is performed, as depicted in Fig. 3 on some L-trajectory matrix $\mathbf{A} \in \mathbb{R}^{L \times K}$, which represents either an elementary or resultant matrix. As shown, by computing the diagonal averages, DA reconstructs the contribution of $\mathbf{A} \in \mathbb{R}^{L \times K}$ to $\tilde{\mathbf{x}} \in \mathbb{R}^N$.

Hence, when this diagonalisation operation or Hankelisation operator is applied to the resultant matrices $\mathbf{R}_j, j = 1, \ldots, M$, it computes the contribution of each resultant matrix towards the reconstructed signal $\tilde{\mathbf{x}}_{\mathbf{j}}$.

2.5 Decomposed Sources

The original signal is now decomposed into trends, weak fluctuations and noise as represented by the reconstructed ETs that can be analysed and interpreted for condition-based maintenance applications. It is important to note that the decomposition uses variance as the proxy to find informative ETs, with the grouping of the ETs, influenced by both variance and periodicity of the ETs.

3 Experimental Case Study

The experimental test rig in the Centre for Asset Integrity Management laboratory of the University of Pretoria is shown in Fig. 4. The setup consists of an electrical motor that drives the system and an alternator to control the applied load to the system using independent speed controllers. Four varying operating

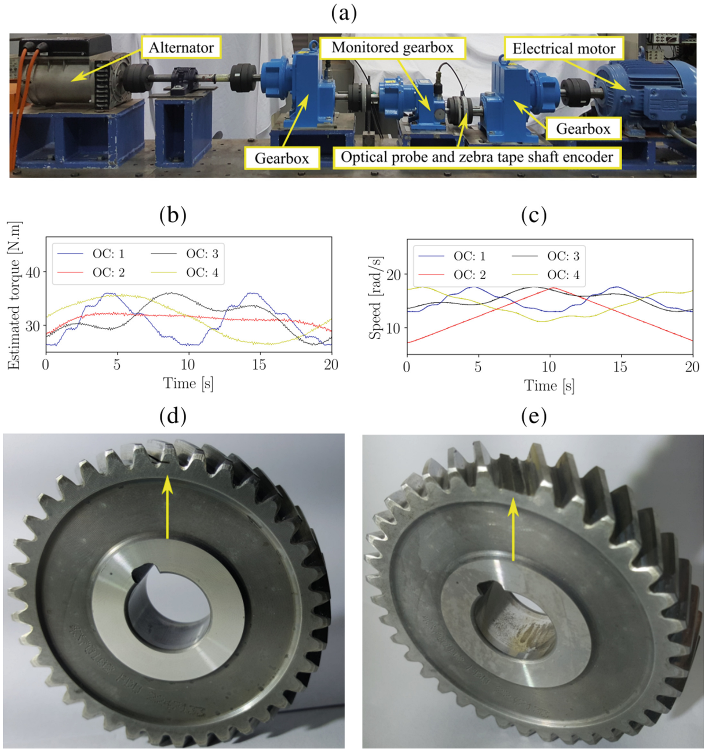


Fig. 4. (a) Experimental test rig. (b) The rotational speed of the input shaft of the monitored gearbox. (c) Load at the input shaft of the monitored gearbox. (d) Gear before the experiment started (0% complete). (e) Gear after the experiment was completed (100% complete).

conditions (OC1 to OC4) are shown in Figs. 4(b) and 4(c). The test rig also contains three helical gearboxes, with the centre gearbox being monitored for damage. For the measurements considered in this work, the gear was damaged with a seeded slot shown in Fig. 4(d) and operated under varying operating conditions until the gear tooth failed, as shown in Fig. 4(e).

The measurement under consideration was acquired after the experiment was approximately 90% complete and used to illustrate the SSA process's important steps. The raw signal, $\mathbf{x}$, of the segment considered in this study is depicted in Fig. 5.

3.1 SSA of the Experimental Signal

One measurement, acquired from OC 1 in Fig. 4, is used to evaluate the SSA method's performance (Sect. 2).

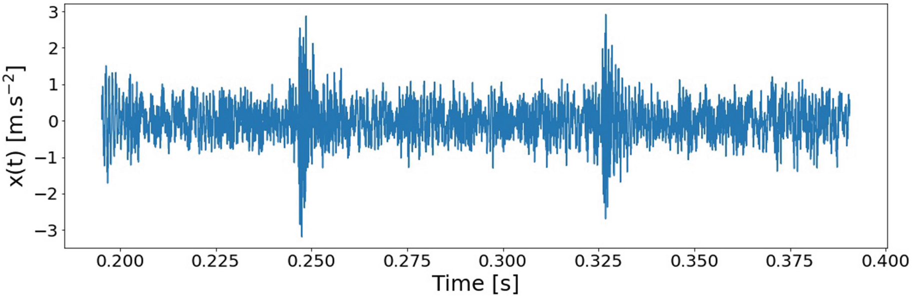


Fig. 5. Raw time-series signal considered in this study.

The trajectory matrix shown in Fig. 7 is computed from the raw time-series signal $\mathbf{x}$, shown in Fig. 5. The raw time-series signal covers 0.195 s and contain 5000 samples, giving a sampling rate (SR) of 25.6 kHz and a Nyquist frequency of 12.8 kHz. In turn, the lowest detectable frequency f_0 is given by

$$f_0 = 5\frac{SR}{L}. \tag{3}$$

A window length of $L = 256$ was used, resulting in frequency $f_0 = 500\,\mathrm{Hz}$ and a time resolution of around 0.002 s.

Lagged covariance matrices are computed from the trajectory matrix.

Eigendecomposition of the lagged covariance matrices gives the ETs in the form of $L \times L$ and $K \times K$ unitary matrices and the $L \times K$ rectangular diagonal matrix with the L singular values of the trajectory matrix.

Reconstruction of the sources requires the computation of the elementary matrices from the ETs. The variance contributed to reconstructing the trajectory matrix by elementary matrix is shown in Fig. 6. Similar variances are shared amongst some elementary matrices that include $\left\{\bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1\right\}$, $\left\{\bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3, \bar{\mathbf{X}}_4, \bar{\mathbf{X}}_5\right\}$ and $\left\{\bar{\mathbf{X}}_6, \bar{\mathbf{X}}_7, \bar{\mathbf{X}}_8, \bar{\mathbf{X}}_9, \bar{\mathbf{X}}_{10}, \bar{\mathbf{X}}_{11}\right\}$.

The first thirty computed elementary matrices from the ETs are depicted in Fig. 8. Although the elementary matrices are mainly distinct, similarities are exhibited amongst some matrices that include $\{\bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1\}$, $\{\bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3\}$, $\{\bar{\bar{\mathbf{X}}}_4, \bar{\mathbf{X}}_5\}$, $\{\mathbf{X}_6, \bar{\mathbf{X}}_7\}$ and $\{\bar{\mathbf{X}}_8, \bar{\mathbf{X}}_9\}$. An additional analysis of the elementary matrices are shown in Fig. 9, which confirms the coherence in the period spectrum amongst $\{\bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1\}$, $\{\bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3\}$, $\{\bar{\mathbf{X}}_4, \bar{\mathbf{X}}_5\}$, $\{\bar{\mathbf{X}}_6, \bar{\mathbf{X}}_7\}$ and $\{\bar{\mathbf{X}}_8, \bar{\mathbf{X}}_9\}$. The energy in the spectra decreases as the variance explained decreases.

Grouping of the ETs is usually done based on strong correlations in the w-correlation matrix. The w-correlation matrix is depicted in Fig. 10. From

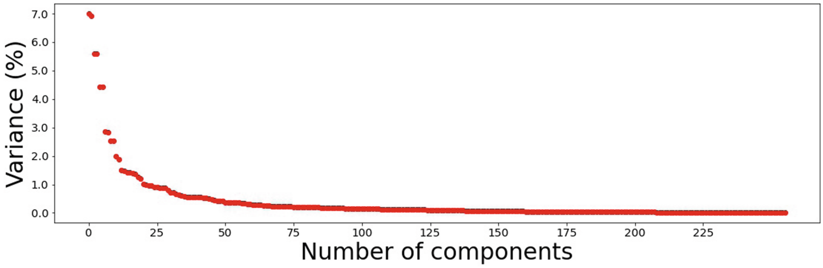


Fig. 6. The relative contributions of elementary matrix $\bar{\mathbf{X}}_i$ in reconstructing the trajectory matrix $\mathbf{X}$.

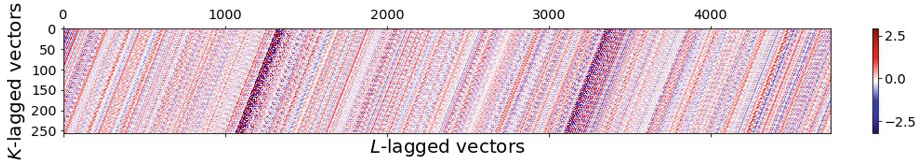


Fig. 7. The trajectory matrix, $\mathbf{X}$.

the w-correlation matrix, 19 resultant matrices $\mathbf{R}_0$–$\mathbf{R}_{18}$, are identified. The components for four selected resultant matrices include $\mathbf{R}_0 : \left\{\bar{\mathbf{X}}_0, \bar{\mathbf{X}}_1\right\}$, $\mathbf{R}_1 : \left\{\bar{\mathbf{X}}_2, \bar{\mathbf{X}}_3, \bar{\mathbf{X}}_4, \bar{\mathbf{X}}_5\right\}$, $\mathbf{R}_3 : \left\{\bar{\mathbf{X}}_8, \bar{\mathbf{X}}_9\right\}$, and $\mathbf{R}_4 : \left\{\bar{\mathbf{X}}_{10}, \bar{\mathbf{X}}_{11}\right\}$.

The magnitude of the transfer functions for all the groupings is depicted in Fig. 11. Just above 7.5 kHz, groups 1, 5, 14, 15, 16, 18 highlight energy in the signal. These frequency bands are associated with the dominant impulsive components that impede damage detection [13]. Groups 1, 5, 16 and 18 have well-separated energy, whereas, for groups 14 and 15, energy is present at lower and higher frequencies.

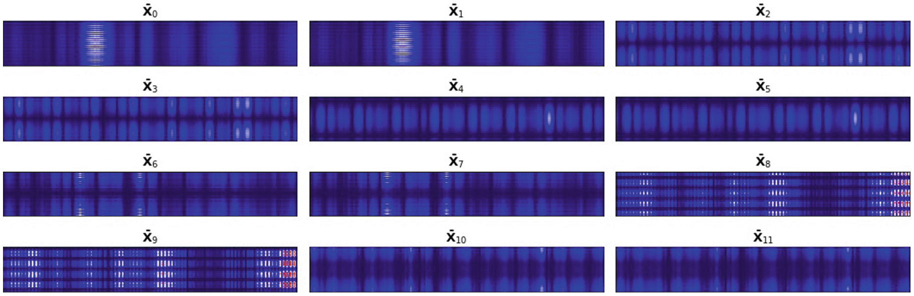


Fig. 8. The magnitude of the first twelve elementary matrices. The x and y axes of each subfigure represent the rows and columns of its associated elementary matrix.

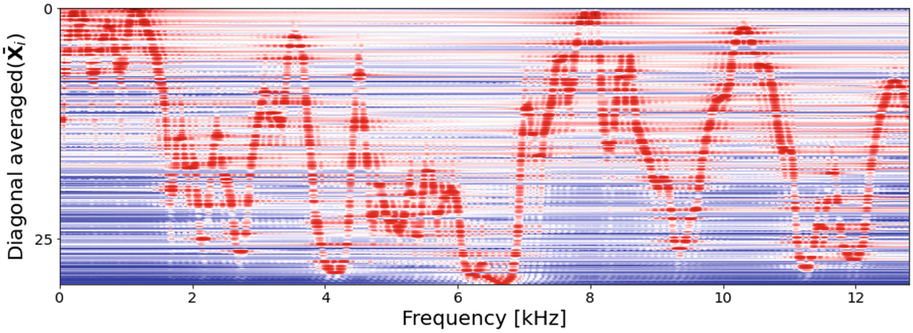


Fig. 9. The magnitude of the fast Fourier transforms (FFTs) of the diagonal averaged elementary matrices $\bar{\mathbf{X}}_i$ for $i = 0, \ldots., 29$, with a window length of $L = 256$.

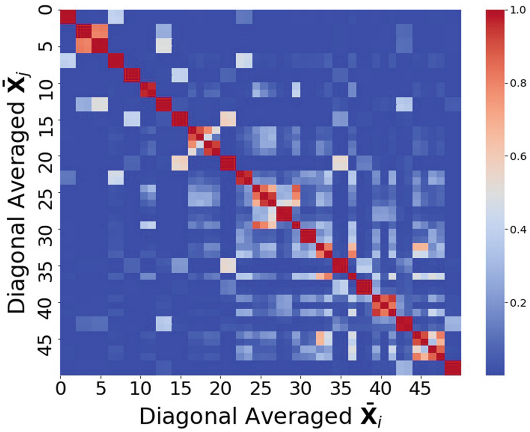


Fig. 10. The weighted correlation ($w-$correlation) matrix indicates the weighted correlation amongst the reconstructed signal of elementary matrices $\bar{\mathbf{X}}_i$.

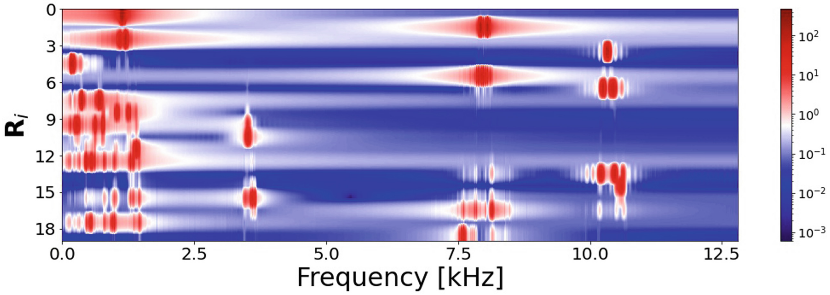


Fig. 11. The magnitude of the transfer functions of the first 19 diagonal averaged resultant matrices.

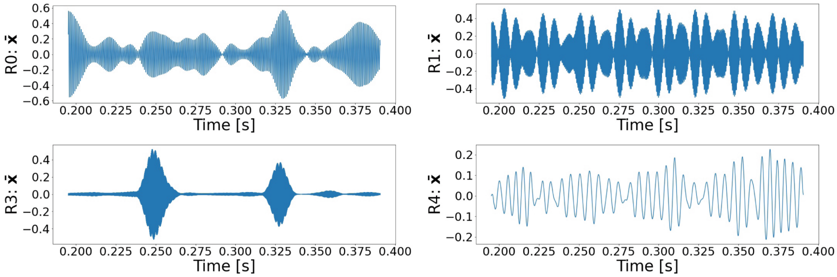


Fig. 12. The diagonalised averaged resultant matrices for $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{R}_3$ and $\mathbf{R}_4$.

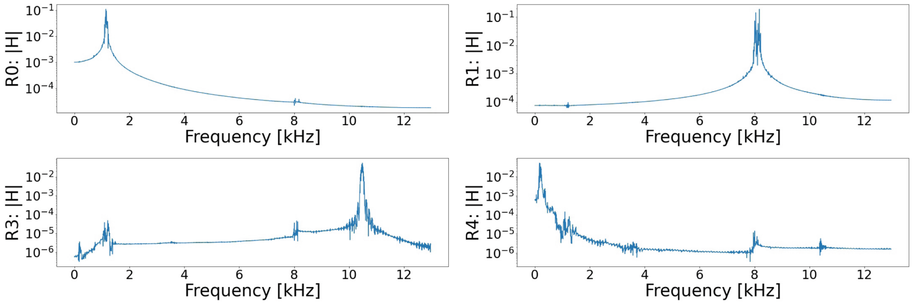


Fig. 13. The magnitude of the transfer functions for the diagonal averaged resultant matrices $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{R}_3$ and $\mathbf{R}_4$.

Consider the diagonalised averaged resultant matrices for $\mathbf{R}_0$, $\mathbf{R}_1$, $\mathbf{R}_3$ and $\mathbf{R}_4$ in Fig. 12, with their respective transfer function magnitudes depicted in Fig. 13. SSA can identify different signal components in the data. This enables the different signal components (e.g. impulsive components at 8 kHz) to be isolated in the signal by using the variance as a basis for decomposition. This is useful for condition monitoring as weak and damaged components can be identified and isolated.

4 Conclusion

Singular spectrum analysis (SSA) is demonstrated to separate a gearbox signal under time-varying operating conditions into predominantly isolated components. This is achieved by grouping elementary matrices to construct resultant matrices. The resultant matrices contribution to signal reconstruction is achieved through diagonal averaging. We showed that SSA could enhance interpretability under time-varying operating conditions given a suitable eigentriples (ETs) grouping. Future work will focus on studying additional separability measures to automatically group the ETs and enhance the interpretation of the components using signal processing techniques. Lastly, SSA offers a convenient framework that could be extended towards other linear latent variable techniques.

References

1. Stetco, A., Dinmohammadi, F., Zhao, X., Robu, V., Flynn, D., Barnes, M., Keane, J., Nenadic, G.: Machine learning methods for wind turbine condition monitoring: a review. Renew. Energy **133**, 620–635 (2019). https://www.sciencedirect.com/science/article/pii/S096014811831231X
2. Faris Elasha, D., Teixeira, J.A.: Condition monitoring philosophy for tidal turbines. Int. J. Perform. Eng. **10**, 521–534 (2014)
3. Atlason, R., Oddsson, G., Unnthorsson, R.: Geothermal power plant maintenance: evaluating maintenance system needs using quantitative kano analysis. Energies **7**, 4169–4184 (2014)
4. Mižáková, J., Mižák, J., Piteľ, J.: Monitoring of operating conditions of biomass combustion process. Appl. Mech. Mater. **308**, 153–158 (2013)
5. Sait, A.S., Sharaf-Eldeen, Y.I.: A review of gearbox condition monitoring based on vibration analysis techniques diagnostics and prognostics. In: Proulx, T. (ed.) Rotating Machinery, Structural Health Monitoring, Shock and Vibration, vol. 5, pp. 307–324. Springer, New York (2011)
6. Abboud, D., Baudin, S., Antoni, J., Rémond, D., Eltabach, M., Sauvage, O.: The spectral analysis of cyclo-non-stationary signals. Mech. Syst. Sig. Process. **75**, 280–300 (2016)
7. Schmidt, S., Heyns, P.S., Gryllias, K.C.: An informative frequency band identification framework for gearbox fault diagnosis under time-varying operating conditions. Mech. Syst. Sig. Process. **158**, 107771 (2021)
8. Kruczek, P., Zimroz, R., Antoni, J., Wyłomańska, A.: Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech. Syst. Sig. Process. **159**, 107737 (2021)
9. Schmidt, S., Zimroz, R., Heyns, P.S.: Enhancing gearbox vibration signals under time-varying operating conditions by combining a whitening procedure and a synchronous processing method. Mech. Syst. Sig. Process. **156**, 107668 (2021)
10. Xu, N., Li, X.: Intelligent fault diagnosis methodology under varying operating conditions using multi-layer domain adversarial learning strategy. Int. J. Dyn. Control 2195–2698 (2021)
11. Abboud, D., Antoni, J., Sieg-Zieba, S., Eltabach, M.: Envelope analysis of rotating machine vibrations in variable speed conditions: a comprehensive treatment. Mech. Syst. Sig. Process. **84**, 200–226 (2017)
12. Antoni, J., Xin, G., Hamzaoui, N.: Fast computation of the spectral correlation. Mech. Syst. Sig. Process. **92**, 248–277 (2017)
13. Schmidt, S., Zimroz, R., Chaari, F., Heyns, P.S., Haddar, M.: A simple condition monitoring method for gearboxes operating in impulsive environments. Sensors **20**, 2115 (2020)
14. Broomhead, D., King, G.P.: Extracting qualitative dynamics from experimental data. Phys. D Nonlinear Phenom. **20**, 217–236 (1986). https://www.sciencedirect.com/science/article/pii/016727898690031X
15. Comon, P.: Independent component analysis, a new concept? Sig. Process. **36**, 287–314 (1994)
16. Kilundu, B., Chiementin, X., Dehombreux, P.: Singular spectrum analysis for bearing defect detection. J. Vibr. Acoust. **133**, 10 (2011)
17. He, Q., Yan, R., Fanrang, K., Du, R.: Machine condition monitoring using principal component representations. Mech. Syst. Sig. Process. **23**, 446–466 (2009)

18. Mazur, K., Borowa, A., Brdyś, M.: Condition monitoring using PCA based method and application to wastewater treatment plant operation. IFAC Proc. **39**, 208–213 (2006). 1st IFAC Workshop on Applications of Large Scale Industrial Systems. https://www.sciencedirect.com/science/article/pii/S1474667015326203
19. Adachi, M.: Embeddings and Immersions. Translations of Mathematical Monographs. American Mathematical Society (1993). https://books.google.co.za/books?id=vvkIuAEACAAJ
20. Huang, N.E., Long, S.R., Shen, Z.: The mechanism for frequency downshift in nonlinear wave evolution. In: Hutchinson, J.W., Wu, T.Y. (eds.) Advances in Applied Mechanics, vol. 32, pp. 59–117. Elsevier (1996). https://www.sciencedirect.com/science/article/pii/S0065215608700760
21. Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H.: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. Roy. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **454**, 903–995 (1998). https://royalsocietypublishing.org/doi/abs/10.1098/rspa.1998.0193
22. Huang, N.E., Hu, K., Yang, A.C.C., Chang, H.-C., Jia, D., Liang, W.-K., Yeh, J.R., Kao, C.-L., Juan, C.-H., Peng, C.K., Meijer, J.H., Wang, Y.-H., Long, S.R., Wu, Z.: On Holo-Hilbert spectral analysis: a full informational spectral representation for nonlinear and non-stationary data. Philos. Trans. Roy. Soc. A Math. Phys. Eng. Sci. **374**, 20150206 (2016). https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2015.0206
23. Wang, H., Ji, Y.: A revised Hilbert-Huang transform and its application to fault diagnosis in a rotor system. Sensors **18** (2018). https://www.mdpi.com/1424-8220/18/12/4329
24. Demachi, K., Rybalko, S., Yusa, N.: Diagnosis of damages of wings of rotor by electromagnetic field measurement. Int. J. Appl. Electromagnet Mech **28**, 313–320 (2008)
25. Garcia, D., Trendafilova, I.: Study on singular spectrum analysis as a data-driven technique for damage diagnosis. Comparison between time and frequency domain. In: MATEC Web of Conferences, vol. 148, p. 14003 (2018). https://doi.org/10.1051/matecconf/201814814003
26. García, D., Tcherniak, D.: An experimental study on the data-driven structural health monitoring of large wind turbine blades using a single accelerometer and actuator. Mech. Syst. Sig. Process. **127**, 102–119 (2019). https://www.sciencedirect.com/science/article/pii/S0888327019301530
27. Diao, Y., Sui, Z., Guo, K.: Structural damage identification under variable environmental/operational conditions based on singular spectrum analysis and statistical control chart. Struct. Control Health Monit. e2721. https://onlinelibrary.wiley.com/doi/abs/10.1002/stc.2721
28. Vautard, R., Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series. Phys. D Nonlinear Phenom. **35**, 395–424 (1989). https://www.sciencedirect.com/science/article/pii/0167278989900778
29. Zhigljavsky, A.: Singular spectrum analysis for time series, pp. 1335–1337. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2_521
30. Golyandina, N., Zhigljavsky, A.: Singular Spectrum Analysis for Time Series. Springer, Heidelberg (2013)

Experimental and Numerical Analysis of Frequency Responses of Sandwich Structures

Hamamed Najah[1,2](✉), Hentati Hamdi[1](✉), Bouaziz Slim[1], Haddar Mohamed[1],
El Guerjouma Rachid[2], and Yaakoubi Nourdin[2]

[1] Mechanics Modeling and Production Research Unit, National School of Engineers of Sfax, Sfax, Tunisia
[2] Laboratoire d'Acoustique de l'Université du Mans LAUM UMR CNR 6613, Le Mans Université, Avenue Olivier Messiaen, 72085 Le Mans, France

Abstract. Among the infinite field of micro system of sandwich structure applications, biomedical devices are mainly vital due to their effect on society and on our health. Many technologies have been industrialized and developed to provide efficient system, including nano and micro electro mechanical systems (NEMS and MEMS) for piezoelectric applications. In this work, experimental and numerical analysis of the frequency responses of two micro electro mechanical systems are presented. The micro piezoelectric cantilever beam was used as a micro electro mechanical system. It is very useful for low frequency vibration sensors. The second system is a transducer. It is as an audio prosthesis and consists of a MEMS membrane made of an isolating interface; a AlN film acting as acoustic active material of the transducer. The finite element analysis is implemented, calculating the natural frequencies and eigenmodes of sandwich structures. We present the effect of the thickness on frequency responses for the cantilever beam. We show a good agreement between numerical results and experimental ones obtained by laser vibrometer measurements.

Keywords: Frequency response · FEM · Experimental results · MEMS

1 Introduction

Dynamic behavior and fracture of components of mechanical structures are an important problem especially in the aeronautical and automobile fields due to dynamic loadings. Numerical and experimental studies are interested in this research field (Silberschmidt 2016, Wu et al. 2020). We are focused on looking at the dynamic behavior of micro-electro-mechanical systems (MEMS), particularly the piezoelectric sandwich structures (Heyliger and Saravanos 1995, Leung et al. 2008). Their dynamic response and mechanical behavior until damage due to vibrations and shocks has been less studied. At material level, the use of composite sandwich structures is developed in the last years, to different engineering problems. This type of structures has been widely used in aerospace

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 173–181, 2022.
https://doi.org/10.1007/978-3-030-85584-0_18

and automotive applications, because of their superior mechanical properties, such as the strength-to-weight ratio. It represents an important factor in the choice of materials and dimensions in the design and manufacturing processes. Finite-element modeling is a powerful tool to predict the dynamic responses of microstructures (Wu and Lin 1990, Shu and Della 2004, Umesh and Ganguli 2008). It allows the optimization of structures before the design and realization steps. Many studies have focused numerically and experimentally on predicting mechanical characteristics of sandwich structures (Tsopanos et al. 2010, Mines et al. 2013). These structures are very complex and include imperfections over the surface of micro beams and different dimensions. These faults affect the mechanical responses of the MEMS structure.

Indeed, this study presents the finite element modeling of two sandwich structures: cantilever beam and a membrane structure. We analyze their frequency responses and their Eigen modes. The finite element (FE) model is implemented in COMSOL software. In addition, an experimental analysis of frequency responses is presented in order to define the vibration response of the both structures. With the aim to validate the numerical model, a comparison between the numerical and experimental results was developed.

2 Geometrical and Material Properties

Two structure will be studied. The firs geometry (Fig. 1) is a cantilever beam.

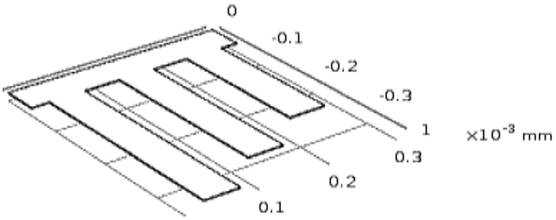


Fig. 1. Cantilever beam geometry

It is made up of two layers that composed on silicon and Cobalt-Nickel with 2.5 µm and 1µm respectively as thickness.

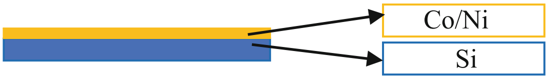


Fig. 2. Different layers composing the cantilever beam

Table 1 represents mechanical characteristics of the cantilever beam materials:

Table 1. Mechanical characteristics of the cantilever beam materials

	Silicon	Co/Ni
Young modulus	170 MPa	180 MPa
Poisson’s ratio	0.28	0.35

The second geometry is the membrane piece, which is shown in Fig. 3.

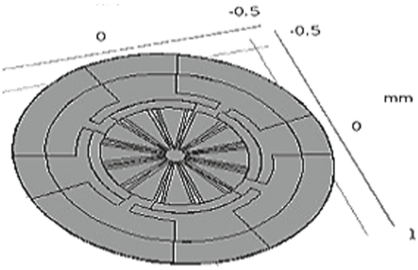


Fig. 3. Membrane geometry

This structure is a MEMS transducer adapted for in vivo implantation as audio prosthesis. The transducer is a piezoelectric sandwich structure. It contains a silicon (Si) membrane with 2 μm as thickness and 500 μm as radius, an isolating oxide interface layer with 2 μm as thickness and 400 nm as radius. It contains also two platinum (Pt) electrodes (thickness of 2 μm and radius of 150 nm), a piezoelectric layer of aluminum nitride (AlN) (thickness of 1 μm and radius of 400 μm) and a passivation layer for device protection Si3N4/SiO2 (700 nm/400 nm).

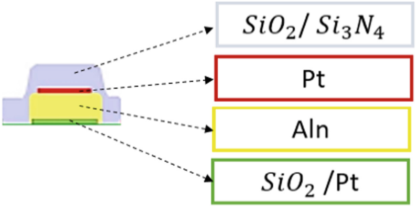


Fig. 4. Different layers composing the membrane

Table 2 represents mechanical characteristics of this membrane:

Table 2. Mechanical characteristics of the cantilever beam

	Pt	Aln	SiO2	Si3N4
Young modulus	100 MPa	169 MPa	70 MPa	250 MPa
Poissons's ratio	0.38	0.28	0.17	0.28

3 Numerical Models

The modelling and study of free vibration response of sandwich MEMS are carried out through the commercial FE code COMSOL Multiphysics 5.3. The FE model consists of two parts: cantilever beam and piezoelectric sandwich structure. The contact between the different parts is considered perfect. The free vibration analysis is carried out, obtaining the displacement and the eigenvalues. The cantilever beam structure is discretized using a fine triangle element and a swept mesh with a distribution (21786 elements). It is fixed from a part and free in the other. Furthermore, in order to validate the membrane sandwich discretization with tetrahedral elements, we implemented the FE model using a 3D element to mesh this structure (29218 elements). In this case, an extra fine tetrahedral volumetric mesh was used. As boundary conditions, the edge of the membrane Si is fixed.

The validation of experimental results (Hamamed et al. 2021), which are taken by the MSA-500 Micro System Analyzer for the membrane and the cantilever beam structures was released by numerical stimulation.

In the first step, a comparison between numerical and experimental results for the 3D and beam structures is carried out, in order to validate the both numerical models. Subsequently, we show the first three eigenmodes of the two undamaged structures. In the second step, the values of the displacement (µm) for the three first modes shape and the frequency responses in the undamaged structures will be used to analyze.

4 Comparison Between Numerical and Experimental Results

4.1 Cantilever Beam Structure

As mentioned above, a simplified model is developed in order to reduce the computational cost. We consider the 2D-node beam elements to represent the Cantilever beam structure. With the aim to validate the beam model, a comparison between the results obtained with numerical and experimental tests is developed. By considering the first three eigenmodes, the comparison is made by determine the Eigenfrequencies of vibration for the beam. Two values of the thickness of beam 250 µm and 300 µm) are studied. The different results are shown in Fig. 5.

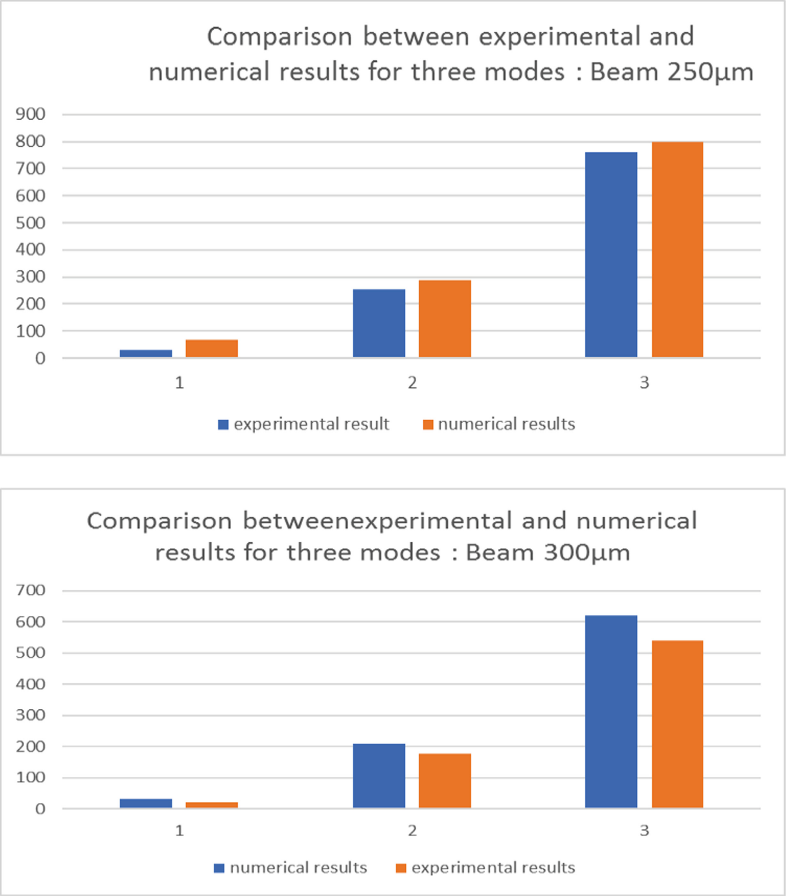


Fig. 5. Comparison between experimental and numerical Eigenfrequencies values

We represent in Fig. 6, the displacement fields for the first three modes of the cantilever beam structures.

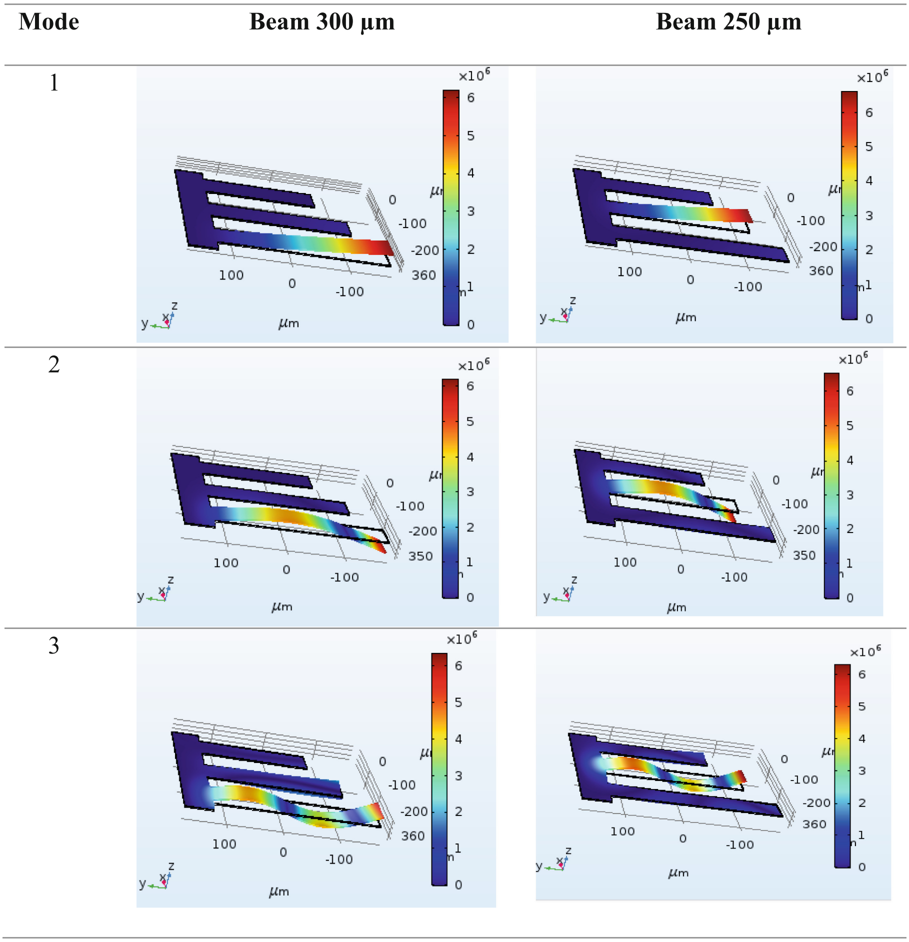


Fig. 6. Displacement (µm) for the first tree modes shape of the cantilever beam

4.2 Membrane Structure

We illustrate in Fig. 7, the first three eigenmodes for the piezoelectric sandwich structure.

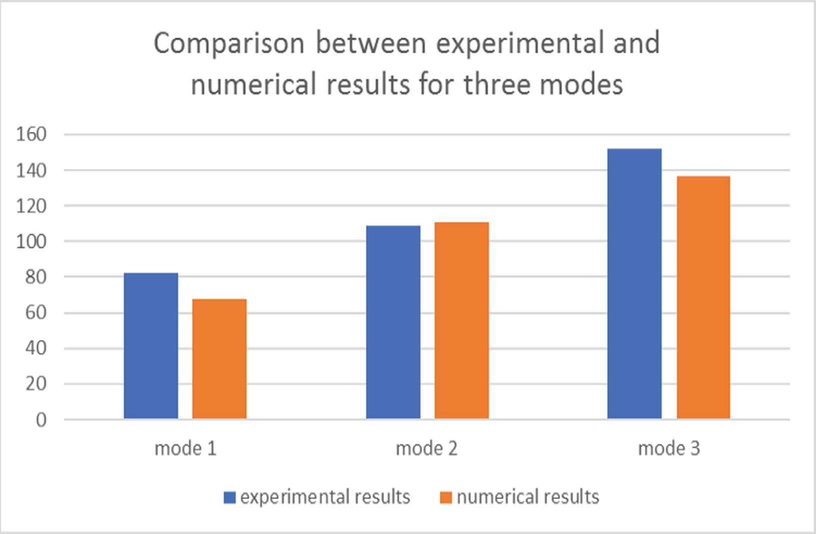


Fig. 7. Comparison between experimental and numerical Eigenfrequencies values

We noticed that the thickness of the silicon layer is the most influencing parameter. For that, the numerical results are sensible for the choice of this geometric parameter. The small difference between experimental and numerical value of the thickness of the silicon layer may modify our results.

Furthermore, after validate the 3D model, we represent in Fig. 8, the displacement fields for the first three modes of the piezoelectric sandwich structure.

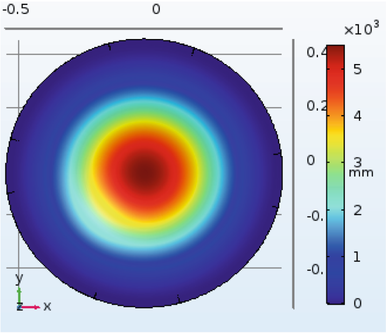
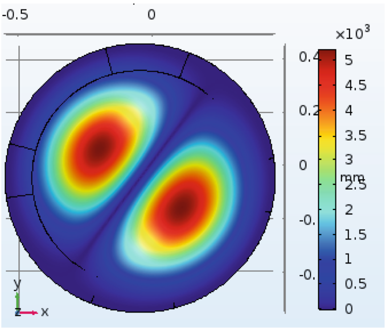
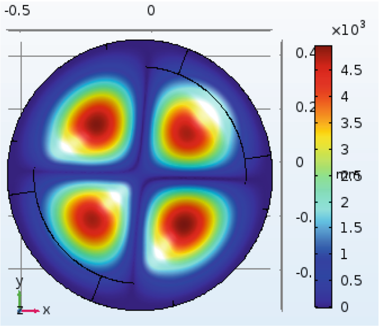
Mode	Shape mode
1	
2	
3	

Fig. 8. Displacement (μm) for the first tree modes shape of the piezoelectric sandwich structure

5 Conclusion

In this work, we have used COMSOL Multiphysics software to design sandwich structures geometries and to validate those models experimentally. The vibration responses of the different sandwich structures was calculated by the finite element method. The different experimental tests are taken by the MSA-500 Micro System Analyzer. It measure the vibrations for different micro structures.

We showed in this paper a good agreement between the 2D and the 3D simulation and experimental results given by the laser vibrometer. This small difference between different results is explained by the fact that we do not know exactly all the properties of all layers (dimensions, homogeneity, isotropy…).

References

Silberschmidt, V.V.: Dynamic Deformation, Damage and Fracture in Composite Materials and Structures. Woodhead Publishing (2016). https://doi.org/10.1016/B978-0-08-100080-9.00001-4

Wu, X., Meng, X., Zhang, H.: An experimental investigation of the dynamic fracture behavior of 3D printed nacre-like composites. J. Mech. Behav. Biomed. Mater. (2020). https://doi.org/10.1016/j.jmbbm.2020.104068

Heyliger, P., Saravanos, D.A.: Exact free-vibration analysis of laminated plates with embedded piezoelectric layers. J. Acoust. Soc. Am. **98**(3), 1547–1557 (1995). https://doi.org/10.1121/1.413420

Leung, A.Y.T., Zheng, J.J., Lim, C.W., Zhang, X., Xu, X.S., Gu, Q.: A new symplectic approach for piezoelectric cantilever composite plates. Comput. Struct. **86**(19–20), 1865–1874 (2008). https://doi.org/10.1016/j.compstruc.2008.04.005

Wu, J.S., Lin, T.L.: Free vibration analysis of a uniform cantilever beam with point masses by an analytical-and-numerical-combined method. J. Sound Vib. **136**(2), 201–213 (1990). https://doi.org/10.1016/0022-460x(90)90851-p

Shu, D., Della, C.N.: Free vibration analysis of composite beams with two non-overlapping delaminations. Int. J. Mech. Sci. **46**(4), 509–526 (2004). https://doi.org/10.1016/j.ijmecsci.2004.05.008

Umesh, K., Ganguli, R.: Shape and vibration control of a smart composite plate with matrix cracks. Smart Mater. Struct. **18**(2), 025002 (2008). https://doi.org/10.1088/0964-1726/18/2/025002

Tsopanos, S., Mines, R., McKown, S., et al.: The influence of processing parameters on the mechanical properties of selectively laser melted stainless steel microlattice structures. J. Manuf. Sci. E-T ASME 2010 **132**(4), 041011 (2010)

Mines, R., Tsopanos, S., Shen, Y., et al.: Drop weight impact behaviour of sandwich panels with metallic micro lattice cores. Int. J. Impact Eng. **60**, 120–132 (2013)

Campanella, H., Camargo, C.J., Lopez Garcia, J., Daza, A., Urquiza, R., Esteve, J.: Thin-film piezoelectric mems transducer suitable for middle-ear audio prostheses. J. Microelectromech. Syst. **21**(6), 1452–1463 (2012). https://doi.org/10.1109/jmems.2012.2211571

Hamamed, N., Hentati, H., Bouaziz, S., Haddar, M., El Guerjouma, R., Yaakoubi, N.: Numerical validation of experimental results for the dynamic behavior of sandwich structures. In: IEEE International Multi-Conference on Systems, Signals and Devices (2021, in press)

Hybrid Diagnostics and Prognostics of Planetary Gearboxes

Douw G. Marx(✉), P. Stephan Heyns, and Stephan Schmidt

Centre for Asset Integrity Management, University of Pretoria, Lynnwood Road,
Hatfield, Pretoria 0002, South Africa
{stephan.heyns,stephan.schmidt}@up.ac.za

Abstract. The effective condition monitoring of planetary gearboxes is important for making informed maintenance decisions and minimizing machine downtime. Traditional physics-based and data-driven approaches for condition monitoring of planetary gearboxes have inherent shortcomings that limit their practical application for remaining useful life prediction. Hybrid approaches show unique potential for performing diagnostics and prognostics of planetary gearboxes in situations where failure data is not available. Hybrid methods incorporate both physics-based and data-driven models that exploit their respective advantages. This work starts with a brief overview of the various challenges associated with vibration-based condition monitoring of planetary gearboxes and the characteristics of planetary gearboxes that complicate the condition monitoring process. Thereafter, the potential of hybrid methods to address these challenges is discussed. The respective advantages and disadvantages of physics-based and data-driven approaches are summarized and a short review of the use of hybrid methods in condition monitoring is presented. Finally, various considerations for developing hybrid methods for planetary gearbox diagnostics and prognostics are discussed. This includes considerations for selecting an appropriate hybrid framework, pre-processing of the measured vibration signal, physics-based modelling of a planetary gearbox, model calibration, health state estimation, and health state prediction.

Keywords: Planetary gearboxes · Diagnostics · Prognostics · Hybrid methods

1 Vibration-Based Condition Monitoring of Planetary Gearboxes

Condition-based maintenance (CBM) of planetary gearboxes has become increasingly important in recent years. The rapidly growing wind energy industry requires that the planetary gearboxes used in wind turbines operate reliably with a minimum amount of downtime [1]. CBM can reduce maintenance cost by eliminating unnecessary preventive maintenance operations [2] and increasing machine availability [3].

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 182–197, 2022.
https://doi.org/10.1007/978-3-030-85584-0_19

Consequently, the development of CBM strategies that can identify damaged components, diagnose fault severity, and predict the remaining useful life (RUL) of planetary gearboxes are an important field of research [4–6].

A common goal of CBM for planetary gearboxes is to infer the gearbox fault severity measurement using a diagnostics model, whereafter the RUL is predicted from the inferred health state by a prognostics model.

Vibration-based condition monitoring is a popular measurement technique for rotating machines such as gearboxes. Vibration-based condition monitoring relies on the assumption that a machine's vibration signature will be altered when the machine health changes [5].

However, the diagnostics and prognostics of planetary gearboxes remain challenging. The overall success of RUL prediction depends on how accurately the heath state is inferred and our ability to predict the future machine operation and machine degradation. This challenge is exacerbated by the unique structure of planetary gearboxes, which invalidate established condition monitoring methods developed for conventional fixed-axis gearboxes [4].

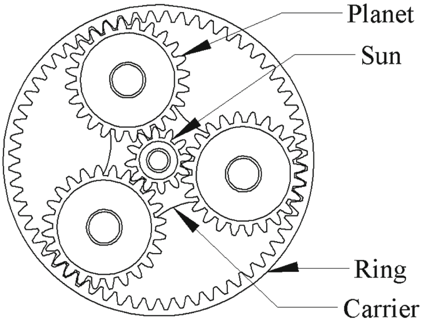


Fig. 1. A schematic of the important components in a planetary gearbox.

Characteristics that complicate the condition monitoring process arise from the planetary gearbox design as shown in Fig. 1. These characteristics include [4]:

1. Planetary gearboxes have multiple planet gears that simultaneously mesh with the sun and ring gears. The presence of multiple synchronous excitation sources can lead to the cancellation of vibration components attributed to damage.
2. There are multiple, time-varying vibration transmission paths between the origin of vibration and an accelerometer mounted to the gearbox housing. This could mask subtle changes in the vibration signature caused by damaged components.
3. The frequency spectra of both healthy and damaged planetary gearboxes is characterised by asymmetric sidebands around the gear mesh frequency and

its harmonics. The presence of sidebands around the gear mesh frequency and its harmonics can therefore not be used to diagnose faults in planetary gearboxes in the same way as done in conventional fixed axis gearboxes.

4. The high transmission ratio of many planetary gearboxes leads to gearbox components rotating at low rotational speeds. Low-frequency components associated with these low rotational speeds are generally masked by heavy noise, further complicating fault diagnostics.

The above-mentioned characteristics of planetary gearboxes require that more advanced condition monitoring techniques should be applied for diagnosing faults and predicting the RUL.

2 Motivation for Hybrid Methods

Two of the most common model types used for diagnostics and prognostics are data-driven models and physics-based models.

Data-driven models rely on previously observed data to create a mapping between measured vibration and the underlying health state (diagnosis) and RUL (prognosis). In contrast, physics-based models are mathematical models derived from first principles, that model the dynamics and degradation of a damaged gearbox. Unknown parameters for the physics-based models can be inferred in a model calibration process. Usually, this is performed with statistical methods such as Bayesian updating or regression [7].

Physics-based and data-driven models have unique advantages and disadvantages for diagnostics and prognostics which are summarized in Table 1.

The purpose of hybrid methods also referred to as integrated methods, is to overcome the shortcomings of the individual models and improve their performance by exploiting their respective advantages. There are several variants of hybrid methods [7] but the combination of physics-based and data-driven models appear to have unique advantages for the planetary gearbox diagnostics and prognostics problem.

For instance, hybrid methods can exploit the extensive research [10] on physics-based vibration and degradation models. Furthermore, data-driven methods can be used to quantify uncertainty in the model, the inferred model parameters and ultimately, the predicted RUL.

In fact, several authors have applied hybrid methods for CBM of different machines with various failure modes. For fixed axis gearboxes, hybrid methods have been studied for the crack growth [11,14,15], and surface wear [16] failure modes. For planetary gearboxes, hybrid methods have been applied to sun gear cracks [17], planet carrier cracks [18,19], planet gear cracks [20] and surface wear [20,21]. Hybrid methods have also been successfully applied outside the domain of gearbox CBM in for instance crack growth in aircraft fuselage panels [22] and Lithium-Ion batteries [23].

Although the hybrid methods applied in the above-mentioned investigations serve as proof that hybrid methods can be effectively be applied for CBM, in

Table 1. The advantages and disadvantages of physics-based and data-driven models

Data-driven	Physics-based
Examples: Bayesian state estimation [8], machine learning methods [9]	**Examples:** Finite element models, Lumped mass models [3,10]
Advantages – Can model non-linear, non-monotonic degradation [7] – An in-depth understanding of the complex physics that governs the degradation is not required [7]	**Advantages** – Can be used in circumstances where failure data is not available – Allows for the incorporation of the operating conditions of the machine [11]. This makes physics-based methods valuable when working with time-varying operating conditions or for making predictions for operating conditions or health states for which training data is not available – Model output has a physical meaning, is understandable and explainable
Disadvantages – Black-box models do not explain the inherent physics of the problem [7]. This makes this approach ill-suited for extrapolation since the influence of a fault on the machine response is not incorporated in the model – Requires large amounts of representative training data over the entire life cycle of the machine [7]. This is generally not available with insufficient historical fault data and many potential modes of failure – Is not guaranteed to generalize well to operating conditions for which training data is not available [12]	**Disadvantages** – Requires in-depth domain knowledge [13] – Physics-based models are intrinsically deterministic. As a result, they cannot be used in isolation for uncertainty quantification and risk assessment – It is difficult to estimate model parameters through the optimisation of non-linear physics-based models [13] – A given physics-based model applies to only a small set of problems [13] and requires extensive modeling effort to create

most cases, either the methods used to infer the health state or the methods used to predict RUL, rely on the availability of historical failure data.

The remainder of this work is dedicated to discussing the challenges and considerations associated with applying hybrid methods to planetary gearbox prognostics where an abundance of healthy data is available, but no failure data is available.

Section 3.1 lists considerations for choosing an appropriate hybrid framework and Sect. 3.2 discusses potential pre-processing of the measured vibration signal. Thereafter, model calibration and health state estimation considerations are discussed in Sect. 3.4 with RUL prediction considerations listed in Sect. 3.5. Finally, recommendation for future work is presented in Sect. 4.

3 Considerations for Using Hybrid Methods for Planetary Gearbox Prognostics

This section lists different considerations for developing hybrid prognostics methods for planetary gearboxes. Figure 2 shows a roadmap of this section based on the hybrid framework suggested in Sect. 3.1. The section numbers associated with a given component are indicated on the figure.

3.1 Hybrid Framework

For the CBM of planetary gearboxes, the prediction of the RUL can be divided into two main sub-problems, namely diagnosis and prognosis (see Fig. 2).

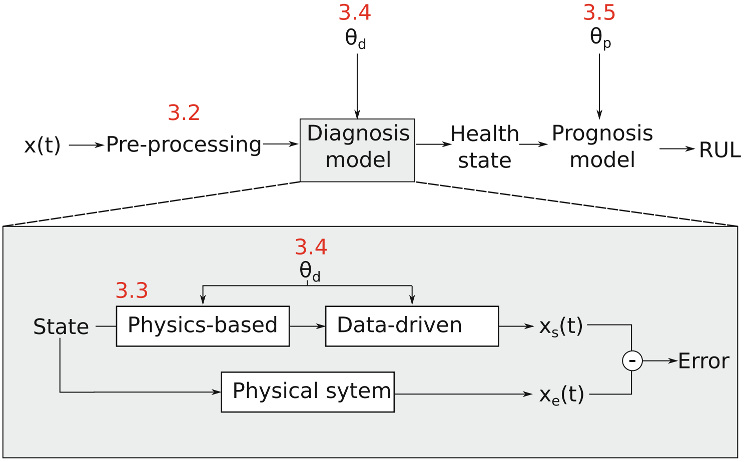


Fig. 2. A hybrid framework for diagnostics and prognostics in planetary gearboxes.

In the diagnosis phase, the current health state (e.g. the crack length in the tooth of a planet gear) is inferred from vibration data (Sect. 3.4). Health state estimation models are therefore used to define a mapping between the (possibly pre-processed, Sect. 3.2) measured condition monitoring data and the health state of the gearbox.

This mapping can consist of an ensemble of physics-based models, a data-driven model, or a combination of physics-based and data-driven models in the form of a hybrid model.

If the health state estimation model is data-driven, failure data is required to establish a mapping between the measured variable and the underlying health state. Although prior work have successfully implemented health state estimation models [18,23,24], knowledge of the underlying heath state was required to create this mapping. Consequently, if failure data is not available, the health state estimation model has to be physics-based or hybrid, incorporating a physics-based component, and a data-driven component trained on healthy data.

For a hybrid health state estimation model, it is important that phenomena modeled by the data-driven component is not dependent on the underlying health state. This would ensure that a data-driven component of the health state estimation model would remain applicable under degraded health state conditions, even though it is trained on healthy data only. Of course, it is also possible to skip the health state estimation step altogether and construct a data-driven mapping between the measured gearbox vibration response and the RUL. However, this would only be possible if large amounts of failure data is available.

The second sub-problem of the CBM process is remaining useful life estimation (Sect. 3.5). If the diagnostics model is capable of reliably estimating the health state from the condition monitoring measurements, a health state prediction or prognostics model can be used to predict future health states, and ultimately, the RUL.

Unless failure data is available and a data-driven state prediction model can be used, a physics-based health state prediction model needs to be used. Therefore, if failure data is not available, an in-depth understanding of the physics of failure is required so that a health state prediction model can be defined. Fortunately, for a planetary gearbox, failure modes like crack growth are accurately modeled by the Paris law [18] and the surface wear failure mode is modeled by the Archard wear model [21]. In circumstances where a failure model has not yet been derived for a given failure mode, physics-based health state prediction is more difficult and a degradation model such as exponential degradation must be assumed.

Additional to the health state estimation and health state prediction models, a hybrid prognostics model can incorporate pre-processing (Sect. 3.2) and uncertainty quantification components. The measured vibration signal is often contaminated with noise and signal components that do not contribute useful information related to the underlying health state. To remove these signal components, the measured variable can be pre-processed before health state estimation. For physics-based models, pre-processing can involve signal processing that removes signal components and noise that are not explicitly modelled in the physics-based model. For data-driven models, pre-processing can include the reduction of the dimensionality of the measured variable through the calculation of features that are expected to correlate well with a change in the gearbox health state.

Additional to signal processing, an uncertainty quantification component can be included in a hybrid method.

Bayesian state estimation techniques such as particle filters and Kalman filters can be used to update the health state and model parameter probability densities as health state measurements become available through the health state estimation model [15,18,21,22]. In this way, measured data under damaged conditions can be incorporated into the model without the availability of run-to-failure data.

Considerations for each of the respective components of a hybrid prognostics framework are now discussed in more detail.

3.2 Pre-processing of the Measured Vibration Signal

Signal processing and analysis methods form an integral part of conventional fault diagnosis methods [4]. In the context of hybrid prognostics of planetary gearboxes, signal processing methods can also be used as the first step in a hybrid framework. Signal processing enhances the fault information in the signal and for some techniques, such as the synchronous average, it ensures that physics-based models do not have to account for complex non-synchronous phenomena that are difficult to model or do not carry fault information.

However, excessive pre-processing could impede the identifiably of the damage component since a non-unique representation exists between the damage and the processed data. This means that the underlying health state cannot be uniquely inferred resulting in an erroneous prediction of the RUL. A vector of features extracted from vibration data is therefore mostly useful when used in conjunction with a data-driven model or when the vector is compared to the same features from a physics-based model response.

A potential pre-processing technique that preserves physics-based meaning in planetary gearboxes is the time-synchronous average [25]. The goal of the time-synchronous average is to separate the vibration components related to the planet gears from the overall vibration of the machine. This technique assumes that for small measurement windows, the periodic transfer function that leads to amplitude changes in the vibration signal as the planet gear passes the accelerometer remains constant [25].

Therefore, if physics-based meaning should be preserved in the measured signal, it can make sense to calculate the synchronous average as an initial pre-processing step for planetary gearboxes prognostics frameworks that employ physics-based health state estimation models. Alternatively, if a data-driven model trained on healthy data is used as the first component of the state estimation model, a vector of features that are expected to correlate well with the damage state can be calculated as a pre-processing step.

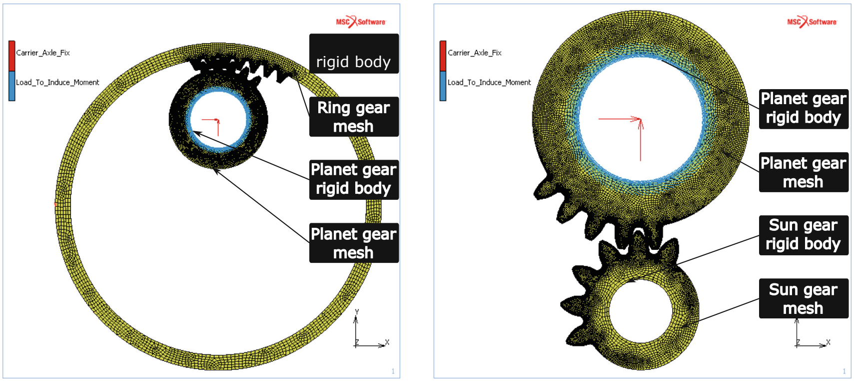
3.3 Physics-Based Modelling

Physics-based models are particularly useful in hybrid prognostics models where historical fault data is not available. Examples of physics-based models that are useful for planetary gearbox prognostics are shown in Figs. 3 and 4. Figure 3a shows a crack growth simulation to calculate the crack growth path and stress intensity factors in a planet gear. The crack growth path is compared to an experimental crack growth path shown in Fig. 3b. The simulated crack profile can then be used in a subsequent time-varying mesh stiffness (TVMS) simulation (Fig. 4) with the stess intensity factors being used in RUL prediction using the Paris crack propagation law.

In this section, considerations for the physics-based modelling of a planetary gearbox are presented.

(a) Simulated crack growth (b) Experimental crack growth

Fig. 3. Crack growth simulation vs. experimental crack growth.



(a) Ring-planet interaction (b) Sun-planet interaction

Fig. 4. TVMS simulation for a planetary gearbox

Physics-based models are useful for defining both the mapping between the health state severity and the measured gearbox response as well as the mapping between the current health state severity and the expected RUL of the gearbox.

Two methods are typically used to model the vibration response of planetary gearboxes subject to a given failure mode and fault severity. These include phenomenological models and dynamic, lumped mass models (LMM). With phenomenological models, the gearbox is described using algebraic equations based on empirical observations. On the other hand, with physics-based dynamic models laws of the gearbox vibration are defined by differential equations [26]).

Although phenomenological models have been successfully applied to explain the frequency and time domain response of planetary gearboxes [27,28], the parameters that govern the output of phenomenological models do not have physical meaning. Phenomenological models are therefore capable of demonstrating the way in which a fault would manifest in the vibration response, but are not useful for inferring the underlying health state of the planetary gearbox.

It is therefore important that the physics-based models used in the health state estimation model are defined by physical parameters so that the underlying health state can be inferred from the measured vibration response. A dynamic lumped mass model is, therefore, a more appropriate choice of physics-based model for hybrid models since there is a connection between the physical parameters of the gearbox and the model response [29].

An example of where lumped mass models are useful for modelling the vibration response is the crack growth failure mode. The physics of cracked gear systems have been studied extensively [10]. A fatigue crack in a gear tooth will lead to a decrease in gear tooth stiffness [30] where the stiffness reduction is a function of the crack length and crack path in the gear tooth. This reduced gear tooth stiffness will in turn influence the time-varying mesh stiffness and lead to a vibration response that is different from that of an undamaged gear system [10,31].

When the TVMS for a damaged gear pair is known, it can be used in an LMM to model the expected vibration response of the gearbox due to the excitation caused by the TVMS.

Considerations that have to be kept in mind when applying physics-based models in the health state estimation part of a prognostics framework are now presented.

- A commonly used family of 2D dynamic lumped mass models for planetary gearboxes is based on that of Kahraman [32] which has been used with some adaption by others [29,33–36]. For a planetary gearbox with three planet gears, a planar LMM with three DOFs for each lumped mass will have 18 degrees of freedom in total. Many degrees of freedom make the numerical solution of the non-linear differential equations that govern the problem, very expensive to compute. This also complicates the solution of the inverse problem (see Sect. 3.4) where the crack length should be inferred from the vibration response.
- All translations of lumped masses in the planetary gearbox LMM are typically defined relative to a rotating reference frame fixed to the planet carrier. As such, the vibration response obtained from the LMM is not directly comparable to the response measured by an accelerometer mounted to the gearbox

housing. Consequently, a transfer function is required to model the transmission path between the gearbox components in the rotating reference frame and the measured response at the accelerometer [37]. This can be modeled in different ways [29,38], but ideally, the complex physics associated with this transmission path would be modeled by a data-driven model trained on healthy data. Alternatively, this mapping could be omitted entirely if torsional vibration is measured rather than axial vibration on the gearbox housing.

- The choice of the appropriate degree of model fidelity is further important. Different authors extend the basic capabilities of planetary gearbox LMMs to model various physical phenomena such as gyroscopic and centrifugal forces [29] or the influence of a flexible ring gear [36]. Similarly, the time-varying mesh stiffness for a faulty gear can be modeled using various analytical and finite element methods [10] of different fidelity. However, as the model fidelity is increased, the required modelling effort and computational cost of solving for the model response also increases. The extent to which increased model fidelity contributes to the accuracy of health state estimation should therefore be critically evaluated. Since there are many sources of uncertainty in the modelling of the planetary gearbox, an increased model fidelity of a given model component could be rendered useless by a different, low fidelity model component.
- Model inadequacy [39] will be present if the physics-based model is not flexible enough to model the complexities of the true system. Although high fidelity models could potentially model more complex systems, it is unlikely that a model would ever perfectly describe a planetary gearbox.

For different failure modes, the physics of the health state degradation is well understood and a physics-based model can be used during health state prediction. For the crack growth failure mode, for instance, the Paris law [40] can be used to predict the crack length in the gear tooth after some time has elapsed.

3.4 Model Calibration and Health State Estimation

After an appropriate model is created to serve as a health state estimation model, the model must be calibrated to ensure that it is representative of reality. Since data in damaged conditions are not available, the model is calibrated on healthy data only, relying on the physics-based components of the model to generalize to damaged conditions. The health state estimation model has many unknown parameters. These unknown parameters either need to be chosen by experts, or they should be inferred through the calibration process.

The calibration process involves solving the inverse problem of finding the optimal model parameters given the pre-processed measured vibration response. After the model is calibrated, the current health state can be inferred in a similar way with all other model parameters remaining fixed at the values obtained from model calibration.

There are several considerations that need to be kept in mind when calibrating the health estimation model.

- For each degree of freedom, there are unknown mass, stiffness, and damping values that must either be chosen and fixed or inferred from measurement. The large number of design variables in the optimisation problem presents some challenges. It is difficult to find the optimal values for many free parameters in the optimisation problem, and it becomes easy to over-fit the model to the data. Furthermore, if the initial conditions for the differential equation need to be inferred from the measured vibration, the optimisation problem is further complicated since each degree of freedom in the LMM requires a displacement and velocity initial condition. Initial conditions can be eliminated from the set of free parameters in the optimisation problem by starting with random initial conditions and discarding the transient part of the solution when comparing the model response to the measured response. For example, the first 50 mesh cycles of a solution can be discarded to obtain the steady-state response of an LMM [41]. Depending on how quickly steady state is reached, it can be computationally expensive to evaluate the solution for a long enough time period for all transients to die out. For complex models, the response can also be sensitive to the choice of initial conditions. [42] show that an LMM of a multi-stage planetary gear system exhibits complex motion that leads to period-doubling, demonstrating that the system is sensitive to the choice of the initial condition.
- The model parameters and health state are not necessarily identifiable. This means that there is uncertainty in the true system parameters and that these parameters cannot be uniquely be inferred from the measured system response [43]. The positioning of accelerometers on the gearbox housing and the time-varying vibration transfer paths are examples of factors that can influence the parameter identifiably.
- The total number of parameters can be reduced by setting parameters that are expected to be equal to the same free parameter or making use of proportional damping to eliminate some of the unknown damping free parameters. Additionally, a sensitivity analysis [44] can be conducted to determine which parameters most greatly affect the model response. The most sensitive parameters can then be used as free parameters in the calibration problem with fewer sensitive parameters remaining fixed.
- During the optimisation process, many function evaluations are required to compute the optimisation cost associated with each candidate solution. For a planetary gearbox lumped mass model, each function evaluation requires the solution of a differential equation. The solution to the differential equation is computationally expensive and reduces the feasibility of using high fidelity LMMs with many degrees of freedom.
- Choosing a good starting point for the optimisation algorithm and defining accurate parameter bounds is critical for converging to the global minimum in a reasonable amount of time. Relative constraints between parameters can be added to express an understanding of the relative size of model parameters. Furthermore, the use of multiple starting points can increase the chances of converging to a global optimum.

- The TVMS used in the LMM also contributes to the calibration challenges. The TVMS introduces a non-linearity in the LMM differential equation. This non-linearity requires that small solution time steps should be taken by a numerical differential equation solver, increasing the overall computational expense of model calibration.
- The solution to LMM tends to be sensitive to certain model parameters such as damping coefficients. In the context of solving the optimisation problem of fitting the LMM to measured data, the algorithm could easily propose a candidate solution that would lead to an unstable LMM response. Although this candidate solution would typically incur a high optimisation cost and would typically not be selected as the optimum, the computation of the unstable solution with the numerical differential equation solver can take a long time.

3.5 Remaining Useful Life Prediction

If the current health state can be successfully inferred through the heath state estimation model, the health state prediction model can be used to predict the remaining useful life of the machine. For the crack growth failure mode, the propagation of a crack in isotropic and linear elastic materials can be modeled by the Paris law [40]. The Paris law is a physics-based model that defines a relationship between the crack growth rate and stress state in materials with cracks in the stable crack growth region.

The simplest variant of the Paris law is given by

$$\frac{da}{dN} = C(\Delta K)^m \tag{1}$$

where $\frac{da}{dN}$ is the crack growth rate, ΔK is the stress intensity range, and C and m are experimentally determined constants.

To achieve accurate RUL predictions, the appropriate Paris law parameters C and m and stress intensity range ΔK must be used. Due to the unavailability of failure data, obtaining the model parameters, C and m from an experimental fit is not an option. Consequently, there exists a degree of uncertainty in the true Paris law parameters. To reduce this uncertainty, Bayesian sequential inference can be used to update prior probability densities of the Paris law parameters as more measurements of the crack length become available during the course of the machine's lifetime.

The sequential Bayesian inference problem consists of two parts. During model-based prediction, an a-priori estimate of a future state is obtained by projecting the current state through the discretized Paris crack growth law. As measurements containing information about the true system state become available, the current belief of the system state, as calculated by the prediction step, can then be updated with the measurement information in an update step [8].

An example of the mean and standard deviation for crack growth RUL predictions as computed by the Bayesian state estimation is shown in Fig. 5.

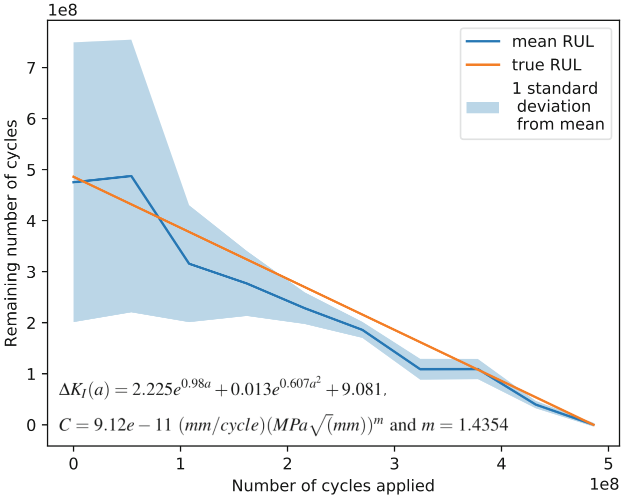


Fig. 5. RUL predictions versus true RUL

4 Recommendation for Future Work in Hybrid Prognostics of Planetary Gearboxes

From the set of considerations listed in the previous section, it is clear that the planetary gearbox prognostics problem present many challenges.

Prognostics approaches that incorporate physics-based health state prediction laws such as the Paris law or the Archard wear model together with Bayesian state estimation techniques are successful in predicting the RUL of a gearbox [15,16,18,21,22]. However, it is likely that the overall success of vibration-based hybrid prognostics methods applied to planetary gearboxes would largely be reliant on the ability of the state estimation model to infer the current health state from the vibration response.

As a result, the development and validation of physics-based or hybrid models capable of inferring the underlying health state in a planetary gearbox from a measured vibration signal can be most valuable to future research in planetary gearbox prognostics.

5 Conclusion

Hybrid approaches for prognostics of planetary gearboxes have the potential to outperform traditional physics-based or data-driven models under conditions where failure data is not available. However, several factors have to be considered when developing hybrid prognostics methods for planetary gearboxes.

Some of the considerations associated with the hybrid prognostics of planetary gearboxes include selecting an appropriate hybrid framework comprised of physics-based and data-driven components, pre-processing of the vibration signal, physics-based modeling of the planetary gearbox, model calibration, health state estimation and health state prediction.

References

1. Nie, M., Wang, L.: Review of condition monitoring and fault diagnosis technologies for wind turbine gearbox. Procedia CIRP **11**, 287–290 (2013)
2. Jardine, A.K.S., Lin, D., Banjevic, D.: A review on machinery diagnostics and prognostics implementing condition-based maintenance. Mech. Syst. Signal Process. **20**, 1483–1510 (2006)
3. Cubillo, A., Perinpanayagam, S., Esperon-Miguez, M.: A review of physics-based models in prognostics: application to gears and bearings of rotating machinery. Adv. Mech. Eng. **8**, 1–21 (2016)
4. Lei, Y., Lin, J., Zuo, M.J., He, Z.: Condition monitoring and fault diagnosis of planetary gearboxes: a review. Meas. J. Int. Meas. Confed. **48**, 292–305 (2014)
5. Salameh, J.P., Cauet, S., Etien, E., Sakout, A., Rambault, L.: Gearbox condition monitoring in wind turbines: a review. Mech. Syst. Signal Process. **111**, 251–264 (2018)
6. Wang, T., Han, Q., Chu, F., Feng, Z.: Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: a review. Mech. Syst. Signal Process. **126**, 662–685 (2019)
7. Liao, L., Köttig, F.: Review of hybrid prognostics approaches for remaining useful life prediction of engineered systems, and an application to battery life prediction. IEEE Trans. Reliab. **63**, 191–207 (2014)
8. Fang, H., Tian, N., Wang, Y., Zhou, M., Haile, M.A.: Nonlinear Bayesian estimation: from Kalman filtering to a broader horizon. IEEE/CAA J. Automatica Sinica **5**, 401–417 (2018)
9. Liu, R., Yang, B., Zio, E., Chen, X.: Artificial intelligence for fault diagnosis of rotating machinery: a review. Mech. Syst. Signal Process. **108**, 33–47 (2018)
10. Ma, H., Zeng, J., Feng, R., Pang, X., Wang, Q., Wen, B.: Review on dynamics of cracked gear systems. Eng. Fail. Anal. **55**, 224–245 (2015)
11. Zhao, F., Tian, Z., Bechhoefer, E., Zeng, Y.: An integrated prognostics method under time-varying operating conditions. IEEE Trans. Reliab. **64**, 673–686 (2015)
12. Kan, M.S., Tan, A.C.C., Mathew, J.: A review on prognostic techniques for non-stationary and non-linear rotating systems. Mech. Syst. Signal Process. **62**, 1–20 (2015)
13. Xia, T., Dong, Y., Xiao, L., Du, S., Pan, E., Xi, L.: Recent advances in prognostics and health management for advanced manufacturing paradigms. Reliab. Eng. Syst. Saf. **178**, 255–268 (2018)
14. Li, C.J., Lee, H.: Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics. Mech. Syst. Signal Process. **19**, 836–846 (2005)
15. Zhao, F., Tian, Z., Zeng, Y.: Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method. IEEE Trans. Reliab. **62**, 146–159 (2013)
16. Feng, K., et al.: Vibration-based updating of wear prediction for spur gears. Wear **426–427**, 1410–1415 (2019)
17. Cheng, Z., Hu, N., Zhang, X.: Crack level estimation approach for planetary gearbox based on simulation signal and GRA. J. Sound Vib. **331**, 5853–5863 (2012)

18. Orchard, M.E., Vachtsevanos, G.J.: A particle filtering approach for on-line failure prognosis in a planetary carrier plate. Int. J. Fuzzy Logic Intell. Syst. **7**, 221–227 (2007)
19. Patrick, R., et al.: An integrated approach to helicopter planetary gear fault diagnosis and failure prognosis. In: 2007 IEEE Autotestcon, pp. 547–552 (2007)
20. Peng, D., Smith, W.A., Borghesani, P., Randall, R.B., Peng, Z.: Comprehensive planet gear diagnostics: use of transmission error and mesh phasing to distinguish localised fault types and identify faulty gears. Mech. Syst. Signal Process. **127**, 531–550 (2019)
21. Zhao, F., Tian, Z., Liang, X., Xie, M.: An integrated prognostics method for failure time prediction of gears subject to the surface wear failure mode. IEEE Trans. Reliab. **67**, 316–327 (2018)
22. Coppe, A., Haftka, R.T., Kim, N.H., Yuan, F.: Uncertainty reduction of damage growth properties using structural health monitoring. J. Aircr. **47**, 2030–2038 (2010)
23. Liao, L., Köttig, F.: A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction. Appl. Soft Comput. **44**, 191–199 (2016)
24. He, D., Bechhoefer, E., Dempsey, P., Ma, J.: An integrated approach for gear health prognostics. Ann. Forum Proc. AHS Int. **3**, 1647–1655 (2012)
25. McFadden, P.D.: Window functions for the calculation of the time domain averages of the vibration of the individual planet gears and sun gear in an epicyclic gearbox. J. Vib. Acoust. **116**, 179–187 (1994)
26. Lei, Y., Liu, Z., Lin, J., Lu, F.: Phenomenological models of vibration signals for condition monitoring and fault diagnosis of epicyclic gearboxes. J. Sound Vib. **369**, 266–281 (2016)
27. McFadden, P.D., Smith, J.D.: An explanation for the asymmetry of the modulation sidebands about the tooth meshing frequency in epicyclic gear vibration. Proc. Inst. Mech. Eng. C J. Mech. Eng. Sci. **199**, 65–70 (1985)
28. Inalpolat, M., Kahraman, A.: A theoretical and experimental investigation of modulation sidebands of planetary gear sets. J. Sound Vib. **323**, 677–696 (2009)
29. Liang, X., Zuo, M.J., Hoseini, M.R.: Vibration signal modeling of a planetary gear set for tooth crack detection. Eng. Fail. Anal. **48**, 185–200 (2015)
30. Chaari, F., Baccar, W., Abbes, M.S., Haddar, M.: Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. Eur. J. Mech. A. Solids **27**, 691–705 (2008)
31. Belsak, A., Flasker, J.: Detecting cracks in the tooth root of gears. Eng. Fail. Anal. **14**, 1466–1475 (2007)
32. Kahraman, A.: Load sharing characteristics of planetary transmissions. Mech. Mach. Theory **29**, 1151–1165 (1994)
33. Lin, J., Parker, R.G.: Analytical characterization of the unique properties of planetary gear free vibration. J. Vib. Acoust. **121**, 316–321 (1999)
34. Chaari, F., Fakhfakh, T., Hbaieb, R., Louati, J., Haddar, M.: Influence of manufacturing errors on the dynamic behavior of planetary gears. Int. J. Adv. Manuf. Technol. **27**, 738–746 (2006)
35. Chen, Z., Shao, Y.: Dynamic simulation of planetary gear with tooth root crack in ring gear. Eng. Fail. Anal. **31**, 8–18 (2013)
36. Chen, Z., Zhu, Z., Shao, Y.: Fault feature analysis of planetary gear system with tooth root crack and flexible ring gear rim. Eng. Fail. Anal. **49**, 92–103 (2015)
37. Randall, R.B.: A new method of modeling gear faults. J. Mech. Des. **104**, 259–267 (1982)

38. Parra, J., Vicuña, C.M.: Two methods for modeling vibrations of planetary gearboxes including faults: comparison and validation. Mech. Syst. Signal Process. **92**, 213–225 (2017)
39. Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. J. Royal Stat. Soc. Ser. B (Stat. Methodol.) **63**, 425–464 (2001)
40. Paris, P., Erdogan, F.: A critical analysis of crack propagation laws. J. Basic Eng. **85**, 528–533 (1963)
41. Ambarisha, V.K., Parker, R.G.: Nonlinear dynamics of planetary gears using analytical and finite element models. J. Sound Vib. **302**, 577–595 (2007)
42. Xiang, L., Gao, N., Hu, A.: Dynamic analysis of a planetary gear system with multiple nonlinear parameters. J. Comput. Appl. Math. **327**, 325–340 (2018)
43. Villaverde, A.F.: Observability and structural identifiability of nonlinear biological systems. Complexity **2019** (2019)
44. Saltelli, A.: Making best use of model evaluations to compute sensitivity indices. Comput. Phys. Commun. **145**, 280–297 (2002)

Vibration Analysis of Planetary Gear Connected by Flexible Coupling Using Substructuring Methodology

Marwa Bouslema[1](✉), Taher Fakhfakh[1], Rachid Nasri[2], and Mohamed Haddar[1]

[1] Mechanical Modeling and Manufacturing Laboratory (LA2MP), National School of Engineers of Sfax, University of Sfax, B.P 1173, 3038 Sfax, Tunisia
`{tahar.fakhfakh,mohamed.haddar}@enis.rnu.tn`
[2] Applied Mechanics and Engineering, University of El-Manar II, 1002 Tunis, Tunisia
`rachid.nasri@enit.rnu.tn`

Abstract. The modeling of a reducer stage coupled the planetary stage through a flexible connection was established in this research. The studied system was made by lumped parameter model. The behavior of connecting elements affect the overall dynamic characteristics for instance natural frequencies and response characteristics. Therefore, the implementation of frequency based sub-structuring methodology (FBS) was aimed to identify the total dynamics behavior of a linked transmission system and to discuss the contribution of an elastic coupling on transmission system vibration. For this reason, this approach was considered for dual numerical issues with rigid and elastic coupling. A comparative examination of a rigid coupling results with one another flexible case was elaborated to discuss their sensitivity on the response. Then, different couplings models was suggested to test their impact on the sub-structuring response. In the interest of determining the goal of this research, a parametric study is conducted to inspect the sensibility of various values of flexible link for instance (kc = 1e+4, 1e+5 and 1e+6) on subsystems FRFs. As a conclusion, the obtained simulations illustrate that the elastic connection affects the achieve results of different subsystem in rotational direction. Additionally, the impact of coupling will relates to the selected of different subsystems.

Keywords: Transmission system · Sub-structuring technique · Different subsystems · Flexible coupling

1 Introduction

A mechanical system composed of many sub-systems, which are connected together through different categories of joints (for example, sliding joints, bolted joints, coupling, etc.). Mechanical coupling are essential parts in the planetary-coupled subsystems.

The global system response is depend on the FRF of all subsystem. Joints might have significant effects on the overall assembly behavior, such as natural frequencies, mode shapes, and response characteristics.

Previously, researchers like the machine tools field Erturk et al. (2006, 2009), have investigated the different joints properties. Researchers presented several approach for

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 198–206, 2022.
https://doi.org/10.1007/978-3-030-85584-0_20

identification of complicated systems. The receptance coupling (RC) was employed to linked subsystems with a joint component made from the FRFs Park (2003). The latter is applied to detect dynamic characteristics coupling. Numerous authors suggested this technique, for instance, Liu and Ewins (2002) have used their to discover the joint properties modeled by nonlinear spring. Others like Ewins (2000) have used the methodology of substructure for the identification of joints dynamic properties using coupled FRFs.

Recently, The singularity issue was implemented by El Mahmoudi et al. (2020) when considered flexible connection by means of substructures inside the (LM-FBS) coupling procedure. Zhao et al. (2016) showed that the transmitted torque amplitude was influenced by disc coupling stiffness in the field of compressor.

Indeed, the inspect of linked system is important for explaining the flexible connection impact. In fact, this kind of coupling minimizes the vibration.

In the other hand, Hmida et al. (2017) investigated the effect of introducing an elastic connection on the modal properties of spur gearbox system.

In the present paper, the receptance coupling like the process of the substructuring is tested in the situation of a reducer stage linked to a planetary gear stage through the flexible coupling to discover its dynamic characteristics.

The goal of this research is to identify the impact of the rotational coupling, which connects the subsystems together on the assembled system FRF.

First, the receptance technique is presented. Then, the studied system description is performed. Finally, a comparison among rigid and elastic coupling FRF responses is implemented.

2 Receptance Coupling Theory

The investigation of the RC method is concentrated on the dynamic result of a flexible-coupled system. The global system FRF is determined by compound the FRF of each subsystem.

2.1 Frequency Response Functions (FRFs)

Zhen et al. (2004) described the FRF coupling method. The motion equation of a linear subsystem is expressed by:

$$[M]\left\{\ddot{U}(t)\right\} + [K]\{U(t)\} = \{F(t)\} \tag{1}$$

where $[M]$ and $[K]$ are the mass and stiffness matrices of subsystem respectively. $\{U(t)\}$ and $\{F(t)\}$ denote the displacement and the excitation force, respectively. A Fourier transformation is applied to Eq. (1). The FRF of study system can be defined by realizing an FRF synthesis depend on a mode shapes and natural frequencies. The FRF matrix expression $H_{jk}(w)$ is defined by:

$$H_{qs}(w) = \sum_{r=1}^{n} \frac{{}_r\Phi_{q\,r}\Phi_s}{w_r^2 - w^2 + j2\xi_r w_r w} \tag{2}$$

where $H_{qs}(w)$ is the steady state displacement at the coordinate q due to a harmonic force excitation at the coordinate s, n is the total of dof. ${}_r\Phi_j$ is the mass-normalized mode shape.

2.2 Receptance Coupling Method

The principal idea of this technique is to utilize individual uncoupled subsystem FRFs to compose the entire system response via either a receptance or an impedance computational equation. The developing of coupling problem is expressed mathematically in Liu and Ewins (2002).

Jetmundsen (1988) expressed the receptance matrices of subsystems 'A' and 'B' in partitioned form as:

$$\begin{bmatrix} {}_AU_i \\ {}_AU_c \end{bmatrix} = \begin{bmatrix} {}_AH_{ii} \; {}_AH_{ic} \\ {}_AH_{ci} \; {}_AH_{cc} \end{bmatrix} \begin{bmatrix} {}_Af_i \\ {}_Af_c \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} {}_BU_j \\ {}_BU_c \end{bmatrix} = \begin{bmatrix} {}_BH_{jj} \; {}_BH_{jc} \\ {}_BH_{cj} \; {}_BH_{cc} \end{bmatrix} \begin{bmatrix} {}_Bf_j \\ {}_Bf_c \end{bmatrix} \tag{4}$$

The subscripts "*i*" and "j" indicate the uncoupled dof and the subscript "c" denotes the coupling dof. The connection of subsystem A to subsystem B by flexible connector at DoF c. A general expression of a both assembled system linked by flexible connector is presented as stated on Zhen et al. (2004):

$$\begin{bmatrix} H_{aa} & H_{ac} & H_{ab} \\ H_{ca} & H_{cc} & H_{cb} \\ H_{ab} & H_{bc} & H_{bb} \end{bmatrix} =$$

$$\begin{bmatrix} {}_AH_{ii} & {}_AH_{ic} & 0 \\ {}_AH_{ci} & {}_AH_{cc} & 0 \\ 0 & 0 & {}_BH_{jj} \end{bmatrix} - \begin{bmatrix} {}_AH_{ic} \\ {}_AH_{cc} \\ {}_BH_{ic} \end{bmatrix} \left[[{}_AH_{cc}] + [{}_BH_{cc}] + [K_c]^{-1} \right]^{-1} \begin{bmatrix} {}_AH_{ic} \\ {}_AH_{cc} \\ {}_BH_{ic} \end{bmatrix}^T \tag{5}$$

Where $[{}_AH]$ and $[{}_BH]$ are the FRF matrices of subsystems A and B.

$[K_c]$ is described by the coupling rigidity matrix within the coupling DoF of dual subsystems.

For the rigid joint, the matrix $[{}_AH_{cc}] + [{}_BH_{cc}] + [K_c]^{-1}$ replaced by $[{}_AH_{cc}] + [{}_BH_{cc}]$. The sub-matrices of assembled system where established by Liu and Ewins (2002). The FRF matrice was determinate using Eq. (5).

3 Studied System

A planetary gear related by flexible coupling forms the present transmission system. The studied system were modeled by adopting the lumped parameter model (Fig. 1). The planetary gear train components are supposed to be rigid bodies. Each one of these components has three DOFs. In this Figure, ui, vi, and wi ((i = r, s, c that can be the ring gear, sun, and the carrier, respectively) illustrate the displacements in two radials and rotational directions.

Two various dynamic models studied the coupling response. A rigid coupling model characterizes the start study. The coupling is modeled as a rigid disc, with equal parts at the coordinates related the coupling to the adjoining shafts corresponding to the coupling inertia. The second study used the connected system with an elastic coupling, which is approached by Krämer model with torsional stiffness (Kθ). The flexible coupling connected the second shaft reducer to the sun shaft.

Bouslema et al. (2017) established the coupling planetary system models related by a rigid coupling.

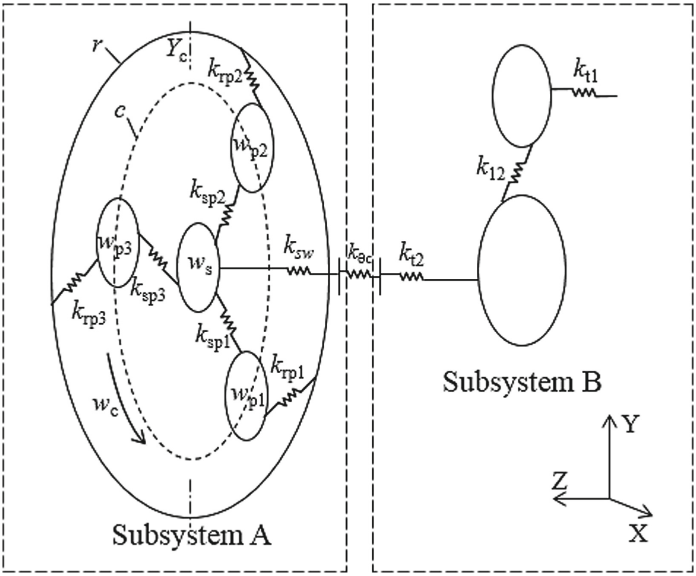


Fig. 1. A planetary gear linked by a flexible coupling to a reducer stage

4 Numerical Applications

The modeling of joint is explored for the coupling systems. The influence of flexible joint on different subsystems FRFs was discussed. Table 1 contained the studied system parameters.

Table 1. Parameters of parallel stage (i = 1, 2) and planetary gear.

	Sun	Ring	Carrier	Planet
Number of Teeth	30	70	-	20
Width of Teeth (mm)	25	25	-	25
Module	1.7	1.7	-	1.7
I/r^2 (Kg)	0.272	0.759	1.5	0.1
Mass (Kg)	0.46	0.588	3	0.177
Radius of base (m)	0.024	0.056	-	0.016
Helix Angle	0	0	-	0
Planet Gear mesh stiffness (N\m)	$K_{sp} = K_{rp} = 2.10^8$			
Bearing stiffness in u and v direction (N/m)	$K_p = K_{su,v} = K_{ru,v} = 10^8$			
Torsional stiffness of the ring, sun and carrier (N/m)	$k_{rw} = 10^9$; $K_{sw} = 10^5$; $K_{cw} = 0$			
Pressure angle (°)	s = r = 21.34			

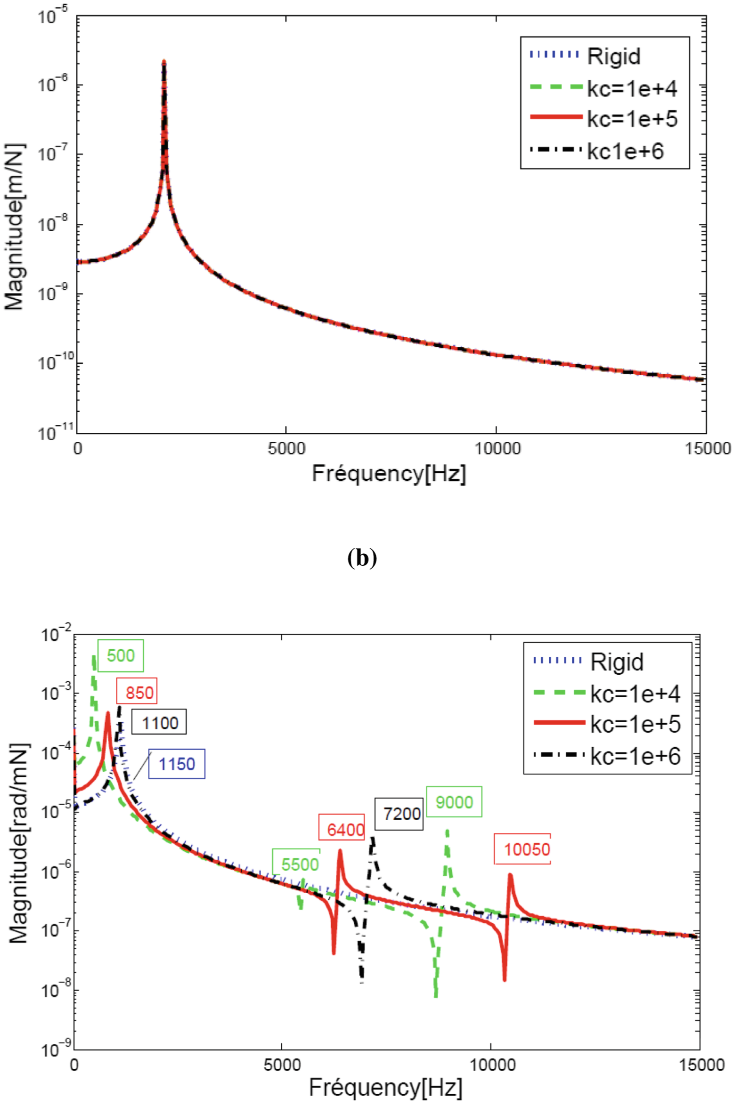
Wheel and pinion mass (kg)	$m_i = 2$
Wheel and pinion inertia moment (kg)	$I_i/(r_i)^2 = 0.58$
Wheel and pinion inertia moment (kg)	$J_i/(r_i)^2 = 1.16$
Torsional shaft flexibilities (N/m)	$k_{bpyi}/(r_i)^2 = 10^8$
Bearings stiffness (N/m)	$k_{yi} = 3.5 \times 10^8$, $k_{zi} = 10^8$
Pressure angle	$\alpha_1 = \alpha_2 = 20^\circ$
Module of Teeth (m)	$m = 4 \times 10^{-3}$
Average mesh stiffness (N/m)	$k_{1moy} = k_{2moy} = 2 \times 10^8$
Inertia coupling (kg m2)	$I_A = 4.48 \times 10^{-8}$

4.1 Coupling Stiffness Effect on the Coupled Subsystem (B) FRF

The sensibility of elastic connection on the frequencies values is investigated for the reducer stage example. Figure 2 depict the FRF of a coupled subsystem B with various stiffness values of coupling in the axial and the rotational directions of the second bearing and the shaft.

The four curves are superposed, which displays that the coupling rigidity don't affect the FRFs frequency of subsystem B in the axial Direction.

In the rotational direction, the both models of coupling represents a difference in the frequency. The decreasing of frequencies as reason of decreasing in coupling stiffness. The frequencies of connection shafts appear in the flexible case because of the inertia value as compared to the rigid case. The two frequency of two values connection shafts which appear in the flexible case for the three value of the torsional stiffness (Kc = 1e+4, 1e+5, 1e+6) are located respectively at (5500 Hz, 9000 Hz), (6400 Hz, 10050 Hz) and 7200 Hz. The high frequencies of the connection shaft was neglected in the rigid case. It is observed that when the rigidity of coupling is important, the system response is close to the rigid case.


(a)

(b)

Fig. 2. FRF of a coupled system B with rigid and flexible coupling: (a) axial direction, (b) rotational direction.

Figure 3 appear the FRF of subsystem B linked to coupling with different stiffness values. The frequencies of connection shafts appear with high amplitude.

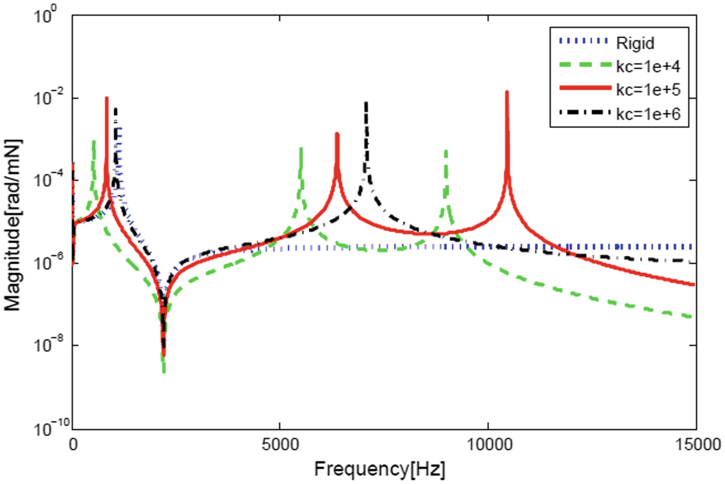


Fig. 3. FRF at coupling DoF.

4.2 Coupling Stiffness Effect on the Coupled Subsystem (A) FRF

The vibratory responses of subsystem A is appeared in Fig. 4. The FRF of flexible coupling model are compared to the rigid coupling.

The FRF of subsystem A is displayed in Fig. 4. The subsystem A FRF is not influenced by the connection rigidity in axial-direction (difference in the magnitude). In the rotational direction, it appears the connection frequency for the flexible coupling case. The peak appearing at around 5000 is due to the coupling DoF. The appearance of a novel frequency because of the elasticity of coupling conform to the connection frequency of planetary gear. In the rigid case, this frequency is localized out of the frequency range.

As a conclusion, the rotational and the coupling directions are sensitive to flexible coupling system. In fact, the curve FRF shows a current eigen-frequency, which represents the connection frequency. The flexibility of coupling affects the transmission reducer because of the low inertia than the inertia of planetary. The sensibility of joint on the distinct subsystems selected.

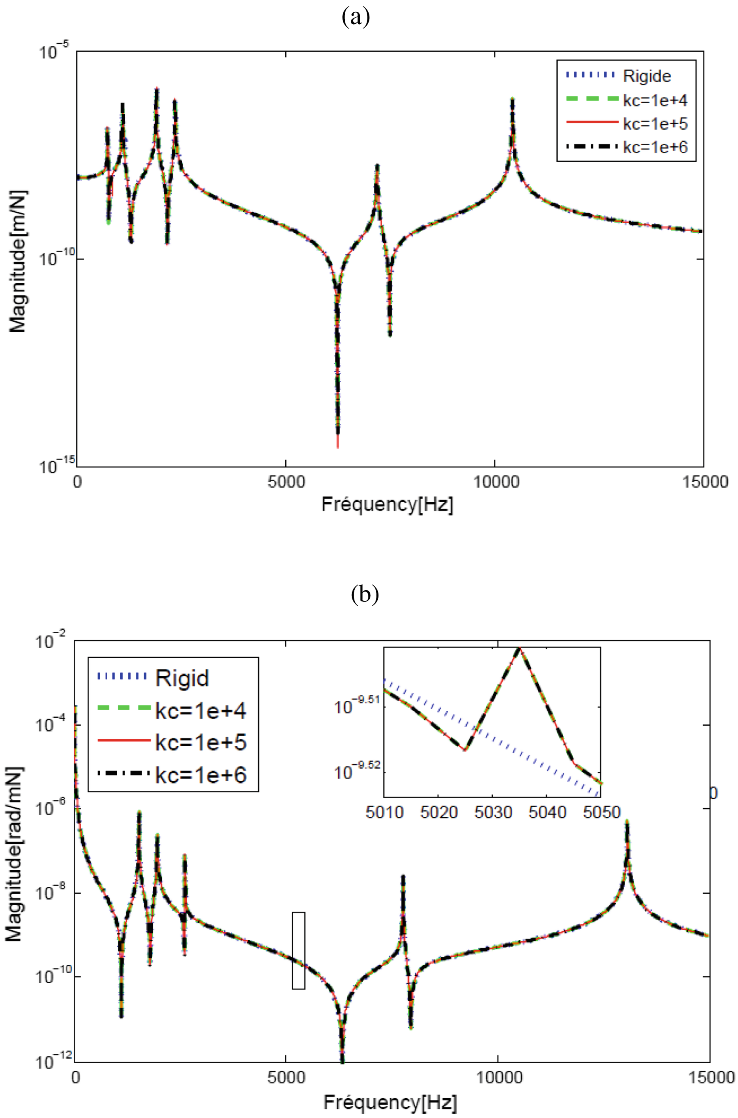


Fig. 4. FRF of coupled system A with various stiffness value: (a) radial direction, (b) sun rotational direction.

5 Conclusion

This manuscript presents the methodology of receptance to be apply for transmission system. For the studied system that composed of planetary gear linked by flexible coupling to parallel stage, this suggested technique is designed to predict their behavior dynamics through subsystems FRF. By comparing the rigid- flexible coupling FRF, we obtained different results in the frequencies. Hence, the elastic coupling affects the achieve results of different subsystem in rotational direction.

The sensibilities of coupling approach to perturbation flexible connection will be the subject of the future work. Furthermore, the analysis of double planetary gear linked with flexible connection will be investigated.

References

Erturk, A., Ozguven, H.N., Budak, E.: Analytical modeling of spindle-tool dynamics on machine tools using Timoshenko beam model and receptance coupling for the prediction of tool point FRF. Int. J. Mach. Tools Manuf. **46**, 1901–19012 (2006)

Park, S.S., Altintas, Y., Movahhedy, M.R.: Receptance coupling for end mills Int. J. Mach. Tools Manuf. **43**, 889–896 (2003)

Liu, W., Ewins, D.J.: Substructure synthesis via elastic media. J. Sound Vib. **257**, 361–379 (2002)

Ozguven, H.N., Budak, E., Ozashin, O., Erturk, A.: A closed form approach for identification of dynamical contact parameters in spindle-holder-tool assemblies. Int. J. Mach. Tools Manuf. **49**, 25–35 (2009)

Ewins, D.J., Liu, W.: Substructure synthesis via elastic media. parti: joint identification. In: IMACXVIII: A Conference on structural dynamics, San Antonio, TX (2000)

Ewins, D.J.: Modal Testing: Theory and Practice. Research Studies Press Ltd., Hertfordshire (2000)

Jetmundsen, B., Bielawa, R.L., Flannelly, W.G.: Generalized frequency domain substructure synthesis. J. Am. Helicopter Soc. **33**(1), 55–64 (1988)

Zhao, B., Zhao, Y., Feng, J., Peng, X.: Numerical and experimental investigation of the torsional stiffness of flexible disc couplings. Int. J. Mech. Sci. **114**, 207–216 (2016)

Bouslema, M., et al.: Effects of modal truncation and condensation methods on the Frequency Response Function of a stage reducer connected by rigid coupling to a planetary gear system. Comptes Rendus Mecanique **345**, 807–823 (2017)

Zhen, J., Lim, T.C., Lu, G.: Determination of system vibratory response characteristics applying a spectral-based inverse sub-structuring approach. Part I: analytical formulation. Int. J. Veh. Noise Vib. **1**, 1–30 (2004)

Hmida, A., Hammemi, A., Khabou, M.T., Charii, F., Haddar, M.: Effect of elastic coupling on the modal characteristics of spur gearbox system. Appl. Acoust. (2017)

Mahmoudi, A.E., Rixen, D.J., Meyer, C.H.: Comparison of different approaches to include connection elements into frequency-based substructuring. Exp. Tech. **44**(4), 425–433 (2020). https://doi.org/10.1007/s40799-020-00360-1

A Comparison Between Independent Component Analysis and Established Signal Processing Methods for Gearbox Fault Diagnosis Under Time-Varying Operating Conditions

Stephan Schmidt(✉), Daniel N. Wilke, and P. Stephan Heyns

Centre for Asset Integrity Management, Department of Mechanical and Aeronautical Engineering, University of Pretoria, Pretoria, South Africa
`stephan.schmidt@up.ac.za`

Abstract. Reliable condition monitoring methods are required for rotating machines operating under time-varying operating conditions. The measured vibration signals typically contain information related to the different interacting components (e.g. gear mesh components, bearing fault components), the transmission paths between the excitation sources and the sensors, the environmental conditions (e.g. changes in temperature) and the operating conditions of the machine. Hence, multiple sources could be present in the measured signals, which could impede the detection of weak sources attributed to incipient damage. Several methods have been proposed to solve this problem, including, synchronous statistics (e.g. time-synchronous averages, synchronous average of the squared envelope, synchronous median of the squared envelope), the squared envelope spectrum, the order-frequency spectral coherence and the integrated squared spectral coherence (e.g. the enhanced envelope spectrum and the improved envelope spectrum). Independent Component Analysis (ICA) is a well-established technique that has not been compared against the aforementioned methods. In this work, we compare the performance of ICA against the performance against established signal analysis methods for fault detection under time-varying operating conditions. We show that ICA performs well against established signal analysis-based condition monitoring methods for machines operating under time-varying conditions.

Keywords: Independent Component Analysis · Order-frequency spectral coherence · Improved envelope spectrum · Gearbox fault detection · Time-varying operating conditions

1 Introduction

Reliable fault diagnosis methods are required for critical assets in the renewable (e.g. wind turbines) and mining (e.g. ball mills) industries [1,2]. Several signal

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 207–223, 2022.
https://doi.org/10.1007/978-3-030-85584-0_21

processing methods have been developed in the past few decades to address the challenges in gearbox diagnostics [3–5].

It is especially challenging to perform fault diagnosis under time-varying and impulsive noise conditions [2,4,6]. This is because the damaged components often manifest in low energy bands, with time-varying operating conditions and extraneous impulsive events dominating the raw signal and its processed representations (e.g. spectral coherence). The Squared Envelope Spectrum [7], the Order-Frequency Spectral Coherence [4], the Improved Envelope Spectrum [8], and the Synchronous Median of the Squared Envelope [9] of the raw and processed signals are well-established methods for fault diagnosis under varying operating conditions.

Independent Component Analysis (ICA) was actively researched in the 1990s for extracting statistically independent non-Gaussian sources from data [10]. ICA is also useful for fault diagnosis, as damaged signals often manifest as non-Gaussian sources in the measured data. This is proven by the papers on ICA for gearbox condition monitoring [11–15], with the first papers using ICA published in the early 2000s. According to the best of our knowledge, ICA has not been compared against established signal processing methods for gearboxes operating under time-varying operating conditions. Hence, in this work, we compare ICA against established signal processing methods for fault detection under time-varying operating conditions.

In Sect. 2, the ICA method is introduced, whereafter a brief overview of the established signal processing methods are presented in Sect. 3. The methods are compared on two experimental signals in Sect. 4, followed by a conclusion of the study in Sect. 5.

2 Independent Component Analysis (ICA)

ICA is used to extracted statistically independent, linear mixed sources from vibration signals. In this work, we define a source as a statistical independent signal component caused by an excitation mechanism (e.g. bearing damage) in the system. In gearbox diagnosis, ICA has been applied on multiple sensor observations [12,13,16], used on the wavelet decomposition of a single accelerometer signal [15], and used on windowed vibration signals [11]. The investigated ICA fault diagnosis methodology is shown in Fig. 1. A similar procedure is used in Ref. [11]. The measured vibration signal $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ is acquired from a rotating machine in an unknown condition. The windowing process transforms the signal from a vector $\boldsymbol{x} \in \mathbb{R}^{1\times N}$ to a matrix $\boldsymbol{X} \in \mathbb{R}^{M\times L}$ with the process shown in Fig. 2. The window length parameter, L, and the stride, S, need to be selected. Before ICA is performed, the windowed signal is whitened. In Fig. 1 the whitened data are denoted $\boldsymbol{X}_w \in \mathbb{R}^{M\times L}$.

Independent component analysis aims to extract Z sources $\boldsymbol{S} \in \mathbb{R}^{M\times Z}$ from the signal $\boldsymbol{X}_w \in \mathbb{R}^{M\times L}$. In the formulation, we assume

$$\boldsymbol{X}_w = \boldsymbol{S}\boldsymbol{W}^T, \tag{1}$$

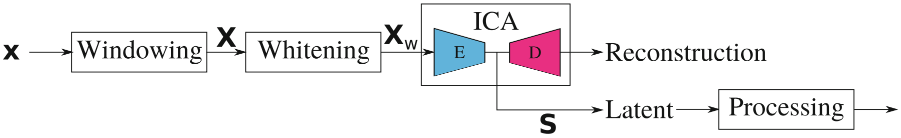


Fig. 1. The investigated methodology is shown for a single measured vibration signal $\boldsymbol{x} \in \mathbb{R}^{1\times N}$. The windowed signal $\boldsymbol{X} \in \mathbb{R}^{M\times L}$ and the sources $\boldsymbol{S} \in \mathbb{R}^{M\times Z}$ are shown, where M is the number of windows and Z is the number of sources. The encoder is denoted by E and the decoder is denoted by D.

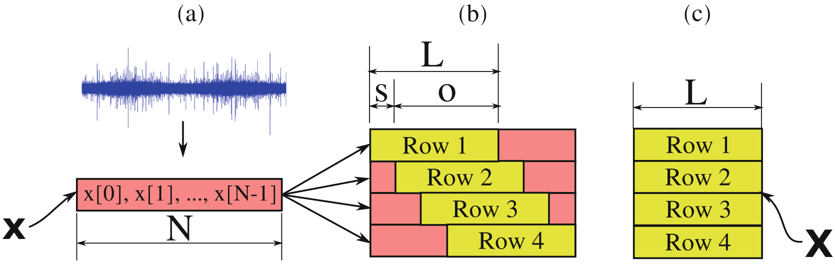


Fig. 2. The windowing process is illustrated. (a) The measured vibration signal $\boldsymbol{x}^{1\times N}$ has N equally spaced points. (b) The vibration signal is windowed into four windows. The window length is denoted L, the stride is denoted s, and the overlap is denoted o, with $L = s + o$. (c) The windowed signal is a matrix $\boldsymbol{X} \in \mathbb{R}^{4\times L}$.

where the unknown weights (also known as the projection matrix or mixing matrix) of Z sources is denoted $\boldsymbol{W} \in \mathbb{R}^{L\times Z}$. The source matrix can be written in terms of Z sources

$$\boldsymbol{S} = [\boldsymbol{s}_1 \ \ \boldsymbol{s}_2 \ \ \ldots \boldsymbol{s}_Z], \tag{2}$$

where the ith source is denoted $\boldsymbol{s}_i \in \mathbb{R}^{M\times 1}$. In ICA, the sources $\boldsymbol{S}$ and the weights $\boldsymbol{W}$ are unknown. Therefore, to find the unknown weight matrix, the sources are assumed to be statistically independent and non-Gaussian [10]. Statistical independence implies that the joint distribution can be factored, i.e. $p(\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_z) = \prod_{i=1}^{Z} p(\boldsymbol{s}_i)$, while non-Gaussianity defines the characteristics of the expected sources. Specifically, the sources are selected to maximise non-Gaussianity according to some selected measure subject to being uncorrelated. Measures of non-Gaussianity include kurtosis and negentropy, with kurtosis tending to be sensitive to outliers. Several algorithms have been developed to compute independent components and include FastICA [17], InfoMax [18] and Extended InfoMax [19] that generalise source distributions to be sub- and super-Gaussian. Vibration signals typically become more leptokurtic as the machine degrades. This is because damaged components such as bearings are impulsive events in the signals [20] and therefore focusing on non-Gaussian sources are sensible.

As opposed to analysing the reconstruction error, we only investigate the sources in this work. Therefore, after the sources $\boldsymbol{S}$ are determined, each source

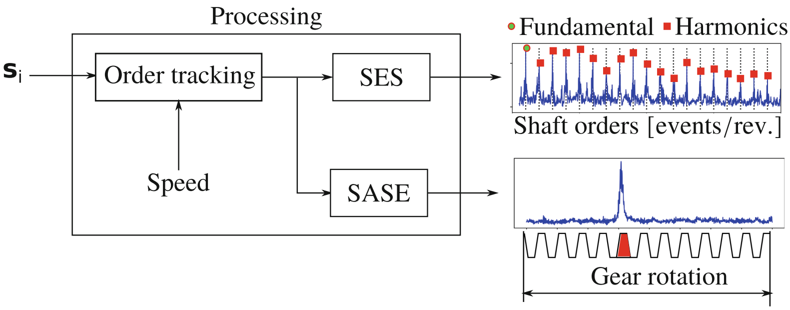


Fig. 3. The processing procedure of the sources: The Squared Envelope Spectrum (SES) and the Synchronous Average of the Squared Envelope (SASE) of the source signal are calculated and used for diagnosis. The unfolded gear, with the highlighted damaged tooth, is shown below the SASE.

$\boldsymbol{s}_i$ is separately processed with the procedure shown in Fig. 3. Firstly, the signal is order tracked to transform the time domain signal to an angle domain representation. This is important as the signals are known to be angle periodic (instead of time-periodic) due to the angle-locked excitations in rotating machines. In a computed order tracked signal, the angle-periodic information is preserved under time-varying speed conditions. Computed order tracking is performed with a phase reference (e.g. optical probe focused on a zebra tape shaft encoder). Tacholess order tracking methods are useful when the reference phase information cannot be measured or it is not available (e.g. it is impractical to install a tachometer) [21].

Damage events (e.g. bearing cracks) often create abrupt excitations, which manifest as amplitude modulation in the signals. The amplitude modulation's periodicity can be seen in the Squared Envelope Spectrum. The SES of the ith source

$$\text{SES}_i(\alpha) = \left|\mathcal{F}_{\theta\rightarrow\alpha}\left\{|s_i(\theta)|^2\right\}\right|, \tag{3}$$

is calculated from the Fourier transform $\mathcal{F}_{\theta\rightarrow\alpha}$ of the squared source $|s_i(\theta)|^2$. The Fourier transform $\mathcal{F}_{\theta\rightarrow\alpha}$, transforms the angle variable θ to an order variable α. The SES indicates the periodicity of the modulation components. Since the signal is order tracked, the x-axis of the SES is measured in number of events per revolution (e.g. number of bearing impulses per shaft rotation).

If the cyclic order of the component-of-interest is known, it is possible to visualise the modulation caused by the damage with the Synchronous Average of the Squared Envelope (SASE). The SASE of the ith source

$$\text{SASE}_i(\theta; \Theta) = \frac{1}{N_r}\sum_{n=0}^{N_r-1} |s_i(\theta + n\cdot\Theta)|^2, \tag{4}$$

is calculated over the preselected period Θ. For example, if the gear rotates at one shaft order (i.e. it is synchronous with the reference shaft), then $\Theta = 1.0$ rotation or 2π radians.

3 Established Signal Processing Methods

Four established signal processing methods are considered in this work, namely, the Synchronous Statistics (SS), the Squared Envelope Spectrum (SES), the Order-Frequency Spectral Coherence (OFSCoh), and the integrated OFSCoh. Only a brief overview of the methods is presented here, with the appropriate references given for more detailed information.

3.1 Synchronous Statistics (SS)

The Synchronous Average of the Squared Envelope (SASE) and the Synchronous Median of the Squared Envelope (SMSE) are capable of visualising the modulation caused by the damaged components [9]. Synchronous statistics are calculated in two steps:

- The synchronous angle domain data, with a predefined period Φ, are collected in a synchronous matrix $\boldsymbol{X}_s \in \mathbb{R}^{N_r \times N_s}$, where N_r is the number of rotations and N_s is the number of samples per rotation of the component-of-interest. The period of rotation Φ is the reciprocal of the cyclic order of the component-of-interest. The matrix $\boldsymbol{X}_s$ can be obtained with the windowing procedure explained in Fig. 2 using a window length of N_s and a window overlap of 0.
- The statistic (e.g. average, median) is taken over the rows of each column to find a synchronous vector with $\mathbb{R}^{N_s}$.

The SMSE is obtained by windowing the squared envelope and calculating the median across the rows of the matrix $\boldsymbol{X}_s$. In contrast to the SASE, the SMSE is not sensitive to extraneous[1] impulsive components and, therefore, more reliable to visualise the modulation due to damage [9]. Both are considered in this work to show that the SASE of the sources perform better than the SASE and the SMSE of the raw signals.

3.2 Squared Envelope Spectrum (SES)

The SES is the most commonly used method for bearing diagnostics [3]. The SES visualises the periodicity of the damaged components making it possible to detect bearing and gear damage. It is also possible to determine the damaged component (e.g. inner race vs. outer race bearing damage). A detailed analysis of the SES for bearing diagnostics under time-varying speed conditions is presented in Ref. [7]. Usually, the signal is pre-processed with cepstrum pre-whitening [22] or the deterministic components are subtracted from the signal [7]. However, the squared envelope spectrum of the raw signal is considered in this work for simplicity's sake.

---

[1] We refer to extraneous events as signal components (deterministic, stationary, or cyclostationary) that impede damage detection.

3.3 Order-Frequency Spectral Coherence (OFSCoh)

The OFSCoh is a cyclic order α spectral frequency f representation of the signal [4]. The OFSCoh, denoted $\gamma(\alpha, f) \in \mathbb{C}$, contains magnitude and phase information. The OFSCoh makes it possible to visualise the periodicity of the modulation caused by the damage and the carrier frequency bands. It is a generalisation of the squared envelope spectrum, where the squared envelope spectrum is obtained from [8]

$$\mathrm{SES}(\alpha) = \left| \int_0^{f_s/2} \gamma(\alpha, f) df \right|, \tag{5}$$

where f_s is the sampling frequency. The OFSCoh contains all of the fault information, however, it can be difficult to interpret. Therefore the integrated squared OFSCoh is often used for diagnosis.

3.4 Integrated Squared OFSCoh

The integrated squared OFSCoh representations are considered here. The Enhanced Envelope Spectrum (EES)

$$\mathrm{EES}(\alpha) = \int_0^{f_s/2} \left| \gamma(\alpha, f) \right|^2 df, \tag{6}$$

is better suited for damage detection than the SES. This is because it is more capable of enhancing small damage components [8]. The Improved Envelope Spectrum (IES) is defined in this work by

$$\mathrm{IES}(\alpha; f_l, f_h) = \int_{f_l}^{f_h} \left| \gamma(\alpha, f) \right|^2 df. \tag{7}$$

If the integration band $[f_l, f_h]$ is appropriately selected, it can improve the signal-to-noise ratio of the damaged components [8]. In this work, we distinguish between the EES and the IES as follows: The EES is calculated over the full frequency bandwidth of the signal, while the IES is obtained by integrating over a predefined frequency band. The IESFOgram is a frequency band identification method that automatically finds the optimal frequency bands to calculate the IES [23]. This is performed by maximising a feature such as the strength of the cyclostationary components. However, to apply the IESFOgram as proposed in Ref. [23], the characteristic cyclic order of the component-of-interest needs to be specified, which is not always possible. In addition to this, the damaged component might not be targeted and therefore the damage will not be detected [23]. Blind informative frequency band identification methods overcome the need to specify the characteristic order of the component-of-interest, but do not perform as well as targeted methods [5,24].

The IES, obtained by selecting an informative frequency band is considered state-of-the-art in this work and will be used to benchmark the performance of the ICA method. Since we know the informative frequency band from previous investigations (e.g. targeted frequency band identification methods were used in Refs. [5,25] to identify the informative frequency band), we used this information to specify the frequency band to calculate the IES in this work.

4 Experimental Case Study

In Fig. 4, the experimental test-rig in the Centre for Asset Integrity Management laboratory of the University of Pretoria is shown. It contains an electrical motor that drives the system and an alternator that applies a load to the system. The motor's speed and the alternator's load are independently controlled from a personal computer and used to enforce the varying operating conditions shown in Figs. 4(b) and 4(c). The test-rig also contains three helical gearboxes, with the centre gearbox being monitored for damage. For the measurements considered in this work, the gear was damaged with seeded slot shown in Fig. 4(d) and operated under varying operating conditions until the gear tooth failed as shown in Fig. 4(e). The gearbox under consideration has 37 gear teeth and 20 pinion teeth, which means that it has a gear ratio of 1.85 and it has a gear mesh order of 37 when using the input shaft, i.e. S2 in Fig. 4, as a reference.

Two measurements are used to compare the performance of the ICA method (Sect. 2) against the established signal processing methods (Sect. 3). The first measurement was taken after the experiment was approximately 30% complete, while the second measurement was taken after the experiment was approximately 90% complete, i.e. the gear tooth was close to failure.

4.1 Measurement 1

The raw vibration signal and its power spectral density are shown in Figs. 5(a) and (b) respectively. The signal is dominated by impulse components that manifest in the 8 kHz frequency band shown in Fig. 5(b). The power spectral density of the order tracked signal is shown in Fig. 5(c) with a zoomed view shown in Fig. 5(d). The lower orders are dominated by components associated with S4 as the associated shaft was slightly unbalanced. The Gear Mesh Order (GMO) is small, since the helical gears result in lower excitations due to its smoother operation.

The SASE, SMSE and the SES are compared in Fig. 6. Extraneous impulsive components, with a cyclic order of approximately 5.72 shaft orders, dominate the SASE and the SES and therefore impede damage detection. The SMSE is more robust to the extraneous impulsive components, with the damage seen at approximately 135°.

The squared magnitude of the OFSCoh is presented on a linear and logarithmic scale in Fig. 7. The extraneous impulsive components at 5.72 shaft orders also dominate the OFSCoh in the linear scale plot, and therefore the logarithmic

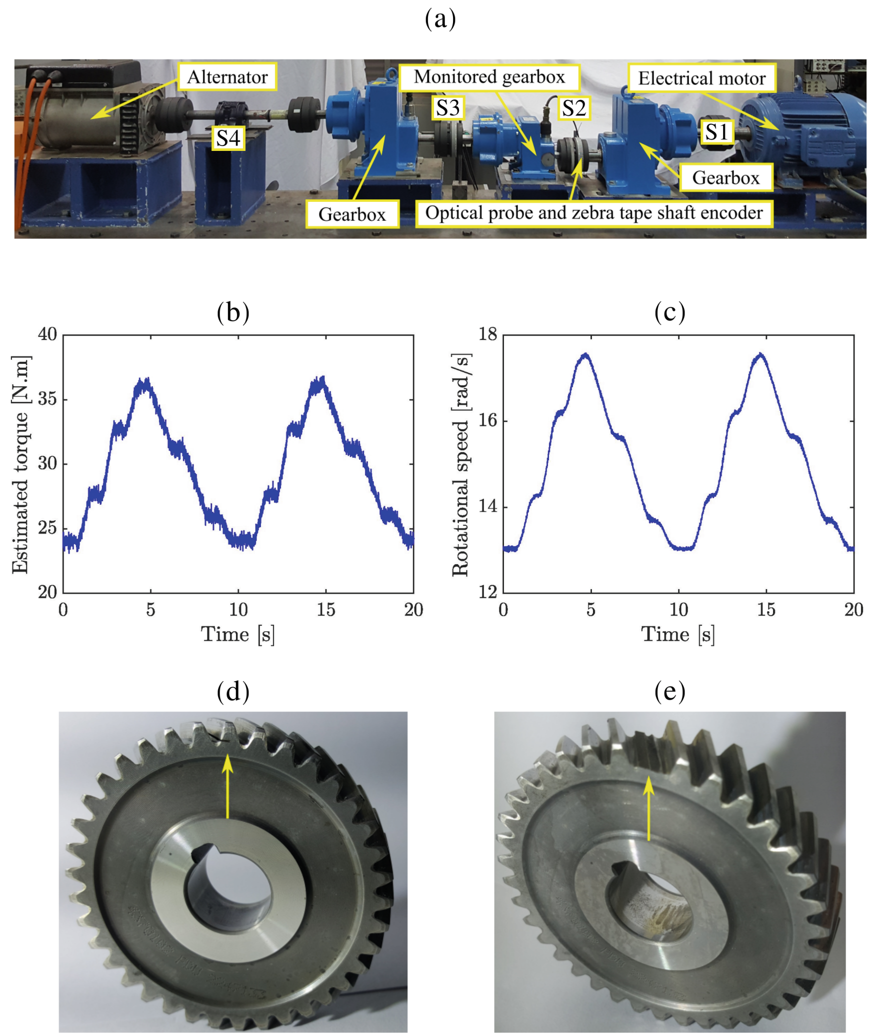


Fig. 4. (a) Experimental test-rig with the shaft numbers S1–S4 superimposed on the figure. (b) Rotational speed of the input shaft of the monitored gearbox (i.e. shaft S2). (c) Load at the input shaft of the monitored gearbox (i.e. shaft S2). (d) Gear before the experiment started (0% complete). (e) Gear after the experiment was completed (100% complete).

scale is used in the subsequent discussion. The OFSCoh contains many dominant frequency bands that are unrelated to the damaged component under consideration. This makes detecting the damage in the lower frequency bands (between 250–700 Hz in Fig. 5) difficult.

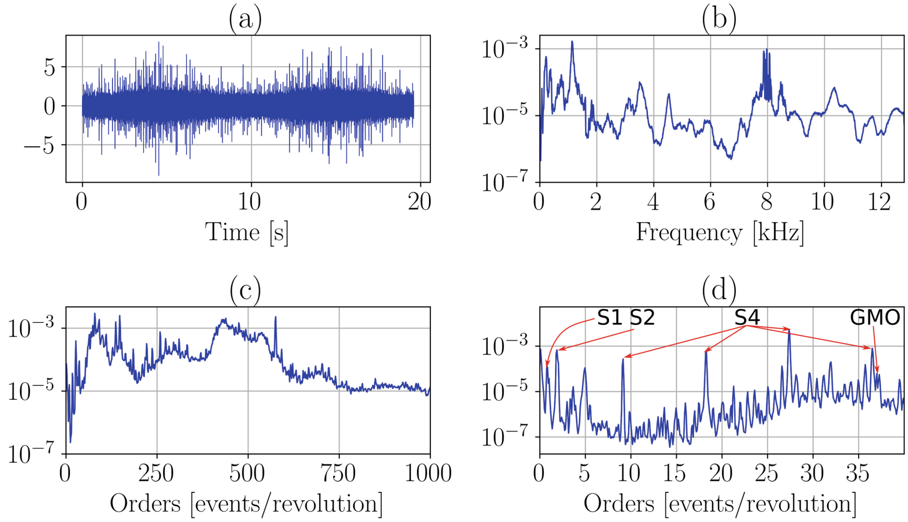


Fig. 5. The raw acceleration signal in [m/s^2] (a); the power spectral density of the raw signal (b); and the power spectral density of the order tracked signal (c - d) are shown for Measurement 1. Abbreviations: S2 - Shaft 2 (input shaft of the monitored gearbox); S3 - Shaft 3 (output shaft of the monitored gearbox) S4 - Shaft 4 (input shaft to the alternator); GMO - Gear Mesh Order. See Fig. 4 for the locations of the shafts.

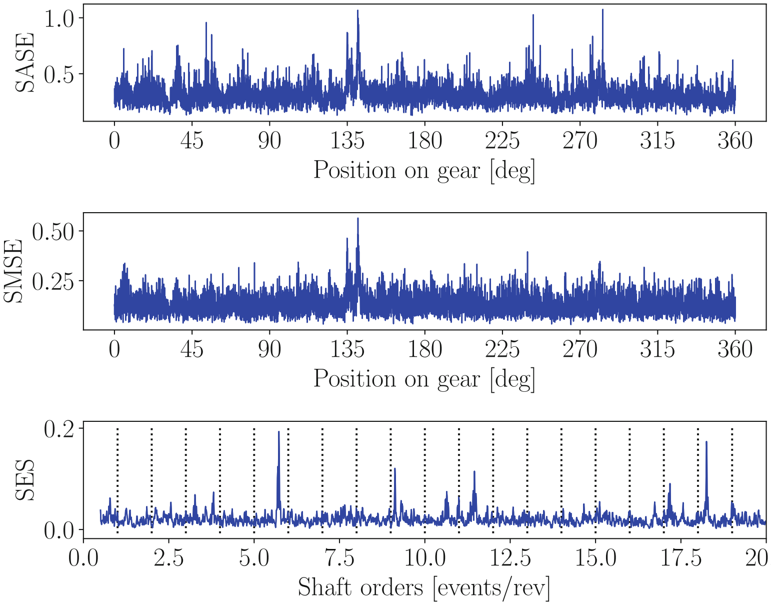


Fig. 6. The Synchronous Statistics (SS) and the Squared Envelope Spectrum (SES) of Measurement 1.

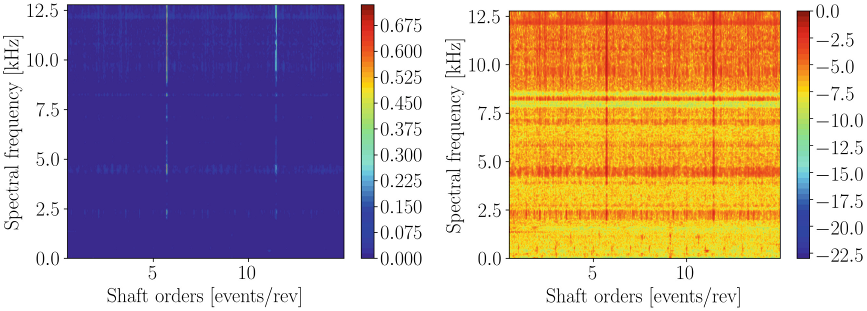


Fig. 7. The Order-Frequency Spectral Coherence (OFSCoh) of Measurement 1 on a linear scale (left) and on a logarithmic scale (right).

Therefore, the integrated squared OFSCoh is investigated. The EES in Fig. 8 is also dominated by the 5.72 shaft order components, and the gear damage at 1.0 shaft order and its harmonics are not observed. The IES in Fig. 8 is calculated over a spectral frequency band between 250 and 700 Hz. The damage is prominent and is therefore easy to detect. We can determine the integration band $[f_l, f_h]$ with the following options:

- Prior knowledge about the damage frequency band can be used. However, the spectral frequency band of the damaged component could vary between damaged components, and we will not know these bands for new machines.
- Targeted informative frequency band identification methods can be used. However, targeted informative frequency band identification methods require the cyclic order of the component-of-interest to be specified, which might not always be possible or we might not target the correct components.
- Blind informative frequency band identification methods can be used. However, these methods do not perform as well as the targeted methods [5,24] and are mostly dominated by the extraneous impulsive component at 5.72 shaft orders in this work.

In this work, 16 sources were extracted with the ICA procedure with a window length of 256 and an overlap of 255. We used FastICA in Ref. [17,26]. The results for the investigated ICA procedure are presented in Fig. 9. The two-dimensional SASE map is obtained by following the procedure in Fig. 3 for each source. The gear damage is most prominent in the fourth source, with the SASE and the SES of the fourth source, also presented in Fig. 9. The damage is more prominent in the SASE of Fig. 9 when compared to the SASE and the SMSE in Fig. 6. The SES in Fig. 9 also contains the damaged gear components at 1.0 shaft order and performs better than the SES in Fig. 6 and the EES in Fig. 8. The IES performs the best of the considered methods. However, its performance depends on our ability to determine the frequency band of interest. If the wrong frequency band is selected or if the frequency band is too large, the gear damage will not be detected.

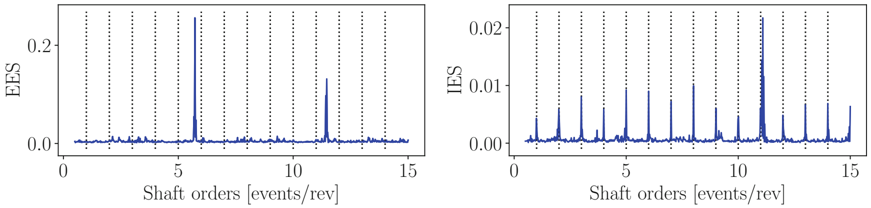


Fig. 8. The Enhanced Envelope Spectrum (EES) and the Improved Envelope Spectrum (IES) of Measurement 1. The dashed line indicates the shaft orders of the theoretical fault components.

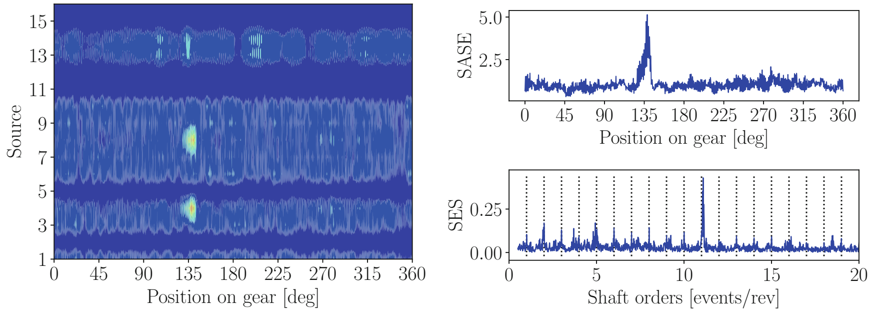


Fig. 9. The Synchronous Average of the Squared Envelope (SASE) of the sources map is shown on the left. On the right, the SASE and the Squared Envelope Spectrum (SES) of the source with the damage information are shown for Measurement 1.

4.2 Measurement 2

For the second measurement, the synchronous statistics, the squared envelope spectrum and the OFSCoh performed similarly to the previous measurement. Therefore, their results are excluded for brevity's sake. Hence, in this section, we only compare the EES (which is better than the SES for damage detection [8]) and the IES against the investigated ICA procedure. The EES and the IES are presented in Fig. 10 for the second measurement. Even though the damage is more severe in this measurement, the EES is still dominated by the extraneous signal components with the damaged gear components at 1.0 shaft order and its harmonics not seen. The IES, which was calculated over the same frequency band as the previous measurement, performs well. The damaged gear components at 1.0 shaft order and its harmonics are easily visualised.

The same ICA procedure is applied in this section as the first measurement, with the results shown in Fig. 11. The SASE sources map in Fig. 11 contains the gear damage in the 1st source. The damaged gear tooth is prominent in the SASE of the 1st source and the cyclic orders of the damaged gear tooth is prominent in the SES.

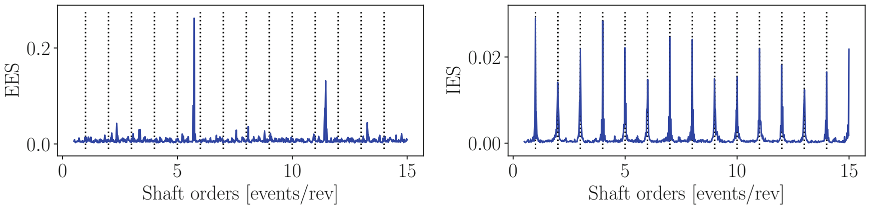


Fig. 10. The Enhanced Envelope Spectrum (EES) and the Improved Envelope Spectrum (IES) of Measurement 2.

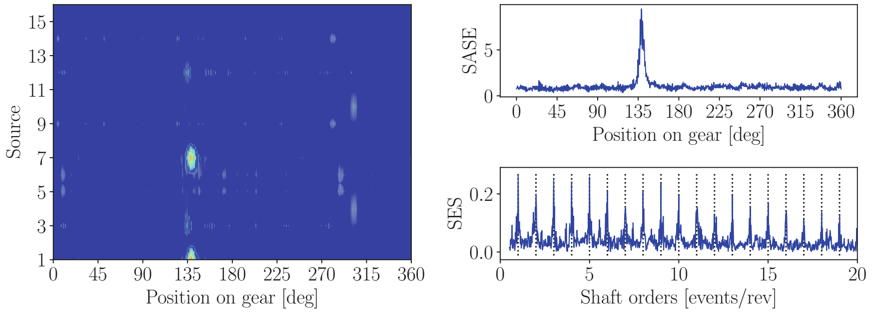


Fig. 11. The Synchronous Average of the Squared Envelope (SASE) of the sources map is shown on the left. On the right, the SASE and the Squared Envelope Spectrum (SES) of the source with the damage information are shown for Measurement 2.

The SASE and the SES of the appropriate ICA sources perform better than the SASE, SMSE, OFSCoh, and the EES of the raw data. Even though ICA does not perform as well as the IES, it has significant potential for damage diagnosis as it does not require the fault frequency to be known a priori. The cyclic order of the fault components can be inferred from the SES of the appropriate source. In the next section, a quantitative comparison between the different cyclic spectra is performed.

4.3 Quantitative Comparison

In this section, we quantitatively compare the performance of the different cyclic spectra:

1. Squared Envelope Spectrum (SES) of the raw signal.
2. Enhanced Envelope Spectrum (EES) of the raw signal.
3. Improved Envelope Spectrum (IES) of the raw signal.
4. SES of each source that was extracted from the signal using ICA. In this work, we extracted 16 sources and therefore 16 SES will be considered.

For a specific cyclic spectrum, we denote the amplitude of the ith harmonic of the component-of-interest as A_i. The average amplitude of the first N_Q components is given by

$$\bar{A} = \frac{1}{N_Q} \sum_{i=1}^{N_Q} A_i. \tag{8}$$

In this work, we calculate $\bar{A}$ using the first five harmonics of the gear component, i.e. $N_Q = 5$.

The noise floor of the cyclic spectrum is denoted by A_n and approximated by the median of the cyclic spectrum. The median is used in this work, since it is more robust to sparse components than the mean. Therefore, the median will be less affected by the presence of large amplitude components (e.g. components due to damage, extraneous cyclostationary components). This median is also used in Ref. [5] to estimate the noise floor. The maximum amplitude in the cyclic spectrum is denoted by A_{max}.

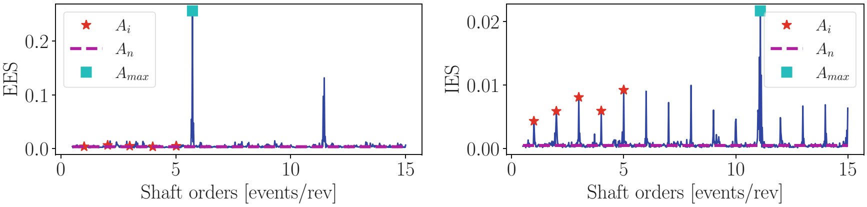


Fig. 12. The amplitude of the component-of-interest A_i (i.e. the ith harmonic of the gear component), the estimated noise level A_n and the maximum component A_{max} are superimposed on the EES and the IES of Measurement 1. These components are used in Eqs. (8)–(10).

To visualise these terms, we present the amplitudes of the first five components A_i, the estimated noise floor A_n, and the maximum amplitude A_{max} for the EES and the IES of measurement 1 in Fig. 12. The damage is much more prominent in the IES than the EES as highlighted by:

- The magnitude of the component-of-interest is much larger than the average noise level in the IES.
- The magnitude of the component-of-interest has the same order of magnitude as the most dominant component. In the EES, the most dominant component, which is not attributed to gear damage, dominates the cyclic spectrum and makes the damaged components difficult to observe.

Two metrics are used to quantify the two observations. Metric 1

$$\phi_1 = \frac{\bar{A}}{A_n}, \tag{9}$$

measures the relative magnitude of the components-of-interest with respect to the noise floor and is an indication of the prominence of the damage. This metric is also used in Ref. [5] to identify informative frequency bands. Metric 2

$$\phi_2 = \frac{\bar{A}}{A_{max}}, \tag{10}$$

compares the average magnitude of the component-of-interest against the magnitude of the most dominant component. The larger this value is, the easier it will be to observe the damage in the spectrum.

We calculated these metrics for the different cyclic spectra and included the results in Table 1. The IES has the largest values for both metrics, which corroborates the previous observations that the IES performs the best when the cyclic order of the component-of-interest is known *a priori*. ICA performs second best, with more prominent damaged components when compared to the SES and the EES of the raw signals. Since the damage components only manifest

Table 1. A quantitative comparison between the different cyclic spectra using metric 1 (ϕ_1), defined in Eq. (9), and metric 2 (ϕ_2), defined in Eq. (10). The bold values indicate the maximum values in the columns. The italic values indicate the ICA sources with the damage components.

	Measurement 1		Measurement 2	
	ϕ_1	ϕ_2	ϕ_1	ϕ_2
SES (Raw)	1.561	0.023	2.498	0.07
EES	1.298	0.019	1.945	0.043
IES	**14.781**	**0.307**	**42.453**	**0.797**
SES (ICA source 1)	1.372	0.041	*6.475*	*0.242*
SES (ICA source 2)	1.895	0.032	2.413	0.071
SES (ICA source 3)	1.365	0.038	2.528	0.107
SES (ICA source 4)	*4.234*	*0.096*	1.783	0.209
SES (ICA source 5)	2.287	0.03	1.515	0.057
SES (ICA source 6)	1.565	0.043	1.525	0.057
SES (ICA source 7)	1.592	0.029	*6.063*	*0.247*
SES (ICA source 8)	*4.031*	*0.095*	2.336	0.051
SES (ICA source 9)	1.521	0.037	1.632	0.06
SES (ICA source 10)	1.625	0.032	1.776	0.207
SES (ICA source 11)	1.76	0.032	2.339	0.05
SES (ICA source 12)	2.289	0.03	2.526	0.105
SES (ICA source 13)	2.088	0.066	2.052	0.087
SES (ICA source 14)	2.12	0.07	1.634	0.06
SES (ICA source 15)	1.751	0.032	2.402	0.071
SES (ICA source 16)	1.885	0.032	2.041	0.088

in a subset of the sources, it is important to identify the informative sources for further processing. Automatic source selection methods will be considered in future work.

5 Conclusion

Independent Component Analysis (ICA) is a well-established technique for gearbox fault diagnosis. However, its performance has not been compared against established fault diagnosis methods under time-varying operating conditions. In this work, we showed for two experimental measurements that ICA is capable of extracting informative signal components from the vibration signals and performs very well when compared to established fault diagnosis methods under time-varying operating conditions. ICA has several benefits, e.g. it is blind (i.e. we do not have to specify the orders of the damaged component) and simple to perform. Hence, its underutilisation by the condition monitoring community is unclear to the authors. We speculate that the lack of mainstream adoption is attributed to ICA being incorrectly perceived as a solution to the cocktail party problem, which implies that ICA is only useful when combining multiple vibration channels, when using time-frequency decompositions or when performing feature selection. Hence, our recommendation is that ICA should be considered when testing new condition monitoring methods under time-varying operating conditions, especially in circumstances where blind methods are required. Future work will focus on automatically resolving the hyperparameters (e.g. number of sources) of the ICA method.

References

1. Salameh, J.P., Cauet, S., Etien, E., Sakout, A., Rambault, L.: Gearbox condition monitoring in wind turbines: a review. Mech. Syst. Sig. Process. **111**, 251–264 (2018)
2. Kruczek, P., Zimroz, R., Antoni, J., Wyłomańska, A.: Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech. Syst. Sig. Process. **159**, 107737 (2021)
3. Randall, R.B., Antoni, J.: Rolling element bearing diagnostics - a tutorial. Mech. Syst. Sig. Process. **25**, 485–520 (2011)
4. Abboud, D., Baudin, S., Antoni, J., Rémond, D., Eltabach, M., Sauvage, O.: The spectral analysis of cyclo-non-stationary signals. Mech. Syst. Sig. Process. **75**, 280–300 (2016)
5. Schmidt, S., Heyns, P.S., Gryllias, K.C.: An informative frequency band identification framework for gearbox fault diagnosis under time-varying operating conditions. Mech. Syst. Sig. Process. **158**, 107771 (2021)
6. Schmidt, S., Zimroz, R., Heyns, P.S.: Enhancing gearbox vibration signals under time-varying operating conditions by combining a whitening procedure and a synchronous processing method. Mech. Syst. Sig. Process. **156**, 107668 (2021)
7. Abboud, D., Antoni, J., Sieg-Zieba, S., Eltabach, M.: Envelope analysis of rotating machine vibrations in variable speed conditions: a comprehensive treatment. Mech. Syst. Sig. Process. **84**, 200–226 (2017)

8. Antoni, J., Xin, G., Hamzaoui, N.: Fast computation of the spectral correlation. Mech. Syst. Sig. Process. **92**, 248–277 (2017)
9. Schmidt, S., Zimroz, R., Chaari, F., Heyns, P.S., Haddar, M.: A simple condition monitoring method for gearboxes operating in impulsive environments. Sensors **20**, 2115 (2020)
10. Comon, P.: Independent component analysis, a new concept? Sig. Process. **36**, 287–314 (1994)
11. He, Q., Feng, Z., Kong, F.: Detection of signal transients using independent component analysis and its application in gearbox condition monitoring. Mech. Syst. Sig. Process. **21**, 2056–2071 (2007)
12. Tian, X., Lin, J., Fyfe, K.R., Zuo, M.J.: Gearbox fault diagnosis using independent component analysis in the frequency domain and wavelet filtering. In: 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. (ICASSP 2003), vol. 2, pp. II-245. IEEE (2003)
13. Wodecki, J., Stefaniak, P., Sawicki, M., Zimroz, R.: Application of independent component analysis in temperature data analysis for gearbox fault detection. In: Cyclostationarity: Theory and Methods III, pp. 187–198. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-51445-1_11
14. Albarbar, A., Gu, F., Ball, A.: Diesel engine fuel injection monitoring using acoustic measurements and independent component analysis. Measurement **43**, 1376–1386 (2010)
15. Zuo, M.J., Lin, J., Fan, X.: Feature separation using ICA for a one-dimensional time series and its application in fault detection. J. Sound Vibr. **287**, 614–624 (2005)
16. Duan, F., Corsar, M., Zhou, L., Mba, D.: Using independent component analysis scheme for helicopter main gearbox bearing defect identification. In: 2017 IEEE International Conference on Prognostics and Health Management (ICPHM), pp. 252–259. IEEE (2017)
17. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Networks **10**, 626–634 (1999)
18. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Proceedings of the 8th International Conference on Neural Information Processing Systems, NIPS 1995, (Cambridge, MA, USA), pp. 757–763. MIT Press (1995)
19. Lee, T.-W., Girolami, M., Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. Neural Comput. **11**, 417–441 (1999)
20. Borghesani, P., Smith, W., Zhang, X., Feng, P., Antoni, J., Peng, Z.: A new statistical model for acoustic emission signals generated from sliding contact in machine elements. Tribol. Int. **127**, 412–419 (2018)
21. Peeters, C., et al.: Review and comparison of tacholess instantaneous speed estimation methods on experimental vibration data. Mech. Syst. Sig. Process. **129**, 407–436 (2019)
22. Borghesani, P., Pennacchi, P., Randall, R., Sawalhi, N., Ricci, R.: Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions. Mech. Syst. Sig. Process. **36**, 370–384 (2013)
23. Mauricio, A., Smith, W.A., Randall, R.B., Antoni, J., Gryllias, K.: Improved envelope spectrum via feature optimisation-gram (IESFOgram): a novel tool for rolling element bearing diagnostics under non-stationary operating conditions. Mech. Syst. Sig. Process. **144**, 106891 (2020)

24. Smith, W.A., Borghesani, P., Ni, Q., Wang, K., Peng, Z.: Optimal demodulation-band selection for envelope-based diagnostics: a comparative study of traditional and novel tools. Mech. Syst. Sig. Process. **134**, 106303 (2019)
25. Schmidt, S., Gryllias, K.C.: Combining an optimisation-based frequency band identification method with historical data for novelty detection under time-varying operating conditions. Measurement. **169**, 108517 (2021)
26. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

Simulation of the Acoustic Fields and Directivities Radiated from a Set of Elementary Sources

Dhouha Tounsi, Maryam Bedoui, Mohamed Taktak(✉), Chafik Karra,
and Mohamed Haddar

Mechanics, Modeling and Production Research Unit, Mechanical Engineering Department,
National School of Engineers of Sfax, BP. 1173, 3038 Sfax, Tunisia
mohamed.taktak@fss.rnu.tn

Abstract. Acoustic radiation of sound sources is a major problem in applied and industrial acoustics which affects many disciplines. This tool can be used for the optimization of these sources (types, distribution, etc.…). In industry, machines and processes are composed of various sources of noise such as rotors, stators, gears, fans, vibrating panels, circulation of turbulent fluid, impact processes, electrical machines, internal combustion engines, etc. To simplify the computation of the acoustic fields radiated from these sources; these ones can be simply modeled by elementary or punctual sources consisting in pulsing spheres. This allows the computation of the radiated acoustic field easily. In the context the proposed work in this paper presents the results of the simulation of the acoustic fields (the acoustic pressure, intensity and directivity) in the case of two and four elementary sources for different configurations (vertical and horizontal linear arrangements). The cases when sources are in phase or in phase opposition are studied. For each configuration, these fields are plotted and analyzed to understand the acoustic behavior of each one. The effect of the geometric distribution of the sources is analyzed also using the directivity which indicates the acoustic behavior of each studied case (isotropic, bipolar, quadripolar).

Keywords: Acoustic radiation · Acoustic radial speed · Acoustic propagation · Directivity

1 Introduction

In machines and industrial environments, there are various sources of noise such as rotors, stators, gears, fans, vibrating panels, turbulent fluid circulation, impact processes, electrical machines, internal combustion engines, etc. Manufacturers and public authorities have been worried about noise from workshops for decades (Goelzer et al. 1995). It is possible to compute the acoustic field radiated of these sources. The acoustic field radiation is a generally a complex problem. To simplify this computation, real sources can be modeled as an arrangement of punctual elementary sources. The acoustic pressure, velocity or intensity field are then obtained by the summation of the corresponding

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 224–232, 2022.
https://doi.org/10.1007/978-3-030-85584-0_22

acoustic quantity radiated from each source at a given position and frequency. The sound pressure level of each source depends on the type of the noise of the source, the distance between the source and the receiver and the nature of the propagation medium. This kind of problem was studied in previous works like (Lim Choi and Lee 2006), (Liu and Yi 2011) and (Wang et al. 2015).

To study acoustic radiation, there are two used approaches:

- Consider elementary sources (monopole sources and their combinations) such as, the pulsating sphere, which has a simple description as presented in (Lim Choi and Lee 2006), (Liu and Yi 2011).
- -Take into account the geometry and the real characteristics of the radiating source as presented in (Wang et al. 2015).

The objective of the present work is the computation of acoustic fields in the case of linear vertical and horizontal antennas approaching real arrangement of acoustic sources. For each case of these configurations, the acoustic pressure radiation speed, acoustic intensityas well as acoustic directivity are computed and plotted.. The studied cases are presented by (MolleJuhln 2011), (Vallier Brasier 2018), (Bruneau 1998) who studied elementary sources much smaller than the wavelength of the used sound. They studied the acoustic radiation from point sources and their interactions. Their work is based primarily on the equations of pressure field, acoustic radial velocity, intensity and of the sound radiated by elementary sources. The studied cases consist on linear antenna composed by two and four elementary sources.

2 Acoustic Radiation from Elementary Sources

The simplest source to be mathematically modeled is an extremely small pulse sphere. Such a source is called a monopoly, point source or simple source. Any source that changes volume as a function of time can be approached by an elementary point source at frequencies where its radius is smaller than the wavelength. A closed loudspeaker is an example of a point source in this category (low frequencies). Any source of noise can be assimilated by a succession of elementary sources. In the proposed work, the two studied cases are the linear arrangements of elementary sources in vertical and horizontal line.

2.1 Case of Two Point Sources

It is easy to calculate the sound field generated by a combination of point sources. Since they are infinitely small, they do not interfere with each other and therefore the sound fields are simply added (Valier Brasier 2018).

Figure 1 present two point sources separated by $2d$ distance oscillating with flow rates q_1 and q_2.

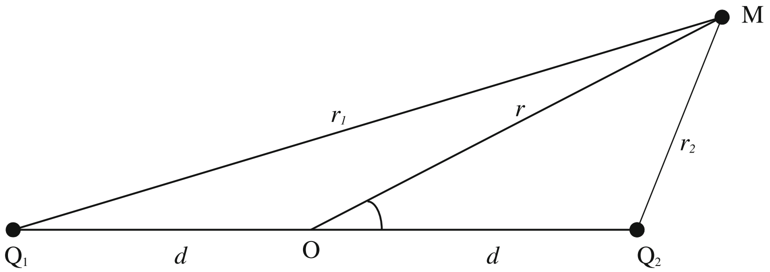


Fig.1. Representation of two distant sources (MollerJuhln 2011)

Two cases can be obtained:

Two Punctual Sources in Phase: The pressure field resulting from the radiation of two sources in phase with the sound flow rates of each are equal $q_1 = q_2 = q$ has the following expression:

$$p(r,\theta) = \frac{\rho_0 q}{4\pi}(\frac{\exp(-ik\sqrt{r^2+d^2+2dr\cos(\theta)})}{\sqrt{r^2+d^2+2dr\cos(\theta)}} + \frac{\exp(-ik\sqrt{r^2+d^2-2dr\cos(\theta)})}{\sqrt{r^2+d^2-2dr\cos(\theta)}}) \tag{1}$$

Taking into account the following approximations:

- far field: $d\,/\,r \;<<\; 1$
- The wavelength is greater than the distance d, that is, if $k.d << 1$, then the sound pressure can be put in the form:

$$p(r) = \frac{i\omega\rho_0 q_0}{4\pi r}\exp(-ikr) \tag{2}$$

The radial component of the acoustic particle velocity having the expression:

$$U(r,\theta) = \frac{q}{4\pi}(\frac{1}{r^2} + \frac{ik}{r})\exp(-ikr) \tag{3}$$

The acoustic intensity is then given by the relation:

$$I(r,t) = \frac{1}{2}\mathrm{Re}[p(r,t) * U_r^*(r,t)] = (\cos(kd\cos(\theta))^2 I_1 \tag{4}$$

Two Punctual Sources in Phase Opposition: The combination of two identical monopolies placed close to each other and expanding and contracting in phase opposition constitutes a dipole (Valier Brasier 2018). In this case, the acoustic flows of each opposed $q_1 = -q_2 = q$.

The acoustic pressure field therefore has the expression:

$$p(r,\theta) = \frac{\rho_0 q}{4\pi r} ikd\cos(\theta)\exp(-ikr) \tag{5}$$

The factor $\cos(\theta)$ represents the directivity of the pressure field radiated by a dipole:

$$D(\theta) = \cos\theta \tag{6}$$

The acoustic particle radial velocity having the expression (MollerJuhl 2011):

$$U_r(r) = ikd\cos(\theta)\frac{q_0}{4\pi}(\frac{1}{r^2} + \frac{ik}{r})\exp(-ikr) \tag{7}$$

The sound intensity is therefore equal to:

$$I(r,\theta) = \frac{1}{2}\mathrm{Re}[p(r,t) * U_r^*(r,t)] = [kd\cos(\theta)]^2 I_1(r) \tag{8}$$

2.2 Case of Four Punctual Sources

The studied cases with four punctual sources are presented in Fig. 2.

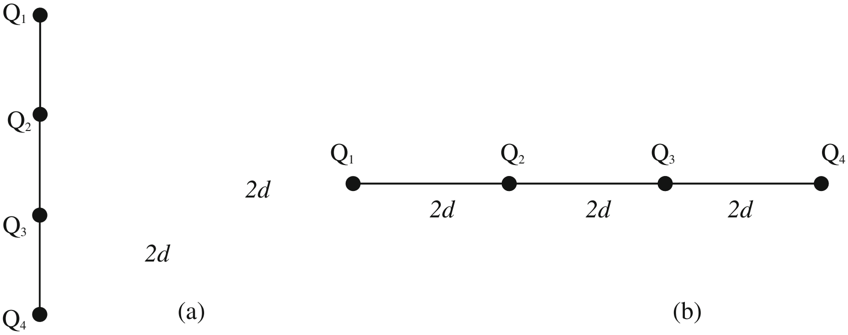


Fig. 2. The two studied linear arrangements: (a) Vertical four sources arrangement (b) Horizontal four sources (MollerJuhl 2011)

Vertical Quadruple

The pressure field resulting from the radiation of the four vertical sources has the approximate expression in the far field (MollerJuhl 2011):

$$p(r,\theta) = \frac{1}{2}\frac{\rho_0\, q}{4\pi r}\exp(-ikr)(kd)^2\sin(2\theta) \tag{9}$$

So the directivity factor is equal to:

$$D(\theta) = \sin(2\theta) \tag{10}$$

The radial component of the acoustic particle velocity having the expression:

$$U(r,\theta) = \frac{1}{2}\frac{q}{4\pi}(\frac{1}{r^2} + \frac{ik}{r})\exp(-ikr)(kd)^2\sin(2\theta) \tag{11}$$

The acoustic intensity is then given by the relation (MollerJuhl 2011):

$$I(r,t) = \frac{1}{2}\mathrm{Re}[p(r,t) * U_r^*(r,t)] = \frac{1}{4}[(kd)^2 sin(2\theta)]^2 I_1(r) \tag{12}$$

Horizontal Quadruple

In the far field, the acoustic pressure generated by four aligned horizontal sources has the approximate expression (MollerJuhl 2011):

$$p(r,\theta) = 2\frac{\rho_0\, q}{4\pi r}(kd)^2 \exp(-ikr)\cos^2(\theta) \tag{13}$$

The horizontal quadruple directivity therefore has the following expression:

$$D(\theta) = cos(\theta)^2$$

The radial component of the acoustic velocity has the expression (MollerJuhl 2011):

$$U(r,\theta) = 2\frac{q}{4\pi}\left(\frac{1}{r^2} + \frac{ik}{r}\right)(kd)^2 \exp(-ikr)\cos^2(\theta) \tag{14}$$

The acoustic intensity is then given by the relation:

$$I(r,q) = \frac{1}{2}\mathrm{Re}[p(r,t) * U_r^*(r,t)] = \frac{1}{4}[(kd)\cos(q)]^4 I_1(r) \tag{15}$$

3 Results and Discussions

3.1 Combination of Two Point Sources

Two spherical sources are placed in the middle of the $10 \times 10\ \mathrm{m}^2$ space. These sources are characterized by the same radius ($a = 0.009$m) and acoustic radial velocity in ($r = a$): $V_0 = 0.5$m/s at the frequency $f = 1$ kHz. The Fig. 3, 4 and 5 present respectively the sound pressure field (in Pa), the pressure level (in dB) and the directivity of each studied case (when the two sources are respectively in phase and phase opposition.

The Figs. 3, 4 and 5 show that the radiation of the acoustic pressure and the acoustic intensity is isotropic for the case of sources in phase, but not isotropic in the case of source in phase opposition. In addition, the amplitudes of the waves radiated at a certain point far from the sources are different in the two cases. These results are confirmed by the acoustic directivities of the studied cases which demonstrate the isotropic behavior in the first case and the dipole one in the second case as presented in Fig. 6.

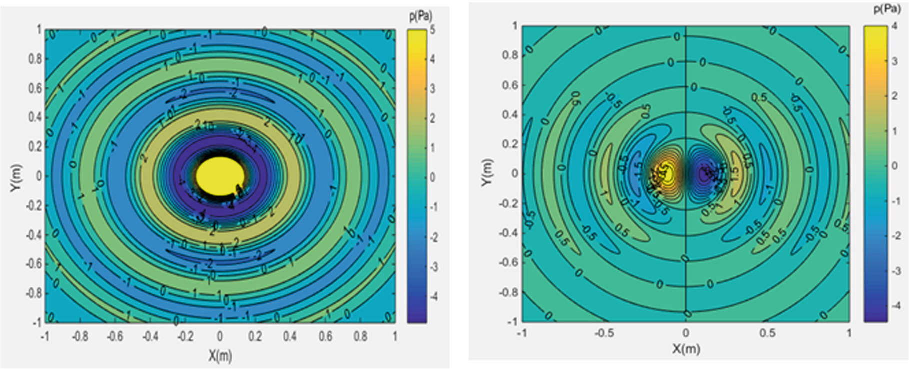
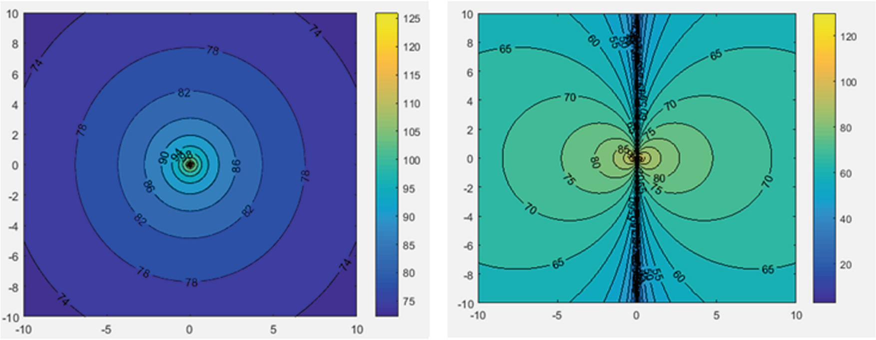


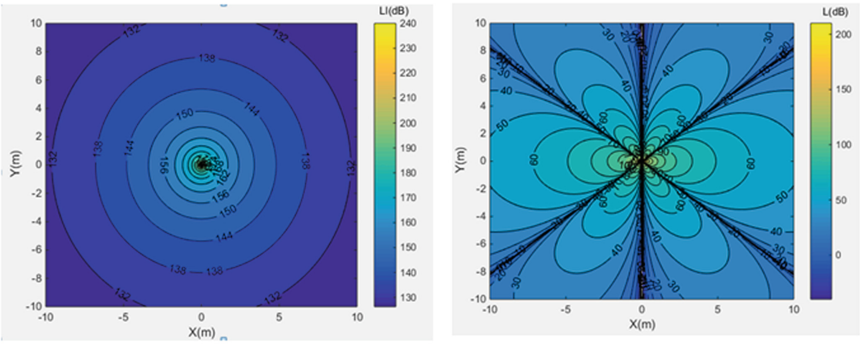
Fig. 3. The acoustic Sound pressure field of: (a) two point sources in phase and (b) two point sources in phase opposition.



a)

b)

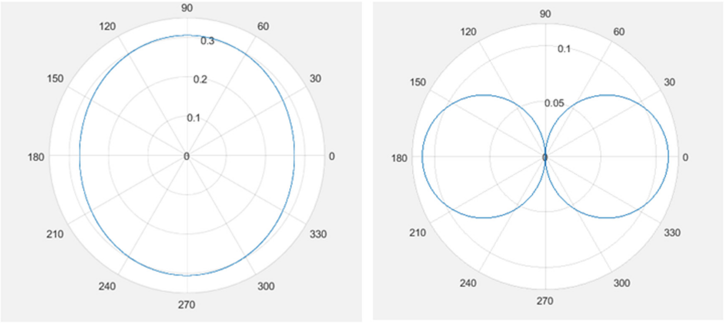
Fig. 4. The acoustic pressure level in dB: (a) two point sources in phase and (b) two point sources in phase opposition



a)

b)

Fig. 5. The acoustic intensity (a) two point sources in phase and (b) two point sources in phase opposition

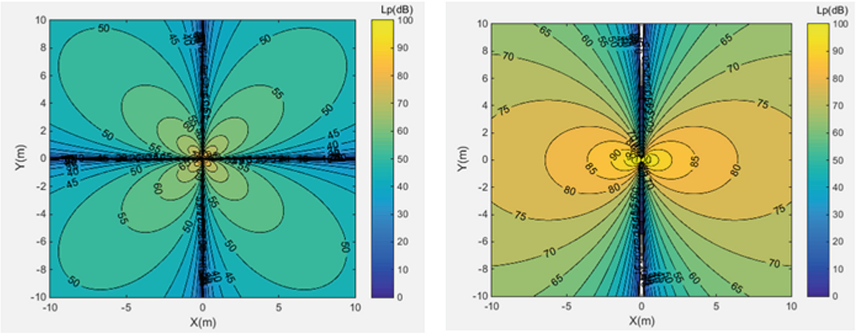


a) b)

Fig. 6. The directivity diagram (a) two point sources in phase and (b) two point sources in phase opposition

3.2 Combination of Four Point Sources

The considered sources are four vertical or horizontal sources as presented in Fig. 2. We consider four spherical sources placed in the middle of space. These sources are characterized by the same radius ($a = 0.05$m) and acoustic radial velocity in ($r = a$) and the frequency $f = 1$ kHz.

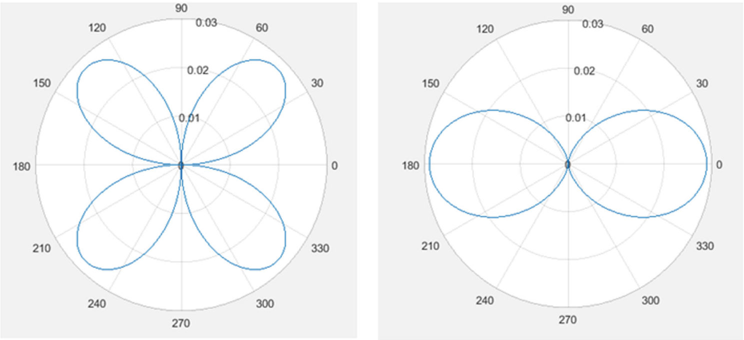


a) b)

Fig. 7. Sound pressure level (a) from four lateral and (b) longitudinal sources.

Figure 7 presents the acoustic pressure level respectively of the vertical arrangement (a) and horizontal one (b). It is observed that the acoustic pressure level of four lateral sources varies between 10 dB and 100 dB and have a large variation in the axes $x = 0$ and $y = 0$. However for longitudinal sources, the acoustic pressure level varied between 0 and 100 dB and it has the same shape like the case of two sources in phase opposition.

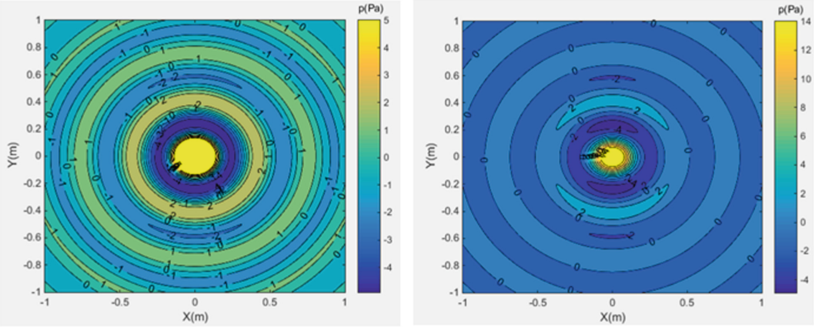
The modulus of the directivity factor D of four side sources is plotted in Fig. 8 as a function of the angle θ. It appears that the acoustic radiation has a maximum in $\theta = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ and is zero in $\theta = 0, \pi/2, \pi, 3\pi/2$. While it is maximum in $\theta = 0$ and π and is zero in $\theta = \pi/2, 3\pi/2$ for four longitudinal sources. So, the behavior of the acoustic radiation from the four sources is quadrupolar.



a) b)

Fig. 8. Directivity diagram (a) of four lateral sources and (b) longitudinal.

3.3 Influence of the Distance Between the Sources

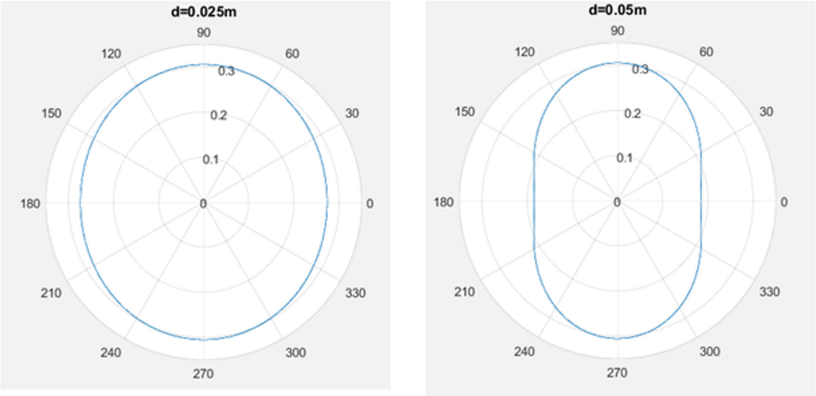


a) b)

Fig. 9. The acoustic pressure field of two sources in phase (a) $d = 0.025\text{m}$ and (b) $d = 0.05\text{m}$.

In this part, the influence of the distance between sources on the acoustic behavior of the sources systems is presented. This distance is varied between the two sources and the influence on the distribution of the pressure field is displaying in Fig. 9.

Figure 9 shows that if the distance between sources increases the pressure field values increase and it is no longer isotropic. This result is confirmed by the directivities plots presented in Fig. 10 which shows that when the distance between the sources increases, the directivity is no longer the same for all angles. It changes shape, and it becomes elliptical, which confirms that the field is no longer isotropic.



a)

b)

Fig. 10. The directivity diagram of two sources in phase (a) $d = 0.025$m and (b) $d = 0.05$m.

4 Conclusion

The simulation of acoustic radiation is a very important issue today. Indeed it allows to optimize the acoustic quality or even to reduce the noise in the workshops. So the work presented in this paper deals with the simulation of acoustic pressure and intensity as well as the determination of directivity radiated by elementary sources which can model real applications in the industry. Also, the influence of the distance between the sources on the radiated pressure level and on the directivity is studied. From the obtained results, it can be seen that the alignment of the sources (longitudinally or even laterally) and the positioning of the sources (in phase or in phase opposition) considerably influences the propagation of acoustic radiation.

References

1. Goelzer, B., Hansen, C.H., Sehrndt, G.A.: Occupational exposure to noise: evaluation, prevention and control. Book chapter (1995)
2. MollerJuhl, P.: Radiation of sound, Institute of Technology and Innovation University of Southern Denmark (2011)
3. Valier-Brasier, T.: Rayonnement acoustique, dissertation , Université Pierre et Marie Curie – Paris 6 (2018)
4. Bruneau, M.: Manuel d'acoustique fondamentale, Hermès Edition (1998)
5. Lim Choi, H., Lee, D.J.: Development of the numerical method for calculating sound radiation from a rotating dipole source in an opened thin duct. J. Sound Vib. **295**, 739–752 (2006). https://doi.org/10.1016/j.jsv.2006.01.017
6. Liu, Z., Yi, C.: On acoustic radiation by rotating dipole sources in frequency domain. Mod. Appl. Sci. **5**(5), 80–86 (2011)
7. Wang, G., Li, W.Y., Li, W.L., Du, J.: Sound Radiation from an elastically restrained plate covered by an acoustic decoupling layer, mathematical problems in engineering, vol. 2015 (2015). Article ID 659472. https://doi.org/10.1155/2015/659472.

Time Synchronous Averaging Based Detection of Bearing Defects

Mohamed Habib Farhat[1,2](✉), Xavier Chiementin[2], Fakher Chaari[1], Fabrice Bolaers[2], and Mohamed Haddar[1]

[1] Laboratory of Mechanics, Modeling and Production (LA2MP), National School of Engineers of Sfax, BP1173, 3038 Sfax, Tunisia
Mohamed-habib.farhat@enis.tn
[2] Institute of Thermics, Mechanics and Material (ITHEMM), University of Reims, Moulin de La Housse, 51687 Reims cedex 2, France

Abstract. Rolling element bearings are the most important rotating guide elements used in industry. They are used in a wide range of speeds and loads. They are found at very low speeds (actuator bearings), at high speeds (turbine shaft bearings) and are able to support high loads while maintaining a reasonable tolerance to misalignment. However, rolling element bearings are recognized among the critical components in rotating machinery and their failures can lead to catastrophic losses. This makes its efficiency a major requirement. Among all the diagnostic tools for machines used in the industry, vibration analysis remains the ultimate tool used for diagnosing bearing health. However, early detection of bearing failures based on vibration analysis remains challenging, especially for machines with multiple excitation sources. In this chapter the time synchronous averaging technique (TSA) is updated and used for the early detection of bearing defects. A dynamic model of a shaft bearing system is used to simulate the vibration response of ball bearing in the case of tiny defect localized in the baring outer race. The simulated signals are used to validate the effectiveness of the proposed technique. The results confirm the effectiveness of the adopted approach in the early detection of bearing defects.

Keywords: Bearings · Time synchronous averaging · Defects · Diagnosis · Detection

1 Introduction

Vibration monitoring of rotating machinery has become a subject of growing interest in modern industry. Bearings are the most frequent component in rotating machines and mastering their vibration behavior can considerably improve the machines diagnostic process, (Farhat et al. 2021). Many signal processing techniques are successfully adopted in the literature to highlight the impact of defects in vibration signals, namely filtering (Boudraa and cexus 2007), trend elimination (Smulko 2019) and de-noising (Chegini et al. 2019). Under special conditions, i.e. low speed shaft rotations, these conventional

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 233–241, 2022.
https://doi.org/10.1007/978-3-030-85584-0_23

processing methods cannot be applied confidently for bearing fault diagnosis (Mishra et al. 2016). In fact, bearings vibration signatures are generally low and dissolving in the operation noise (Farhat et al. 2020). In addition, under real operation conditions bearing defects usually begin with infinitesimal dimensions, making it difficult to early reveal their presence from vibration signals.

In general, the vibration signals measured from rotating machines contain three important components: the periodic frequencies, their harmonics and the operation noise (Wang et al. 2014). Averaging the signal synchronously to the rotation of the drive shaft can highlight the cyclic components of the signal and reduce its random disturbances. This is the principle of the Time-Synchronous Averaging technique (TSA), which has proven its effectiveness in the early detection of gear faults under constant or nearly constant operation speed condition (Halim et al. 2008). The synchronization with respect to the shaft rotation angle adopted in TSA highlights the components that occur in exactly the same way during each complete revolution of the shaft. This makes from this technique very effective for the detection of localized defects in gears. The impulses distribution of localized bearing defects, in contrast, are not the same for all shaft rotation cycles (same shape but shifted) (Mishra et al. 2016). This limit the effectiveness of the conventional shaft synchronous rotation-based TSA in the early detection of localized bearing defects. In this chapter, TSA has been updated and used as a part of an original unsupervised classification-based diagnosis approach developed for the early detection of bearing defects. The unsupervised classification algorithm used is DBSCAN (density-based spatial clustering of applications with noise). The latter has proven to be effective in the early detection of bearing defects in the work of (Kerroumi et al. 2013).

The paper is divided in 3 sections without considering the introduction. Section 2 presents the proposed methodology used for the early detection of bearing defects. In Sect. 3 the proposed approach has been validated. Conclusions are performed in Sect. 4.

2 Methodology

In this work, the TSA technique is updated and used in bearing early fault detection approach.

2.1 Update of TSA

Bearing failures are usually caused by defects localized in the outer ring, inner ring, or the rolling element. Under stationary conditions, bearing defect pulses occur at constant frequencies (Wang et al. 2014). The typical frequency associated with bearing are: the Ball Pass Frequency Outer Race (BPFO), the Ball Pass Frequency Inner Race (BPFI), the Ball Spin Frequency (BSF) and the fundamental Train Frequency (FTF).

The basic TSA algorithm consists to divide the signal into time intervals corresponding to the shaft rotation time periods and then compute their average. This method has proven to be effective in gear defects detection for the case of a constant operation speed. In fact, gear defect impulses exhibit at the same angular position in each shaft revolution. Therefore, averaging the signal corresponding to complete shaft revolution intervals will reinforce their presence and attenuate the non-synchronous components, such as measurement and mechanical noises. Bearing localized defects frequencies are proportional to the frequency of the shaft rotation. Despite this, and contrary to the case of gear defects, the distribution of bearing fault impulses is not the same at each complete revolution of the shaft. In fact, depending on the location of the fault (in the outer ring, inner ring or ball), bearing defect impulses tend to shift by a constant angle within each complete shaft revolution (while keeping the same occurring order). This shift would cause a loss in bearing defects vibration signature using the classical TSA.

To overcome this problem, TSA is updated in this work. The update consists to average the signal synchronously with the defects accruing cycles instead of shaft revolutions, avoiding the smearing due to accumulation of minor phase shifts. For the case of bearing race defects, the defect accruing cycle is defined as the intervals between 2 consecutive passage of a same ball over the defect. In the case of an outer race defect, the latter correspond to a complete revolution of the cage. Figure 1 illustrates (with exaggeration) the difference between the application of conventional and updated TSA in the case of an outer race defect.

The rotational speed of the cage is deduced mathematically from the speed of the driving shaft (speed of the inner ring of the bearing). The relationship between the instantaneous speed of the cage and the shaft is:

$$\mathrm{FTF} = \frac{\mathrm{R_f}}{2}\left(1 - \frac{\mathrm{B_d}}{\mathrm{P_d}}.\cos(\alpha)\right) \tag{1}$$

Where: $\mathrm{B_d}$ is the ball diameter, $\mathrm{P_d}$ and α represents the pitch diameter of the bearing and its contact angle respectively, (α is equal to zero in the current case), and $\mathrm{R_f}$ is the shaft rotation frequency.

2.2 Approach for the Early Detection of Bearing Defects

The approach proposed for the early detect of bearing degradation is described in Fig. 2.

The latter consists to preprocess the bearing vibration data using the updated TSA in order to highlight faults and minimize noise. Then, referring to the literature, effective bearing fault diagnosis features are extracted from the preprocessed data. Once features are extracted, Kernel Principal Component Analysis (KPCA) is applied to obtain the features combination that better presents the signals by eliminating redundancy. The extracted combination is taken as input to the DBSCAN classifier, used to perform the diagnosis. This classifier is chosen for its ability to be used without prior knowledge of data. It takes a decision on the newly acquired data if it corresponds to a healthy state or if it is degraded. For more details on DBSCAN, see the work of (Kerroumi et al. 2013).

3 Result and Discussion

This section is devoted to validate the proposed bearing defects early detection approach. The vibration signals used in the validation process are simulated by a dynamic model of a shaft bearing system previously developed by authors. The effectiveness of the updated TSA will be evaluated by comparing the performance of two DBSCAN classifiers fed with data processed using conventional TSA and updated TSA respectively in early bearing fault Detection.

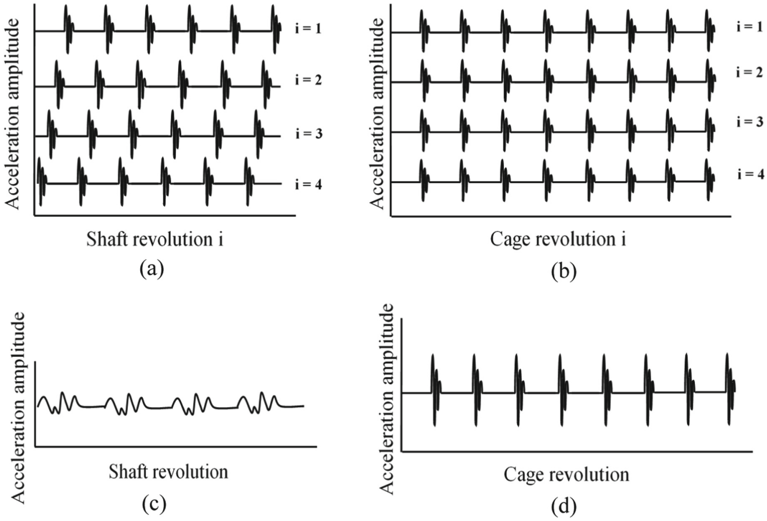


Fig. 1. Exaggerated example of conventional and updated TSA applied to outer race defect signals: (a) four signal with imperfect shifts with respect to the shaft rotation, (b) four signals with no shifts with respect to the cage rotation, (c) signal averaged synchronously to the shaft rotation (d) signal averaged synchronously to the cage rotation

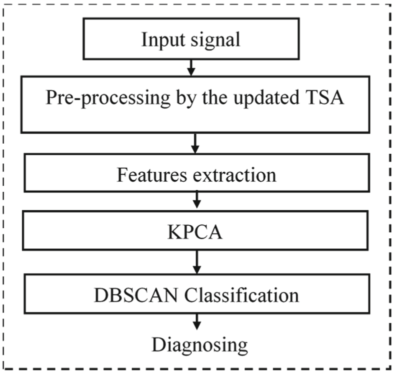


Fig. 2. Proposed bearing defects early detection approach.

3.1 Dynamic Model Simulation

In this study, the dynamic model of a shaft bearing system, proposed by (Farhat et al. 2021), shown in Fig. 3, is used to simulate bearing vibration signals. The latter consists of a synchronous electric motor driving a rotating shaft, supported by two ball bearings type 6206 and charged radially by a load-controllable hydraulic actuator.

This model dynamic is presented using 9 nodes (see Fig. 3). Here authors are limited to 3° of freedom (DOFs) in each node: two translation (DOFs) which represents deflection in $\vec{v}$ and $\vec{w}$ direction and one rotational (DOF) representing the torsion in

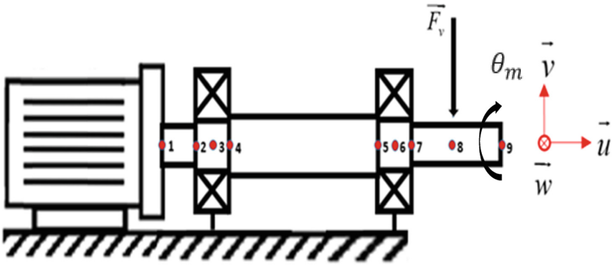


Fig. 3. Dynamic model of a shaft bearing system, (Farhat et al. 2021).

$\overrightarrow{u}$ direction. All the simulation carried out in this chapter are realized in the node 6, direction $\overrightarrow{v}$. The sampling frequency is fixed to 20 kHz. The motor rotation speed is fixed to 2000 rpm and the radial load F_v is fixed to 100 KN.

In total 100 signals of 1s each are generated: 20 signals corresponding to healthy bearing and 80 signals corresponding to 80 severities of a localized defect in the outer race of the load-side bearing. The first considered defect present a width d = 0.01mm. Then the defect width increases by 0.025 mm from one severity to another, so that the last severity corresponds to $0.01 + 79 * 0.025 \approx 2$mm. White noise of constant amplitude is added to all generated signals to simulate real operation noise.

3.2 Pre-processing

All the simulated signals corresponding to the healthy and defective cases, respectively, are preprocessed using the conventional and the updated TSA. The preprocessing process is further clarified in this section using the vibration signal shown in Fig. 4, which corresponds to a simulated vibration data for a defective bearing with an outer race defect of severity 40 (defect width = 0.985 mm).

From the raw vibration signal shown in Fig. 4, it is difficult to detect the presence of fault impulses, those are mostly masked by noise. A single peak corresponding to the fundamental frequency of the bearing outer race defect, BPFO = 118.9Hz, is detected in the raw signal spectrum. Conventional and updated TSA are respectively applied to the considered signal. The conventional TSA-based pre-processed signal and its envelop spectrum are given in Fig. 5. The updated TSA-based pre-processed signal and its envelop spectrum are given in Fig. 6.

From Fig. 5 and Fig. 6, the defect impulses in the time signal, the fundamental defect frequency, BPFO = 118.9Hz, in the spectrum and its harmonics are clearer when performing the updated TSA pre-processing.

3.3 Features Extraction

Considering their effectiveness in the diagnosis of bearing vibration signals, 10 features from the time domain are considered to be extracted from simulated signals. The expressions of the chosen features are summarized in Table 1.

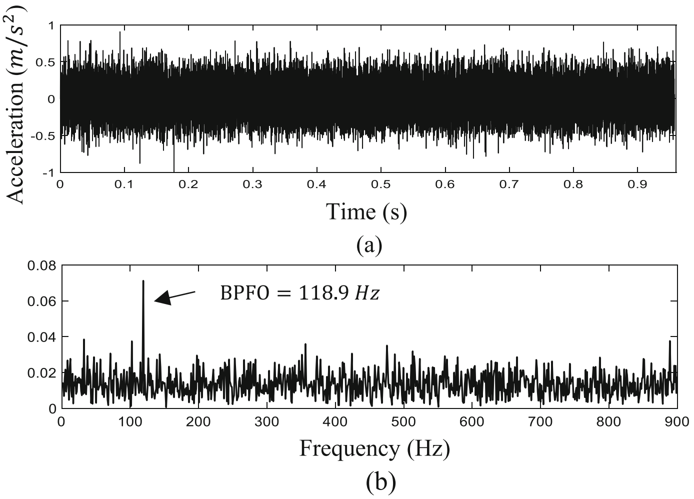


Fig. 4. Case of outer race defect of severity 40: (a) raw vibration signal and (b) squared envelope spectrum

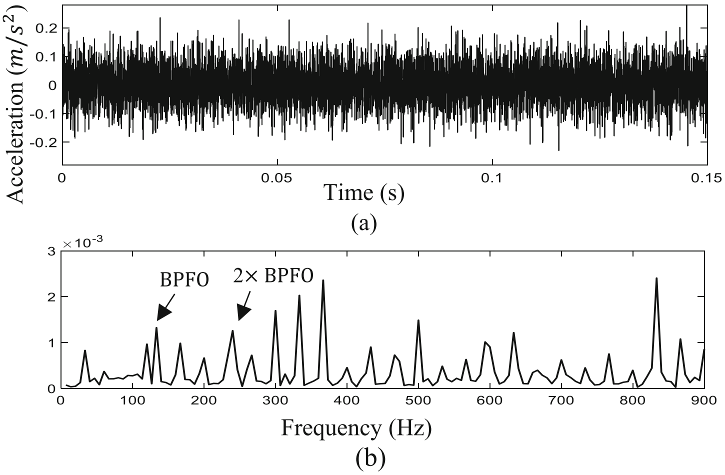
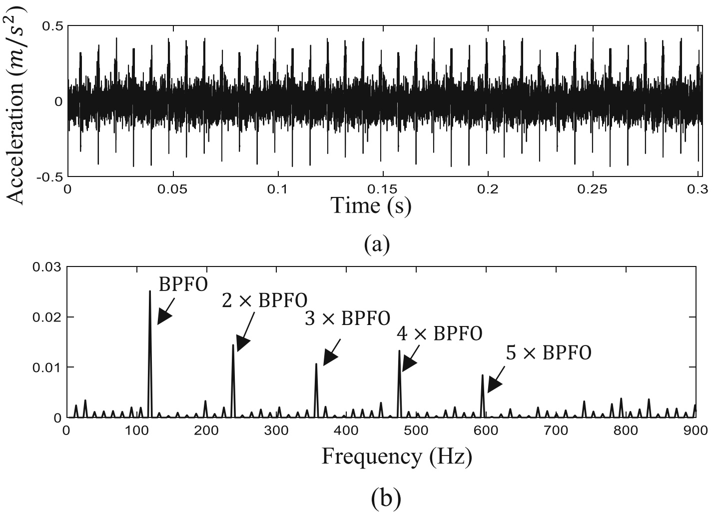


Fig. 5. Case of outer race defect of severity 40: (a) conventional TSA averaged vibration signal and (b) squared envelope spectrum

3.4 Classification

The DBSCAN classification algorithm is used to detect the degradation from the considered bearing signals. Referring to (Kerroumi et al. 2013), the DBSCAN imputation parameters *Eps* and *Minpts* are set to 10 and 4 respectively. Figure 7(a) and Fig. 7(b) shows the classification results of the DBSCAN algorithm for the case of input signals processed with the conventional TSA and with the updated TSA, respectively. The green stars correspond to signals classified as healthy and the red stars correspond to signals classified as defective.



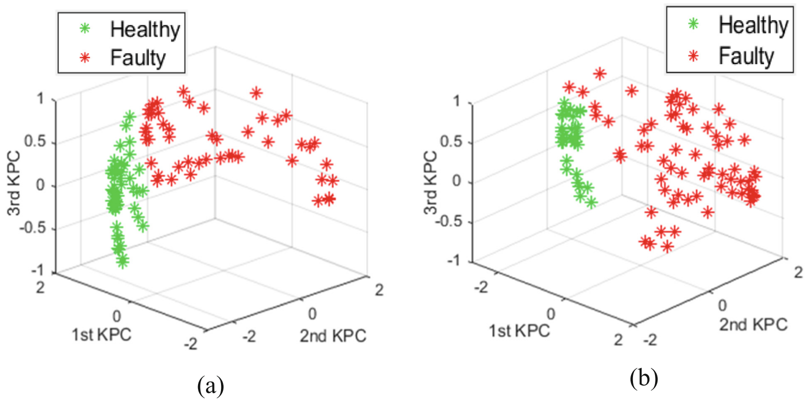
(a)

(b)

Fig. 6. Case of outer race defect of severity 40: (a) conventional TSA averaged vibration signal and (b) squared envelope spectrum

Table 1. The computed features, x represents the digitized signal, x_i is the sample number. $i \in (1, 2 \ldots N)$.

Root mean square	Mean	Energy
$\mathrm{RMS} = \left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{\frac{1}{2}}$	$\mathrm{mean} = \frac{1}{N}\sum_{i=1}^{N} x_i$	$E = \sum_{i=1}^{N} x_i^2$
Kurtosis	Shape factor	Skewness
$\mathrm{Ku} = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{x_i - \bar{x}}{\sigma}\right)^4$	$\mathrm{Sf} = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} x_i^2\right)^{\frac{1}{2}}}{\frac{1}{N}\sum_{i=1}^{N}\lvert x_i\rvert}$	$\mathrm{SK} = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - \bar{x})^3}{(N-1)\sigma^3}$
Peak	Standard Deviation	Crest factor
$\mathrm{Peak} = \max(\lvert x_i\rvert)$	$\mathrm{Std} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})}$	$\mathrm{CF} = \frac{\mathrm{Peak}}{\mathrm{RMS}}$



(a) (b)

Fig. 7. DBSCAN classification result: (a) case of conventional TSA-based preprocessing and (b) case of updated TSA-based processing. KPC: Kernel Principal Component.

For the classifier fed with data processed using the classical TSA, the degradation is detected from signal number 46, i.e. the first detected fault is of severity 26. On the other hand, for the one fed with data processed with the updated TSA, the first detected fault is of severity 9. Confirming the effectiveness of the updated TSA technique in the early detection of bearing defects.

4 Conclusion

In this chapter, an original approach is investigated for the early detection of bearing defects. First, the conventional TSA preprocessing technique is updated and applied to bearing outer ring defect signals. Then, the preprocessed signals are used to feed a DBSCAN classifier used for fault detection. Therefore, the effectiveness of the proposed features is evaluated based on the performance of the constructed classifier. The proposed approach has proven to be effective in early detection of bearing faults.

In the following work, the proposed approach will be tested on experimental degradation data.

References

1. Boudraa, A.O., Cexus, J.C.: EMD-based signal filtering. IEEE Trans. Instrum. Meas. **56**(6), 2196–2202 (2007)
2. Chegini, S.N., Bagheri, A., Najafi, F.: Application of a new EWT-based denoising technique in bearing fault diagnosis. Measurement **144**, 275–297 (2019)
3. Farhat, M.H., Hentati, T., Chiementin, X., Bolaers, F., Chaari, F., Haddar, M.: Numerical model of a single stage gearbox under variable regime. Mechanics Based Design of Structures and Machines, pp. 1–28 (2020)
4. Farhat, M.H., Chiementin, X., Chaari, F., Bolaers, F., Haddar, M.: Digital twin-driven machine learning: ball bearings fault severity classification. Measur. Sci. Technol. **32**(4), 044006 (2021)

5. Halim, E.B., Choudhury, M.S., Shah, S.L., Zuo, M.J.: Time domain averaging across all scales: a novel method for detection of gearbox faults. Mech. Syst. Signal Process. **22**(2), 261–278 (2008)
6. Kerroumi, S., Chiementin, X., Rasolofondraibe, L.: Dynamic classification method of fault indicators for bearings' monitoring. Mech. Ind. **14**(2), 115–120 (2013). https://doi.org/10.1051/meca/2013058
7. Smulko, J.: Methods of trend removal in electrochemical noise data–overview. Measurement **131**, 569–581 (2019)
8. Mishra, C., Samantaray, A.K., Chakraborty, G.: Rolling element bearing defect diagnosis under variable speed operation through angle synchronous averaging of wavelet de-noised estimate. Mech. Syst. Signal Process. **72**, 206–222 (2016)
9. Wang, T., Liang, M., Li, J., Cheng, W.: Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis. Mech. Syst. Signal Process. **45**(1), 139–153 (2014)

A Short Review of Gear Fault Modelling in a Hybrid Modelling Context

Luke van Eyk(✉), P. Stephan Heyns, and Stephan Schmidt

Centre for Asset Integrity Management, University of Pretoria,
Pretoria, South Africa

Abstract. Gears are commonly used in many critical engineering applications and vulnerable to various modes of failure. Prognostic health management (PHM) of these components is therefore often desirable. Accurate physical models of the gearboxes and its failure modes that are fast to compute have several benefits for PHM. Therefore, analytical gear fault modelling methods, obtained from the potential energy method (PEM) received significant attention. This is because it aims to deliver the accuracy of finite element methods, whilst being much faster to solve. Hence, it is suitable for PHM. This chapter gives a short overview of the modelling philosophy behind spur gears. This philosophy includes the theory behind the PEM and the methods used to induce faults to these physical gear models. The chapter continues by reviewing state-of-the art root crack, chip, surface pit, spall and broken tooth models and showcases the expected meshing stiffness phenomena expected for each fault type. Finally, this chapter gives recommendations for the gear fault modelling field and proposes the use of a generalised stiffness model, which would simultaneously benefit both the PHM task and the gear modelling research community.

Keywords: Gear fault modelling · Physical modelling · Potential energy method · Prognostic health management · Time-varying mesh stiffness

1 Introduction

Gears are very commonly found in many critical engineering applications such as large wind turbines. Prognostic health management (PHM) of these gears is therefore often highly desirable and financially justified. PHM usually require some form of physical modelling of the gears and their dynamic interactions with each other. For gears, these models may be complex because they are often subjected to various modes of failure.

Conventional PHM methods are either data-driven or physics-based modelling, which have several shortcomings for practical applications. Recently, hybrid methods for PHM of geared systems have taken root (Sun et al. 2020). Such hybrid methods combine physics-based models with data-driven models. This requires generalized low computational cost physics-based gear models that

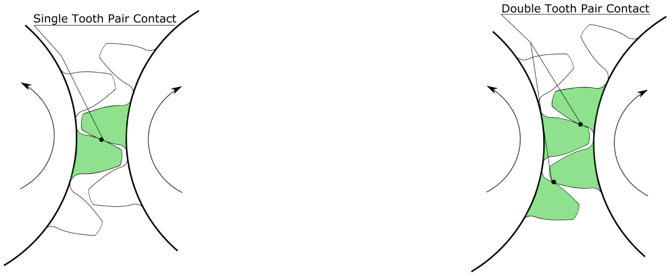
A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 242–258, 2022.
https://doi.org/10.1007/978-3-030-85584-0_24

are digital representations of the physical system. This enables the generation of digital condition monitoring data under different health and operating conditions and therefore makes it possible to perform the inverse diagnosis problem, i.e. infer the health state from the available measurements. The models need to be calibrated to ensure that the predicted response is the same as the measured response of the physical system under equivalent measurement conditions. Once a hybrid model has been created and calibrated, this forms an integral part of the remaining useful life estimation part of PHM.

The vibration signals acquired from gearboxes are dominated by gear mesh interactions. The interactions are usually modelled with one of the following methods:

- The Square Waveform Method (SWM)
- The Potential Energy Method (PEM)
- The Finite Element Method (FEM)

As the gears rotate, the gear teeth transmit forces between shafts. Depending on the rotational angle of the shaft, either one or two gear teeth pairs are in contact (See Fig. 1). This causes the stiffness between gears to discontinuously jump up and down as a single or a double tooth pair comes into contact. This behaviour is modelled by the SWM (see Fig. 2, left). A second phenomenon occurs, namely that as a gear rotates, the angle and position of the applied force changes over the tooth face. This causes a secondary stiffness variation effect, which the PEM tries to account for (see Fig. 2, middle). The FEM automatically accounts for both meshing effects as a result of the stiffness matrix from the FEM (see Fig. 2, right).



(a) Single tooth pair meshing (b) Double tooth pair meshing

Fig. 1. Illustration of the two gear tooth pair meshing cases that need to be modelled: on the left one observes that only a single pair of teeth are in contact. On the right one observes that two pairs of gear teeth are simultaneously in contact. The teeth in contact are shown in green shadings.

The SWM is limited in its flexibility to model different damage modes but serves as a good phenomenological starting point that is very easy to implement.

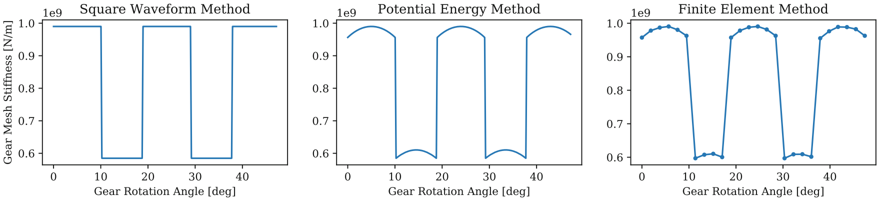


Fig. 2. Time-varying gear mesh stiffnesses for the same gear pair given for the three main methods. On the left is the square waveform method, in the middle, the potential energy method, and on the right, the finite element method

In the SWM, the damage is modelled as a percentage reduction in gear mesh stiffness. However, the gear mesh stiffness reduction is not connected to quantifiable damage parameters (e.g. crack sizes), which makes it ill-suited for health severity estimation and prognostics.

The FEM is considered to be the most accurate method of determining gear meshing stiffness and has the advantage that it can be used to calculate critical stresses for remaining useful life estimation. However, due to the non-linear phenomena in the gearbox (e.g. contacts between gear teeth), it is much more difficult to solve and is also computationally slow.

The PEM is capable of modelling the gear mesh stiffness with the same accuracy as the FEM while having the same computational speed as the SWM. The PEM relates changes in the physical dimensions of the damage (e.g. crack length) to changes in gear mesh stiffness, which makes it is especially well-suited for PHM. For this reason, a large amount of literature is available on the PEM and this constitutes the focus of this chapter.

2 The Potential Energy Method

The potential energy method aims to derive the stiffness terms between gears by calculating the stored potential energy between the two gear teeth during meshing. This method was first introduced by Yang and Lin (1987). The meshing stiffness is inversely proportional to the potential energy, and thus through applying the potential energy method we may indirectly also determine the meshing stiffness. The meshing stiffness is usually derived through three main physical means, namely a beam stiffness term, a Hertzian contact stiffness term and a gear foundation stiffness term (Mohammed and Rantatalo 2020).

The beam stiffness term is derived directly from cantilever beam theory, where one models the gear tooth is modelled as a non-uniform cantilevered beam (See Fig. 3). From beam theory, three beam stiffness terms are considered, namely the axial compressive, bending and shear stiffnesses. The comprehensive equations of these beam stiffness terms (based on Fig. 3) are given below:

$$U_a = \frac{F^2}{2k_a} = \int_0^d \frac{F_a^2}{2EA_x} dx \tag{1}$$

$$U_b = \frac{F^2}{2k_b} = \int_0^d \frac{\left[F_b(d-x) - F_a h\right]^2}{2EI_x} dx \tag{2}$$

$$U_s = \frac{F^2}{2k_s} = \int_0^d \frac{1.2F_b^2}{2GA_x} dx = \int_0^d \frac{1.2(1+\nu)F_b^2}{2EA_x} dx \tag{3}$$

In the model of Fig. 3, d_1, x and d refer to the horizontal fillet length, the horizontal distance from the tooth root and the horizontal distance of the applied force from the tooth root, respectively. The h_x and h parameters refer to the tooth height from the centreline at the horizontal distance x and the tooth height from the centreline and the applied force point. R_r, R_b and R_h respectively refer to the tooth root radius, base radius and hub radius. The angular values of α_1, α, α_2 and α_3 refer to the angular equivalents of h, h_x, d_1 and R_r respectively, where the angles are always determined by extending from the contact point on the tooth surface to be tangent to the tooth base. The only exception is α_3, which is defined in reference to the root circle. Finally, F_a and F_b refer to the decomposed force vectors experienced by the tooth.

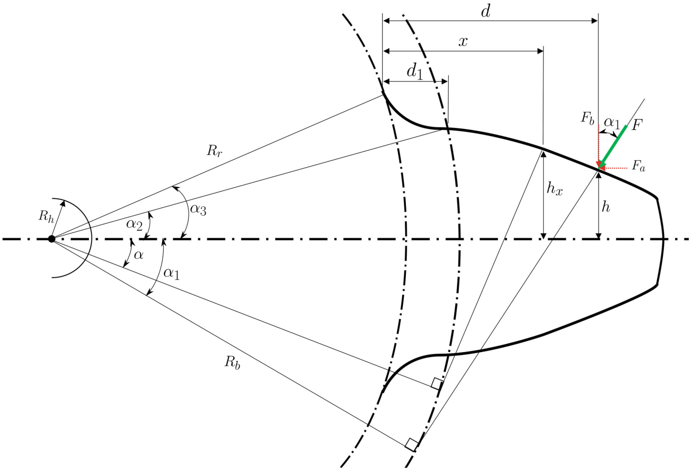


Fig. 3. Gear tooth modelled as a cantilever beam.

Given the non-uniform nature of the gear tooth, the beam stiffness integrals do not have any clear anti-derivatives and are thus reported in integral form and solved numerically.

The Hertzian contact stiffness is focussed on describing the contact line stiffness which occurs due to local deformation at the contact line across the width of the gear tooth pair. The contact surfaces are assumed to be cylindrical. The Hertzian contact potential energy may thus be described as:

$$U_h = \frac{F^2}{2k_h} = \frac{F^2}{2} \cdot \frac{4(1-\nu^2)}{\pi EL} \tag{4}$$

Finally, the gear foundation stiffness tries to account for the deflection the gear body (the portion beneath the tooth root and above the hub radius) undergoes due to forces applied to the gear teeth. This is given as:

$$U_f = \frac{F^2}{2k_f} = \frac{F^2}{2} \cdot \frac{\cos^2 \alpha_1}{EL} \left[L^* \left(\frac{\mu_f}{S_f} \right)^2 + M^* \left(\frac{\mu_f}{S_f} \right) + P^* (1 + Q^* \tan^2 \alpha_1) \right] \tag{5}$$

where μ_f, S_f, L^*, M^*, P^* and Q^* are determined by following the approach in Sainsot et al. (2004).

It is important to take note that the foundation stiffness is derived under the assumption of single tooth pair engagement. Thus, for the case of double tooth pair engagement, the foundation stiffness contribution is overestimated, since it is added to the overall stiffness of the gear mating process twice - one term for each gear pair. This double addition overestimates the overall stiffness, as it assumes that each mating gear pair has its own gear body, where in reality both teeth pairs are sharing the same bodies. Since no analytical correction exists, it is important to follow the procedure outlined in Ma et al. (2016), where the double tooth region fillet terms are calibrated through FEM, leading to a more realistic overall stiffness.

This concludes the description of the meshing stiffness terms, which are solved by substituting expressions for F_a, F_b, d, x, A_x and I_x into Eqs. (1)–(3). Clear definitions of these expressions are given in Meng et al. (2021). Notice, however, that the beam terms in Eqs. (1)–(3) have A_x and I_x terms, and the Hertzian term in Eq. (4) has an L term. It will shortly be shown that these are the key quantities that are varied over the tooth profile to simulate faults.

3 Development of Gear Fault Modelling

Gear fault modelling using the potential energy method has received significant attention. Many fault types have been modelled and in this work, the state-of-the-art for five types of gear faults are summarised, namely:

- Root Crack Faults
- Broken Tooth Faults
- Chipping Faults
- Spalling Faults
- Surface Pitting Faults

For all of these fault types, the h_x term in Fig. 3 is firstly parameterised as a function of the location on the tooth x. Thereafter, the A_x and I_x terms in the integrals of Eqs. (1)–(3) are updated to reflect the local surface area and area moment of inertia of the damaged tooth. This manifests as a reduction in the mesh stiffness.

Furthermore, some fault types aim to reduce the effective width of the tooth (L), which again directly affects Eqs. (1)–(3), but also the Hertzian term from Eq. (4). The gear foundation term given in Eq. (5) is not reduced for the considered fault types.

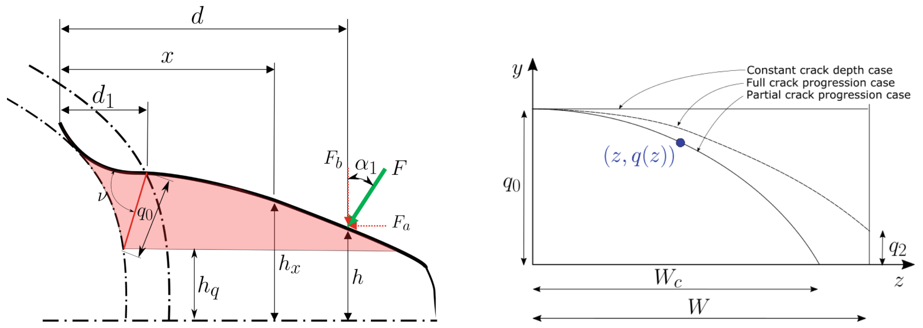
The following sub-sections will give a short overview of the state-of-the-art models being employed for each fault type mentioned above, with comparisons given against older methods. These sub-sections will conclude with gear mesh stiffness profiles to indicate what the expected effect of a fault type may be on the meshing stiffness. Although it is generally of great importance to validate any numerical results, the validation is beyond the scope of this work. This is because the purpose of this chapter is to review existing models for gear fault modelling. In this chapter, we show the behaviour of the gear mesh stiffness profiles for different damage models. These models were published in the literature and presented without experimental validation.

3.1 Crack Fault Modelling

Gear crack modelling received much attention in the literature, and the state-of-the-art gear crack models are quite flexible and advanced (Mohammed et al. 2013). The models discussed here refer to gear tooth root cracks, which if allowed to progress, lead to another fault type, namely a broken tooth fault.

Before reviewing the literature briefly, it is useful to describe how crack faults are generally implemented. Referring to Fig. 4(a), note that the height of the gear tooth above the centreline is defined as h_x. However, one assumes that the effective useful area of the tooth is reduced when a crack enters the tooth root - thus the effective tooth material that has been lost due to the crack is highlighted in red. Notice the parameter h_q, which defines the new effective height of the tooth. Thus, for any integration values where h_x would have otherwise been larger than h_q, we set $h_x = h_q$.

The first gear crack model utilising the PEM was proposed by Tian (2004), where the gear crack is assumed to be a full width crack, with a constant crack angle. The constant crack angle assumption is generally deemed accurate enough according to experimental results from Lewicki (2002). In the model, the depth of the crack could be varied, in turn affecting only the h_x term during integration as discussed above. However, due to the way in which forces are applied to the gear tooth, the fault is only allowed to affect the bending and shear terms, not the axial compressive stiffness. It is assumed that the axial compressive stiffness will be unaffected as the crack is assumed to still bear the load of axial compressive forces.



(a) Illustration of how the effective height of the tooth is assumed to be reduced within the crack depth region seen in red. We define a crack depth (q_0) and angle (ν) which determine how much material is virtually removed from the tooth stiffness. Figure adapted from Chen & Shao (2011).

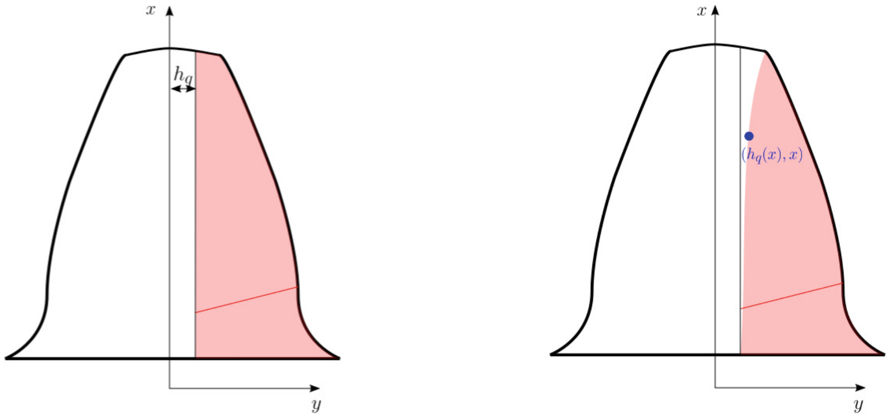
(b) Illustration of how the varying crack width is modelled.Here z describes the position on the tooth width, and $q(z)$ defines the variable crack depth (which was previously assumed as a constant q_0) as a function of the tooth width. Figure adapted from Chen & Shao (2011).

Fig. 4. Illustration of the model used to define constant width cracks (a) and the methodology proposed by Chen and Shao (2011) to improve upon this method (b).

As Tian (2004) only considered a crack depth above that of the central line of the tooth, Wu et al. (2008) extended the model to include cracks beyond the tooth midpoint, extending to below the centreline as well. Up to this point, neither Tian (2004) nor Wu et al. (2008) considered the fillet foundation effect. In Chaari et al. (2009) this effect is considered, and the crack methodology using PEM is shown to correlate very well with FEM results. However, Chaari et al. (2009) do not assume that a crack affects the full effective height of the tooth, but only affects the root area where the crack occurs.

Chen and Shao (2011) extended the crack formulation to include a variable width crack by utilising a parabolic distribution along the tooth width (See Fig. 4b). This allowed for more flexible crack shapes to be modelled. This is an important addition to the crack modelling process, as from practical experience, teeth do not always experience a uniform load across the tooth face, leading to cracks that propagate in depth and width. Chen and Shao (2011) managed to validate their updated model against the FEM results from Chaari et al. (2009).

Mohammed et al. (2013) developed a novel method to reduce the effective height of the tooth. Previously, in the cracked region, a constant height value of h_q was assumed across all x values as seen in Fig. 5(a). However, Mohammed et al. (2013) noted that as the tooth cracks, it will deflect, also changing the area available for carrying axial compressive, bending and shear loads. They generated a formulation which generates a new parabolic curve for h_q as a function of x, which was shown to be noticeably more accurate for cracks that are deeper than 30% of the gear thickness. An illustration of the new improved method is given in Fig. 5(b).



(a) Previous method of effective tooth height reduction assuming a straight reduction line

(b) Updated effective tooth height reduction according to a parabolic shape

Fig. 5. Illustration of the old (a) and new (b) method of modelling effective tooth height as described by Mohammed et al. (2013).

Finally, Ma et al. (2014) did an in-depth study comparing different combinations of gear fault modelling methodologies to FEM. The conclusion was that the method described by Mohammed et al. (2013) should be used. Furthermore, their works show that there is no appreciable difference between modelling the crack propagation path as a curve as opposed to a straight line. This is in line with what has been found experimentally from Lewicki (2002).

Although many other improvements have been made in the literature, they are mainly focussed on the overall stiffness modelling and not the crack modelling itself. Therefore, the model from Mohammed et al. (2013) is still deemed the standard on which gear root crack models should be based.

A mesh stiffness sample from a cracked tooth (varying width effect but not using parabolic limiting line and assuming a linearly growing crack) may be seen in Fig. 6.

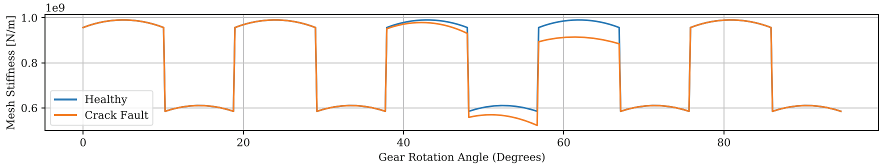


Fig. 6. Effect of a cracked gear tooth on the gear meshing stiffness

3.2 Broken Tooth Modelling

There are two simple methods to model a broken tooth. Firstly, a broken tooth can be modelled as a full-width crack with the crack model described in the previous section. Alternatively, one may recognise that for a broken tooth, one tooth's entire meshing period will contribute no stiffness. Thus, one may set up the healthy gear meshing stiffness array and simply subtract the broken tooth's contribution to the stiffness array, as implemented in Tian (2004). Effectively we set Eqs. (1)–(5) to ∞ for the broken tooth, as the stiffness is inversely proportional to these energy terms. What one sees in the stiffness results, is that during the first double tooth pair meshing region, the stiffness starts decreasing rapidly, as only one tooth pair is carrying the load. During the single tooth meshing region, there is no stiffness contribution, and the stiffness terms are zero. Finally, during the last double meshing period, the one tooth is still unable to contribute, and the other tooth pair is left to carry all the load, leading to a reduced stiffness. A mesh stiffness example from a broken tooth is shown in Fig. 7.

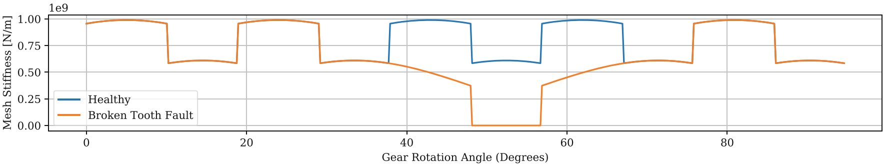


Fig. 7. Effect of a broken tooth on the gear meshing stiffness.

3.3 Chip Modelling

Tooth chip modelling has received relatively little attention in literature and the progression of this field can briefly be summarised in three key works. Initially, Tian (2004) proposed a methodology that assumed that the depth of chips were negligible, and the only effect to be accounted for would be the effective tooth width (L) reduction on the Hertzian stiffness term.

Since then, works from Chaari et al. (2008) accounted for the depth of the chipping fault, further affecting the axial compressive, bending and shear stiffnesses due to the newly affected A_x and I_x terms. However, in their modelling of the chipping parameters, Chaari et al. (2008) assume the tooth surface fracture curves to be linear in nature.

Liu et al. (2021) mention that the linear fracture curve assumption from Chaari et al. (2008) may lead to large deviations from the true fracture shape when the chips are deep. Therefore, Liu et al. (2021) developed an analytical geometry method that describes the fracture curves precisely. This method was validated against FEM results, and is currently the state of the art in chip modelling. An illustration of the chipping modelling may be seen in Fig. 8. Notice that we have full control over setting the location (on the vertices) of crack nodes A, B and C. The method automatically determines the connecting curves between these points. Furthermore, we note that the red region corresponds to the now missing material, which as mentioned before affects the A_x, I_x and L terms.

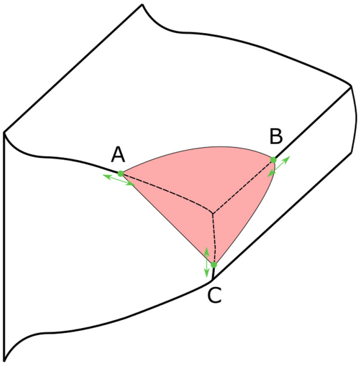


Fig. 8. Illustration of the state-of-the-art chip model. Each letter represents a different parameter that can be varied to describe the chip shape. The resulting fracture curves are an inherent property of the analytical geometry method. Figure adapted from Liu et al. (2021).

A mesh stiffness sample from a chipped tooth (as implemented by Tian (2004)) may be seen in Fig. 9.

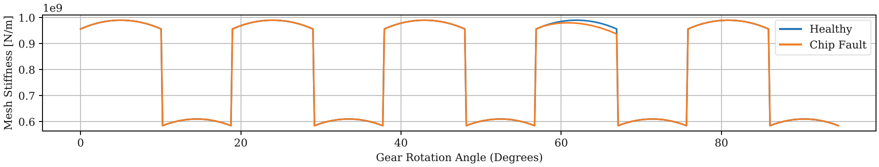


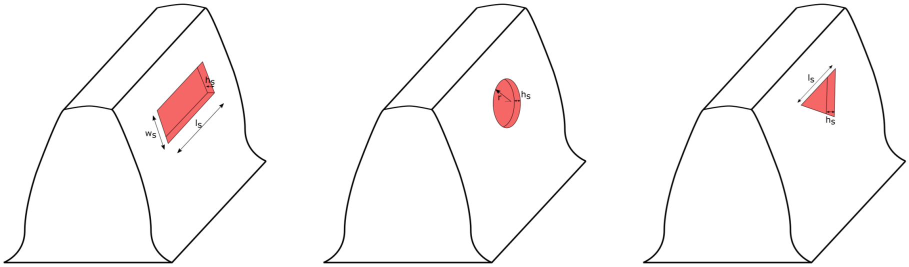
Fig. 9. Effect of a chipped gear tooth on the gear meshing stiffness

3.4 Spalling Fault Modelling

From the outset, it should be mentioned that there is an overlap between spall and pit modelling. For this review we assume that a pit refers to a small piece of regularly shaped material that is removed from a gear tooth. Pits generally occur in groups. Spall refers to a large piece of material that is missing from a tooth face, usually caused by the accumulation of the smaller pits. Thus, for this review, although spalling and pitting are covered in different sections, they are somewhat interchangeable with the main difference between the two being the fault size and amount.

Chaari et al. (2008) proposed the first analytical method to induce a spall defect on a gear tooth. This was done by defining the spall as a rectangular cut-out from the tooth profile, meaning that one can fully define the fault through the rectangle width, height, depth and starting point along the tooth length. Note once again that for Eqs. (1)–(3), the h_x term will be affected, and in turn thus the A_x and I_x terms. Furthermore, for Eqs. (1)–(4), the effective length L will also be affected, therefore affecting the stiffness terms.

The next major work on the PEM would only come in 2016, when Saxena et al. (2016) took all of the previously disjunct literature and formulated it into the PEM implementation. Furthermore, they derived a new stiffness term for torsional effects due to unsymmetric defects on gear teeth. They presented three spall models in this work: The rectangular shape from Chaari et al. (2008) (Fig. 10a), a circular shape from Jia and Howard (2006) (Fig. 10b) and finally a V-shape from various other authors (Fig. 10c). This work included a well documented analytical method of implementing these three fault shapes.



(a) Rectangular spall model (b) Circular spall model (c) V-shape spall model

Fig. 10. Illustration of the three constant depth modelling techniques with the specified parameters. These figures are adapted from Saxena et al. (2016).

Luo et al. (2018) further improved the spall modelling field by introducing a spherical cut-out instead of the rectangular, circular or V-shaped models. The spherical cut-out could be modelled using four positioning and shape parameters. The spherical model was more realistic, as it implied a gradually deepening spall depth, as opposed to the abrupt depth changes introduced by the models presented by Saxena et al. (2016). In the same year, Luo et al. (2018) improved upon this model, by adding another parameter to allow the spherical cut-out shape to be extended to an ellipsoidal shape. This made the shape of the spall more flexible and realistic. This ellipsoidal model was also validated against FEM, and correlated well with the FEM results.

Luo et al. (2019) did important work in generalising the spall model. The authors generated a formulation where virtually any spall shape could be modelled by using only the effective length information and effective depth information of a spall. The authors showed that this method reproduces older spall models, but was also capable of modelling very complex spall shapes with good agreement with FEM. The only limiting assumption this method made was that the varying spall depth is uniform throughout the tooth's width. For the more complicated spall shapes with a non-analytical description of spall width and depth variation, a discrete sampling and fit method (DSFM) was employed, where regularly measured intervals of the spalls were fitted with a polynomial curve to approximate the shape. This was shown to work well. This model is further capable of modelling pitting defects. For the pit modelling, the authors presented a Gaussian-based approach to distribute a limited number of fixed fault sizes by sampling their locations. Multiple Gaussians could further be used to describe multiple different pitting zones. This method showed promising results.

El Yousfi et al. (2019) presented a novel method to determine spalling faults' effect on the gear mesh stiffness of a gear. Their method was capable of modelling everything that Luo et al. (2019) could, with the addition that a varying fault depth in the tooth width direction could also be accounted for. In this method, the tooth was discretised as a two-dimensional matrix. The width and length of the tooth is given as an $m \times n$ matrix with the entries in position $[m, n]$ corresponding to the tooth height at that coordinate. This implies that integration must now be done across two directions. The benefit of this approach is that analytical equations for the fault types are no longer needed, with the fault shape modelled by modifying the entries of the $m \times n$ matrix. It is found that during the integration phase (still employing PEM stiffness terms), the stiffness reduction is an inherent property of the discretised modelling approach. It must be mentioned that due to the formulation of this approach, a new algorithm, namely the contact detection algorithm needed to be built, complicating the modelling process slightly. This method was validated against Chaari et al. (2009), which inherently also serves as a FEM validation. The computational speed of this method is subject to how finely the tooth discretisation is done, implying it provides the researcher control over the solving speed/simulation accuracy trade-off.

To complete this section, a mesh stiffness sample from a spalled tooth using the method from Chaari et al. (2008) may be seen in Fig. 11.

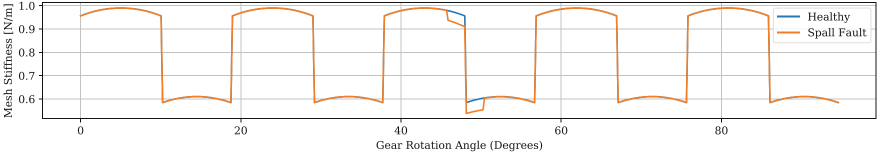


Fig. 11. Effect of a spalled gear tooth on the gear meshing stiffness

3.5 Surface Pit Modelling

Surface pit modelling has received some attention by authors, but less than crack modelling. This is surprising because pitting faults are more common than cracking faults (Liang et al. 2016; Liu et al. 2020). Pitting is generally seen to start beneath the pitch line of the tooth due to this being the region where single tooth pair meshing starts. The single tooth pair meshing region is associated with higher stresses and thus offers a more conducive environment for pits (Lei et al. 2018).

Traditionally, pitting models were initially employed as spalling models, which as we have already seen has been handled as a separate occurrence in the previous section. Essentially, pits were modelled as a single fault/dent on the tooth surface, with the dent shape being different across different approaches. These spalling-based approaches are however unrealistic as pitting faults have several dents in various locations across the tooth.

The first instance that the authors found where pitting defects were extended to multiple pits was in the work of Liang et al. (2016). The authors modelled a single pit as a cylindrical piece of material removed from the gear tooth. Three parameters were used to define a single fault, namely the pit radius, pit depth and pit location (in the length direction) on the tooth. The width information for the pit is not required, as it is assumed that the location of the pit across the width of the tooth is of no consequence to the overall stiffness. The fault was induced by subtracting the effective length (L), cross-sectional area (A_x) and moment of inertia (I_x) of material removed from the beam (Eqs. (1)–(3)) and Hertzian stiffness (Eq. (4)) terms. After setting up a formulation for the singular pit case, Liang et al. (2016) extended the pit modelling approach by inducing multiple spall defects, all with identical radii and depth. The authors superimposed the multitude of pits' effective length, area and moment of inertia reductions to achieve an equivalent length, area and moment of inertia reduction term which could be subtracted from the healthy case. The authors proved their case to coincide very well with a FE model of their geometry.

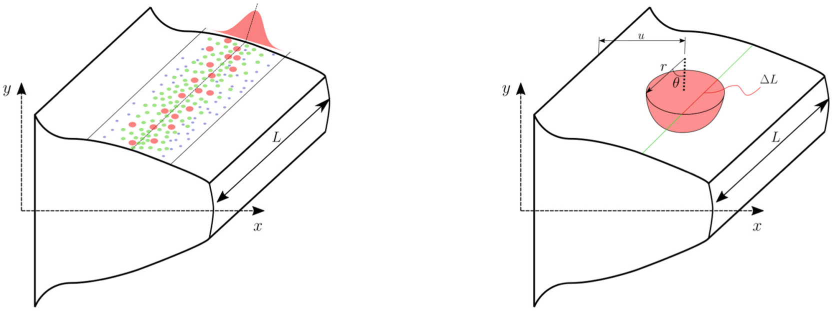
Lei et al. (2018) improved upon the work from Liang et al. (2016) by noting that pits should be distributed probabilistically below the pitch line of the tooth. Liang et al. (2016) 'hardcoded' the pits around the pitch line, whereas Lei et al. (2018) defined the pit location parameters along the tooth length using a normal distribution (See Fig. 12a).

Furthermore, Lei et al. (2018) addressed a further shortcoming: Liang et al. (2016) assumed constant pit sizes for all severity cases. Lei et al. (2018) presented an analytical method by which one can increase a fault severity number and in turn grow the size of existing pits (in depth and radius) but also generate new smaller pits according to the normal distribution mentioned above (See the different pit sizes corresponding to different colours in Fig. 12a). This was seen to be a more realistic fault severity progression modelling approach. When validated against FEM, this approach differed by less than 0.65%.

Chen et al. (2019) also utilised the cylindrical pitting model with multiple faults, but generated a new methodology of describing where the pits should be formed, modelling pits as local clusters described by two-dimensional overlapping Gaussian distributions.

Liu et al. (2020) followed an approach similar to Lei et al. (2018) in terms of methodology but implemented spherical pits as opposed to cylindrical shaped pits. Unfortunately, the parametrisation of this method was not detailed.

Meng et al. (2021) also proposed a spherical pitting model, which is described in great detail. Their model is very similar to that of Lei et al. (2018), but only employing a spherical model as opposed to a cylindrical model. Meng et al. (2021) also showed that their model correlates well with FEM. The authors' model utilised only three parameters to define the fault, but instead of the radius and depth parameter, a spherical radius and angle parameter were utilised. This model can be seen in Fig. 12(b).



(a) Illustration of how pits are distributed across the tooth face using a Gaussian distribution. Figure adapted from Lei et al. (2018).

(b) Illustration of the state of the art pitting model. Only three parameters, namely u, r and θ are required to fully define a fault. Figure adapted from Meng et al. (2021).

Fig. 12. Illustration of the distribution of pitting defects in (a), and the model used to describe the pits themselves in (b).

An implementation of the model by Meng et al. (2021) was done, and a meshing stiffness sample from a severely pitted tooth may be seen in Fig. 13.

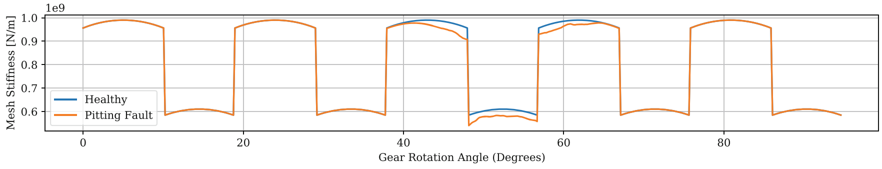


Fig. 13. Effect of a severely pitted gear tooth on the gear meshing stiffness

4 Prospects for Gear Fault Modelling Going Forward

Significant advances have been made in the field of gear fault modelling. However, due to a large number of novel ideas and approaches, the literature is quite disjunct in its formulation of the meshing stiffness terms. Although all fault types inherently aim to modify the tooth height (h_x) or width (L) to model damage to a gear tooth, it is found that different authors utilise different symbols, and utilise different substitutions for formulating faults in Eqs. (1)–(4).

This makes the field of gear tooth fault modelling daunting to neophytes. Furthermore, it may be found that not all authors agree upon which terms to apply the damage parameters to, as is the case with gear crack modelling, where some authors do not consider the axial compressive stiffness as being affected by a root crack. Finally, it is also seen that not all authors consider the fillet foundation effect (Eq. (5)) in their works, and as such have results which can not be compared to one another.

For a hybrid-based PHM approach, it would be extremely useful if a simpler, generalised model existed where one could simply ‘plug and play’ different fault types into the stiffness equations and get the resulting stiffness effects. Currently one has to implement each author’s model with their own unique formulation to the fault modelling to obtain the correct gear meshing stiffness results.

From this perspective, the focus should be geared more towards modelling the stiffness reduction shape of different faults accurately, rather than getting the exact amplitude of the stiffness reduction correct. This will anyhow be adjusted during a calibration process to get correspondence between the models and experimental data as part of the hybrid modelling process. Therefore, the authors believe that gear condition monitoring could greatly benefit from a generalised stiffness model, which can account for several fault types that can simply be induced with a term in Eqs. (1)–(4).

A generalised model would have a twofold benefit:

1. It would lead to a stronger, more flexible physics-based approach that could more easily integrate with data-driven approaches, leading to a powerful hybrid model.
2. It would allow for authors to more easily compare their works with one another if they utilised the same generalised model.

Although a generalised model is advantageous, it is important to consider that many works do not validate their results against any physical experiments. Thus, to increase the value of the generalised model, it must not only be validated against existing works or FEM, but also against experiments.

Work from Luo et al. (2019) shows a good first step towards a generalised model, but is currently limited to spalling faults. The work from El Yousfi et al. (2019) also addresses the generalisation idea, however, the method utilises node-based approaches which could suffer from the same computational burden of FEM.

Even though some work has been done towards generalising fault models for spalling failures, it is the authors’ opinion that a PEM model which can simultaneously model various fault types is missing in the field of gear fault modelling.

The models considered here focus on capturing the dynamic response of the gears correctly. For remaining useful life calculations a further step is required to build a degradation model, which takes force/stress inputs from the hybrid model and utilises these values to propagate faults. The RUL can then be determined as an output from the fault progression prediction. This equips the user with powerful diagnostic knowledge from the hybrid model, being able to identify fault types and intensities, and powerful prognostic knowledge from the degradation model, which can be used to predict the RUL of an asset.

References

Chaari, F., Baccar, W., Abbes, M.S., Haddar, M.: Effect of spalling or tooth breakage on gearmesh stiffness and dynamic response of a one-stage spur gear transmission. Eur. J. Mech. A/Solids **27**(4), 691–705 (2008)

Chaari, F., Fakhfakh, T., Haddar, M.: Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness. Eur. J. Mech. A/Solids **28**(3), 461–468 (2009). https://doi.org/10.1016/j.euromechsol.2008.07.007. https://www.sciencedirect.com/science/article/pii/S0997753808000867. ISSN: 0997-7538.

Chen, T., Wang, Y., Chen, Z.: A novel distribution model of multiple teeth pits for evaluating time-varying mesh stiffness of external spur gears. Mech. Syst. Sig. Process.**129**, 479–501 (2019) . https://doi.org/10.1016/j.ymssp.2019.04.029

Chen, Z., Shao, Y.: ‘Dynamic simulation of spur gear with tooth root crack propagating along tooth width and crack depth. Eng. Fail. Anal.**18**(8), 2149–2164 (2011). http://dx.doi.org/10.1016/j.engfailanal.2011.07.006

El Yousfi, B., Soualhi, A., Medjaher, K., Guillet, F.: New approach for gear mesh stiffness evaluation of spur gears with surface defects.. Eng. Fail. Anal. **116**, 104740 (2019)

Jia, S., Howard, I.: Comparison of localised spalling and crack damage from dynamic modelling of spur gear vibrations. Mech. Sys. Sig. Process. **20**(2), 332–349 (2006)

Lei, Y., Liu, Z., Wang, D., Yang, X., Liu, H., Lin, J.: A probability distribution model of tooth pits for evaluating time-varying mesh stiffness of pitting gears. Mech. Syst. Sig. Process. **106**, 355–366 (2018). https://doi.org/10.1016/j.ymssp.2018.01.005

Lewicki, D.G.: Gear crack propagation path studies-guidelines for ultra-safe design. J. Am. Helicop. Soc. **47**(1), 64–72 (2002)

Liang, X., Zhang, H., Liu, L., Zuo, M.J.: The influence of tooth pitting on the mesh stiffness of a pair of external spur gears. Mech. Mach. Theor. **106**, 1–15. https://doi.org/10.1016/j.mechmachtheory.2016.08.005

Liu, J., Wang, C., Wu, W.: Research on meshing stiffness and vibration response of pitting fault gears with different degrees. Int. J. Rotat. Mach.2020

Liu, Y., Shi, Z., Shen, G., Zhen, D., Wang, F., Gu, F.: Evaluation model of mesh stiffness for spur gear with tooth tip chipping fault. Mech. Mach. Theor. **158**, 104238. https://doi.org/10.1016/j.mechmachtheory.2020.104238

Luo, Y., Baddour, N., Han, G., Jiang, F., Liang, M.: Evaluation of the time-varying mesh stiffness for gears with tooth spalls with curved-bottom features. Eng. Fail. Anal. **92**(June), 430–442 (2018). https://doi.org/10.1016/j.engfailanal.2018.06.010

Luo, Y., Baddour, N., Liang, M.: Evaluation of time-varying mesh stiffness of gears with tooth spalls modeled as spherical shapes., In: Proceedings - 2018 Prognostics and System Health Management Conference. PHM-Chongqing, pp. 94–99 (2018)

Luo, Y., Baddour, N., Liang, M.: A shape-independent approach to modelling gear tooth spalls for time varying mesh stiffness evaluation of a spur gear pair. Mech. Syst. Sig. Process. **120**, 836–852.https://doi.org/10.1016/j.ymssp.2018.11.008

Ma, H., Song, R., Pang, X., Wen, B.: Time-varying mesh stiffness calculation of cracked spur gears. Eng. Fail. Anal. **44**, 179–194 (2014). https://doi.org/10.1016/j.engfailanal.2014.05.018

Ma, H., Zeng, J., Feng, R., Pang, X., Wen, B.: An improved analytical method for mesh stiffness calculation of spur gears with tip relief. Mech. Mach. Theor.**98**, 64–80 (2016). https://doi.org/10.1016/j.mechmachtheory.2015.11.017

Meng, Z., Wang, F., Shi, G.: A novel evolution model of pitting failure and effect on time -varying meshing stiffness of spur gears. Eng. Fail. Anal. **120**, 105068 (2020). https://doi.org/10.1016/j.engfailanal.2020.105068

Mohammed, O.D., Rantatalo, M.: Gear fault models and dynamics-based modelling for gear fault detection - a review. Eng. Fail. Anal.**117**, 104798 (2020). https://doi.org/10.1016/j.engfailanal.2020.104798

Mohammed, O.D., Rantatalo, M., Aidanpää, J.O.: Improving mesh stiffness calculation of cracked gears for the purpose of vibration-based fault analysis. Eng. Fail. Anal. **34**, 235–251 (2013). https://doi.org/10.1016/j.engfailanal.2013.08.008

Sainsot, P., Velex, P., Duverger, O.: Contribution of gear body to tooth deflections - a new bidimensional analytical formula. J. Mech. Des. Trans. ASME **126**(4), 748–752 (2004)

Saxena, A., Parey, A., Chouksey, M.: Time varying mesh stiffness calculation of spur gear pair considering sliding friction and spalling defects. Eng. Fail. Anal.**70**, 200–211 (2016). https://doi.org/10.1016/j.engfailanal.2016.09.003

Sun, R.B., Yang, Z.B., Yang, L.D., Qiao, B.J., Chen, X.F., Gryllias, K.:Planetary gearbox spectral modeling based on the hybrid method of dynamics and LSTM. Mech. Syst. Sig. Process.**138**, 106611 (2020). https://doi.org/10.1016/j.ymssp.2019.106611

Tian, X.: Dynamic simulation for system response of gearbox including localized gear faults, Masters dissertation, University of Alberta (2004)

Wu, S., Zuo, M.J., Parey, A.: Simulation of spur gear dynamics and estimation of fault growth. J. Sound Vib. **317**(3–5), 608–624 (2008)

Yang, D.C., Lin, J.Y.: Hertzian damping, tooth friction and bending elasticity in gear impact dynamics. J. Mech. Desi. Trans. ASME **109**(2), 189–196 (1987)

Author Index

A. Hammami et al. (Eds.): MOSCOSSEE 2021, ACM 20, pp. 259–260, 2022.
https://doi.org/10.1007/978-3-030-85584-0