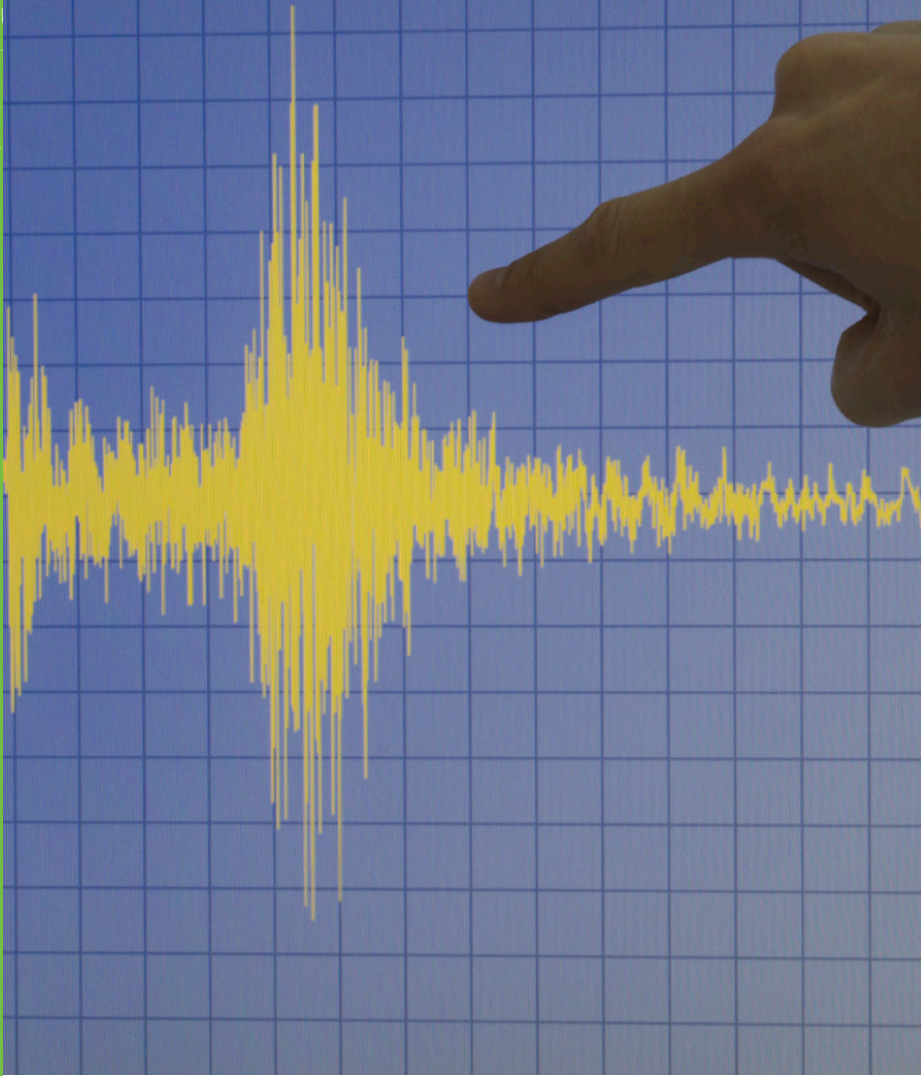# Dynamic modelling and fault analysis of wear evolution in rolling bearings

Idriss El-Thalji

VTT

Dynamic modelling and fault analysis of wear evolution in rolling bearings

Idriss El-Thalji

Thesis for the degree of Doctor of Science in Technology to be presented with due permission for public examination and criticism in Festia Building, Auditorium Pieni Sali 1, at Tampere University of Technology, on the 26th of May 2016, at 12 noon.

Preface

April 20th, 2016

Idriss El-Thalji

Academic dissertation

List of publications

This thesis is based on four peer-refereed original publications which are referred to in the text as I–IV. The publications are reproduced with kind permission from the publishers.

I   El-Thalji, I., & Jantunen, E., "A summary of fault modelling and predictive health monitoring of rolling element bearings," *Mechanical Systems and Signal Processing*, vols. 60–61, pp. 252–272, 2015.

II   El-Thalji, I., & Jantunen, E., "A descriptive model of wear evolution in rolling bearings," *Engineering Failure Analysis*, vol. 45, pp. 204–224, 2014.

III   El-Thalji, I., & Jantunen, E., "Dynamic modelling of wear evolution in rolling bearings," *Tribology International*, vol. 84, pp. 90–99, 2015.

IV   El-Thalji, I., & Jantunen, E., "Fault analysis of the wear fault development in rolling bearings," *Engineering Failure Analysis*, vol. 57, pp. 470–482, 2015.

Author's contributions

The author was responsible for the wear evolution model, simulation model, and the fault analysis. The author's contributions are illustrated as follows with respect to each publication:

I   In publication I, the author has designed the review study and developed the methodology with the second author. The author has collected the reviewed papers, performed the review, and written the manuscript.

II   In publication II, the author has designed the study and developed the methodology with the second author. The author has developed the descriptive model of wear evolution, and written the manuscript.

III In publication III, the author has designed the study and developed the methodology with the second author. The author has performed the simulation model of the wear evolution, and written the manuscript.

IV   In publication IV, the author has designed the study and developed the methodology with the second author. The author has performed the fault analysis, and written the manuscript.

Contents

List of abbreviations

AE	acoustic emission
ANN	artificial neural network
BPFI	ball pass frequency for inner race fault
BPFO	ball pass frequency for outer race fault
BSF	ball spin frequency
CBM	condition-based maintenance
DOF	degree-of-freedom
EHL	elasto-hydrodynamic lubrication
FEM	finite element method
FFT	fast Fourier transform
FTF	fundamental train frequency
ISO	international organization for standardization
PHM	predictive health monitoring
RC	rolling contact
REB	rolling element bearing
RMS	root mean square
SIF	stress intensity factor
SK	spectral kurtosis
SP	signal processing
SPM	shock-pulse measurements
SVM	support vector machine
WT	wavelet transform

1. Introduction

1.1 Background and motivation

In order to make condition-based maintenance (CBM) an effective option for different industrial machines, the health measurements e.g. vibration, debris, should be automatically processed and diagnosed in the correct way and as early as possible. Thus, the maintenance procedures can be planned in a cost-effective manner. The diagnosis and prognosis procedures are essential so as to determine the health status i.e. severity of the machine in question and to predict the remaining lifetime.

The basic approaches to predicting the remaining lifetime are based on data, physical model or a combination of data and physical model. The drawback of the data-driven prognosis e.g. statistical and artificial neural networks, appears in the cases where the system conditions are rapidly and heavily fluctuating. The wear evolution is a complex process of fault development i.e. a degradation process where it influences the bearing dynamic response. As the fault is developing in a non-linear manner, then the dynamic response is also influenced non-linearly by that. Therefore, several studies (Al-Ghamd & Mba 2006), (Sassi et al. 2006), (Nakhaeinejad & Bryant 2011), have introduced the degradation process as a localized fault within the dynamic models, with different defect sizes to represent the development of defect severity. It is clear that this kind of approach assumes a linear relationship between the defect size and the obtained dynamic response. In fact, Al-Ghamd & Mba (2006) showed the change in the defect topography over time by introducing different defect shapes and sizes. This study shows that the defect shape is significantly important together with defect size. The defect development is a continuous process which involves changes in the fault shape over time that make the dynamic impact response differ a great deal with the above assumption (linear relationship between the defect size and the obtained dynamic response). Therefore, the condition monitoring and diagnosis of wear and degradation of the machine can be improved and made more reliable if the degradation process and its physics is understood (Jantunen 2004).

In fact, the rolling element bearing (REB) is one of the most critical components that determines the machine health and its remaining lifetime in modern production machinery. Robust predictive health monitoring (PHM) tools are needed to guarantee the healthy state of REBs during the operation. Therefore, the following reasons underlie the motivation to enhance the monitoring of REBs:

- REBs are one of the most critical components in many industrial applications due to their failure and severity rates. REBs are all wearing components and inevitably produce some debris from their natural operation. For example, the REB failure rate in wind turbines (based on figures in Ribrant & Bertling (2007) is about 3.5% of total failures, which lead to 9.5% of total downtimes.
- The size of modern REBs in large-scale rotating applications becomes a critical lifetime issue. For example, Tavner et al. (2008) observed that large-scale wind turbines (>800 kW) have in general higher failure rates compared to small- (<500 kW) and medium-sized ones.
- The modern REBs are allocated in complex design configurations which lead to different failure rates and downtimes, for example, Tavner et al. (2008) observed that a direct drive has a higher failure rate than its indirect drive partners of the same size.
- The modern REBs are operated at harsher installation sites and in tougher conditions e.g. offshore and cold climate sites. Such sites produce load irregularities and sudden impacts (Holttinen 2005).
- Early detection of REBs becomes a critical requirement for modern renewable energy applications e.g. wind energy due to their limited maintenance window related to the site and seasonal conditions. The maintenance actions outside the limited windows are extremely costly and in some cases are impossible. Thus, detecting the failure just before it occurs does not offer a cost-effective option for maintenance action. The CBM scenario tries to utilise the predictive capabilities in order to monitor an industrial asset and later perform the maintenance work as a planned maintenance (i.e. annual maintenance) or opportunistic maintenance (i.e. low production seasons). That means, in principle, the prediction must be able to cover the time interval between two planned maintenance intervals i.e. one-year-before in the case of wind turbines.

1.2   Research question

In order to understand and overcome the challenges of monitoring the wear evolution in REBs, more detailed dynamic and fault analyses of the wear evolution are needed. This leads to the research question: *How can the wear evolution process of the rolling bearings be effectively modelled?*

The bearing wear modelling is difficult due to the non-linear wear evolution progress in REBs. Therefore, in the beginning, there is a need for a descriptive model

that simply illustrates the wear evolution over the REB's lifetime. Later, a developed numerical model based on that descriptive model is required to provide the dynamic response of the modelled REB over its lifetime. The simulated outcome of the developed dynamic model will be analysed in order to determine the fault features and their changes due to the wear development process throughout the entire lifetime. The developed dynamic model can be also used for remaining useful lifetime prediction.

1.3   Objectives of the research

The main objective of the thesis is to develop a dynamic model which can represent the wear evolution process in REBs over the entire lifetime. Thus, it can be used as a tool to analyse the fault features and to potentially verify the signal analysis methods and the prognosis techniques. For this purpose, a number of sub-goals have to be reached.

- It is necessary to describe the wear evolution in REBs.
- It is necessary to develop a dynamic model of wear evolution in REBs.
- It is necessary to identify the fault features and their changes over the entire lifetime in order to provide effective indicators to track the wear evolution progress.

1.4   Contents of the thesis

The thesis is divided into seven further chapters as follows:

- Chapter 2 reviews the state of the art of wear monitoring in REBs.
- Chapter 3 builds a descriptive model of the wear evolution in REBs based on a wide range of the experimental findings that have been presented in the contact mechanics literature.
- Chapter 4 develops a numerical dynamic model of the wear evolution in REBs.
- Chapter 5 presents the experimental testing and the used data set for validation process of the developed dynamic model.
- Chapter 6 presents the fault analysis based on the simulated vibration data.
- Chapter 7 discusses the results and their contributions with respect to current contributions.
- Chapter 8 provides the thesis conclusions and suggestions for future work.

1.5   Scope of the research

The thesis covers the dynamic modelling, testing work and fault analysis of rolling bearings under a wear evolution process. Therefore, it is a multi-disciplinary thesis where the accumulated knowledge of wear and contact mechanics has been utilised for dynamic modelling. The dynamic model aims ultimately to give the insights and knowledge for further enhancements of current monitoring practices.

Wear scope: The thesis covers quite a wide range of wear issues. However, the focus is the wear in rolling contact, in particularly, the REBs. The work is delimited to mechanical wear mechanisms i.e. fatigue wear, abrasive wear and adhesive wear and not corrosive wear. The actual wear evolution progress in REBs as a physical phenomenon is covered in this work and summarized in a descriptive wear evolution model. Therefore, the descriptive wear evolution model is based on the experimental findings in the literature which are provided by both direct and indirect monitoring techniques. However, it describes the evolution of wear process rather than the wear process itself. Therefore, it utilises the idea of describing the wear evolution based on stages. It also tries to describe the most probable wear evolution scenario and determines the key parameters that might influence such a scenario.

Modelling approach scope: Even though this thesis is based on a comprehensive review which covers most of the literature on wear in rolling contact, the modelling work is simplified by several means. First, the developed model does not try to model the individual wear mechanisms e.g. a model of fatigue, abrasive and adhesive wear. It merely tries to model the overall wear process that covers the interactions and competitions of the individual wear mechanisms with respect to the topographical changes. Second, a lot of emphasis is given to the considerations of how the wear evolution model can be made computationally simple. Therefore, the developed model determines a number of topographical changes based on the wear process stages. The topographical changes influence the dynamics and contact mechanics outcomes. It is a simplified modelling approach instead of developing a continuous model that provides updated topological changes e.g. using finite element methods. A continuous topological change means heavy computational and time consuming tasks. Third, the developed model starts with a single-degree-of-freedom (DOF) under stationary operating conditions to provide simple explanations of the outcomes and reduce the risk of computational errors in such a complex problem. The complexity comes from the integration of several dynamics and contact mechanics models covering the whole wear evolution by means of wear stages. However, the developed model can easily be scaled up to cover higher degrees of freedom, load variation, lubrication film, and debris.

Testing approach scope: Neither the artificially introduced defect nor rolling contact apparatus (e.g. ball on disc) are used. The experimental tests are based on the natural accelerated testing of REB to provide more insights and knowledge of the whole wear evolution process and based on component that are used in indus-

trial applications. However, only the indirect monitoring measurements have been collected.

Fault analysis scope: Neither new measuring nor signal processing techniques are developed. However, the commonly used monitoring and signal processing techniques have been discussed in the light of the wear evolution model, and some future enhancements are proposed. The research work used a simple fault analysis technique to illustrate the effect of the fault topography on the bearing dynamic behaviour over its entire lifetime, which might establish methods of better suitability for wear evolution monitoring in REBs. The purpose is simply to illustrate the changes in the fault features over the lifetime, which indicates the evolution of the wear severity.

1.6   Scientific contribution of the thesis

The scientific contribution of the thesis can be summarised as the development of wear evolution model that can be used for monitoring purposes. The new approach is based on the integration of dynamics and contact mechanics models to involve several wear mechanisms (i.e. fatigue, abrasive, adhesive) and stress concentration mechanisms (i.e. dent, asperities, debris, sub-surface inclusions) over the REB's lifetime. These involvements and their interactions and competitions produce a wear evolution progress which varies significantly with respect to surface topographical and tribological changes. These involvements provide the fluctuations in the dynamic response that represent real data. The wear evolution model is simple and does not require heavy computational and time-consuming tasks. The research work consists of:

- A descriptive wear evolution model has been established which can be used to describe the most probable wear evolution scenario in REBs and illustrate their physical phenomena.
- A simplified dynamic model of wear evolution has been developed which can be used to generate simulated data with features similar to real data. Thus, the model helps to understand what the expected behaviour of a faulty REB is over its lifetime. This model can be used in the physical explanation, training and testing the predictive health monitoring tools.
- Exploring the changes in the fault features due to the wear evolution process over its entire lifetime.
- A simplified dynamic model is also counted as a potential prediction model which can be a part of the prognosis approach. Therefore, the prediction model can be used to prognosis the remaining useful lifetime once the health state is effectively diagnosed.

2. Rolling bearings: Faults, models and analytical techniques

In modern production machinery the rolling bearing is one of the most critical components that determine the machinery's health and its remaining lifetime. Robust PHM tools are needed to guarantee the healthy state of REBs during their operation. The PHM tool indicates the upcoming failures which provides more time for maintenance planning. The PHM tool aims to monitor the deterioration i.e. wear evolution, rather than just detecting the defects. There are a number of literature reviews which are related to the condition monitoring of REBs (Howard 1994), (Tandon & Choudhury 1999), (Jardine et al. 2006), (Jantunen 2006), (Halme & Andersson 2009), (Randall & Antoni 2011). These reviews explain very well the developed signal processing (SP), diagnosis and prognosis analysis methods and their challenges, enhancements and limitations. Many experiments and studies have been made to explore the nature of bearing defects with the help of several monitoring techniques such as vibration, acoustic emission (AE), oil-debris, ultrasound, electrostatic, Shock-Pulse Measurements (SPM), etc. Some simple signal/data processing techniques have been applied to process the signals, such as root mean square (RMS), kurtosis, Fast Fourier Transform (FFT), etc. However, there are several challenges that require more advanced SP methods, e.g. to remove the background noise effect, the smearing effect and the speed fluctuation effect. The most important challenge is to deal with the signal response due to defective REBs. Bearing faults are assumed to generate impulses due to the passing of the rolling element over the defective surface. The difficulty is to detect and track such impulses, especially, in the early stage of wear process where the defect is quite small and can easily be hidden by other vibration phenomena. Therefore, most of the PHM studies have concentrated on the development of more advance SP techniques such as envelope detection, cyclostationary analysis, wavelets, data-driven methods, expert systems, fuzzy logic techniques, etc.

In the field of machinery vibration monitoring and analysis, a variety of relevant standards are developed and published by ISO (International Organization for Standardization). A wide variety of ISO standards describe acceptable vibration limits, such as the ISO/7919 series (5 parts) "Mechanical vibration of non-reciprocating machines – Measurements on rotating shafts and evaluation criteria"

and the ISO/10816 series (6 parts) "Mechanical vibration – Evaluation of machine vibration by measurements on non-rotating parts". The scope covers the methods of measurement, handling and processing of the data required to perform condition monitoring and diagnostics of machines. In industry, the most commonly used techniques are RMS, crest factor, probability density functions, correlation functions, band pass filtering prior to analysis, power and cross power spectral density functions, transfer and coherence functions as well as Cepstrum analysis, narrow band envelope analysis and shock pulse methods. These methods try to extract the expected defect features. The frequency equations of the bearing defects (i.e. for outer-race, inner-race and rolling elements, cage defects) are the main way to provide a theoretical estimate of the frequencies to be expected when various defects occur on the REB. They are based on the assumption that sharp force impacts will be generated whenever a bearing element encounters a localized bearing fault such as spall and pitting. These techniques have continued to be used and have been further developed over time (Howard 1994).

The ultimate purpose of the PHM system is to indicate the upcoming failures which provide sufficient lead time for maintenance planning. Therefore, apart from the experimental studies, there are several analytical and numerical models to: (1) simulate the faulty REBs; (2) verify the ability of SP and diagnosis methods to extract the defect features; and (3) predict the remaining useful lifetime of the faulty REBs. Several studies have explored data-driven and model-based prognosis methods for REBs applications.

Publication I gives the fundamentals of rolling bearing and their modelling techniques, monitoring techniques, SP, diagnostic methods and prognosis analysis. The following subsections give a good summary of what have been published and how the wear in rolling bearing is understood, analysed and diagnosed.

2.1 Bearing faults

A rolling bearing is a mechanical component which carries a load and reduces the sliding friction by placing rolling elements i.e. balls or rollers between two bearing rings i.e. outer and inner raceway. Depending on the internal design, rolling bearings may be classified as radial bearings i.e. carrying radial loads or thrust bearings i.e. carrying axial loads. Practically all rolling bearings consist of four basic parts: inner ring, outer ring, rolling elements, and cage, as illustrated in Figure 1.
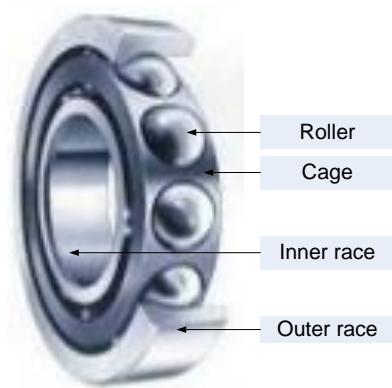
Figure 1. Elements of rolling bearing.

Therefore, the bearing faults may be classified by their locations as outer, inner, rolling element and cage fault. The general reason behind these faults is the rolling contact stresses that might increase due to increased operating loads, additional loads due to faults i.e. imbalance, misalignment, bent shaft, looseness, and/or distributed defects i.e. high degree of surface roughness and waviness, contaminations, inclusions. Therefore, some topographical changes might occur. These topographical changes in the contact area generate stress concentration points and lubrication film disturbances and lead to the wear evolution process.

2.2 Dynamic simulation models

Over the years, several dynamic models have been developed to investigate the dynamic behaviour and features of REBs. The dynamic models of REB were first introduced by Palmgren (1947) and Harris (1966). However, total non-linearity and time varying characteristics were not addressed at that time. After that, Gupta (1975) provided the first completed dynamic model of REB and later Fukata et al. (1985) presented a comprehensive non-linear and time-variant model. The more advanced issues of time-variant characteristics and non-linearity were raised and studied by several authors. For example, Wijnat et al. (1999) reviewed the studies concerning the effect of the Elasto-Hydrodynamic Lubrication (EHL) on the dynamics of REB. Tiwari & Vyas (1995), Tiwari et al. (2000a) and Tiwari et al. (2000b) studied the effect of the ball bearing clearance on the dynamic response of a rigid rotor. Sopanen & Mikkola (2003) reviewed different dynamic models with the discussion of the effect of waviness, EHL, and localised faults and clearance effect. Later, the finite element method (FEM) was used to provide more accurate results. Kiral & Karagülle (2003) presented a defect detection method using FEM vibration analysis for REBs with single and multiple defects. The vibration signal includes impulses produced by the fault, modulation effect due to non-uniform load

distribution, bearing induced vibrations, and machinery induced vibrations and the noise which is encountered in any measurement system. Sopanen & Mikkola (2003) implemented the proposed ball bearing model using a commercial multi-body system software application, MSC.ADAMS. First, the FEM model was utilized to simulate the variation of the mesh stiffness for two types of faults under varying static load conditions. Then the model was integrated into the lumped parameter dynamic model. The study obtained the dynamic transmission error and acceleration responses under different loads and speeds. Sawalhi & Randall (2008) developed a 34-DOF model of a gearbox in order to simulate spall and cracks in the REB. Massi et al. (2010) studied the wear that results from false brinelling at the contact surfaces between the balls and races of the bearings. Several models have been developed to study the effects of several distributed and localized defect on REB dynamics: clearance effect (Tiwari et al. 2000b), (Sopanen & Mikkola 2003), (Purohit & Purohit 2006), (Cao & Xiao 2008), (Nakhaeinejad 2010), waviness effect (Jang & Jeong 2002), (Sopanen & Mikkola 2003), (Cao & Xiao 2008), disturbances effect of EHL (Sopanen & Mikkola 2003), (Sawalhi & Randall 2008), and the effect of localized faults (McFadden & Smith 1984), (McFadden & Smith 1985),(Tandon & Choudhury 1997), etc.

The largest proportion of the studies has focused on the localized faults using different modelling techniques. McFadden & Smith (1984), McFadden & Smith (1985), Tandon & Choudhury (1997) and Sawalhi & Randall (2008) simulated the defect as a signal function of an impulsive train into the modelled system. For example, Tandon & Choudhury (1997) have introduced the defect as pulse function with three different pulse shapes: rectangular, triangular and half-sine pulse. Wang & Kootsookos (1998) introduced defects as a function of a basic impulse series. Ghafari et al. (2007) have virtually introduced a defect into the equation of motion as a triangular impulse train at the related characteristic frequencies of a defect. Rafsanjani et al. (2009) modelled the localized defects as a series of impulses having a repetition rate equal to the characteristics frequencies. The amplitude of the generated impulses is related to the loading and angular velocity at the point of contact. Malhi (2002), Kiral & Karagülle (2003), Sopanen & Mikkola (2003), Massi et al. (2010) and Liu et al. (2012) introduced the defect as force function into their FEM models i.e. as a constant impact factor. More preciously, Liu et al. (2012) introduced the localized defect as a piecewise function.

Ashtekar et al. (2008), Sassi et al. (2007), Cao & Xiao (2008), Rafsanjani et al. (2009), Patil et al. (2010), and Tadina & Boltezar (2011) modelled the defect based on its geometrical features i.e. as a surface bump or a dent that has length, width and depth. Tadina & Boltezar (2011) modelled the defect as an impressed ellipsoid on the races and as flattened sphere for the rolling elements. Nakhaeinejad (2010) utilised the bond graphs to study the effects of defects on bearings vibrations. The model incorporated gyroscopic and centrifugal effects, contact deflections and forces, contact slip and separations, and localized faults. Dents and pits on inner race, outer race and balls were modelled through surface profile changes i.e. type, size and shape of the localized faults. The main difficulty with

the use of complex dynamic models lies in experimentally verifying the predicted results (Howard 1994).

El-Thalji & Jantunen (2014) reviewed the most relevant studies and experimental findings in order to describe the wear process over the lifetime for the rolling bearings. In summary, the wear evolution process is quite complex due to the involvement of several wear mechanisms (i.e. fatigue, abrasive, adhesive, corrosive) and several stress concentration mechanisms (i.e. dent, asperities, debris, sub-surface inclusions). These mechanisms and their interactions and competitions produce a wear evolution progress which varies significantly with respect to surface topographical and tribological changes. As the fault topography is changing over the lifetime that simply means the fault features are changing over time. In this sense, the fault topography that is assumed in the simulation models should change. Moreover, there is a need to clearly determine the fault features of specific wear evolution stages and to understand how different signal analysis methods cope with such features, in order to effectively track the detected fault features.

In this sense, the dynamic models deal with the wear phenomenon as a localized defect with fixed features over the lifetime. The reason is that the purpose of these models is to detect the defect within the generated vibration signals and not the incremental deterioration process i.e. wear evolution. These dynamic models start from the point where the defect is localized as a simulated defect in the models or artificially introduced into the experiments. That ignores the prior stages of the localization process. The localized defects and their associated impact remain constant over the whole lifetime. That ignores the topographical and tribological changes of the defected surface.

In order to model the wear evolution, an incremental numerical procedure should be developed which is able to integrate the contact information continuously into the dynamic model. This means that the applied force due to the wear progress and its associated topological and tribological conditions should be iteratively updated in the dynamic model.

2.3   Monitoring methods

Several experiments have been conducted in order to study specific monitoring techniques such as vibration, acoustic emission, oil-debris, ultrasound, electrostatic, shock-pulse measurements, and their use in faulty REBs detection. Many studies have used simple signal/data processing techniques such as RMS, kurtosis, FFT, etc. However, the largest proportion of studies has focused on the development of the advanced SP e.g. envelope, wavelets, and decision making techniques e.g. expert systems, fuzzy logic techniques. The majority of the advanced SP techniques are related to vibration measurements, and these studies will be discussed in the next section. There are basically two testing approaches. The first is the naturally accelerated testing with the help of applying overload, adjusting the lubricant film thickness or adding contaminated oil. The second approach is the

artificially introduced defects by cutting, false-brinelling, electric charge (i.e. erosion dent) and scratching.

2.3.1 Testing techniques

Several experiments with the help of vibration measurements have been conducted on bearings and on other rolling contact mechanisms. Quite large numbers of studies have explored the effectiveness of the AE technique for rolling contact mechanisms. Other AE experiments have been performed on bearings. Comparative studies that combine vibration and AE measurements have been conducted in order to explore defect features with the help of rolling contact test, and some other with the help of REBs. Some studies have investigated the capability of electrostatic charge measurements (when a charged particle passes the sensor) in detecting a bearing defect. The studies have investigated the capability of ultrasound measurements in detecting a bearing defect, in particular, for the low speed bearings.

Many detection issues were studied, such as the effect of surface roughness (Tandon & Choudhury 1999), the influence of running parameters on the AE of grease lubricated REB (Miettinen 2000), the effect of λ factor (i.e. film thickness/surface roughness) (Serrato et al. 2007), the running-in process (Massouros 1983),(Peng & Kessissoglou 2003), the effects of low speed, the large scale bearings and operating conditions (lubrication type, temperature) (Momono & Noda 1999) and the effects of geometrical imperfections (i.e. variation of roller diameters, inner ring waviness), abrasive and fatigue wear (Sunnersjö 1985). The effect of contaminant concentration on vibration was also studied (Maru et al. 2007), (Boness & McBride 1991), (Momono & Noda 1999).

2.3.2 Wear evolution

Jantunen (2006) and Yoshioka & Shimizu (2009) observed two main stages of wear progress: steady state and instability. The steady-state stage is roughly stable. However, a clear offset in the RMS values of monitoring signals is observed at the instability stage, together with instability and a rapid increase of these values before the final failure. Schwach & Guo (2006) and Harvey et al. (2007) observed three stages of wear progress. Moreover, the instability stage is observed to follow a steeply-offset propagation. Harvey et al. (2007) observed that electrostatic charge measurements indicate the wear initiation as a region of high signal amplitude (with respect to normal signal state), where it disappears (i.e. goes back to normal single state) until the failure occurs. Therefore, electrostatic measurement indicates instantaneous occurrences of wear mechanisms in the region of high signal amplitude rather than progressive stages. Manoj et al. (2008) observed that the 3rd harmonic of the roller contact frequency of vibration has very good correlation with the wear and, when the pitting takes place, the amplitude of

the 3rd harmonic of contact frequency increases to nearly four to five times the amplitude of other harmonics. In the same manner, the frequency analysis of sound signals shows that the 3rd and 1st harmonics of roller contact frequency have good correlation to the wear trend. Zhi-qiang et al. (2012) observed two stages of wear progress using vibration measurements. However, four stages of wear progress were observed using AE: running-in, steady-state, a stage of minor-instability due to distributed defects, and finally a stage of major-instability due to pitting and spall. Sawalhi & Randall (2011) investigated the trend of kurtosis values of faulty signals, with relation to the development of the fault size. The kurtosis increases almost linearly in the early stage of testing time as the defect size increases. However, it stabilizes later as the defect size slowly extends. It could be due either to the existence of a smoothing process or the surface becoming very rough where the effectiveness has become weak.

The artificially introduced defect approach is widely used due to its simplicity. The researchers can virtually introduce a well-known shape and size of a defect. what is more, they can artificially introduce the same defect features in the validation experiment. Furthermore, this approach delimits the testing complexity, as it focuses on a single artificial defect, compared to the natural defect propagation approach. However, the natural wear process highlights that the bearing defect is changing over the time with respect to the topological and tribological changes due to different wear and stress concentration mechanisms. The drawback of the artificially introduced defect approach is that the damage criterion is somehow artificially determined, which might be totally different from the defect in real operation. Therefore, the artificially defective bearing tests are helpful in the development of new analysis and diagnosis techniques; however, they are not the best way to investigate the evolution of real wear progress.

The impulsive response is clearly seen when the impact of the rolling element that passes over a defect is strong. The impact severity is related to the size of the impact area and the sharpness of the defect edges. The impact area is the area on the trailing edge of the dent, asperity or the defect that comes into contact with the rolling element. Based on the literature, this area is quite small at the defect initiation stage and depends on the length, depth and width of the defect. Al-Ghamd & Mba (2006) observed that increasing the defect width increased the ratio of burst amplitude-to-operational noise (i.e. the burst signal was increasingly more evident above the operational noise levels). It was also observed that increasing the defect length increased the burst duration. The first observation indicates that the width of defect increases the impact area on the trailing edge and, therefore, stronger amplitude and high signal-to-noise ratio was observed. In fact, most of the studies show the ability of the envelope analysis to detect such a size and feature of the impact area. However, the problem with wear evolution is when the impact areas are rapidly and continuously changing due to the loading and wear progress.

In the early stage of the wear process, the defect is quite small and can easily be concealed by other vibration phenomena. Most of the SP methods are validated

based on experimental tests in which the defects are introduced artificially into the bearings. Such a testing approach guarantees the availability of the impulsive response due to the introduced defect and somehow its severity is quite enough to be detected. The natural accelerated testing experiments (Jantunen (2006) and Yoshioka & Shimizu (2009)) show that it is quite hard to detect the impulsive response at an early stage and much harder to track its evolution. The basic reason behind the difficulty is that the relation between the defect growth i.e. to become larger is not linear with its dynamic impact. It is a nonlinear relation due to the high stochastic nature of defect growth i.e. wear evolution. Also, it depends greatly on the wear and stress contraction mechanisms that are involved. The experiments show that the wear process is slow in nature and can hardly produce detectable impacts at an early stage. Moreover, the experiments in the literature show that the impulsive response of bearing defect is changing over the time with respect to the topographical and tribological changes.

The natural accelerated tests show fluctuations in the impulsive response of bearing defects and at some time intervals it is hard to detect them. For example, the over-rolling and abrasive wear effects make the defected surface smoother and the impact events softer. These empirical facts are quite important to explain the capabilities and limitations of the applied monitoring methods, in order to enhance their suitability for wear evolution monitoring in REBs.

2.4 Signal analysis methods

2.4.1 Statistical measures

At the beginning, the SP methods were very simple and mainly based on the statistical parameters i.e. RMS, mean, kurtosis, crest factor, etc. The trending based on RMS value is one of the most used methods, which shows the correlation between vibration acceleration and the REB wear over the whole lifetime (Jantunen 2006), (Schwach & Guo 2006), (Harvey et al. 2007), (Yoshioka & Shimizu 2009), (Zhi-qiang et al. 2012). Kurtosis and crest factors increase as the spikiness of the vibration increases. In this sense, the kurtosis and the crest factor are very sensitive to the shape of the signal. However, the third central moment (Skewness) was found to be a poor measure of fault features in rolling bearings (Tyagi 2008), in general Skewness can be an effective measure for signals that contain unsymmetrical signals i.e. non-linearity. The kurtosis is sensitive to the rotational speed and the frequency bandwidth. It is efficient in narrow bands at high frequencies especially for incipient defects (Pachaud et al. 1997), (Djebala et al. 2007). More advanced approaches of time-domain analysis are the parameter identification methods, where a time series modelling is applied to fit the waveform data to a parametric time series model and extract the features (Jardine et al. 2006). Baillie & Mathew (1996) introduced the concept of an observer bank of autoregressive time series models for fault diagnosis of slow speed machinery under transient

conditions, where a short set of vibration data is needed. Due to instantaneous variations in friction, damping, or loading conditions, machine systems are often characterised by non-linear behaviours. Therefore, techniques for non-linear parameter estimation provide a good alternative for extracting defect-related features hidden in the measured signals (Yan & Gao 2007). A number of non-linear parameter identification techniques have been investigated, such as Correlation Dimension (Logan & Mathew 1996), (Logan & Mathew 1996), (Yan & Gao 2007) and Complexity (the degree of regularity of a time series) Measure (Yan & Gao 2004). As the bearing system deteriorates due to the initiation and/or progression of defects, the vibration signal will increase, resulting in a decrease in its regularity and an increase in its corresponding entropy value (Yan & Gao 2007). In the early stage of machinery faults, the signal-to-noise ratio is very low due to relatively weak characteristic signals. Therefore, a chaotic oscillator was proposed (Logan & Mathew 1996), (Logan & Mathew 1996), (Wei et al. 2008) to extract the fault bearing features due to its sensitiveness to weak periodic signals. The complexity measure analysis shows that the inception and the growth of faults in the machine could be correlated with the changes in the complexity value (Yan & Gao 2004). The biggest drawback of statistical methods is the need for a suitable quantity of data for training and testing the system during the development phase. The large quantity of data points that need to be calculated leads to lengthy computational time unsuitability for on-line, real-time applications (Yan & Gao 2007).

2.4.2 Frequency domain methods

The frequency domain methods have been introduced to provide another way to detect the fault-induced signal. FFT is one of the most common methods to transform the signal from time domain into the frequency components and produce a spectrum. However, it is often not clear enough to observe the fault peak, because of slip and masking by other stronger vibrations, apart from the effects of the defect frequency harmonics and sidebands (Ocak et al. 2007). Moreover, the FFT method is actually based on the assumption of periodic signals, which is not suitable for non-stationary signals. The output signals of running REB contain non-stationary components due to the changes in the operating conditions and faults from the machine and bearing itself (Peng & Chu 2004). Time–frequency analysis is the most popular method to deal with non-stationary signals. The Wigner–Ville distribution, the short time Fourier transform and Wavelet transform (WT) represent a sort of compromise between the time- and frequency-based views of a signal and contain both time and frequency information. Mori et al. (1996) applied the discrete wavelet transform so as to predict the occurrence of spall in REBs. Shibata et al. (2000) used the WT to analyse the sound signals generated by bearings. Peng et al. (2005) highlighted that the Hilbert–Huang transform has good computational efficiency and does not involve the frequency resolution and the time resolution.

2.4.3 Challenges of feature extraction process

There are several challenges to remove the speed fluctuations, the smearing effect of signal transfer path and the background noise. The effect of speed fluctuation, e.g. chirp signals, is important and needs to be removed. The chirp signal or sweep signal, i.e. a signal in which the frequency increases ('up-chirp') or decreases ('down-chirp') with time, might be generated due to speed fluctuations, running-in, cut-out operations. It should be noticed that there have been several methods proposed to deal with the chirp signals such as chirp z-transform, chirp Fourier transform, adaptive chirplet transforms, and high-order estimations. Moreover, The order tracking methods are used to avoid the smearing of discrete frequency components due to speed fluctuations (Randall & Antoni 2011). To solve the smearing effect due to the signal transmission path, the Minimum entropy deconvolution method has been developed (Endo & Randall 2007), (Sawalhi et al. 2007).

For the background noise problem, different de-noising filters have been developed such as discrete/random separation (Antoni & Randall 2004b), adaptive noise cancellation (Chaturvedi & Thomas 1982), (Tan & Dawson 1987), self-adaptive noise cancellation (Ho 2000), (Antoni & Randall 2001), (Antoni & Randall 2004a) or linear prediction. However, for a situation, where the noise type and frequency range are unknown, the traditional filter designs could become computationally intense processes (Qiu et al. 2003). For example, the WT methods perform very well on Gaussian noise and can almost achieve optimal noise reduction while preserving the signal. However, it is still a challenge how to select an optimum wavelet for a particular kind of signal i.e. to select the optimum wavelet basis, to select the corresponding shape parameter and scale level for a particular application. Moreover, how to perform thresholding is another challenge. There are two major wavelet-based methods, which are used for mechanical fault diagnosis: The first method focuses on selecting a suitable wavelet filter, e.g. the Morlet wavelet, impulse response wavelet, and the second method focuses on selecting a suitable decomposition process e.g. adaptive network based fuzzy inference. Based on the WT, many kinds of fault features can be obtained, all of which can be classified as the wavelet coefficients-based, wavelet energy-based, singularity-based and wavelet function-based (Peng & Chu 2004). The continuous WT of the Morlet wavelet function has been used (Lin & Qu 2000). Junsheng et al. (2007) proposed the impulse response wavelet base function to describe the vibration signal characteristics of the REB with fault, instead of the Morlet wavelet function. Liu et al. (2008) proposed a weighted Shannon function in order to synthesize the wavelet coefficient functions to enhance the feature characteristics, i.e. optimal wavelet shape factor and minimizes the interference information. Djebala et al. (2007) presented a denoising method of the measured signals-based on the optimization of wavelet multi-resolution analysis based on the kurtosis value. Liu et al. (1997) proposed a wavelet packet-based method for the fault diagnostics of REB, where the wavelet packet coefficients were used as features. Altmann & Mathew

(2001) presented a method based on an adaptive network-based fuzzy inference system, so as to select the wavelet packets of interest as fault features automatically, to enhance the detection and diagnostics of low speed REB faults. Su et al. (2010) presented a new hybrid method based on optimal Morlet wavelet filter and autocorrelation enhancements i.e. to eliminate the frequency associated with interferential vibrations, reduce the residual in-band noise and highlight the periodic impulsive feature.

2.4.4 Bearing fault signals

Some studies (Sun & Tang 2002), (Peng et al. 2007), (Altmann & Mathew 2001), (Hao & Chu 2009) highlight the fact that the most relevant information of a signal is often carried by the singularity points, such as the peaks, the discontinuities, etc. Therefore, singularity detection methods are proposed (Sun & Tang 2002), (Peng et al. 2007) based on calculating the Lipschitz exponents of the vibration signals. A large Lipschitz exponent indicates a regular point in the signal, while a small Lipschitz exponent indicates a singular point. The WT is very successful in singularity detection, however before the singularity is detected, the signal pre-processing must be carried out, so as not to overlook some singularities (Peng & Chu 2004). Hao & Chu (2009) observed that the impulse components cannot be seen clearly due to the existence of harmonic waves. The WT filtering removes the noise, but; the harmonic waves are not suppressed, since the impulse frequency was very close to the harmonic wave frequencies (Hao & Chu 2009). Therefore, the scalogram (i.e. a visual method of displaying a wavelet transform) is proposed to reveal more information about the signal.

Several methods try to extract the periodic information of the impulsive response of faulty REB such as the time synchronous average (McFadden 1987), (Dalpiaz et al. 2000), (Miller 1999). The bearing fault signals have a deterministic part and a quasi-cyclostationary part, where the envelope and the squared envelope of the bearing vibration signal is the way to overcoming this problem (Randall et al. 2001). The envelope analysis utilizes the idea of detecting the fault impulses that are amplified by structural resonance. However, it is a challenge to determine the spectrum band which contains the highest signal-to-noise ratio. Randall (2011) has highlighted that determining the suitable demodulation band is recently solved by means of e.g. spectral kurtosis (SK) (Sawalhi & Randall 2005), (Antoni 2006), (Antoni & Randall 2006), (Sawalhi 2007). Tse et al. (2001) compared the effectiveness of the wavelet and the envelope detection methods for REBs fault diagnosis. The results showed that both the wavelet and envelope detection methods are effective in finding the bearing fault, but the wavelet method is less time-consuming. The shortcoming of the envelope detection approach is the increasing difficulty in analysing the vibration spectrum when the signal-to-noise ratio is low (Chiementin et al. 2007), in which case the fault-imposed frequencies can be masked by noise and other frequency components. To overcome this problem, some morphological operators are proposed (Hao & Chu 2009) with the aim of

extracting the envelope of impulsive type periodic vibration signals by modifying (i.e. using morphological operators such as dilation, erosion, opening, closing) the geometrical features of the signals in the time domain. That constructs a kind of envelope which accentuates information corresponding to the impact series produced by a fault.

The impacts on the fault do not occur exactly in a periodic manner, because of random slips, possible speed fluctuations, and variations of the axial to radial load ratio. Therefore, the bearing fault signals are more likely to be described as cyclostationary (McCormick & Nandi 1998), (Kilundu et al. 2011), as pseudo-cyclostationary (Randall 2011), as quasi-cyclostationary (Randall et al. 2001), (Antoni et al. 2004) and as poly-cyclostationary (Antoni et al. 2004). The cyclostationary signal is defined as a random signal in which the statistical parameters vary in time with single or multiple periodicities (Da Costa 1996) and as a signal which, although not necessarily periodic, is produced by a hidden periodic mechanism (Randall & Antoni 2011). The quasi-cyclostationary signal is generated when the existence of a common cycle is not allowed due to the fact that some rotating components are not locked together such as in REBs. Antoni et al. (2004) highlighted that poly-cyclostationary signals are generated since many mechanical components in the machinery introduce various different periodicities, so they are a combination of cyclostationary processes with different basic cycles. Antoni et al. (2004) explained that all kinematical variables in the machinery that are periodic with respect to some rotation angles are intrinsically angle-cyclostationary rather than time-cyclostationary. The synchronous averaging, comb-filters, blind filters and adaptive comb-filters are of the type of first-order cyclostationary methods. Synchronous auto-covariance function, Instantaneous variance, and spectral correlation density are second-order cyclostationary methods. The spectral correction is proposed by Gardner (1986) where the second-order periodicity can be characterized i.e. the degree of coherence of a time series. Several studies have discussed the cyclostationary and spectral correlation technique for fault detection in REBs, such as (Randall et al. 2001), (Antoni et al. 2004), (Antoni 2007), (Antoni 2007), (Antoni 2009). The envelope analysis gives the same result as the integration of the cyclic spectral density function over all frequencies, thus establishing the squared envelope analysis as a valuable tool for the analysis of (quasi-) cyclostationary signals more generally (Randall et al. 2001). Moreover, since the autocorrelation of a periodic signal is both periodic vs. time and time-lag, it produces a spectral correlation function discrete in both the "f" and "α" directions like a ''bed of nails''. The higher-order spectra describe the degree of phase correlation among different frequencies present in the signal (Liu et al. 1997). Therefore, Li & Ma (1997) used bi-coherence spectra so as to derive features that relate to the condition of a bearing. Collis et al. (1998) explained that the bi-spectrum can be viewed as a decomposition of the third moment 'skewness' of a signal over frequency, and that it proves useful for analysing systems with asymmetric non-linearities. However, this statistical approach requires a rather large set of data in order to obtain a good estimation (Mori et al. 1996). Pineyro et al. (2000) com-

pared the second-order power spectral density, the bi-spectral technique and the WT, and found this last to be useful in the short transient detection, since it could eliminate the background noise.

The signal analysis methods, which are applied to signals measured from bearings with an artificially introduced defect, are quite effective. However, a careful comparison between the defect features of a natural wear process and the artificially introduced defect should be taken in consideration. Simply put, the artificially introduced defects are the dominated damage mechanism and in general they are large, sharp and strictly localised. A natural fault is smaller, less sharp and has evolved with the help of different wear and stress concentration mechanisms. In fact, the impulse due to wear defect is changing over the whole lifetime (El-Thalji & Jantunen 2015a).

It is also clear from the literature, the definition of bearing fault signal type i.e. stationary, cyclostationary, non-stationary, etc. is the main reason and motivation for the variety of SP methods. Some methods are just for specific types of signals and it is hard to illustrate their outcomes or there is no point in using them if the fault-induced signals are not of that type. Thus, it is more realistic to illustrate how the bearing fault-induced signal evolves over the whole REB's lifetime. The fault-induced signal is usually of the impulsive signal type due to the impact event when the rolling element passing over e.g. an asperity, dent or defect. The defect topology affects the impact severity when a rolling element passes over it. Therefore, the impulsive nature of wear is changing as the wear defect evolves. Moreover, at some wear progression intervals, there is no clear impulsive impact, where some monitoring and diagnosis techniques are not effective.

2.5 Fault diagnosis methods

The fault diagnosis task consists of the determination of fault type with as many details as the fault size, location, and severity. Since a machine has many components and is highly complex, diagnosis of a machine fault usually requires technical skill and experience. It also requires extensive understanding of the machine's structure and operation, general concepts of diagnosis and an expert engineer to have domain-specific knowledge of maintenance and to know the 'ins-and-outs' of the system. In reality, the expert is either too busy with several tasks or a specific component expert is not available at all (Yang et al. 2005). In order to automatize the diagnosis procedures and provide a decision about the REB's health state, a number of automatic feature diagnosis methods have been developed. Several diagnosis methods are proposed to diagnose the faulty REBs such as artificial neural network (ANN), expert systems, fuzzy logic, support vector machine (SVM), state observes, and model-based methods.

The ANN methods have been applied to diagnose the REB's fault and such as (Roemer et al. 1996), (Paya et al. 1997). Larson et al. (1997) performed the phase demodulation by means of neural networks. Li et al. (2000) utilised the FFT as a

pre-processor for Feed-Forward Neural network to perform fault detection. Samanta & Al-Balushi (2003) developed a back-propagation neural network model to reduce the number of inputs which leads to faster training requiring far fewer iterations. Moreover, Baillie & Mathew (1996) illustrated the better noise rejection capabilities of the back-propagation networks compared to traditional linear methods. However, noise still remains a problem, and the best way to combat this is to use longer data lengths so that the noise can effectively be cancelled by the signal-to-noise ratio averaging process. Alternatively, it also highlights the importance of signal pre-processing techniques, such as amplitude demodulation in the case of REBs (Baillie & Mathew 1996). The cascade correlation algorithm offers the advantage that the number of hidden units does not have to be determined prior to training. Spoerre (1997) applied the cascade correlation algorithm so as to predict the imbalance fault in rotor-bearing configuration. Radial basis functions are used (Baillie & Mathew 1996) for REB, and compared to back-propagation networks these show superior outcome due to their rapid training time. Since the unsupervised learning does not require external inputs, Wang & Too (2002) applied the unsupervised neural networks, self-organising map and learning vector quantisation to rotating machine fault detection. Tallam et al. (2002) proposed some self-commissioning and on-line training algorithms for feed-forward a neural network with particular application to electric machine fault diagnostics. However, the building and training of artificial neural networks typically requires a trial and error approach and some experience (Baillie & Mathew 1996). The scope of the reviewed ANN methods is to classify the following fault features i.e. health state, defect type, defect location, defect severity, etc. Paya et al. (1997) used the ANN to differentiate between each fault and establish the exact position of the fault occurring in the drive-line. Samanta & Al-Balushi (2003) developed a back-propagation neural network model which obtains the fault features directly using very simple pre-processing i.e. root mean square, variance, skewness, kurtosis and normalised sixth central moment of the time-domain vibration signals, to classify the status of the machine in the form of normal or faulty bearings.

Different expert systems have been proposed for diagnosing abnormal measurements such as rule-based reasoning (Yang et al. 2005), case-based reasoning (Yang et al. 2004), and model-based reasoning. It would be wise to present the cause-symptom relationship in a tabular form for quick comprehension and a concise representation. Yang et al. (2005) developed a decision table i.e. IF (symptom) and THEN (cause) to link the causes of fault and symptoms from an empirical knowledge gained either by direct experience with the system or through another expert in the field. The ANNs are required to gradually learn knowledge in the operating process, and to have the adaptive function expanding the knowledge continuously without the loss of the previous knowledge during learning new knowledge. Therefore, Yang et al. (2004) proposed the integrated approach of Adaptive Resonance Theory and Kohonen Neural Network. However, the previous cases may influence a case-based reasoning system in different directions without giving it many hints on which cases to consider as more important. This problem,

associated with other difficulties in case-based indexing and retrieval, suggests that combining the case-based reasoning with complementary forms of reasoning, such as rule-based, model-based or neural network, may be fruitful (Yang et al. 2004).

In order to have flexible classification practices, the fuzzy logic approach is introduced. Fuzzy logic has gained wide acceptance as a useful tool for blending objectivity with flexibility. Fuzzy logic is also proving itself to be a powerful tool when used for knowledge modelling, particularly when used in condition monitoring and diagnostics applications. Liu et al. (1996) developed a fuzzy logic based expert system for rolling bearing faults. Mechefske (1998) applied fuzzy logic method to classify frequency spectra representing various REB faults. Unlike other neural networks, fuzzy neural networks adopt bidirectional association. They make use of the information from both fault symptoms and fault patterns and improve recognition rate. Therefore, Zhang et al. (2003) applied a neural network to diagnosing the fault on a rotary machine. Jantunen (2004) proposed the use of a simplified fuzzy logic for automated prognosis. It saves the history of measured parameters and gives a prognosis of further development.

In practice, the huge number of possible loading conditions, i.e. measuring situations, makes the ANN task more complicated. Therefore, it is always a question of whether the training results can be moved from one machine to others. An SVM is another classification technique based on statistical learning theory. Three methods were used to find the separating hyper-plane, namely Quadratic Programming, Least-Squares and Sequential Minimal Optimization method. Yang et al. (2007) used intrinsic mode function envelope spectrum as input to SVMs for the classification of bearing faults. Hu et al. (2007) used improved wavelet packets and SVMs for the bearing fault detection. Abbasion et al. (2007) used the SVM as a classifier to compute optimum wavelet signal decomposition level, in order to find an effective method for multi-fault diagnosis. Gryllias & Antoniadis (2012) proposed the hybrid two stage one-against-all SVM approach for the automated diagnosis of defective REBs. In SVM approach, it is quite necessary to optimize the parameters which are the key factors impacting the classification performance. Li et al. (2013) proposed an improved ant colony optimization (IACO) algorithm to determine these parameters, and then the IACO-SVM algorithm is applied to the REB fault detection. Liu et al. (2013) proposed a multi-fault classification model based on Wavelet SVM. Particle swarm optimization is applied in order to seek the optimal parameters of Wavelet SVM and pre-processed using empirical model decomposition. Guo et al. (2009) investigated the SVM method based on envelope analysis to diagnose REB with a ball fault, inner race fault or outer race fault. The SVM is originally designed for two-class classification problem; however bearing fault diagnosis is a multi-class case. Tyagi (2008) observed that more accurate classification of bearing conditions is achieved by using SVM classifiers as compared to ANN. In fact, the ANN uses traditional empirical risk minimization principles to minimize the error in training data, while SVM utilizes structural risk minimization principles to minimize the upper boundary of expected risk (Guo et al.

2009). Pan et al. (2009) proposed a combined method based on improved wavelet packet decomposition and support vector data description to achieve better speed in training. However, Jack & Nandi (2002) observed that the ANN tends to be faster to train and slightly more robust than the SVM.

The other non-linear classifiers such as the Gaussian Mixture Model and Hidden Markov Model have been used for classification problems in specific applications. Therefore, Nelwamondo et al. (2006) introduced the Gaussian mixture model and hidden Markov model to diagnose faults in rolling bearing features, based on extracted features using Multi-Scale Fractal Dimension, Mel frequency Cepstral Coefficients and kurtosis. However, the major drawback of the hidden Markov model classifier is that it is computationally expensive, taking more than 20 times longer than the time required to train the GMM. Ocak et al. (2007) developed a new scheme based on wavelet packet decomposition and the hidden Markov model for tracking the severity of bearing faults. Zhang & Kang (2010) proposed and hidden Markov model to represent the states of bearing through partition substate for the five states.

The model-based methods utilise the physics models to diagnose the health of the monitored REB. Vania & Pennacchi (2004) proposed a diagnostic technique where the fault is obtained by evaluating the system of excitations that minimizes the error i.e. residual, between the machine experimental response and the numerical response evaluated with the model. Söffker et al. (2013) introduced Proportional-Integral Observer method to detect a crack by detecting small stiffness changes. The very detailed and physically-oriented understanding that is provided by the model-based approach enhances the interpretation problem of signal-based approaches. However, the necessity for fault models and the hypotheses about the location of the fault is a limitation. The majority of real industrial processes are nonlinear and are not effective to be modelled by using linear models for all operating conditions.

In summary, one important issue is the early detection of the fault i.e. earliness. There are wide and qualitative definitions of earliness of detection within the literature. In fact, many studies which utilised the artificially introduced defects can be recognised to be a severe state in the real application and, therefore, can also be detected with simple methods. It is clear that the signal analysis methods aim to detect the defect as early as possible. Most of the feature diagnosis methods classify the REB's state into either a healthy or faulty state. Some other methods aim to classify the defect types i.e. imbalance, defect, and defect locations i.e. outer race, inner race, rolling element. Few studies classify the defect evolution in terms of wear stages.

2.6 Prognosis analysis

Several researchers have reviewed prognosis contributions (Engel et al. 2000), (Jardine et al. 2006), (Lee et al. 2006), (Heng et al. 2009), (Peng et al. 2010),

(Jammu & Kankar 2011), etc. There are two types of methods of prognosis: a physics 'model'-based and data-driven method, i.e. statistical and artificial intelligence. Physics-based prognostic models describe the physics of the system and failure modes based on mathematical models such as Paris' law, Forman law, fatigue spall model, contact analysis and 'stiffness-based damage rule' model. Data-driven prognostic models attempt to be driven by routinely and historically collected data (condition monitoring measurements). Data-driven prognostic models cover a high number of different techniques and artificial intelligence algorithms such as the simple trend projection model, time series prediction model, exponential projection using (ANN, data interpolation using ANN, particle filtering, regression analysis and fuzzy logic, recursive Bayesian technique, hidden Markov model, hidden semi-Markov model, system identification model, etc. Data-driven methods utilize data from past operations and current machine conditions, in order to forecast the remaining useful life. There are several reviews concerning the data-driven approaches (Schwabacher 2005), (Schwabacher & Goebel 2006) and (Camci et al. 2012).

2.6.1 Statistical approach

Yan et al. (2004) explored a method to assess the performance of assets and to predict the remaining useful life. At first, a performance model is established by taking advantage of logistic regression analysis with maximum-likelihood technique. Two kinds of application situations, with or without enough historical data, are discussed in detail. Then, real-time performance is evaluated by inputting features of online data to the logistic model. Finally, the remaining life is estimated using an Auto-Regressive–Moving Average model based on machine performance history; the degradation predictions are also upgraded dynamically. Vlok et al. (2004) proposed a residual life estimation method based on a proportional intensity model for non-repairable systems which utilise historic failure data and corresponding diagnostic measurements i.e. vibration and lubrication levels. Yang & Widodo (2008) proposed a prognosis method using SVM. The statistics-based models assume that historical data is representative of the future wear progress, which is not always the case. Probabilistic-based models assume that the whole wear evolution progress is represented by a probability distribution function i.e. Weibull.

2.6.2 AI approach

Li et al. (1999) utilized a recurrent neural network approach. Yam et al. (2001) proposed a model based on the recurrent neural network approach for the critical equipment of a power plant. Dong et al. (2004) proposed a model that combines condition prediction for equipment in a power plant based on grey mesh GM (1,1) model and BPNN on the basis of characteristic condition parameters extraction. Wang et al. (2004) evaluated the performance of recurrent neural networks and

neuro-fuzzy systems. By comparison, it was found that if a neuro-fuzzy system is properly trained, it performs better than recurrent neural networks in both forecasting accuracy and training efficiency. However, they often suffer from the need for complex training due to the huge number of possible combinations of damage scenarios that might occur in the case of rolling contact wear.

2.6.3 Physics-based approach

Physics-based prognostic models describe the physics of the system and failure modes based on mathematical models such as Paris' law, Forman's law, fatigue spall model, contact analysis and the stiffness-based damage rule model. Physics-based prognostic models are based on crack length, and defect area as illustrated in (Li et al. 1999), and (Li et al. 2000), or relations of stiffness as shown by Qiu et al. (2002). However, the most challenging issue within physics-based prognostic is to define the loading-damage relationship and to model it. There are models based on damage rules such as the linear damage rule, damage curve rule, and double-linear damage rule (Qiu et al. 2002). The drawback of these simplified functions is that they all use the constant damage factor, which is hard to estimate or measure. Moreover, these functions are either linear or multi-linear functions. That means that the estimated results might seem to match, with the overall measured results; however, both of they might describe different damage scenarios in behind. Therefore, the prediction based on such functions makes the prognosis a risky task. Recently, some model-based models have been utilised for the contact stress analysis to illustrate the wear evolution progress. These models provide more accurate predictions. Some models are based on contact stress analysis (Marble & Morton 2006) and some are based on system dynamics (Begg et al. 1999), (Begg et al. 2000). Chelidza & Cusumano (2004) proposed a method based on a dynamic systems approach to estimate the damage evolution. The results of these models also depend on the stress-damage function and the constant damage factor in use. These models assume that each wear mechanism generates stresses that in total equal the overall measured stresses. Therefore, the wear mechanics interactions and competitions are somehow ignored.

In summary, the survey shows that the data-driven approach is more suited to prognosis of rolling bearings than the physics-based approach. All prognosis approaches i.e. physics-based or data-driven have advantages and drawbacks in different applications and operating cases, specially, in the case of variable operating conditions. Moreover, the prognosis models try to control or delimit the effect of some operational variables. However, that is somehow possible in the experimental tests but not in real applications. The prediction based on simplified experimental tests, i.e. the ball on disc test, is easier than tests that use REBs. The statistical models represent the wear evolution as one function with the possibility of inserting weights. The statistical models assume that the past history profile represents the future failure mechanism of a specific component. However, the failure mechanisms are changing with respect to the failure evolution and the

involvement of failure mechanisms. This means that the statistical approach is not fully valid and might not represent wear progress, especially, if the evolution stages are highly varying, as they are in the case of wear evolution stages. ANN models use specific functions and multiple weights. However, ANN models have drawbacks once the system conditions are rapidly fluctuating. The model-based models are still representing the wear evolution with two stages. Moreover, the damage is represented as a damage factor. This really is a dramatic simplification to describe the wear evolution as a two-stage stabile phenomenon, whereas by nature it has a complex evolution process. Consequently, this kind of approach is very far from reality and its usability can be strongly criticized. The prognosis models can be improved remarkably by understanding the physics of the wear evolution progress and its associated measured outcomes. Actually, that will help the model-based approaches to provide better results and the data-driven approach to have better interpretations of the results and training inputs.

3. A descriptive model of wear evolution

The review and the discussions in publication I summarise the capabilities, advantages and disadvantages of bearing modelling and monitoring procedures i.e. SP, diagnosis and prognosis methods. It is observed that several experimental agreements with their associated analytical models are valid for specific wear definitions. The most commonly used method for wear testing is to artificially introduce a localised defect into the surface i.e. sharp, large enough and at specific radial location. However, in reality, the fault features change over the lifetime due to topographical and tribological changes. These changes are due to the fact that wear progresses in rolling contact because of the involvement of different wear and stress concentration mechanisms. Therefore, the review highlights the evolution monitoring challenge of the wear fault over the REB's lifetime. Moreover, it discusses the directions and implications of understanding the natural evolution of the wear process in REB and how can it be monitored and modelled effectively.

The discussion in the previous chapter highlights the need for a wear evolution model e.g. incremental numerical procedure which would be able to simulate and integrate the contact information continuously into a dynamic model. This means that the applied force due to the wear progress and its associated topographical and tribological conditions should be iteratively updated into the dynamic model. The review part of the monitoring and experimental testing methods highlights the need to understand the dominant physical damage mechanism of each testing method in order to interpret the measured signal in the correct way. In terms of signal processing and diagnosis method, it is very clear that several methods might show considerable capabilities at specific time intervals within the whole lifetime. However, they might have poor capability to indicate the fault at specific time intervals due to the change in the surface topography. Therefore, there is a need to study and validate these methods with the data of time intervals that contain several topographical changes due to wear evolution.

Generally, it is hard to describe the wear evolution progress due to the variety of the wear and contact mechanisms involved, which might produce different wear evolution scenarios. Therefore, the development of the wear fault over the lifetime is described in detailed in publication II. Publication II provides a new descriptive model of wear interaction and evolution which combines and integrates the large

experimental and numerical findings that have been accumulated in the literature. The descriptive model aims to generalise the most probable wear evolution scenario in REBs. First, it explains the interaction of multiple wear mechanisms in rolling contact. Second, it explains both the wear evolution on and beneath the contact surface from the running-in stage until spall occurs. Third, it explains the wear transition points during the wear evolution progress. Therefore, this chapter describes the wear evolution in rolling bearings over the whole lifetime in terms of the wear progression stages, surface topography evolution, wear mechanics interaction and influence factors of wear progression.

The detailed illustration of the new descriptive model with the supportive experimental findings in the literature is presented in publication II. In this section, a summary of the descriptive model and its characteristics are provided.

3.1.1 Wear evolution process

The bearing lifetime is described based on five stages. The wear evolution model assumes that, at a certain time interval of the steady-state stage, a transition into the defect initiation stage will take a place. Later, the evolution model describes the wear progress with the help of two assumptions: the existence of multiple stages that have specific transition events and existence of multiple wear and stress concentration mechanisms that are acting in each progress stage. Therefore, the wear evolution model that has been described in details with the help of literature (i.e. experimental findings) can be summarized as follows:

Some topographical changes might occur when the stresses in rolling contact increase due to the increased operating loads, additional loads due to faults i.e. imbalance, misalignment, bent shaft, looseness, and/or distributed defects i.e. high degrees of surface roughness and waviness, contaminations, inclusions. These topographical changes in the contact area generate stress concentration points and lubrication film disturbances.

At early stage, the concentrated stresses are not strong enough to produce a defect. It is mainly located in the loading zone and in the normal running track i.e. pure rolling points. Later, some sliding events, lubrication film transfers, false brinelling events due to stand-still events, might occur and introduce some degree of surface interaction. These surface interactions might appear as a reduction in the lubrication film, gapping, miss-matching between the rolling element profile and race profile, etc. Therefore, such surface conditions allow some abrasive wear events, some contaminations to enter the contact zone and minor vibration impacts. As a result of these actions, some surface dents might be generated. Therefore, the model describes how dents and in particular their asperities has the main role in the defect initiation and propagation process. The main assumption is that, as long as the asperity is large and sharp, the contact force between the rolling element and the asperity is large. The contact force and other loading forces contribute to the applied tangential force. The tangential force and the fric-

tion force are the main forces that generate a sufficient stress intensity factor (SIF) for crack opening and later for the crack propagation. However, since the rolling elements are rolling over the asperity and the asperity can be abraded, the original shape of the dent might change over the time. Therefore, the impact force and the crack opening and propagation progress are influenced. However, although the asperities are plastically deformed during the over-rolling cycles and degraded by abrasive and adhesive wear actions, they remain sufficiently high to produce tensile surface stresses. In the end, the propagated crack needs a secondary crack to reach the surface or it can attach to the rolling element once a sufficient adhesive bonding exists. During this process, the lubrication film is distorted and transfers into surface crack where another mechanical and chemical actions might accelerate the defect creation. When the defect is completed and the material is detached from the surface, new asperity is generated, new debris is generated, severe disturbances of lubrication film are generated, and the less hardening material (i.e. the material that was below the removed defect material) became the new surface.

This new descriptive model highlights the features of the generated defect and its asperity. The length of the generated defect depends on how long the crack could propagate in parallel to the surface before the detachment process occurred. The depth of the generated defect depends on how deep the crack was opened before it matches with a secondary or sub-surface inclusion. It also depends on the friction force and its depth of stress concentration. The width of the generated defect depends how far the force trajectories were propagating. The defect's length, depth and width are the basic elements of the new impact area. Therefore, a new impact force will be generated when the rolling element passes over the new defect i.e. especially at the trailing edge of new defect. The new defect will generate a number of defect serials, and the generated debris will generate several dents in different locations. First, the tangential force will be larger, since the asperity is larger and rougher than the initial dent asperities. Second, the new debris will generate a more severe dent, since it is larger and sharper than contamination particles. Debris might act as moving and distributed asperity. Moving asperities have a more random and non-linear way of action compare to the fixed ones, i.e. dent asperities. Large portion of debris will be pressed into the surface and generate more dents. Moreover, debris can act as an asperity and minor indenter and generate abrasive wear.

The five stages are schematically illustrated in Figure 2. The wear evolution progress produces several surface topographical changes. These topographical changes have a significant effect on the physical measurements, in particular, for condition monitoring purposes. At the end of the running-in stage, the roughness of the surface becomes smooth. Therefore, the steady-state stage is characterized by uniform lubricant film and contact mechanics, under normal operating conditions. When the surface is dented, the surface looks like peelings. Later, the shapes of the dents change due to the over-rolling and wear actions. However, the trailing edge acts as a stress raiser and the micro-cracking is initiated and opened

on and under the surface. The micro-cracks propagate from the surface downward with inclination, which depends on the rolling direction. The crack propagates later in parallel with rolling direction until it meets a secondary crack and connects to the surface or detaching process occur. Therefore, a relatively large material will be detached from the surface as debris particles. After the first pit, a number of pits and spalls are expected to occur in a serial pattern and extend in a wider and deeper manner, where the defect area becomes larger and rougher.
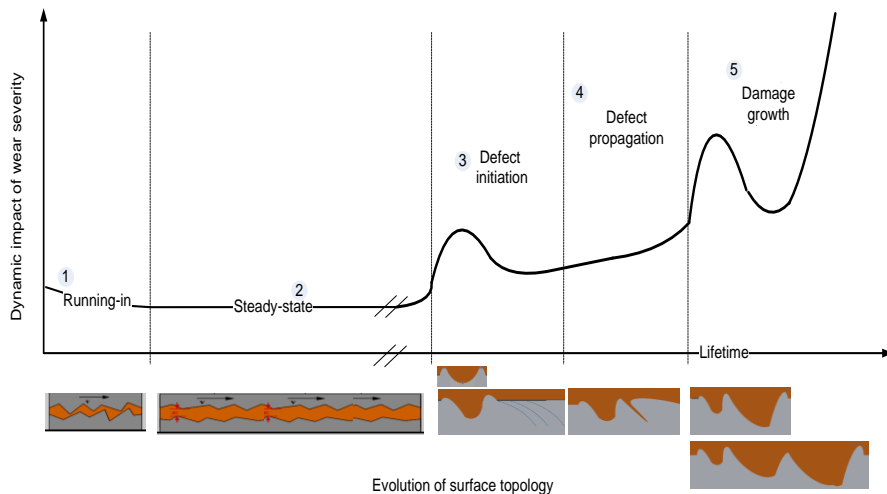
Figure 2. Evolution of dynamic behaviour and surface topography due to wear evolution.

It is worth explaining that several bearing manufacturers indicate that their lifetime estimation formula (e.g. the SKF bearing rating life formula) is based on the load and capacity assuming that the lubrication, oil contamination and operating conditions are ideal. Based on this assumption, such lifetime prediction shows quite a long lifetime and they fit very well with applications where the influences of the lubrication, oil contamination and operating conditions are controlled and neglected. The damage in the bearings of such applications are related to sub-surface defect which appear after a long period of operation i.e. after high repeated stress cycles, inclusions in the sub-surfaces and the degradation of the material properties.

However, several studies in the literature (Xu & Sadeghi 1996), (Mota & Ferreira 2009), (Morales-Espejel & Brizmer 2011), etc. show that under the normal loading conditions, the damage might appear much earlier once the lubrication, oil contamination and operating conditions are disturbed. Moreover, the rating life formula expresses the capacity to load ratio with an exponential factor that is related to contact type i.e. ball contact, rolling contact. Therefore, the studies show that lubrication disturbances might generate boundary lubricated contact, abrasive actions, pressing particles between surfaces, etc. and these actions in fact initiate

the surface dents. Therefore, the "service life factors" have been introduced to the life rating formula including lubrication, the degree of contamination, misalignment, proper installation and environmental conditions. These service life factors can easily influence the bearing lifetime at an early stage and initiate surface dents much earlier than sub-surface defects, in particular, for bearings that are operated under normal loading conditions (with some degree of variation) and assuming high material quality (which have been used in other applications and showed a long lifetime). In fact, several studies (Dwyer-Jones 1999), (Maru et al. 2007), (Morales-Espejel & Brizmer 2011) highlighted the fact that the probability of getting disturbances in lubrication, contamination degree and operating conditions are higher than having sub-surface inclusions at an early stage.

3.1.2 Rolling wear interactions

The new descriptive model of wear evolution i.e. five-stages emphasise the dynamic nature of wear. This dynamic nature is produced as a combination of wear mechanisms, stress concentration mechanisms, and operating conditions. This dynamic nature changes significantly the wear evolution scenario over the lifetime. Therefore, there is a need to understand the interactions and competitions among the involved wear mechanisms. The wear interaction process involves multi-stress concentration mechanisms (dent, debris and asperity) and multi wear mechanisms (fatigue, abrasive, and adhesive). The interactions among those different mechanisms might follow several stages as follows:

Interaction between stresses i.e. loading and fatigue wear: Assume that the rolling bearing is running under good operational conditions. The interaction process at this stage depends on the loading patterns and lubrication regimes. Therefore, the cyclic load and friction coefficient are the main attributes for contact damage criteria. The stress concentration might vary based on friction coefficient. Low friction means that stresses will most probably be concentrated in the subsurface. High friction means that stresses will be concentrated on the surface. The low-friction produces subsurface stresses which might accumulate and initiate a sub-surface crack. However, the high-friction produces micro-cracking. Later, the sub-surface crack propagates until a secondary or branching crack occurs before it reaches the surface. Therefore, based on the interaction process, the location of the fatigue wear might differ.

Interaction between stresses i.e. dent and different wear mechanisms: The dent-related wear mechanism can be produced by three different mechanisms: due to high stresses, local hardening (where one surface presses against the other surface in stand-still operations) and debris particles. The dent usually increases the surface roughness and might lead to the lubrication disconnect points during rolling, as the lubrication film move toward the bottom of the dent. Once the dent is generated, the interaction between the dent as stress mechanism and the fatigue wear might start. The generated dent might increase the contact stresses which

might initiate a new surface crack or might be accumulated to propagate an existing crack. On the other hand, the dent is considered as a stress concentration point which has the potential to be over-rolled by repetitive rolling mechanism (rolling elements in rolling bearing) or to be abraded by abrasive wear mechanism. The over-rolling and abrasive wear might smooth the dent sharpness and reduce the stresses which influence the fatigue wear mechanism. However, the abrasive wear might produce debris which generates another type of stress concentration point. Debris might act as a minor movable indenter (cutting object) where more abrasive wear might occur. Therefore, the wear interactions are important in order to understand how the wear and stress concentration mechanism might accelerate or de-accelerate the entire wear progress.

3.1.3 Influencing factors

The new descriptive model of wear evolution, i.e. five-stages, emphasises the influencing factors that can vary greatly within each stage. Therefore, it is important to consider influencing factors within each stage. Kappa (2006) had illustrated the influencing factors on the lubricant film quality. Therefore, in this study, the influencing factors are illustrated for the whole wear evolution in the same manner. In summary, the wear progress depends on the evolution and the changes of the loading conditions, surface quality, lubrication film quality and subsurface quality. These four factors generate together specific patterns of stress concentration and the damage progression scenarios.

In this sense, the wear modelling can utilise the descriptive model to acquire the required fundamentals about the wear mechanisms, stages transitions. The wear testing can also utilise the descriptive model to design experiments that can delimit the mechanisms underlying the natural wear evolution process. In fact, there are two approaches to wear testing, either using natural accelerated damage by applying overloads, adjusting the lubricant film thickness, and adding contaminated oil, or introducing an artificial defect into the surface by false-brinelling, erosion charge and scratching. The latter approach is widely used due to it is simplicity of modelling and validations, where the researchers can virtually introduce a well-known shape and size of the defect. They can also artificially introduce the same defect in the validation experiment. Furthermore, this approach delimits the testing complexity, as it focuses on a single artificial defect, compared to the natural defect approach where a number of defects might be produced. The drawback of this approach is that the damage criterion is somehow artificially determined which might be totally different from the defect in the real operation. Therefore, the artificially defected bearing tests have not much to tell us about the evolution of real wear progress. Finally, the wear monitoring can utilise the descriptive model to understand the physical phenomenon and surface topographical changes underlying the measured signals and provide better ability to diagnosis the wear severity over the time.

4.   Simulation model of wear evolution

The new descriptive model, which was presented in the previous chapter, explains and links the pre-stages with the crack initiation in the rolling contact i.e. steady state, defect localization and dent. In particular, the explanation of defect localization provides five potential stress localization mechanisms beside the contamination effect, lubrication disturbances and impact events. Moreover, it explains and links the post-stage of crack propagation i.e. damage growth. These stages and their associated issues provide the whole wear evolution progress, supported by the scientific experimental findings within the literature. The new descriptive model enhances the understanding of the fatigue wear i.e. crack initiation and propagation by considering the wear interactions i.e. over-rolling, abrasive, and adhesive. The model shows that the interaction among the involved wear mechanisms might accelerate or decelerate the wear evolution. That provides better understanding of the measured data and better estimation for the remaining useful time. In summary, the new descriptive model highlights the evolution stages, the transition event between the stages, the involved mechanisms in each stage, and the surface topographical changes over the lifetime and the influencing factors. The implication of the new descriptive model is to provide the basis for more realistic modelling, testing, and monitoring of the wear evolution in REBs.

The development of artificial forces influenced by the wear evolution and a simplified dynamic model that can be used for producing vibration simulation data are described in detail in publications III and IV. The purpose of developing this simple model is to gain a better understanding of the dynamics that influence rolling bearing, specially, due to the wear evolution process. The simulated signals over the whole lifetime can be used in the testing and training the automatic diagnosis tools. The forces due the wear evolution process are based on simple single degree of freedom, therefore, is not supposed to predict the absolute level of rolling bearing forces correctly. In fact, in the following sub-sections, the main assumptions and simplifications of the simulation model are presented.

4.1 Bearing force model

A real shaft-bearing system is generally very complicated and difficult to model (Arslan & Aktürk 2008). In addition to that, the modelling of wear evolution in REBs makes it even more complicated. Thus, the developed model obtains simple equations of motion based on single-degree-of-freedom system. The dynamic simulation models are basically described with help of the following equation of motion:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F_{ex}\} \ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \ (1)$$

Where $[M], [C], [K]$, and $\{F_{ex}\}$ are respectively matrices of system mass(es), damping coefficient(s), stiffness(es) and external forces. M is 11 kg, K is 5.96×10^7N/m (For outer race), and C is 5×10^7N.s/m. In fact, the dynamic simulation model covers four main phenomena and their forces: machine fault, surface imperfections, and defect fault.

The forces due to imbalance, surface roughness and waviness and surface defect faults can be represented in the system equation of motion as:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F_{imb} + F_s + F_d\} \ \dots\dots\dots\dots\dots\dots \ (2)$$

Where, F_{imb} is the imbalance force, F_s is the force due to surface roughness and waviness and F_d is the force due to surface defect. In the following sub-sections, the force magnitudes and frequencies of F_{imb}, F_s and F_d will be determined.

4.1.1 Force due to imbalance

If the structure holding the bearings in such a system is infinitely rigid, the centre of rotation is constrained from moving, and the centripetal force resulting from the imbalance mass can be found by the following formula:

$$F_{imb} = I_m * r * \omega^2 * \sin{(2\pi f_m t)} \ \dots\dots\dots\dots\dots\dots\dots\dots \ (3)$$

Where F_{imb} is the imbalance force, I_m is the mass, r is distance from the pivot, ω is the angular frequency, and f_m is the machine fault frequency i.e. the frequency at first speed order. For this model, the imbalance mass is considered to be 0.005 kg, r is 0.1 m, and the angular frequency is 160 rad/sec.

4.1.2 Force due to surface imperfections

The surface peaks and valleys influence the rolling contact forces. Therefore, the force generated due to surface texture is related to the frequency of surface waves

i.e. waviness and roughness. Harsha et al. (2004) presented a formula to calculate the restoring force due to waviness, which can be represented as the relation between local Hertzian contact force and deflection, as follows:

$$F_s = k(r_{\theta i})^p \dots\dots\dots\dots\dots\dots\dots (4)$$

Where k is the material deflection factor and p is related to contact type (i.e. it is 3/2 for ball bearings and 10/9 for rolling bearings). The parameter $r_{\theta i}$ is the radial deflection at the contact angle θ_i, which is given as:

$$r_{\theta i} = (x \cos \theta_i + y \sin \theta_i) - \{\gamma + \Pi_i\} \dots\dots\dots\dots\dots (5)$$

Where, x is the horizontal direction, y is the vertical direction, γ is the internal radial clearance, and Π_i is the amplitude of the waves at the contact angle and is given as:

$$\Pi_i = \Pi_0 + \Pi_p \sin(N\theta_i) \dots\dots\dots\dots\dots (6)$$

Where Π_p is the maximum amplitude of waves, Π_0 is the initial wave amplitude and N is the number of wave lobs. The contact angle is a function of the number of rolling elements and cage speed, which is given as:

$$\theta_i = \frac{2\pi}{N_b}(i - 1) + \omega_{cage}t \dots\dots\dots\dots\dots (7)$$

Where i = 1,…, N_b. The number of waves for the entire surface length L is given as:

$$N = \frac{L}{\lambda} \dots\dots\dots\dots\dots\dots\dots (8)$$

Where $L = \pi D$, and λ is the wavelength of roughness or waviness. The total restoring force is the sum of restoring forces from all of the rolling elements. Therefore, the total restoring force (considering waviness and clearance) is given as:

$$F_s = \sum_{i=1}^{N_b} k[(x \cos \theta_i + y \sin \theta_i) - \{\gamma + \Pi_i\}]^{\frac{3}{2}} \dots\dots\dots\dots\dots (9)$$

When the magnitude of contact angle is substituted in the total restoring force formula, the effects of the cage speed ω_{cage} and bearing diameter D becomes clear.

In this research work, the waviness order is around 80. The number of waves N is defined (Harsha et al. 2004) as follows:

$$N = \frac{2 * \pi * r}{\lambda} \dots\dots\dots\dots\dots\dots\dots (10)$$

Where r is the radius of outer or inner race, and λ is the wavelength. The common wavelength from bearing manufacture is measured to be about 0.8 mm. Therefore, the waviness order of an outer race of e.g. 10 mm in diameter and wavelength of 0.8 mm is about 78.5. With the same logic, the roughness waves can be estimated, where the roughness wavelength is 0.02 mm.

4.1.3 Force due to bearing defect

The force amplitude due to a surface defect can be represented as contact deflection or impact force. Several dynamic models simulate the force due to the defect as an expression of deflection:

$$F_d = k(\delta_{o,i,b})^p \dots\dots\dots\dots\dots\dots\dots (11)$$

Where k is the material deflection factor and p is related to the contact type, it is 3 for ball bearing and 10/3 for roller bearing. However, the radial deflection $\delta_{o,i,b}$ depends on the defect location. Therefore δ_o, δ_i , and δ_b represent the deflection due to an outer race defect, inner race defect and rolling eminent defect, respectively. In this paper, the focus is to consider only the outer race defect, therefore, it is given (Tadina & Boltezar 2011) as:

$$\delta_0 = \rho_j + \rho_b - (R_o + D_d) \dots\dots\dots\dots\dots (12)$$

Where, ρ_j is the radial position of the rolling element, ρ_b is the radius of the rolling element, R_o is the radius of the non-deformed outer race and D_d is the depth of the defect at the contact position. Epps (1991) shows that the radial deflection depends on contact mode at the entry and exit points of the defect. However, Sassi et al. (2007) described the total impacting force F_d as a sum of the static and dynamic components, as follows:

$$F_d = F_{static} + F_{dynamic} \dots\dots\dots\dots\dots (13)$$

The static component of the impact force depends on the maximum load Q_{max}, separation angle between rolling elements Ψ_i and load distribution factor ε:

$$F_{static} = Q_{max} \left(1 - \frac{1-\cos(\Psi_i)}{2\varepsilon}\right)^{1.5} \dots\dots\dots\dots\dots (14)$$

The static component of the impact force is a function of the static component of the impacting force, shock velocity ΔV and impacting material factor K_{impact}:

$$F_{dynamic} = F_{static} K_{impact} \Delta V^2 \dots\dots\dots\dots\dots (15)$$

Where ΔV^2 is given as:

$$\Delta V^2 = \frac{10}{28} g \left(\frac{d_{def}}{B_d}\right)^2 \dots\dots\dots\dots\dots\dots (16)$$

Where d_{def} is the defect length (in the rolling direction), and g is the gravitational constant and equal to 9.81 m/s^2.

Basically, the changes are related to the leading and trailing edge of the dent/defect. Therefore, a load function of five steps was defined, to represent the transient impacts, i.e. due to defect into the dynamic model, as shown Figure 3.
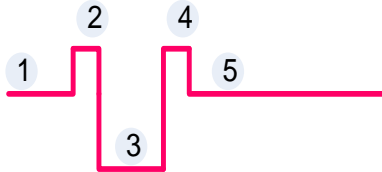
Figure 3. The simulated force diagram of wear defect.

4.1.4 Force due to wear evolution

The evolution of the wear defect stages is described in Figure 4, where the main defect topologies are illustrated with their force diagrams (on the right-hand side). The solid lines in Figure 4 show either dent or defect, while the dotted lines illustrate the potential cracking trajectories. To the right of Figure 4, the force diagrams of each defect topology are schematically illustrated and used to generate impulsive force series of the topographical features.
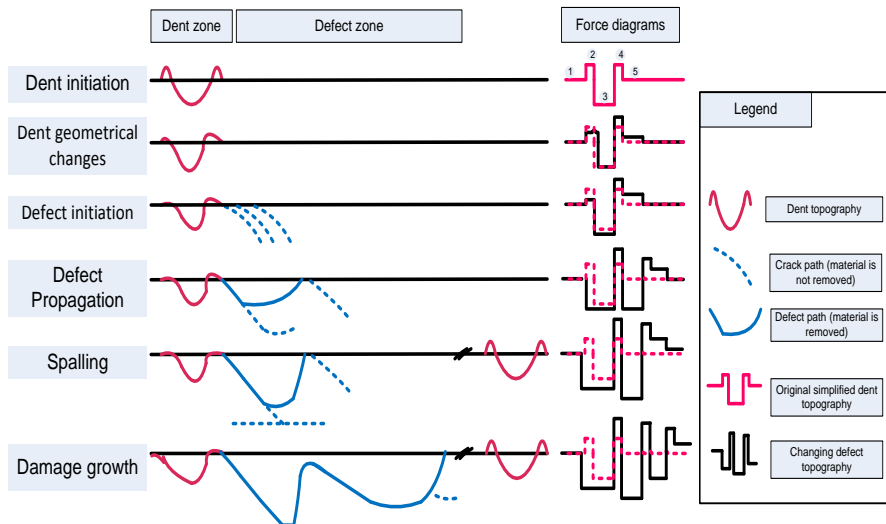
Figure 4. Wear defect and surface topology evolution (El-Thalji & Jantunen 2015b).

These load functions vary based on the variation of the surface topography as illustrated in Figure 4. Each stages of wear evolution progress were introduced

into the dynamic model by its corresponding force diagram (as shown in Figure 4 and with dimensions illustrated in Table 1. For each stage of wear evolution progress, different parameters of the load function are given in order to simulate the surface topographical variations due to wear progress.

Table 1. Specific events of wear evolution process.

| #	Wear evolution cases	Asperity height of the defect (in µm)	defect diameter/ size (in µm)	Reference
1	Real dent	0.4	50	(Dwyer-Joyce 2005)
2	Smoothed dent	0	50	(Dwyer-Joyce 2005)
2	Mild defect	20	300	(Dwyer-Joyce 2005)
3	Medium defect	20	1000	(Mota et al. 2008)
4	Large defect	30	3000	(Al-Ghamd & Mba 2006), (Nakhaeinejad 2010)
5	Smoothed large defect	4	3000	(Al-Ghamd & Mba 2006)
6	Damage growth, Multiple defects, two large defects	30	3000	(Al-Ghamd & Mba 2006)

There are two issues that influence the non-linear phenomena of wear evolution in rolling bearing. First, the bearing defect phenomenon is quite complex to be modelled due to the nonlinear evolution of wear defect. For example, the over-rolling and mild abrasive wear might smooth the trailing edge of the new defect. Therefore, the measured impact severity might be reduced as the asperity height and impact area is reduced.

Second, the width of the defect is the key parameter in impact severity. The wider the defect is the larger the impact area at the trailing edge. A wider impact area means higher peak amplitudes at the bearing defect frequency zone (Al-Dossary et al. 2009).

Several studies explain wear aspects and their evolutions with different rates and frequencies, depending on loading and operating conditions (Wang & Hadfield 1999), (Chue & Chung 2000), (Datsyshyn & Panasyuk 2001). Therefore, the wear progress is a nonlinear physical phenomenon.

4.1.5 Bearing fault frequency

The bearing fault frequency is basically based on the impact events which are generated when the rolling elements pass over a dent or defect. The fundamental train frequency FTF is given (Harris 1966) as:

$$FTF = \frac{F}{2} \times \left(1 - \frac{B_d}{P} \cos \theta_i\right) \dots\dots\dots\dots\dots\dots\dots (17)$$

Where, F is the shaft frequency, B_d is the rolling element diameter, P is the pitch diameter, and θ_i is the contact angle. The ball pass frequency for outer race fault BPFO is given (Harris 1966) as:

$$BPFO = \frac{N}{2} \times F \times \left(1 - \frac{B_d}{P} \cos \theta_i\right) = N \times FTF \dots\dots\dots\dots\dots (18)$$

Where, N is the number of rolling elements. The ball pass frequency for inner race fault BPFI is given (Harris 1966) as:

$$BPFI = \frac{N}{2} \times F \times \left(1 + \frac{B_d}{P} \cos \theta_i\right) = N \times (F - FTF) \dots\dots\dots\dots\dots (19)$$

The ball spin frequency BSF is given (Harris 1966) as:

$$BSF = \frac{P}{2B_d} \times F \times \left(1 - \left(\frac{B_d}{P} \cos \theta_i\right)^2\right) \dots\dots\dots\dots\dots\dots (20)$$

4.1.6 Bearing natural frequency

The amplitudes of spectral components at bearing fault frequencies are very weak at early stages of the fault. Epps (1991) has observed that the amplitudes of the spectral components at bearing fault frequency are very low, since the defect is very small at an early stage of degradation. Therefore, it is more effective to detect the frequency peaks at a high frequency band (bearing natural frequency) which already exists there at an early stage of fault development. Simply put, when the rolling element passes over a defect, the contact and impact events between the rolling elements and the raceways might excite the natural frequency of the raceways. Therefore, the frequency peaks at the bearing natural frequency might appear. Over time, when the defect becomes larger and the deflection amount also becomes larger, higher peaks at the bearing fault frequency become clear. Therefore, the bearing natural frequency is quite important as well as the bearing fault frequencies.

The contact events with the help of a small degree of loose fit between the raceway and the housing might excite the natural frequency of the raceway. Sassi et al. (2007) have shown that natural frequency ω_n of bearings for flexural vibration mode 'n' can be given as:

$$\omega_n = \frac{n[n^2-1]}{\sqrt{1+n^2}} \sqrt{\frac{EI}{\mu R^4}} \quad \text{in Rad/s} \dots\dots\dots\dots\dots\dots\dots (21)$$

Where E is the modulus of elasticity, I is the moment of inertia of the race cross section, μ is the mass per unit length, R is the radius of the ring and n is the order

of flexural vibration. In fact, n is the number of deformation waves in each mode (i+1). Sassi et al. (2007) highlighted that n_0 and n_1 correspond to rigid modes, therefore the flexural vibration modes start from n=2.

However, the value of the sectional secondary moment is needed before using Equation (21). Due to the difficulties in obtaining the exact value for a bearing ring with a complicated cross- sectional shape, equation (22) is suggested by NSK as best used when the radial natural frequency is known approximately for the outer ring of a radial ball bearing (NSK 2013).

$$f_n = 9.41 \times 10^5 \times \frac{K(D-d)}{\{D-K(D-d)\}^2} \times \frac{n[n^2-1]}{\sqrt{n^2-1}} \quad \text{in Hz} \ldots\ldots\ldots\ldots\ldots\ldots (22)$$

Where d is the bearing bore in mm, D is the bearing outside diameter in mm, K is the cross-sectional constant (K=.15 for outer ring of an open type) (NSK 2013). Therefore, the approximate natural frequency of the raceway depends greatly on the effective diameter of the raceway. For simplicity, the natural frequency of bearing raceway can be estimated based on ring vibration theory by the following formula (Irvine 2014):

$$\omega_n = \frac{E}{\pi D} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (23)$$

Where, E is the speed of sound in the material.

It is worth mentioning that the time between impact events is known as the epicyclical frequency. However, the topographical change due to the wear evolution most probably might change the drag and driving tangential forces, which makes the cage and rolling element travel more slowly than its epicyclical value, which is one of the main causes of slippage phenomenon. The simulation model introduces some degree of random slippage. However, the amount is related to the wear evolution stages. It is basically related to topographical features at each stage which influences the tribological features i.e. drag and driving tangential forces.

4.2 Wear mechanics

4.2.1 Wear interaction events

The over-rolling and abrasive actions have the potential to take place within the wear process once the asperity interaction degree allows this. The over-rolling and abrasive actions act as asperity degradation processes. Therefore, their actions are clear when the asperities of dents and defects are generated. It is clear that the impact effect of the asperity will be reduced as long as the asperity degrades. The developed model considers the degradation effect i.e. smoothing effect as degradation function. Sahoo & Banerjee (2005) present an analysis of the effect of

asperity interaction in elastic-plastic contact. Therefore, the developed model adopted their equation to estimate the asperity deformation due to the contact pressure at the asperity, which is given as follows:

$$\omega = z - d + 1.12\frac{\sqrt{w_a p_a}}{E} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(24)$$

Where, ω is the deformation due to the contact pressure at the asperity, z is the height of a given asperity, d is the mean separation between the rigid flat and the rough surface, w_a is the contact pressure on a single asperity, p_a is the global mean contact pressure on the surface and E is the modulus of elasticity. More details of the deformation of asperity can be found in (Sahoo & Banerjee 2005) and (Kim et al. 2013).

Abrasive wear depends on the lambda ratio (i.e. degree of asperity interaction). This is the ratio of lubricant film thickness to composite surface roughness and is given by the expression (Begg et al. 1999):

$$\lambda = \frac{h}{\sqrt{R_{\text{Surface 1}}{}^2 + R_{\text{Surface 2}}{}^2}} \quad\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(25)$$

Where λ is degree of asperity interaction, h is the lubricant film thickness, $R_{\text{Surface 1}}$ is the RMS roughness of the roller surface, and $R_{\text{Surface 2}}$ is the RMS roughness of the raceway. If λ is less than unity, it is unlikely that the bearing will attain its estimated design life because of the surface distress, which can lead to a rapid fatigue failure of the rolling surfaces. In general, λ ratios greater than three indicate complete surface separation. A transition from full elasto hydrodynamic lubrication (EHL) to mixed lubrication (partial EHL film with some asperity contact) occurs in the λ range between 1 and 3 (Begg et al. 1999).

Therefore, under specific tribological conditions, the abrasive wear is taking place in the asperity deformation as well. The abrasive wear depends of the height of the rolling element asperity which cuts part of the surface asperity of the raceway. Therefore, the developed model has adopted the abrasive wear model provided (Masen et al. 2005). It is assumed that the abrasive wear process gradually cuts over time a specific amount of the surface asperity with specific depth (d). This process will continue until the surface asperity reaches a specific height where no interaction with rolling element asperities.

4.2.2 Wear progression stages

The first transition event within the whole wear evolution process is the transition from running-in stage to steady-state stage. However, this transition event is not of interest for the developed model. Therefore, the second transition event is between the steady state and defect initiation stage, in particular, when the dentations occur. There are several potential defect locations and four denting mechanisms i.e. high stress, contamination, vibration, and lubrication disturbances. In

general, the dent might occur the applied stresses reach the yield stress limit (Y). In the elastic-plastic stage, the plastic deformation is small enough to be accommodated by an expansion of the surrounding area. The model has adopted the formula in (Harris 1991) to estimate the yield stresses for permanent deformation. Therefore, the model can estimate when the dent occurs by accumulating the applied stresses until the yield stress limit is reached.

The third transition point is between the dent and crack opening stage. Several models show that the stress fields resulting in elastic and plastic deformations introduce some changes to the shape of the surface (Kim & Olver 1998), (Karmakar et al. 1996), (Jaing & Sehitoglu 1996), (Franklin et al. 2001), (Masen et al. 2005), (Holmberg et al. 2005). These models help to estimate how much induced stresses (to initiate and accelerate the crack opening) are applied on the surface due to the surface asperity. The asperity-induced stresses are accumulated either by the time for crack opening process or the accumulated loading cycles. Different subscripts are used to designate the stress intensity factor for different modes. The stress intensity factor for mode I is designated K_I and applied to the crack opening mode. The asperity-induced stresses σ_0 can be estimated by the formula of Hannes & Alfredsson (2011). The model assumes that the crack is straight. Therefore, the developed model utilises the asperity model to estimate the asperity-induced stresses and accumulate the stresses until the crack opening limit is reached. The crack opening limit is determined by the stress intensity factor of mode I.

The fourth transition point is between the crack propagation and defect completion. Ringsberg & Bergkvist (2003) studied crack length, crack angle, crack face friction and coefficient of surface friction near the contact load. Tsushima (2007), Liu et al. (2007), Liu & Choi (2008), Donzella & Petrogalli (2010) and Leonel & Venturini (2011) have defined a number of issues that are required while modelling crack propagation: high stress location, depth below surface and the direction and angle of crack inclination. Therefore, based on these preliminary studies, quite a large number of models related to crack propagation have been developed (Glodez & Ren 2000), (Bogdanski & Brown 2002), (Bogdański & Trajer 2005), (Choi & Liu 2006), (Fajdiga et al. 2007), (Raje et al. 2008), (Canadinc et al. 2008), (Sadeghi et al. 2009), (Slack & Sadeghi 2010), (Slack & Sadeghi 2011), (Warhadpande et al. 2012), (Weinzapfel & Sadeghi 2012), (Santus et al. 2012), (Tarantino et al. 2013). The developed model in this thesis assumes that the crack propagation process will continue and the induced stresses will be accumulated until the crack propagation limit is reached. The crack propagation speed depends on the applied SIF. The crack propagation distance is accumulated until it reaches a pre-specified crack length. Navarro & Rios (1988) and Sun et al. (1991) proposed the model where the crack growth rate da/dN is assumed to be proportional to the crack tip plastic displacement δ_{pl}.

$$\frac{da}{dN} = C_0(\delta_{pl})^{m_0} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots(26)$$

where C_0, and m_0 are material constants that are determined experimentally. The total number of stress cycles N required for a short crack to propagate from the initial crack length a_0 to any crack length 'a' can then be determined as (Sun et al. 1991):

$$N = \sum_{j=1}^{z} N_j \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.(27)$$

Later, the damage growth stage will start and continue iteratively in a rapid growing manner. The developed model assumes that whenever the accumulated stress reaches the crack propagation limit, a new defect is generated.

4.3   Results of the simulation model

The basic estimation principle of the induced stresses due to wear evolution is based on the influence of the surface asperities that are generated by the topographical change. A surface asperity generates concentrated stresses at a certain point of line in the rolling contacted area.

Therefore, the fatigue wear theory is highly dependent on the asperity size and shape. However, as it is described in the literature (El-Thalji & Jantunen 2014), the wear process is a process of multiple wear and stress concentration mechanisms, and there are interactions and competitions among those mechanisms. For example, the surface asperity-induced stresses which are sufficient to initiate and propagate the fatigue wear process. At the same time, this asperity might be smoothed by the over-rolling and wear actions, which in return causes a reduction in the induced stresses. Therefore, the wear interactions and competitions should be considered as well. However, the wear modes that are modelled in this research work are fatigue and abrasive wear, besides the influence of over-rolling mechanism.

A groove ball bearing (SKF 61810-4) was used for the modelling and testing purposes. The bearing inner diameter is 50 mm, the outer diameter is 65 mm and the weight is 0.052 kg. As can be seen in Figure 5, the simulated time response of the bearing status over the entire lifetime indicates the healthy state i.e. good condition for a long period. Then, some degree of distortion appears in the simulated signal due to small degree of dent effect. The dent effect is introduced into the dynamic model as impact force function, as illustrated in (El-Thalji & Jantunen 2015b). At the dent stage, the topographical shape of the dent might generate some impact forces when the rolling element passes over the dent. Even though the edges of the dent are quite small and less sharp compare to a defect, the accumulated impact forces can initiate and propagate a crack which ends as a defect. Therefore, after the dent stage, the defect stage appears in Figure 5, where the impact force effect becomes stronger than at the previous stage.

It is worth mentioning that the new generated defect produces a rough topography i.e. larger edges, sharper edges, and deeper emptiness, which might initiate and sequentially propagate another defect at its trailing edge. Therefore, the simulated

signals show an increase in the impact force effect due to the number of the defects and their size in the defect stage. Later, the simulated signals show a decrease in the impact force effect due to some degree of smoothing effect. The smoothing effect is explained in (El-Thalji & Jantunen 2014) to be due to the interaction among different wear mechanisms i.e. abrasive, adhesive and over-rolling. Those smoothing actions change the defect features to generate weaker impact forces due to smaller and less sharp asperities within the defected area. However, the smoothing actions are responsible to change the surface topography where more asperities might generate and be in contact with the rolling element. Therefore, as the simulated signal shows, the noise degree increases over the smoothing period. In fact, during the smoothing period more debris is generated which also act as moving asperities and produce higher noise. Even though, the impact forces due the defected area have decreased over the smoothing period, but the accumulated impact forces are enough to initiate and propagate few cracks that end up with large defects. Those defects are large in terms of length, width and depth as illustrated in the experimental studies in the literature (Al-Ghamd & Mba 2006).
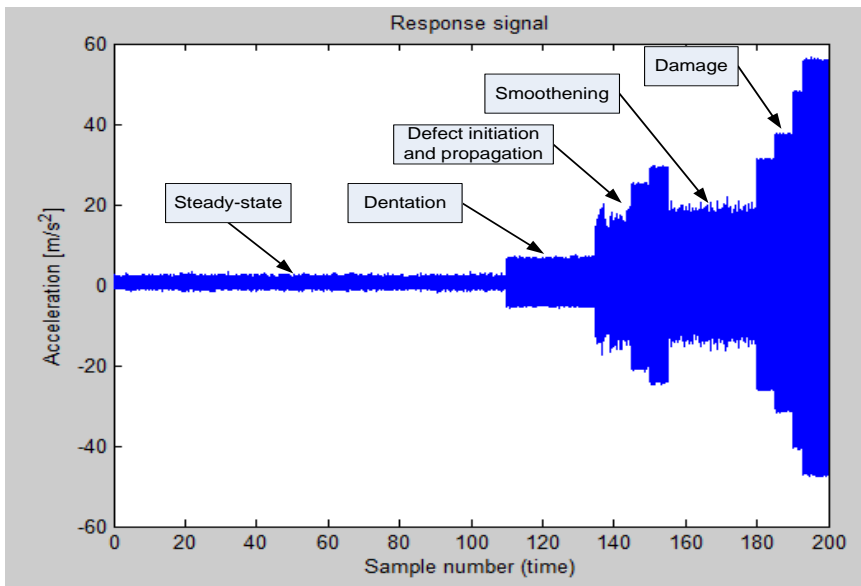

Figure 5. The simulated time response of bearing over the entire lifetime.

5.   Experimental findings

The experimental data set that has been used to validate the simulation model is described in publication III and in (Jantunen 2006). The discussion of a different and wide range of testing methods in publication I highlighted several significant testing aspects that should be considered so as to verify and validate the developed simulation model of this thesis.

First, the defect should be naturally created in order to let different wear and stress concentration mechanisms interact with each other and progress the wear development. This means, the approach of artificially introduced defect into the rolling bearings, is not applicable for this thesis. Since it is de-limiting the experimental tests of wear evolution in terms of ignoring some actors that are needed to produce the wear interaction and the evolution phenomenon that we are looking for.

Second, as we are looking for the entire lifetime, there is no chance of stopping the test and have a look or even take a sectional cut of the bearings raceways to prove the relation between the measured impacts and the topographical surface changes.

As mentioned in the modelling part, a groove ball bearing (SKF 61810-4) was used for the modelling and testing purposes. The bearing inner diameter is 50 mm, the outer diameter is 65 mm and the weight is 0.052 kg. The results of the developed model are shown in Figure 6 where the main wear events have been illustrated. There is a sudden increase in the response amplitude after the dent occurs, approximately around the middle of the lifetime, as shown in Figure 6. The effect of over-rolling and abrasive actions is clearly observed after the dent and defect completion stages. Later, the system response rapidly increased when the defect was completed and material removed. The response rapidly increases due to the generation of new asperities and wear-debris, which means the damage growth will continue. A number of bearings were tested in accelerated laboratory tests and the normalised RMS acceleration response is shown in Figure 7.
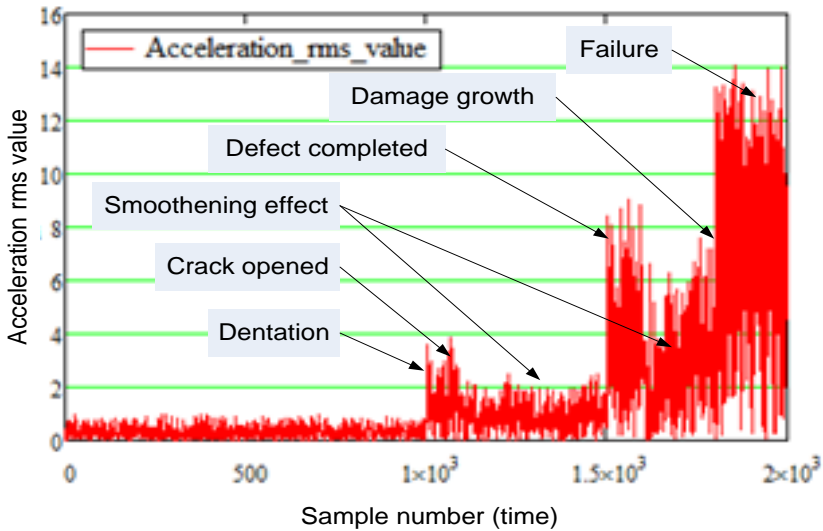
Figure 6. Normalized RMS value of vibration acceleration of simulated system.
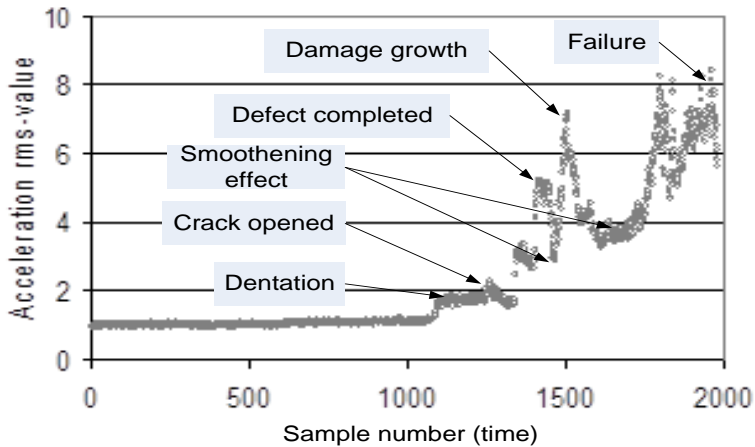
Figure 7. Normalized RMS value of vibration acceleration of laboratory tests (Jantunen 2006).

Both results in Figure 6 and Figure 7 indicate the steady-state nature of vibration acceleration during the first half of the lifetime. However, the simulated results show some variation due to the surface dent. Later, during the second half of the lifetime, a lot of variation is shown. Therefore, the simulated data illustrates the principal phenomena of dynamic response due to wear evolution process. The

53

principal agreement that can be observed in the predicted response compared to the measured response is related to the principal stages and events of wear evolution process. However, the simulation is based on multiple models that estimate the response, transition conditions and stress accumulations with certain degree of uncertainties. These models that are used, e.g. Hertzian contact theory, approximated impact area, stress intensity factors, are approximated models which can accumulate the uncertainties and give origin to error propagation.

Moreover, the measured data contains several issues related to different applications, boundary conditions, environments, third-body features, etc. At this stage of the model development, the model aims to simulate the wear evolution process in a simplified manner. These issues, which are mentioned above, might significantly influence the level of uncertainty of such a simplified model. Therefore, further development of the model should address them in order to enhance the accuracy level.

This study intends to provide reliable interpretations of the measured data and to predict the future wear progress. Therefore, at this stage, it aims firstly to achieve a principle agreement between the simulated and measured data. However, for quantitative comparison purposes, determining the shift between the simulated and measured curves is important. It requires having a comparable timeline for both simulated and measured data sets. At this stage of model development, it is a heavy computational task to generate the simulated dynamic response for the whole lifetime e.g. 6 million cycles. Keep in mind that the ultimate goal is to interpret and predict at certain time intervals by samples, rather than to get a well fitted curve over the whole lifetime. This model might help to tune the multiple models that are used to estimate the response, transition conditions and stress accumulations and to gain better accuracy degree, once the entire lifetime progress is considered.

6. Fault analysis

The detailed descriptions of the fault analysis and results are given in publication IV. The simulated data shown in Figure 5 is analysed in this chapter.

A deep groove ball bearing (SKF 61810-4) was used for modelling purposes. Because of various assumptions made in developing this model, the model was continuously verified with the model in (Sassi et al. 2007), both in their numerical and experimental figures. However, the dimension and feature differences between the modelled bearing type in this model (SKF 661819-4) and the type (SKF 6206) which considered in (Sassi et al. 2007) should be noticed.

In industry, the amplitude peaks at the bearing fault frequencies in the spectrum is the most commonly used indicator. Therefore, the amplitude peaks are extracted and tracked over the time. Several studies e.g. (STI 2012), (Graney & Starry 2012) have represented the evolution of the amplitude peaks in the spectrum.

In fact, these spectrum charts also represent the main characteristic frequencies related to the high frequency zone, the bearing natural frequency, and the bearing fault frequency zone, as shown in Figure 8.
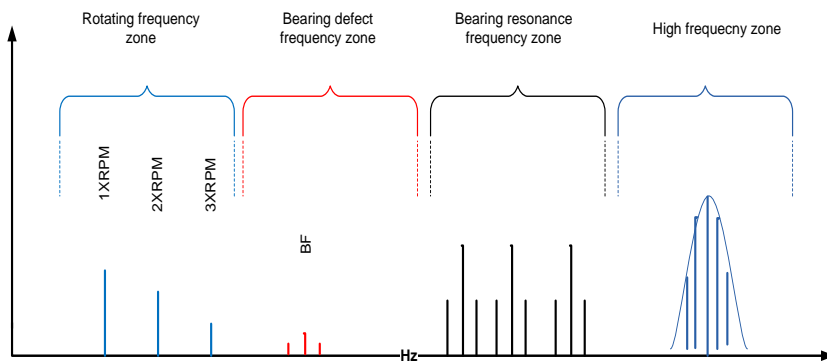

Figure 8. Schematic spectrum of vibration signal that contain bearing fault.

The spectrum correlates the amplitude change in these peaks at certain frequencies to the wear evolution progress. However, the topographical and tribological features e.g. defect shape, size and the uniformity of the lubrication film might change over the lifetime. Therefore, when the rolling element is passing over the defected area, the impulsive nature of the contact is changing as well. Moreover, the topographical and tribological changes might disturb the time between the impacts (between rolling element and the defected area), which introduces some degree of slippage. The slippage might delay the rotational time of the rolling element and, therefore, disturb the periodic phenomenon of rolling element contacts/impacts. Therefore, the frequency peaks which are related to those impact events might not be as clear (i.e. amplitudes) as if the impacts occurred in a perfectly periodic manner.

The time history signal is sampled at specific events of wear evolution process in order to apply several signal processing techniques. The aim is to illustrate the defect features and their changes over the time. There are several events which represent the wear evolution in rolling bearings. Eight different time intervals were sampled so as to study eight events of wear evolution: (1) normal operation; (2) mild surface roughness and waviness; (3) excitation of bearing natural frequency; (4) dent stage; (5) dent smoothing; (6) slippage case; (7) defect completion; (8) defect smoothing; (9) damage growth. The results of each case event are briefly illustrated as follows.

6.1 Machine imbalance fault

Under normal operation, the REBs are expected to work in a steady state manner, especially after the running-in stage.
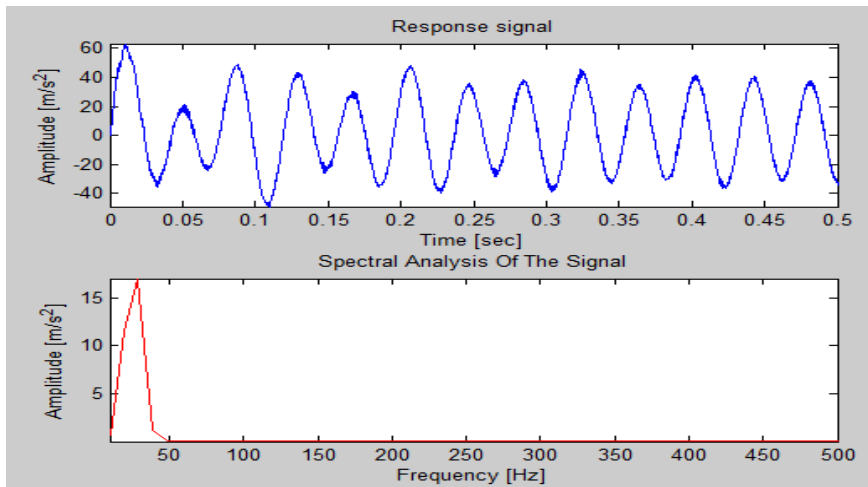
Figure 9. Time response and spectrum of the bearing under imbalance fault.

However, due to machine faults such as imbalance, bent shaft, misalignment, a number of distortions might be detected. For example, the imbalance fault can be detected once the spectrum starts showing peak amplitudes at 1xrpm (25 Hz), as shown in Figure 9.

6.2   Dented surface fault

When a dent is localised, it might be expected to see peak amplitudes at the bearing defect frequency zone in the spectrum, as shown in Figure 10. However, these are very weak peaks (compare to the defect stage). The amplitude peaks are observed at bearing natural frequency zone (around 1100 Hz) due the impact phenomenon when the rolling element passes over the dent. Epps (1991) explained the appearance of amplitude peak at bearing natural frequency (around 1100 Hz) and the amplitude peak in the bearing fault frequency zone (around 170 Hz). It was observed that the defect signal is of two parts: the first part originates from the entry of the rolling element into the defect, which generates an amplitude peak at bearing fault frequency zone. The second part originates due to the impact events between the rolling element and the trailing edge of the defect, which generates an amplitude peak at high frequency zone. However, the first part of the impact signals depends to a great extent on the defect size, which is very small at dent stage.
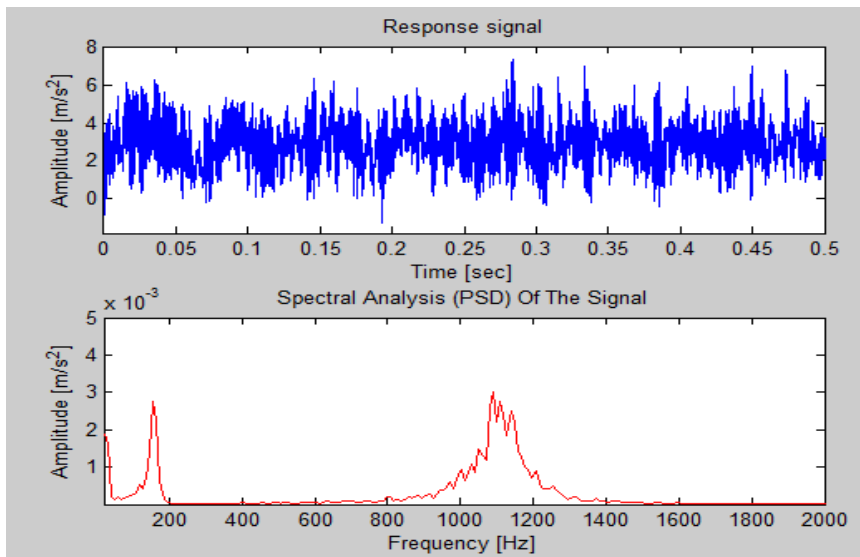
Figure 10. Time response and spectrum of the dent event.

Over the time, the over-rolling and mild abrasive wear make the dent impacts to become smaller. The reduction of peak amplitudes at bearing natural frequency zone due to smoothing actions is noticeable, as shown in Figure 11.
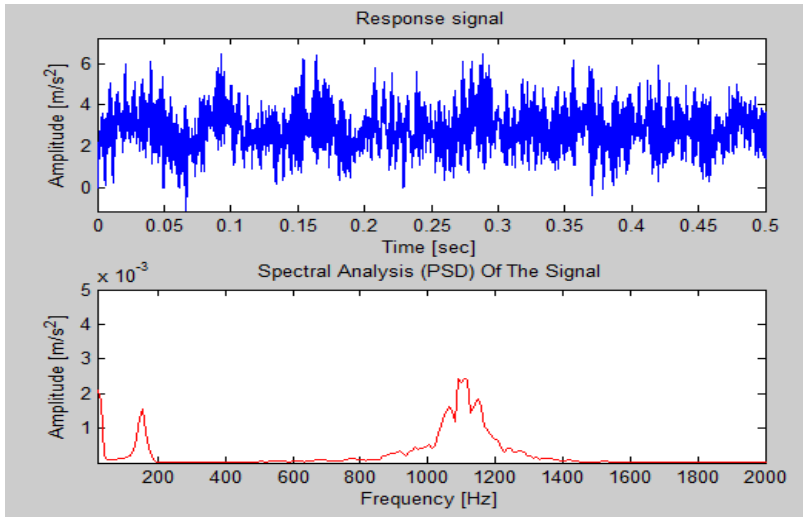
Figure 11. Time response and spectrum of the smoothed dent.

6.3   Defected surface fault

Even though, the smoothing process is taking place and reducing the impact forces, the stresses at the trailing edge of the dent might still be enough to initiate a crack. The crack will propagate and eventually end as a defect. The impact events at the trailing edge of the defect are much higher compared to the ones due to the dent. Therefore, it is expected to see higher peak amplitudes at the bearing defect frequency zone in the spectrum, as shown in Figures 12, 13 and 14. In fact, the impact events when the rolling element passes over the trailing edge do not only generate impulsive impact, but also distort the rolling element motion i.e. signal flatness. This distortion phenomenon is responsible for producing peak amplitudes at the harmonics of the bearing defect frequencies.
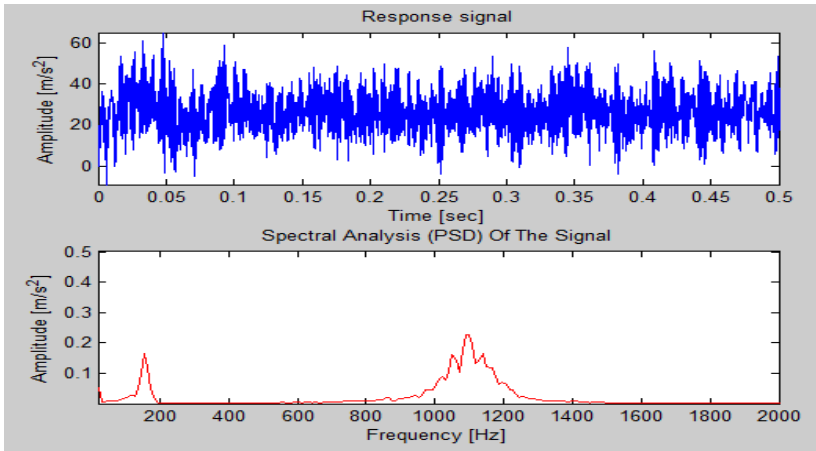
Figure 12. Time response and spectrum of the bearing with defect (0.3 mm).
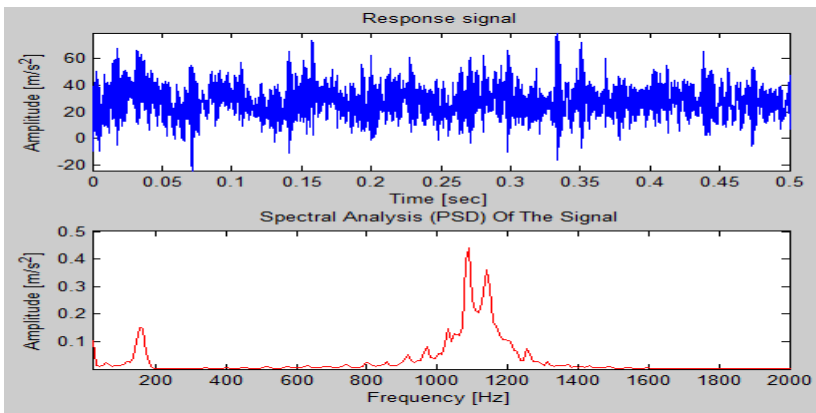
Figure 13. Time response and spectrum of the bearing with defect (1 mm).
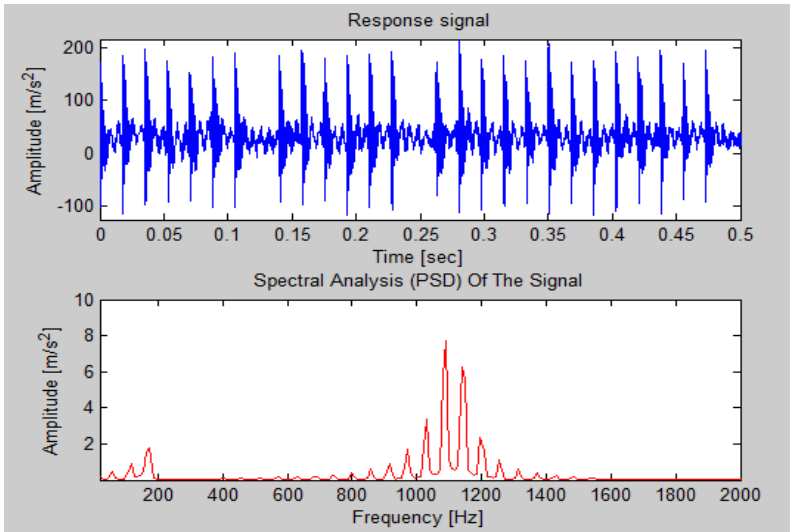
Figure 14. Time response and spectrum of the bearing with defect (3 mm).

6.4 Smoothed defect fault

The defect impacts also become smaller due to the over-rolling and mild abrasive wear. The reduction of peak amplitudes at bearing natural frequency zone can be clearly noticeable due to smoothing actions, as shown in Figure 15.
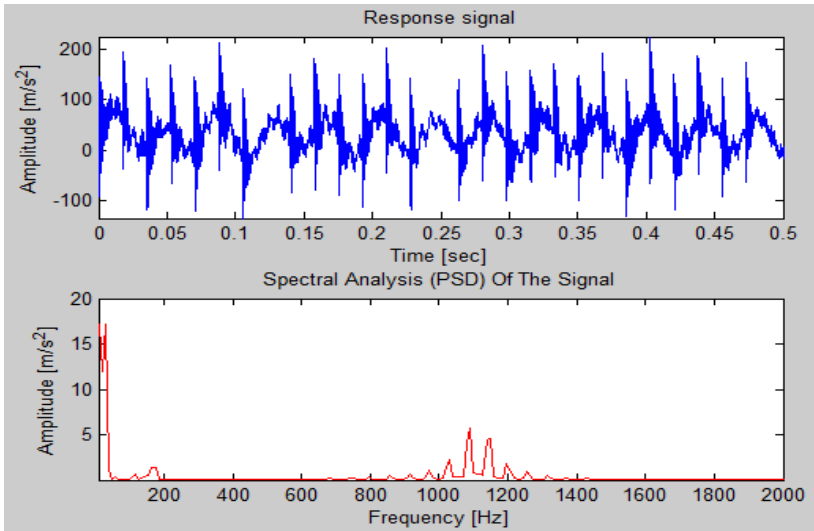
Figure 15. Time response and spectrum of the bearing with smoothed defect.

6.5   Damage growth fault

One important physical feature in the damage growth is the defect width. A wider defect means a wider impact area at the trailing edge. A wider impact area means higher peak amplitudes at the bearing fault frequency zone. Moreover, the length and the depth of the defect are expected to propagate over time until a failure occurs. In Figure 16, the wear fault features are illustrated. It is worth mentioning that Figure 16 represents an early-stage of damage growth. As the damage grows, it becomes more complicated to visualise the features of the spectrum due to the complex distortions associated with damage growth process.
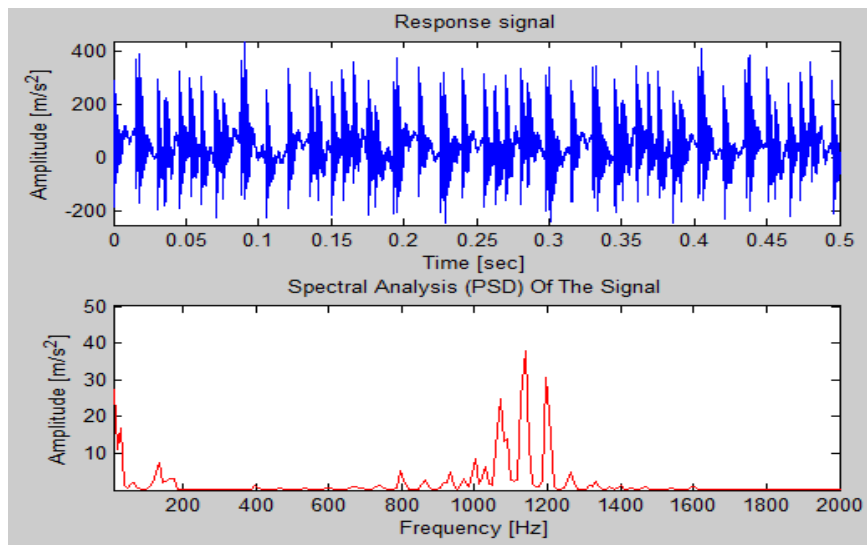
Figure 16. Time response and spectrum of the bearing under damage growth state.

In summary, the wear evolution stages can be illustrated in the following descriptions, with physical justification related to the topographical and tribological evolution of the defected bearing.

High frequency zone: The experimental findings that are discussed in (El-Thalji & Jantunen 2013) have shown that surface roughness is the main reason behind the amplitude peaks at the high frequency zone of the spectrum. The surface roughness and waviness generate the surface peaks and valleys which increase the probability of race surface peaks contact with rolling element surface. These contact events present amplitude peaks in the high frequency zone. Over time, the contact events between the surface peaks of the race and rolling element surface smooth the surface. Moreover, the film thickness will be stabilised and become uniform. However, the bearing geometrical and tribological characteristics might change due to the loading and operating conditions. For example, there is all the

time some degree of error in the contact profile between the race and the rolling element which disturbs the pure rolling contact (Morales-Espejel & Brizmer 2011). Another example is the lubrication transfer into surface valleys (Kotzalas & Doll 2010). Such contact disturbances produce high stresses when the rolling elements pass over the surface asperities. These stresses will appear as amplitude peaks in the high frequency zone in the spectrum. These contact events might involve minor abrasive wear which make the surface smoother, and higher stresses will be generated from those events. Later, the distributed wear debris will be the main issue in generating high stresses. That means higher amplitude peaks in the high frequency zone are expected at a later stage.

Bearing natural and bearing fault frequency zones: The impact events when the rolling element passes over the defective area (i.e. with the help of a small degree of loose fit between the raceway and the housing might) excite the natural frequency of the raceway. Epps (1991) observed that the defect signal is of two parts: the first part originates from the entry of the rolling element into the defect, which generates an amplitude peak in the bearing fault frequency zone. The second part originates due to the impact event between the rolling element and the trailing edge of the defect, which generates an amplitude peak at bearing natural frequency zone. However, the defect signal might change as the wear (defected area) is evolved over time.

The abrasive wear generates some internal debris, which might be transferred with the oil lubrication into the valleys of either the surface waviness or the contact deflection. At the moment when the rolling element passes over the valleys that contain debris, the rolling element might press the debris into the surface and generate a localised dent. This localised dent generates impact events that might excite the bearing natural frequency. Therefore, the peak amplitudes at the bearing natural frequency might be seen. In fact, the dent acts as a stress riser in particular at the trailing edge of the dent (Alfredsson et al. 2008). However, the impact event which is generated when the rolling element passes over the new dent is very small. Moreover, it becomes even smaller due to the over-rolling and mild abrasive wear of the asperity at the trailing edge of the new dent. However, the high stresses at the trailing edge are still enough to initiate a crack. The crack will propagate and end eventually become a defect. The defect will have leading and trailing edges as well.

However, the impact events at the trailing edge of the defect are much higher compared to that of the dent. Therefore, it is expected to see higher peak amplitudes at the bearing natural frequency in the spectrum. In fact, the impact event (i.e. generated when the rolling element passes over the trailing edges) generates impulsive impact and distorts the rolling element motion. The distortion phenomenon is responsible for producing peak amplitudes at the harmonics of the bearing defect frequencies. It depends on whether the impact area (the trailing edge area in contact) is large enough to distort the impact signals.

The over-rolling and mild abrasive wear will act again to smooth the trailing edge of the new defect. Therefore, a clear reduction in peak amplitudes at both the bearing defect and bearing natural frequency zones is expected. However, the high stresses at the trailing edge are enough to initiate a crack for the next defect. The width of the defect is the key parameter in the impact severity. In fact, the impact area of a dent is smaller than the impact area of the defect. Therefore, the crack trajectories of the defect will be further from each other when compared to the crack trajectories of the dent. The crack trajectories are the main issue that determines the width of the new defect. A wider impact area means higher peak amplitudes at both bearing defect and bearing natural frequency zones.

A significantly important issue of this simulation model is the consideration of the slippage phenomenon and its effect on frequency domain features. The time between impact events is known as the epicyclical frequency, which we try to extract from the spectrum. For example, the topographical change due to the wear evolution might most probably change the drag and driving tangential forces which make the cage and rolling element travel more slowly than its epicyclical value. In fact, as the wear become more severe, the topographical and tribological features of the surface generate and influence stronger the drag and driving tangential forces, which means more slip and disturbances. Therefore, the amplitude peaks at the defect frequency might not be clear, and several harmonics and sideband peaks will appear.

Rotating frequency zone: The machine faults e.g. imbalance, bent shaft, misalignment, looseness have specific characteristic features at the rotating speeds and their orders. In fact, those machine faults have in-direct influence on the wear initiation and evolution, as they introduce some changes in the contact characteristics and load distribution; for example, the misalignment introduces higher contact stresses at specific part of the raceway, which is one of the main reasons for wear initiation.

7. Discussion

The largest proportion of the studies which have been reviewed in publication I focused on the localized faults at a certain time interval rather than the fault development over the entire lifetime. In summary, the wear evolution process is quite complex due to the involvement of several wear mechanisms (i.e. fatigue, abrasive, adhesive, corrosive) and several stress concentration mechanisms (i.e. dent, asperities, debris, sub-surface inclusions). These involvements and their interactions and competitions produce a wear evolution progress which varies significantly with respect to surface topographical and tribological changes. As the fault topography is changing over the lifetime that simply means the fault features are changing over time. In this sense, the fault topography, which is assumed to be constant in several simulation models, should be changing over time. Moreover, there is a need to clearly determine the fault features of specific wear evolution stages and to understand how different signal analysis methods can detect such features, in order to effectively track the fault features detected.

The approach described in publication II and modelled in publication III simply tries to show the possible influence of various artificial dynamic loads (i.e. forces due to the fault topography) related to wear evolution progress. Therefore, the influence of wear has been introduced into the excitation forces. It should be noted that the simulation described in Chapter 4 and publications III and IV is limited to the first radial vibration mode, as it is the most important vibration mode. Moreover, with this kind of simplified model i.e. a one-degree-of-freedom model, the vibration at the natural frequency is very dominant.

The industrial applications utilise the in-field measurements in order to plan the required maintenance actions in a cost-effective manner. The available signal processing techniques have been developed to deal with speed fluctuation effect, signal noise, and smearing effect. However, the main challenge of fault analysis in this thesis is to illustrate the expected fault features due to the wear evolution process over the lifetime. The fault features can serve as diagnostic indicators, which can be used to track the wear evolution progress over the lifetime of the bearing. The simulation dynamic model of wear evolution, which is described in publications III and IV, provides a simplified way to mimic the wear process in rolling bearings. In fact, it is capable of mimicking the topographical surface evolu-

tion due to wear fault in rolling bearings over their lifetime. However, the study is based on multiple models that estimate the response, transition conditions and stress accumulations with certain degree of uncertainties. Thus, the potential fault features of wear defect have been analysed with the help of simple fault analysis. The dynamic nature of the wear process, i.e. interaction among different wear mechanisms, makes the clarity of the fault features vary over the time.

The results represent and illustrate the fluctuation in the clarity of the indicators at different deterioration stages of the bearing. Thus, the tracking process based on specific indicators might be not sufficient to diagnose the wear evolution and the bearing health. In this sense, the study shows how fault indicators might be influenced by the wear evolution process i.e. surface topography changes due to wear. Thus, it provides the potential changes in the tracked indicator, and highlights the need to combine different tracked indicators in order to enhance the tracking effectiveness.

The wear fault development is quite complex evolution process that can involve different wear mechanisms. The interactions among the wear mechanisms might accelerate or de-accelerate the overall wear process. Therefore, the assumption that fault size is increasing over time might not be the best way to represent the wear process over the entire lifetime. In this sense, the simulation model could in a simplified manner represent the real wear process. The real wear process contains several aspects e.g. steady-state conditions, dent stage, defect initiation and propagation, defect completion, defect smoothing, multi defects generation, and failure. These aspects have different physical phenomena where the signal analysis methods might vary in their effectiveness in extracting and detecting such phenomena.

The fault analysis shows the basic challenges that might influence the effectiveness of extracting the required fault features, e.g. at the dent stage due the high health state variations, at the defect completion stage which lasts for a somewhat short time interval and diminishes again, at the smoothing period where the impulsive phenomenon is somewhat diminished, and at the rapid increase of the wear severity at the end of the lifetime. The simulation model can predict the bearing dynamic response for the remaining lifetime as long as the fault follows the assumed wear evolution scenario. Therefore, the developed dynamic model can be used in the future to represent the wear evolution process in more realistic way than just considering the increase in the fault size.

8.   Conclusions

The modern machineries demand reliable and effective predictive health monitoring tools. Prognosis of machine health is of the greatest importance to achieve cost-effective production and maintenance. There is great potential in enhancing the prognosis of bearing health if the degradation process is understood.

A comprehensive review of rolling contact wear was carried out in order to describe the wear evolution process in rolling bearings over the whole lifetime. Based on the experimental tests in the literature, a descriptive wear evolution model is developed to represent the most probable wear evolution scenario that might occur for the rolling bearing over the lifetime. The scenario described is suggested to cover several wear mechanisms, stress concentration mechanisms and their interactions to represent real wear fault development.

A dynamic model is developed based on the descriptive wear evolution model. The simulation model shows, in other words predicts the dynamic behaviour of the rolling bearing during the entire wear evolution progress. The results of the simulation model were compared with the data sets of comprehensive experimental tests (as described in chapter 5) in order to validate the simulation model. The results show a principle agreement with the experimental results. The simulation model is beneficial, as it contains different physical phenomena that might occur during the real wear evolution process. The results show that fault and contact topographies, significantly influence the dynamic response and not just the size of the fault (i.e. as is the commonly used assumption in the literature). The data from laboratory tests represent the overall dynamic behaviour; however it is complicated to trace back the influencing factors. In this sense, the simplified simulation model is an effective tool for understanding the dynamic behaviour of the rolling bearing which is influenced by the wear evolution progress, loading and operating factors.

The importance of the wear mechanisms and their interactions is apparent in the simplified simulation model, experimental tests and fault analysis that were carried out. The use of the simulated data, i.e. which covers the wear evolution aspects, is suggested to be used in the future for verifying the effectiveness of fault detection methods. Moreover, the use of, simulated data can be suggested for the future development of a novel diagnosis method that can effectively extract the health

state of the bearing. The dynamic model which was used to generate the simulated data can also be used in the future to develop a prognosis model that can simulate the response of the remaining lifetime.

Even though the simulation data provides a promising understanding of the bearing behaviour and health during the wear evolution progress, there are no test results from a real production environment. It should be noted that the simulation results are based on simplified stress accumulation approach and unfortunately the stress accumulation based on finite element approach is more demanding. It should also be noted that the simulation model has been developed with several simplifications, and unfortunately the modelling of wear evolution over the whole lifetime is more demanding. Thus, the simulation is based on multiple models that estimate the response, transition conditions and stress accumulations with certain degree of uncertainties. However, the simulation model has two significant and potential benefits. First, it can be used as a tool so as to verify the diagnostic techniques that are available for wear monitoring. Therefore, it can illustrate several diagnostic indicators that might be utilised for tracking the wear evolution process in rolling bearings. Second, the simulation model can be used as a part of a prognosis model to predict the response of remaining lifetime during wear evolution process.

Based on the research reported in this thesis and the above conclusions, some suggestions can be made for further work:

- Wider testing of the developed approach both in the laboratory and in industry is suggested. In these tests the benefits of the descriptive wear evolution model could be tested more thoroughly, including the variation in the wear mechanisms and influencing factors i.e. loading and operating factors. Furthermore, the tests could also be used to study the scale influence on the dynamic behaviour of the rolling bearings.
- One further step in gaining a better understanding of bearing wear monitoring could be the extension of the simulation approach to represent real machines i.e. higher degrees of freedom and several influencing factors of real environments.
- Further work could also be done in verifying several signal analysis and diagnosis methods that are suggested in the literature.

References

Abbasion, S. et al., 2007. Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine. *Mechanical Systems and Signal Processing*, 21(7), pp.2933–2945.

Al-Dossary, S., Hamzah, R.I.R. & Mba, D., 2009. Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing. *Applied Acoustics*, 70, pp.58–81.

Alfredsson, B., Dahlberg, J. & Olsson, M., 2008. The role of a single surface asperity in rolling contact fatigue. *Wear*, 264, pp.757–762.

Al-Ghamd, A.M. & Mba, D., 2006. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. *Mechanical Systems and Signal Processing*, 20, pp.1537–1571.

Altmann, J. & Mathew, J., 2001. Multiple Band-Pass Autoregressive Demodulation for Rolling-Element Bearing Fault Diagnosis. *Mechanical Systems and Signal Processing*, 15(5), pp.963–977.

Antoni, J., 2007. Cyclic spectral analysis in practice. *Mechanical Systems and Signal Processing*, 21(2), pp.597–630.

Antoni, J., 2007. Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions. *Journal of Sound and Vibration*, 304(3-5), pp.497–529.

Antoni, J., 2009. Cyclostationarity by examples. *Mechanical Systems and Signal Processing*, 23(4), pp.987–1036.

Antoni, J. et al., 2004. Cyclostationary modelling of rotating machine vibration signals. *Mechanical Systems and Signal Processing*, 18(6), pp.1285–1314.

Antoni, J., 2006. The spectral kurtosis: a useful tool for characterising nonstationary signals. *Mechanical Systems and Signal Processing*, 20(2), pp.282–307.

Antoni, J. & Randall, R.B., 2001. Optimisation of SANC for separating gear and bearing signals. In *Proceedings of COMADEM conference*. Manchester, pp. 89–96.

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

Antoni, J. & Randall, R.B., 2004a. Unsupervised noise cancellation for vibration signals: part I-evaluation of adaptive algorithm. *Mechanical Systems and Signal Processing*, 18, pp.89–101.

Antoni, J. & Randall, R.B., 2004b. Unsupervised noise cancellation for vibration signals: part II-a novel frequency-domain algorithm. *Mechanical Systems and Signal Processing*, 18, pp.103–117.

Arslan, H. & Aktürk, N., 2008. An investigation of rolling element vibrations caused by local defects. *Journal of Tribology*, 130.

Ashtekar, A., Sadeghi, F. & Stacke, L.-E., 2008. A New Approach to Modeling Surface Defects in Bearing Dynamics Simulations. *Journal of Tribology*, 130.

Baillie, D.C. & Mathew, J., 1996. A Comparison of Autoregressive Modeling Tech-

niques for Fault Diagnosis of Rolling Element Bearings. *Mechanical Systems and Signal Processing*, 10(1), pp.1–17.

Begg, C.D. et al., 1999. Dynamics Modeling for Mechanical Fault Diagnostics and Prognostics. In *Mechanical system modeling for failure diagnosis and prognosis, Maintenance and reliability conference*. Gatlinburg, Tennessee.

Begg, C.D., Byington, C.S. & Maynard, K.P., 2000. Dynamic simulation of mechanical fault transition. In *Proceedings of the 54th meeting of the society for machinery failure prevention technology*. Virginia Beach, VA., pp. 203–212.

Bogdanski, S. & Brown, M.W., 2002. Modelling the three-dimensional behaviour of shallow rolling contact fatigue cracks in rails. *Wear*, 253, pp.17–25.

Bogdański, S. & Trajer, M., 2005. A dimensionless multi-size finite element model of a rolling contact fatigue crack. *Wear*, 258, pp.1265–1272.

Boness, R.J. & McBride, S.L., 1991. Adhesive and abrasive wear studies using acoustic emission techniques. *Wear*, 149, pp.41–53.

Camci, F. et al., 2012. Feature Evaluation for Effective Bearing Prognostics. *Quality and Reliability Engineering International*.

Canadinc, D., Sehitoglu, H. & Verzal, K., 2008. Analysis of surface crack growth under rolling contact fatigue. *International Journal of Fatigue*, 30, pp.1678–1689.

Cao, M. & Xiao, J., 2008. A comprehensive dynamic model of double-row spherical roller bearing—Model development and case studies on surface defects, preloads, and radial clearance. *Mechanical Systems and Signal Processing*, 22, pp.467–489.

Chaturvedi, G.. & Thomas, D.W., 1982. Bearing fault detection using adaptive noise cancelling. *Journal of Sound and Vibration*, 104, pp.280–289.

Chelidza, D. & Cusumano, J.P., 2004. A dynamical systems approach to failure prognosis. *Journal of vibration and acoustics*, 126.

Chiementin, X., Bolaers, F. & Dron, J.-P., 2007. Early Detection of Fatigue Damage on Rolling Element Bearings Using Adapted Wavelet. *Journal of Vibration and Acoustics*, 129(4), p.495.

Choi, Y. & Liu, C.R., 2006. Rolling contact fatigue life of finish hard machined surfaces Part 1. Model development. *Wear*, 261, pp.485–491.

Chue, C.. & Chung, H.., 2000. Pitting formation under rolling contact. *Theoretical and Applied Fracture Mechanics*, 34.

Collis, W.B., White, P.R. & Hammond, J.K., 1998. higher-order spectra: the bispectrum and trispectrum. *Mechanical Systems and Signal Processing (1998)*, 12(3), pp.375–394.

Da Costa, E.L., 1996. *Detection and identification of cyclostationary signals*.

Dalpiaz, G., Rivola, A. & Rubini, R., 2000. Effectiveness and Sensitivity of Vibration Processing Techniques for Local Fault Detection in Gears. *Mechanical Systems and Signal Processing*, 14(3), pp.387–412.

Datsyshyn, O.P. & Panasyuk, V.V., 2001. Pitting of the rolling bodies contact surface. *Wear*, 251, pp.1347–1355.

Djebala, A., Ouelaa, N. & Hamzaoui, N., 2007. Detection of rolling bearing defects using discrete wavelet analysis. *Meccanica*, 43(3), pp.339–348.

Dong, Y. et al., 2004. A combining condition prediction model and its application in power plant. In *Proceedings of the Third international conference on Machine learning and cybernetics*. Shanghai, pp. 26–29.

Donzella, G. & Petrogalli, C., 2010. A failure assessment diagram for components subjected to rolling contact loading. *International Journal of Fatigue*, 32, pp.256–268.

Dwyer-Jones, R.S., 1999. Predicting the abrasive wear of ball bearings by lubricant debris. *Wear*, 233-235, pp.692–701.

Dwyer-Joyce, R.S., 2005. The life cycle of a debris particle. *Tribology and Interface Engineering series*, 48, pp.681–690.

El-Thalji, I. & Jantunen, E., 2014. A descriptive model of wear evolution in rolling bearings. *Engineering Failure Analysis*, 45, pp.204–224.

El-Thalji, I. & Jantunen, E., 2015a. A summary of fault modelling and predictive health monitoring of rolling element bearings. *Mechanical Systems and Signal Processing*, pp.1–21.

El-Thalji, I. & Jantunen, E., 2015b. Dynamic modelling of wear evolution in rolling bearings. *Tribology International*, 84, pp.90–99.

El-Thalji, I. & Jantunen, E., 2013. Wear of rolling element bearings. In *Proceedings conference of Condition Monitoring and Diagnostic Engineering Management (COMADEM)*. Helsinki, Finland.

Endo, H. & Randall, R.B., 2007. Application of a minimum entropy deconvolution filter to enhance autogressive model based gear tooth fault detection technique. *Mechanical Systems and Signal Processing*, 21(2), pp.906–919.

Engel, S.J. et al., 2000. Prognostics , The Real Issues Involved With Predicting Life Remaining. In *IEEE Aerospace Conference*. New York, pp. 457–469.

Epps, I.K., 1991. *An investigation into the characteristics of vibration excited by discrete faults in rolling element bearings*. Christchurch: University of Canterbury, New Zealand.

Fajdiga, G., Ren, Z. & Kramar, J., 2007. Comparison of virtual crack extension and strain energy density methods applied to contact surface crack growth. *Engineering Fracture Mechanics*, 74, pp.2721–2734.

Franklin, F.J., Widiyarta, I. & Kapoor, A., 2001. Computer simulation of wear and rolling contact fatigue. *Wear*, 251, pp.949–955.

Fukata, S. et al., 1985. On the vibration of ball bearings. *Bulletin of JSME*, 28(239), pp.899–904.

Gardner, W.A., 1986. The spectral correlation theory of cyclostationary time-series. *Signal processing*, 11, pp.13–36.

Ghafari, S.H., Golnaraghi, F. & Ismail, F., 2007. Effect of localized faults on chaotic vibration of rolling element bearings. *Nonlinear Dynamics*, 53(4), pp.287–301.

Glodez, S. & Ren, Z., 2000. Modelling of crack growth under cyclic contact loading. *Theoretical and Applied Fracture Mechanics*, 30, pp.159–173.

Graney, B. & Starry, K., 2012. Rolling element bearing analysis. *Material evaluation*, 70(1), pp.78–85.

Gryllias, K.C. & Antoniadis, I. a., 2012. A Support Vector Machine approach based

on physical model training for rolling element bearing fault detection in industrial environments. *Engineering Applications of Artificial Intelligence*, 25(2), pp.326–344.

Guo, L., Chen, J. & Li, Xi., 2009. Rolling Bearing Fault Classification Based on Envelope Spectrum and Support Vector Machine. *Journal of Vibration and Control*, 15(9), pp.1349–1363.

Gupta, P.K., 1975. Transient ball motion and skid in ball bearings. *Transactions of the ASME, Journal of lubrication technology*, pp.261–269.

Halme, J. & Andersson, P., 2009. Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics-state of the art. *Journal of Engineering Tribology*, 224, pp.377–393.

Hannes, D. & Alfredsson, B., 2011. Rolling contact fatigue crack path prediction by the asperity point load mechanism. *Engineering Fracture Mechanics*, 78, pp.2848–2869.

Hao, R. & Chu, F., 2009. Morphological undecimated wavelet decomposition for fault diagnostics of rolling element bearings. *Journal of Sound and Vibration*, 320, pp.1164–1177.

Harris, F.J., 1991. *Rolling bearing analysis* Third edit., New York: John Wiley & Sons, Inc.

Harris, F.J., 1966. *Rolling bearing analysis I* First., New York: John Wiley & Sons, Inc.

Harsha, S.P., Sandeep, K. & Prakash, R., 2004. Non-linear dynamic behaviors of rolling element bearings due to surface waviness. *Journal of Sound and Vibration*, 272(3-5), pp.557–580.

Harvey, T.J., Wood, R.J.K. & Powrie, H.E.G., 2007. Electrostatic wear monitoring of rolling element bearings. *Wear*, 263, pp.1492–1501.

Heng, A. et al., 2009. Rotating machinery prognostics: State of the art, challenges and opportunities. *Mechanical Systems and Signal Processing*, 23(3), pp.724–739.

Ho, D., 2000. *Bearing diagnostics and self adaptive noise cancellation*. UNSW.

Holmberg, K. et al., 2005. Tribological analysis of fracture conditions in thin surface coatings by 3D FEM modelling and stress simulations. *Tribology International*, 38(11-12), pp.1035–1049.

Holttinen, H., 2005. Impact of hourly wind power variations on the system operation in the Nordic countries. *Wind Energy*, 8(2), pp.197–218.

Howard, I., 1994. *A review of rolling element bearing vibration "detection, diagnosis and prognosis,"* Melbourne, Australia.

Hu, Q. et al., 2007. Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, 21(2), pp.688–705.

Irvine, T., 2014. *Ring vibration modes*, Available at: http://www.vibrationdata.com/tutorials2/ringmode.pdf.

Jack, L.B. & Nandi, A.K., 2002. Fault Detection Using Support Vector Machines and Artificial Neural Networks, Augmented By Genetic Algorithms. *Mechanical Systems and Signal Processing*, 16(2-3), pp.373–390.

Jaing, Y. & Sehitoglu, H., 1996. Rolling contact stress analysis with the application of a new plasticity model. *Wear*, 191, pp.35–44.

Jammu, N.S. & Kankar, P.K., 2011. A Review on Prognosis of Rolling Element Bearings. *Inrernational Journal of engineering science and technology*, 3(10), pp.7497–7503.

Jang, G.H. & Jeong, S.W., 2002. Nonlinear Excitation Model of Ball Bearing Waviness in a Rigid Rotor Supported by Two or More Ball Bearings Considering Five Degrees of Freedom. *Journal of Tribology*, 124, pp.82–90.

Jantunen, E., 2006. How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods. *International Journal of COMADEM*, 9(3), pp.24–38.

Jantunen, E., 2004. Intelligent Monitoring and Prognosis of Degradation of Rotating Machinery. In *Proceedings of the intelligent maintenance systems IMS'2004*.

Jardine, A.K.S., Lin, D. & Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), pp.1483–1510.

Junsheng, C., Dejie, Y. & Yu, Y., 2007. Application of an impulse response wavelet to fault diagnosis of rolling bearings. *Mechanical Systems and Signal Processing*, 21, pp.920–929.

Kappa, B., 2006. *Predicting bearing failures and measuring lubrication film thickness in your plants rotating equipment*,

Karmakar, S., Rao, U.R.K. & Sethuramiah, A., 1996. An approach towards fatigue wear modelling. *Wear*, 198, pp.242–250.

Kilundu, B. et al., 2011. Cyclostationarity of Acoustic Emissions (AE) for monitoring bearing defects. *Mechanical Systems and Signal Processing*, 25(6), pp.2061–2072.

Kim, H. et al., 2013. Multiscale contact mechanics model for RF–MEMS switches with quantified uncertainties. *Modelling and Simulation in Materials Science and Engineering*, 21(8).

Kim, T.H. & Olver, A.V., 1998. Stress history in rolling-sliding contact of rough surfaces. *Tribology International*, 31(12), pp.727–736.

Kiral, Z. & Karagülle, H., 2003. Simulation and analysis of vibration signals generated by rolling element bearing with defects. *Tribology International*, 36(9), pp.667–678.

Kotzalas, M.N. & Doll, G.L., 2010. Tribological advancements for reliable wind turbine performance. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, 368, pp.4829–4850.

Larson, E.C., Wipf, D.P. & Parker, B.E., 1997. Gear and bearing diagnostics using neural network-based amplitude and phase demodulation. In *Proceedings of the 51st meeting of the Society for Machinery Prevention Technology*. Virginia, pp. 511–521.

Lee, J. et al., 2006. Intelligent prognostics tools and e-maintenance. *Computers in Industry*, 57(6), pp.476–489.

Leonel, E.D. & Venturini, W.S., 2011. Multiple random crack propagation using a

boundary element formulation. *Engineering Fracture Mechanics*, 78, pp.1077–1090.

Li, B. et al., 2000. Neural-Network-Based motor rolling bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, 47(5), pp.1060–1069.

Li, C.J. & Ma, J., 1997. Wavelet decomposition of vibrations for detection of bearing-localized defects. *ND*, 30(3), pp.143–149.

Li, X. et al., 2013. Rolling element bearing fault detection using support vector machine with improved ant colony optimization. *Measurement*, 46, pp.2726–2734.

Li, Y. et al., 1999. Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing*, 13(1), pp.103–113.

Li, Y., Kurfess, T. & Liang, Y., 2000. Stochastic prognostics for rolling element bearings. *Mechanical Systems and Signal Processing*, 14(5), pp.747–762.

Lin, J. & Qu, L., 2000. Feature Extraction Based on Morlet Wavelet and Its Application for Mechanical Fault Diagnosis. *Journal of Sound and Vibration*, 234, pp.135–148.

Liu, B., Ling, S.-F. & Meng, Q., 1997. Machinery Diagnosis Based on Wavelet Packets. *Journal of Vibration and Control January*, 3, pp.5–17.

Liu, J. et al., 2008. Wavelet spectrum analysis for bearing fault diagnostics. *Measurement Science and Technology*, 19.

Liu, J., Shao, Y. & Lim, T.C., 2012. Vibration analysis of ball bearings with a localized defect applying piecewise response function. *Mechanism and Machine Theory*, 56, pp.156–169.

Liu, T.I., Singonahalli, J.H. & Iyer, N.R., 1996. Detection of Roller Bearing Defects Using Expert System and Fuzzy Logic. *Mechanical Systems and Signal Processing*, 10, pp.595–614.

Liu, Y., Liu, L. & Mahadevan, S., 2007. Analysis of subsurface crack propagation under rolling contact loading in railroad wheels using FEM. *Engineering Fracture Mechanics*, 74, pp.2659–2674.

Liu, Z. et al., 2013. Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings. *Neurocomputing*, 99, pp.399–410.

Logan, D. & Mathew, J., 1996. Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings—I. Basic Concepts. *Mechanical Systems and Signal Processing*, 10(3), pp.241–250.

Logan, D.B. & Mathew, J., 1996. Using the Correlation Dimension for Vibration Fault Diagnosis of Rolling Element Bearings—II. Selection of Experimental Parameters. *Mechanical Systems and Signal Processing*, 10(3), pp.251–264.

Malhi, A.S., 2002. *Finite element modelling of vibrations caused by a defect in the outer ring of a ball bearing*, University of Massachusetts, Amherst.

Manoj, V., Manohar Shenoy, K. & Gopinath, K., 2008. Developmental studies on rolling contact fatigue test rig. *Wear*, 264, pp.708–718.

Marble, S. & Morton, B.P., 2006. Predicting the Remaining Life of Propulsion System Bearings. In *Proceedings of the 2006 IEEE Aerospace Conference*. Big

sky, MT: IEEE.

Maru, M.M., Castillo, R.S. & Padovese, L.R., 2007. Study of solid contamination in ball bearings through vibration and wear analyses. *Tribology International*, 40(3), pp.433–440.

Masen, M.A., de Rooij, M.B. & Schipper, D.J., 2005. Micro-contact based modelling of abrasive wear. *Wear*, 258, pp.339–348.

Massi, F. et al., 2010. Coupling system dynamics and contact behaviour: Modelling bearings subjected to environmental induced vibrations and "false brinelling" degradation. *Mechanical Systems and Signal Processing*, 24(4), pp.1068–1080.

Massouros, G.P., 1983. Normal vibration of a plain bearing working under boundary lubrication conditions. *Tribology International*, 16(5), pp.235–238.

McCormick, A.C. & Nandi, A.K., 1998. Cyclostationarity in Rotating Machine Vibrations. *Mechanical Systems and Signal Processing*, 12(2), pp.225–242.

McFadden, P.D., 1987. A revised model for the extraction of periodic waveforms by time domain averaging. *Mechanical Systems and Signal Processing*, 1(1), pp.83–95.

McFadden, P.D. & Smith, J.D., 1984. Model for the vibration produced by a single point defect in a rolling element bearing. *Journal of Sound and Vibration*, 96(1), pp.69–82.

McFadden, P.D. & Smith, J.D., 1985. The vibration produced by multiple point defects in a rolling element bearing. *Journal of Sound and Vibration*, 98(2), pp.263–273.

Mechefske, C.K., 1998. Objective Machinery Fault Diagnosis Using Fuzzy Logic. *Mechanical Systems and Signal Processing*, 12, pp.855–862.

Miettinen, J., 2000. The Influence of the Running Parameters on the Acoustic Emission of Grease Lubricated Rolling Bearings. *Maintenance & Asset Management*.

Miller, A.J., 1999. *A new wavelet basis for the decomposition of gear motion error signals and its application to gearbox diagnostics*. The Pennsylvania State University, State College, PA,.

Momono, T. & Noda, B., 1999. Sound and Vibration in rolling bearings. *Motion & Control*, 6, pp.29–37.

Morales-Espejel, G.E. & Brizmer, V., 2011. Micropitting Modelling in Rolling–Sliding Contacts: Application to Rolling Bearings. *Tribology Transactions*, 54(4), pp.625–643.

Mori, K. et al., 1996. Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals. *Wear*, 195, pp.162–168.

Mota, V. & Ferreira, L.A., 2009. Influence of grease composition on rolling contact wear: Experimental study. *Tribology International*, 42, pp.569–574.

Mota, V., Moreira, P. & Ferreira, L., 2008. A study on the effects of dented surfaces on rolling contact fatigue. *International Journal of Fatigue*, 30, pp.1997–2008.

Nakhaeinejad, M., 2010. *Fault Detection and Model-Based Diagnostics in Nonlinear Dynamic Systems*. University of Texas, Austin.

Nakhaeinejad, M. & Bryant, M.J., 2011. Dynamic modelling of rolling element bearings with surface contact defects using bond graphs. *Journal of Tribology*, 133.

Navarro, A. & Rios, E.R., 1988. Short and long fatigue crack growth – a unified model. *Philos Mag*, 57, pp.15–36.

Nelwamondo, F. V., Marwala, T. & Mahola, U., 2006. Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, Mel-Frequency cepstral coefficients and fractals. *International Journal of Innovative Computing, Information and Control*, 2(6), pp.1281–1299.

NSK, 2013. General miscellaneous information. In *NSK technical report No. E728g*. pp. 240–267.

Ocak, H., Loparo, K. a. & Discenzo, F.M., 2007. Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics. *Journal of Sound and Vibration*, 302, pp.951–961.

Pachaud, C., Salvetas, R. & Fray, C., 1997. Crest factor and kurtosis contributions to identify defects inducing periodical impulsive forces. *Mechanical Systems and Signal Processing*, 11(6), pp.903–916.

Palmgren, A., 1947. *Ball and roller bearing engineering*, Philadelphia: S.H.Burbank and Co. Inc.

Pan, Y., Chen, J. & Guo, L., 2009. Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description. *Mechanical Systems and Signal Processing*, 23, pp.669–681.

Patil, M.S. et al., 2010. A theoretical model to predict the effect of localized defect on vibrations associated with ball bearing. *International Journal of Mechanical Sciences*, 52, pp.1193–1201.

Paya, B.A., Esat, I.I. & Badi, M.N.M., 1997. Artifical neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor. *Mechanical Systems and Signal Processing*, 11(5), pp.751–765.

Peng, Y., Dong, M. & Zuo, M.J., 2010. Current status of machine prognostics in condition-based maintenance: a review. *The International Journal of Advanced Manufacturing Technology*, 50(1-4), pp.297–313.

Peng, Z. & Kessissoglou, N., 2003. An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis. *Wear*, 255(7-12), pp.1221–1232.

Peng, Z.K. & Chu, F.L., 2004. Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography. *Mechanical Systems and Signal Processing*, 18, pp.199–221.

Peng, Z.K., Chu, F.L. & Tse, P.W., 2007. Singularity analysis of the vibration signals by means of wavelet modulus maximal method. *Mechanical Systems and Signal Processing*, 21(2), pp.780–794.

Peng, Z.K., Tse, P.W. & Chu, F.L., 2005. A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, 19, pp.974–988.

Pineyro, J., Klempnow, A. & Lescano, V., 2000. Effectiveness of new spectral

tools in the anomaly detection of rolling element bearings. *Journal of Alloys and Compounds*, 310, pp.276–279.

Purohit, R.K. & Purohit, K., 2006. Dynamic analysis of ball bearings with effect of preload and number of balls. *International journal of applied machines and engineering*, 11(1), pp.77–91.

Qiu, H. et al., 2003. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Advanced Engineering Informatics*, 17, pp.127–140.

Qiu, J. et al., 2002. Damage Mechanics Approach for Bearing Lifetime Prognostics. *Mechanical Systems and Signal Processing*, 16, pp.817–829.

Rafsanjani, A. et al., 2009. Nonlinear dynamic modeling of surface defects in rolling element bearing systems. *Journal of Sound and Vibration*, 319, pp.1150–1174.

Raje, N. et al., 2008. A Numerical Model for Life Scatter in Rolling Element Bearings. *Journal of Tribology*, 130.

Randall, R.B., 2011. *Vibration-based Condition Monitoring*, Chichester, UK: John Wiley & Sons, Ltd.

Randall, R.B. & Antoni, J., 2011. Rolling element bearing diagnostics—A tutorial. *Mechanical Systems and Signal Processing*, 25(2), pp.485–520.

Randall, R.B., Antoni, J. & Chobsaard, S., 2001. The Relationship Between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals. *Mechanical Systems and Signal Processing*, 15(5), pp.945–962.

Ribrant, J. & Bertling, L., 2007. Survey of failures in wind power systems with focus on Swedish wind power plants during 1997-2005". In *IEEE Power Engineering Society , General Meeting*. pp. 1–8.

Richard Liu, C. & Choi, Y., 2008. A new methodology for predicting crack initiation life for rolling contact fatigue based on dislocation and crack propagation. *International Journal of Mechanical Sciences*, 50, pp.117–123.

Ringsberg, J.W. & Bergkvist, A., 2003. On propagation of short rolling contact fatigue cracks. *Fatigue fracture engineering material structure*, 26, pp.969–983.

Roemer, M.J., Hong, C. & Hesler, S.H., 1996. Machine health monitoring and life management using finite-element-based neural networks. *Journal of engineering for gas turbines and power*, 118, pp.830–835. Available at: http://cat.inist.fr/?aModele=afficheN&cpsidt=3260060 [Accessed June 20, 2012].

Sadeghi, F. et al., 2009. A Review of Rolling Contact Fatigue. *Journal of Tribology*, 131(4).

Sahoo, P. & Banerjee, A., 2005. Asperity interaction in elastic–plastic contact of rough surfaces in presence of adhesion. *Journal of Physics D: Applied Physics*, 38(16), pp.2841–2847.

Samanta, B. & Al-Balushi, K.R., 2003. Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features. *Mechanical Systems and Signal Processing*, 17, pp.317–328.

Santus, C. et al., 2012. Surface and subsurface rolling contact fatigue characteristic depths and proposal of stress indexes. *International Journal of Fatigue*, 45, pp.71–81.

Sassi, S., Badri, B. & Thomas, M., 2007. A Numerical Model to Predict Damaged Bearing Vibrations. *Journal of Vibration and Control*, 13(11), pp.1603–1628.

Sassi, S., Badri, B. & Thomas, M., 2006. TALAF and THIKAT as innovative time domain indicators for tracking BALL bearings. In *CMVA -PROCEEDINGS-; 31 Machinery vibration Seminar; 24th, Machinery vibration*. pp. 404–419.

Sawalhi, N., 2007. *Diagnostics, Prognostics and fault simulation for rolling element bearings*. UNSW, Sydney.

Sawalhi, N. & Randall, R.B., 2008. Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults. *Mechanical Systems and Signal Processing*, 22(8), pp.1924–1951.

Sawalhi, N. & Randall, R.B., 2005. Spectral kurtosis optimization for rolling element bearings. In *Proceedings of the ISSPA conference*. Sydney, Australia.

Sawalhi, N. & Randall, R.B., 2011. Vibration response of spalled rolling element bearings: Observations, simulations and signal processing techniques to track the spall size. *Mechanical Systems and Signal Processing*, 25(3), pp.846–870.

Sawalhi, N., Randall, R.B. & Endo, H., 2007. The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. *Mechanical Systems and Signal Processing*, 21, pp.2616–2633.

Schwabacher, M. & Goebel, K., 2006. A Survey of Artificial Intelligence for Prognostics. *NASA Ames Research Center*, pp.107–114.

Schwabacher, M.A., 2005. A Survey of Data-Driven Prognostics. , (September), pp.1–5.

Schwach, D. & Guo, Y., 2006. A fundamental study on the impact of surface integrity by hard turning on rolling contact fatigue. *International Journal of Fatigue*, 28, pp.1838–1844.

Serrato, R., Maru, M.M. & Padovese, L.R., 2007. Effect of lubricant viscosity grade on mechanical vibration of roller bearings. *Tribology International*, 40, pp.1270–1275.

Shibata, K., Takahashi, A. & Shirai, T., 2000. Fault Diagnosis of Rotating Machinery Through Visualisation of Sound Signals. *Mechanical Systems and Signal Processing*, 14(2), pp.229–241.

Slack, T. & Sadeghi, F., 2011. Cohesive zone modeling of intergranular fatigue damage in rolling contacts. *Tribology International*, 44, pp.797–804.

Slack, T. & Sadeghi, F., 2010. Explicit finite element modeling of subsurface initiated spalling in rolling contacts. *Tribology International*, 43, pp.1693–1702.

Söffker, D., Saadawia, M. & Wei, C., 2013. Model- and feature-based diagnostics in rotating machinery. In M. A. Savi, ed. *Proceedings of the XV International Symposium on Dynamic Problems of Mechanics*. ABCM, Buzios, RJ, Brazil, February 17-22, 2013.

Sopanen, J. & Mikkola, A., 2003. Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: theory. *Proceedings of the Institution of Mechanical Engineers, Part K: Journal of Multi-body Dynamics*, 217(3), pp.201–211.

Spoerre, J.K., 1997. Application of the cascade correlation algorithm (CCA) to bearing fault classification problems. *Computers in Industry*, 32, pp.295–304.

STI, 2012. Field Application Note: Rolling Element Bearings. Available at: http://www.stiweb.com/appnotes/Rolling-Element-Bearings.html [Accessed February 10, 2015].

Su, W. et al., 2010. Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement. *Mechanical Systems and Signal Processing*, 24, pp.1458–1472.

Sun, Q. & Tang, Y., 2002. Singularity Analysis Using Continuous Wavelet Transform for Bearing Fault Diagnosis. *Mechanical Systems and Signal Processing*, 16, pp.1025–1041.

Sun, Z., Rios, E.R. & Miller, K.J., 1991. Modelling small fatigue cracks interacting with grain boundaries. *Fatigue Fracture Engineering Meterlogy*, 14, pp.277–291.

Sunnersjö, C.S., 1985. Rolling bearing vibrations- the effects of geometrical imperfections and wear. *Journal of Sound and Vibration*, 98(4), pp.455–474.

Tadina, M. & Boltezar, M., 2011. Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up. *Journal of Sound and Vibration*, 300(17), pp.4287–4301.

Tallam, R.M., Habetler, T.G. & Harley, R.G., 2002. Self-commissioning training algorithms for neural networks with applications to electric machine fault diagnostics. *IEEE Transactions on Power Electronics*, 17(6), pp.1089–1095.

Tan, C.C. & Dawson, B., 1987. An adaptive noise cancellation approach for condition monitoring of gearbox bearings. In *Processdings of the international tribology conference*. Melbourne.

Tandon, N. & Choudhury, A., 1999. A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings. *Tribiology International*, 32, pp.469–480.

Tandon, N. & Choudhury, A., 1997. An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect. *Journal of Sound and Vibration*, 205(3), pp.275–292.

Tarantino, M.G. et al., 2013. Experiments under pure shear and rolling contact fatigue conditions: Competition between tensile and shear mode crack growth. *International Journal of Fatigue*, 46, pp.67–80.

Tavner, P., Spinato, F. & van Bussel, G.J.W., 2008. Reliability of Different Wind Turbine Concepts with Relevance to Offshore Application. In *European Wind Energy Conference (EWEC 2008)*.

Tiwari, M., Gupta, K. & Prakash, O., 2000a. Dynamic response of an unbalanced rotor supported on ball bearings. *Journal of Sound and Vibration*, 238, pp.757–779.

Tiwari, M., Gupta, K. & Prakash, O., 2000b. Effect of radial internal clearance of ball bearing on the dynamics of a balanced horizontal rotor. *Journal of Sound and Vibration*, 238, pp.723–756.

Tiwari, R. & Vyas, N.S., 1995. Dynamic response of an unbalanced rotor supported on ball bearings. *Journal of Sound and Vibration*, 187(2), pp.229–239.

Tse, P.W., Peng, Y.H. & Yam, R., 2001. Wavelet Analysis and Envelope Detection For Rolling Element Bearing Fault Diagnosis—Their Effectiveness and Flexibilities. *Journal of Vibration and Acoustics*, 123(3), p.303.

Tsushima, N., 2007. Crack Propagation of Rolling Contact Fatigue in Ball Bearing Steel Due to Tensile Strain. *NTN Technical review*, (75), pp.128–138.

Tyagi, S., 2008. A Comparative Study of SVM Classifiers and Artificial Neural Networks Application for Rolling Element Bearing Fault Diagnosis using Wavelet Transform Preprocessing. *World Academy of Science, Engineering and Technology*, 43, pp.309–317.

Vania, A. & Pennacchi, P., 2004. Experimental and theoretical application of fault identification measures of accuracy in rotating machine diagnostics. *Mechanical Systems and Signal Processing*, 18, pp.329–352.

Vlok, P.-J., Wnek, M. & Zygmunt, M., 2004. Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical Systems and Signal Processing*, 18, pp.833–847.

Wang, C. & Too, G., 2002. Rotating machine fault detection based on HOS and artificial neural networks. *Journal of Intelligent Manufacturing*, 13, pp.283–293.

Wang, W.Q., Golnaraghi, M.F. & Ismail, F., 2004. Prognosis of machine health condition using neuro-fuzzy systems. *Mechanical Systems and Signal Processing*, 18, pp.813–831.

Wang, Y. & Hadfield, M., 1999. Rolling contact fatigue failure modes of lubricated silicon nitride in relation to ring crack defects. *Wear*, 225-229, pp.1284–1292.

Wang, Y.-F. & Kootsookos, P.J., 1998. Modeling of Low Shaft Speed Bearing Faults for Condition Monitoring. *Mechanical Systems and Signal Processing*, 12(3), pp.415–426.

Warhadpande, A. et al., 2012. Effects of plasticity on subsurface initiated spalling in rolling contact fatigue. *International Journal of Fatigue*, 36, pp.80–95.

Wei, W., Qiang, L. & Guojie, Z., 2008. Novel approach based on chaotic oscillator for machinery fault diagnosis. *Measurement*, 41(8), pp.904–911.

Weinzapfel, N. & Sadeghi, F., 2012. Numerical modeling of sub-surface initiated spalling in rolling contacts. *Tribology International*.

Wijnat, Y.H., Wensing, J.A. & van Nijen, G.C., 1999. The influence of lubrication on the dynamic behaviour of ball bearings. *Journal of Sound and Vibration*, 222(4), pp.579–596.

Xu, G. & Sadeghi, F., 1996. Spall initiation and propagation due to debris denting. *Wear*, 201, pp.106–116.

Yam, R.C.M. et al., 2001. Intelligent Predictive Decision Support System for Con-

dition-Based Maintenance. *The International Journal of Advanced Manufacturing Technology*, 17, pp.383–391.

Yan, J., Koç, M. & Lee, J., 2004. A prognostic algorithm for machine performance assessment and its application. *Production Planning & Control*, 15(8), pp.796–801.

Yan, R. & Gao, R.X., 2007. Approximate Entropy as a diagnostic tool for machine health monitoring. *Mechanical Systems and Signal Processing*, 21, pp.824–839.

Yan, R. & Gao, R.X., 2004. Complexity as a Measure for Machine Health Evaluation. *IEEE Transactions on instrumentation and measurement*, 53(4), pp.1327–1334.

Yang, B., Lim, D. & Tan, A., 2005. VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table. *Expert Systems with Applications*, 28, pp.735–742.

Yang, B.-S., Han, T. & Kim, Y.-S., 2004. Integration of ART-Kohonen neural network and case-based reasoning for intelligent fault diagnosis. *Expert Systems with Applications*, 26, pp.387–395.

Yang, B.-S. & Widodo, A., 2008. Support Vector Machine for Machine Fault Diagnosis and Prognosis. *Journal of System Design and Dynamics*, 2(1), pp.12–23.

Yang, Y., Yu, D. & Cheng, J., 2007. A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM. *Measurement*, 40, pp.943–950.

Yoshioka, T. & Shimizu, S., 2009. Monitoring of Ball Bearing Operation under Grease Lubrication Using a New Compound Diagnostic System Detecting Vibration and Acoustic Emission. *Tribology Transactions*, 52(6), pp.725–730.

Zhang, S. et al., 2003. Fault Diagnosis System for Rotary Machines Based on Fuzzy Neural Networks. *JSME International journal. Series C: Mechanical Systems, Machine Elements and Manufacturing*, 46, pp.1035–1041.

Zhang, X. & Kang, J., 2010. Hidden Markov Models in bearing fault diagnosis and prognosis. In *Second International Conference on Computational Intelligence and Natural Computing (CINC)*. pp. 364–367.

Zhi-qiang, Z. et al., 2012. Investigation of rolling contact fatigue damage process of the coating by acoustics emission and vibration signals. *Tribology International*, 47, pp.25–31.

A summary of fault modelling and predictive health monitoring of rolling element bearings

Review

A summary of fault modelling and predictive health monitoring of rolling element bearings

Idriss El-Thalji *, Erkki Jantunen

Industrial Systems, VTT Technical Research Centre of Finland, Espoo, Finland

ABSTRACT

The rolling element bearing is one of the most critical components that determine the machinery health and its remaining lifetime in modern production machinery. Robust Predictive Health Monitoring tools are needed to guarantee the healthy state of rolling element bearing s during the operation. A Predictive Health Monitoring tool indicates the upcoming failures which provide sufficient lead time for maintenance planning. The Predictive Health Monitoring tool aims to monitor the deterioration i.e. wear evolution rather than just detecting the defects. The Predictive Health Monitoring procedures contain detection, diagnosis and prognosis analysis, which are required to extract the features related to the faulty rolling element bearing and estimate the remaining useful lifetime. The purpose of this study is to review the Predictive Health Monitoring methods and explore their capabilities, advantages and disadvantage in monitoring rolling element bearings. Therefore, the study provides a critical review of the Predictive Health Monitoring methods of the entire defect evolution process i.e. over the whole lifetime and suggests enhancements for rolling element bearing monitoring.

© 2015 Elsevier Ltd. All rights reserved.

Contents

* Corresponding author.
  E-mail addresses: idriss.el-thalji@vtt.fi (I. El-Thalji), erkki.jantunen@vtt.fi (E. Jantunen).

1. Introduction

The rolling element bearing (REB) is one of the most critical components that determine the machinery health and its remaining lifetime in modern production machinery. Robust Predictive Health Monitoring (PHM) tools are needed to guarantee the healthy state of REBs during the operation. The PHM tool indicates the upcoming failures and provides more time for maintenance planning. The PHM tool aims to monitor the deterioration i.e. wear evolution rather than just detecting the defects. There are a number of literature reviews which are related to the condition monitoring of REBs [1–6]. These reviews explain very well the developed signal processing (SP), diagnosis and prognosis analysis methods and their challenges, enhancements, and limitations. Many experiments and studies have been performed to explore the nature of bearing defects with the help of several monitoring techniques such as vibration, AE, oil-debris, ultrasound, electrostatic, Shock-Pulse Measurements (SPM), etc. Some simple signal/data processing techniques have been applied to process the signals such as root mean square (RMS), kurtosis, Fast Fourier Transform (FFT), etc. However, there are several challenges that require more advanced SP methods, e.g. to remove the background noise effect, the smearing effect and the speed fluctuation effect. The most important challenge is to deal with the signal response due to defected REBs. Bearing faults are assumed to generate impulses due to the passing of the rolling element over the defected surface. The difficulty is to detect and track such impulses, specially, in the early stage of wear process where the defect is quite small and can be easily buried by other vibration phenomena. Therefore, most of the PHM studies have concentrated to the development of more advanced SP techniques such as envelope detection, cyclostationary analysis, wavelets, data-driven methods, expert systems, fuzzy logic techniques, etc.

In the field of machinery vibration monitoring and analysis, a variety of relevant standards are developed and published by ISO (International Organization for Standardization). A number of ISO standards describe acceptable vibration limits, such as the ISO/7919 series (5 parts) "Mechanical vibration of non-reciprocating machines—Measurements on rotating shafts and evaluation criteria" and the ISO/10816 series (6 parts) "Mechanical vibration—Evaluation of machine vibration by measurements on non-rotating parts". The standards cover the methods of measurement, handling, and processing of the data required to perform condition monitoring and diagnostics of machines. In the industry, the most commonly used techniques are RMS, crest factor, probability density functions, correlation functions, band pass filtering prior to analysis, power and cross power spectral density functions, transfer and coherence functions as well as Cepstrum analysis, narrow band envelope analysis and shock pulse method. These methods try to extract the expected defect features. The frequency equations of the bearing defects (i.e. for outer-race, inner-race, rolling elements, and cage defects) are the main way to provide a theoretical estimate of the frequencies to be expected when various defects occur in the REB. They are based upon the assumption that sharp force impacts will be generated whenever a bearing element encounters a localized bearing fault such as caused by spalling and pitting. These techniques have continued to be used and have been further developed over the time [1].

The ultimate purpose of the PHM system is to indicate the upcoming failures in order to provide sufficient lead time for maintenance planning. Beside the experimental studies, there are several analytical and numerical models to (1) simulate the faulty REBs; (2) verify the ability of SP and diagnosis methods to extract the defect features; and (3) to predict the remaining useful lifetime of the faulty REBs. Several studies have explored the data-driven and model-based prognosis methods for REBs applications. Therefore, the purpose of this study is to review and discuss the entire PHM procedures i.e. detection, diagnosis and prognosis based on experimental studies and simulation models that have been made available in the literature.

The study begins with presenting the fundamentals of rolling bearing and their modelling techniques. Then, the monitoring techniques, SP, diagnostic methods and prognosis analysis for REB are reviewed. Later, all these issues are critically discussed in order to draw some the conclusions of current research, emerging trends and the areas where more work and research is needed.
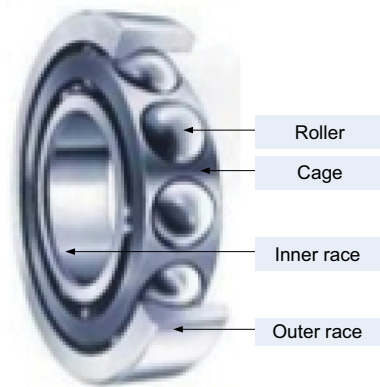
Fig. 1. Elements of rolling bearing.

2. Bearing faults and dynamic simulation methods

Rolling bearing is a mechanical component which carries loads and eliminates the sliding friction by placing rolling elements i.e. balls or rollers between two bearing rings i.e. outer and inner raceway. Depending on the internal design, rolling bearings may be classified as radial bearings i.e. carry radial loads or thrust bearings i.e. carry axial loads. Practically all rolling bearings consist of four basic parts: inner ring, outer ring, rolling elements, and cage, as illustrated in Fig. 1.

The bearings faults may be classified based on their locations as outer, inner, rolling element and cage fault. The general reason behind these faults is the rolling contact stresses that might increase due to the increase of operating loads, additional loads due to faults i.e. imbalance, misalignment, bent shaft, looseness, and/or distributed defects i.e. high degree of surface roughness and waviness, contamination, and inclusions. Therefore, some topological changes might occur. These topological changes in the contact area generate stress concentration points and lubrication film disturbances and lead to the wear evolution process. El-Thalji and Jantunen [7] reviewed the most relevant studies and experimental findings in order to describe the wear evolution process over the lifetime for the rolling bearings. In summary, the wear evolution process is quite complex due to the involvement of several wear mechanisms (i.e. fatigue, abrasive, adhesive, and corrosive) and several stress concentration mechanisms (i.e. dent, asperities, debris, and sub-surface inclusions). These involvements and their interactions and competitions produce a wear evolution progress which varies significantly with respect to the surface topographical and tribological changes. As the fault topography is changing over the lifetime, the fault features are changing over the lifetime. In this sense, the fault topology that is assumed in the simulation models should also be changing. Moreover, there is a need to clearly determine the fault features of specific wear evolution stages and to understand how different signal analysis methods cope with such features, in order to effectively track the detected fault features. Therefore, the review starts with the simulation models, followed by the complete predictive monitoring procedures i.e. monitoring, signal analysis, diagnosis, and prognosis.

2.1. Dynamic simulation models

Over the years, several dynamic models have been developed to investigate the dynamic behaviour and features of REBs, as shown in Table 1. The dynamic models of REB were first introduced by Palmgren [8] and Harris [9]. However, total non-linearity and time varying characteristics were not addressed at that time. After that Gupta [3,10] provided the first complete dynamic model of REB and later Fukata et al. [11] presented a comprehensive non-linear and time variant model. The more advanced issues of time variant characteristics and non-linearity were raised and studied by several authors. For example, Wijnat et al. [12] reviewed the studies concerning the effect of the Elasto-Hydrodynamic Lubrication (EHL) on the dynamics of REB. Tiwari and Vyas [13] and Tiwari et al. in [14] and [15] studied the effect of the ball bearing clearance on the dynamic response of a rigid rotor. Sopanen and Mikkola [16] reviewed different dynamic models with the discussion of the effect of waviness, EHL, and localised faults and clearance effect. Later, the finite element method (FEM) was used to provide more accurate results. Kiral and Karagülle [17] presented a defect detection method using FEM vibration analysis for REBs with single and multiple defects. The vibration signal includes impulses produced by the fault, modulation effect due to non-uniform load distribution, bearing induced vibrations, and machinery induced vibrations and the noise which is encountered in any measurement system. Sopanen and Mikkola [16] implemented the proposed ball bearing model using a commercial multi-body system software application MSC.ADAMS. First, the FEM model was utilized to simulate the variation of the mesh stiffness for two types of faults under varying static load conditions. Then the model was integrated into the lumped parameter dynamic model. The study obtained the dynamic transmission error and acceleration responses under different loads and speeds. Sawalhi and Randall [18] developed a 34-DOF model of a gearbox in order to simulate spalling and cracks in the REB based on Endo's model of 16-DOF. This model includes additional 18-DOF due to the consideration of a five DOF bearing model and the consideration of translational DOF both along the line of action and perpendicular to it. Massi et al. [19] studied the wear that is resulting from false Brinelling at the contact surfaces between the balls and races of the bearings.

Table 1
Summary of dynamic models for bearing faults.

Reference	Bearing contact	Clearances	EHL contact	Distributed defect	Localised defect
[26]	x				x
[13,14]	x				
[15]	x	x			
[23]	x			X	
[16]	x	x	x	X	x
[27]	x				
[20]	x	x			
[28]	x				x
[18,29]	x		x		x
[21]	x	x		X	x
[30–35]	x				x
[22]	x	x			x

Table 2
Dynamic models of localised defect.

Reference	System dynamic modelling		Fault dynamic modelling		
	Analytical dynamic	FEM	Geometrical defect function	Force defect function	defect function
[24–26]	X				X
[17,37,38]		X		X	
[16,27]	X			X	
[18]		X			X
[21,22,28,30,32,33]	X		X		
[19]	X	X		X	
[34]	X			X	

Several models have been developed to study the effects of distributed and localized defects on REB dynamics: clearance effect [15,16,20–22], waviness effect [16,21,23], disturbances effect of EHL [16,18], and the effect of localized faults [24–26], etc.

2.2. Dynamic models of localised defect

The largest share of the studies has focused on the localized faults using different modelling techniques, as shown in Table 2. McFadden and Smith [24], McFadden and Smith [25], Tandon and Choudhury [26] and Sawalhi and Randall [18] simulated the defect as a signal function of impulsive train in the modelled system. For example, Tandon and Choudhury [26] have introduced the defect as pulse function with three different pulse shapes: rectangular, triangular and half-sine pulse. Wang and Kootsookos [171] introduced defects as a function of a basic impulse series. Ghafari et al. [36] have virtually introduced a defect into the equation of motion as a triangular impulse train at the related characteristic frequencies of a defect. Rafsanjani et al. [31] modelled the localized defects as a series of impulses having a repetition rate equal to the characteristics frequencies. The amplitude of the generated impulses is related to the loading and angular velocity at the point of contact. Malhi [37], Kiral and Karagülle [17], Sopanen and Mikkola [16], Massi et al. [19], and Liu et al. [34] introduced the defect as force function in their FEM models i.e. as a constant impact factor. More precisely Liu et al. [25] introduced the localized defect as a piecewise function.

Ashtekar and Sadeghi [30], Sassi et al. [28], Cao and Xiao [21], Rafsanjani et al. [31], Patil et al. [32], and Tadina and Boltezar [33] modelled the defect based on its geometrical features i.e. as a surface bump or a dent that has length, width and depth. Tadina and Boltezar [33] modelled the defect as an impressed ellipsoid on the races and as flattened sphere for the rolling elements. Nakhaeinejad [22] utilised the bond graphs to study the effects of defects on bearing vibrations. The model incorporated gyroscopic and centrifugal effects, contact deflections and forces, contact slip and separations, and localized faults. Dents and pits on inner race, outer race, and balls were modelled through surface profile changes i.e. type, size and shape of the localized faults. The main difficulty with the use of complex dynamic models lies in experimentally verifying the predicted results [1].

3. Monitoring methods

Several experiments have been conducted to study specific monitoring techniques such as vibration, acoustic emission (AE), oil-debris, ultrasound, electrostatic, SPM, etc. and their use in faulty REBs detection. Many studies have used simple signal/data processing techniques such as RMS, kurtosis, FFT, etc. However, the largest share of studies has focused to develop

Table 3
Wear monitoring techniques for rolling bearings.

Reference	Test type	Measurement type				
		Vibration/SPM	AE	Electro-static	Ultrasound	Oil/debris
[39–41]	RC	x				
[42,43]	RC	x	x			
[18,44,45]	REB	x				
[46–48]		REB	x	x		
[49–61]	RC		x			
[62–70]	REB		x			
[71]	REB	x				x
[72]	RC		x	x		
[73]	REB	x			x	
[4]	REB	x	x			x
[74]	REB	x		x		x
[75]	RC	x			x	

new SP techniques: envelope, wavelets, data-driven methods, expert systems, fuzzy logic techniques, etc. The majority of the advanced SP techniques are related to vibration measurements and these studies will be discussed in the next section. There are basically two testing approaches. The first one is the naturally accelerated testing with the help of applying overload, adjusting the lubricant film thickness or adding contaminated oil. The second approach is the artificially introduced defects by cutting, false-Brinelling, electric charge (i.e. erosion dent) and scratching. Moreover, some studies are based on the use of REBs and others utilise rolling contact test rigs e.g. ball on disc apparatus. In Table 3, a summary is given of the tests that have been applied to deteriorate either REB or other rolling contact (RC) mechanisms.

3.1. Testing techniques

Several experiments with the help of vibration measurements have been conducted with bearings such as [18,44,45], and on other rolling contact mechanisms such as [39–41]. Quite large number of studies have been interested to explore the AE technique on rolling contact mechanisms [49–61]. Other AE experiments have been performed with bearings [62–70]. Comparative studies that combine vibration and AE measurements have been conducted to explore defect features with the help of rolling contact test [42,43] and some other with the help of REBs [46–48]. Some studies [72,74] have investigated the capability of electrostatic charge measurements (when a charged particle passes the sensor) in detecting a bearing defect. The studies [73,75] have investigated the capability of ultrasound measurements in detecting a bearing defect, in particular, for the low speed bearings. In Table 1, it is clear that AE and vibration measurements are the most studied monitoring techniques. Also, quite a large share of the experimental tests have been done with the help of RC test rigs e.g. ball on disc. The oil/debris technique is commonly used as supportive techniques in several studies [4,71,74].

3.2. Influencing factors

Many detection issues have been studied such as the effect of surface roughness [2], the influence of running parameters on the AE of grease lubricated REB [76], the effect of λ factor (i.e. film thickness/surface roughness) [77], the running-in process [71,78], the effects of low speed, the large scale bearings and operating conditions (lubrication type, temperature) [79] and the effects of geometrical imperfections (i.e. variation of roller diameters, inner ring waviness), abrasive and fatigue wear [44]. The effect of contaminant concentration on vibration has also been studied [79–81].

3.3. Crack detection

Few studies have been conducted with the aim to detect the surface crack initiation [43,82] with the help of AE measurements. Sawalhi and Randall [45] observed the impulse effect of dent shoulders to the dynamic response. Few studies have been conducted to study the subsurface crack detection [49,83,84]. It was observed that AE is capable to detect subsurface cracks prior to pitting, however, based on the assumption that the detected bursts in the signal indicate near-surface cracking. The detection of defect propagation has been studied in [47,73,85,86]. It was observed that first, increasing the defect width increased the ratio of burst amplitude-to-operational noise (i.e. the burst signal was increasingly more evident above the operational noise levels). Second, it was reasoned that increasing the defect length increased the burst duration. The ultrasound signals were observed at low speeds and display a number of impulses which are generated by localized defects [73].

Table 4
A summary of SP analysis that have been studied for REBs.

References	Statistical measures	Time synchronous average	Morphological analysis	Parametric identification methods	Non-linear parameter identification method	FFT	Cyclostationary analysis	High-order spectra methods	Envelope Analysis and associated techniques e.g. ANC, SANC, DRS.	Wavelet transforms	Other transformers	Value-type data analysis	Combined event-measured data analysis
[4,41,43,46–49,53,55,57,–61–65,71–75,87,88]	X					X							
[89–91]		X											
[92,93]			X										
[94]				X									
[95–98]					X								
[99–106]							X						
[107–111]								X					
[29,99,112–123]									X				
[124–130]										X			
[131]													
[132,133]									X	X	X		
[134]										X			
[135]	X												
[136–138]												X	
[139–141]													X

3.4. Wear evolution

Jantunen [3] and Yoshioka and Shimizu [20] observed two main stages of wear progress: steady state and instability. The steady-state stage is roughly stable. However, a clear offset in the root mean square (RMS) values of monitoring signals is observed at instability stage, together with instability and rapid increase of these values before the final failure. Schwach and Guo [84] and Harvey et al. [74] observed three stages of wear progress. Moreover, the instability stage is observed to follow a steep-offset propagation. Harvey et al. [74] observed that electrostatic charge measurement indicate the wear initiation as a region of high signal amplitude (with respect to normal signal state), where it disappears (i.e. goes back to normal single state) until the failure occurs. Therefore, electrostatic measurement indicates instantaneous occurrences of wear mechanisms in the region of high signal amplitude rather than progressive stages. Manoj et al. [75] observed that the 3rd harmonic of the roller contact frequency of vibration has very good correlation with wear and when the pitting takes place, the amplitude of the 3rd harmonic of contact frequency increases to nearly four to five times the amplitude of other harmonics. In the same manner, the frequency analysis of sound signal shows that the 3rd and 1st harmonics of roller contact frequency have good correlation with the wear trend. Zhi-qiang et al. [43] observed two stages of wear progress using vibration measurements. However, four stages of wear progress were observed using AE: running-in, steady-state, a stage of minor-instability due to distributed defects, and finally a stage of major-instability due to pitting and spalling. Sawalhi and Randall [45] investigated the trend of kurtosis values of faulty signals, with relation to the development of the fault size. The kurtosis increases almost linearly in the early stage of testing time as the defect size increases. However, it stabilizes later as the defect size slowly extends. It could be either due to the existence of a smoothing process or the surface becoming totally rough without pronounced peaks/values when the effectiveness becomes weak.

4. Signal analysis methods

Over the years, several SP methods have been developed to extract the detect features from the raw signal of faulty REBs. A summary of signal analysis methods that have been used for the fault detection in REBs is given in Table 4.

4.1. Statistical measures

In the beginning, the SP methods were very simple and mainly based on statistical parameters such as RMS, mean, kurtosis, crest factor, etc. The trending based on RMS value is one of the most used methods which shows the correlation between vibration acceleration and the REB wear over the whole lifetime [4,43,74,84,87]. Kurtosis and crest factors increase as the spikiness of the vibration increases. In this sense, the kurtosis and the crest factor are very sensitive to the shape of the signal. However, the third central moment (Skewness) was found to be a poor measure of fault features in rolling bearings [142], in general skewness can be an effective measure for signals that are unsymmetrical i.e. non-linearity. The kurtosis is sensitive to the rotational speed and the frequency bandwidth. It is efficient in narrow bands at high frequencies especially for incipient defects [127,143]. More advanced approaches of time-domain analysis are the parameter identification methods, where a time series modelling is applied to fit the waveform data to a parametric time series model and extract the features [3]. Baillie and Mathew [94] introduced the concept of an observer bank of autoregressive (AR) time series models for fault diagnosis of slow speed machinery under transient conditions, where a short set of vibration data is needed. Due to instantaneous variations in friction, damping, or loading conditions, machine systems are often characterised by non-linear behaviour. Therefore, techniques for non-linear parameter estimation provide a good alternative for extracting defect-related features hidden in the measured signals [97]. A number of non-linear parameter identification techniques have been investigated, such as Correlation Dimension [95–97] and Complexity (the degree of regularity of a time series) Measure [144]. As the bearing system deteriorates due to the initiation and/or progression of defects, the vibration signal will increase, resulting in a decrease in its regularity and an increase in its corresponding entropy value [97]. In the early stage of machinery faults, the signal-to-noise ratio (SNR) is very low due to relatively weak characteristic signals. Therefore, chaotic oscillator was proposed [95,96,98] to extract the fault bearing features due to its sensitiveness to weak periodic signals. The complexity measure analysis shows that the inception and the growth of faults in the machine could be correlated with the changes in the Complexity value [144]. The biggest drawback of statistical methods is the need for a suitable quantity of data for training and testing the system during the development phase. A high number of data points that are needed to be calculated, which leads to lengthy computational time unsuitable for on-line, real-time applications [97].

4.2. Frequency domain methods

The frequency domain methods have been introduced to provide another way to detect the fault induced signals. FFT is commonly used method to transform the signal from time domain into its frequency components and produce a spectrum. However, it is often not clear enough method to observe the fault peaks, because of slip and masking by other stronger vibrations, beside the effects of harmonics of the defect frequencies and sidebands [88]. Moreover, the FFT method is actually based on the assumption of periodic signal, which is not suitable for non-stationary signals. The output signals of running REB contain non-stationary components due to the changes in the operating conditions and faults of the machine and bearing itself [145]. Time–frequency analysis is the most popular method to deal with non-stationary signals. The Wigner–Ville distribution

(WVD), the short time Fourier transform (STFT) and Wavelet transform (WT) represent a sort of compromise between the time and frequency based views of a signal and contain both time and frequency information. Mori et al. [130] applied the discrete wavelet transform (DWT) to predict the occurrence of spalling in REBs. Shibata et al. [146] used the WT to analyse the sound signals generated by bearings. Peng et al. [98] highlighted that Hilbert–Huang transform (HHT) has good computational efficiency and does not involve challenges with the frequency resolution and the time resolution.

4.3. Challenges of feature extraction process

There are several challenges to remove the speed fluctuations, the smearing effect of signal transfer path and the background noise. The effect of speed fluctuation (e.g. chirp signals) is important and need to be removed. The chirp signal or sweep signal i.e. a signal in which the frequency increases ('up-chirp') or decreases ('down-chirp') with time might be generated due to speed fluctuations, running-in, cut-out operations. It should be noticed that there has been several methods proposed to deal with the chirp signals such as chirp z-transform, chirp Fourier transform, adaptive chirplet transforms, and high-order estimations. Moreover, the order tracking methods are used to avoid the smearing of discrete frequency components due to speed fluctuations [6]. To solve the smearing effect problem, the Minimum entropy de-convolution (MED) method [121,122] has been developed.

For the background noise problem, different de-noising filters have been developed such as Discrete/random separation (DRS) [117], adaptive noise cancellation (ANC) [112,113], self-adaptive noise cancellation (SANC) [114–116] or linear prediction methods. However, for a situation, where the noise type and frequency range are unknown, the traditional filter designs could become computationally intense processes [123]. For example, the WT methods perform very well on Gaussian noise and can almost achieve optimal noise reduction while preserving the signal. However, it is still a challenge how to select an optimum wavelet for a particular kind of signal i.e. to select the optimum wavelet basis, to select the corresponding shape parameter and scale level for a particular application. Moreover, how to perform thresholding is another challenge. There are two major wavelet-based methods which are used for mechanical fault diagnosis: wavelet decomposition-based and wavelet filter-based method. Based on the WT, many kinds of fault features can be obtained, all of which can be classified as the wavelet coefficients based, wavelet energy based, singularity based and wavelet function based [145]. The continuous WT of Morlet wavelet functions have been used by Lin and Qu [126]. Junsheng et al. [128] proposed the impulse response wavelet base function to describe the vibration signal characteristics of the REB with fault, instead of the Morlet wavelet function. Liu et al. [129] proposed a weighted Shannon function to synthesize the wavelet coefficient functions to enhance the feature characteristics i.e. optimal wavelet shape factor and minimize the interference information. Djebala et al. [127] presented a denoising method of the measured signals based on the optimization of wavelet multi-resolution analysis based on the kurtosis value. Liu et al. [125] proposed a wavelet packet based method for the fault diagnostics of REB, where the wavelet packet coefficients were used as features. Altmann and Mathew [141] presented a method based on adaptive network based fuzzy inference system to select the wavelet packets of interest as fault features automatically, to enhance the detection and diagnostics of low speed REB faults. Su et al. [134] presented a new hybrid method based on optimal Morlet wavelet filter and autocorrelation enhancements i.e. to eliminate the frequency associated with interferential vibrations, reduce the residual in-band noise and highlight the periodic impulsive feature.

4.4. Bearing fault signals

Some studies [93,139–141] highlight that the most relevant information of a signal is often carried by the singularity points, such as the peaks, the discontinuities, etc. Therefore, singularity detection methods are proposed [139,140] based on calculating the Lipschitz exponents of the vibration signals. A large Lipschitz exponent indicates a regular point in the signal while a small Lipschitz exponent indicates a singular point. The WT is very successful in singularity detection, however before the singularity is detected, the signal pre-processing must be carried out, not to overlook some singularities [145]. Hao and Chu [93] observed that the impulse components cannot be seen clearly due to the existence of harmonic waves. The WT filtering removes the noise, however, the harmonic waves were not suppressed, since the impulse frequency was very close to the harmonic wave frequencies [93]. Therefore, the scalogram (i.e. a visual method of displaying a wavelet transform) is proposed to reveal more information about the signal.

Several methods try to extract the periodic information of the impulsive response of faulty REB such as the time synchronous average (TSA) [89–91]. The bearing fault signals have a deterministic part and a quasi-cyclostationary part, where the envelope and the squared envelope of the bearing vibration signal is the way to overcome this problem [99]. The envelope analysis utilizes the idea of detecting the fault impulses that are amplified by structural resonance. However, it is a challenge to determine the spectrum band which contains the highest signal-to-noise ratio (SNR). Randall [16] has highlighted that determining the suitable demodulation band is recently solved by means of e.g. spectral kurtosis (SK) [29,118–120]. Tse et al. [132] compared the effectiveness of the wavelet and the envelope detection methods for REBs fault diagnosis. The results showed that both the wavelet and envelope detection methods are effective in finding the bearing fault, but the wavelet method is less time expensive. The shortcoming of the envelope detection approach is the increasing difficulty in analysing the vibration spectrum when the signal-to-noise ratio is low [147], in which case the fault-imposed frequencies can be masked by noise and other frequency components. To overcome this problem, some morphological operators are proposed [93] with the aim to extract the envelope of impulsive type periodic vibration signals by modifying

Table 5
A summary of diagnostic methods that have been studied for REBs.

	Signal analysis methods					Diagnosis methods								
	Raw signals	Time-domain analysis	FFT	Envelope analysis	Wavelet analysis	Supervised ANN	Unsupervised ANN	Expert systems	Genetic algorithm	Fuzzy logic	Neuro-fuzzy	SVM	State observers	Model-based approach
[153,154]	X					X								
[94,155,156]		X				X								
[157]			X			X								
[158]					X	X								
[159,160]	X						X							
[151,152,161]								X						
[134]					X									
[152,162]									X					
[163,164]		X								X				
[165–169]											X	X		
[170,171]				X								X		
[172,173]					X							X		
[142]						X						X		
[88,174–176]					X								X	
[177–179]														X

(i.e. using morphological operators such as dilation, erosion, opening, closing) the geometrical features of the signals in the time domain. This constructs a kind of envelope which accentuates information corresponding to the impact series produced by a fault.

The impacts on the fault do not occur exactly in a periodic manner, because of random slips, possible speed fluctuations, and variations of the axial to radial load ratio. Therefore, the bearing fault signals are more likely described as cyclostationary [105,106], as pseudo-cyclostationary [104], as quasi-cyclostationary [99,100] and as poly-cyclostationary [100]. The cyclostationary is defined as a random signal in which the statistical parameters vary in time with single or multiple periodicities [148] and as a signals which, although not necessarily periodic, are produced by a hidden periodic mechanism [6]. The quasi-cyclostationary signal is generated when the existence of a common cycle is not allowed since some rotating components are not locked together such as is the case in REBs. Antoni et al. [100] highlighted that poly-cyclostationary signals are generated since many mechanical components in the machinery introduce various different periodicities, so they are a combination of cyclostationary processes with different basic cycles. Antoni et al. [100] explained that all kinematical variables in the machinery, which are periodic with respect to some rotational angles, are intrinsically angle-cyclostationary rather than time-cyclostationary. The synchronous averaging, comb-filters, blind filters and adaptive comb-filters are of the type of first-order cyclostationary methods. Synchronous auto-covariance function, Instantaneous variance, and Spectral correlation density are second-order cyclostationary methods. The Spectral correlation is proposed by Gardner [149] where the second-order periodicity can be characterized as the degree of coherence of a time series. Several studies have been discussed based on the cyclostationary and spectral correlation technique for fault detection in REBs [99–103]. The envelope analysis gives the same result as the integration of the cyclic spectral density function over all frequencies, thus establishing the squared envelope analysis as a valuable tool for the analysis of (quasi-) cyclostationary signals more generally [99]. Moreover, since the autocorrelation of a periodic signal is both periodic vs. time and time-lag, it produces a spectral correlation function which is discrete in both f and α directions like a "bed of nails". The Higher-order spectra describe the degree of phase correlation among different frequencies present in the signal [83]. Therefore, Li and Ma [124] used Bi-coherence spectra to derive features that relate to the condition of a bearing. Collis et al. [150] explained that the bi-spectrum can be viewed as a decomposition of the third moment 'skewness' of a signal over frequency and that it proves useful for analysing systems with asymmetric non-linearities. However, this statistical approach requires a rather large set of data to obtain a good estimation [130]. Pineyro et al. [111] compared the second-order power spectral density, the bi-spectral technique and the WT and found the later to be useful in the short transient detection, since it could eliminate the background noise.

5. Feature diagnosis methods

The fault diagnosis task consists of the determination of fault type with such many details as the fault size, location, and time to detection. Since a machine has many components and is highly complex, diagnosis of a machine fault usually requires technical skill and experience. It also requires extensive understanding of the machine's structure and operation, general concepts of diagnosis and an expert engineer to have domain specific knowledge of maintenance and to know the 'ins-and-outs' of the system. In reality, the expert is either too busy with several tasks or a specific component expert is not available at all [151]. In order to automatise the diagnosis procedures and provide the decision about the REB's health state, a number of automatic feature diagnosis methods have been developed. Several diagnosis methods are proposed to diagnose the faulty REBs such as artificial neural network (ANN), expert systems, fuzzy logic, support vector machine (SVM), state observes, and model-based methods. A summary of diagnostic methods that have been studied for REBs is presented in Table 5. However, since most of the diagnosis methods are utilising specific signal analysis methods, these combinations are illustrated in Table 5 as well. Moreover, many machinery fault diagnostic techniques have been developed to in order to increase the accuracy and reduce the errors caused by subjective human judgment [152].

5.1. Artificial neural network methods

The ANN methods have been applied to diagnose the REB's fault such as [153,157]. Larson et al. [154] performed the phase demodulation by means of neural networks. Li et al. [158] utilised the FFT as a pre-processor for Feed-Forward Neural network (FFNN) to perform fault detection. Samanta and Al-Balushi [156] developed a back-propagation neural network BPNN model, to reduce the number of inputs which leads to faster training requiring far less iterations. Moreover, Baillie and Mathew [94] illustrated the better noise rejection capabilities of the back-propagation networks compared to traditional linear methods. However, noise still remains a problem, and the best way to combat this is to use longer data lengths so that the noise can be effectively cancelled by the SNR averaging process. Alternatively, it also highlights the importance of signal pre-processing techniques, such as amplitude demodulation in the case of REBs [94]. The cascade correlation algorithm (CCA) offers the advantage of that the number of hidden units does not have to be determined prior to training. Spoerre [155] applied the CCA to predict the imbalance fault in rotor-bearing configuration. Radial basis functions are used by Baillie and Mathew [94] for REB, and compared to back-propagation networks, they show superior outcome due to their rapid training time. Since the unsupervised learning does not require external inputs, Wang and Too [159] applied the unsupervised neural networks, self-organising map (SOM) and learning vector quantisation to rotating machine fault detection. Tallam et al. [160] proposed some

self-commissioning and on-line training algorithms for FFNN with particular application to electric machine fault diagnostics. However, the building and training of artificial neural networks typically requires a trial and error approach and some experience [94]. The scope of the reviewed ANN methods is to classify the following fault features i.e. health state, defect type, defect location, defect severity, etc. Paya et al. [157] used the ANN to differentiate between each fault and establish the exact position of the fault occurring in the drive-line. Samanta and Al-Balushi [156] developed a BPNN model which obtains the fault features directly using very simple pre-processing i.e. root mean square, variance, skewness, kurtosis and normalised sixth central moment of the time-domain vibration signals, to classify the status of the machine in the form of normal or faulty bearings.

5.2. Expert systems

Different expert systems (ES) have been proposed for diagnosing abnormal measurements such as rule-based reasoning [151], case-based reasoning (CBR) [161], and model-based reasoning. It would be wise to present the cause-symptom relationship in a tabular form for quick comprehension and a concise representation. Yang et al. [151] developed a decision table i.e. IF (symptom) and THEN (cause) to link causes of fault and symptoms from an empirical knowledge gained either by direct experience with the system or through another expert in the field. The ANNs are required to learn gradually the knowledge in operating process, and to have the adaptive function expanding the knowledge continuously without the loss of the previous knowledge during learning new knowledge. Therefore, Yang et al. [161] proposed the integrated approach of Adaptive Resonance Theory and Kohonen Neural Network (ART-KNN). However, the previous cases may influence a CBR system in different directions without giving it many hints on which cases to consider as more important. This problem, associated with other difficulties in case-based indexing and retrieval, suggests that to combine the CBR with complementary forms of reasoning, such as rule-based, model-based or neural network, may be fruitful [161].

5.3. Fuzzy logic

In order to have flexible classification practices, the Fuzzy logic approach has been introduced. Fuzzy logic has gained wide acceptance as a useful tool for blending objectivity with flexibility. Fuzzy logic is also proving itself to be a powerful tool when used for knowledge modelling particularly when used in condition monitoring and diagnostics applications. Liu et al. [162] developed a fuzzy logic based expert system for rolling bearing faults. Mechefske [152] applied fuzzy logic method to classify frequency spectra representing various REB faults. Unlike other neural networks, fuzzy neural networks adopt bidirectional association. It makes use of the information from both fault symptoms and fault patterns and improves recognition rate. Therefore, Zhang et al. [140] applied fuzzy-neural network to diagnosis the fault of rotary machine. Jantunen [141] proposed the use of simplified fuzzy logic for automated prognosis. It saves the history of measured parameters and prognoses the further development.

5.4. Support vector machine

In practice, the huge number of possible loading conditions i.e. measuring situations makes the ANN task very complicated. Therefore, it is always a question that can the training results be moved from one machine to another. A SVM is another classification technique based on statistical learning theory. Three methods have been used to find the separating hyper-plane namely Quadratic Programming, Least-Squares and Sequential Minimal Optimization method. Yang et al. [170] used intrinsic mode function envelop spectrum as input to SVMs for the classification of bearing faults. Yang et al. [172] used improved wavelet packets and SVMs for the bearing fault detection. Abbasion et al. [166] used the SVM as a classifier to compute optimum wavelet signal decomposition level, in order to find an effective method for multi-fault diagnosis. Gryllias et al. [167] proposed the hybrid two stage one-against-all SVM approach for the automated diagnosis of defective REBs. In SVM approach, it is quite necessary to optimize the parameters which are the key factors impacting the classification performance. Li et al. [137] proposed an improved ant colony optimization (IACO) algorithm to determine these parameters, and then the IACO-SVM algorithm is applied on the REB fault detection. Liu et al. [138] proposed a multi-fault classification model based on Wavelet SVM (WSVM). Particle swarm optimization (PSO) is applied to seek the optimal parameters of WSVM and pre-processed using empirical model decomposition (EMD). Guo et al. [171] investigated the SVM method based on envelope analysis to diagnose REB with ball fault, inner race fault and outer race fault. The SVM is originally designed for two-class classification problem, while bearing fault diagnosis is a multi-class case. Tyagi [101] observed that more accurate classification of bearing condition is achieved by using SVM classifiers as compared to ANN. In fact, the ANN uses traditional empirical risk minimization principles to minimize the error on training data, while SVM utilizes structural risk minimization principles to minimize the upper bound of expected risk [171]. Pan et al. [173] proposed a combined method based on improved wavelet packet decomposition (IWPD) and support vector data description (SVDD) to gain better speed of training. However, Jack et al. [165] observed that the ANN tends to be faster to train and slightly more robust than the SVM.

The other non-linear classifiers like Gaussian Mixture Model (GMM) and Hidden Markov Model (HMM) have been used for classification problems in specific applications. Nelwamondo et al. [180] introduced GMM and HMM to diagnose fault in rolling bearings, based on extracted features using Multi-Scale Fractal Dimension (MFD), Mel frequency Cepstral Coefficients and kurtosis. However, the major drawback of HMM classifier is that it is computationally expensive, taking more than 20

Table 6
A summary of approaches adopted for prognosis of REBs.

Reference	Data-driven models									Physics-based models			
	Regression analysis	Probability estimation	Principal component analysis	Support vector machine	Neural networks	Self-organizing map	Recurrent neural networks	Neuro-fuzzy	Markov modelling, Genetic algorithm	Deterministic damage	Stochastic damage	Contact stress analysis	Dynamic analysis
[189,190],													X
[191,192]					X		X			X			
[193–199]					X								
[200]							X			X			
[201–204]							X						
[205,206]		X											
[138]		X									X		
[207]							X						
[208–212]								X					
[213–217]								X				X	
[218,219]	X							X					
[164]	X							X					
[220]			X		X								
[221]					X	X			X				
[222,223]		X							X				
[224–226]				X					X				
[173]			X	X									
[227]	X												

times longer than the duration required to train the GMM. Ocak et al. [88] developed a new scheme based on wavelet packet decomposition and HMM for tracking the severity of bearing faults. Zhang and Kang [176] proposed HHM model to represent the states of bearing through partition sub-state for the five states.

5.5. Model-based methods

The model-based methods utilise the physics models to diagnose the health of the monitored REB. Vania and Pennacchi [178] proposed a diagnostic technique where the fault is obtained by evaluating the system of excitations that minimizes the error i.e. residual, between the machine experimental response and the numerical response evaluated with the model. Söffker et al. [179] introduced Proportional-Integral Observer (PIO) method to detect a crack by detecting small stiffness changes. The very detailed and physical-oriented understanding that is provided by the model-based approach enhances the interpretation problem of signal based approaches. However, the necessity of fault models and the hypotheses about the location of the fault is a limitation. The majority of real industrial processes are nonlinear and is not effective to be modelled by using linear models for all operating conditions.

6. Prognosis analysis

Several researchers have reviewed the prognosis contributions such as Engel et al. [181], Jardine et al. [3], Lee et al. [182], Heng et al. [183], Peng et al. [184], Jammu and Kankar [185], etc. There are two types of methods of prognosis: physics 'model'-based and data-driven i.e. statistical and artificial intelligence (AI). In Table 6, an updated summary of the approaches adopted for prognosis of REBs is represented. The physics-based prognostic models describe the physics of the system and failure modes based on mathematical models such as Paris' law, Forman law, fatigue spall model, contact analysis and 'stiffness based damage rule' model. The data-driven prognostic models attempt to be driven by routinely and historically collected data (condition monitoring measurements, SCADA measurements, etc.). The data-driven prognostic models cover a high number of different techniques and artificial intelligence algorithms such as simple trend projection model, time series prediction model, exponential projection using (ANN, data interpolation using ANN, particle filtering, regression analysis and fuzzy logic, recursive Bayesian technique, HMM, hidden semi-Markov model, system identification model, etc. The data driven methods utilize data from past operations and current machine conditions, in order to forecast the remaining useful life. There are several reviews concerning the data-driven approaches such as [186–188].

6.1. Statistical approach

Yan et al. [218] explored a method to assess the performance of assets and to predict the remaining useful life. At first, a performance model is established by taking advantage of logistic regression analysis with maximum-likelihood technique. Two kinds of application situations, with or without enough historical data, are discussed in detail. Then, real-time performance is evaluated by inputting features of online data to the logistic model. Finally, the remaining life is estimated using an Auto-Regressive–Moving Average (ARMA) model based on the machine performance history; the degradation predictions are also upgraded dynamically. Vlok et al. [138] proposed a residual life estimation method based on proportional intensity model for non-repairable systems which utilise historic failure data and corresponding diagnostic measurements i.e. vibration and lubrication levels. Yang and Widodo [224] proposed a prognosis method using SVM. The statistics-based models assume that historical data is representative for the future wear progress, which is not always the case. Probabilistic-based models assume that the whole wear evolution progress is represented by a probability distribution function i.e. Weibull.

6.2. AI approach

Li et al. [191] utilized recurrent neural network (RNN) approach. Yam et al. [200] proposed a model based on the RNN approach for the critical equipment of a power plant. Dong et al. [197] proposed a model that combines condition prediction for equipment in a power plant based on grey mesh GM (1,1) model and BPNN on the basis of characteristic condition parameters extraction. Wang et al. [207] evaluated the performance of RNNs and neuro-fuzzy (NF) systems. Through comparison, it was found that if an NF system is properly trained, it performs better than RNNs in both forecasting accuracy and training efficiency. However, they often suffer from the need for complex training due to the huge number of possible combinations of damage scenarios that might take place in the case of rolling contact wear.

6.3. Physics-based approach

Physics-based prognostic models describe the physics of the system and failure modes based on mathematical models such as Paris' law, Forman law, fatigue spall model, contact analysis and stiffness based damage rule model. Physics-based prognostic models are based on crack length, and defect area as illustrated by Li et al. [191], and Li et al. [192], or relations of stiffness as shown by Qiu et al. [201]. However, the most challenging issue within physics-based prognostic is to define the loading-damage relationship and to model it. There are models based on damage rules such as linear damage rule, damage curve rule, and double-linear damage rule [201]. The drawback of these simplified functions is that they all use the constant

damage factor which is hard to estimate or measure. Moreover, these functions are either linear or multi-linear functions. That means that the estimated results might seem matching with the overall measured results, however, they might describe different damage scenarios. Therefore, the prediction based on such functions makes the prognosis a risky task. Recently, some model-based models have been utilised for the contact stress analysis to illustrate the wear evolution progress. These models provide more accurate predictions. Some models are based on contact stress analysis [215] and some are based on system dynamics [189,190]. Chelidza and Cusumano [217] proposed a method based on a system dynamics approach to estimate the damage evolution. The results of these models depend also on the stress-damage function and the constant damage factor that are in use. These models assume that each wear mechanism generates stresses that in total equal to the overall measured stresses. Therefore, the wear mechanics interactions and competitions are somehow ignored.

7. Discussions

7.1. Modelling techniques

The dynamic models deal with the wear phenomenon as a localized defect with fixed features over the lifetime. The reason is that the purpose of these models is to detect the defect within the generated vibration signals and not the incremental deterioration process, i.e. wear evolution. These dynamic models start from the point where the defect is localized as a simulated defect in the models or artificially introduced into the experiments. That ignores the prior stages of the localization process. The localized defects and their associated impact stay constant over the whole lifetime. This kind of approach ignores the topological and tribological changes of the defected surface.

In order to model the wear evolution an incremental numerical procedure should be developed which would be able to integrate the contact information continuously into the dynamic model. This means that the applied force due to the wear progress and its associated topological and tribological conditions should be iteratively updated into the dynamic model.

7.2. Monitoring techniques

The artificially introduced defect approach is widely used due to its simplicity. The researchers can virtually introduce a well-known shape and size of a defect. Besides, they can artificially introduce the same defect features in the validation
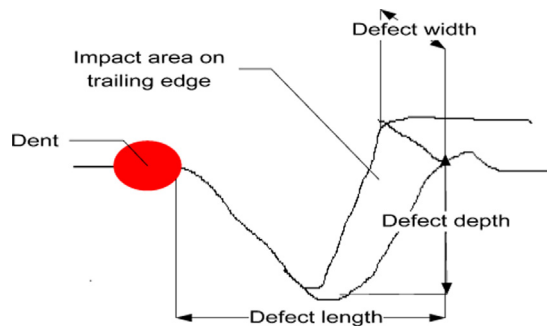

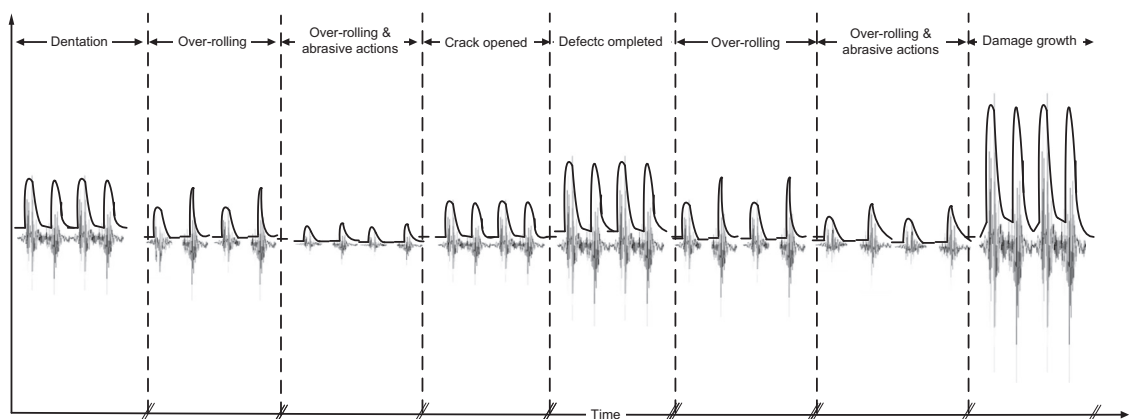
Fig. 2. The impact area of a defect.

Fig. 3. The evolution of the faulty bearing response.

experiment. Furthermore, this approach delimits the testing complexity, as it focuses on a single artificial defect, compared to the natural defect propagation approach. However, the natural wear process highlights that the bearing defect is changing over the time with respect to the topological and tribological changes due to different wear and stress concentration mechanisms. The drawback of the artificially introduced defect approach is that the damage criterion is somehow artificially determined which might be totally different than the defect in the real operation. Therefore, the artificial defected bearing tests are helpful in the development of new analysis and diagnosis techniques; however they are not helpful in the investigation of the evolution of real wear progress.

The impulsive vibration response is clearly seen when the impact of the roller element that passes over a defect is strong. The impact severity is related to the size of impact area and the sharpness of the edges of the defect. The impact area is the area on the trailing edge of the dent, asperity or the defect that becomes in contact with rolling element, as shown in Fig. 2. Based on the literature, this area is quite small at the defect initiation stage and depends on the length, depth and width of the defect. Al-Ghamd and Mba [47] observed that increasing the defect width increased the ratio of burst amplitude-to-operational noise (i.e. the burst signal was increasingly more evident above the operational noise levels). It was also observed that increasing the defect length increased the burst duration. The first observation indicates that the width of defect increases the impact area on the trailing edge and therefore stronger amplitude and high SNR was observed. In fact, most of the studies show the ability of the envelope analysis to detect such influence of the impact area. However, the challenge with wear evolution is when the impact areas are rapidly and continuously changing due to the loading and wear progress.

In the early stage of wear process, the defect is quite small and can be easily buried by other vibration phenomena. Most of the SP methods are validated based on experimental tests in which the defects are introduced artificially into the bearings. Such testing approach guarantees the availability of the impulsive response due to the introduced defect and somehow its severity is quite enough to be detected. The natural accelerated testing experiments [4,87] show that it is quite hard to detect the impulsive response at an early stage and much harder to track its evolution. The basic reason behind the difficulty is that the relation between the defect growth i.e. to become larger is not linear with its dynamic impact. Also it significantly depends on the wear and stress contraction mechanisms that are involved. The experiments show that the wear process is slow in nature and can hardly produce detectable impacts at early stage. Moreover, the experiments in literature show that the impulsive response of bearing defect is changing over the time with respect to the topological and tribological changes, as shown in Fig. 3. The natural accelerated tests show fluctuations in the impulsive response of bearing defects and in some time intervals it is hard to detect them. For example, the over-rolling and abrasive wear effects make the defected surface smoother and the impact events softer.

These empirical facts are quite important to explain the capabilities and limitations of the applied monitoring methods, in order to enhance their suitability for wear evolution monitoring in REBs.

7.3. Signal analysis and diagnosis methods

In many cases, the measured signals have very low signal-to-noise ratio (SNR) which makes the feature extraction of the studied components difficult. In fact, in the early stage of bearing failures, the bearing characteristic frequencies contain very little energy and are often overwhelmed by noise and higher level structural vibrations. An effective SP method would be necessary to remove such corrupting noise and interference [134]. The overlapping and interference of signals might mislead the analysis of signals [145]. The interference terms should be reduced in order to improve the readability of many methods. It is usually difficult to explain the results due to the REB complexity in terms of structure and operate under noisy or uncertain environment [129]. As summarised by Sawalhi [29] the signal analysis methods try to cope with five main challenges as have been highlighted by many studies: (1) remove the speed fluctuation; (2) remove the noise effect; (3) remove the smearing effect of transfer path; (4) select optimal band of high Signal-to-Noise ratio; and (5) extract clear fault features. However, the evolution of the fault induced signal requires effective tracking of the extracted fault features over the bearing lifetime. This is the sixth challenge. The important issue of tracking technique is the signal and feature extraction analysis method it uses and how effective and clear indication it can provide. Tracking the evolution progress demand more features than just the normal detection of a bearing fault.

The signal analysis methods, which are applied on signals measured from bearings with artificially introduced defect, are quite effective. However, a careful comparison between the defect features of a natural wear process and the artificial introduced defect should be taken into consideration. Simply, the artificially introduced defects are in general large, sharp and strictly localised. A natural fault is smaller, less sharp and has evolved with the help of different wear and stress concentration mechanisms. In fact, the impulse due to wear defect is changing over the whole lifetime as schematically illustrated in Fig. 3.

It is also clear from the literature that the definition of the bearing fault signal type i.e. stationary, cyclostationary, non-stationary, etc. is the main reason and motivation for the variety of SP methods. Some methods are just for specific type of signals and it is hard to illustrate their outcomes or there is no point to use them if the fault-induced signals are not from that specific type. Thus, it is more realistic to illustrate how the bearing fault-induced signal is evolved over the whole REB's lifetime. The fault-induced signal is usually of the impulsive signal type due to the impact event when the rolling element is passing over e.g. asperity, dentation or defect. The defect topology affects the impact severity when a rolling element passes over it.

Therefore, the impulsive nature of wear is changing as the wear defect evolves. Moreover, in some wear progression intervals, there is no clear impulsive impact, and consequently some monitoring and diagnosis techniques are not effective during that period.

One important issue is the early detection of the fault i.e. earliness. There are wide and qualitative definitions of the detection earliness within the literature. In fact, many studies which have utilised the artificially introduced defects can be recognised to represent a severe wear state in the real application which might also be detected with simple SP methods. It is clear that the signal analysis methods should detect the defect as early as possible. Most of the feature diagnosis methods classify the REB's state into healthy or faulty state. Some other methods aim to classify the defect types i.e. imbalance, defect, and defect locations i.e. outer race, inner race, rolling element. Very few studies classify the defect evolution in terms of wear stages.

7.4. Prognosis methods

The survey shows that the data driven approach is more adopted in the prognosis of rolling bearings than the physics based approach. All prognosis approaches i.e. physics based or data driven have advantages and drawbacks in different applications and operating cases, specially, in case of variable operating conditions. Moreover, the prognosis models try to control or delimit the effect of some operational variables. However, that is somehow possible in the experimental tests but not in real applications. The prediction based on simplified experimental tests i.e. ball on disc test is easier than tests that use REBs. The statistical models represent the wear evolution as one function with the possibility to insert weights. The statistical models assume that the past history profile represents the future failure mechanism of a specific component. However, the failure mechanisms are changing with respect to the failure evolution and the involvement of failure mechanisms. It means that the statistical approach is not fully valid and might not represent wear progress, especially, if the evolution stages are highly varying, as they are in the case of wear evolution stages. ANN models use specific functions and multiple weights. However, ANN models have drawbacks once the system conditions are rapidly fluctuating. The model based models are still representing the wear evolution with two stages. Moreover, the damage is represented as a damage factor. This really is a dramatic simplification to describe the wear evolution as a two stage stabile phenomenon, whereas it by nature has a complex evolution process. Consequently, this kind of approach is very far from reality and the usability can be highly criticized. The prognosis models can be improved remarkably by understanding the physics of the wear evolution progress and its associated measured outcomes. Actually, that will help the model based approaches to provide better results and the data driven approach to have better interpretations of the results and training inputs.

8. Conclusions

The survey reviews the capabilities, advantages and disadvantages of bearing modelling and monitoring procedures i.e. SP, diagnosis and prognosis methods. It emphasises the main PHM capabilities, limitation and challenges for REBs. It is observed that several experimental agreements with their associated analytical models are valid for specific wear definitions. The most commonly used definition of wear is to artificially introduce a localised defect into the surface i.e. sharp, large enough and at specific radial location. However, in reality, the fault features are changing over the lifetime due to topological and tribological changes. These changes are due to the wear progression in rolling contact because of the involvement of different wear and stress concentration mechanisms. Therefore, the review highlights the evolution monitoring challenge of the wear fault over the REB's lifetime. Moreover, it discusses the directions and implications of understanding the natural evolution of the wear process in REB and how can it be monitored and modelled effectively.

The review of modelling methods highlights the need for a wear evolution model e.g. incremental numerical procedure which would be able to simulate and integrate the contact information continuously into a dynamic model. This means that the applied force due to the wear progress and its associated topological and tribological conditions should be iteratively updated into the dynamic model. The review part of the monitoring and experimental testing methods highlights the need to understand the dominant physical damage mechanism of each testing method in order to interpret the measured signal in the right way. In terms of signal processing and diagnosis method, it is very clear that several methods might show high capabilities at specific time intervals within the whole lifetime. However, they might have poor capability to indicate the fault at other time interval, due to the change in the surface topography. Therefore, there is a need to study and validate these methods with data at time intervals that contain the several topographical changes due to wear evolution. Moreover, the prognosis models can be also improved remarkably by understanding the physics of the wear evolution progress and its associated measured outcomes.

References

[1] I. Howard, A Review of Rolling Element Bearing Vibration 'Detection, Diagnosis and Prognosis', Melbourne, Australia, 1994.
[2] N. Tandon, A. Choudhury, A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings, Tribol. Int. 32 (1999) 469–480.
[3] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Process. 20 (7) (Oct. 2006) 1483–1510.
[4] E. Jantunen, How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods, Int. J. COMADEM 9 (3) (2006) 24–38.
[5] J. Halme, P. Andersson, Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics-state of the art, J. Eng. Tribol. 224 (2009) 377–393.
[6] R.B. Randall, J. Antoni, Rolling element bearing diagnostics—a tutorial Feb., Mech. Syst. Signal Process. 25 (2) (2011) 485–520.
[7] I. El-Thalji, E. Jantunen, A descriptive model of wear evolution in rolling bearings, Eng. Fail. Anal. 45 (2014) 204–224.
[8] A. Palmgren, Ball and Roller Bearing Engineering, S.H. Burbank and Co. Inc., Philadelphia, PA, 1947.
[9] F.J. Harris, Rolling Bearing Analysis I, First, John Wiley & Sons, Inc., New York, 1966.
[10] P.K. Gupta, Transient ball motion and skid in ball bearings, J. Lubr. Technol.: Trans. ASME (1975) 261–269.
[11] S. Fukata, E.H. Gad, T. Kondou, T. Ayabe, H. Tamura, On the vibration of ball bearings, Bull. JSME 28 (239) (1985) 899–904.
[12] Y.H. Wijnat, J.A. Wensing, G.C. van Nijen, The influence of lubrication on the dynamic behaviour of ball bearings, J. Sound Vibr. 222 (4) (1999) 579–596.
[13] R. Tiwari, N.S. Vyas, Dynamic response of an unbalanced rotor supported on ball bearings, J. Sound Vib. 187 (2) (1995) 229–239.
[14] M. Tiwari, K. Gupta, O. Prakash, Dynamic response of an unbalanced rotor supported on ball bearings, J. Sound Vib. 238 (2000) 757–779.
[15] M. Tiwari, K. Gupta, O. Prakash, Effect of radial internal clearance of ball bearing on the dynamics of a balanced horizontal rotor, J. Sound Vib. 238 (2000) 723–756.
[16] J. Sopanen, A. Mikkola, Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: Theory Jan., Proc. Inst. Mech. Eng. Part K J. Multi-body Dyn. 217 (3) (2003) 201–211.
[17] Z. Kiral, H. Karagülle, Simulation and analysis of vibration signals generated by rolling element bearing with defects Sep., Tribol. Int. 36 (9) (2003) 667–678.
[18] N. Sawalhi, R.B. Randall, Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults Nov., Mech. Syst. Signal Process. 22 (8) (2008) 1924–1951.
[19] F. Massi, J. Rocchi, A. Culla, Y. Berthier, Coupling system dynamics and contact behaviour: Modelling bearings subjected to environmental induced vibrations and 'false brinelling' degradation May, Mech. Syst. Signal Process. 24 (4) (2010) 1068–1080.
[20] R.K. Purohit, K. Purohit, Dynamic analysis of ball bearings with effect of preload and number of balls, Int. J. Appl. Mech. Eng. 11 (1) (2006) 77–91.
[21] M. Cao, J. Xiao, A comprehensive dynamic model of double-row spherical roller bearing—model development and case studies on surface defects, preloads, and radial clearance Feb., Mech. Syst. Signal Process. 22 (2008) 467–489.
[22] M. Nakhaeinejad, Fault Detection and Model-Based Diagnostics in Nonlinear Dynamic Systems, University of Texas, Austin, 2010.
[23] G.H. Jang, S.W. Jeong, Nonlinear excitation model of ball bearing waviness in a rigid rotor supported by two or more ball bearings considering five degrees of freedom, J. Tribol. 124 (2002) 82–90.
[24] P.D. McFadden, J.D. Smith, Model for the vibration produced by a single point defect in a rolling element bearing, J. Sound Vib. 96 (1) (1984) 69–82.
[25] P.D. McFadden, J.D. Smith, The vibration produced by multiple point defects in a rolling element bearing, J. Sound Vib. 98 (2) (1985) 263–273.
[26] N. Tandon, A. Choudhury, An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect, J. Sound Vib. 205 (3) (1997) 275–292.
[27] S.P. Harsha, Non-linear dynamic response of a balanced rotor supported on rolling element bearings May, Mech. Syst. Signal Process. 19 (3) (2005) 551–578.
[28] S. Sassi, B. Badri, M. Thomas, A numerical model to predict damaged bearing vibrations Nov., J. Vib. Control 13 (11) (2007) 1603–1628.
[29] N. Sawalhi, Diagnostics, Prognostics and Fault Simulation for Rolling Element Bearings, UNSW, Sydney, 2007.
[30] A. Ashtekar, F. Sadeghi, L.-E. Stacke, A new approach to modeling surface defects in bearing dynamics simulations, J. Tribol. 130 (2008).
[31] A. Rafsanjani, S. Abbasion, A. Farshidianfar, H. Moeenfard, Nonlinear dynamic modeling of surface defects in rolling element bearing systems Jan., J. Sound Vib. 319 (2009) 1150–1174.
[32] M.S. Patil, J. Mathew, P.K. Rajendrakumar, S. Desai, A theoretical model to predict the effect of localized defect on vibrations associated with ball bearing Sep., Int. J. Mech. Sci. 52 (2010) 1193–1201.
[33] M. Tadina, M. Boltezar, Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up, J. Sound Vib. 300 (17) (2011) 4287–4301.
[34] J. Liu, Y. Shao, T.C. Lim, Vibration analysis of ball bearings with a localized defect applying piecewise response function Oct., Mech. Mach. Theory 56 (2012) 156–169.
[35] Y.-F. Wang, P.J. Kootsookos, Modeling of low shaft speed bearing faults for condition monitoring May, Mech. Syst. Signal Process. 12 (3) (1998) 415–426.
[36] S.H. Ghafari, F. Golnaraghi, F. Ismail, Effect of localized faults on chaotic vibration of rolling element bearings Dec., Nonlinear Dyn. 53 (4) (2007) 287–301.
[37] A.S. Malhi, Finite Element Modelling of Vibrations Caused by a Defect in the Outer Ring of a Ball Bearing, University of Massachusetts, Amherst, 2002.
[38] H. Endo, A Study of Gear Faults by Simulation and the Development of Differential Diagnostic Techniques, UNSW, Sydney, 2005.
[39] T.E. Tallian, O.G. Gustafsson, Progress in rolling bearing vibration research and control, ASLE Trans. 8 (1965) 195–207.
[40] E.M. Yhland, Waviness measurement-an instrument for quality control in rolling bearing industry, Proc. Inst. Mech. Eng. 182 (1967) 438–445.
[41] M.N. Kotzalas, T.A. Harris, Fatigue Failure progression in ball bearings, Trans. ASME, J. Tribol. 123 (2001) 238–242.
[42] N.S. Swansson, S.C. Favalord, Application of Vibration Analysis to the Condition Monitoring of Rolling Element bearings Melbourne, Austrralia, 1984.
[43] Z. Zhi-qiang, L. Guo-lu, W. Hai-dou, X. Bin-shi, P. Zhong-yu, Z. Li-na, Investigation of rolling contact fatigue damage process of the coating by acoustics emission and vibration signals Mar., Tribol. Int. 47 (2012) 25–31.
[44] C.S. Sunnersjö, Rolling bearing vibrations—the effects of geometrical imperfections and wear, J. Sound Vib. 98 (4) (1985) 455–474.
[45] N. Sawalhi, R.B. Randall, Vibration response of spalled rolling element bearings: observations, simulations and signal processing techniques to track the spall size Apr., Mech. Syst. Signal Process. 25 (3) (2011) 846–870.
[46] N. Tandon, B.C. Nakra, Comparison of vibration and acoustic measurement techniques for the condition monitoring of rolling element bearings Jun., Tribol. Int. 25 (3) (1992) 205–212.
[47] A.M. Al-Ghamd, D. Mba, A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size Oct., Mech. Syst. Signal Process. 20 (2006) 1537–1571.
[48] N. Tandon, G.S. Yadava, K.M. Ramakrishna, A comparison of some condition monitoring techniques for the detection of defect in induction motor ball bearings Jan., Mech. Syst. Signal Process. 21 (2007) 244–256.
[49] T. Yoshioka, Detection of rolling contact subsurface fatigue cracks using acoustic emission technique, J. Soc. Tribol. Lubr. Eng. 49 (4) (1992) 303–308.
[50] D. Nelias, T. Yoshioka, Location of an acoustic emission source in a radially loaded deep groove ball-bearing, Proc. Inst. Mech. Eng. Part J J. Eng. Tribol. 212 (1998) 33–45.
[51] R.C. Dommarco, P.C. Bastias, G.T. Hahn, C.A. Rubin, The use of artificial defects in the 5-ball-rod rolling contact fatigue experiments Mar., Wear 252 (2002) 430–437.

[52] T.F. Page, B.A. Shaw, Scanning electron acoustic microscopy (SEAM): a technique for the detection of contact-induced surface & sub-surface cracks, J. Mater. Sci. 39 (2004) 6791–6805.
[53] E.D. Price, A.W. Lees, M.I. Friswell, Detection of severe sliding and pitting fatigue wear regimes through the use of broadband acoustic emission, Proc. Inst. Mech. Eng. Part J J. Eng. Tribol. 219 (2005) 85–98. Jan.
[54] Y. Guo, D. Schwach, An experimental investigation of white layer on rolling contact fatigue using acoustic emission technique Sep., Int. J. Fatigue 27 (2005) 1051–1061.
[55] A. Warren, Y. Guo, Acoustic emission monitoring for rolling contact fatigue of superfinished ground surfaces Apr., Int. J. Fatigue 29 (2007) 603–614.
[56] Z. Rahman, H. Ohba, T. Yamamoto, T. Yoshioka, A study on incipient damage monitoring in rolling contact fatigue process using acoustic emission, Tribol. Trans. 51 (2008) 543–551.
[57] Z. Rahman, H. Ohba, T. Yoshioka, T. Yamamoto, Incipient damage detection and its propagation monitoring of rolling contact fatigue by acoustic emission Jun., Tribol. Int. 42 (6) (2009) 807–815.
[58] F. Hort, P. Mazal, F. Vlasic, Monitoring of acoustic emission signal of loaded axial bearings, J. Mater. Sci. Eng. A 1 (1) (2011) 717–724.
[59] F. Hort, P. Mazal, Application of acoustic emission for measuring of contact fatigue of axial bearing, Eng. Mech. 18 (2) (2011) 117–125.
[60] L. Nohal, P. Mazal, F. Hort, Analysis of surface initiated damage in thrust bearings with acoustic emission, in: 30th European Conference on Acoustic Emission Testing & Seventh International Conference on Acoustic Emission, 2012, no. September.
[61] L. Guo-lu, Z. Zhi-qiang, W. Hai-dou, X. Bin-shi, P. Zhong-yu, Z. Li-na, Acoustic emission monitoring and failure mechanism analysis of rolling contact fatigue for Fe-based alloy coating May, Tribol. Int. 61 (2013) 129–137.
[62] A. Choudhury, N. Tandon, Application of acoustic emission technique for the detection of defects in rolling element bearings Jan., Tribol. Int. 33 (1) (2000) 39–45.
[63] N. Jamaludin, D. Mba, Monitoring extremely slow rolling element bearings: Part I Sep., NDT E Int. 35 (6) (2002) 349–358.
[64] N. Jamaludin, D. Mba, Monitoring extremely slow rolling element bearings: Part II Sep., NDT E Int. 35 (6) (2002) 359–366.
[65] M. Elforjani, D. Mba, Observations and location of acoustic emissions for a naturally degrading rolling element thrust bearing May, J. Fail. Anal. Prev. 8 (4) (2008) 370–385.
[66] M. Elforjani, D. Mba, Assessment of natural crack initiation and its propagation in slow speed bearings, Nondestr. Test. Eval. 24 (3) (2009) 261–275.
[67] M. Elforjani, D. Mba, Accelerated natural fault diagnosis in slow speed bearings with acoustic emission Jan., Eng. Fract. Mech. 77 (1) (2010) 112–127.
[68] Y. He, X. Zhang, M.I. Friswell, Defect diagnosis for rolling element bearings using acoustic emission, J. Vib. Acoust. 131 (6) (2009).
[69] Y. He, X. Zhang, M.I. Friswell, Observation of time–frequency characteristics of the acoustic emission from defects in rolling element bearings Non-Destructive Testing and Condition Monitoring, Insight 52 (8) (2010) 412–418.
[70] L. Solazzi, C. Petrogalli, M. Lancini, Vibration based diagnostics on rolling contact fatigue test bench Jan., Procedia Eng. 10 (2011) 3465–3470.
[71] Z. Peng, N. Kessissoglou, An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis Aug., Wear 255 (7–12) (2003) 1221–1232.
[72] J. Sun, R.J.K. Wood, L. Wang, I. Care, H.E.G. Powrie, Wear monitoring of bearing steel using electrostatic and acoustic emission techniques, Wear 259 (2005) 1482–1489.
[73] Y.-H. Kim, A.C.C. Tan, J. Mathew, B.-S. Yang, Condition monitoring of low speed bearings: a comparative study of the ultrasound technique versus vibration measurements, World Congr. Eng. Asset Manage. (2006).
[74] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Electrostatic wear monitoring of rolling element bearings Sep., Wear 263 (2007) 1492–1501.
[75] V. Manoj, K. Manohar Shenoy, K. Gopinath, Developmental studies on rolling contact fatigue test rig Mar., Wear 264 (2008) 708–718.
[76] J. Miettinen, The influence of the running parameters on the acoustic emission of grease lubricated rolling bearings, Mainten. Asset Manage. (2000).
[77] R. Serrato, M.M. Maru, L.R. Padovese, Effect of lubricant viscosity grade on mechanical vibration of roller bearings Aug., Tribol. Int. 40 (2007) 1270–1275.
[78] G.P. Massouros, Normal vibration of a plain bearing working under boundary lubrication conditions Oct., Tribol. Int. 16 (5) (1983) 235–238.
[79] T. Momono, B. Noda, Sound and Vibration in rolling bearings, Motion Control 6 (1999) 29–37.
[80] M.M. Maru, R.S. Castillo, L.R. Padovese, Study of solid contamination in ball bearings through vibration and wear analyses Mar., Tribol. Int. 40 (3) (2007) 433–440.
[81] R.J. Boness, S.L. McBride, Adhesive and abrasive wear studies using acoustic emission techniques Sep., Wear 149 (1991) 41–53.
[82] R. Kocich, M. Cagala, J. Crha, P. Kozelsky, Character of acoustic emission signal generated during plastic deformation, in: 30th European Conference on Acoustic Emission Testing & Seventh International Conference on Acoustic Emission, 2012, vol. 1, no. September.
[83] E.D. Price, A.W. Lees, M.I. Friswell, B.J. Roylance, Online detection of subsurface distress by acoustic emissions, Key Eng. Mater. 246–246 (2003) 451–460.
[84] D. Schwach, Y. Guo, A fundamental study on the impact of surface integrity by hard turning on rolling contact fatigue Dec., Int. J. Fatigue 28 (2006) 1838–1844.
[85] T. Yoshioka, T. Fujiwara, Measurement of propagation initiation and propagation time of rolling contact fatigue cracks by observation of acoustic emission and vibration, Tribol. Ser. 12 (1987) 29–33.
[86] H. Kakishima, T. Nagatomo, H. Ikeda, T. Yoshioka, A. Korenaga, Measurement of acoustic emission and vibration of rolling bearings with an artificial defect, QR RTRI 41 (3) (2000) 127–130.
[87] T. Yoshioka, S. Shimizu, Monitoring of ball bearing operation under grease lubrication using a new compound diagnostic system detecting vibration and acoustic emission Oct., Tribol. Trans. 52 (6) (2009) 725–730.
[88] H. Ocak, K.a. Loparo, F.M. Discenzo, Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: a method for bearing prognostics May, J. Sound Vib. 302 (2007) 951–961.
[89] P.D. McFadden, A revised model for the extraction of periodic waveforms by time domain averaging, Mech. Syst. Signal Process. 1 (1) (1987) 83–95.
[90] G. Dalpiaz, A. Rivola, R. Rubini, Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears May, Mech. Syst. Signal Process. 14 (3) (2000) 387–412.
[91] A.J. Miller, A New Wavelet Basis for the Decomposition of Gear Motion Error Signals and its Application to Gearbox Diagnostics, The Pennsylvania State University, State College, PA, 1999.
[92] N.G. Nikolaou, I.a. Antoniadis, Application of morphological operators as envelope extractors for impulsive-type periodic signals Nov., Mech. Syst. Signal Process. 17 (6) (2003) 1147–1162.
[93] R. Hao, F. Chu, Morphological undecimated wavelet decomposition for fault diagnostics of rolling element bearings Mar., J. Sound Vib. 320 (2009) 1164–1177.
[94] D.C. Baillie, J. Mathew, A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings Jan., Mech. Syst. Signal Process. 10 (1) (1996) 1–17.
[95] D. Logan, J. Mathew, Using the correlation dimension for vibration fault diagnosis of rolling element bearings—I. Basic concepts May, Mech. Syst. Signal Process. 10 (3) (1996) 241–250.
[96] D.B. Logan, J. Mathew, Using the correlation dimension for vibration fault diagnosis of rolling element bearings—II. Selection of experimental parameters May, Mech. Syst. Signal Process. 10 (3) (1996) 251–264.
[97] R. Yan, R.X. Gao, Approximate entropy as a diagnostic tool for machine health monitoring Feb., Mech. Syst. Signal Process. 21 (2007) 824–839.
[98] W. Wei, L. Qiang, Z. Guojie, Novel approach based on chaotic oscillator for machinery fault diagnosis Oct., Measurement 41 (8) (2008) 904–911.
[99] R.B. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals Sep., Mech. Syst. Signal Process. 15 (5) (2001) 945–962.

[100] J. Antoni, F. Bonnardot, A. Raad, M. El Badaoui, Cyclostationary modelling of rotating machine vibration signals Nov., Mech. Syst. Signal Process. 18 (6) (2004) 1285–1314.
[101] J. Antoni, Cyclic spectral analysis in practice Feb., Mech. Syst. Signal Process. 21 (2) (2007) 597–630.
[102] J. Antoni, Cyclic spectral analysis of rolling-element bearing signals: facts and fictions Jul., J. Sound Vib. 304 (3–5) (2007) 497–529.
[103] J. Antoni, Cyclostationarity by examples May, Mech. Syst. Signal Process. 23 (4) (2009) 987–1036.
[104] R.B. Randall, Vibration-based Condition Monitoring, John Wiley & Sons, Ltd., Chichester, UK, 2011.
[105] A.C. McCormick, A.K. Nandi, Cyclostationarity in rotating machine vibrations, Mech. Syst. Signal Process. 12 (2) (1998) 225–242.
[106] B. Kilundu, X. Chiementin, J. Duez, D. Mba, Cyclostationarity of acoustic emissions (AE) for monitoring bearing defects Aug., Mech. Syst. Signal Process. 25 (6) (2011) 2061–2072.
[107] A.C. McCormick, A.K. Nandi, Bispectral and trispectral features for machine condition diagnosis, IEEE Proc.—Vis. Image Signal Process. 146 (5) (1999) 229.
[108] T.D. Almeida, S.A. da, S. Vicente, L.R. Padovese, New technique for evaluation of global vibration levels in rolling bearings, Shock Vibr. 9 (2002) 225–234.
[109] D.-M. Yang, A.F. Stronach, P. Macconnell, J. Penman, Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks Mar., Mech. Syst. Signal Process. 16 (2–3) (2002) 391–411.
[110] L. Qu, X. Liu, G. Peyronne, Y. Chen, The holospectrum: a new method for rotor surveillance and diagnosis, Mech. Syst. Signal Process. 3 (3) (2003) 255–267.
[111] J. Pineyro, A. Klempnow, V. Lescano, Effectiveness of new spectral tools in the anomaly detection of rolling element bearings, J. Alloys Compd. 310 (2000) 276–279.
[112] G. Chaturvedi, D.W. Thomas, Bearing fault detection using adaptive noise cancelling, J. Sound Vib. 104 (1982) 280–289.
[113] C.C. Tan, B. Dawson, An adaptive noise cancellation approach for condition monitoring of gearbox bearings, Process. Int. Tribol. Conf. (1987).
[114] D. Ho, Bearing Diagnostics and Self Adaptive Noise Cancellation, UNSW, 2000.
[115] J. Antoni, R.B. Randall, Optimisation of SANC for separating gear and bearing signals, Proc. COMADEM Conf. (2001) 89–96.
[116] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals: Part I—Evaluation of adaptive algorithm, Mech. Syst. Signal Process. 18 (2004) 89–101.
[117] J. Antoni, R.B. Randall, Unsupervised noise cancellation for vibration signals: Part II—A novel frequency-domain algorithm, Mech. Syst. Signal Process. 18 (2004) 103–117.
[118] N. Sawalhi, R.B. Randall, Spectral kurtosis optimization for rolling element bearings, Proc. ISSPA Conf. (2005).
[119] J. Antoni, The spectral kurtosis: a useful tool for characterising nonstationary signals, Mech. Syst. Signal Process. 20 (2) (2006) 282–307.
[120] J. Antoni, R.B. Randall, The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines, Mech. Syst. Signal Process. 20 (2) (2006) 308–331.
[121] H. Endo, R.B. Randall, Application of a minimum entropy deconvolution filter to enhance autogressive model based gear tooth fault detection technique, Mech. Syst. Signal Process. 21 (2) (2007) 906–919.
[122] N. Sawalhi, R.B. Randall, H. Endo, The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis, Mech. Syst. Signal Process. 21 (2007) 2616–2633.
[123] H. Qiu, J. Lee, J. Lin, G. Yu, Robust performance degradation assessment methods for enhanced rolling element bearing prognostics Jul., Adv. Eng. Inf. 17 (2003) 127–140.
[124] C.J. Li, J. Ma, Wavelet decomposition of vibrations for detection of bearing-localized defects, ND 30 (3) (1997) 143–149.
[125] B. Liu, S.-F. Ling, Q. Meng, Machinery diagnosis based on wavelet packets, J. Vib. Control 3 (January) (1997) 5–17.
[126] J. Lin, L. Qu, Feature extraction based on morlet wavelet and its application for mechanical fault diagnosis Jun., J. Sound Vib. 234 (2000) 135–148.
[127] A. Djebala, N. Ouelaa, N. Hamzaoui, Detection of rolling bearing defects using discrete wavelet analysis Nov., Meccanica 43 (3) (2007) 339–348.
[128] C. Junsheng, Y. Dejie, Y. Yu, Application of an impulse response wavelet to fault diagnosis of rolling bearings, Mech. Syst. Signal Process. 21 (2007) 920–929.
[129] J. Liu, W. Wang, F. Golnaraghi, K. Liu, Wavelet spectrum analysis for bearing fault diagnostics Jan., Meas. Sci. Technol. 19 (2008).
[130] K. Mori, N. Kasashima, T. Yoshioka, Y. Ueno, Prediction of spalling on a ball bearing by applying the discrete wavelet transform to vibration signals Jul., Wear 195 (1996) 162–168.
[131] Z.K. Peng, P.W. Tse, F.L. Chu, A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing Sep., Mech. Syst. Signal Process. 19 (2005) 974–988.
[132] P.W. Tse, Y.H. Peng, R. Yam, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—their effectiveness and flexibilities, J. Vib. Acoust. 123 (3) (2001) 303.
[133] S. Tyagi, Wavelet analysis and envelope detection for rolling element bearing fault diagnosis—a comparative study, Lonavla (2001).
[134] W. Su, F. Wang, H. Zhu, Z. Zhang, Z. Guo, Rolling element bearing faults diagnosis based on optimal Morlet wavelet filter and autocorrelation enhancement Jul., Mech. Syst. Signal Process. 24 (2010) 1458–1472.
[135] I.K. Fodor, A Survey of Dimension Reduction Techniques, 2002.
[136] C. Bunks, D. Mccarthy, T. Al-Ani, Condition-based maintenance of machines using hidden Markov models Jul., Mech. Syst. Signal Process. 14 (4) (2000) 597–612.
[137] D. Lin, V. Makis, Recursive filters for a partially observable system subject to random failure, Adv. Appl. Probab. 35 (1) (2003) 207–227.
[138] P.-J. Vlok, M. Wnek, M. Zygmunt, Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions Jul., Mech. Syst. Signal Process. 18 (2004) 833–847.
[139] Q. Sun, Y. Tang, Singularity analysis using continuous wavelet transform for bearing fault diagnosis Nov., Mech. Syst. Signal Process. 6 (2002) 1025–1041.
[140] Z.K. Peng, F.L. Chu, P.W. Tse, Singularity analysis of the vibration signals by means of wavelet modulus maximal method Feb, Mech. Syst. Signal Process. 21 (2) (2007) 780–794.
[141] J. Altmann, J. Mathew, Multiple band-pass autoregressive demodulation for rolling-element bearing fault diagnosis Sep., Mech. Syst. Signal Process. 15 (5) (2001) 963–977.
[142] S. Tyagi, A comparative study of SVM classifiers and artificial neural networks application for rolling element bearing fault diagnosis using wavelet transform preprocessing, World Acad. Sci. Eng. Technol. 43 (2008) 309–317.
[143] C. Pachaud, R. Salvetas, C. Fray, Crest factor and Kurtosis contributions to identify defects inducing periodical impulsive forces, Mech. Syst. Signal Process. 11 (6) (1997) 903–916.
[144] R. Yan, R.X. Gao, Complexity as a measure for machine health evaluation, IEEE Trans. Instrum. Meas. 53 (4) (2004) 1327–1334.
[145] Z.K. Peng, F.L. Chu, Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography, Mech. Syst. Signal Process. 18 (2004) 199–221.
[146] K. Shibata, A. Takahashi, T. Shirai, Fault diagnosis of rotating machinery through visualisation of sound signals Mar., Mech. Syst. Signal Process. 14 (2) (2000) 229–241.
[147] X. Chiementin, F. Bolaers, J.-P. Dron, Early detection of fatigue damage on rolling element bearings using adapted wavelet, J. Vib. Acoust. 129 (4) (2007) 495.
[148] E.L. Da Costa, Detection and Identification of Cyclostationary Signals, 1996.
[149] W.A. Gardner, The spectral correlation theory of cyclostationary time-series, Signal Process. 11 (1986) 13–36.
[150] W.B. Collis, P.R. White, J.K. Hammond, Higher-order spectra: the bispectrum and trispectrum, Mech. Syst. Signal Process. (1998) 12 (3) (1998) 375–394.

[151] B. Yang, D. Lim, A. Tan, VIBEX: an expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table May, Expert Syst. Appl. 28 (2005) 735–742.
[152] C.K. Mechefske, Objective machinery fault diagnosis using fuzzy logic Nov., Mech. Syst. Signal Process. 12 (1998) 855–862.
[153] M.J. Roemer, C. Hong, S.H. Hesler, Machine health monitoring and life management using finite-element-based neural networks, J. Eng. Gas Turbines Power 118 (1996) 830–835.
[154] E.C. Larson, D.P. Wipf, B.E. Parker, Gear and bearing diagnostics using neural network-based amplitude and phase demodulation, in: Proceedings of the 51st Meeting of the Society for Machinery Prevention Technology, 1997, pp. 511–521.
[155] J.K. Spoerre, Application of the cascade correlation algorithm (CCA) to bearing fault classification problems Mar., Comput. Ind. 32 (1997) 295–304.
[156] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features Mar., Mech. Syst. Signal Process. 17 (2003) 317–328.
[157] B.A. Paya, I.I. Esat, M.N.M. Badi, Artificial neural network based fault diagnostics of rotating machinery using wavelet transforms as a preprocessor, Mech. Syst. Signal Process. 11 (5) (1997) 751–765.
[158] B. Li, M. Yuen Chow, Y. Tipsuwan, J.C. Hung, Neural-network-based motor rolling bearing fault diagnosis, IEEE Trans. Ind. Electron. 47 (5) (2000) 1060–1069.
[159] C. Wang, G. Too, Rotating machine fault detection based on HOS and artificial neural networks, J. Intell. Manuf. 13 (2002) 283–293.
[160] R.M. Tallam, T.G. Habetler, R.G. Harley, Self-commissioning training algorithms for neural networks with applications to electric machine fault diagnostics Nov., IEEE Trans. Power Electron. 17 (6) (2002) 1089–1095.
[161] B.-S. Yang, T. Han, Y.-S. Kim, Integration of ART-Kohonen neural network and case-based reasoning for intelligent fault diagnosis Apr., Expert Syst. Appl. 26 (2004) 387–395.
[162] T.I. Liu, J.H. Singonahalli, N.R. Iyer, Detection of roller bearing defects using expert system and fuzzy logic Sep., Mech. Syst. Signal Process. 10 (1996) 595–614.
[163] S. Zhang, T. Asakura, X.L. Xu, B.j. Xu, Fault diagnosis system for rotary machines based on fuzzy neural networks, JSME Int. J., Ser. C: Mech. Syst. Mach. Ele. Manuf. 46 (2003) 1035–1041.
[164] E. Jantunen, Intelligent monitoring and prognosis of degradation of rotating machinery, in: Proceedings of the Intelligent Maintenance Systems IMS'2004, 2004.
[165] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms Mar., Mech. Syst. Signal Process. 16 (2–3) (2002) 373–390.
[166] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine Oct., Mech. Syst. Signal Process. 21 (7) (2007) 2933–2945.
[167] K.C. Gryllias, I.a. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments Mar., Eng. Appl. Artif. Intell. 25 (2) (2012) 326–344.
[168] X. Li, A. Zheng, X. Zhang, C. Li, L. Zhang, Rolling element bearing fault detection using support vector machine with improved ant colony optimization Oct., Measurement 46 (2013) 2726–2734.
[169] Z. Liu, H. Cao, X. Chen, Z. He, Z. Shen, Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings Jan., Neurocomputing 99 (2013) 399–410.
[170] Y. Yang, D. Yu, J. Cheng, A fault diagnosis approach for roller bearing based on IMF envelope spectrum and SVM Nov., Measurement 40 (2007) 943–950.
[171] L. Guo, J. Chen, X.i. Li, Rolling bearing fault classification based on envelope spectrum and support vector machine Jul., J. Vib. Control 15 (9) (2009) 1349–1363.
[172] Q. Hu, Z. He, Z. Zhang, Y. Zi, Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble Feb., Mech. Syst. Signal Process. 21 (2) (2007) 688–705.
[173] Y. Pan, J. Chen, L. Guo, Robust bearing performance degradation assessment method based on improved wavelet packet—support vector data description Apr., Mech. Syst. Signal Process. 23 (2009) 669–681.
[174] Q. Miao, V. Makis, Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models Feb., Mech. Syst. Signal Process. 21 (2) (2007) 840–855.
[175] T. Xinmin, D. Baoziang, X. Yong, Bearings fault diagnosis based on HMM and fractal dimensions spectrum, in: Proceedings of the 2007 IEEE International Conference on Mechatronics and Automation, 2007, pp. 1671–1676.
[176] X.-Hui Zhang, J.-She Kang, Hidden Markov Models in bearing fault diagnosis and prognosis, in: Second International Conference on Computational Intelligence and Natural Computing (CINC), 2010, pp. 364–367.
[177] D.C. Baillie, J. Mathew, Non-linear model-based fault diagnosis of bearings, Proc. Cond. Monit. (1994) 241–252.
[178] A. Vania, P. Pennacchi, Experimental and theoretical application of fault identification measures of accuracy in rotating machine diagnostics Mar., Mech. Syst. Signal Process. 18 (2004) 329–352.
[179] D. Söffker, M.-sami Saadawia, C. Wei, Model- and feature-based diagnostics in rotating machinery, in: Proceedings of the XV International Symposium on Dynamic Problems of Mechanics, 2013.
[180] F.V. Nelwamondo, T. Marwala, U. Mahola, Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, Mel-frequency cepstral coefficients and fractals, Int. J. Innovative Comput. Inf.Control 2 (6) (2006) 1281–1299.
[181] S.J. Engel, B.J. Gilmartin, K. Bongort, A. Hess, Prognostics, the real issues involved with predicting life remaining, IEEE Aerosp. Conf. (2000) 457–469.
[182] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, H. Liao, Intelligent prognostics tools and e-maintenance Aug., Comput. Ind. 57 (6) (2006) 476–489.
[183] A. Heng, S. Zhang, A.C.C. Tan, J. Mathew, Rotating machinery prognostics: state of the art, challenges and opportunities Apr., Mech. Syst. Signal Process. 23 (3) (2009) 724–739.
[184] Y. Peng, M. Dong, M.J. Zuo, Current status of machine prognostics in condition-based maintenance: a review Jan., Int. J. Adv. Manuf. Technol. 50 (1–4) (2010) 297–313.
[185] N.S. Jammu, P.K. Kankar, A review on prognosis of rolling element bearings, Int. J. Eng. Sci. Technol. 3 (10) (2011) 7497–7503.
[186] M.A. Schwabacher, A Survey of Data-Driven Prognostics, no. September, pp. 1–5, 2005.
[187] M. Schwabacher, K. Goebel, A survey of artificial intelligence for prognostics, NASA Ames Research Center (2006) 107–114.
[188] F. Camci, K. Medjaher, N. Zerhouni, P. Nectoux, Feature evaluation for effective bearing prognostics Mar., Qual. Reliab. Eng. Int. (2012).
[189] C.D. Begg, T. Merdes, C. Byington, K. Maynard, Dynamics modeling for mechanical fault diagnostics and prognostics, in: Mechanical System Modeling for Failure Diagnosis and Prognosis, Maintenance and Reliability Conference, no. Marcon 99, 1999.
[190] C.D. Begg, C.S. Byington, K.P. Maynard, Dynamic simulation of mechanical fault transition, in: Proceedings of the 54th Meeting of the Society for Machinery Failure Prevention Technology, 2000, pp. 203–212.
[191] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, S. Liang, Adaptive prognostics for rolling element bearing condition Jan., Mech. Syst. Signal Process. 13 (1) (1999) 103–113.
[192] Y. Li, T. Kurfess, Y. Liang, Stochastic prognostics for rolling element bearings, Mech. Syst. Signal Process. 14 (5) (2000) 747–762.
[193] Y. Shao, K. Nezu, Prognosis of remaining bearing life using neural networks, Proc. Inst. Mech. Eng. Part I J. Syst. Control Eng. 214 (2000) 217–230.
[194] A.K. Mahamad, S. Saon, T. Hiyama, Predicting remaining useful life of rotating machinery based artificial neural network Aug., Comput. Math. Appl. 60 (2010) 1078–1087.
[195] P. Wang, G. Vachtsevanos, Fault prognosis using dynamic wavelet neural networks, AI EDAM—Artif. Intell. Eng. Des. Anal. Manuf. 15 (2001) 349–365.
[196] T. Khawaja, G. Vachtsevanos, B. Wu, Reasoning about uncertainty in prognosis: a confidence prediction neural networks approach, in: IEEE Annual meeting of the North American Fuzzy information Processing society, 2005, pp. 7–12.

[197] Y.-lung Dong, W. J. I. O. N. G. Gu, K.U.N. Yang, W.-kun Zhang, A combining condition prediction model and its application in power plant, in: Proceedings of the Third International Confenrence on Machine Learning and Cybernetics, 2004, no. August, pp. 26–29.

[198] N. Gebraeel, M. Lawley, R. Liu, V. Parmeshwaran, Residual life predictions from vibration-based degradation signals: a neural network approach Jun, IEEE Trans. Ind. Electron. 51 (3) (2004) 694–700.

[199] N.Z. Gebraeel, M.A. Lawley, A Neural network degradation model for computing and updating residual life distributions, IEEE Trans. Autom. Sci. Eng. 5 (1) (2008) 154–163.

[200] R.C.M. Yam, P.W. Tse, L. Li, P. Tu, Intelligent predictive decision support system for condition-based maintenance Feb., Int. J. Adv. Manuf. Technol. 17 (2001) 383–391.

[201] J. Qiu, C. Zhang, B.B. Seth, S.Y. Liang, Damage mechanics approach for bearing lifetime prognostics Sep., Mech. Syst. Signal Process. 16 (2002) 817–829.

[202] K. Goebel, P. Bonanni, N. Eklund, Towards an integrated reasoner for bearings prognostics, IEEE Aerosp. Conf. (2005) 3647–3657.

[203] K. Goebel, N. Eklund, P. Bonanni, Fusing competing prediction algorithms for prognostics, IEEE Aerosp. Conf. (2006).

[204] B. Zhang, C. Sconyers, M. Orchard, R. Patrick, G. Vachtsevanos, Fault progression modeling: an application to bearing diagnosis and prognosis, Am. Control Conf. (2010) 6993–6998.

[205] W. Wang, A model to predict the residual life of rolling element bearings given monitored condition information to date, J. Manage. Math. 13 (2002) 3–16.

[206] E. Myötyri, U. Pulkkinen, K. Simola, Application of stochastic filtering for lifetime prediction Feb., Reliab. Eng. Syst. Saf. 91 (2) (2006) 200–208.

[207] W.Q. Wang, M.F. Golnaraghi, F. Ismail, Prognosis of machine health condition using neuro-fuzzy systems Jul., Mech. Syst. Signal Process. 18 (2004) 813–831.

[208] R.B. Chinnam, P. Baruah, A neuro-fuzzy approach for estimating mean residual life in condition-based maintenance systems, Int. J. Mater. Prod. Technol. 20 (1–3) (2004) 166–179.

[209] W. Wang, An adaptive predictor for dynamic system forecasting Feb., Mech. Syst. Signal Process. 21 (2007) 809–823.

[210] J. Liu, W. Wang, F. Golnaraghi, A multi-step predictor with a variable input pattern for system state forecasting Jul., Mech. Syst. Signal Process. 23 (2009) 1586–1599.

[211] W. Wang, D.Z. Li, J. Vrbanek, An evolving neuro-fuzzy technique for system state forecasting Jun., Neurocomputing 87 (2012) 111–119.

[212] D. Li, W. Wang, F. Ismail, Enhanced fuzzy-filtered neural networks for material fatigue prognosis Jan., Appl. Soft Comput. 13 (2013) 283–291.

[213] R. Orsagh, M. Roemer, J. Sheldon, C.J. Klenke, A comprehensive prognostics approach for predicting gas turbine engine bearing life, Proc. IGTI Turbo Expo (2004).

[214] G.J. Kacprzynski, A. Sarlashkar, M.J. Roemer, Predicting remaining life by fusing the physics of failure modeling with diagnostics, J. Met. 56 (2004) 29–35.

[215] S. Marble, B.P. Morton, Predicting the remaining life of propulsion system bearings, in: Proceedings of the 2006 IEEE Aerospace Conference, 2006.

[216] N. Lybeck, S. Marble, B. Morton, Validating prognostic algorithms: a case study using comprehensive bearing fault data, in: 2007 IEEE Aerospace Conference, 2007.

[217] D. Chelidza, J.P. Cusumano, A dynamical systems approach to failure prognosis, J. Vib. Acoust. 126 (2004).

[218] J. Yan, M. Koç, J. Lee, A prognostic algorithm for machine performance assessment and its application Dec., Prod. Plann. Control 15 (8) (2004) 796–801.

[219] H.T. Pham, B.-S. Yang, Estimation and forecasting of machine health condition using ARMA/GARCH model Feb., Mech. Syst. Signal Process. 24 (2010) 546–558.

[220] R. Xu, S.Y. Liang, L. Haynes, An integrated approach to bearing fault diagnostics and prognostics, Proc. Am. Control Conf. (2005) 2750–2755.

[221] R. Huang, L. Xi, X. Li, C. Richard Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods Jan., Mech. Syst. Signal Process. 21 (2007) 193–207.

[222] M. Dong, D. He, A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology Jul., Mech. Syst. Signal Process. 21 (2007) 2248–2266.

[223] D.Z. Li, W. Wang, An enhanced GA technique for system training and prognostics Jul., Adv. Eng. Softw. 42 (2011) 452–462.

[224] B.-S. Yang, A. Widodo, Support vector machine for machine fault diagnosis and prognosis, J. Syst. Des. Dyn. 2 (1) (2008) 12–23.

[225] A. Widodo, et al., Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Syst. Appl. 36 (2009) 7252–7261.

[226] A. Widodo, B.-S. Yang, Machine health prognostics using survival probability and support vector machine Jul., Expert Syst. Appl. 38 (2011) 8430–8437.

[227] E. Sutrisno, H. Oh, A.S.S. Vasan, M. Pecht, Estimation of remaining useful life of ball bearings using data driven methodologies, in: IEEE Conference on Prognostics and Health Management, Jun. 2012.
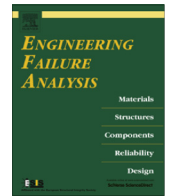
CrossMark

A descriptive model of wear evolution in rolling bearings

Idriss El-Thalji *, Erkki Jantunen

Industrial Systems, VTT Technical Research Centre of Finland, Finland

A R T I C L E   I N F O

A B S T R A C T

Rolling contact wear is a complex phenomenon that might involve different wear mechanisms (adhesive, abrasive, fatigue and corrosive) and different stress concentration mechanisms (asperity, dent, debris, inclusions, etc.). The interactions among these mechanisms might accelerate or decelerate the overall wear progress. Therefore, it is complicated to model and monitor the fluctuations of wear progress. The current descriptive models are either describing individual physical phenomena within rolling contact wear or describing a specific stage of wear progress. Thus, the interactions among different wear mechanisms and the transition events among different stages of wear progress are not sufficiently addressed. Therefore, the purpose of this paper is to propose a descriptive model of the wear evolution process in rolling bearings over the whole lifetime. The descriptive model utilises a wide range of empirical findings in the literature to describe the wear interactions and evolution in the five-stage scenario: running-in, steady-state, defect initiation, defect propagation, and damage growth. The new descriptive model provides the most probable scenario of wear evolution in rolling bearings, which is useful for modelling and monitoring the wear progress. It illustrates the wear evolution stages, the involved wear mechanisms in each stage, the interaction among wear mechanisms in each stage, the surface topology changes and the influencing factors within each stage. For design, condition monitoring and prognosis purposes, these aspects are significantly important to understand, model, test and monitor the wear evolution process.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The rolling element bearing (REB) is one of the most critical components that determine the machinery health and its remaining lifetime in modern production machinery. REBs are all wearing components and inevitably the REB health will degrade over the lifetime. Therefore, there is significant interest in the whole wear evolution progress over the REB's lifetime, to gain the benefits for REB design and condition monitoring purposes. Over the years, the interest in exploring the wear evolution has rapidly increased. The determination of transition point between mild and severe wear was the starting point. Mild wear is considered as acceptable wear state whereas the transition to severe conditions often represents a change to commercially unacceptable situations [1]. Therefore, the wear maps were developed [2–5]. However, there are more information about wear evolution needs to be revealed than just single transition point. Therefore, Voskamp [6] described the material degradation due to rolling contact fatigue as a three-stage process: shakedown, steady-state elastic response, and instability. These stages are corresponding to what is known as three-stage model: running-in, steady-state and damage stages of wear progress. However, the three stages model represents roughly the evolution of rolling contact wear. A number

---

of complex damage phenomena are taking place in each of these stages and require better understanding. The result of these complex damage phenomena is the probability to produce different wear progression scenarios. The wear progression scenario is significantly important to predict the length and propagation rate of each stage and to estimate the remaining lifetime.

Therefore, more studies have been conducted to explore the detailed phenomena of wear in rolling contact. In fact, a number of researchers have specifically studied the evolution of surface deterioration due to wear and material response over the lifetime, as listed in Table 1. Different wear progression scenarios are described in the literature. Longching et al. [7] and Raje et al. [8] described the wear progression of subsurface inclusions. Gao et al. [9] and Mota [10] investigated the wear progression of dented surfaces. These models illustrate several defect initiation and propagation aspects: determine the location of first micro-crack, direction of propagation, and crack branching and propagation. Other studies illustrate the complex damage phenomena such as dentation, stress concentration, distress, micro-cracking, adhesive and abrasive actions. In Table 1, a summary of wear evolution studies are presented. The direct and indirect measurements, i.e. under the 'point of view' column, are also illustrated in Table 1. Direct measurement refers to the experimental validation methods which deal directly with the damaged surface such as surface topography. Indirect measurement refers to experimental validation method which utilise condition monitoring techniques such as vibration, acoustic emission, and oil-debris.

The indirect measurement studies describe the overall dynamic impacts of wear evolution process. Therefore, El-Thalji and Jantunen [11] discussed how the indirect measurement methods could help to monitor he wear evolution process. Jantunen [12] and Yoshioka and Shimizu [13] could describe the wear evolution process based on the indirect measurements. It was observed two main stages of wear progress: steady state and instability. The steady-state stage is roughly stable and a clear offset in the root mean square (RMS) values of measured signals is observed at instability stage, later a rapid increase of these values occurs before the final failure. Dempsey et al. [14] observed that no debris is detected (using filter of 10 μm) in the early stage (around 25%) of testing time. All spall lengths and wear volumes increase sharply at the beginning and then take stable increment state before it increase sharply again. However, it is still hard either to describe the whole wear evolution process based on the direct measurements or to provide a reliable speculation based on the indirect measurements.

In fact, the models in Table 1 give great insight into defect initiation and propagation in REBs over the lifetime and from different points of views such as material microstructure [16,17,19,8], stress wave response [21], dynamic response [12] and debris particles [18]. However, these models consider (1) either subsurface wear progression or surface wear progression, (2) a single specific damage stage e.g. crack initiation or crack propagation, (3) a single wear mechanism e.g. fatigue wear, abrasive wear, and (4) a single stress concentration mechanism e.g. dent, asperity, debris, inclusion. Thus, the interactions among different wear mechanisms and the transition events among different stages of wear progress are not sufficiently addressed.

Fortunately, in the literature, these complex issues have been separately investigated by several experimental studies i.e. related to the abrasive wear and rolling contact fatigue. In summary, different studies have developed several descriptions and models to explain the wear in rolling contact. These studies have investigated the wear mechanisms i.e. fatigue, adhesive, abrasive and corrosive. The studies have mainly investigated the effect of different stress concentration mechanisms e.g. asperity and dent [24–26], debris [27–31], inclusions [32–35]. Moreover, a wide range of studies i.e. as classified in Table 2 have investigated several issues of wear behaviour: lubrication regimes i.e. elasto-hydrodynamic (EHL), debris, denting, micro-cracking, asperity formation, subsurface crack formation, crack propagation, adhesive wear, abrasive wear, pitting, spalling.

However, these studies explain separately several individual physical phenomena. In reality, almost every wear process in REBs involves all these aspects and their evolutions with different rates and frequencies, depending on loading and operating conditions [36–38]. Therefore, the wear progress is a nonlinear physical phenomenon. In fact, even though having a good

Table 1
Summary of wear evolution studies in rolling contact.

Reference based on date of publication	Running-in stage	Debris involvement	Dentation	Micro cracking		Pitting	Abrasive wear	Crack propagating		Spalling	Point of view	Validation
				Surface	Subsurface			Surface	Subsurface			
[6]	X				X				X	X	M	X
[15]			X		X			X		X	M	X
[16]					X				X	X	M	X
[9]			X	X		X		X		X	M	X
[17]			X		X				X	X	M	X
[18]		X	X	X				X			P	
[10]			X	X			X	X		X	M	X
[12]		X				X	X			X	D,P,S	X
[19,20]			X	X				X			M	X
[21]			X			X				X	D,P,S	X
[22]			X	X						X	M	
[8]					X				X	X	M	
[14]			X	X				X		X	P	X
[23]			X			X				X	D	X

M: material, D: dynamic, S: stress wave, P: debris-particles.

Table 2

Classified literature survey of wear in rolling bearings.

Wear issue	Reference(s)
General wear	[44–51]
Adhesive wear	[52–54]
Abrasive wear	[54–60,1,61–67,1,68–73]
Fatigue wear	[74–78,39,79–82]
Stress modelling and RCF	[83–90,22,91,92]
Multi-axial fatigue criteria	[93–97]
Crack initiation	[32–35,98–115]
Crack propagation	[116–127]
Crack propagation modelling	[128–136,79,137–141,43]
Damage mechanics	[142–147]
Surface roughness and asperity effect	[24–26,128,129,148–151]
Debris formation	[152–154]
EHL contact	[155–167]
EHL-rolling contact, rolling friction	[168–174]
Lubrication regimes	[175,176]
EHL modelling	[177–179]
Hydrodynamic contact	[180–184,156]
Elastic–plastic contact	[185,186]
Visco-elastohydrodynamic	[187–189]
Starvation effect	[190,191]

understanding of specific wear aspects is needed, it might not be sufficient to understand the whole wear evolution process. A good understanding of wear evolution requires to define how the wear is progressing over the lifetime [39]. Therefore, the interactions among the wear and stress concentration mechanisms have been discussed to explore how these mechanisms accelerate each other or compete with each other [39]. For example, it is widely observed that the tendency for the spalled area to become worn is increasing over the time [40], however, the measured impacts due spall might be much smaller in certain time interval than in the early stages. This phenomenon is called wear competition and explained physically in few studies [41–43] based on the interaction of abrasive wear and over-rolling action with surface asperities, that leads to surface smoothing. Moreover, the studies in the literature have partially illustrated the evolution of rolling contact wear.

In fact, there are slightly large differences between the simulated and the actual results in most of the studies that focus on wear progression over the specific lifetime [40]. Up to date, it is hard to describe the wear evolution progress due to the variety of the involved wear and contact mechanisms, which might produce different wear evolution scenarios. Therefore, there is a need for a descriptive model that is able to address the wear evolution over the whole REB's lifetime with the above mentioned complexity. The purpose of the paper is to develop a new descriptive model of wear interaction and evolution which combines and integrates the large experimental and numerical findings that have been accumulated in the literature. The descriptive model aims to generalise the most probable wear evolution scenario in REBs. First, it explains the interaction of multiple wear mechanisms in rolling contact. Second, it explains both the wear evolution on and beneath the contact surface from running-in stage until spalling occurs. Third, it explains the wear transition points during the wear evolution progress. Therefore, the paper describes the wear evolution in rolling bearings over the whole lifetime in terms of the wear progression stages, surface topology evolution, wear mechanics interaction and influence factors of wear progression.

2. The descriptive model of wear evolution in rolling bearings

The descriptive model is developed with the support of experimental findings published in the literature listed in Table 2. It illustrates the wear evolution progress based on five stages. First, in each stage, it illustrates the potential involvement of different wear mechanisms (adhesive, abrasive, fatigue and corrosive) and different stress concentration mechanisms (asperity, dent, debris, inclusions, etc.). Second, different mechanisms require definitely the consideration of the interactions among them, which helps to address the non-linear evolution of wear in REBs. Third, the descriptive model illustrates the transitions events between the stages, since it covers the whole lifetime. The new descriptive model is presented in detail in the following sub-sections with the help of experimental findings in the literature.

However, before the presentation of the descriptive model, two issues are worth to be highlighted. First, there are many industrial applications where wear behave differently, even though, the rolling contact in common [39] e.g. hard steels (i.e. REB) and soft steels (i.e. railway track). The multi-axial stresses of compressive type are also common. However, the distinguished factor is the friction coefficient, whether it is low or high. In the case of REB, the lubrication provides low friction rolling contact result in low friction coefficient and elastic deformation. The railway rolling contact is characterized with high friction coefficient and plastic deformation. Such differences in the wear phenomena generate difference in the applied stress regime, defect origin, defect propagation stages and other affecting factors. The wear in REB originates from surface asperities and subsurface inclusions. The crack propagates through inclined short growth in REB, in compared to the inclined long crack growth type in railway track case. The most significant factors that affect the wear in REB are: the slide-to-roll ratio,

hoop stress, surface roughness and operating environment effects (i.e. lubricant, hydrogen, and debris). It is widely accepted that pits propagate when fluid is present due to fluid pressurization and/or fluid entrapment.

Second issue is related to the system where the REB is allocated. Mukras et al. [51] observed that wear prediction of the component without considering the system as whole will lead in most of cases to inaccurate predictions. The formulas and models that are developed to describe wear mechanisms are valid under specific assumptions. For example, adhesive wear is described based on sliding actions. However, it is very hard to estimate the sliding events in the rolling contact as the pure rolling is dominant. Morales-Espejel and Brizmer [92] indicated also that low speed applications have thinner film thicknesses. There are a number of critical aspects have been studied to enhance the wear understanding with respect to the rolling bearing as a system. The considerations are mainly related to the operating conditions and system configurations: thermal effect [192–198], machine-induced and process-induced residual stress effect [199–202], wear particles and debris contamination [27–31,203], interference fits, environment and component configuration [204–207], cleanliness, filtration [208,209], large scale effect [210–212]. Therefore, it is really important to be aware of these characteristics of the rolling contact in REBs.

2.1. Running-in stage

In general, the REBs have some degrees of surface waviness and surface roughness due to the surface manufacturing processes. Therefore the surface asperities have potential to interact with each other in running-in stage. Arakere and Subhash [213] explained that running-in stage involves localized microplastic flow, work hardening and shakedown. Therefore, the residual stresses are induced with increase in material strengthening and microyield stress. Surface properties during the running-in stage have great importance and influence on lubricant film, friction and state of vibration. There is a non-linear relationship between contact deformation and force at the beginning due to the surface asperities [213]. However, running-in contact cycles smoothen the surface asperities and the lubrication film become uniform over time. The subsurface hardness increases during first stage of cyclic loading (shakedown response) [213]. However, the surface hardness increase until the work hardening saturates which means the steady-state stage started. Andersson [214] states that interacting surfaces show only some mild wear during the running-in phase followed by a nearly no wear phase, which indicate the steady-stage stage. The interaction of major asperities during running-in stage might be detected with help of vibration measurements, however, it might not be clearly appeared in the presence of lubricant additives [215].

2.2. Steady state stage

In the steady-state stage, the response is steady state due to uniform Hertzian stresses, uniform lubrication film, formation of dark etching region and white etching bands and texture strengthening. Arakere and Subhash [213] mentioned that the length of steady-state stage is a function of maximum load induced stress, material characteristics and operating temperature. The steady state is well known as the healthy stage of the lifetime of any component. Therefore, it is significantly important to understand their physical outcomes in order to detect the abnormalities once the wear progress reaches the instability stage. The relationship between RMS vibration values in high frequency band and lambda (λ) factor (i.e. film thickness/surface roughness) is observed [216]. However, the detection of defects is more difficult in real cases due to the high and variable influences of operating parameters compare to experimental testing environment. Massouros [217] observed shifts in the vibration frequencies in the range of 0.2–10 kHz due to running-in process. The vibration progress due to the wear process in running-in stage has high amplitudes at the shaft's rotational frequency during the first interval of testing time, later the vibration amplitudes are reduced due to the smoothening effect of the wear process [218]. The mild and severe wear can be distinguished in terms of the operating conditions and the involvement of the fundamental wear mechanisms [214]. The effect of operating conditions and different damage scenarios of rolling contact wear can be observed in several studies, due to differences in lubrication type and film thickness [10].

2.3. Defect initiation

In reality, the REBs have quite long lifetime. Most of the bearing lifetime equations [219] show that the lifetime is an exponential function of loading i.e. approximately exponential of 3. This means, if the load increases, the life decreases with approximate three times. The loading might increase due to increasing in applied load or some changes in the contact surface, which concentrate more loads at specific locations rather than keeping the designed distribution of the contact loads. In fact, Dizdar and Andersson [220] indicated that the properties of the boundary layers and the changes in contact conditions are much more important factors than the degree of plastic deformation. Refereeing to bearing lifetime equations, Ioannides et al. [219] showed that the evolution of the equation over the time and the importance of estimating accurate life adjustment factors e.g. reliability factor, material factor and operating condition factor, are highly influencing the whole bearing lifetime. The literature e.g. [27,28,204,205] shows that most of the causes of the wear phenomena have a random and distributed nature such as surface roughness, waviness, debris, and inclusions.

Therefore, the descriptive model explains what is taking place in the end of the steady state stage and how defect start to appear on and under the surface. It explains how the distributed defects initiate and/or influence the localized defects. The new descriptive model considers five stress localization phenomena to be the responsible mechanisms for raising the cyclic

stresses at specific locations in the REBs. These five phenomena are extracted from the literature and related to the contact feature of REBs i.e. running track, roller profile. Furthermore, the descriptive model considers three fundamental issues i.e. contamination, lubrication disturbances and impact vibrations. These issues are significantly important as external mechanisms that have additional contributions in the defect initiation progress.

2.3.1. Defect localization

The five phenomena that are responsible to raise the cyclic stresses highlight the highly potential locations of wear. First, it is observed that defect is most probably located in the loading contact zone of the REBs. Second, the investigations of highest stress location show that the normal running track of the rolling elements is one of most probably location of wear defects. It is observed that the normal running track of the rolling elements is usually located on the pure-rolling points (zero sliding points). The spall is usually located at the normal running track of the rolling elements, but also there are more severe spalls at the edges of the normal running track of the rolling elements [221]. Under abnormal operation e.g. shaft imbalance and misalignment, or other patterns of high stresses can influence the defect localization process.

Third, the Heathcote slip phenomenon might explain the zones of greater damage which is located mainly outside the normal running track. Kotzalas and Doll [221] defined the heath-coat slip as a geometrical constraint suffered by all spherical roller bearings, as illustrated in Fig. 1, if surface velocities between the inner ring and the rollers match at locations R1 and R3, then the surface velocities must differ at location R2, which means that there is sliding between the roller and the centre of the raceway. Therefore, the inner raceway might have micro-pit as the slower velocity component in the contact. Olofsson et al. [222] found that wear in the area outside the zero sliding points is less than in the area of the zero sliding points during the early life of a bearing. In the later stage during long-term wear process, the wear is almost the same in both areas. Nilsson et al. [72] studied later the abrasive wear of spherical roller and observed a trend of worn coating depth with respect to roller profile as shown by the two schematics curves in Fig. 1. Furthermore, Massi et al. [223] concluded with help of the finite element modelling (FEM) that a concentration of the local stress distribution is located at the boundaries of the contact between ball and race. Moreover, a local sliding is also predicted in this area, with a consequent large value of the shear stress. The severe indentation damage which was observed relatively outside the running track has same space distance as between rolling elements.

Fourth, some empirical observations show that defect locations might have specific spacing [224]. The specific spacing between multiple defects is due to static false-brinelling occurred at stand-still events or dynamic impact loads. One of the reason behind the false-brinelling is due to the fact that EHL contact behave as dry contact in case of non-rotating situations of bearing [224]. The false-brinelling might initiated defects as same as denting by debris. It is also observed that the false-brinelling defects allocate usually some debris particles.

Fifth, the localization possibility might be due to a subsurface defect. In fact, the higher the sliding imposed on the EHL contact, the higher the maximum shear stresses in the subsurface due to the frictional force [92]. Thus, phase transformations might produce 'butterflies' and 'dark and white etching' regions. It might not be critical for the crack initiation but it can involve some degree of strain localization, however, it leads to accumulation of plastic strain in the local region. It will be continually accompanied with changes in residual stresses and local mechanical properties such as phase transformations, surface intrusions and extrusions.
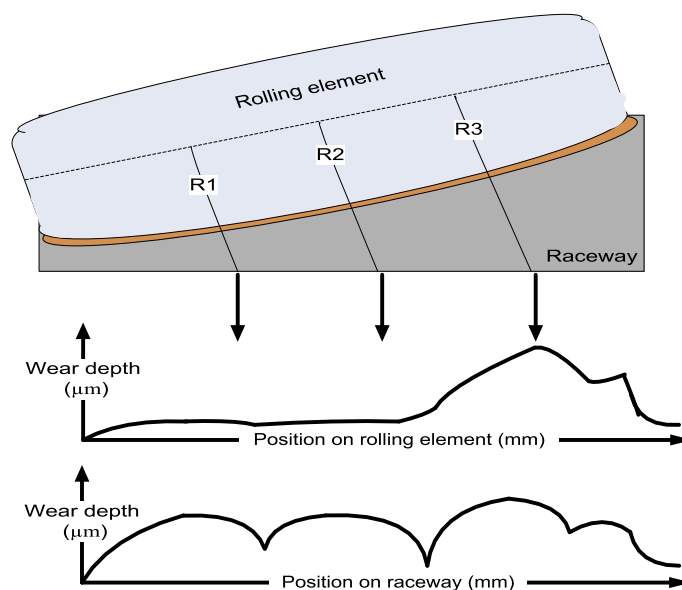

Fig. 1. Physical description of Heathcote slip phenomenon in a spherical roller bearing.

These five phenomena explain where the defect might be potentially localized. Besides that, some experimental observations in the literature help us to explain how dentation might happen. Dentation might occur due to: (1) debris particles which enter the contact zone, (2) lubrication disturbances and (3) impact vibrations.

Contamination particles might be naturally produced within running-in and in latest stage of steady state stage. Contamination particles of lubricant oil is one of the main reason of distributed defects [225]. It is observed that when solid hard particles go into the contact region, surface dentation is inevitable. Contaminations might generate non-linear film thickness and stress concentration. Basically, when particles of sizes (5–20 μm) larger than the oil film thickness pass through the contact region, localized pressure peaks have greater chance to occur in the contact region. However, it is important to notice that when the particle size is large, there is less chance to enter the contact region [225]. The grease of small thickener particles (that enter the contact region) will increase the risk of pitting damage. However, the large particles can originate oil starvation in the inlet of the contact zone, once they are big or gathered. The thickener particles can cause deep elastic indentations accompanied by large pressure fluctuations and cause high stress concentration [226]. Hence, increased stress is expected. The pressure level will be doubled due to thickener particles entering the contact region. The pressures fluctuations due to large thickener particles are less significant compare to small ones, since the load is distributed over a larger area. Once the particles enter the contact region and the loads are high enough, the particles might be squashed into either platelets or fragments. It depends on particles ductility. Ductile particles cause smooth rounded and relatively shallow indents, while the brittle particles cause deep steep sided dents [18]. The dent size that is initiated by smashed particles is usually varied with depth range of 2.5–32 μm and 30–360 μm in width. Moreover, the dent has usually shoulders of 0.4–4 μm height. It is noted also that the surfaces indentations vary in depth and spatial extent. In general, the dent is analytically characterized by depth, diameter, distance between shoulders peaks and height of its shoulders [227].

The surface and operating imperfections (waviness, clearances, preload, etc.) disturb the lubrication film uniformity. The lubrication film transfers to the surface valleys. The peaks of wavy surfaces act as preload where the clearances between contact bodies become smaller. Therefore, boundary lubricated contact is more possible to exist between the rolling element and the race [225]. The stress concentrations are caused by local film thickness fluctuations and pressure ripples [92]. Therefore, the surface imperfections and EHL conditions can generate local film thickness fluctuations and pressure ripples, as causes for stress concentration. Moreover, the higher the sliding imposed on the EHL contact, the higher the maximum shear stresses in the subsurface [92]. The increasing correlation between the influence of surface layer thickness and viscosity on the operating load and friction coefficient is observed [158]. This observation is quite important for the applications which are running in conditions of high loads, low speeds and lack of lubricant, where insufficient hydrodynamic forces may enable the asperities on the opposing surfaces to interact. Halme and Andersson [228] illustrated the lubrication regime effects and their detectable symptoms of rolling bearing defects. In Table 2, the literature that illustrates different lubrication regimes is classified for more detailed information.

Vibrations due to sudden impact events might dent the contact surfaces [229]. Vibrations might generate high contact stresses which cause permanent plastic deformation of balls, raceways and lubricant degradation. Söchting et al. [229] highlighted that the vibration forces are much higher than the applied bearing preload, in which the bearing might momentarily offload and allow a gap between the rolling element and raceway to occur, this phenomenon named as gapping. The small bearing stiffness change result in relatively large variations in gapping. When gap (28–48 μm) is occurred, large dynamic loadings (shock or hammering) can result on the rolling element and raceway when they come into contact. The vibration force might produce gapping phenomenon normally in starting-up, launching [229], cut-off situations, and disturbances during operations. The lubricant type is significantly important in gapping phenomenon, it was experimentally noted that natural frequency of solid lubricant (grease) was 50–100 Hz higher than the recorded for the liquid lubricant (base oil) [230]. The solid lubricants produced powdery particles as a result of vibration forces, which also reduce the lubricant durability. Moreover, it is quite possible for bearings debris particles to be created during period of subjected vibration forces [229]. These debris particles could become entrained into the lubricant and may damage the bearing surface later.

Therefore, the surface properties have the great importance and influence on bearing endurance, lubricant film, friction and vibration. Vibrations caused by varying compliance arise in radially loaded bearings because of the non-linear relationship between contact deformation and forces. The spectrum of vibration contains bins at the rolling element passing frequency and its harmonics. Therefore, surface properties and their non-linear effects make the generated vibration of such contact is not strictly periodic and quite complex [231]. The vibration characteristics are highly dependent on the effects of geometrical imperfections (i.e. variation of roller diameters, inner ring waviness), abrasive and fatigue wear [232].

2.3.2. Dentation

The transition from running-in stage to steady-state stage represents the first transition event within the whole wear evolution progress. The second transition event is from the steady state to the defect initiation stage. In the previous section, five potential defect locations and four mechanisms of how the dentation might occur i.e. high stress, contamination, vibration, and lubrication disturbances were presented. The remaining question is when the dentation might start within the REB's lifetime? It is basically related to the yield stress limit (Y). This limit is required in order to generate a dentation by the applied forces. There are three theories predict when the dentation might occur. The first theory is related to contact mechanics. The other two dentation theories are related to the contamination and impact vibration mechanisms, and both are dependent on external initiator i.e. contamination in oil or vibration due to sudden impact events. Contact mechanics approach offers the possibility of treating the contact region as a continuous dynamic phenomenon. It is done either with

the restitution coefficient model or with the contact force approach. The contact force approach relies heavily on a mathematical model for the force-indentation and indentation rate relationship. Hertz theory predicts the stress distribution in the contact zone between two bodies having a surface of revolution [233]. The common force–indentation relation for the sphere to sphere contact is [233]:

$$F = K\delta^{\frac{3}{2}}$$

where F is the normal force pressing the solids together, δ is the deflection of the two spheres i.e. the total of deformation of both surfaces, K is a constant depending on the sphere radii and the elastic properties of the sphere materials. Beyond the elastic loading we consider two other stages i.e. elastic–plastic and fully plastic. In the elastic–plastic stage, the plastic deformation is small enough to be accommodated by an expansion of the surrounding area. As the load increases, the plastic zone grows and the displaced material flows to the sides of the indenter. For example, the sphere contact yield will initiate when the mean contact pressure is 1.1Y and the flow will become fully plastic at about 3.0Y [234].

2.3.3. Crack initiation

Once the surface is dented, the asperities act as the stress raisers. Thus, the crack initiation process will have the sufficient stress intensity factor to start the progress. The crack initiation process is considered as the core of the fatigue wear process. Fatigue wear process generates wear particles after repeated cycles of contact. It is related to whether the applied contact is high or low-cycle [76]. Lundgren and Palmgren [74] defined experimentally the critical number of rolling cycles for generation of wear particles. The formula depends on applied load and shape factor of rolling element. It is assumed that contact pressure is high enough to introduce yield in the contact region, however, it is not. The local yield is generated in the contact region because of the existence of material micro-defects e.g. inclusions, vacations, boundaries, asperities. Later, the fatigue wear progress will have several life cycle processes: incubation, formation, propagation, and removal process [77]. Sadeghi et al. [79] have reviewed and discussed fatigue wear due to subsurface inclusions in rolling contact and the bearing life prediction models i.e. probabilistic and deterministic. The material microstructure has influence on the fatigue wear, due to the inhomogeneous and random nature of material microstructure, related to randomly shaped, sized and oriented grains [79]. However, the descriptive model try to consider the effect of other mechanisms such as asperities, dents and debris on the whole fatigue wear, to reveal their roles in the wear evolution progress.

The wear progression of dented surfaces have investigated [15,9,10] to determine the location of first micro-crack. Several empirical observations of these studies are presented below to draw up the fundamental mechanism of crack initiation phenomenon that supports the descriptive model. Gao et al. [9] observed that the presence of a surface indentation can modify the EHL condition and the film thickness, which increase the pressure and stress concentration on the trailing edge of the defect. The defect starts with a crack on the surface near to the trailing edge of original dent, instead of the edge of the deformed dent [9]. Mota [10] observed the damaged area around the dents. The dent was found to be always larger for the tests using base oils that for those using greases [10]. The dent itself remains as prominent feature and creates relatively high local surfaces stress around the dent edges, where the new defect later initiate [27]. The high surface stresses around the dent shoulders create correspondingly high shear stresses in the subsurface. When the lubrication is poor, surface distress (micro-spalls) can be observed in the trailing edge of the dent, while the substantial wear occurs at the leading edge, possibly because of the local film collapse [92]. In summary, the high surface contact pressure and sub-surface stresses from the dentation shoulders are the main cause of defect [9,10]. The residual compressive stress around the dents can prolong the contact fatigue life [9]. The same effect might exist due to the over-rolling and abrasive/adhesive wear, which is well-known as smoothening phenomenon. It explains how the asperities that generate stresses might be smoothened. It was observed in many studies that there are shape changes of the dents' asperities during several contact cycles [9]. Mota [10] refers the surface smoothening of damaged area to the abrasive and adhesive wear actions. Ueda and Mitamura [20] refer the shape changes of the dent to the over-rolling phenomenon, where asperities become smaller during the first few contact cycles. It can be concluded that these asperities and their topological changes clarify the physical mechanisms of crack opening. Moreover, it is concluded that the descriptive model should consider three key mechanisms to illustrate the crack initiation in REB. These three mechanisms are: tangential forces acting on the dented surface, friction forces, and the impact forces due to associated vibrations. Therefore, the descriptive model illustrates and considers the crack initiation phenomenon as the following:

First, for healthy surface, tensile stress is generated at the entrance of contact point, while compressive stress is generated at the exit. Therefore, the magnitude of tensile and compressive stress generated by the tangential force is assumed to be the same at each contact point, due to a constant tangential force. In the case of dented surface, two issues are important to understand the Leading–Trailing Impulse Phenomenon:

- At the leading edge: the compressive stress which is generated at exit of the leading edge is larger than the tensile stress which is generated at the entrance of the leading edge.
- At the trailing edge: the tensile stress which is generated at entrance of the trailing edge is larger than the compressive stress which is generated at the exit of the trailing edge.

The direction of tangential forces depends on whether the surface is driven or driving, since the surface friction direction is different [10]. The direction of tangential forces is the same as the rolling direction for driven surface, while it is in opposite

direction for driving surface. Micro-cracks tend to occur at the one edge of dents with respect to rolling direction. It occurs mainly at the trailing edge of the dent of driven surfaces, due to the influence of large tensile stress.

Second, the friction force is one of the most important issues that is pulling the lubricant film and causing tensile stresses in the sliding direction [235–237]. The wear and possibly surface distress can be expected at the leading edge of the dent, due to thinner film thicknesses. The sliding might result due to larger traction forces when the roller rotational speed is reduced and the roller become the slower component at the roller–raceway contacts [221]. The zone of surface distress was observed in the trailing edge of the dent, while substantial wear occurs at the leading edge. When the asperity friction increases, the cracks go deeper [150]. Therefore, the frictional forces are quite significant issue to design durable REB. The coating material and its effect on wear process have been further illustrated in [238,25,239,240,208,211]. Moreover, several stress estimation models based on multi-axial fatigue criteria have been proposed in [93–97].

Third, the impact forces that are generated when the rolling element hits the trailing edge of the dent, explain the stress concentration and the initiation of several micro-cracks near to that edge. In fact, Kotzalas and Doll [221] observed the number of micro-cracks after the trailing edges that have relative spaces between them. It indicates the dynamic effect of rolling element after hitting the trailing edge, which might result as hammering loads with specific spacing once enough gapping is available. Kocich et al. [241] investigated the white-noise with low energy character of AE signals, in order to detect the plastic deformation event. The results show that a sudden rising of AE counts rate occurs near the yield point. It is induced from the movement of several dislocation bands. However, the dislocation density rises in the surface layer, but not in a continuous manner. Zhi-qiang et al. [242] highlighted the great influence of strain rate on the level of emitted AE signal. It is worth to highlight that the experiments which introduced artificial defects somehow ignored the earlier stages of wear progress. In such studies and tests, the further wear progress is totally dependent on the artificial defect size (length, width and depth), shape (dent, line defect) and its radial location. Therefore, it might not represent the real wear evolution progress. In order to represent a real wear evolution progress, the wear interactions should be considered. The descriptive model adopts the asperity theory proposed by Alfredsson et al. [148] to illustrate the wear interactions. Alfredsson et al. [148] introduce the asperity as a stress raiser mechanism that initiate the surface cracks. However, the new descriptive model considers also the degradation effect of the asperity due to over-rolling, adhesive and abrasive wear. Moreover, the descriptive model considers the effect of asperity-lubrication film on the subsurface crack. These novel considerations present how the descriptive model is able to address the wear interaction. These novel considerations are based on several studies that are discussed below.

2.3.3.1. Asperity as stress raiser. Kaneta et al. [24], Polonsky et al. [25], Dahlberg and Alfredsson [26] and Jouini et al. [151] have studied the asperity-induced stresses. Alfredsson et al. [148] observed that the single asperity served as stress raisers, when it enters and leaves the dry rolling contact. The critical asperity height was found to be small approximately 2 μm. The largest principle stresses in the un-cracked materials determine the cracking trajectories (in depth, width and length) [149], as illustrated in Fig. 2. In fact, the crack trajectories can explain how the wear defect might propagate in the width direction. The asperity size, friction and residual surface stress are all have influences on the crack initiation. The asperity height and local asperity friction have the largest effect on the crack initiation risk. Therefore, any reduction in the asperity friction will lengthen the surface lives and decrease the progression of the fatigue wear process. When the asperity friction increases, the cracks go deeper [150]. However, the crack depth is mainly influenced by the residual surface stress. Therefore, the defect size is best reduced by increasing the compressive residual stress, due to the fact that compressive residual stress tends to give more shallow fatigue wear. Mota et al. [227] highlighted that the shoulder sharpness is one of the most important parameters of dent characteristics. Shoulder sharpness is defined as ratio of shoulder height to the difference between the dent diameter and the diameter of shoulders peaks. From dynamic point of view, the dent edges are the potential impact
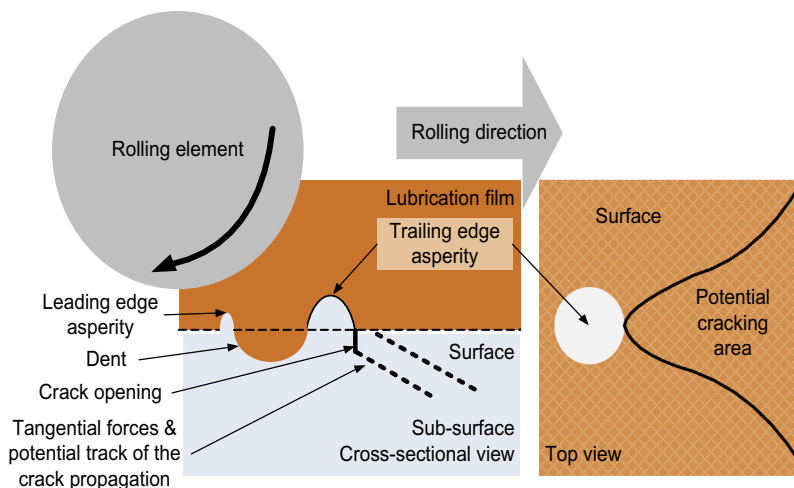

Fig. 2. Schematics of the dent, asperity and crack trajectories.

area when the rolling elements pass over the dent. These impact events produce sudden high stresses which are function of rolling element speed and impact area. In fact, there are impulsive effect of shoulders on the dynamic response due to the impact effects of entry edge and exit edge of the defect zone. This phenomenon named as which called the double-impulse phenomenon [23]. The entry edge is called leading edge and the exit edge is called as trailing edge in contact mechanics field. It is commonly observed that the impact of leading edge is quite smaller than trailing edge. The impulse interval in time domain could indicate the dent size (i.e. diameter, dent length). Moreover, the impact amplitude at leading edge corresponds to the dent's shoulder height. The impact amplitude at trailing edge corresponds relatively also to the dent depth and size. In this sense, the impulsive responses at leading and trailing edges are not symmetrical.

2.3.3.2. The effects of over-rolling, adhesive and abrasive wear. The shape of the dent changes to become smaller, during the first few contact cycles [20]. The material at leading edge of the dent deforms to broaden the dent due to high tensile stress, while the material at the trailing edge of the dent deforms and fills in the dent due to high compressive stress [20]. Thus, the measured dynamic impact of the dent is much smaller when its shoulders are smoothened. The impact at trailing edge might be also higher in case the roller element hits the modified bulge material at that edge [20]. Therefore, a plastic flow of the material is generated around the dent. Kim et al. [243] highlighted that the effective softening of the contact is due to asperity plasticity. Therefore, the over-rolling process modifies the surface topography i.e. asperity and plastic deformation. The over-rolling on the surface topography usually flattens the top of the asperity. However, although the asperities are plastically deformed during the over-rolling cycles, they remain sufficiently high to produce tensile surface stresses during a second steady-state roll cycle [148].

The effects of adhesive and abrasive wear on the asperity degradation are quite significant in rolling contact. Kato and Adachi [76] explained that the probability of wear particle generation depends on several issues: (1) microscopic shape of the contact, (2) microstructure of the material in the contact region, (3) microscopic surface contamination, and (4) disturbances in the surroundings. The scanning electron microscope (SEM) examination of rubbing surfaces showed that the gradual transfer of material from the rough rotating cylinder to the smoother stationary ball. The adhesive wear might act in the contact interface which has enough adhesive bonding strength and introduce a plastic deformation. This plastic deformation is due to dislocations which occur in the contact interface region under compression and shearing. Adhesive wear coefficient varies between 10^{-7} and 10^{-2} [76].

The abrasion actions are peeling-like mechanism the produce abrasion particles with sizes up to 200 μm. However, it is strongly affected by the material properties and geometry of contacting surfaces. Therefore, it is the most complicate parameter to estimate as well. The abrasive wear occurs when a single contact point is hard and sharp enough to remove a surface material of other contact point [54,55]. Abrasive cutting results in ploughing edges, groove and removed material (i.e. debris). Therefore, the hard abrasive asperity is assumed to act as indenter with three different modes: micro-cutting, wedge forming and ploughing. Hamblin and Stachowiak [66] observed a good correlation between quadratic spike parameter and abrasive wear rates. Abrasive wear resistance is linearly proportional to hardness of wearing materials [76]. Abrasive wear coefficient varies between 10^{-4} and 10^{-1} [76]. The wear particles are approximately spheres with diameter of 1 μm and the number of wear particles is about 18×10^{-9} for 0.6 g weight loss due to wear [65]. The correlation between bearing wear and the debris particle size is observed [68]. The accumulated RMS signal is seen to be directly proportional to the wear scar volume [215]. A wear test with a polished cylinder (i.e. that established a full elasto-hydrodynamic film) shows no RMS signal above the noise level. This observation indicates that the asperity contact is the prime source of the AE signals. Wear tests with particle inclusions (1 μm) show both RMS signal and wear volume were reduced by the inclusion of the particles [215]. Such observation illustrates the potential competition between adhesive and abrasive wear. The increase in RMS signal in the beginning of testing time was associated with material transfer from the cylinder to the ball by adhesion, later, the abrasive wear was dominant.

The volume of the abrasive wear can be predicted based on several modes: micro-cutting, wedge forming and ploughing. Several enhancements have been conducted to enhance the abrasive wear predictions such as the shear and attack angle based model [56], two-body abrasive wear model [57,58,66], slip-line field model [59,60], subsequent abrasion approach [62], artificial abrading surface model [63], coating-abrasion model [64]. However, the analytical estimation of wear based on metal cutting theory overestimates the wear [1]. This is due to the assumption that all the material from the groove is lost from the surface. The observations show that in general only some of the displaced material is actually detached while the remainder is piled up at the edges of resultant groove.

These asperity degradation mechanisms i.e. over-rolling, adhesive and abrasive wear are depending on the amount of the asperity interactions. A common parameter used to estimate the degree of asperity interaction is the lambda ratio λ as the ratio of lubricant film thickness to composite surface roughness. It is given by the following expression [244]

$$\lambda = \frac{h}{\sqrt{R_{Surface\ 1}^2 + R_{Surface\ 2}^2}}$$

where λ is degree of asperity interaction, h is the lubricant film thickness, $R_{Surface\ 1}$ is the RMS roughness of the roller surface, and $R_{Surface\ 2}$ is the RMS roughness of the raceway. If λ less than unity it is unlikely that the bearing will attain its estimated design life because of surface distress, which can lead to a rapid fatigue failure of the rolling surfaces. In general λ ratios greater than three indicate complete surface separation. A transition from full elastohydrodynamic lubrication (EHL) to

mixed lubrication (partial EHL film with some asperity contact) occurs in the λ range between 1 and 3 [244]. These asperity degradation mechanisms might de-accelerate the crack opening phenomenon as the asperity will be partially removed. A correlation between vibration levels and presence of wear particles (i.e. concentration) is observed over the test lifetime [225]. The AE data may be related to the source mechanisms due to different phenomena such as asperity contact, micro-crack initiation and propagation, plastic deformation and flow [215]. However, since any combination of these mechanisms may be active at any time, the interpretation of AE signals can be complicated. Peng and Kessissoglou [218] detected wear debris by vibration spectrum analysis in the low-frequency range, up to 0.3 kHz. Moreover, the wear was represented in the frequency domain as narrowband of increasing energy content (offset) around 260–280 Hz.

2.3.3.3. Subsurface crack initiation. It was discussed previously that the sub-surface crack either initiation by inclusion or stress concentration due to frictional forces. Larger size roller bearings operating typically at low to moderate Hertzian pressure are most susceptible to frictional surface loading [245]. Tangential forces by sliding friction acting on a rolling contact increase the von Mises equivalent stress and shift its maximum, i.e. the position of incipient plastic deformation, toward the surface [246,247]. Moreover, the subsurface crack can propagate due to dentation load [128,129]. The correlation between the AE location results (based on cumulative RMS values) and the subsurface cracks locations is observed [248]. The cracks were propagated approximately parallel to the surface with the maximum length of 200 μm and were distributed between 50 and 200 μm below the surface. Price et al. [249] observed that AE is capable to detect subsurface cracks prior to pitting. Schwach and Guo [202] observed bursts in the signal that indicate near-surface cracking. Also, the noise in the AE signal may indicate that crack initiation and propagation is occurring. The noise feature depends on the crack propagation features (shape, direction, angle, etc.). The detection of spalls depends on the thickness of the white layer of the surface. The white layer is a brittle layer often associated with high tensile residual stresses which may allow sudden fracture at the crack tip. Therefore, the signals detect the aggressive and rapid crack growth.

2.3.3.4. Wear mechanisms interaction. Watson et al. [48] highlighted the difficulty to develop wear prediction models as the ability to consider several wear and stress concentration mechanisms. The prediction based on individual wear mechanism is usually not a good assumption for practical applications. The most complicated issue is to consider the interactions among these wear and stress concentration mechanisms. This descriptive model illustrates the effect of asperities-induced stresses on the crack initiation, in the same time, the effect of over-rolling, adhesive wear and abrasive wear which smoothen the asperities. Smoothening the asperities prolong the surface life. Such physical interactions among the wear and stress concentration mechanisms lead to non-linear wear evolution which is well observed in the experimental and field data [12,13]. Therefore, there is a need to determine the deformation rate i.e. degradation rate of the generated asperities over the time. Olver [39] stated that mild wear at the surface can act to shorten cracks as time progresses. The influence of wear on the crack length can be observed as mitigation effect of fatigue. Moreover, the small changes in the shape of the contacting surfaces have a very large effect on the stresses and fluid pressures generated in each subsequent passes of rollers. Therefore, the asperity topography should be modified continuously with respect to the interacted mechanisms.

2.4. Crack initiation to crack opening state

A defect is expected to take place after a critical number of plastic strain cycles in the wave [250]. Jiang and Sehitoglu [251] and Jiang and Sehitoglu [252] stated that at the point where the maximum contact pressure is above the critical value of the elastic shakedown limit, the repeated plastic deformation takes places. Therefore, the fatigue wear coefficient is a function of the attack angle and normalised shear strength. This plastic deformation phenomenon is known as cyclic plastic or ratchetting, which have been further studied [139,253,254]. The several models show that the stress fields result in elastic and plastic deformations introduce some changes to the shape of the surface [83–87,235]. Therefore, the new descriptive model utilises these observations to define key factors of the crack opening phenomenon. First, the loading pattern must contain minimum and maximum peak values with large enough variation or fluctuation. The peak values may be in tension or compression and may change over time but the reverse loading cycle must be sufficiently great for the fatigue crack initiation. Second, the peak stress levels must be of sufficiently high. If the peak stresses are too low, no crack initiation will occur. Third, the material must experience a sufficiently large number of cycles of applied stress. The number of cycles required to initiate and grow a crack is largely dependent on the first two factors. These three conditions help to estimate how much the surface asperity is applying additional stresses that will initiate and accelerate the crack opening. Moreover, it helps to estimate how the asperity-induced stresses are accumulated either by the time for crack opening process or the accumulated loading cycles. Different subscripts are used to designate the stress intensity factor for different modes. The stress intensity factor for mode I is designated $\mathbf{K_I}$ and applied to the crack opening mode. The mode II stress intensity factor $\mathbf{K_{II}}$ applies to the crack sliding mode and the mode III stress intensity factor $\mathbf{K_{III}}$ applies to the tearing mode. Hannes and Alfredsson [149] defined a formula of asperity induced-stresses σ_0 assumes that the crack is straight.

2.5. Defect propagation

As the crack is opened, the focus will shift from plasticity at pre-existing stress–strain concentration points into the plasticity at the tip of the crack [39]. The crack is described by the strength of the stress intensity factor (SIF) at the tip. The most

straight-forward assumption of crack propagation is the linear elastic fracture mechanics [116–118]. The crack is described by the strength of the SIF at the tip. There are three main stages of propagation process: incubation, stable, and crack-to-surface [39]. It has been observed that the pit does not result from the return of the crack itself to the surface but rather the secondary cracks connecting the original crack to the surface [255]. Before that the crack continues to extend so that it often eventually becomes parallel to the surface. Wang and Hadfield [119] and Mota [10] state that the cracks depend on the direction of tractive or friction forces. Therefore, cracks extend usually in depth at about 20° to the surface, beside the role of crystallographic texture [34,7]. The crack continues to extend so that it often eventually becomes parallel to the surface [256,124]. Later, the pits results once the secondary crack(s) connecting the original one to the surface. The direction of cracking in stable fatigue stage (second stage) controlled by the direction of the motion of the contact zone cross the affected surface [39]. When cracks grow due to the increased stresses and unite with other cracks, debris is formed when it detach from the surface. This is the origin of the fatigue wear process [88–90,22,91].

There are three mechanisms that are driving the cracking process: shear stresses (mode II and III), fluid pressurization due to lubrication [162,163,170] and fluid entrapment [156,157], as illustrated in Fig. 3. Moreover, enough adhesive bonding influence and propagates the crack in a combined damage mode of tensile and shearing. The crack propagates downward to the depth of the maximum stress in a smooth Hertzian contact [9]. Olver [39] stated that mild wear at the surface can act to shorten cracks as time progresses. The influence of wear on the crack length can be observed as mitigation effect of fatigue [41]. Moreover, the small changes in the shape of the contacting surfaces have a very large effect on the stresses and fluid pressures generated in each subsequent passes of rollers. Therefore, we can conclude that the shape and size of pits and spalls are significant aspects to determine the new defected surface topology. Fluid entrapment refers to the closure of the mouth sealed fluid to the crack. It generates fluid pressure result in a high SIF on the crack tip. Balcombe et al. [156,157] observed that the pressurized fluid is the factor driving the crack opening. It is directly linked to the severity of crack propagation in the normal mode. In fact, that is due to an increase in SIF, which is directly related to the severity of crack opening displacement. The SIF is the govern crack propagation and critical to determine the crack propensity to grow. Furthermore, the combined effect of the external loading and crack opening generates shearing at the crack tip, which is strongly affected by the angle of inclination. It was also observed that the magnitude of SIF is directly related to the length of the crack. Therefore, longer cracks yield higher SIF, in which the acceleration of crack growth rate might occur. Balcombe et al. [156] summarized the second group i.e. crack lubrication film and listed four main lubrication influences on crack growth: (1) it reduces the friction between crack faces. (2) It applies direct pressure on the crack faces. (3) It causes hydrostatic pressure at crack tip. (4) It causes transient effecting due to the squeeze fluid layer mechanism.

The dent-initiated micro-crack(s) propagates down to the subsurface until either it meets another surface crack or an adhesive action removes the material. The first option depends on the high surface contact pressure and sub-surface stresses which are generated at the indentation edges. The tangential forces might be enough to initiate secondary cracks. The secondary cracks are usually responsible to connect the subsurface crack with the surface. The effects of material flow and grain size are observed in the defect clustering process [257,258]. The white etching layer phenomenon has significant role in the secondary crack initiation [259,255]. The development of heavily branching and widely spreading trans-crystalline crack systems is also depends on the chemically assisted crack growth by corrosion fatigue [245]. Chang and Jackson [152], Heyes [153], and Li et al. [154] provide detailed explanations of debris formation and asperity interactions. The formation of debris particles and natural dents are observed [10]. However, the wear volume was always higher for the tests using base oils, for each pair of grease/base oil (i.e. have almost same λ) [10]. In general, the defected surface results in more aggressive contact which increase the applied loading [12]. Yoshioka and Fujiwara [260] measured the defect propagation time as the time difference between the times at which AE signals increase at the position of failure on the raceway to the time when vibration acceleration increases at the same position of failure.
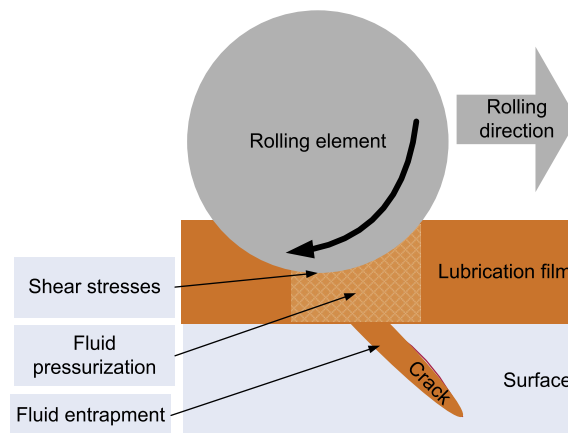

Fig. 3. Schematics of three mechanisms that are driving the cracking process.

2.6. Defect completion to damage growth

Wang and Hadfield [119] studied surface crack defect characteristics. Ringsberg and Bergkvist [120] studied crack length, crack angle, crack face friction and coefficient of surface friction near the contact load. Zhao et al. [121] and Zhao et al. [122] studied the ring crack propagation. Further, Tsushima [123], Liu et al. [124], Richard Liu and Choi [125], Donzella & Petrogalli [126] and Leonel and Venturini [127] have defined a number of issues that are required while modelling crack propagation: high stress location, depth below surface and the direction and angle of crack inclination. Therefore, based on these preliminary studies, quite large number of models related to crack propagation have been developed [130–136,79,137–141,43].

The crack propagation process will continue and the damage factor will be accumulated until the defect is completed and material is removed from the surface. The defect completion means the damage growth stage will start and continue iteratively in rapid growing manner. The crack propagation speed depends on the applied SIF. The crack propagation distance will be accumulated until it reaches a pre-specified crack length. Therefore, it is important to estimate how long time the crack propagation process will take. Also, it is important to know how the crack propagation phenomenon accumulates the asperity-induced stresses. Navarro and Rios [261] and Sun et al. [262] proposed the model where the crack growth rate da/dN is assumed to be proportional to the crack tip plastic displacement δ_{pl}.

$$\frac{da}{dN} = C_0(\delta_{pl})^{m_0}$$

where C_0, and m_0 are material constants that are determined experimentally. It is known that short cracks do not behave in accordance with linear elastic fracture model (LEFM). However, in view of the numerical simulation, it is beneficial to express the plastic displacement δ_{pl} in terms of the stress intensity factor K [106]. The total number of stress cycles N required for a short crack to propagate from the initial crack length a_0 to any crack length 'a' can then be determined as [262]:

$$N = \sum_{j=1}^{z} N_j$$

As the crack extends through ten or more grains, the influence of the material structure on the crack growth becomes negligible and the linear elastic fracture mechanic theory can be applied thereafter. In the framework of the LEFM the propagation rate of long cracks can be described by the Paris' equation [262].

$$\frac{da}{dN} = C_0(\Delta K)^{m_0}$$

where C_0, and m_0 are experimentally determined material parameters. However, this equation holds only before the crack reaches some critical crack length, when the crack growth rapidly increases until it becomes uncontrollable. Expressing the plastic displacement δ_{pl} in terms of the stress intensity range K enables treatment of short and long cracks in a similar fashion. If the relationship between the stress intensity range and the crack length $K = f(a)$ can be derived in some way, the remaining service life of a mechanical element with the crack can be estimated with appropriate integration of rate da/dN. Considering small crack lengths observed during pitting in the contact area of mechanical elements, only the theory of short cracks is usually needed for describing crack propagation from the initial to the critical crack length. The effect of over-rolling and abrasion on the SIF has essential role in crack prorogation process.

2.7. Damage growth

After the propagation stage, a complete defect is existing. The new descriptive model explains damage growth stage as a combination of defect growth in three directions (length, width and depth). In the depth direction, the spall phenomena can occur and deeper defect might be generated. Spalling is defined by ISO 15234 standards as advanced stage of flaking or rolling/sliding contacting surfaces during service. Therefore, the defect might propagate as extended area or propagates downward of the depth of the maximum stress in a smooth Hertzian contact. One important factor to be considered in the wear progress is the material response and hardness. In fact, it is well known that the instability stage might exposed to: decrease in yield stresses due to material softening, increase in the volume of deformed subsurface, microstructure changes and increase the radial tensile stress. The hardness profile of the virgin material is determined as a function of depth [238,25,239,240,208,211]. Therefore, it is quite important to consider the material softening, since some region of the first surface layers (i.e. the most hardened layer), are already partially removed. It means that in this stage the rolling element will contact softer and rougher surfaces. Therefore, it is expected that the defect will generates more stresses and propagates faster to become wider and deeper. In the length direction, the first defect has also leading and trailing edges which can initiate a new micro-crack and once again new defect is occurred. Therefore, the new trailing edge of pits will function as the stress raiser and initiate further spall [15]. However, at this stage, several wear mechanisms are involved and therefore the wear progress is accelerated. The wear curve tends to have aggressive and rapid increase in this stage. Al-Ghamd and Mba [263] observed that increasing the defect length increased the burst duration. In the width direction, the defect area is defined by the width and depth of the defect. Therefore, as the defect area become larger the rolling element hits stronger.

Al-Ghamd and Mba [263] observed that increasing the defect width increased the ratio of burst amplitude-to-operational noise (i.e. the burst signal was increasingly more evident above the operational noise levels).

The surface become rougher after some material related to first defect is removed. Therefore, adhesive and abrasive wear have higher possibility to occur. Moreover, the lubrication film is not uniform anymore and a number of surface disturbances are easily expected to happen. The defect also generates a number of larger sizes and sharper of surface asperities. Altogether produce clearance between the contacting surface which become gradually larger and therefore the contact gapping is expected. Moreover, the impact, which is generated while rolling element hit the defect area (the area of defect trailing edge), become stronger. The wear debris particles of first defect(s), which have large size and sharper shape, act a stress raiser and abrasive particles. Once the defect is completed and material is removed from surface as debris, new asperities are generated and the damage growth will continue and repeat slightly in the same damage scenario. Thus, the crack opening, crack propagation, over-rolling and abrasion will repeat again and again until a new defect is completely removed from surface, and so on. Moreover, the effect of multiple asperities is significantly important in damage growth stage.

Most of the experiments that have studied this stage have introduced a quite large size artificial defect which is even larger than the real size of spall defects. Dempsey et al. [14] observed that the spall length increases sharply at the beginning and then settles to a stable increment state before increasing sharply again. Sawalhi and Randall [23] investigated the trend of kurtosis values of faulty signals, with relation to the development of the fault size. The kurtosis increases almost linearly in the early stage of testing time interval as the defect size increases. However, it stabilizes later as the defect size is slowly extended. It could be either due to the existence of smoothing process or the surface becoming highly rough which the kurtosis might not be any more effective to detect. Kakishima et al. [264] utilised the vibration and AE measurements to detect artificial defects for both ball and rolling bearings. The AE sidebands were smaller in the case of wider artificial defect. In general, the AE and vibration magnitudes of the defect frequency component increase with increasing defect size. The vibration magnitude of the defect frequency component is more sensitive to the width of defect. Therefore, the vibration magnitude is related to the impact area (the contact area where the rolling element hits the raceway). This area can be calculated by the defect width and the inclined depth of defect. The length of the defect has indirect impact on the impact area, due to the fact that a longer defect space lets the rolling element fall deeper and hit a larger area of the trailing edge of the defect. Moreover, the roller element contacts/hits wider area of the defect than the ball element. Kim et al. [265] have observed that the ultrasound signals are clear at low speeds and display a number of impulses which are generated by localized defects.

3. The evolution of rolling contact wear and surface topology features

The detailed illustration of the new descriptive model with the supportive experimental findings in the literature is presented in previous section. In this section, a summary of the descriptive model and its characteristics are provided.

The bearing lifetime is described based on five stages. The wear evolution model assumes that at certain time interval of the steady-state stage, a transition into the defect initiation stage will take a place. Later, the evolution model describes the wear progress with help of two assumptions: the existence of multiple stages that have specific transition events and existence of multiple wear and stress concentration mechanisms that are acting in each progress stage. Therefore, the wear evolution model that has been described in details with help of literature (i.e. experimental findings) can be summarized as follows. When the stresses in rolling contact increase due to the increase operating loads, additional loads due to faults i.e. imbalance, misalignment, bent shaft, looseness, and/or distributed defects i.e. high degrees surface roughness and waviness, contaminations, inclusions, some topological changes might occur. These topological changes in the contact area generate stress concentration points and lubrication film disturbances. At early stage, the concentrated stresses are not strong enough to produce a defect. It is mainly located in the loading zone and in the normal running track i.e. pure rolling points. Later, some sliding events, lubrication film transfers, false brinelling events due to stand-still events might occur and introduce some degree of surface interaction. These surface interactions might appear as reduction in the lubrication film, gapping, miss-matching between the rolling element profile and race profile, etc. Therefore such surface conditions allow some abrasive wear event, some contaminations to enter the contact zone and minor vibration impacts. As a result of these actions, some surface dentations might be generated. Therefore, the model describes how dentations and in particular their asperities has the main role in the defect initiation and propagation process. The main assumption is that as long as the asperity is larger and sharper, the contact force is larger when the rolling element passes over the asperity. The contact force and other loading forces increase and concentrate the applied tangential force. The tangential force and the friction force are the main forces that generate sufficient SIF for crack opening and later for the crack propagation. However, since the rolling element is rolling over the asperity over the time and the asperity can be abraded, the original shape of dent is changing over the time. Therefore, the impact force can be degraded and the crack opening and propagation progress as well. However, although the asperities are plastically deformed during the over-rolling cycles and degraded by abrasive and adhesive wear actions, they remain sufficiently high to produce tensile surface stresses. In the end, the propagated crack need a secondary crack to reach the surface or it can attach to the rolling element once sufficient adhesive bonding is exist. During this process, the lubrication film is distorted and transfer into surface crack where another mechanical and chemical actions might accelerate the defect completion. When the defect is completed and the material is detached from the surface: new asperity is generated, new debris is generated, sever disturbances of lubrication film is generated, and the less hardening material (i.e. the material that was below the removed defect material) became the new surface.

The new descriptive model highlights the features of the generated defect and its asperity. The length of the generated defect depends on how long the crack could propagate in parallel to the surface before the detachment process occurred. The depth of the generated defect depends on how deep the crack was opened before it matches with a secondary or sub-surface inclusion. It also depends on the friction force and its depth of stress concentration. The width of the generated defect depends how far the force trajectories were propagating. The defect's length, depth and width are the basic elements of the new impact area. Therefore, a new impact force will be generated when the rolling element passes over the new defect i.e. specially at the trailing edge of new defect. The new defect will generate a number of defect serials and the generated debris will generate several dentations in different locations. First, the tangential force will be larger since the asperity is larger and rougher than initial dent asperities. Second, the new debris will generate more sever dentation since it is larger and sharper than contamination particles. Debris might act as moving and distributed asperity. Moving asperities have more random and non-linear way of action compare to the fixed ones i.e. dent asperities. Large portion of debris will be pressed into the surface and generate more dents. Moreover, debris can act as asperity and minor indenter and generate abrasive wear. Third, the surface hardness of the new appeared surface is less than the initial surface hardness.

The five stages are schematically illustrated in Fig. 4. The wear evolution progress produces several surface topology changes. These topology changes have significant effect on the physical measurements, in particular, for condition monitoring purposes. At the end of running-in stage, the roughness of the surface becomes smoother. Therefore, the steady-state stage is characterized by uniform lubricant film and contact mechanics, under normal operating conditions. When the surface is dented, the surface look likes peelings. Later, the shapes of dents change due to the over-rolling and abrasive wear actions. However, the trailing edge acts as stress raiser and the micro-cracking is initiated and opened on and under the surface. The micro-cracks propagate from the surface downward with inclination, which depends on rolling direction. The crack propagate later in parallel with rolling direction until it meet a secondary crack and connect to the surface or detaching process occur. Therefore, relatively large material will be detached from the surface as debris particles. After first pit a number of pits and spalls are expected to occur in a serial pattern and extend in wider and deeper manner, where the defect area becomes larger and rougher.

Rolling wear interactions: the new descriptive model of wear evolution i.e. five-stages emphasise the dynamic nature of wear. This dynamic nature is produced as a combination of wear mechanisms, stress concentration mechanisms, and operating conditions. Thus, rolling contact wear is not generated by single and individual wear mechanism. This dynamic nature changes significantly the wear evolution scenario over the lifetime. Therefore, we need to understand the interactions and competitions among the involved wear mechanisms. In Fig. 5, the schematic description of surface mechanism, wear interactions and wear evolution are simply illustrated within three frameworks respectively. The wear interaction framework presents the involved wear mechanisms (fatigue, abrasive, and adhesive) together with the dent, debris and asperity mechanisms. First, fatigue wear mechanism depends on the loading patterns and lubrication regimes. Therefore, the cyclic load and friction coefficient are main attributes for contact damage criteria. The stress concentration might vary based on friction coefficient. Low friction means that stress will be most probably concentrated in subsurface. High friction means that stress will be concentrated on surface. The low-friction produces subsurface stresses which might accumulate and initiate a
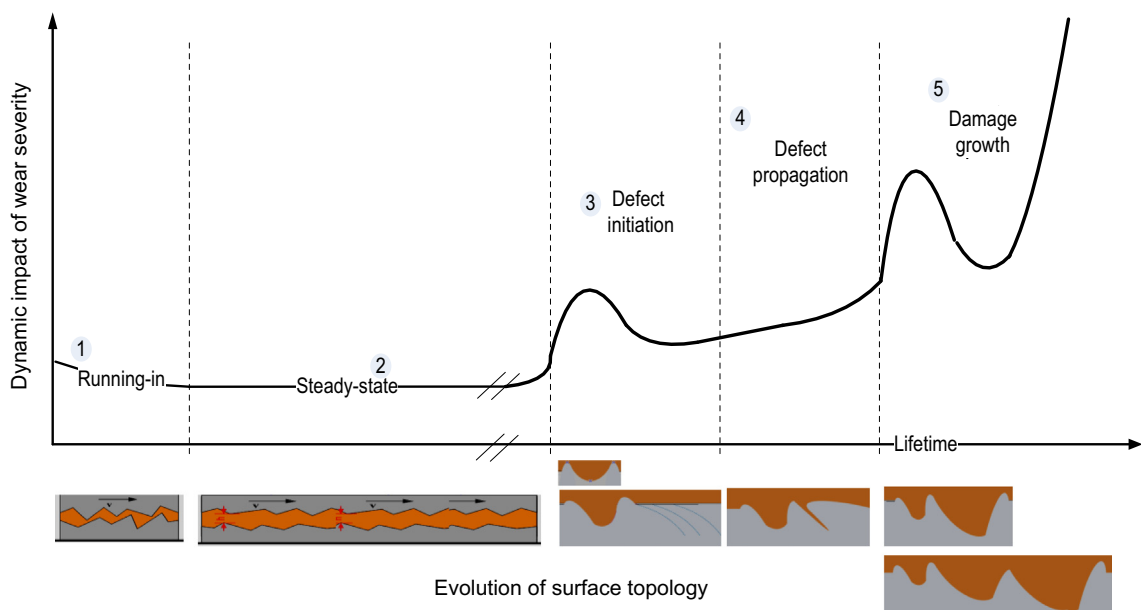

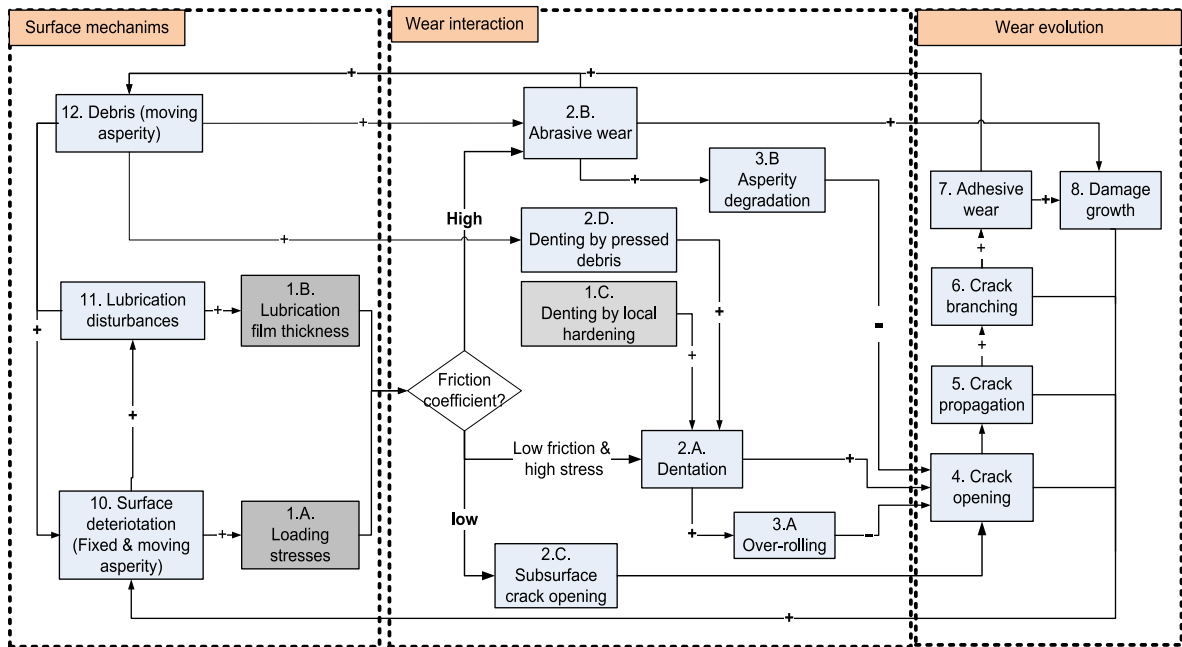**Fig. 4.** Evolution of dynamic behaviour and surface topology due to wear evolution.

Fig. 5. Interaction and competition of wear mechanisms.

crack, as shown in 2C-4 track in Fig. 5. However, the high-friction produces micro-cracking. Later this crack propagates until a secondary or branching crack occurs before it reaches the surface, as illustrated in the track 2.A-8 in Fig. 5.

The dent-related wear mechanism can be produced by three different mechanisms: (i) due to high stresses as represented by (2.A), (ii) due to local hardening as (1.C), where one surface press (i.e. as brinelling) the other surface in stand-still operations and (iii) due to debris particles (2.D) which is pressed and left the surface. The dent usually increases surface roughness and might lead to lubrication disconnect points during rolling, as the lubricated film move toward the bottom of the dent in (11). Third, debris might be produced by abrasive wear, (2.B). On other hand, debris produce a stress concentration points (as illustrated in track 12–10), also debris might act as minor indenter (cutting object) where more abrasive wear might occur (as illustrated in track 12-2B). Moreover, the debris produce moving particles (12), and fixed asperities (10) on the surface, in terms of ploughing edges on the sides of abrasion track. Fourth, asperities are represented once the dentation is occurring on the surface. They are also considered a stress concentration points and as potential objects for abrasive wear. The interactions are represented in Fig. 5 as (+) sign which shows how the different mechanisms interact and accelerate the performance and consequence of each other. However, mechanism competitions are represented by (−) sign. One of the clear examples of the mechanism competition is the crack initiation and abrasive wear. In this case, the over-rolling and abrasive wear mechanism smoothen the rough asperities which are generated by dentation.

Influencing factors: the new descriptive model of wear evolution i.e. five-stages emphasise the influencing factors that can highly vary within each stage. Therefore, it is important to consider influencing factors within each stage. Kappa [266] had illustrated the influencing factors on the lubricant film quality. Therefore, in the study the influencing factors are illustrated for the whole wear evolution in the same manner. In summary, the wear progress depends on the evolution of loading conditions, surface quality, lubrication film quality and subsurface quality as illustrated in Fig. 6. These four factors generate together specific patterns of stress concentration and the damage progression scenarios. In Fig. 6, these four factors which influence the wear evolution are illustrated.

In this sense, the wear modelling can utilise the descriptive model to gain the required fundamentals about the wear mechanisms, stages transitions. The wear testing can also utilise the descriptive model to design experiments that can delimit the mechanisms behind the natural wear evolution process. In fact, there are two approaches for wear testing either using natural accelerated damage by applying overloads, adjusting the lubricant film thickness, and adding contaminated oil, or Introducing artificial defect into the surface by false-brinelling, erosion charge and scratching. The later approach is widely used due to it is simplicity of modelling, where the researchers can virtually introduce a well-known shape and size of the defect. They can also artificially introduce the same defect in the validation experiment. Furthermore, this approach delimits the testing complexity, as it focuses on a single artificial defect, compared to the natural defect approach where a number of defects might be produced. The drawback of this approach is that the damage criterion is somehow artificially determined which might be totally different than the defect in the real operation. Therefore, the artificial defected bearing tests have not much to tell us about the evolution of real wear progress. Finally, the wear monitoring can utilise the
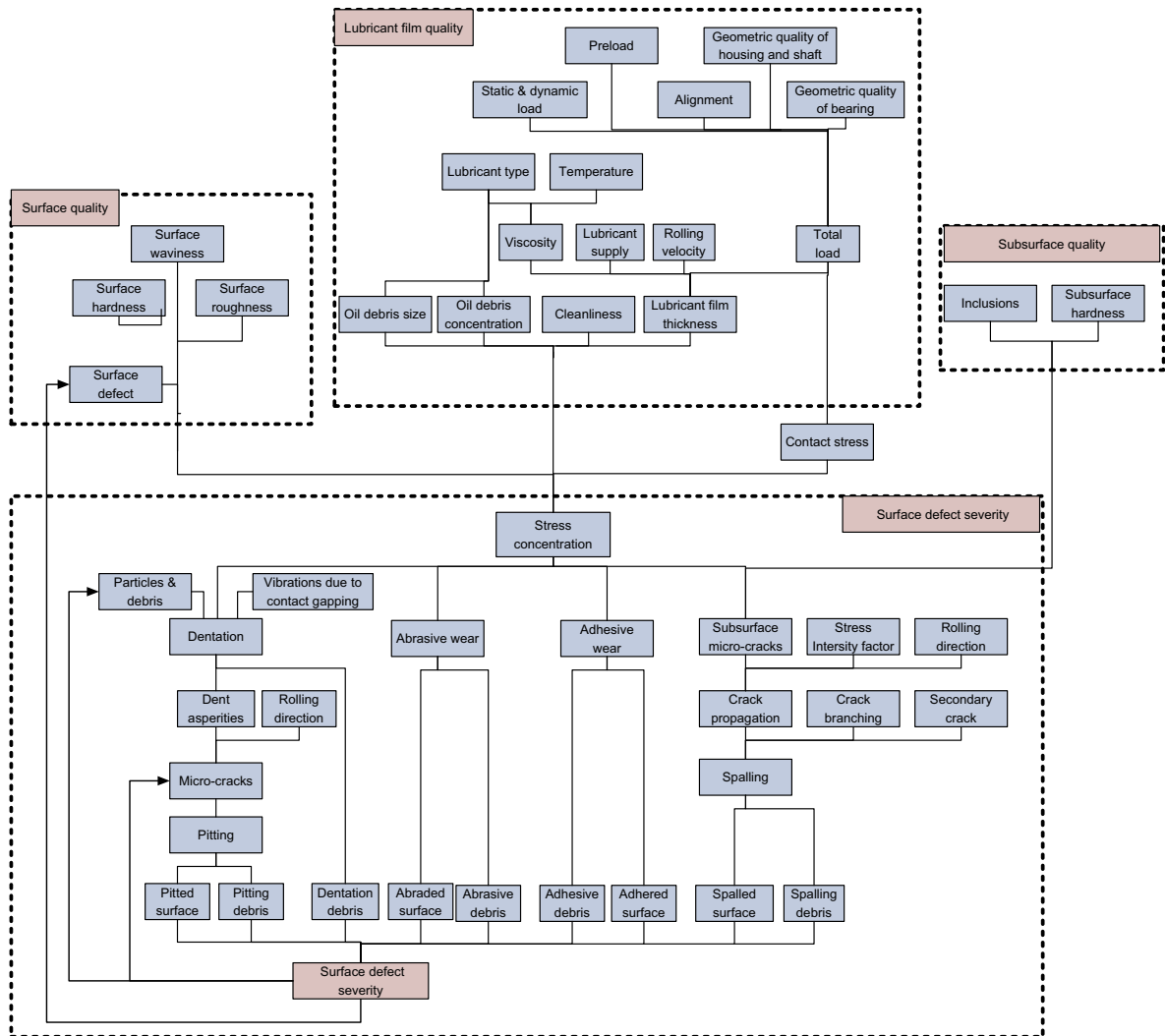
Fig. 6. Key operating factors of rolling contact wear.

descriptive model to understand the physical phenomenon and surface topology changes behind the measured signals and provide better ability to diagnosis the wear severity over the time.

4. Conclusions

The published literature contains a wide range of numerical and experimental findings related mainly to crack initiation and propagation. These two phenomena are important and central phenomena to explain the wear evolution. However they are not sufficient. Therefore, the new descriptive model explains and links the pre-stages with the crack initiation in the rolling contact i.e. steady state, defect localization and dentation. In particular, the explanation of defect localization provides five potential stress localization mechanisms beside the contamination effect, lubrication disturbances and impact events. Moreover, it explains and links the post-stage of crack propagation i.e. damage growth. These stages and their associated issues provide the whole wear evolution progress, supported by the scientific experimental findings within the literature. The new descriptive model enhances the understanding of the fatigue wear i.e. crack initiation and propagation by considering the wear interactions i.e. over-rolling, abrasive, and adhesive. The model shows that the interaction among the involved wear mechanisms might accelerate or decelerate the wear evolution. That provides better understanding of the measured data and better estimation for the remaining useful time. In summary, the new descriptive model highlights the evolution stages, the transition event between the stages, the involved mechanism in each stage, and the surface topology changes over the lifetime and the influencing factors. The implication of the new descriptive model is to provide the basis for more realistic modelling, testing, and monitoring of the wear evolution in REBs.

Acknowledgements

References

[1] Williams J. Wear modelling: analytical, computational and mapping: a continuum mechanics approach. Wear 1999;225–229:1–17.
[2] Lim S. Recent developments in wear-mechanism maps. Tribol Int 1998;31(1–3):87–97.
[3] Lim SC, Ashby MFA. Wear mechanism maps. Acta Metall 1987;35.
[4] Cantizano A, Carnicero A, Zavarise G. Numerical simulation of wear-mechanism maps. Comput Mater Sci 2002;25(1–20):54–60.
[5] Bosman R, Schipper DJ. Mild wear prediction of boundary-lubricated contacts. Tribol Lett 2011;42(2):169–78.
[6] Voskamp AP. Material response to rolling contact loading. J Tribol 1985;107:359–64.
[7] Longching C, Qing C, Eryu S. Study on initiation and propagation angles of subsurface cracks in GCr15 bearings steel under rolling contact. Wear 1989;133:205–18.
[8] Raje N, Slack T, Sadeghi F. A discrete damage mechanics model for high cycle fatigue in polycrystalline materials subject to rolling contact. Int J Fatigue 2009;31:346–60.
[9] Gao N, Dwyer-Joyce RS, Beynon JH. Effects of surface defects on rolling contact fatigue of 60/40 brass. Wear 1999;225–229:983–94.
[10] Mota V. Investigations on surface damage by rolling contact fatigue in elastohydrodynamic contacts using artificial dents. Faculdade de Engenharia da Universidade do Porto; 2005.
[11] El-Thalji I, Jantunen E. Wear of rolling element bearings. In: Proceedings conference of condition monitoring and diagnostic engineering management (COMADEM); 2013.
[12] Jantunen E. How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods. Int J COMADEM 2006;9(3):24–38.
[13] Yoshioka T, Shimizu S. Monitoring of ball bearing operation under grease lubrication using a new compound diagnostic system detecting vibration and acoustic emission. Tribol Trans 2009;52(6):725–30.
[14] Dempsey PJ, Bolander N, Haynes C, Toms AM. Investigation of bearing fatigue damage life prediction using oil debris monitoring; 2011.
[15] Xu G, Sadeghi F. Spall initiation and propagation due to debris denting. Wear 1996;201:106–16.
[16] Ding Y, Jones R, Kuhnell BT. Elastic-plastic finite element analysis of spall formation in gears. Wear 1996;197:197–205.
[17] Ahmed R. Contact fatigue failure modes of HVOF coatings. Wear 2002;253:473–87.
[18] Dwyer-Joyce RS. The life cycle of a debris particle. Tribol Interface Eng Ser 2005;48:681–90.
[19] Marble S, Morton BP. Predicting the remaining life of propulsion system bearings. In: Proceedings of the 2006 IEEE aerospace conference; 2006.
[20] Ueda T, Mitamura N. Mechanism of dent initiated flaking and bearing life enhancement technology under contaminated lubrication condition Part I: effect of tangential force on dent initiated flaking. Tribol Int 2008;41:965–74.
[21] Harvey TJ, Wood RJK, Powrie HEG. Electrostatic wear monitoring of rolling element bearings. Wear 2007;263:1492–501.
[22] Arakere NK, Branch N, Levesque G, Svendsen V, Forster NH. Rolling contact fatigue life and spall propagation of AISI M50, M50NiL, and AISI 52100, Part II: stress modeling. Tribol Trans 2009;53(1):42–51.
[23] Sawalhi N, Randall RB. Vibration response of spalled rolling element bearings: observations, simulations and signal processing techniques to track the spall size. Mech Syst Signal Process 2011;25(3):846–70.
[24] Kaneta M, Sakai T, Nishikawa H. Effects of surface roughness on point contact EHL. Tribol Trans 1993;36(4):605–12.
[25] Polonsky IA, Chang TP, Keer LM, Sproul WD. An analysis of the effect of hard coatings on near-surface rolling contact fatigue initiation induced by surface roughness. Wear 1997;208:204–19.
[26] Dahlberg J, Alfredsson B. Influence of a single axisymmetric asperity on surface stresses during dry rolling contact. Int J Fatigue 2007;29:909–21.
[27] Sayles RS, Ioannides E. Debris damage in rolling bearings and its effects on fatigue life. Trans ASME, J Tribol 1988;110:26–31.
[28] Nikas GK, Sayles RS, Ioannides E. Effects of debris particles in sliding/rolling elastohydrodynamic contacts. Proc Int Mech Eng 1998;212(Part J):333–43.
[29] Ai X. Effect of debris contamination on the fatigue life of roller bearings. Proc Inst Mech Eng, Part J: J Eng Tribol 2001;215(Part J):563–775.
[30] Hiratsuka K, Muramoto K. Role of wear particles in severe – mild wear transition. Wear 2005;259:467–76.
[31] Stachowiak GP, Stachowiak GW, Podsiadlo P. Automated classification of wear particles based on their surface texture and shape features. Tribol Int 2008;41(1):34–43.
[32] Yamashita N, Mura T. Contact fatigue crack initiation under repeated oblique force. Wear 1983;91:235–50.
[33] Cheng W, Cheng HS, Keer LM. Experimental investigation on rolling/sliding contact fatigue crack initiation with artificial defects. Tribol Trans 1994;37(1):1–12.
[34] Wong SL, Bold PE, Brown MW, Allen RJ. A branch criterion for shallow angled rolling contact fatigue cracks in rails. Wear 1996;191:45–53.
[35] Ringsberg JW, Loo-morrey M, Josefson BL. Prediction of fatigue crack initiation for rolling contact fatigue. Int J Fatigue 2000;22:205–15.
[36] Wang Y, Hadfield M. Rolling contact fatigue failure modes of lubricated silicon nitride in relation to ring crack defects. Wear 1999;225–229:1284–92.
[37] Chue C, Chung H. Pitting formation under rolling contact. Theor Appl Fract Mech 2000;34.
[38] Datsyshyn OP, Panasyuk VV. Pitting of the rolling bodies contact surface. Wear 2001;251:1347–55.
[39] Olver AV. The mechanism of rolling contact fatigue: an update. J Mech Eng, Eng Tribol 2005;219(Part J):313–29.
[40] Randall RB, Antoni J. Rolling element bearing diagnostics—a tutorial. Mech Syst Signal Process 2011;25(2):485–520.
[41] Fan H, Keer LM, Cheng W, Cheng HS. Competition between fatigue crack propagation and wear. J Tribol 1993;115:141–57.
[42] Oliveira S, Bower A. An analysis of fracture and delamination in thin coatings subjected to contact loading. Wear 1996;198:15–32.
[43] Tarantino MG, Beretta S, Foletti S, Papadopoulos I. Experiments under pure shear and rolling contact fatigue conditions: competition between tensile and shear mode crack growth. Int J Fatigue 2013;46:67–80.
[44] Harris FJ. Rolling bearing analysis I, First. New York: John Wiley & Sons Inc.; 1966.
[45] Harris FJ. Wear. In: Rolling bearing analysis. New York: John Wiley & Sons Inc.; 1991. p. 891–917.
[46] Hockenhull BS, Kopalinsky EM, Oxley PLB. Mechanical wear models for metallic surfaces in sliding contact. J Phys 1992;25:266–72.
[47] Engel PA. Failure models for mechanical wear modes and mechanisms. IEEE Trans Reliab 1993;42(2):262–7.
[48] Watson M, Byington C, Edwards D, Amin S. Dynamic modeling and wear-based remaining useful life prediction of high power clutch systems. STLE Tribol Trans 2005;48(2):208–17.
[49] Holmberg K, Ronkainen H, Laukkanen A, Wallin K. Friction and wear of coated surfaces—scales, modelling and simulation of tribomechanisms. Surf Coat Technol 2007;202:1034–49.
[50] Montgomery S, Kennedy D, O'Dowd N. Analysis of wear models for advanced coated materials. In: International conference on materials, tribology, recycling; 2009.
[51] Mukras S, Kim NH, Mauntler NA, Schmitz TL, Sawyer WG. Analysis of planar multibody systems with revolute joint wear. Wear 2010;268(5–6):643–52.
[52] Suh NP. An overview of the delamination theory of wear. Wear 1977;44:1–16.
[53] Fouvry S, Kapsa P. An energy description of hard coating wear mechanisms. Surf Coat Technol 2001;138(2–3):141–8.
[54] Kruschov MM. Resistance of metals to wear by abrasion, as related to hardness. In: Proceedings conference on lubrication and wear; 1957. p. 655–9.
[55] Richardson RCD. The wear of metals by hard abrasives. Wear 1967;291.

[56] Challen JM, Oxley PLB. An explanation of the different regimes of friction and wear using asperoty deformation models. Wear 1979;17:229–43.
[57] Zum-Gahr K-H. Micro-structure and wear of materials. Amsterdam; 1987.
[58] Zum-Gahr K-H. Modelling of two-body abrasive wear. Wear 1988;124:87–102.
[59] Petryk H. Slip-line field solutions for sliding contact. In: Proceedings conference on friction, lubrication and wear; 1987. p. 987–94.
[60] Torrance AA, Buckley TR. A slip-line field model of abrasive wear. Wear 1996;196(1–2):35–45.
[61] Jacobson S, Wallen P, Hogmark S. Correlation between groove size, wear rate and topography of abraded surfaces. Wear 1987;115:83–93.
[62] Garrison WM. Abrasive wear resistance: the effect of ploughing and the removal of plouged material. Wear 1986;114:239–47.
[63] Gåhlin R. Micro-scale studies of wear. University of Upsala; 1998.
[64] Kassman Å, Jacobson S, Erickson L, Hedenqvist P, Olsson M. A new test method for the intrinsic abrasion resistance of thin coatings. Surf Coat Technol 1991;50(1):75–84.
[65] Olofsson U. Characterisation of wear in boundary lubricated spherical roller thrust bearings. Wear 1997;208:194–203.
[66] Hamblin M, Stachowiak G. Characterisation of surface abrasivity and its relation to two-body abrasive wear. Wear 1997;206(1–2):69–75.
[67] Jiang J, Sheng F, Ren F. Modelling of two-body abrasive wear under multiple contact conditions. Wear 1998;217:35–45.
[68] Dwyer-Jones RS. Predicting the abrasive wear of ball bearings by lubricant debris. Wear 1999;233–235:692–701.
[69] Stachowiak G, Podsiadlo P. Characterization and classification of wear particles and surfaces. Wear 2001;249:194–200.
[70] Torrance AA. Modelling abrasive wear. Wear 2005;258:281–93.
[71] Nilsson R. On wear in rolling/sliding contacts. Stockholm, Sweden: Royal Institute of Technology (KTH); 2005.
[72] Nilsson R, Svahn F, Olofsson U. Relating contact conditions to abrasive wear. Wear 2006;261(1):74–8.
[73] Savio G, Meneghello R, Concheri G. A surface roughness predictive model in deterministic polishing of ground glass moulds. Int J Mach Tools Manuf 2009;52(2):611–33.
[74] Lundgren G, Palmgren A. Dynamic capacity of roller bearings. Ingenicorsrenten skapsakedemiens 1952;210.
[75] Giannakopoulos AE, Lindley TC, Suresh S. Aspects of equivalence between contact mechanics and fracture mechanics: theoretical connections and a life-prediction methodology for fretting-fatigue. Acta Mater 1998;46(9):2955–68.
[76] Kato K, Adachi K. Wear mechanisms. In: Bhushan B, editor. Modern tribology handbook. CRC Press; 2001.
[77] Bayer R. Mechanical wear fundamentals and testing. 2nd ed. Marcel Dekker, Inc.; 2004.
[78] Ekberg A, Kabo E. Fatigue of railway wheels and rails under rolling contact and thermal loading—an overview. Wear 2005;258:1288–300.
[79] Sadeghi F, Jalalahmadi B, Slack TS, Raje N, Arakere NK. A review of rolling contact fatigue. J Tribol 2009;131(4).
[80] Joseph Ebert F. Fundamentals of design and technology of rolling element bearings. Chin J Aeronaut 2010;23:123–36.
[81] Pattabhiraman S, Levesque G, Kim NH, Arakere NK. Uncertainty analysis for rolling contact fatigue failure probability of silicon nitride ball bearings. Int J Solids Struct 2010;47:2543–53.
[82] Lewis M, Tomkins B. A fracture mechanics interpretation of rolling bearing fatigue. Proc Inst Mech Eng, Part J: J Eng Tribol 2012;226(5):389–405.
[83] Kim TH, Olver AV. Stress history in rolling-sliding contact of rough surfaces. Tribol Int 1998;31(12):727–36.
[84] Karmakar S, Rao URK, Sethuramiah A. An approach towards fatigue wear modelling. Wear 1996;198:242–50.
[85] Jaing Y, Sehitoglu H. Rolling contact stress analysis with the application of a new plasticity model. Wear 1996;191:35–44.
[86] Franklin FJ, Widiyarta I, Kapoor A. Computer simulation of wear and rolling contact fatigue. Wear 2001;251:949–55.
[87] Masen MA, de Rooij MB, Schipper DJ. Micro-contact based modelling of abrasive wear. Wear 2005;258:339–48.
[88] Khan ZA, Hadfield M, Tobe S, Wang Y. Residual stress variations during rolling contact fatigue of refrigerant lubricated silicon nitride bearing elements. Ceram Int 2006;32:751–4.
[89] Raje N, Sadeghi F, Rateick RG. A discrete element approach to evaluate stresses due to line loading on an elastic half-space. Comput Mech 2007;40:513–29.
[90] Zaretsky EV, Poplawski JV, Root LE. Relation between hertz stress-life exponent, ball-race conformity, and ball bearing life; 2008.
[91] Branch N, Arakere N, Svendsen V, Forster N. Stress field evolution in a ball bearing raceway fatigue. J ASTM Int 2010;7(2):1–18.
[92] Morales-Espejel GE, Brizmer V. Micropitting modelling in rolling-sliding contacts: application to rolling bearings. Tribol Trans 2011;54(4):625–43.
[93] Karolczuk A, Macha E. A review of critical plane orientations in multiaxial fatigue failure criteria of metallic materials. Int J Fract 2005;134:267–304.
[94] Ciavarella M, Monno F, Demelio G. On the Dang Van fatigue limit in rolling contact fatigue. Int J Fatigue 2006;28:852–63.
[95] Desimone H, Bernasconi A, Beretta S. On the application of Dang Van criterion to rolling contact fatigue. Wear 2006;260:567–72.
[96] Ciavarella M, Monno F. A comparison of multiaxial fatigue criteria as applied to rolling contact fatigue. Tribol Int 2010;43:2139–44.
[97] Conrado E, Foletti S, Gorla C, Papadopoulos IV. Use of multiaxial fatigue criteria and shakedown theorems in thermo-elastic rolling–sliding contact problems. Wear 2011;270:344–54.
[98] Ringsberg JW. Life prediction of rolling contact fatigue crack initiation. Int J Fatigue 2001;23:575–86.
[99] Harris TA, Barnsby RM. Life ratings for ball and roller bearings. Proc Inst Mech Eng, Part J: J Eng Tribol 2001;215(6):577–95.
[100] Akama M, Mori T. Boundary element analysis of surface initiated rolling contact fatigue cracks in wheel/rail contact systems. Wear 2002;253:35–41.
[101] Bormetti E, Donzella G, Mazzù A. Surface and subsurface cracks in rolling contact fatigue of hardened components. Tribol Trans 2002;45(3):274–83.
[102] Sraml M, Flasker J, Potrc I. Numerical procedure for predicting the rolling contact fatigue crack initiation. Int J Fatigue 2003;25:585–95.
[103] Gessesse YB, Attia MH, Osman MOM. On the mechanics of crack initiation and propagation in elasto-plastic materials in impact fretting wear. J Tribol 2004;126:395–403.
[104] Bernasconi A, Davoli P, Filippini M, Foletti S. An integrated approach to rolling contact sub-surface fatigue assessment of railway wheels. Wear 2005;258:973–80.
[105] Li Y, Kang G, Wang C, Dou P, Wang J. Vertical short-crack behavior and its application in rolling contact fatigue. Int J Fatigue 2006;28:804–11.
[106] Fajdiga G, Sraml M. Fatigue crack initiation and propagation under cyclic contact loading. Eng Fract Mech 2009;76:1320–35.
[107] Nagatani H. Improved method of rolling bearing fatigue life prediction under edge loading conditions. NTN Tech Rev 2010;78:83–90.
[108] Grabulov A, Petrov R, Zandbergen HW. EBSD investigation of the crack initiation and TEM/FIB analyses of the microstructural changes around the cracks formed under rolling contact fatigue (RCF). Int J Fatigue 2010;32:576–83.
[109] Alley ES, Sawamiphakdi K, Anderson PI, Neu RW, Beswick J, Dean SW. Modeling the influence of microstructure in rolling contact fatigue. J ASTM Int 2010;7(2).
[110] Alley ES, Neu RW. Microstructure-sensitive modeling of rolling contact fatigue. Int J Fatigue 2010;32:841–50.
[111] Beheshti A, Khonsari MM. On the prediction of fatigue crack initiation in rolling/sliding contacts with provision for loading sequence effect. Tribol Int 2011;44:1620–8.
[112] Sandström J, de Maré J. Probability of subsurface fatigue initiation in rolling contact. Wear 2011;271:143–7.
[113] Lu L, Wang X, Gao Z, Jiang Y. Influence of the contact pressure on rolling contact fatigue initiation of 1070 steel. Proc Eng 2011;10:3000–5.
[114] Datsyshyn OP, Panasyuk VV, Glazov AY. Modeling of fatigue contact damages formation in rolling bodies and assessment of their lifetime. Wear 2011;271:186–94.
[115] Kadin Y, Rychahivskyy AV. Modeling of surface cracks in rolling contact. Mater Sci Eng 2012;541:143–51.
[116] O'Regan SD, Hahn GT, Rubin CA. The driving force for mode II crack growth under rolling contact. Wear 1985;101:333–46.
[117] Salehizadeh H, Saka N. Crack propagation in rolling line contacts. Trans ASME, J Tribol 1992;114:690–6.
[118] Panasyuk VV, Datsyshyn OP, Marchenko HP. The crack propagation theory under rolling contact. Eng Fract Mech 1995;52(1):179–91.
[119] Wang Y, Hadfield M. Ring crack propagation in silicon nitride under rolling contact. Wear 2001;250:282–92.
[120] Ringsberg JW, Bergkvist A. On propagation of short rolling contact fatigue cracks. Fatigue Fract Eng Mater Struct 2003;26:969–83.

[121] Zhao P, Hadfield M, Wang Y, Vieillard C. Subsurface propagation of partial ring cracks under rolling contact: Part I. Experimental studies. Wear 2006;261:382–9.
[122] Zhao P, Hadfield M, Wang Y, Vieillard C. Subsurface propagation of partial ring cracks under rolling contact: Part II. Fracture mechanics analysis. Wear 2006;261:390–7.
[123] Tsushima N. Crack propagation of rolling contact fatigue in ball bearing steel due to tensile strain. NTN Tech Rev 2007;75:128–38.
[124] Liu Y, Liu L, Mahadevan S. Analysis of subsurface crack propagation under rolling contact loading in railroad wheels using FEM. Eng Fract Mech 2007;74:2659–74.
[125] Richard Liu C, Choi Y. A new methodology for predicting crack initiation life for rolling contact fatigue based on dislocation and crack propagation. Int J Mech Sci 2008;50:117–23.
[126] Donzella G, Petrogalli C. A failure assessment diagram for components subjected to rolling contact loading. Int J Fatigue 2010;32:256–68.
[127] Leonel ED, Venturini WS. Multiple random crack propagation using a boundary element formulation. Eng Fract Mech 2011;78:1077–90.
[128] Komvopoulos K. Subsurface crack mechanisms under indentation loading. Wear 1996;199:9–23.
[129] Komvopoulos K, Cho S-S. Finite element analysis of subsurface crack propagation in a half-space due to a moving asperity contact. Wear 1997;209:57–68.
[130] Glodez S, Ren Z. Modelling of crack growth under cyclic contact loading. Theoret Appl Fract Mech 2000;30:159–73.
[131] Bogdanski S, Brown MW. Modelling the three-dimensional behaviour of shallow rolling contact fatigue cracks in rails. Wear 2002;253:17–25.
[132] Bogdański S, Trajer M. A dimensionless multi-size finite element model of a rolling contact fatigue crack. Wear 2005;258:1265–72.
[133] Choi Y, Liu CR. Rolling contact fatigue life of finish hard machined surfaces Part 1. Model development. Wear 2006;261:485–91.
[134] Fajdiga G, Ren Z, Kramar J. Comparison of virtual crack extension and strain energy density methods applied to contact surface crack growth. Eng Fract Mech 2007;74:2721–34.
[135] Raje N, Sadeghi F, Rateick RG, Hoeprich MR. A numerical model for life scatter in rolling element bearings. J Tribol 2008;130.
[136] Canadinc D, Sehitoglu H, Verzal K. Analysis of surface crack growth under rolling contact fatigue. Int J Fatigue 2008;30:1678–89.
[137] Slack T, Sadeghi F. Explicit finite element modeling of subsurface initiated spalling in rolling contacts. Tribol Int 2010;43:1693–702.
[138] Slack T, Sadeghi F. Cohesive zone modeling of intergranular fatigue damage in rolling contacts. Tribol Int 2011;44:797–804.
[139] Warhadpande A, Sadeghi F, Kotzalas MN, Doll G. Effects of plasticity on subsurface initiated spalling in rolling contact fatigue. Int J Fatigue 2012;36:80–95.
[140] Weinzapfel N, Sadeghi F. Numerical modeling of sub-surface initiated spalling in rolling contacts. Tribol Int 2012.
[141] Santus C, Beghini M, Bartilotta I, Facchini M. Surface and subsurface rolling contact fatigue characteristic depths and proposal of stress indexes. Int J Fatigue 2012;45:71–81.
[142] Lin JF, Chu HY. A numerical solution for calculating elastic deformation in elliptical-contact EHL of rough surface. Trans ASME, J Tribol 1991;113:12–21.
[143] Biao LS, Ju MJ, Yang CX. Elliptoc-paraboloid method for calculating surface elastic deformation in EHL. Tribol Int 1993;26(6):443–8.
[144] Fatemi A, Vangt L. Cumulative fatigue damage and life prediction theories: a survey of the state of the art for homogeneous materials. Int J Fatigue 1998;20(1):9–34.
[145] Xiao Y, Li S, Gao Z. A continuum damage mechanics model for high cycle fatigue. Int J Fatigue 1998;20(7):503–8.
[146] Shang D-Guang, Yao W-Xing. A nonlinear damage cumulative model for uniaxial fatigue. Int J Fatigue 1999;21:187–94.
[147] Liu Y, Liang LH, Hong QC, Antes H. Non-linear surface crack analysis by three dimensional boundary element with mixed boundary conditions. Eng Fract Mech 1999;63:413–24.
[148] Alfredsson B, Dahlberg J, Olsson M. The role of a single surface asperity in rolling contact fatigue. Wear 2008;264:757–62.
[149] Hannes D, Alfredsson B. Rolling contact fatigue crack path prediction by the asperity point load mechanism. Eng Fract Mech 2011;78:2848–69.
[150] Hannes D, Alfredsson B. Surface initiated rolling contact fatigue based on the asperity point load mechanism—a parameter study. Wear 2012;294–295:457–68.
[151] Jouini N, Revel P, Mazeran P-E, Bigerelle M. The ability of precision hard turning to increase rolling contact fatigue life. Tribol Int 2012.
[152] Chang L, Jackson A. A study of asperity interactions in EHL line contacts. Tribol Trans 1993;36(4):679–85.
[153] Heyes M. Molecular lubrication aspects of boundary. Tribol Int 1996;29(8):627–9.
[154] Li Q, Loughran J, Peng ZX, Osborne J. A fracture based model for wear debris formation. Key Eng Mater 2006;324–325:1157–60.
[155] Fletcher DI, Hyde P, Kapoor A. Modelling and full-scale trials to investigate fluid pressurisation of rolling contact fatigue cracks. Wear 2008;265:1317–24.
[156] Balcombe R, Fowell MT, Dini D. Modelling rolling contact fatigue cracks in the hydrodynamic lubrication regime: a coupled approach. Proc Eng 2009;1:245–8.
[157] Balcombe R, Fowell MT, Olver AV, Ioannides S, Dini D. A coupled approach for rolling contact fatigue cracks in the hydrodynamic lubrication regime: the importance of fluid/solid interactions. Wear 2011;271:720–33.
[158] Tichy JA. A surface layer model for thin film lubrication. Tribol Trans 1995;38(3):577–82.
[159] Fernandez Rico JE, Hernandez Battez A, Garcia Cuervo D. Rolling contact fatigue in lubricated contacts. Tribol Int 2003;36:35–40.
[160] Lainé E, Olver AV, Beveridge TA. Effect of lubricants on micropitting and wear. Tribol Int 2008;41:1049–55.
[161] Glodež S, Potočnik R, Flašker J, Zafošnik B. Numerical modelling of crack path in the lubricated rolling–sliding contact problems. Eng Fract Mech 2008;75:880–91.
[162] Zhu WS, Neng YT. A theoretical and experimental study of EHL lubricated with grease. Trans ASME, J Tribol 1988;110:38–43.
[163] Chang L, Webster MN. A study of elastohydrodynamic lubrication of rough surfaces. Trans ASME, J Tribol 1991;113:110–5.
[164] Khonsari MM, Hua DY. Generalised non-Nowtonian elastohydrodynamic lubrication. Tribol Int 1993;26(6):405–11.
[165] Dowson D. Elastohydrodynamic and micro-elastohydrodynamic lubrication. Wear 1995;190:125–38.
[166] Dowson D, Ehret P. Past, present and future studies in elastohydrodynamics. Int Mech Eng 1999;213(J):317–33.
[167] Sugimura J, Okumura T, Yamamoto Y, Spikes HA. Simple equation for elastohydrodynamic film thickness under acceleration. Tribol Int 1999;32(2):117–23.
[168] Demirel AL, Granick S. Transition from static to kinetic friction in a model lubricated system. J Chem Phys 1998;109(16):6889–97.
[169] Stewart S, Ahmed R. Rolling contact fatigue of surface coatings — a review. Wear 2002;253:1132–44.
[170] Hwang D, Gahr KZ. Transition from static to kinetic friction of unlubricated or oil lubricated steel/steel, steel/ceramic and ceramic/ceramic pairs. Wear 2003;255:365–75.
[171] Wang J, Hashimoto T, Nishikawa H, Kaneta M. Pure rolling elastohydrodynamic lubrication of short stroke reciprocating motion. Tribol Int 2005;38:1013–21.
[172] Maeda Y, Iwasaki M, Kawafuku M, Hirai H. Nonlinear modeling and evaluation of rolling friction. In: IEEE international conference on mechatronics, vol. 00; 2009.
[173] Ciulli E, Stadler K, Draexl T. The influence of the slide-to-roll ratio on the friction coefficient and film thickness of EHD point contacts under steady state and transient conditions. Tribol Int 2009;42:526–34.
[174] Edwin L-J. Numerical model to study of contact force in a cylindrical roller bearing with technical mechanical event simulation. J Mech Eng Automat Aug. 2011;1(1).
[175] Lu X, Khonsari MM, Gelinck ERM. The stribeck curve: experimental results and theoretical prediction. Trans ASME, J Tribol 2006;128:789–94.
[176] Lord J, Larsson R. Film-forming capability in rough surface EHL investigated using contact resistance. Tribol Int 2008;41:831–8.
[177] Schlijper AG, Scales LE, Rycroft JE. Current tools and techniques for EHL modelling. Tribol Int 1996;29(8):669–73.

[178] Tasgetiren S, Aslantas K. A numerical study of the behaviour of surface cracks under dry-sliding conditions. Mater Des 2003;24:273–9.
[179] Evans HP, Snidle RW, Sharif KJ, Bryant MJ. Predictive modelling of fatigue failure in concentrated lubricated contacts. Roy Soc Chem 2012;156:105–21.
[180] Yang Z, Chung Y-W, Cheng HS. Lubricant effects in the transition from boundary to microelastohydrodynamic lubrication. Tribol Trans 1996;39(4):974–8.
[181] Polonsky IA, Keer LM. A numerical method for solving rough contact problems based on the multi-level multi-summation and conjugate gradient techniques. Wear 1999;231:206–19.
[182] Jubault I, Molimard J, Lubrecht AA, Mansot JL, Vergne P. In situ pressure and film thickness measurements in rolling/sliding lubricated point contacts. Tribol Lett 2003;15(4):421–9.
[183] Molimard J, Querry M, Vergne P, Krupka I, Hartl M. Calculation of pressure distribution in EHD point contacts from experimentally determined film thickness. Tribol Int 2005;38(4):391–401.
[184] Popov VL, Psakhie SG. Numerical simulation methods in tribology. Tribol Int 2007;40:916–23.
[185] Wang Z, Nakamura T. Simulations of crack propagation in elastic–plastic graded materials. Mech Mater 2004;36:601–22.
[186] Zhang Z, Liu G, Liu TX, Wang QJ. Integrated adaptive element free Galerkin model for elastoplastic contact of engineering surfaces. Tribology 2007;1(2):87–97.
[187] Elsharkawy A. Visco-elastohydrodynamic lubrication of line contacts. Wear 1996;199:45–53.
[188] Luo JB, Liu S. The investigation of contact ratio in mixed lubrication. Tribol Int 2006;39:409–16.
[189] Choo JW, Olver AV, Spikes HA. The influence of transverse roughness in thin film, mixed elastohydrodynamic lubrication. Tribol Int 2007;40(2):220–32.
[190] Anandan N, Pandey RK, Jagga CR. An efficient numerical analysis of starved thermohydrodynamically lubricated rolling line contacts. Tribol Int 2008;41(9–10):940–6.
[191] Carli M, Sharif KJ, Ciulli E, Evans HP, Snidle RW. Thermal point contact EHL analysis of rolling/sliding contacts with experimental comparison showing anomalous film shapes. Tribol Int 2009;42:517–25.
[192] Jiang J, Stott FH, Stack MM. A mathematical model for sliding wear of metals at elevated temperatures. Wear 1995;183:20–31.
[193] Michael PC, Rabinowicz E, Iwasa Y. Thermal activation in boundary lubricated friction. Wear 1996;193:218–25.
[194] Pandey RK, Ghosh MK. A thermal analysis of traction in elastohydrodynamic rolling/sliding line contacts. Wear 1998;216(2):106–14.
[195] Pandey RK, Ghosh MK. Temperature rise due to sliding in rolling / sliding elastohydrodynamic lubrication line contacts: an efficient numerical analysis for contact zone temperatures. Tribol+Schmierungstechnik 1999;31(12):745–52.
[196] Jeng Y-R, Huang PY. Temperature rise of hybrid ceramic and steel ball bearings with oil-mist lubrication. Lubr Eng 2000:18–23.
[197] Stott F, Jordan M. The effects of load and substrate hardness on the development and maintenance of wear-protective layers during sliding at elevated temperatures. Wear 2001;250:391–400.
[198] Stott FH. High-temperature sliding wear of metals. Tribol Int 2002;35:489–95.
[199] Liu CR, Choi Y. Rolling contact fatigue life model incorporating residual stress scatter. Int J Mech Sci 2008;50:1572–7.
[200] Choi Y. A study on the effects of machining-induced residual stress on rolling contact fatigue. Int J Fatigue 2009;31:1517–23.
[201] Guo Y, Barkey M. Modeling of rolling contact fatigue for hard machined components with process-induced residual stress. Int J Fatigue 2004;26:605–13.
[202] Schwach D, Guo Y. A fundamental study on the impact of surface integrity by hard turning on rolling contact fatigue. Int J Fatigue 2006;28:1838–44.
[203] Sedlaček M, Podgornik B, Vižintin J. Influence of surface preparation on roughness parameters, friction and wear. Wear 2009;266:482–7.
[204] Coe HH, Zaretsky EV. Effect of interference fits on roller bearing fatigue life. Cleveland, Ohio, USA; 1986.
[205] Cookson JM, Mutton PJ. The role of the environment in the rolling contact fatigue cracking of rails. Wear 2011;271:113–9.
[206] Mosleh M, Bradshaw K. Role of component configuration in evaluation of accelerated rolling contact fatigue of ball bearings. Wear 2011;271:2681–6.
[207] Mosleh M, Bradshaw K, Belk JH, Waldrop JC. Fatigue failure of all-steel and steel–silicon nitride rolling ball combinations. Wear 2011;271:2471–6.
[208] Nilsson RÅ, Olofsson U, Sundvall K. Filtration and coating effects on self-generated particle wear in boundary lubricated roller bearings. Tribol Finn J Tribol 2005;38:145–50.
[209] Guan von C, Schweitzer F, Köttritsch H. Cleanliness of rolling bearings a new environmental approach. Tribologie+Schmierungstechnik 2009;56(6):19–23.
[210] Fysh SD, Oravec EM, Medley JB. An experimental simulation of the tribology of large spherical roller bearings in paper machines. Tribol Int 1990;23(5):317–27.
[211] Kunc R, Zerovnik A, Prebil I. Verification of numerical determination of carrying capacity of large rolling bearings with hardened raceway. Int J Fatigue 2007;29:1913–9.
[212] Kim TH, Olver AV, Pearson PK. Fatigue and fracture mechanisms in large rolling element bearings. Tribol Trans 2008;44(4):583–90.
[213] Arakere NK, Subhash G. Work hardening response of M50-NiL case hardened bearing steel during shakedown in rolling contact fatigue. Mater Sci Technol 2012;28:34–8.
[214] Andersson S. Prediction of wear in rolling and sliding contacts; 1999.
[215] Boness RJ, McBride SL. Adhesive and abrasive wear studies using acoustic emission techniques. Wear 1991;149:41–53.
[216] Serrato R, Maru MM, Padovese LR. Effect of lubricant viscosity grade on mechanical vibration of roller bearings. Tribol Int 2007;40:1270–5.
[217] Massouros GP. Normal vibration of a plain bearing working under boundary lubrication conditions. Tribol Int 1983;16(5):235–8.
[218] Peng Z, Kessissoglou N. An integrated approach to fault diagnosis of machinery using wear debris and vibration analysis. Wear 2003;255(7–12):1221–32.
[219] Ioannides E, Bergling G, Gabelli A. The SKF formula for rolling bearing life. Evol – Bus Technol Mag, SKF 2001:25–8.
[220] Dizdar S, Andersson S. Influence of plastic deformation on seizure initiation in a lubricated sliding contact. Wear 1999;232:151–6.
[221] Kotzalas MN, Doll GL. Tribological advancements for reliable wind turbine performance. Philos Trans. Ser A, Math, Phys, Eng Sci 2010;368:4829–50.
[222] Olofsson U, Andersson S, Bjorklund S. Simulation of mild wear in boundary lubricated spherical roller thrust bearings. Wear 2000;241:180–5.
[223] Massi F, Rocchi J, Culla A, Berthier Y. Coupling system dynamics and contact behaviour: modelling bearings subjected to environmental induced vibrations and 'false brinelling' degradation. Mech Syst Signal Process 2010;24(4):1068–80.
[224] Wijnat YH, Wensing JA, van Nijen GC. The influence of lubrication on the dynamic behaviour of ball bearings. J Sound Vib 1999;222(4):579–96.
[225] Maru MM, Castillo RS, Padovese LR. Study of solid contamination in ball bearings through vibration and wear analyses. Tribol Int 2007;40(3):433–40.
[226] Larsson PO, Larsson R, Jolkin A, Marklund O. Pressure fluctuations as grease soaps pass through an EHL contact. Tribol Int 2000;33:211–6.
[227] Mota V, Moreira P, Ferreira L. A study on the effects of dented surfaces on rolling contact fatigue. Int J Fatigue 2008;30:1997–2008.
[228] Halme J, Andersson P. Rolling contact fatigue and wear fundamentals for rolling bearing diagnostics-state of the art. J Eng Tribol 2009;224:377–93.
[229] Söchting S, Sherrington I, Lewis SD, Roberts EW. An evaluation of the effect of simulated launch vibration on the friction performance and lubrication of ball bearings for space applications. Wear 2006;260:1190–202.
[230] Miettinen HL. The influence of the running parameters on the acoustic emission of grease lubricated rolling bearings. Maint Asset Manage 2000.
[231] Yhland EM. Waviness measurement-an instrument for quality control in rolling bearing industry. Proc Inst Mech Eng 1967;182:438–45.
[232] Sunnersjö CS. Rolling bearing vibrations – the effects of geometrical imperfections and wear. J Sound Vib 1985;98(4):455–74.
[233] Faik S, Witteman H. Modeling of impact dynamics : a literature survey. In: International ADAMS user conference; 2000.
[234] Johnson KL. Contact mechanics. Cambridge University Press; 1985.
[235] Holmberg K, Laukkanen A, Ronkainen H, Wallin K. Tribological analysis of fracture conditions in thin surface coatings by 3D FEM modelling and stress simulations. Tribol Int 2005;38(11–12):1035–49.

[236] Jacobson B. The Stribeck memorial lecture. Tribol Int 2003;36:781–9.
[237] Blau PJ. The significance and use of the friction coefficient. Tribol Int 2008;34:585–91.
[238] Borruto A, Taraschi I. Wear dependence on some factors characterizing the surface state: the hardness, the roughness and the surface degreasing. Wear 1995;184:119–24.
[239] Chue CH, Chung HH, Lin JF, Chou CC. The effects of strain hardened layer on pitting formation during rolling contact. Wear 2001;249:108–15.
[240] Aslantas K, Tasgetiren S. Debonding between coating and substrate due to rolling–sliding contact. Mater Des 2002;23:571–6.
[241] Kocich R, Cagala M, Crha J, Kozelsky P. Character of acoustic emission signal generated during plastic deformation. In: 30th European conference on acoustic emission testing & 7th international conference on acoustic emission, vol. 1; 2012.
[242] Zhi-qiang Z, Guo-lu L, Hai-dou W, Bin-shi X, Zhong-yu P, Li-na Z. Investigation of rolling contact fatigue damage process of the coating by acoustics emission and vibration signals. Tribol Int 2012;47:25–31.
[243] Kim H, Shaik NH, Xu X, Raman A, Strachan A. Multiscale contact mechanics model for RF–MEMS switches with quantified uncertainties. Modell Simul Mater Sci Eng 2013;21(8):085002.
[244] Begg CD, Merdes T, Byington C, Maynard K. Dynamics modeling for mechanical fault diagnostics and prognostics. In: Mechanical system modeling for failure diagnosis and prognosis, maintenance and reliability conference; 1999. no. Marcon 99.
[245] Gegner J. Tribological aspects of rolling bearing failures. In: Physics, Germany; 2011.
[246] Nierlich W, Gegner J. Material response models for sub-surface and surface rolling contact fatigue; 2007. p. 182–92.
[247] Gegner J, Wolfgang N. Operational residual stress formation in vibration-loaded rolling contact. In: JCPCS- International centre of diffraction data; 2009. p. 722–31.
[248] Yoshioka T. Detection of rolling contact subsurface fatigue cracks using acoustic emission technique. J Soc Tribol Lubr Eng 1992;49(4):303–8.
[249] Price ED, Lees AW, Friswell MI, Roylance BJ. Online detection of subsurface distress by acoustic emissions. Key Eng Mater 2003;246–246:451–60.
[250] Challen JM, Oxley PLB, Hockenhull BS. Prediction of Archard's wear coefficient for metallic sliding friction assuming a low cycle fatigue wear mechanism. Wear 1986;111:275–88.
[251] Jiang Y, Sehitoglu H. Rolling contact stress analysis with the application of a new plasticity model. Wear 1996;191:35–44.
[252] Jiang Y, Sehitoglu H. A model for rolling contact failure. Wear 1999;224:38–49.
[253] Kotzalas MN, Harris TA. Fatigue failure progression in ball bearings. Trans ASME, J Tribol 2001;123:238–42.
[254] Ciavarella M, Maitournam H. Letter to the editor: On the Ekberg, Kabo and Andersson calculation of the Dang Van high cycle fatigue limit for rolling contact fatigue. Fatigue fracture engineering material structure 2004;27:523–8.
[255] Wang Y, Hadfield M. A mechanism for nucleating secondary fractures near a pre-existing flaw subjected to contact loading. Wear Apr. 2003;254:597–605.
[256] Magalhães L, Seabra J, Sá C. Experimental observations of contact fatigue crack mechanisms for austempered ductile iron (ADI) discs. Wear Nov. 2000;246(1–2):134–48.
[257] Kabo E. Material defects in rolling contact fatigue- influence of overloads and defect clusters. Int J Fatigue 2002;24:887–94.
[258] Alexandre F, Deyber S, Pineau A. Modelling the optimum grain size on the low cycle fatigue life of a Ni based superalloy in the presence of two possible crack initiation sites. Scripta Mater Jan. 2004;50:25–30.
[259] Seo J, Kwon S, Jun H, Lee D. Numerical stress analysis and rolling contact fatigue of White Etching Layer on rail steel. Int J Fatigue Feb. 2011;33:203–11.
[260] Yoshioka T, Fujiwara T. Measurement of propagation initiation and propagation time of rolling contact fatigue cracks by observation of acoustic emission and vibration. Tribology Series 1987;12:29–33.
[261] Navarro A, Rios ER. Short and long fatigue crack growth – a unified model. Philos Mag 1988;57:15–36.
[262] Sun Z, Rios ER, Miller KJ. Modelling small fatigue cracks interacting with grain boundaries. Fatigue Fract Eng Meteorol 1991;14:277–91.
[263] Al-Ghamd AM, Mba D. A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size. Mech Syst Signal Process 2006;20:1537–71.
[264] Kakishima H, Nagatomo T, Ikeda H, Yoshioka T, Korenaga A. Measurement of acoustic emission and vibration of rolling bearings with an artificial defect. QR RTRI 2000;41(3):127–30.
[265] Kim Y-H, Tan ACC, Mathew J, Yang B-S. Condition monitoring of low speed bearings: a comparative study of the ultrasound technique versus vibration measurements. In: World congress of engineering asset management; 2006.
[266] Kappa B. Predicting bearing failures and measuring lubrication film thickness in your plants rotating equipment; 2006. p. 5969–74.

ELSEVIER

TRIBOLOGY
INTERNATIONAL

Dynamic modelling of wear evolution in rolling bearings

CrossMark
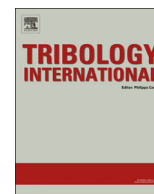
Idriss El-Thalji *, Erkki Jantunen

Industrial Systems, VTT Technical Research Centre of Finland, Espoo FI-02044 VTT, Finland

A B S T R A C T

Condition monitoring tools aim to monitor the deterioration process i.e. wear evolution of defects. The wear evolution is quite complex process due to the involvement of several wear and stress concentration mechanisms. Therefore, the purpose of this paper is to provide a dynamic model of wear evolution that considers the topographical and tribological changes over the lifetime. The model suggests the use of multiple force diagrams to simulate the dynamic impact and utilises several models of contact mechanics to estimate the transition points between the wear evolution stages. The simulated results of the developed evolution model are in principal agreement with the experimental results.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

The rolling element bearing (REB) is one of the most critical components that determine the machinery health and its remaining lifetime in modern production machinery. Robust condition monitoring (CM) tools are needed to guarantee the healthy state of REB during the operation. CM tools indicate the upcoming failures which provide more lead time for maintenance planning. CM tools aim to monitor the deterioration i.e. wear evolution rather than just detecting the defects. One way to get significant knowledge and insights concerning the wear process and how to monitor its evolution is by dynamic modelling. However, the current dynamic models in the literature lack to model the dynamic response over the whole life time. The wear evolution process is quite complex due to the involvement of several wear mechanisms (i.e. fatigue, abrasive, adhesive, corrosive) and several stress concentration mechanisms (i.e. dent, asperities, debris, sub-surface inclusions). These involvements and their interactions and competitions produce a wear evolution progress which varies significantly with respect to surface topographical and tribological changes. Thus, the wear evolution model requires a continuous two-way feedback between the contact mechanics and the dynamic mechanics processes over the REB's lifetime.

Over the years, several dynamic models have been developed to investigate the dynamic behaviour and features of REBs. The dynamic models of REB were first introduced by Palmgren [1] and Harris [2]. However, total non-linearity and time varying

characteristics were not addressed at that time. After that Gupta [3] completed the first dynamic model of REB and later Fukata et al. [4] presented a comprehensive non-linear and time variant model. The more advanced issues of time variant characteristics and non-linearity were raised and studied by several authors. For example, Wijnat et al. [5] reviewed the studies concerning the effect of the Elasto-Hydrodynamic Lubrication (EHL) on the dynamics of REB. Tiwari and Vyas [6] and Tiwari et al. in [7,8] studied the effect of the ball bearing clearance on the dynamic response of a rigid rotor. Sopanen and Mikkola [9] reviewed different dynamic models with the discussion of the effect of waviness, EHL, and localised faults and clearance effect. Later, the finite element method (FEM) was used to provide more accurate results. Kiral and Karagülle [10] presented a defect detection method using FEM vibration analysis for REBs with single and multiple defects. The vibration signal includes impulses produced by the fault, modulation effect due to non-uniform load distribution, bearing induced vibrations, and machinery induced vibrations and the noise which is encountered in any measurement system. Sopanen and Mikkola [9] implemented the proposed ball bearing model using a commercial multi-body system software application MSC.ADAMS. Endo [11] developed a 16-degree-of-freedom (DOF) model of a gearbox in order to simulate spall and cracks in the gear teeth. First, the FEM model was utilized to simulate the variation of the mesh stiffness for two types of faults under varying static load conditions. Then the model was integrated into the lumped parameter dynamic model. The study obtained the dynamic transmission error and acceleration responses under different loads and speeds. Sawalhi and Randall [12] developed a 34-DOF model of a gearbox in order to simulate spall and cracks in the REB based on Endo's model of 16-DOF. This model includes extra 18-DOF due to the consideration of a five-

DOF bearing model and the consideration of translational DOF both along the line of action and perpendicular to it. Massi et al. [13] studied the wear that is resulting from false Brinelling at the contact surfaces between the balls and races of the bearings. Several models have been developed to study the effects of several distributed and localized defect on REB dynamics: clearance effect [8,9,14–16], waviness effect [17,9,15], disturbances effect of EHL [9,12], and the effect of localized faults [18–20] etc.

In the past, the largest share of the studies has focused on the localized faults using different modelling techniques. McFadden and Smith [18], McFadden and Smith [19], Tandon and Choudhury [20] and Sawalhi and Randall [12] simulated the defect as a signal function of impulsive train into the modelled system. For example, Tandon and Choudhury [20] have introduced the defect as pulse function with three different pulse shapes: rectangular, triangular and half-sine pulse. Wang and Kootsookos [21] introduced defects as a function of basic impulse series. Ghafari et al. [22] have virtually introduced a defect into the equation of motion as a triangular impulse train at the related characteristic frequencies of a defect. Rafsanjani et al. [23] modelled the localized defects as a series of impulses having a repetition rate equal to the characteristics frequencies. The amplitude of the generated impulses is related to the loading and angular velocity at the point of contact. Malhi [24], Kiral and Karagülle [10], Sopanen and Mikkola [9], Massi et al. [13] and Liu et al. [25] introduced the defect as force function into their FEM models i.e. as a constant impact factor. More preciously Liu et al. [25] introduced the localized defect as a piecewise function. Ashtekar and Sadeghi [26], Sassi et al. [27], Cao and Xiao [15], Rafsanjani et al. [23], Patil et al. [28], and Tadina and Boltezar [29] modelled the defect based on its geometrical features i.e. as a surface bump or a dent that has length, width and depth. Tadina and Boltezar [29] modelled the defect as an impressed ellipsoid on the races and as flattened sphere for the rolling elements. Nakhaeinejad [16] utilised the bond graphs to study the effects of defects on bearings vibrations. The model incorporated gyroscopic and centrifugal effects, contact deflections and forces, contact slip and separations, and localized faults. Dents and pits on inner race, outer race and balls were modelled through surface profile changes i.e. type, size and shape of the localized faults.

However, these dynamic models deal with wear phenomenon as a localized defect with fixed features over the lifetime. The reason is that the purpose of these models is to detect the defect within the generated vibration signals and not the incremental deterioration process i.e. wear evolution. These dynamic models start from the point where the defect is localized as simulated defect in the models or artificially introduced into experiments. That ignores the prior-stage(s) of the localization process. These dynamic models assume that the localized defects and their associated impact stay constant over the whole lifetime. That ignores the topographical and tribological changes of the defected surface. In order to model the wear evolution, an incremental numerical procedure should be developed that is able to integrate the contact information continuously into the dynamic model. This means that the applied force due to the wear progress and its associated topographical and tribological conditions should be iteratively updated into the dynamic model.

Physics-based prognostic models describe the physics of the system and failure modes based on mathematical models such as Paris' law, Forman law, fatigue spall model, contact analysis and stiffness based damage rule model. Physics-based prognostic models are based on crack length, and defect area as illustrated by Li et al. [30], and Li et al. [31], or relations of stiffness as shown by Qiu et al. [32]. However, the most challenging issue within physics-based prognostic is to define the loading–damage relationship and to model it. There are models based on damage rules as linear damage rule, damage curve rule, and double-linear damage rule [32]. The drawback of these simplified functions is

that they all use the constant damage factor which is hard to estimate or measure. Moreover, these functions are either linear or multi-linear functions. That means that the estimated results might seem matching with the overall measured results, however, both of them might describe different damage scenarios in behind. Therefore, the prediction based on such functions makes the prognosis a risky task. Recently, some model-based models have been utilised for the contact stress analysis to illustrate the wear evolution progress. These models provide more accurate predictions. Some models are based on contact stress analysis [33] and some are based on system dynamics [34], [35]. Chelidza and Cusumano [36] proposed a method based on a dynamical systems approach to estimate the damage evolution. The results of these models depend also on the stress-damage function and the constant damage factor that are in use. These models assume that each wear mechanism generates stresses that in total equal to the overall measured stresses. Therefore, the wear mechanics interactions and competitions are somehow ignored.

Recently, El-Thalji and Jantunen [37] presented an updated framework of wear evolution in REBs. It is based on the well-known three stage model (i.e. proposed by Voskamp [38]) and most recent studies in rolling contact wear. The updated framework is simply described as a five-stage scenario: running-in, steady-state, defect initiation (dentation, micro-cracking, inclusions), defect propagation (pits, propagated cracks), and damage growth (spalling). Thus, the advantage is to explain more effectively the instability stage which is sub-divided into three stages. That helps us to explain the general nature of wear evolution i.e. the wear interactions and competitions in more detailed manner for modelling and monitoring purposes. It provides better understanding of wear evolution stages, the involved wear mechanisms in each stage, the interaction among wear mechanisms in each stage, and the influencing factors within each stage. The challenge is how to develop a dynamic model that can illustrate this five stage model that considering the wear interactions and competitions due to the involvement of wear and stress concentration mechanisms. First, the modelling procedure should be capable to integrate the contact mechanics model with dynamic model, for iterative feedbacks. El-Thalji and Jantunen [39] have discussed such modelling integration for REBs. Second, the modelling procedure should be able to estimate the wear interactions and competitions over the lifetime. Third, it should be able to determine the transition events and their impact on the system dynamic behaviour e.g. dentation event as a transition between steady-state stage and defect initiation stage. Therefore, the purpose of this paper is to provide a dynamic model of wear evolution. The dynamic model is presented for studying the dynamic behaviour and properties of a single-degree-of-freedom system under the effects of wear evolution process over the whole lifetime. The dynamic model requires a number of contact mechanics models in order to model the transition between wear evolution stages. Therefore, the contact mechanics models will be discussed to illustrate the determination of the wear transition events, beside the wear interactions and competitions. In the end, the study aims to develop a model that can generate non-linear wear evolution which can be used for prognosis purposes. The paper begins with presenting a description of wear evolution. Later, the analytical dynamic model is explained. The results part describes the simulated dynamic response of each wear progression stages and discusses the simulated with respect to the measured data.

2. The developed dynamic model of wear evolution

A real shaft-bearing system is generally very complicated and difficult to model [40]. In addition to that, the modelling of wear evolution in REBs makes it even more complicated. Thus, the developed model obtains simple equations of motion based on

single-degree-of-freedom system. The dynamic simulation models are basically described with help of the following equation of motion:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F\} \qquad (1)$$

where $[M], [C]$, and $[K]$ are respectively matrices of system mass (es), damping coefficient(s) and stiffness(es). $\{F\}$ is the excitation force matrix.

Surface defects i.e. wear influence the dynamic response of the system, in particular, the mass, damping ratio, stiffness and excitation force. The removed material from the surface influences the mass, either reduce or change the uniform distribution. The surface defect disturbs the uniformity of the lubrication film and leads to some changes in damping ratio. When a dent or a defect occur, it will leave an empty space on the surface, which changes the stiffness i.e. curvature difference and summation of the rolling contact. The defect also introduces an excitation force when the rolling element passes through the edges (i.e. impact areas) of the defected surface. However, the developed model considers the effects of mass and damping ratio are small and therefore they are ignored in this study. Moreover, the model considers the surface dentations as damage initiator rather than the sub-surface defect. The sub-surface cracking is considered in the crack propagation stage of the model.

It is worth to explain that several bearing manufacturers indicate with their lifetime estimation formula, e.g. the SKF bearing rating life formula, that the load and capacity are the main influencing factors, assuming that the lubrication, oil contamination and operating conditions are ideal. Based on this assumption, their lifetime prediction show quite long lifetime and they fit very well with applications where the influences of the lubrication, oil contamination and operating conditions are controlled and neglected. The damage in the bearings of such applications are related to sub-surface defect which appear after long time of operation i.e. after high repeated stress cycles, inclusions in the sub-surfaces and the degradation of the material properties.

However, several studies in the literature [41–43], etc. show that under the normal loading conditions, the damage might appear much earlier once the lubrication, oil contamination and operating conditions are disturbed. Moreover, the rating life formula expresses the capacity to load ratio with an exponential factor that is related to contact type i.e. ball contact, rolling contact. Therefore, the studies show that lubrication disturbances might generate boundary lubricated contact, abrasive actions, pressing particles between surfaces, etc. and these actions in fact initiate the surface dentations. Therefore, the "service life factors" have been introduced to the life rating formula including lubrication, the degree of contamination, misalignment, proper installation and environmental conditions. These service life factors can easily influence the bearing lifetime at early stage and initiate surface dentations much earlier than sub-surface defects, in particular, for bearings that are operated in normal loading conditions (with some degree of variation) and assuming high material quality (which have been used in other applications and showed long lifetime). In fact, several studies [44,45,43], etc. highlighted that the probability of getting disturbances in lubrication, contamination degree and operating conditions are higher than having sub-surface inclusions at early stage.

Moreover, the model relies on the surface dentations which might influence the dynamic response in more measurable manner compare to the sub-surface defect at early stage of wear evolution process.

Thus, two effects of wear evolution are considered in the dynamic simulation model: stiffness and excitation forces. Therefore, the effect of surface defect can be introduced to the previous equation as following:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K+K_F]\{y\} = \{F+F_F\} \qquad (2)$$

where, K_F is the stiffness change due to specific defect and F_F is the defect-induced forces. Therefore, most of modelling studies that discussed in the introduction part have implemented the same approach of considering surface defect into the dynamic models. The changes were in the size and shape of the modelled surface defect.

However, the size and shape of the surface defect due to wear are changing over the lifetime. One of the most probable scenario of wear evolution is presented in [46]. The scenario highlights that there are mainly four stress concentration mechanisms which involved in rolling contact: subsurface inclusions, asperities, dents and debris (as moving asperities). These mechanisms concentrate and allocate the stresses at a specific point or region. Cracks are initiated due to material in-homogeneities and cyclic stresses. Later, cracks propagate toward the surface. The cracks propagate in parallel to the surface, until it meets secondary cracks that connect it to the surface. Asperities might be expressed in the form of surface roughness. However, the most significant asperities are initiated by dents, abraded points, ploughed points, etc. A dent can be described in two dimensions as the stress raiser of two up-ward asperities (wedges). The asperity acts as a stress raiser which results in more cracking actions. When the new cracks reach the surface, new material will be detached and new asperities will be initiated. That might explain the phenomena of stress increasing while passing the detected surface zone over time. Moreover, the asperity might act as minor indenter once it is hard and sharp enough. However, the asperity is also subjected to over-rolling, adhesive and abrasive actions, once the operating conditions are allowing that. The adhesive and abrasive action might reduce the asperity and its effect as stress raiser. This issue considered as a wear competition with rolling contact and should be considered to model the non-linear progress of the wear evolution. This brief description of wear evolution is schematically illustrated in Fig. 1.

The wear evolution model in Fig. 1 provides a good idea of how the defect size and shape change over the time. This information is required to simulate the changes in the stiffness and defect-induced forces (excitation forces) within the dynamic model. The literature which have been discussed in [46] show that the surface topography change due to wear can be represented as force functions. In Fig. 2, the schematic representation of the surface topography change due to wear is shown together with approximate force functions (right side). The surface topography changes i.e. dentation, cracking, defect, and damage growth are considered. This representation is capable to consider the defect features (length, depth and width of the defect), issues related to the trailing edge features i.e. shoulder of asperity, sharpness, and other issue related to the number of defects i.e. multi asperities and their chain effects i.e. speed difference. However, in order to simulate the wear evolution over the whole lifetime a couple of issues should be considered:

1. The K_F and F_F should be estimated with respect to the surface topographical change due to wear evolution process. Thus, the K_F and F_F are functions of time and will be inserted to the equation of motion continuously over the time. However, it is clear that the change in the K_F and F_F are related to the change of the surface topography. Therefore, the changes are related to certain events of the wear evolution process e.g. dentation, cracking, defect, and damage growth. Knowing when those event might occur, it provides us the time when the K_F and F_F should be changed to the updated ones.

2. The transition points between wear evolution stages should be estimated with the help of stress analysis. The basic idea at first, is to estimate the stress levels which are sufficient to generate a certain surface topographical change. After the first step the stress which is generated by the dynamic model over the time should be accumulated and compared with the determined
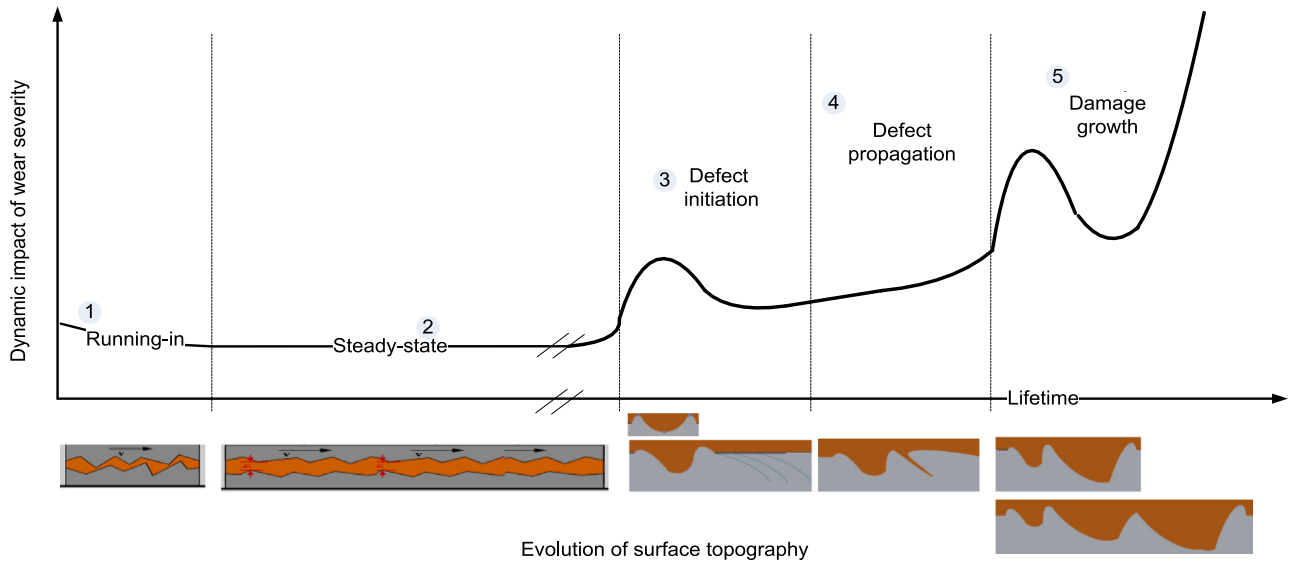
Fig. 1. Lifetime progress of rolling contact wear (the below surface topographies are explained further in Fig. 2).
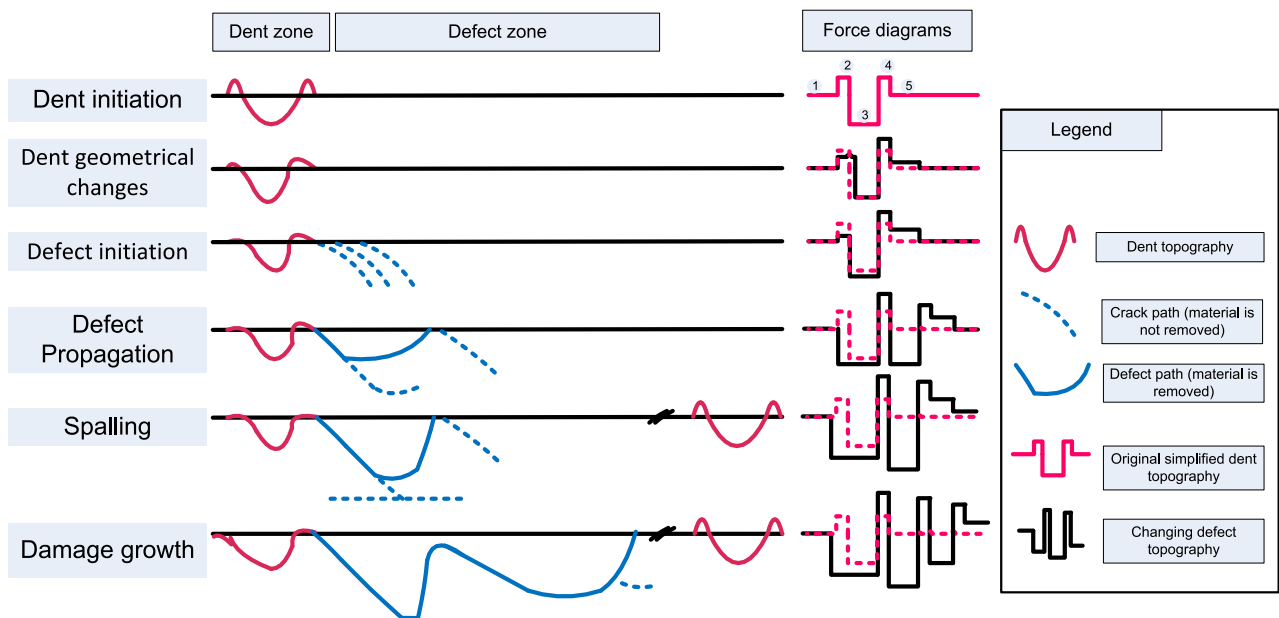

Fig. 2. The surface topographical changes during the wear evolution process.

stress levels. The stress history can be extracted based on the time history to estimate when certain surface topographical events might happen. In order to determine these three issues, contact mechanics and dynamic analyses are needed. The contact mechanics analysis will provide the induced stress due to the different defected surface topologies. In addition the contact mechanics analysis is needed to estimate the accumulated stresses in order to determine the transition events between the five stages of the wear evolution progress. El-Thalji and Jantunen [39] have discussed the modelling integration between dynamic and contact mechanics for wear evolution in REBs.

The basic estimation principle of the induced stresses due to wear evolution is based on influence of the surface asperities that are generated by the topographical change. A surface asperity generates concentrated stresses at certain point of line in the rolling contacted area. Therefore, the fatigue wear theory is highly dependent on the asperity size and shape. However, as it is

described in literature [46], the wear process is a process of multiple wear and stress concentration mechanisms and there are interactions and competitions among those mechanism. For example, the surface asperity induced stresses is sufficient to initiate and propagate the fatigue wear process. At the same time, this asperity might be smoothen by the over-rolling and abrasive wear actions, which cause a reduction in the induced stresses in return. Therefore, the wear interactions and competition should be considered as well. However, the wear modes that are modelled in this paper are fatigue and abrasive wear, besides the influence of over-rolling mechanism.

In summary, this modelling procedure requires iterative update of stiffness parameters and excitation forces (which change due to the wear evolution progression) in order to provide the system response over the studied lifetime. In order to do that, the modelling procedure should be capable to determine four issues:

1. The stiffness parameter for each certain defected surface topography.

2. The excitation force for each certain defected surface topography.
3. The interaction events i.e. wear competition, over-rolling smoothing, abrasive wear.
4. The transition events i.e. conditions between the five stages of the wear evolution progress.

The flowchart in Fig. 3 illustrates the procedures of the dynamic simulation model. The flowchart shows that the model starts and accumulates the stresses. When the accumulated stresses reach certain value, this means a dentation process is started as well. The generated dent will have specific asperity that produce stress concentration and again the generated stresses will be accumulated. When the accumulated stresses reach certain value, the crack is opened. In the same time, the asperities will be under over-rolling and abrasive wear actions. That means the generated stress due to dents' asperities will be reduced, however, they will still sufficient to open the crack. The crack will propagate and the associated stresses will be again accumulated. When the accumulated stresses reach a certain value, the defect will be completed. The model will keep repeating the previous procedures and the damage will grow rapidly. The accumulated stresses due to the damage growth process reach a certain value, the failure is occurred.

2.1. Stiffness estimation for certain surface topography

The Hertzian contact theory is used to estimate the stiffness parameter due to certain defected surface topography. The contact stiffness depends on the curvature sum and the difference. Therefore, the defect is considered as an object generates clearance. The developed model follows Harris' method [47] of how to estimate the elastic and permanent deformation in rolling bearings. Later, the change in the curvature parameters due to the deformation is considered in the equation of motion as presented by Tadina and Boltezar [29]. A very commonly used formula is the force-indentation relation for sphere to sphere contact [48]:

$$F = K\delta^{3/2} \tag{3}$$

where, F is normal force pressing the solids together. δ is the deflection of the two spheres, i.e. the total of the deformation of both surfaces. K is a constant depending on the sphere radii and elastic properties of the sphere materials.

2.2. Excitation force modelling due to certain surface topography

A shock is the transmission of kinetic energy to a system, occurring in a relatively short time. Sassi, et al. [27] modelled the impact force due to a defect based on the defect length. Assuming that the system is conservative, the conservation of mechanical energy between state 1 (before shock) and state 2 (after shock) is given as follow [27]:

$$\left[\frac{1}{2}mV_1^2 + \frac{1}{2}I\omega_1^2\right] + mg\frac{B_d}{2} = \left[\frac{1}{2}mV_2^2 + \frac{1}{2}I\omega_2^2\right] + mg\frac{B_d}{2}\cos(\theta) \tag{4}$$

where m the rolling element is mass, I is the rolling element mass moment of inertia. V_2 and V_2 present the linear velocities of the rolling element, before and after the shock events, respectively. ω_1 and ω_2 present the angular velocities of the rolling element, before and after the shock events, respectively. The B_d is the rolling element diameter and θ is the contact/impact angle as shown Fig. 4. The equation highlights that the impulse is a change in momentum. In an impact of very short time duration (between an initial time t_i and a final time t_f), the impact force $F_{Impulse}$ is typically very large. With help of Fig. 4, the impact force $F_{Impulse}$ can be estimated.

The mg (the gravitational force) is much smaller than $F_{Impulse}$. Therefore, the gravity can be ignored for the impulse calculation. For planar motion in the xy plane, the equations for impulse and linear momentum are [49]:

$$mv_{Gxi} + \int_{ti}^{tf} F_{Px}dt = mv_{Gxf}, \text{ and } mv_{Gyi} + \int_{ti}^{tf} F_{Py}dt = mv_{Gyf}$$

$$\int_{ti}^{tf} F_{Px}dt = mv_{Gxf} - mv_{Gxi} \tag{5}$$

$$\int_{ti}^{tf} F_{Py}dt = mv_{Gyf} - mv_{Gyi} \tag{6}$$

where v_{Gxi} is the velocity of the centre of mass G in the x-direction before impact, and v_{Gxf} is the velocity of the centre of mass G in the x-direction after impact.

$$v_{Gxi} = -w_1 r \tag{7}$$

w_1 is the initial velocity of the rolling element and r is the rolling element radius. The negative sign accounts for the fact that positive angular velocity means the ball rolls to the left (in the negative x-direction). v_{Gyi} is the velocity of the centre of mass G in the y-direction before the impact, and v_{Gyf} is the velocity of the centre of mass G in the y-direction after the impact. Since the ball initially rolls on a flat horizontal surface [49], $v_{Gyi} = 0$, where $v_{Gxf} = -v_{Gf}\cos\theta$, $v_{Gyf} = v_{Gf}\sin\theta$, where v_{Gf} is the velocity of the center of mass, after impact, which is defined as $v_{Gf} = w_f r$, where w_f is angular velocity about point P on the tip of the asperity of the defect.

In order to define the impact forces with respect to the defect parameters e.g. length, depth, the following angles definitions will be substituted.

$$\cos\theta = \frac{r-h}{r}, \sin\theta = \frac{d_{length}/2}{r} \tag{8}$$

where h is the height of the asperity of the defect and d_{length} is the length of the defect. Therefore, the impact forces can be defined:

$$\int_{ti}^{tf} F_{Px}dt = -mw_1\left(1 - \frac{5h}{7r}\right)(r-h) + mw_1 r$$

$$\int_{ti}^{tf} F_{Py}dt = mw_1\left(1 - \frac{5h}{7r}\right)\left(\frac{d_{length}}{2}\right) \tag{9}$$

It can be noted that the impact force depends on defect depth, defect length, rolling element parameters i.e. mass, speed, and radius.

In this study, it is significantly important to estimate the induced stresses due the impact events. The impact stresses can be estimated based on the stress–force relationship.

$$F_{Impact} = \sigma_{Impact} \times A_{Impact} \tag{10}$$

The impact area is quite complicated to be estimated. However, it was assumed to consider the impact area as a pie slice that might be generated when the rolling element hits the asperity. The "pie" slice is illustrated schematically in Fig. 5. The segment is only the small partially curved figure left when the triangle is removed. The idea is to provide a simplified method to estimate the change in the impact area once the asperity and defect depth changes over the wear evolution process.

The impact area A_{Impact} and defect width d_{width} can be estimated as follows:

$$A_{Impact} = A_{segment} = A_{sector} - A_{triangle} = \frac{2\theta}{360}\pi r^2 - \frac{1}{2}r^2 \tag{11}$$

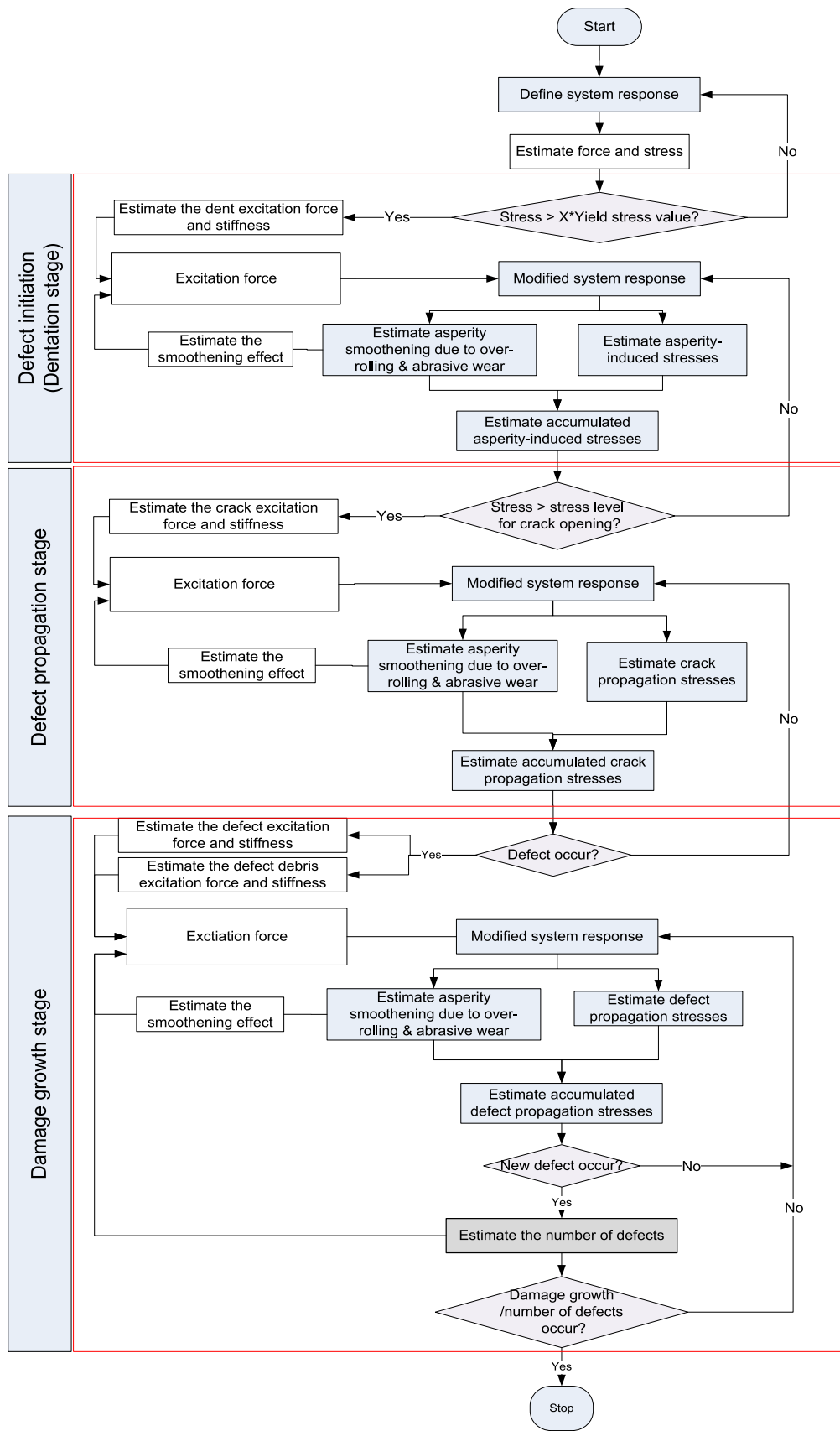$$d_{width} = 2r\sin\theta \tag{12}$$

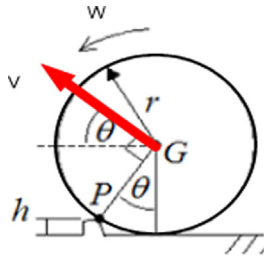**Fig. 3.** The dynamic simulation model of wear evolution process.

Fig. 4. Impact phenomenon in rolling contact.
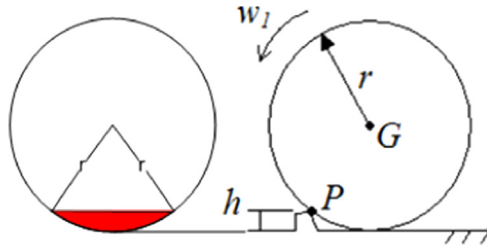

Fig. 5. Impact area of asperity and defect.

This part of analysis provides an illustration of how the induced stresses might be influenced by the defect size parameters and shape features (as shown in Fig. 6).

In the end, for each of surface topographical change, the K_F and F_F can be estimated as described in previous two sections.

2.3. Wear interaction events

The over-rolling and abrasive actions have potential to take a place within the wear process once the asperity interaction degree is allowing that. The over-rolling and abrasive actions act as asperity degradation process. Therefore, their actions are clear when the asperities of dents and defects are generated. It is clear that the impact effect of the asperity will be reduced as long as the asperity degrades. The developed model considers the degradation effect i.e. smoothening effect as degradation function. Sahoo and Banerjee [50] present an analysis of the effect of asperity interaction in elastic–plastic contact. Therefore, the developed model adopted their equation to estimate the asperity deformation due to the contact pressure at the asperity, which is given as follows:

$$\omega = z - d + 1.12\frac{\sqrt{w_a p_a}}{E} \tag{13}$$

where, ω is the deformation due to the contact pressure at the asperity, z is the height of a given asperity, d is the mean separation between the rigid flat and the rough surface, w_a is the contact pressure on a single asperity, p_a is the global mean contact pressure on the surface and E is the modulus of elasticity. More details of the deformation of asperity can be found in [50,51].

Abrasive wear depends on the lambda ratio (i.e. degree of asperity interaction). This is the ratio of lubricant film thickness to composite surface roughness and is given by the expression [34]:

$$\lambda = \frac{h}{\sqrt{R_{Surface\ 1}^2 + R_{Surface\ 2}^2}} \tag{14}$$

where λ is degree of asperity interaction, h is the lubricant film thickness, $R_{Surface\ 1}$ is the RMS roughness of the roller surface, and $R_{Surface\ 2}$ is the RMS roughness of the raceway. If λ is less than unity, it is unlikely that the bearing will attain its estimated design life because of the surface distress, which can lead to a rapid fatigue failure of the rolling surfaces. In general λ ratios greater than three indicate complete surface separation. A transition from full elasto
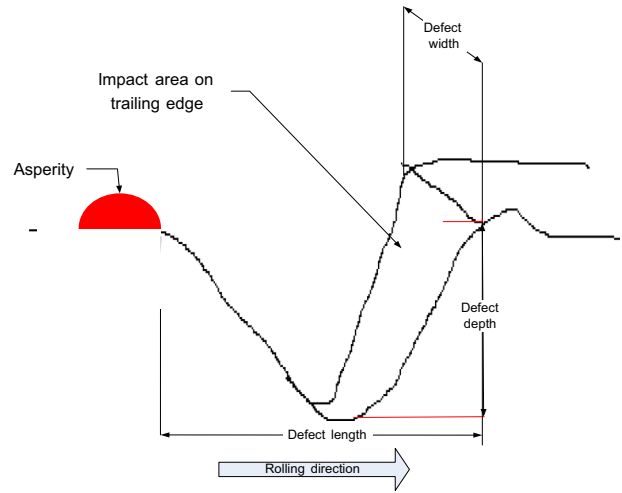

Fig. 6. The defect size parameters.

hydrodynamic lubrication (EHL) to mixed lubrication (partial EHL film with some asperity contact) occurs in the λ range between 1 and 3 [34].

Therefore, under specific tribological condition, the abrasive wear is taking a place in the asperity deformation as well. The abrasive wear depends of the height of the rolling element asperity which cut part of the surface asperity of the raceway. Therefore, the developed model has adopted the abrasive wear model provided by Masen et al. [52]. It is assumed that the abrasive wear process cut gradually over the time a specific amount of the surface asperity with specific depth (d). This process will continue until the surface asperity reach specific height where no interaction with rolling element asperities.

2.4. Wear transition points

The first transition event within the whole wear evolution progress is the transition from running-in stage to steady-state stage. However, this transition event is not with interest of the developed model. Therefore, the second transition event is between the steady state and defect initiation stage, in particular, when the dentations occur. There are several potential defect locations and four mechanisms of dentation i.e. high stress, contamination, vibration, and lubrication disturbances. In general, the dentation might occur the applied stresses reach the yield stress limit (Y). In the elastic–plastic stage, the plastic deformation is small enough to be accommodated by an expansion of the surrounding area. The model has adopted the formula in [47] to estimate the yield stresses for permanent deformation. Therefore, the model can estimate when the dentation occur by accumulating the applied stresses until the yield stress limit is reached.

The third transition point is between the dentation and crack opening stage. Several models show that the stress fields result in elastic and plastic deformations introduce some changes to the shape of the surface [52–57]. These models help to estimate how much induced stresses (to initiate and accelerate the crack opening) are applying on the surface due to the surface asperity. The asperity-induced stresses are accumulated either by the time for crack opening process or the accumulated loading cycles. Different subscripts are used to designate the stress intensity factor for different modes. The stress intensity factor for mode I is designated K_I and applied to the crack opening mode. The asperity induced-stresses σ_0 can be estimated by the formula of Hannes and Alfredsson [58]. The model assumes that the crack is straight. Therefore, the developed model utilise the asperity model to estimate the asperity-induced stresses and accumulate the

stresses until the crack opening limit is reached. The crack opening limit is determined by the stress intensity factor of mode I.

The fourth transition point is between the crack propagation and defect completion. Ringsberg and Bergkvist [59] studied crack length, crack angle, crack face friction and coefficient of surface friction near the contact load. Tsushima [60], Liu et al. [61], Liu and Choi [62], Donzella and Petrogalli [63] and Leonel and Venturini [64] have defined a number of issues that are required while modelling crack propagation: high stress location, depth below surface and the direction and angle of crack inclination. Therefore, based on these preliminary studies, quite large number of models related to crack propagation have been developed [65–78]. The developed model assumes that the crack propagation process will continue and the induced stresses will be accumulated until the crack propagation limit is reached. The crack propagation speed depends on the applied SIF. The crack propagation distance is accumulated until it reaches a pre-specified crack length. Navarro and Rios [79] and Sun et al. [80] proposed the model where the crack growth rate da/dN is assumed to be proportional to the crack tip plastic displacement δ_{pl}.

$$\frac{da}{dN} = C_0(\delta_{pl})^{m_0} \tag{15}$$

where C_0, and m_0 are material constants that are determined experimentally. The total number of stress cycles N required for a short crack to propagate from the initial crack length a_0 to any crack length 'a' can then be determined as [80]:

$$N = \sum_{j=1}^{z} N_j \tag{16}$$

Later, the damage growth stage will start and continue iteratively in a rapid growing manner. The developed model assumes that whenever the accumulated stress reach the crack propagation limit, a new defect is generated.

3. Results and discussion

A groove ball bearing (SKF 61810-4) was used for the modelling and testing purposes. The bearing inner diameter is 50 mm, the outer diameter is 65 mm and the weight is 0.052 kg. The results of the developed model are shown in Fig. 7 where the main wear events have been illustrated. The sudden increase in response amplitude after the dentation occurs, approximately around the middle of the lifetime, as shown in Fig. 7. The effect of over-rolling and abrasive actions is clearly observed after the dentation and defect completion stages. Later, the system response rapidly increased when the defect is completed and material is removed. The response rapidly increases due to the generation of new asperities and wear-debris, which means the damage growth, will continue. A number of bearings were tested in accelerated laboratory tests. The test arrangement is explained in [81] and the normalised RMS acceleration response is shown in Fig. 8. The acceleration measurements were accompanied with particle counting (HIAC PM4000) in some of the tests.

Both results in Figs. 7 and 8 indicate the steady-state nature of vibration acceleration during first half of the lifetime. However, the simulated results show some variation due to the surface dentation. Later, during the second half of the lifetime, a lot of variation is presented. Therefore, the simulated data illustrate the principal phenomena of dynamic response due to wear evolution process. The principal agreement that can be observed in the predicted response compared to the measured response is related to the principal stage and events of wear evolution process. However, the simulation is based on multiple models that estimate the response, transition conditions and stress accumulations with certain degree of uncertainties. These models that are used in the
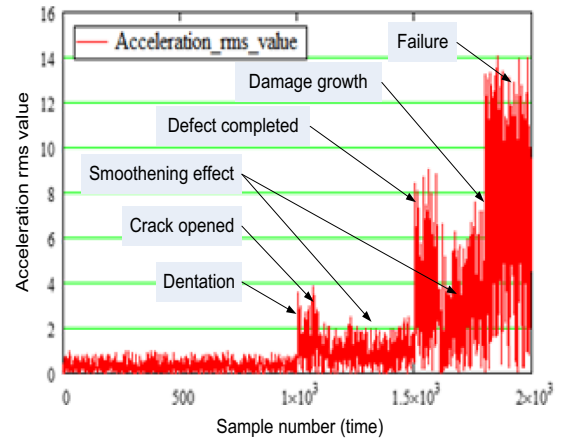

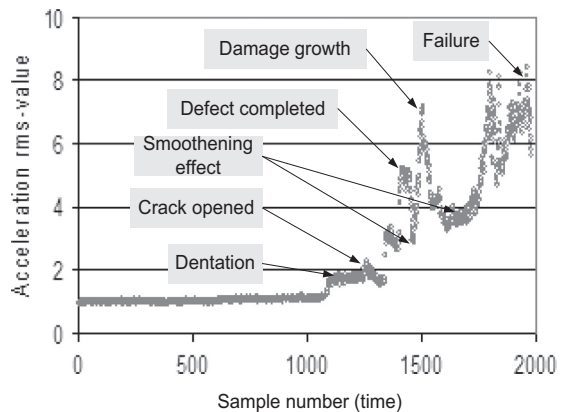Fig. 7. Normalized RMS value of vibration acceleration of simulated system.

Fig. 8. Normalized RMS value of vibration acceleration in laboratory tests. [76].

approach Hertz theory, approximated impact area, stress intensity factors, etc. are approximated models which can cumulate the uncertainties and give origin to error propagation.

Moreover, the measured data contains several issues related to different applications, boundary conditions, environments, third-body features, etc. At this stage of the model development the model aims to simulate the wear evolution process in simplified manner. These issues which are also discussed in detail in [46] might significantly influence the level of uncertainty of such a simplified model. Therefore, the further development of the model should address them in order to enhance the accuracy level.

This study intends to provide reliable interpretations of the measured data and to predict the future wear progress. Therefore, at this stage, it aims to achieve a principle agreement between the simulated and measured data in the first hand. However, for quantitative comparison purposes, the shift between simulated and measured curves is needed. It requires having a comparable timeline for both simulated and measured data sets. At this stage of model development, it is a heavy computational task to generate the simulated dynamic response for the whole lifetime e.g. 6 million cycles. Keep in mind, the ultimate goal is to interpret and predict at certain time intervals by samples, rather than to get a well fitted curve over the whole lifetime. This model might help to tune the multiple models that are used to estimate the response, transition conditions and stress accumulations and to gain better accuracy degree, once the whole lifetime picture is considered. In other words, these multiple models might tune each other to provide a better fit to the measured data in the end.

4. Conclusions

The industrial applications utilise the in-field measurements in order to plan the required maintenance actions in a cost effective manner. In the literature, the developed wear evolution models of for bearings have been developed by introducing a virtual defect into the model and then the dynamic responses have been calculated. The defect size parameters have been varied to present the development of the defect. This approach seems as discrete evolution model. This study describes a dynamic model of wear evolution that is capable to mimic the wear process in rolling bearings in a continuous manner. In fact, it is capable to mimic the topographical surface evolution due to wear in rolling bearings over the lifetime. The simulated results of the developed evolution model are in principal agreement with the experimental results. The wear evolution progress has been simulated in a continuous manner. The model suggests multiple force diagrams that are capable to simulate the wear evolution. The model is capable to simulate the dynamic nature of wear process, i.e. interaction among different wear mechanisms. The model utilises several models of contact mechanism in order to estimate the transition points between the wear evolution stages. Moreover, the developed model considers the surface asperity as key issue to: simulate its dynamic impact, estimate the induced stresses on the surface, estimate the wear competition due to over-rolling, and abrasive wear actions. The results represent and illustrate the fluctuation in the dynamic response of the rolling bearing under wear process over the whole lifetime. In this sense, the study shows how different wear and stress mechanisms can influence surface topography and the dynamic response of the system. Thus, it provides potential benefits for future enhancements in condition monitoring, diagnosis, and prognosis of rolling element bearings.

Acknowledgement

References

[1] Palmgren A. Ball and roller bearing engineering. Philadelphia: S.H. Burbank and Co. Inc; 1947.
[2] Harris FJ. Rolling bearing analysis I, First. New York: John Wiley & Sons, Inc.; 1966.
[3] Gupta PK. Transient ball motion and skid in ball bearings. Trans ASME, J Lubr Technol 1975:261–9.
[4] Fukata S, Gad EH, Kondou T, Ayabe T, Tamura H. On the vibration of ball bearings. Bull JSME 1985;28(239):899–904.
[5] Wijnat YH, Wensing JA, van Nijen GC. The influence of lubrication on the dynamic behaviour of ball bearings. J Sound Vib 1999;222(4):579–96.
[6] Tiwari R, Vyas NS. Dynamic response of an unbalanced rotor supported on ball bearings. J Sound Vib 1995;187(2):229–39.
[7] Tiwari M, Gupta K, Prakash O. Dynamic response of an unbalanced rotor supported on ball bearings. J Sound Vib 2000;238:757–79.
[8] Tiwari M, Gupta K, Prakash O. Effect of radial internal clearance of ball bearing on the dynamics of a balanced horizontal rotor. J Sound Vib 2000;238:723–56.
[9] Sopanen J, Mikkola A. Dynamic model of a deep-groove ball bearing including localized and distributed defects. Part 1: Theory. Proc Inst Mech Eng. Part K J Multi-body Dyn 2003;217(3):201–11 (Jan.).
[10] Kiral Z, Karagülle H. Simulation and analysis of vibration signals generated by rolling element bearing with defects. Tribol Int Sep. 2003;36(9):667–78.
[11] Endo H. A study of gear faults by simulation and the development of differential diagnostic techniques. Sydney: UNSW; 2005.
[12] Sawalhi N, Randall RB. Simulating gear and bearing interactions in the presence of faults: Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults. Mech Syst Sig Process 2008;22(8):1924–51 (Nov.).
[13] Massi F, Rocchi J, Culla A, Berthier Y. Coupling system dynamics and contact behaviour: modelling bearings subjected to environmental induced vibrations and 'false brinelling' degradation. Mech Syst Sig Process 2010;24(4):1068–80 (May).
[14] Purohit RK, Purohit K. Dyanmic analysis of ball bearings with effect of preload and number of balls. Int J Appl Mech Eng 2006;11(1):77–91.
[15] Cao M, Xiao J. A comprehensive dynamic model of double-row spherical roller bearing—Model development and case studies on surface defects, preloads, and radial clearance. Mech Syst Sig Process Feb. 2008;22:467–89.
[16] Nakhaeinejad M. Fault detection and model-based diagnostics in nonlinear dynamic systems. Austin: University of Texas; 2010.
[17] Jang GH, Jeong SW. Nonlinear excitation model of ball bearing waviness in a rigid rotor supported by two or more ball bearings considering five degrees of freedom. J Tribol 2002;124:82–90.
[18] McFadden PD, Smith JD. Model for the vibration produced by a single point defect in a rolling element bearing. J Sound Vib 1984;96(1):69–82.
[19] McFadden PD, Smith JD. The vibration produced by multiple point defects in a rolling element bearing. J Sound Vib 1985;98(2):263–73.
[20] Tandon N, Choudhury A. An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect. J Sound Vib 1997;205(3):275–92.
[21] Wang Y-F, Kootsookos PJ. Modeling of low shaft speed bearing faults for condition monitoring. Mech Syst Sig Process 1998;12(3):415–26 (May).
[22] Ghafari SH, Golnaraghi F, Ismail F. Effect of localized faults on chaotic vibration of rolling element bearings. Nonlinear Dyn 2007;53(4):287–301 (Dec.).
[23] Rafsanjani A, Abbasion S, Farshidianfar A, Moeenfard H. Nonlinear dynamic modeling of surface defects in rolling element bearing systems. J Sound Vib 2009;319:1150–74 (Jan.).
[24] Malhi AS. Finite element modelling of vibrations caused by a defect in the outer ring of a ball bearing. Amherst: University of Massachusetts; 2002.
[25] Liu J, Shao Y, Lim TC. Vibration analysis of ball bearings with a localized defect applying piecewise response function. Mech Mach Theory 2012;56:156–69 (Oct.).
[26] Ashtekar A, Sadeghi F, Stacke L-E. A new approach to modeling surface defects in bearing dynamics simulations. J Tribol 2008;130.
[27] Sassi S, Badri B, Thomas M. A numerical model to predict damaged bearing vibrations. J Vib Control 2007;13(11):1603–28 (Nov.).
[28] Patil MS, Mathew J, Rajendrakumar PK, Desai S. A theoretical model to predict the effect of localized defect on vibrations associated with ball bearing. Int J Mech Sci 2010;52:1193–201 (Sep.).
[29] Tadina M, Boltezar M. Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up. J Sound Vib 2011;300(17):4287–301.
[30] Li Y, Billington S, Zhang C, Kurfess T, Danyluk S, Liang S. Adaptive prognostics for rolling element bearing condition. Mech Syst Sig Process 1999;13(1):103–13 (Jan.).
[31] Li Y, Kurfess T, Liang Y. Stochastic prognostics for rolling element bearings. Mech Syst Sig Process 2000;14(5):747–62.
[32] Qiu J, Zhang C, Seth BB, Liang SY. Damage mechanics approach for bearing lifetime prognostics. Mech Syst Sig Process 2002;16:817–29 (Sep.).
[33] Marble S, Morton BP. Predicting the remaining life of propulsion system bearings. In: Proceedings of the 2006 IEEE aerospace conference; 2006.
[34] Begg CD, Merdes T, Byington C, Maynard K. Dynamics modeling for mechanical fault diagnostics and prognostics. In: Mechanical system modeling for failure diagnosis and prognosis, Maintenance and reliability conference, no. Marcon 99; 1999.
[35] Begg CD, Byington CS, Maynard KP, Dynamic simulation of mechanical fault transition. In Proceedings of the 54th meeting of the society for machinery failure prevention technology; 2000, p. 203–212.
[36] Chelidza D, Cusumano JP. A dynamical systems approach to failure prognosis. J Vib Acoust 2004;126.
[37] El-Thalji I, Jantunen E, Wear of rolling element bearings. In: Proceedings conference of Condition Monitoring and Diagnostic Engineering Management (COMADEM); 2013.
[38] Voskamp AP. Material response to rolling contact loading. J Tribol 1985;107:359–64.
[39] El-Thalji I, Jantunen E, On the integration of wear model into dynamic analysis for rolling element bearing. In: Proceedings of the second international workshop and congress on eMaintenance; 2012, p. 137-143.
[40] Arslan H, Aktürk N. An investigation of rolling element vibrations caused by local defects. J Tribol 2008;130.
[41] Xu G, Sadeghi F. Spall initiation and propagation due to debris denting. Wear 1996;201:106–16.
[42] Mota V, Ferreira LA. Influence of grease composition on rolling contact wear: experimental study. Tribol Int 2009;42:569–74 (Apr.).
[43] Morales-Espejel GE, Brizmer V. Micropitting modelling in rolling–sliding contacts: application to rolling bearings. Tribol Trans 2011;54(4):625–43 (Jul.).
[44] Dwyer-Jones RS. Predicting the abrasive wear of ball bearings by lubricant debris. Wear 1999;233–235:692–701.
[45] Maru MM, Castillo RS, Padovese LR. Study of solid contamination in ball bearings through vibration and wear analyses. Tribol Int 2007;40(3):433–40 (Mar.).
[46] El-Thalji I, Jantunen E. A descriptive model of wear evolution in rolling bearings. Eng Fail Anal 2014;45:204–24.
[47] Harris FJ. Rolling bearing analysis, 3rd ed. New York: John Wiley & Sons, Inc.; 1991.
[48] Faik S, Witteman H, Modeling of impact dynamics: a literature survey. In: International ADAMS user conference; 2000.

[49] F. Normani. Principle of impulse and momentum; 2012. [Online]. Available: ⟨http://www.real-world-physics-problems.com/impulse-and-momentum.html⟩.

[50] Sahoo P, Banerjee A. Asperity interaction in elastic–plastic contact of rough surfaces in presence of adhesion. J Phys D: Appl Phys 2005;38(16):2841–7 (Aug.).

[51] Kim H, Shaik NH, Xu X, Raman A, Strachan A. Multiscale contact mechanics model for RF–MEMS switches with quantified uncertainties. Modell Simul Mater Sci Eng 2013;21(8) (Dec.).

[52] Masen MA, de Rooij MB, Schipper DJ. Micro-contact based modelling of abrasive wear. Wear 2005;258:339–48 (Jan.).

[53] Kim TH, Olver AV. Stress history in rolling-sliding contact of rough surfaces. Tribol Int 1998;31(12):727–36.

[54] Karmakar S, Rao URK, Sethuramiah A. An approach towards fatigue wear modeling. Wear 1996;198:242–50 (Oct.).

[55] Jaing Y, Sehitoglu H. Rolling contact stress analysis with the application of a new plasticity model. Wear 1996;191:35–44.

[56] Franklin FJ, Widiyarta I, Kapoor A. Computer simulation of wear and rolling contact fatigue. Wear 2001;251:949–55 (Oct.).

[57] Holmberg K, Laukkanen A, Ronkainen H, Wallin K. Tribological analysis of fracture conditions in thin surface coatings by 3D FEM modelling and stress simulations. Tribol Int 2005;38(11–12):1035–49 (Nov.).

[58] Hannes D, Alfredsson B. Rolling contact fatigue crack path prediction by the asperity point load mechanism. Eng Fract Mech 2011;78:2848–69 (Dec.).

[59] Ringsberg JW, Bergkvist A. On propagation of short rolling contact fatigue cracks. Fatigue Fract Eng Mater Struct 2003;26:969–83.

[60] Tsushima N. Crack propagation of rolling contact fatigue in ball bearing steel due to tensile strain. NTN Tech Rev 2007;75:128–38.

[61] Liu Y, Liu L, Mahadevan S. Analysis of subsurface crack propagation under rolling contact loading in railroad wheels using FEM. Eng Fract Mech 2007;74:2659–74 (Nov.).

[62] Richard Liu C, Choi Y. A new methodology for predicting crack initiation life for rolling contact fatigue based on dislocation and crack propagation. Int J Mech Sci 2008;50:117–23 (Feb.).

[63] Donzella G, Petrogalli C. A failure assessment diagram for components subjected to rolling contact loading. Int J Fatigue 2010;32:256–68 (Feb.).

[64] Leonel ED, Venturini WS. Multiple random crack propagation using a boundary element formulation. Eng Fract Mech 2011;78:1077–90 (Apr.).

[65] Glodez S, Ren Z. Modelling of crack growth under cyclic contact loading. Theor Appl Fract Mech 2000;30:159–73.

[66] Bogdanski S, Brown MW. Modelling the three-dimensional behaviour of shallow rolling contact fatigue cracks in rails. Wear 2002;253:17–25.

[67] Bogdański S, Trajer M. A dimensionless multi-size finite element model of a rolling contact fatigue crack. Wear 2005;258:1265–72 (Mar.).

[68] Choi Y, Liu CR. Rolling contact fatigue life of finish hard machined surfaces Part 1. Model development. Wear 2006;261:485–91 (Sep.).

[69] Fajdiga G, Ren Z, Kramar J. Comparison of virtual crack extension and strain energy density methods applied to contact surface crack growth. Eng Fract Mech 2007;74:2721–34 (Nov.).

[70] Raje N, Sadeghi F, Rateick RG, Hoeprich MR. A numerical model for life scatter in rolling element bearings. J Tribol 2008;130.

[71] Canadinc D, Sehitoglu H, Verzal K. Analysis of surface crack growth under rolling contact fatigue. Int J Fatigue 2008;30:1678–89 (Sep.).

[72] Sadeghi F, Jalalahmadi B, Slack TS, Raje N, Arakere NK. A review of rolling contact fatigue. J Tribol 2009;131(4).

[73] Slack T, Sadeghi F. Explicit finite element modeling of subsurface initiated spalling in rolling contacts. Tribol Int 2010;43:1693–702 (Sep.).

[74] Slack T, Sadeghi F. Cohesive zone modeling of intergranular fatigue damage in rolling contacts. Tribol Int 2011;44:797–804 (Jul.).

[75] Warhadpande A, Sadeghi F, Kotzalas MN, Doll G. Effects of plasticity on subsurface initiated spalling in rolling contact fatigue. Int J Fatigue 2012;36:80–95 (Mar.).

[76] Weinzapfel N, Sadeghi F. Numerical modeling of sub-surface initiated spalling in rolling contacts. Tribol Int 2012 (Mar.).

[77] Santus C, Beghini M, Bartilotta I, Facchini M. Surface and subsurface rolling contact fatigue characteristic depths and proposal of stress indexes. Int J Fatigue 2012;45:71–81 (Dec.).

[78] Tarantino MG, Beretta S, Foletti S, Papadopoulos I. Experiments under pure shear and rolling contact fatigue conditions: competition between tensile and shear mode crack growth. Int J Fatigue 2013;46:67–80 (Jan.).

[79] Navarro A, Rios ER. Short and long fatigue crack growth—a unified model. Philos Mag 1988;57:15–36.

[80] Sun Z, Rios ER, Miller KJ. Modelling small fatigue cracks interacting with grain boundaries. Fatigue Fract Eng Meterl 1991;14:277–91.

[81] Jantunen E. How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods. Int J COMADEM 2006;9(3):24–38.

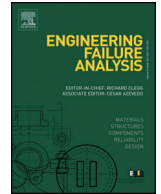# Fault analysis of the wear fault development in rolling bearings

Fault analysis of the wear fault development in rolling bearings

Idriss El-Thalji *, Erkki Jantunen

Industrial Systems, VTT Technical Research Centre of Finland, Finland

A R T I C L E   I N F O

A B S T R A C T

Signal processing methods are required to extract the features related to the wear process and how to track its evolution. Several signal processing methods are commonly applied in the experimental and real field tests. The generated signals of these tests are quite complex due to the dynamic nature of wear process, i.e., interaction among different wear mechanisms. Therefore, a dynamic model is required to explain the physical phenomena behind the detected signals. However, the current dynamic models in the literature lack to model the dynamic response under wear deterioration process over the whole lifetime, due to the complexity. Therefore, the purpose of this paper is to illustrate the evolution of the fault features with respect to the wear evolution process. It utilities a newly developed dynamic model and applies different commonly used signal processing methods to extract the diagnostic features of the whole wear evolution progress. The statistical time domain parameters and spectrum analysis are used in this study. Numerical results illustrate several issues related to wear evolution i.e., capabilities, weaknesses and indicators. The results show the extracted fault features and how they change with respect to the wear evolution process i.e., how the topological and tribological changes influence the extracted defect features. In this sense, the study helps to justify the experimental results in literature. The study provides a better understanding of the capability of different signal processing methods and highlights future enhancement.

© 2015 Published by Elsevier Ltd.

1. Introduction

The rolling element bearing (REB) is one of the most critical components that determine the machinery health and its remaining lifetime in modern production machinery. Robust condition monitoring (CM) tools are needed to guarantee the healthy state of REBs during the operation. CM tools indicate the upcoming failures which provide more time for maintenance planning. CM tools aim to monitor the deterioration i.e., wear evolution rather than just to detect the defects. Signal processing (SP) methods are required to extract the defect features. Over the years, several SP methods have been developed to extract the defect features from the raw signals of faulty REBs.

In summary, the current signal processing methods try to overcome several challenges [1]: (1) remove the speed fluctuation; (2) remove the noise effect; (3) remove the smearing effect of transfer path; (4) select optimal band of high Signal-to-Noise ratio; and (5) extract clear defect features. The order tracking methods are used to avoid the smearing of discrete frequency components due to speed fluctuations. To handle the smearing effect of transfer path, the minimum entropy de-convolution method has been developed. For background noise problem, different de-noising filters have been developed such as discrete/random separation, adaptive noise cancellation, self-adaptive noise cancellation or linear prediction methods to remove the noise background. Different feature extraction methods have been conducted to study specific monitoring techniques such as vibration, acoustic emission (AE),

---

* Corresponding author.
  E-mail addresses: idriss.el-thalji@vtt.fi (I. El-Thalji), erkki.jantunen@vtt.fi (E. Jantunen).

oil-debris, ultrasound, electrostatic, shock-pulse measurements (SPM) and their use in faulty REB detection. Many studies have used simple signal/data processing techniques such as root mean square (RMS), kurtosis and FFT. However, the largest share of studies has focused to develop new SP techniques: envelope, wavelets, data-driven methods, expert systems, fuzzy logic techniques, etc.

However, the development of physical defect i.e., wear is an evolution process taking place over a specific time interval within the lifetime. Therefore, the sixth challenge is to track the defect features over the whole lifetime in order to diagnose the bearing health. The tracking methods utilise the previous feature extraction methods to track the extracted feature(s) over the whole lifetime. Therefore, the most important issue of tracking technique is the reliability of the signal analysis method and how effective indication it can provide. The detection of defect propagation has been studied in [2–5]. First, it was observed that increasing the defect length increases the burst duration. Second, it was observed also that increasing the defect width increases the ratio of burst amplitude. In fact, the signal analysis methods are verified by introducing a virtual defect into the dynamic bearing model and validated by artificially introduced defect in test rigs. Then, it is assumed that increasing the defect size gradually is correlated with the defect severity (i.e., wear evolution). For example, Nakhaeinejad [6] utilised the bond graphs to study the effects of defects on bearing vibration. A localised fault has been introduced into the dynamic model with different defect size to represent the development of defect severity. It is clear that this kind of approach assumes a linear relationship between the defect size and the obtained impact response.

However, the experimental studies show some significant degree of nonlinearity that appears in the measured impact. The tests which have used RMS as tracking measure [7–10] show nonlinear propagation curve of the wear evolution. Moreover, the spectral kurtosis was used to track the severity. The results show also a stable state of spectral kurtosis when the defected surface is getting very rough or very smooth [11]. In the industry, the amplitude peaks at the bearing fault frequencies in the spectrum is the most commonly used indicator. Therefore, the amplitude peaks are extracted and tracked over the time. Several studies e.g., [12,13] have represented the evolution of the amplitude peaks in the spectrum.

In fact, these spectrum charts represent also the main characteristic frequencies related to the high frequency zone, the bearing natural frequency, and the bearing fault frequency zone, as shown in Fig. 1. The spectrum correlates the amplitude change in these peaks at certain frequencies to the wear evolution progress. However, the topological and tribological features e.g., defect shape, size and the uniformity of the lubrication film might change over the lifetime. Therefore, when the rolling element is passing over the defected area, the impulsive nature of the contact is changing as well. Moreover, the topological and tribological changes might disturb the time between the impacts (between rolling element and the defected area) which introduces some degree of slippage. The slippage might delay the rotational time of the rolling element and therefore disturb the periodic phenomenon of rolling element contacts/impacts. Therefore, the frequency peaks which are related to those impact events might not be as clear (i.e., amplitudes) as if the impacts would occur in a perfectly periodic manner.

Therefore, in order to model the wear evolution, a defect development model should be introduced in a way that can represent the surface topology changes over the lifetime. Recently, El-Thalji & Jantunen [14,15] presented an updated framework of wear evolution in REBs. The updated framework of wear evolution is thoroughly illustrated in [15] and schematically shown in Fig. 2, where five-stages are presented: running-in, steady-state, defect initiation (indentation, micro-cracking, inclusions), defect propagation (pits, propagated cracks), and damage growth (spalling).

The wear evolution model can be used to give insights and knowledge for further enhancements of current monitoring practices for REBs. Therefore, the purpose of this paper is to illustrate the evolution of the fault features with respect to the wear evolution process. It utilities a newly developed dynamic model and applies different commonly used signal processing methods to extract the diagnostic features of the whole wear evolution progress. The dynamic model of a single-degree-of-freedom system is presented to study the dynamic behaviour and properties under the effects of wear evolution process for the whole lifetime. The paper begins with presenting a description of the dynamic model and the main considerations that illustrate the wear evolution. The results part presents the extracted defect features of each wear progression stages and critical events.
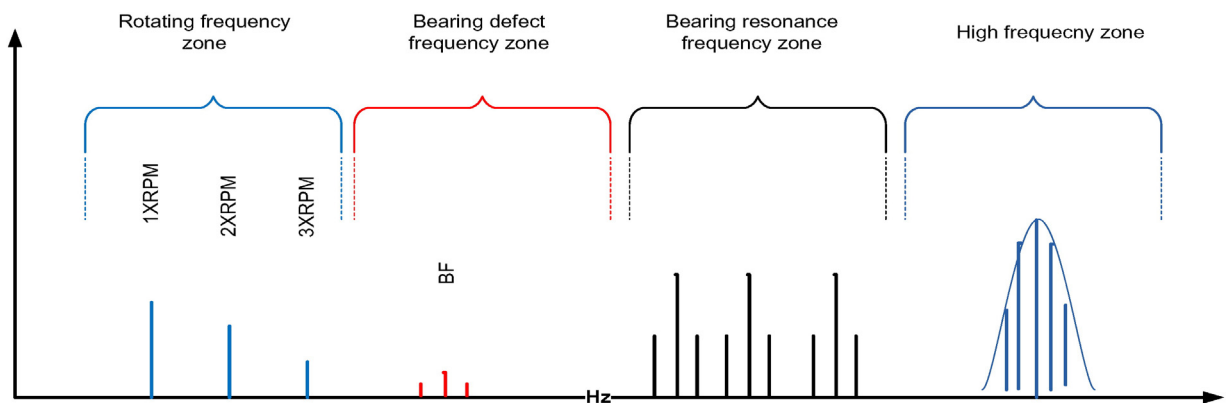

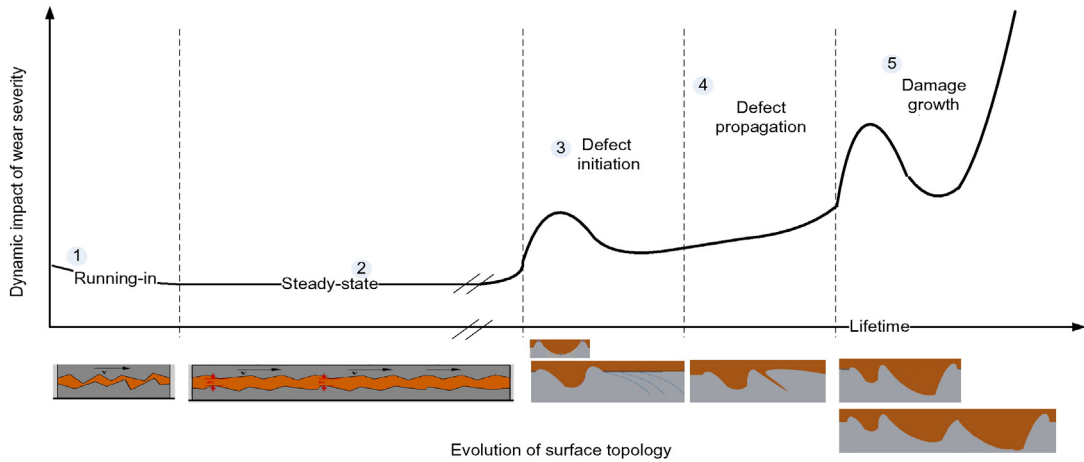**Fig. 1.** Schematic spectrum of vibration signal that contain bearing fault.

Fig. 2. Dynamic impact of the wear severity during the lifetime of a bearing [15].

2. The developed model

The developed model is an integrated model of fault dynamics and wear mechanics. Such integration has introduced a couple of different terminologies due to different perspectives. For example, the researchers of fault dynamics modelling are interested in the effect of the fault existence and its severity level (e.g., fault, damage, failure). However, the researchers of contact/wear mechanics are interested in the fault type (e.g., fatigue, wear) and the involved mechanisms (fatigue wear, adhesive wear, abrasive wear, corrosive wear). Thus, the definition of fault is not sufficient to describe the fault type and its mechanism from the perspective of contact/wear mechanics studies. Therefore, the common terminologies of contact/wear mechanics are used in this paper in the parts where the mechanics of the fault are described. However, the common terminologies of fault modelling are used in the parts where the dynamic effects of the fault are described.

It is worth to mention that the purpose of the paper is to apply tracking analysis of the signal that is generated by the developed dynamic model that is described in [16]. Moreover, the developed dynamic model is based on a descriptive model of wear evolution that is described in [15]. Therefore, the main focus of this paper is on the signal analysis. However, in order to illustrate the wear evolution effect on the physical characteristics of rolling bearings, a parametric dynamic model is described. This model contains the scale parameters to illustrate their effect on e.g., natural frequencies, displacements, velocities, and forces due to defect. In fact, the bearing dimensions influence natural frequencies, displacements, velocities, and forces due to surface disturbances e.g., waviness, and defect.
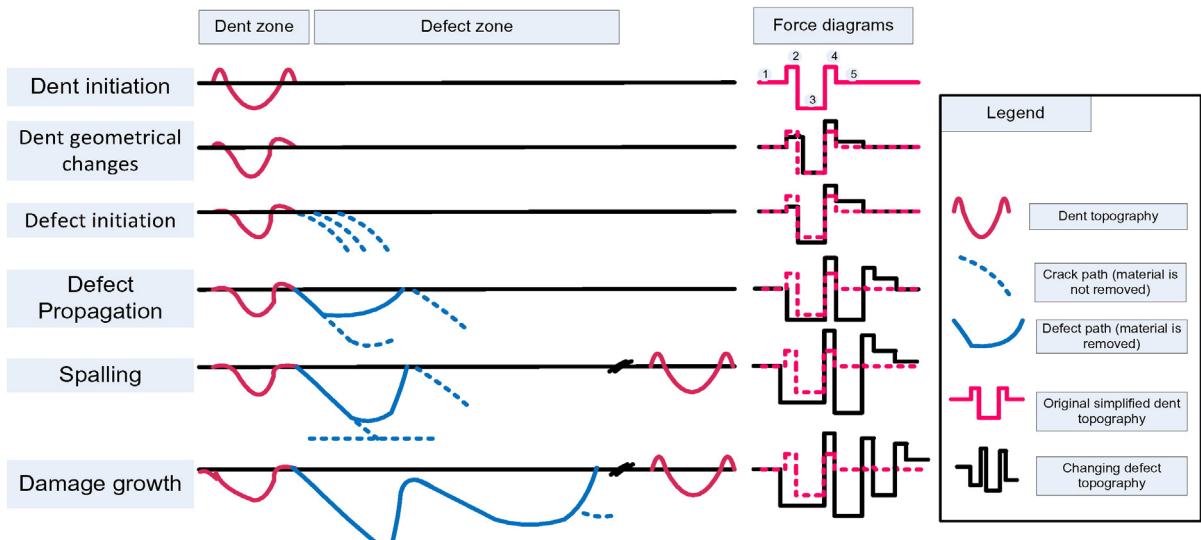

Fig. 3. Schematic representation of the modelled system.

Therefore, the paper utilises a model of a single-degree of freedom (as shown in Fig. 3) to illustrate how the bearing dimensions influence the system response under faulty conditions. The model can be represented as the following:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F_{ex}\} \tag{1}$$

where [M], [C], [K] and $\{F_{ex}\}$ are respectively matrices of system mass(es), damping coefficient(s), stiffness(es) and external forces. M is 0.052 kg, K is 5.96×10^7 N/m (For outer race), and C is 5×10^7 N·s/m which are related to the modelled deep groove ball bearing (SKF 61810-4).

The model can illustrate the effect of bearing size on the following issues:

1. natural frequency
2. magnitude and frequency of force due to waviness
3. magnitude and frequency of force due to defect
4. response in displacement, velocity and acceleration measurements.

In fact, these four issues are significantly important to understand the scale effect on the measured system response. The scale effect on the response magnitudes due to waviness and defect might explain the monitoring difficulties. For simplicity, in some part of the analysis, the model considers the bearing outer race.

2.1. Natural frequency of the raceway

The contact events with the help of small degree of loose fit between the raceway and the housing might excite the natural frequency of the raceway. Sassi et al. [17] showed that natural frequency ω_n of bearings for flexural vibration mode 'n' can be given as:

$$\omega_n = \frac{n\left[n^2 - 1\right]}{\sqrt{1 + n^2}} \sqrt{\frac{EI}{\mu R^4}} \quad \text{in rad/s} \tag{2}$$

where E is the modulus of elasticity, I is the moment of inertia of the race cross section, μ is the mass per unit length, R is the radius of the ring and n is the order of flexural vibration. In fact, n is the number of deformation waves in each mode (i + 1). Sassi et al. [17] highlighted that n_0 and n_1 correspond to rigid modes, therefore the flexural vibration modes start from n = 2.

However, the value of the sectional secondary moment is needed before using Eq. (2). Due to the difficulties in obtain the exact value for a bearing ring with a complicated cross-sectional shape. Eq. (3) is suggested by NSK as best used when the radial natural frequency is known approximately for the outer ring of a radial ball bearing [18].

$$f_n = 9.41 \times 10^5 \times \frac{K(D - d)}{\{D - K(D - d)\}^2} \times \frac{n\left[n^2 - 1\right]}{\sqrt{n^2 - 1}} \quad \text{in Hz} \tag{3}$$

where d is the bearing bore in mm, D is the bearing outside diameter in mm, K is cross sectional constant (K = .15 for outer ring of an open type) [18]. Therefore, the approximate natural frequency of the raceway depends highly on the effective diameter of the raceway. For simplicity, the natural frequency of bearing raceway can be estimated based on ring vibration theory by the following formula [19]:

$$\omega_n = \frac{c}{\pi D} \tag{4}$$

where, c is the speed of sound in the material.

2.2. The potential bearing forces

The distributed fault e.g., waviness and localised fault e.g., defect are expected to influence the system dynamic response. These faults can be represented in the system equation of motion as:

$$[M]\{\ddot{y}\} + [C]\{\dot{y}\} + [K]\{y\} = \{F_{imb} + F_s + F_d\} \tag{5}$$

where, F_{imb} is the imbalance force, F_s is the force due to surface roughness and waviness and F_d is the force due to surface defect. In the following sub-sections, the scale effect on the force magnitudes and frequencies of F_s and F_d will be discussed.

2.2.1. Force due to imbalance

If the structure holding the bearings in such a system is infinitely rigid, the centre of rotation is constrained from moving, and the centripetal force resulting from the imbalance mass can be found by the following formula:

$$F_{imb} = I_m * r * \omega^2 * \sin(2\pi\omega t) \tag{6}$$

where F_{imb} is the imbalance force, I_m is the mass, r is distance from the pivot, and ω is the angular frequency. For this model, the imbalance mass considered to be 0.005 kg, r is 0.1 m, and the angular frequency is 160 rad/s.

2.2.2. Force due to surface imperfections

The surface peaks and valleys influence the rolling contact forces. Therefore, the generated force due to surface texture is related to the frequency of surface waves i.e., waviness and roughness. Harsha et al. [20] presented a formula to calculate the restoring force due to waviness, which can be represented as the relation between local Hertzian contact force and deflection, as follows:

$$F_s = k(r_{\theta i})^p \tag{7}$$

where k is the material deflection factor and p is related to contact type (i.e., it is 3/2 for ball bearing and 10/9 for rolling bearings). The parameter $r_{\theta i}$ is the radial deflection at the azimuth angle θ_i, which is given as:

$$r_{\theta i} = (x \cos\theta_i + y \sin\theta_i) - \{\gamma + \Pi_i\} \tag{8}$$

where, x is the horizontal direction, y is the vertical direction, γ is the internal radial clearance, and Π_i is the amplitude of the waves at the contact angle and is given as:

$$\Pi_i = \Pi_0 + \Pi_p \sin(N\theta_i) \tag{9}$$

where Π_p is the maximum amplitude of waves, Π_0 is the initial wave amplitude and N is the number of wave lobes. The contact angle is a function of the number of rolling elements and cage speed, which is given as:

$$\theta_i = \frac{2\pi}{N_b}(i-1) + \omega_{cage}t \tag{10}$$

where $i = 1, \dots, N_b$. The number of waves for the entire surface length L is given as:

$$N = \frac{L}{\lambda} \tag{11}$$

where $L = \pi D$, and λ is the wavelength of roughness or waviness. The total restoring force is the sum of restoring forces from all of the rolling elements. Therefore, the total restoring force (considering waviness and clearance) is given as:

$$F_s = \sum_{i=1}^{N_b} k[(x \cos\theta_i + y \sin\theta_i) - \{\gamma + \Pi_i\}]^{\frac{3}{2}}. \tag{12}$$

When the magnitude of contact angle is substituted in the total restoring force formula, the effects of the cage speed ω_{cage} and bearing diameter D become clear.

In this paper, the waviness order is around 80. The number of waves N is defined [20] as follows:

$$N = \frac{2 * \pi * r}{\lambda} \tag{13}$$

where r is the radius of outer or inner race, and λ is the wavelength. The common wave length from bearing manufacture is measured to be about 0.8 mm. Therefore, the waviness order of an outer race of e.g., 10 mm in diameter and wavelength of 0.8 mm is about 78.5. With the same logic, the roughness waves can be estimated, where the roughness wavelength is 0.02 mm.

2.2.3. Force due to bearing defect

The bearing fault frequency is basically based on the impact events which are generated when the rolling elements pass over a dent or defect. The fundamental train frequency FTF is given [21] as:

$$FTF = \frac{F}{2} \times \left(1 - \frac{B_d}{P} \cos\varnothing_i\right) \tag{14}$$

where, F is the shaft frequency, B_d is the rolling element diameter, P is the pitch diameter, and $\varnothing_i$ is the contact angle. The ball pass frequency for outer race fault BPFO is given [21] as:

$$BPFO = \frac{N}{2} \times F \times \left(1 - \frac{B_d}{P} \cos\varnothing_i\right) = N \times FTF \tag{15}$$

where, N is the number of rolling elements. The ball pass frequency for inner race fault BPFI is given [21] as:

$$BPFI = \frac{N}{2} \times F \times \left(1 + \frac{B_d}{P} \cos\varnothing_i\right) = N \times (F - FTF). \tag{16}$$

The ball pass frequency BSF is given [21] as:

$$BSF = \frac{P}{2B_d} \times F \times \left(1 - \left(\frac{B_d}{P} \cos\varnothing_i\right)^2\right). \tag{17}$$

However, the force amplitude due to a surface defect can be represented as contact deflection or impact force. Several dynamic models simulate the force due to the defect as an expression of deflection:

$$F_d = k\left(\delta_{o,i,b}\right)^p \tag{18}$$

where k is the material deflection factor and p is related to the contact type, it is 3 for ball bearing and 10/3 for roller bearing. However, the radial deflection $\delta_{o,i,b}$ depends on the defect location. Therefore δ_o, δ_i, and δ_b represent the deflection due to outer race defect, inner race defect and rolling eminent defect, respectively. In this paper, the focus is to consider only the outer race defect, therefore, it is given [22] as:

$$\delta_0 = \rho_j + \rho_b - (R_o + D_d) \tag{19}$$

where, ρ_j is the radial position of the rolling element, ρ_b is the radius of the rolling element, R_o is the radius of the non-deformed outer race and D_d is the depth of the defect at the contact position. Epps [23] shows that the radial deflection depends on contact mode at the entry and exit points of the defect. However, Sassi et al. [17] described the total impacting force F_d as a sum of the static and dynamic components as follows:

$$F_d = F_{static} + F_{dynamic}. \tag{20}$$

The static component of the impact force depends on the maximum load Q_{max}, separation angle between rolling elements Ψ_i and load distribution factor ε:

$$F_{static} = Q_{max}\left(1 - \frac{1 - \cos(\Psi_i)}{2\varepsilon}\right)^{1.5}. \tag{21}$$

The static component of the impact force is a function of the static component of the impacting force, shock velocity ΔV and impacting material factor K_{impact}:

$$F_{dynamic} = F_{static} K_{impact} \Delta V^2 \tag{22}$$

where ΔV^2 is given as:

$$\Delta V^2 = \frac{10}{28} g\left(\frac{d_{def}}{B_d}\right)^2 \tag{23}$$

where d_{def} is the defect length (in the rolling direction), and g is the gravitational constant and equal to 9.81 m/s².
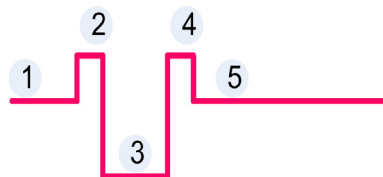

Fig. 4. Wear defect and surface topology evolution [16].

2.2.4. Force due to fault development

The forces due to machine fault (F_m) and due to surface roughness and waviness F_s can be simply introduced to the dynamic model. However, the force due to bearing defect F_d is changing over the time due to the topographical and tribological changes. The evolution of the wear defect stages is described in Fig. 4.

In Fig. 4, the main defect topologies are illustrated with their force diagrams (to the right side). The solid lines in Fig. 4 show either dent or defect, while the dotted lines illustrate the potential cracking trajectories. To the right of Fig. 4, the force diagrams of each defect topology are schematically illustrated and used to generate impulsive force series of the topology features. Basically, the changes are related to the leading and trailing edge of the dent/defect. Therefore, a load function of five steps was defined, to represent the transient impacts i.e., due to defect into the dynamic model, as shown in Fig. 5.

These load function varies based on the variation of the surface topology as illustrated in Fig. 2. Each stages of wear evolution progress were introduced into the dynamic model by its corresponding force diagram (as shown in Fig. 2) and with dimensions illustrated in Table 1. For each stage of wear evolution progress, different parameters of the load function is given in order to simulate the surface topology variations due to wear progress.

There are three issues that influence the non-linear phenomena of wear evolution in rolling bearing. First, the bearing defect phenomenon is quite complex to be modelled due to the nonlinear evolution of wear defect. For example, the over-rolling and mild abrasive wear smoothen the trailing edge of the new defect. Therefore, the measured impact severity might be reduced as the asperity height and impact area is reduced. Second, the width of the defect is the key parameter in the impact severity. The wider the defect is the larger the impact area at the trailing edge. Wider impact area means higher peak amplitudes at the bearing defect frequency zone [26]. Third, the time between impact events is known as the epicyclical frequency. However, the topological change due to the wear evolution most probably might change the drag and driving tangential forces which make the cage and rolling element travel more slowly than its epicyclical value, which is one of the main causes of slippage phenomenon. The simulation model introduces some degree of random slippage. However, the amount is related to the wear evolution stages. It is basically related to topology features at each stage which influence the tribological features i.e., drag and driving tangential forces.

3. Results

A deep groove ball bearing (SKF 61810-4) was used for the modelling purposes. The bearing inner diameter is 50 mm, the outer diameter is 65 mm and the weight is 0.052 kg. Because of various assumption made in developing this model, this model was continuously verified with the model in [17], both in their numerical and experimental figures. However, the dimension and feature differences between the modelled bearing type in this model (SKF 661819-4) and the type (SKF 6206) which considered in [17] should be noticed.

The time history signal is sampled at specific events of wear evolution process in order to apply several signal processing techniques. The aim is to illustrate the defect features and their changes over the time. There are several events which represent the wear evolution in rolling bearings. Eight different time intervals were sampled to study eight events of wear evolution: (1) normal operation; (2) mild surface roughness and waviness; (3) excitation of bearing natural frequency; (4) indentation stage; (5) dent smoothening; (6) slippage case; (7) defect completion; (8) defect smoothening; (9) damage growth. The results of each case event are briefly illustrated as follows:
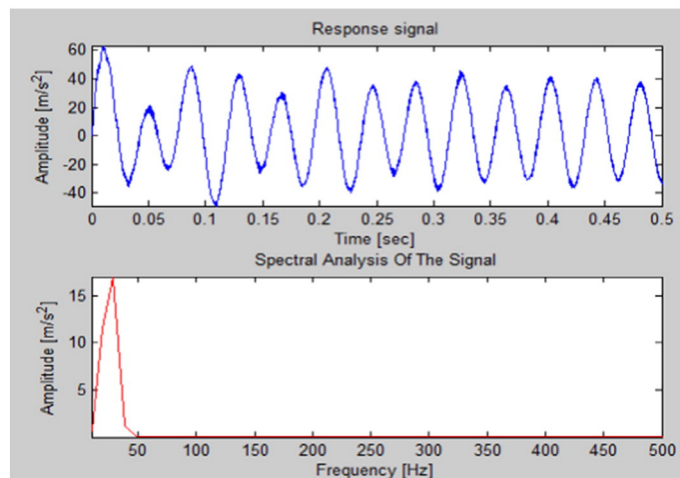

Fig. 5. The simulated force diagram of wear defect.

Table 1
Specific events of wear evolution process.

#	Wear evolution cases	Asperity height of the defect (in μm)	Defect diameter/size (in μm)	Reference
1	Real dent	0.4–4	50	[24]
2	Smoothened dent	0	50	[24]
2	Mild defect	20	300	[24]
3	Medium defect	20	1000	[25]
4	Large defect	30	3000	[4,6]
5	Smoothened large defect	4	3000	[4]
6	Damage growth, multi-defects, two large defects	30	3000	[4]

3.1. Machine fault case

Under normal operation, the REBs are expected to work in a steady state manner, especially after the running-in stage. However, due to machine faults such as imbalance, bent shaft, misalignment, a number of distortions might be detected. For example, the imbalance fault can be detected once the spectrum starts showing peak amplitudes at 1xrpm, as shown in Fig. 6.

3.2. Dented surface case

When a dent is localised, it might be expected to see peak amplitudes at the bearing defect frequency zone in the spectrum, as shown in Fig. 7. However, it is very weak peaks. The amplitude peaks are observed at bearing natural frequency zone (around 1100 Hz) due the impact phenomenon when the rolling element passes over the dent. Epps [23] explained the appearance of amplitude peak at bearing natural frequency (around 1100 Hz) and the amplitude peak at bearing fault frequency zone (around 170 Hz). It was observed that the defect signal is of two parts: first part originates from the entry of the rolling element into the defect, which generates an amplitude peak at bearing fault frequency zone. Second part originates due to the impact event between the rolling element and the trailing edge of the defect, which generates an amplitude peak at high frequency zone. However, the first part of the impact signal is depends highly on the defect size, which is very small at indentation stage.

Over the time, the over-rolling and mild abrasive wear make the dent impacts to become smaller. The reduction of peak amplitudes at bearing natural frequency zone due to smoothening actions is noticeable, as shown in Fig. 8.

3.3. Defected surface case

Even though, the smoothening process will take place and reduce the impact forces, the stresses at the trailing edge of the dent will still be enough to initiate a crack. The crack will propagate and end eventually as a defect. The impact event at the trailing edge of the defect is much higher compare to the dent. Therefore, it is expected to see higher peak amplitudes at the bearing defect frequency zone in the spectrum, as shown in Figs. 9, 10, and 11. In fact, the impact event when the rolling element passes over the trailing edge does not only generate impulsive impact but also distort the rolling element motion i.e., signal flatness. This distortion phenomenon is responsible for producing peak amplitudes at the harmonics of the bearing defect frequencies.
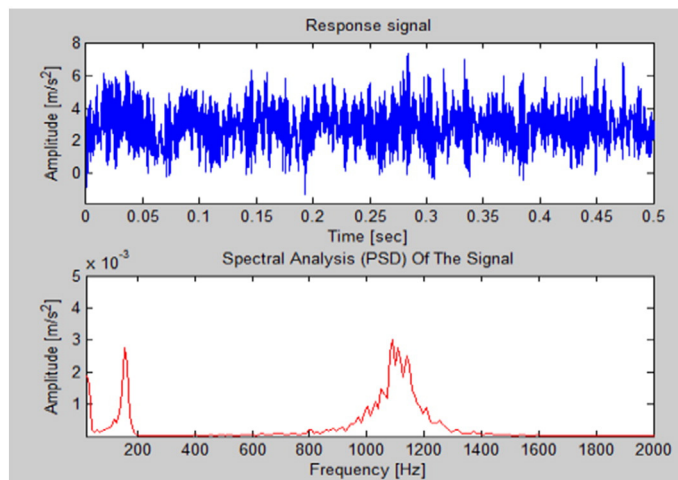
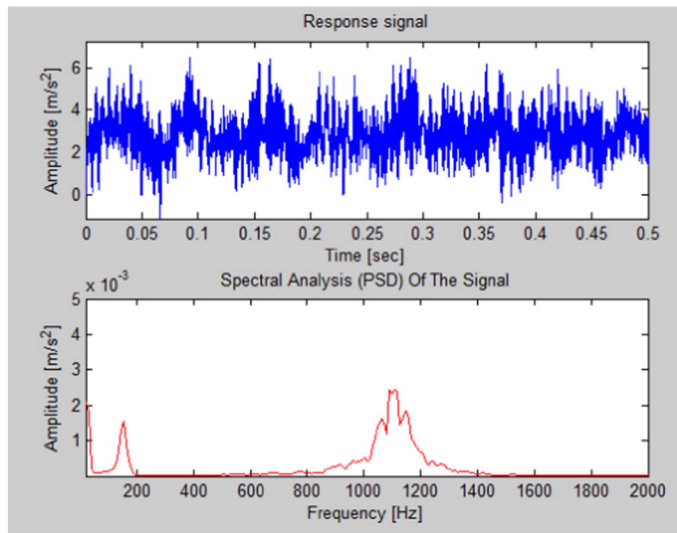**Fig. 6.** Time response and spectrum of the bearing under misalignment fault.

Fig. 7. Time response and spectrum of the indentation event.

3.4. Smoothened defect case

The defect impacts become also smaller due to the over-rolling and mild abrasive wear. The reduction of peak amplitudes at bearing natural frequency zone can be clearly noticeable due to smoothening actions, as shown in Fig. 12.

3.5. Case of damage growth

One important physical feature in the damage growth is the defect width. The wider the defect is the larger the impact area at the trailing edge. Wider impact area means higher peak amplitudes at the bearing defect frequency zone. Moreover, the length and the depth of the defect are expected to propagate over the time until a failure will occur. In Fig. 13, the wear fault features are illustrated. It is worth to mention, that Fig. 13 represents early-stage of damage growth. As the damage grows, the spectrum becomes more complicated to visualise the features, due to the complex distortions associated with damage growth process.

4. Discussion

The frequency domain methods such as FFT have been introduced to detect the fault induced signal. It is also used to track the wear evolution as presented in [12]. However, this presentation of wear evolution features can be updated based on the simulated data of this paper. The idea is to enhance our understanding of the non-linear phenomena of wear evolution in rolling bearings. The wear
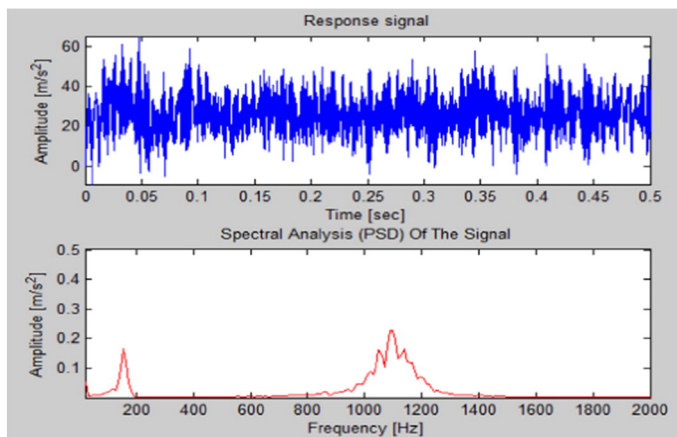

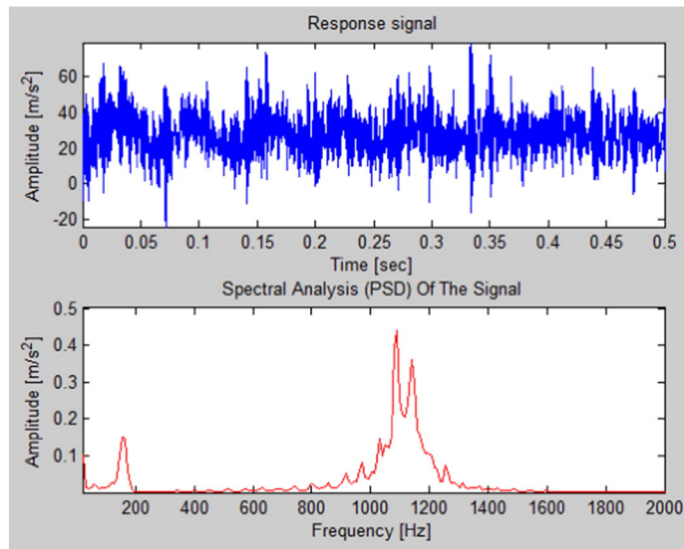**Fig. 8.** Time response and spectrum of the smoothened dent.

Fig. 9. Time response and spectrum of the bearing with defect (0.3 mm).

evolution stages are illustrated in the following descriptions, with physical justification related to the topological and tribological evolution of the defected bearing.

4.1. High frequency zone

The experimental findings that are discussed in [14] have shown that surface roughness is the main reason behind the amplitude peaks at the high frequency zone of the spectrum. The surface roughness and waviness generate the surface peaks and valleys which increase the probability of race surface peaks contact with rolling element surface. These contact events present amplitude peaks at the high frequency zone. Over the time, the contact events, between the surface peaks of the race and rolling element surface, smoothen the surface. Moreover, the film thickness will be stabilised and become uniform. However, the bearing geometrical and tribological characteristics might change due to the loading and operating conditions. For example, there is all the time some degree of errors in the contact profile between the race and the rolling element which disturb the pure rolling contact [27]. Another example is the lubrication transfer into surface valleys [28]. Such contact disturbances produce high stresses when the rolling elements pass over the surface asperities. These stresses will appear as amplitude peaks at the high frequency zone in the spectrum. These contact events
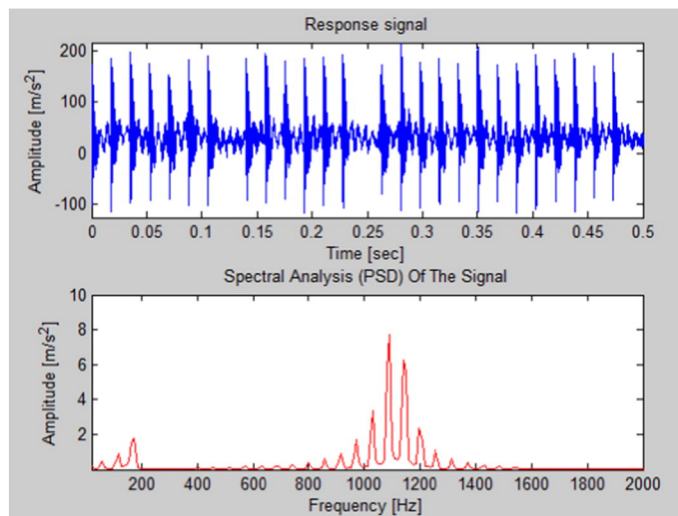

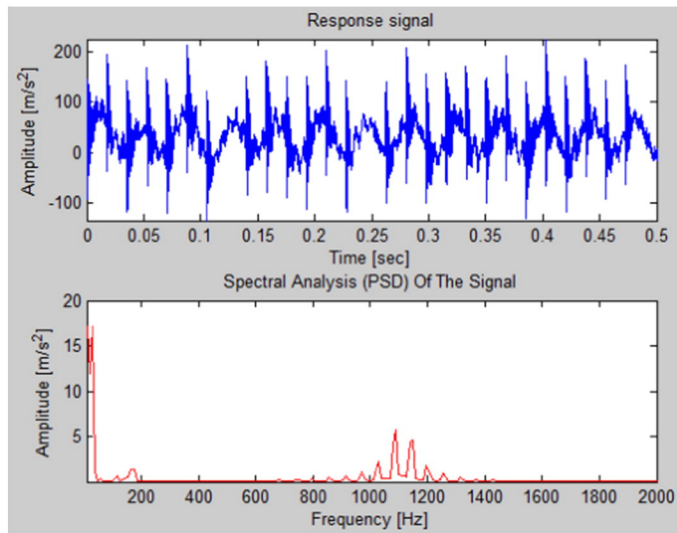**Fig. 10.** Time response and spectrum of the bearing with defect (1 mm).

Fig. 11. Time response and spectrum of the bearing with defect (3 mm).

might involve minor abrasive wear which make the surface smoother and higher stresses will be generated from those events. Later, the distributed wear debris will be the main issue in generating high stresses. That means higher amplitude peaks at the high frequency zone are expected in later stage.

4.2. Bearing natural and bearing defect frequency zones

The impact events when the rolling element passes over the defected area (i.e., with the help of small degree of loose fit between the raceway and the housing might) excite the natural frequency of the raceway. Epps [23] observed that the defect signal is of two parts: first part originates from the entry of the rolling element into the defect, which generates an amplitude peak at bearing fault frequency zone. Second part originates due to the impact event between the rolling element and the trailing edge of the defect, which generates an amplitude peak at bearing natural frequency zone. However, the defect signal might change as the wear (defected area) is evolved over time.

The abrasive wear generates some internal debris which might be transferred with the oil lubrication into the valleys of either the surface waviness or the contact deflection. At the moment when the rolling element passes over the valleys that contain debris, the rolling element might press the debris into the surface and generate a localised dent. This localised dent generates impact events that might excite the bearing natural frequency. Therefore, the peak amplitudes at the bearing natural frequency might be seen. In
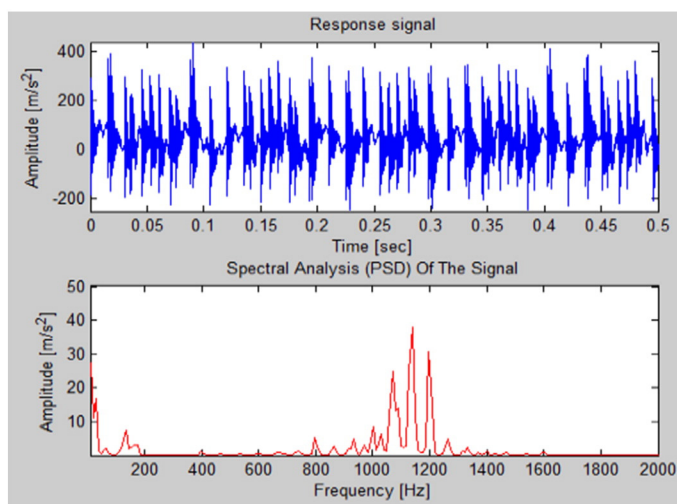

Fig. 12. Time response and spectrum of the bearing with smoothened defect.
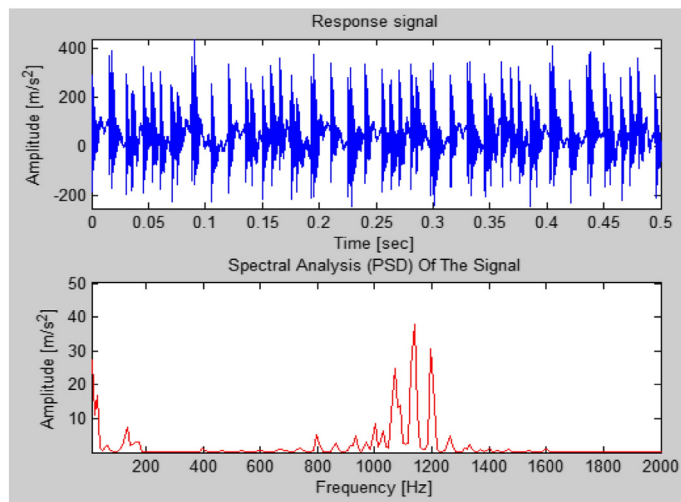
Fig. 13. Time response and spectrum of the bearing under damage growth state.

fact, the dent acts as stress riser in particular at the trailing edge of the dent [29]. However, the impact event which is generated when the rolling element passes over the new dent is very small. Moreover, it becomes even smaller due to the over-rolling and mild abrasive wear of the asperity at the trailing edge of the new dent. However, the high stresses at the trailing edge are still enough to initiate a crack. The crack will propagate and end eventually as a defect. The defect will have leading and trailing edges as well.

However, the impact event at the trailing edge of the defect is much higher compared to the dent one. Therefore, it is expected to see higher peak amplitudes at the bearing natural frequency in the spectrum. In fact, the impact event (i.e., generated when the rolling element passes over the trailing edges) generates impulsive impact and distorts the rolling element motion. The distortion phenomenon is responsible for producing peak amplitudes at the harmonics of the bearing defect frequencies. It depends on how much the impact area (the trailing edge area in contact) is large enough to distort the impact signals.

The over-rolling and mild abrasive wear will act again to smoothen the trailing edge of the new defect. Therefore, a clear reduction in peak amplitudes at both the bearing defect and bearing natural frequency zones are expected. However, the high stresses at the trailing edge are enough to initiate a crack of next defect. The width of the defect is key parameter in the impact severity. The wider the defect is the larger the impact area at the trailing edge. In fact, the impact area of a dent is smaller than the impact area of the defect. Therefore, the crack trajectories of the defect will be further from each other when compared to the crack trajectories of the dent. The crack trajectories are the main issue that determines the width of the new defect. Wider impact area means higher peak amplitudes at both bearing defect and bearing natural frequency zones.

A significantly important issue of this simulation model is the consideration of the slippage phenomenon and its effect on frequency domain features. The time between impact events is known as the epicyclical frequency that we try to extract from the spectrum. For example, the topological change due to the wear evolution might most probably change the drag and driving tangential forces which make the cage and rolling element travel more slowly than of its epicyclical value. In fact, as the wear become more severe, the topological and tribological features of the surface generate and influence stronger the drag and driving tangential forces, which means more slip and disturbances. Therefore, the amplitude peaks at the defect frequency might not be clear and several harmonics and sideband peaks will appear.

4.3. Rotating frequency zone

The machine faults e.g., imbalance, bent shaft, misalignment, looseness have specific characteristic features at the rotating speeds and their orders. In fact, those machine faults have in-direct influence on the wear initiation and evolution as they introduce some change in the contact characteristics and load distribution, for example, the misalignment introduces higher contact stresses at specific part of the raceway, which is one of the main reasons of wear initiation.

5. Conclusion

The industrial applications utilise the in-field measurements in order to plan the required maintenance actions in a cost effective manner. The available signal processing techniques have been developed to deal with speed fluctuation effect, signal noise, smearing effect and the challenge of the defect feature extraction in order to extract clear diagnostic indicator e.g., RMS. It is expected that the selected diagnostic indicator can be used to track the wear evolution progress over the lifetime of the bearing. The study describes a dynamic model of wear evolution that is capable to mimic the wear process in rolling bearings. In fact, it is capable to mimic the topological surface evolution due to wear in rolling bearings over the lifetime. However, the study is based on multiple models that

estimate the response, transition conditions and stress accumulations with certain degree of uncertainties. Thus, the potential diagnostic indicators of wear defect have been tracked with the help of simple signal processing techniques. The dynamic nature of wear process, i.e., interaction among different wear mechanisms, makes some indicators hard to be extracted at specific interval.

The statistical time domain parameters and amplitude at bearing defect frequency are not sufficient indicators for the whole wear evolution process. The other frequency zones include defect features that are required to be extracted and tracked. In general, one indicator i.e., change in amplitude, is not even enough to detect the fault and therefore it is hard to track the evolution. The results represent and illustrate the fluctuation in the clarity of the indicators over the whole lifetime i.e., some indicators are effective at specific time interval. Thus, the tracking process based on specific indicators might be not sufficient to diagnose the wear evolution and the bearing health. In this sense, the study show how different indicators were influenced by the wear evolution process i.e., surface topology changes due to wear. Thus, it provides the potential changes in the tracked indicator and highlights the need to combine different tracked indicators in order to enhance the tracking effectiveness.

References

[1] N. Sawalhi, Diagnostics, Prognostics and Fault Simulation for Rolling Element Bearings, UNSW, Sydney, 2007.
[2] T. Yoshioka, T. Fujiwara, Measurement of propagation initiation and propagation time of rolling contact fatigue cracks by observation of acoustic emission and vibration, Tribol. Ser. 12 (1987) 29–33.
[3] H. Kakishima, T. Nagatomo, H. Ikeda, T. Yoshioka, A. Korenaga, Measurement of acoustic emission and vibration of rolling bearings with an artificial defect, QR RTRI 41 (3) (2000) 127–130.
[4] A.M. Al-Ghamd, D. Mba, A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size, Mech. Syst. Signal Process. 20 (Oct. 2006) 1537–1571.
[5] Y.-H. Kim, A.C.C. Tan, J. Mathew, B.-S. Yang, Condition monitoring of low speed bearings: a comparative study of the ultrasound technique versus vibration measurements, World Congress of Engineering Asset Management, 2006.
[6] M. Nakhaeinejad, Fault Detection and Model-Based Diagnostics in Nonlinear Dynamic Systems, University of Texas, Austin, 2010.
[7] J. Sun, R.J.K. Wood, L. Wang, I. Care, H.E.G. Powrie, Wear monitoring of bearing steel using electrostatic and acoustic emission techniques, Wear 259 (2005) 1482–1489.
[8] E. Jantunen, How to diagnose the wear of rolling element bearings based on indirect condition monitoring methods, Int. J. COMADEM 9 (3) (2006) 24–38.
[9] T.J. Harvey, R.J.K. Wood, H.E.G. Powrie, Electrostatic wear monitoring of rolling element bearings, Wear 263 (Sep. 2007) 1492–1501.
[10] Z. Zhi-qiang, L. Guo-lu, W. Hai-dou, X. Bin-shi, P. Zhong-yu, Z. Li-na, Investigation of rolling contact fatigue damage process of the coating by acoustics emission and vibration signals, Tribol. Int. 47 (Mar. 2012) 25–31.
[11] N. Sawalhi, R.B. Randall, Vibration response of spalled rolling element bearings: observations, simulations and signal processing techniques to track the spall size, Mech. Syst. Signal Process. 25 (3) (Apr. 2011) 846–870.
[12] STI, Field application note: rolling element bearings[Online]. Available: http://www.stiweb.com/appnotes/Rolling-Element-Bearings.html2012 (Accessed: 10-Feb-2015).
[13] B. Graney, K. Starry, Rolling element bearing analysis, Mater. Eval. 70 (1) (2012) 78–85.
[14] I. El-Thalji, E. Jantunen, Wear of rolling element bearings, Proccedings Conference of Condition Monitoring and Diagnostic Engineering Management (COMADEM), 2013.
[15] I. El-Thalji, E. Jantunen, A descriptive model of wear evolution in rolling bearings, Eng. Fail. Anal. 45 (2014) 204–224.
[16] I. El-Thalji, E. Jantunen, Dynamic modelling of wear evolution in rolling bearings, Tribol. Int. 84 (2015) 90–99.
[17] S. Sassi, B. Badri, M. Thomas, A numerical model to predict damaged bearing vibrations, J. Vib. Control. 13 (11) (Nov. 2007) 1603–1628.
[18] NSK, General miscellaneous information, NSK Technical Report No. E728g, 2nd ed. 2013, pp. 240–267.
[19] T. Irvine, Ring Vibration Modes, 2014.
[20] S.P. Harsha, K. Sandeep, R. Prakash, Non-linear dynamic behaviors of rolling element bearings due to surface waviness, J. Sound Vib. 272 (3–5) (May 2004) 557–580.
[21] F.J. Harris, Rolling Bearing Analysis I, First, John Wiley & Sons, Inc., New York, 1966.
[22] M. Tadina, M. Boltezar, Improved model of a ball bearing for the simulation of vibration signals due to faults during run-up, J. Sound Vib. 300 (17) (2011) 4287–4301.
[23] I.K. Epps, An Investigation into the Characteristics of Vibration Excited by Discrete Faults in Rolling Element Bearings, University of Canterbury, New Zealand, Christchurch, 1991.
[24] R.S. Dwyer-Joyce, The life cycle of a debris particle, Tribol. Interface Eng. Ser. 48 (2005) 681–690.
[25] V. Mota, P. Moreira, L. Ferreira, A study on the effects of dented surfaces on rolling contact fatigue, Int. J. Fatigue 30 (Oct. 2008) 1997–2008.
[26] S. Al-Dossary, R.I.R. Hamzah, D. Mba, Observations of changes in acoustic emission waveform for varying seeded defect sizes in a rolling element bearing, Appl. Acoust. 70 (Jan. 2009) 58–81.
[27] G.E. Morales-Espejel, V. Brizmer, Micropitting modelling in rolling–sliding contacts: application to rolling bearings, Tribol. Trans. 54 (4) (Jul. 2011) 625–643.
[28] M.N. Kotzalas, G.L. Doll, Tribological advancements for reliable wind turbine performance, Philos. Transact. A Math. Phys. Eng. Sci. 368 (Oct. 2010) 4829–4850.
[29] B. Alfredsson, J. Dahlberg, M. Olsson, The role of a single surface asperity in rolling contact fatigue, Wear 264 (Apr. 2008) 757–762.

Title	**Dynamic modelling and fault analysis of wear evolution in rolling bearings**
Author(s)	Idriss El-Thalji
Abstract	The rolling element bearing is one of the most critical components that determines the health of the machine and its remaining lifetime in modern production machinery. Robust condition monitoring tools are needed to guarantee the healthy state of rolling element bearings during the operation. The condition of the monitoring tools indicates the upcoming failures which provides more time for maintenance planning by monitoring the deterioration i.e. wear evolution rather than just detecting the defects. Several methods for diagnosis and prognosis that are commonly used in practise have challenge to track the wear fault over the whole lifetime of the bearing. The measurements in the field are influenced by several factors that might be ignored or de-limited in the experimental laboratory tests where those advanced diagnosis and prognosis methods are usually validated. Moreover, those advanced methods are verified with the help of simulation models that are based on specific definition of faults and not on considering the fault development process during the lifetime of the component. The purpose of this thesis is to develop a new dynamic model that represents the evolution of the wear fault and to analyse the fault features of a rolling bearing under the entire wear evolution process. The results show the extracted defect features and how they change over the entire wear evolution process. The results show how the topographical and tribological changes due to the wear evolution process might influence the bearing dynamics over the entire lifetime of the bearing and the effectiveness of the fault detection process.
ISBN, ISSN, URN	
Date	May 2016
Language	English
Pages	80 p. + app. 75 p.
Name of the project	
Commissioned by	
Keywords	Dynamic modelling, Wear evolution, Condition monitoring, Fault development, Fault analysis, Rolling bearings
Publisher	

Dynamic modelling and fault analysis of wear evolution in rolling bearings

The rolling element bearing is one of the most critical components that determines the health of the machine and its remaining lifetime in modern production machinery. Robust condition monitoring tools are needed to guarantee the healthy state of rolling element bearings during the operation. The condition of the monitoring tools indicates the upcoming failures which provides more time for maintenance planning by monitoring the deterioration i.e. wear evolution rather than just detecting the defects.

Several methods for diagnosis and prognosis that are commonly used in practise have challenge to track the wear fault over the whole lifetime of the bearing. The measurements in the field are influenced by several factors that might be ignored or de-limited in the experimental laboratory tests where those advanced diagnosis and prognosis methods are usually validated. Moreover, those advanced methods are verified with the help of simulation models that are based on specific definitions of fault and not on considering the fault development process during the lifetime of the component.

Therefore, in this thesis a new dynamic model was developed to represent the evolution of the wear fault and to analyse the fault features of a rolling bearing under the entire wear evolution process. The results show the extracted defect features and how they change over the entire wear evolution process. The results show how the topographical and tribological changes due to the wear evolution process might influence the bearing dynamics over the entire lifetime of the bearing and the effectiveness of the fault detection process.