LOW COST CONDITION MONITORING UNDER TIME-VARYING OPERATING CONDITIONS

by

Theo Heyns

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor (Engineering)

in the

Faculty of Engineering, Built Environment and Information Technology
Department of Electrical, Electronic and Computer Engineering
UNIVERSITY OF PRETORIA

February 2013

SUMMARY

LOW COST CONDITION MONITORING UNDER TIME-VARYING OPERATING CONDITIONS

by

Theo Heyns

Supervisor:	Dr JP de Villiers
Department:	Electrical, Electronic and Computer Engineering
University:	University of Pretoria
Degree:	Philosophiae Doctor (Engineering)
Keywords:	Condition monitoring, Time-varying operating conditions, Discrepancy analysis, Waveform reconstruction

Advances in machine condition monitoring technologies are driven by the rise in complexity of modern machines and the increased demand for product reliability. Condition monitoring research tends to focus on the development of signal processing algorithms that are sensitive to machine faults, robust under time-varying operating conditions, and informative regarding the nature and extent of machine faults. A significant challenge remains for monitoring the condition of machines that are subject to time-varying operating conditions. The here presented work is concerned with the development of cost effective condition monitoring algorithms. It is investigated how empirical models (including probability density distributions and regression functions) may be used to extract diagnostic information from machine response signals that have been generated under fluctuating operating conditions. The proposed methodology is investigated on a number of case studies, including gearboxes, alternator end windings, and haul roads. It is shown how empirical models for machine condition monitoring may generally be implemented according to one of two basic approaches. The two approaches are referred to as *discrepancy analysis* and *waveform reconstruction*.

Discrepancy analysis is concerned with the comparison of a novel signal to a reference model. The reference model is sufficiently expressive to represent vibration response as measured on a healthy machine over a range of operating conditions. The novel signal is compared to the reference model in such a manner that a discrepancy signal transform is obtained. A discrepancy signal is sensitive to faults, robust to time-varying operating conditions, and inherently simple. As such it may further be analysed in order to extract periodicities and magnitudes as diagnostic markers.

Waveform reconstruction implements a regression function to model machine response as a function of different state space variables. The regression function may subsequently be exploited to extract diagnostic information. The machine response may for instance be reconstructed at a specified steady state operating condition. This renders the signal wide-sense stationary so that Fourier analysis may be applied.

OPSOMMING

LAE KOSTE KONDISIE MONITERING ONDER TYD-VARIEERENDE BEDRYFSTOESTANDE

deur

Theo Heyns

Promotors:	Dr JP de Villiers
Departement:	Elektriese, Elektroniese en Rekenaaringenieurswese
Universiteit:	Universiteit van Pretoria
Graad:	Philosophiae Doctor (Ingenieurswese)
Sleutelwoorde:	Kondisie monitering, Tyd-varierende gebruiks toestande, Afwykings analyse, Golfvorm herkonstruksie

Die ontwikkeling van toestand-gebaseerde masjien-moniteringstegnologie word gedryf deur die groeiende kompleksiteit van moderne masjiene, sowel as die aanvraag na hoër betroubaarheid. Navorsing in toestandmonitering fokus veral op die ontwikkeling van seinverwerkingsalgoritmes wat sensitief is vir masjienskade, wat robuust is onder tyd-variërende bedryfstoestande, en algoritmes wat goeie inligting kan ontgin ten opsigte van die aard en omvang van die masjienskade. Die ontwikkeling van masjien-moniteringsalgoritmes wat onder tyd-variërende omstandighede toegepas kan word is ’n aansienlike uitdaging vir navorsers in die veld van toestandmonitering. Die werk wat in hierdie proefskrif aangebied word fokus op die ontwikkeling van koste-effektiewe seinverwerking algoritmes vir toestandmonitoring van masjiene wat blootgestel word aan tyd-variërende toestande. Daar word ondersoek hoe empiriese modelle (waarskynlikheidsdigtheidverdelings en regressiefunksies) gebruik kan word om diagnostiese inligting vanuit die masjienresponsieseine te ontgin. Die voorgestelde metodiek word verken in die konteks van ’n paar gevallestudies, insluitend; ratkaste, ’n alternator, en paaie. Twee benaderings word in hierdie proefskrif verken, naamlik *afwykingsverskilanalise* en *golfvormrekonstruksie*.

Afwykingsverskilanalise is ’n metodiek waar ’n nuwe sein vergelyk word met ’n verwysingsmodel. Die verwysingsmodel is kompleks genoeg sodat dit die gedrag van ’n gesonde masjien (soos gemeet vir verskillende bedryfstoestande) kan voorstel. Die nuwe sein word op so ’n manier met die verwysingsmodel vergelyk dat ’n afwykingsverskilsein gegenereer word. Die afwykingsverskilsein is sensitief vir masjienfoute, is robuust onder tyd-variërende omstandighede en is inherent eenvoudig.

As gevolg hiervan is dit moontlik om die periodisiteit en omvang van die afwykings in die sein verder te ontgin sodat dit as effektiewe diagnostiese kenmerke kan dien.

Golfvormrekonstruksie maak gebruik van regressiefunksies om masjiengedrag as ’n funksie van verskillende toestandveranderlikes te modelleer. Die regressiefunksie kan dan gebruik word om diagnostiese ingligting te ontgin. Die masjiengedrag kan ook rekonstrueer word onder die gewensde bedryfsomstandighede. Een voorbeeld van golfvormrekonstruksie is waar ’n alternator se gedrag onder ’n gestadigde as-rotasiesnelheid rekonstrueer word, sodat standaard tegnieke soos Fourier-analise weer eens toegepas kan word.

ACKNOWLEDGEMENTS

I would like to sincerely thank my supervisor, Dr Pieter de Villiers, for the opportunity to grow through this research; my family (and especially my father) for support and encouragement; Dr Cornelius Stander and Mr Rudi Kroch for making their gear data available; Dr KeSheng Wang for making his alternator end winding data available; Dr Fanus van der Westhuizen at Anglo American who made the road measurement tests possible; Professor Simon Godsill who hosted me in his signal processing laboratory, and Dr Radoslaw Zimroz with whom I enjoyed co-authoring a paper. Last, but certainly not the least, I would like to thank all the reviewers who have taken the time to either consider this thesis, or the parts thereof that were submitted as papers.

List of publications based on this reseach

1. T. Heyns, S.J. Godsill, J.P. De Villiers, P.S. Heyns, "Statistical gear health analysis which is robust to fluctuating loads and operating speeds", *Mechanical Systems and Signal Processing*, 24, pp. 651-666, 2012
2. T. Heyns, J.P. de Villiers, P.S. Heyns, "A method for real-time condition monitoring of haul roads based on Bayesian parameter estimation", *Journal of Terramechanics*, 49 (2), pp. 103-113, 2012
3. T. Heyns, P.S. Heyns, J.P. de Villiers, "Consistent haul road condition monitoring by means of vehicle response normalisation with Gaussian processes", *Engineering Applications of Artificial Intelligence*, Available online 26 January 2012
4. T. Heyns, P.S. Heyns, J.P. De Villiers, "Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox", *Mechanical Systems and Signal Processing*, 32, pp. 200-215, 2012
5. T. Heyns, P.S. Heyns, "Gear fault detection under fluctuating operating conditions by means of discrepancy analysis", *Condition Monitoring of Machinery in Non-Stationary Operations, The second International Conference on*, March 26-28, 2012
6. T. Heyns, P.S. Heyns, R. Zimroz, "Combining discrepancy analysis with sensorless signal resampling for condition monitoring of rotating machines under fluctuating operations", accepted for presentation at *the ninth international conferenc on condition monitoring and machinery failure prevention technologies (CM 2012 and MFPT 2012)*, upgraded to a specialist keynote paper, and subsequently also accepted for publication in *the International Journal for Condition Monitoring (IJCM)*
7. T. Heyns, J.P. de Villiers, P.S. Heyns, "Neural network based waveform reconstruction for condition monitoring under fluctuating operating conditions", submitted to *Mechanical Systems and Signal Processing*, awaiting feedback

LIST OF ABBREVIATIONS

AE	Auto-encoder
AR	Autoregressive
ARMA	Autoregressive Moving-Average
ARD	Automatic Relevance Detection
COT	Computed Order Tracking
FFT	Fast Fourer Transform
GMM	Gaussian Mixture Model
GP	Gaussian Process
GPS	Global Positioning System
IRI	International Road Roughness Index
LSE	Least Squares Error
MAP	Maximum a Posteriori
NN	Neural Network
NLL	Negative Log Likelihood
PSD	Power Spectral Density
PSI	Present Seviceability Index
RMS	Root Mean Square
RSM	Road Surface Monitoring
RTRRMS	Response-Type Road Roughness Measuring Systems
SE	Squared Exponential
TSA	Time Synchronous Averaging
UV	Utility Vehicle

TABLE OF CONTENTS

CHAPTER 1 Introduction, literature survey and document layout 1

1.1 Introduction 1

1.2 Literature survey 1

1.2.1 Applications of condition monitoring 2

1.2.2 Data processing 3

1.2.3 Empirical modelling 15

1.3 Research objectives 18

1.4 Document layout 19

CHAPTER 2 A methodology for cost effective condition monitoring 20

2.0.1 Overview 20

CHAPTER 3 Residual signal and statistical analysis for condition monitoring of a gearbox 24

3.1 Introduction 24

3.2 The gear casing waveform 25

3.2.1 Modeling the healthy waveform 25

3.2.2 Analysing the residual signal 28

3.3 Study on simulated data 31

3.3.1 Dynamic gear model 31

3.3.2 Implementation 36

3.3.3 Results 39

3.4 Study on experimental data 40

3.4.1 Test setup 40
3.4.2 Implementation 44
3.4.3 Results 46
3.5 Chapter conclusion 47

CHAPTER 4 Synchronous averaged Gaussian mixture model discrepancies 49
4.1 Introduction 49
4.2 Methodology 50
4.3 Analytic investigation 56
4.4 Experimental setup 63
4.5 Chapter conclusion 73

CHAPTER 5 Discrepancy analysis with sensorless signal resampling 76
5.1 Introduction 76
5.2 Methodology 76
5.2.1 Time-series modelling 78
5.2.2 Hilbert transform envelope 78
5.2.3 Sensorless relative speed estimation and resampling 79
5.3 Case study 80
5.3.1 Gear model 80
5.3.2 Operating conditions 80
5.4 Implementation and results 82
5.4.1 Residual signals 82
5.4.2 Sensorless relative speed estimation 83
5.5 Chapter conclusion 85

CHAPTER 6 Neural network based waveform reconstruction for an alternator 87
6.1 Introduction 87
6.2 Methodology 88

6.2.1 Time synchronous averaging 89
6.2.2 The feed forward neural network 89
6.3 Simulated case study 91
6.4 Experimental setup 93
6.5 Implementation and results 96
6.5.1 Synchronous averaging 96
6.5.2 Extended modelling of the deterministic signal components 98
6.5.3 Modelling the signal variance 102
6.5.4 Extracting individual componets of interest 103
6.6 Chapter conclusion 105

CHAPTER 7 Bayesian parameter estimation for monitoring a haul road 106

7.1 Introduction 106
7.2 Methodology 106
7.2.1 Response-type road roughness measuring systems (RTRRMS) 107
7.2.2 Bayesian parameter estimation 108
7.3 Experimental setup 111
7.3.1 Measurements 111
7.3.2 Road profiles and signal alignment 112
7.3.3 Parameter prior 113
7.3.4 Bayesian model selection 114
7.3.5 Iterative updating 116
7.4 Results 117
7.5 Chapter conclusion 120

CHAPTER 8 Vehicle response normalisation with Gaussian Processes 122

8.1 Introduction 122
8.2 Metric calibration for consistent road quality evaluation 122
8.3 Gaussian Process Regression 124

8.4 Experimental investigation 126
8.4.1 Generating training data and optimising hyperparameters 128
8.4.2 Function interpretation 130
8.4.3 Model evaluation and results 132
8.5 Chapter conclusion 136

CHAPTER 9 Conclusions and future research 138

Chapter 1

Introduction, literature survey and document layout

1.1 INTRODUCTION

Machine condition monitoring technologies have enjoyed a surge of research interest in recent years due to the rise in complexity of many modern machines, along with the increased demand for product reliability [1, 2, 3]. Machine maintenance is important not only because of its financial implications, but also due to the significant risk that malfunctioning machines may pose to humans [2].

A condition monitoring system typically comprises three key aspects, namely data acquisitioning, data processing, and maintenance decision support [1, 3]. This thesis is closely aligned with the second stage of data processing and interpretation. It is in particular investigated how diagnostic information may be extracted from vibration signals that are generated on machines when the machines are subject to time-varying operating conditions.

The literature study commences with a brief background on the domains related to the case studies presented later in this work. The bulk of the literature study then focuses on signal processing and feature extraction methodologies. Finally the literature study concludes with a discussion on empirical modelling, including probability density and regression functions.

1.2 LITERATURE SURVEY

The literature survey is conducted in three sections:

1. The first section briefly explores condition monitoring applications. In particular, this section includes general background information on the problem areas related to the case studies presented in this thesis.

2. The second section is concerned with signal processing and feature extraction techniques. Some of these techniques are fairly specific to the field of condition monitoring, but in general they

are also applicable to many other signal processing applications.

3. The last section briefly considers empirical modelling, and in particular the application of probability density functions and regression functions. The empirical models will be discussed in greater detail in the respective chapters where they will be used as signal processing tools.

1.2.1 Applications of condition monitoring

Condition-based maintenance (CBM) is a rapidly expanding field [1] with research and applications in diverse fields of machinery, including manufacturing tools [4], jet engine health monitoring [5], draglines [6], gearboxes [7], bearings [8], and induction machines such as alternators [9]. One relevant example is the aeromotive industry. Helicopter operating costs are extremely high, of the operating cost 24% is accounted for due to maintenance, while insurance accounts for an additional 29%, [2]. Proper condition based maintenance can offer significant financial and human safety rewards.

The case studies presented in this work include the condition monitoring of gearboxes, alternator end windings, and haul roads. The case studies pose unique challenges, but will also be seen to share many similarities. They will be used to illustrate how empirical models of different machine response signals may be generated and how these empirical models may effectively be used to extract diagnostic information from the signals.

1.2.1.1 Rotating machines

Condition monitoring of rotating machines, and especially drive train components (including bearings and gears) are especially important in the field of condition based maintenance, not only because these parts are common to many types of machinery, but also because these load carrying parts are especially prone to failure [2, 3, 7, 10, 11].

Time varying torques and shaft rotational speeds may result in significant amplitude, phase and frequency modulation [10]. Baydar and Ball [12] indicate that not only do the amplitudes and sideband content of frequency components change under varying load conditions, but some non-fault related components may also appear or disappear completely. As such there are significant ambiguity and noise in many vibration signals which are generated under time-varying operating conditions, such that it is often challenging to extract effective diagnostic information [10].

1.2.1.2 Haul roads

Various road maintenance policies, such as routine, periodic and urgent maintenance [13] have been investigated towards the goal of minimising both road and vehicle maintenance costs. However haul roads are often unpaved and constructed from materials with diverse qualities. Not only are these roads subject to variable traffic volumes, but are also used by vehicle types which cover a range of different characteristics and payloads. Standard haul road maintenance management systems are generally poorly suited to deal with these complex and dynamic environments [14]. Poor maintenance strategies result in increased vehicle maintenance costs and reduction in service [14]. Recently the use of machine response signals to assess the condition of haul roads on a real time basis has enjoyed significant attention in an attempt to reduce the substantial maintenance costs (and potential lost production time) associated with the roads and the haul trucks [15, 14, 16, 17].

It has been found that operating conditions (such as the speed of the vehicle, its acceleration and its payload) significantly affect the machine response signals and that isolated interpretation of the vibration data, without considering the vehicle operating conditions, could result in ambiguous road classification.

1.2.2 Data processing

A good data (or signal) processing stage will result in a compact (low-dimensional) signal representation, which reflects the presence of machine damage, is free of signal errors and noise, and which is robust to time-varying operating conditions. Ideally the processed data should also be represented in an intuitive manner, so that a physical understanding of the results may either serve for direct interpretation by a maintenance engineer, or may be used to implement an appropriate diagnostic algorithm.

1.2.2.1 Descriptive metrics

Some of the earliest condition based maintenance systems relied on descriptive metrics which were used to describe certain characteristics from time domain measurements.

Literature on condition monitoring of rotating machinery refer to a number of such time domain descriptive metrics that could be used to detect machine faults based on the increased energy or impulsivity of the vibration waveform. A number of such metrics that are popular in condition based maintenance include: peak values, peak to peak values, root mean square (RMS) values, skewness, shape factor, crest factor, impulse factor, clearance factor, kurtosis, energy ratio, energy operator and many more [18, 2, 1].

While these metrics are generally very simple to implement, they are often insensitive to machine gear faults, while being prone to ambiguity in the presence of time varying operating conditions [1, 2, 19, 20]. These metrics are often dominated by the effects from the normal operational vibration components, such as the gear meshing frequency in a gearbox. The normal operational vibration components, which are generally affected by time-varying operating conditions, and which exhibits some noise, tend to mask the relatively small energy deviations associated with many machine faults [19, 20].

Descriptive metrics also have a long history in the context of haul road monitoring, where they are also referred to as RTRRMS (response-type road roughness measuring systems). Some of the metrics are concerned with ride comfort (human response to vibration) such as the ISO 2631 standard [21]. This metric estimates the average power absorbed by the driver and is computed based on a frequency weighted acceleration RMS value as measured on the driver seat. Other metrics may be more concerned with the vehicle-road interaction forces, such as the metric proposed by Quinn and Wilson [22] which aims to represent the dynamic tire forces (and subsequently the road quality) based on the measured tyre pressures. The international road roughness index (IRI) [23] was proposed as a standardised metric indication of road roughness based on a large scale study financed by the World Bank [23]. The IRI is based on the ratio of the accumulated suspension motion of a standard vehicle to the distance travelled by the vehicle during the measurement. The IRI may be estimated directly from a known road profile by simulating the response of a standardised quarter car vehicle at a constant speed of 80km/h. The IRI has widely been adapted as a standardised measure of road roughness. A roughness measurement which is based on the suspension stroke was partly chosen for convenience during the early development of the "Via-Log" and roadmeters [24]. It later turned out to be a reasonable approximation to the ride quality of a vehicle as commonly measured by acceleration on the body [24]. A road severity metric which is deduced from vehicle response is inherently sensitive to the operating conditions under which the measurements were performed. In order to consistently compare vehicle response-type metrics from different measurements, it is required to perform those measurements under similar operating conditions with a standardised vehicle or vehicle model [23, 25].

1.2.2.2 Digital filtering algorithms

Signal filtering is concerned with the removal of vibration components and signal noise which do not contain diagnostic information and which only obscures the results. Modern digital filters can provide very high performance high pass, low pass, band pass or band stop performance [26]. Digital filters may be implemented in the time domain based on convolution where the raw signal is convolved by the filter impulse response, or it could be implemented based on line elimination and resynthesis by means of the fast Fourier transform [26].

Adaptive filters may be used to denoise non-stationary signals which are generated under fluctuating operating conditions. Khemili and Choachane [27] investigates adaptive noise cancellation and the adaptive-tuning filter on both simulated and measured vibration signals with the aim of detecting rolling element bearing defects. It was found that the adaptive filters are capable of increasing the signal to noise ratio for both signals, which in turn allowed for improved fault detection. Ettefagh and Sadeghi [28] investigates how a time-varying autoregressive model (ARMA) could be used to extract time-varying parameters which represent the signal characteristics and which may be used as features for identifying gear fault conditions.

Blind source separation (BSS) is closely related to the concept of adaptive filters and noise cancelation. BSS allows for the recovering of multiple independent sources from their mixtures [29]. Gelle et al. [30] applied BSS to vibration data from a test bed which consisted of two motors running simultaneously. Data were recorded by means of multiple sensors which were attached to various positions of the motors. The speeds of the motors varied independently over time. BSS was used to separate the influences of different excitation sources. Popescu [29] used BSS on data from a pump that is driven by an electromotor. The different signal excitation sources were separated so that each of the sources could be assessed against a change detection criterion which indicated the onset of damage to the motor.

Signal filtering techniques may also be used to distinguish between deterministic and random components in a signal. It is often useful to identify and to separate the deterministic and random components in a vibration signal [31]. One important example is to differentiate between the deterministic gear mesh components and the more stochastic bearing related components in a gearbox response signal [31]. Time synchronous averaging (TSA) is a powerful digital filtering technique for extracting deterministic components from a vibration signal [32]. Synchronous averaging (SA) does not have to be limited to the time domain, but may also be extended to the resampled order domain to better operate on signals which are generated under fluctuating shaft speeds [10]. Synchronous averaging is recognized as an essential algorithmic tool for condition monitoring of rotating equipment [1, 33, 32]. The SA is computed as the ensemble average of the time or order domain signal over a number of shaft revolutions. This reduces noise and non-synchronous signal components so that the signal characteristics which are of diagnostic interest may be enhanced.

Mdlazi et al. [34] conduct an interesting study where the TSA of a gear vibration signal is estimated by means of supervised regression functions, including a feed-forward neural network (FFNN), and a support vector machine. A reference signal that is synchronous with the shaft angular position is provided as input, and the vibration response is used as target output. It is found that the FFNN can generate robust estimates of the TSA. In one case study the neural network required only a quarter of the amount of data used by direct averaging to produce a TSA of similar quality.

1.2.2.3 Residual signal analysis

Residual signal analysis is a methodology which aims to remove the normal (or healthy) components from a signal so that only those components which are damage related are considered. Residual signal analysis is thus closely related to digital filtering techniques.

In literature residual analysis has only been investigated with respect to rotating machines. A number of different approaches of obtaining residual signals have been investigated. Stewart [35] computes a residual signal by eliminating the gear meshing harmonics and its adjacent sidebands from the signal average spectrum before converting the signal back to the angle domain. Lin et al. [36] investigate a residual signal based on the work by Miller [37], where a family of Morlet Wavelets is composed based on an assumed gear motion model. These wavelets are used as a comb filter to decompose the signal into the gear motion and residual signals. Metrics computed from the residual based on the Wavelet approach have been shown to be sensitive to propagating gear faults. Wang and Wong [38] propose that an autoregressive (AR) model may be used to represent the vibration response as measured on a machine when it is still in a healthy condition. A residual signal is subsequently obtained as the difference between a novel signal and the one step ahead predictions made by an autoregressive filter. The residual signal is shown to be an effective tool in the detection and diagnosis of gear faults.

Apart from mentioned above techniques, it should be recalled that Antoni and Randall used so-called self adaptive noise cancelation (SANC) techniques based on a least mean squares (LMS) adaptive filter [39]. Zirmoz and Bartelmus have applied normalised LMS in order to extract a residual signal from a vibration signal measured on a belt conveyor pulley drive unit in presence of time varying speed/load conditions [40]. An advanced approach that uses adaptive Schur filter was recently proposed by Makowski and Zimroz [41].

In the literature it is typical to compute a simple metric of the residual signal which can be used for comparison and change detection. Stewart [35] defined a non-dimensional parameter (FM4) which was found to be especially sensitive as a general fault detector when detected to the residual signal.

Some additional metrics which have specifically been devised for use with a residual signal include NA4, M6A, M8A, NB4, NA4, along with a number of revised versions of these metrics [2].

1.2.2.4 Spectral analysis

Fourier analysis is based upon the assumption that a signal may be represented by the super positioning of multiple sinusoidal components of different frequencies and phases [42]. The Fourier transform may very effectively be implemented for discretised (digital) signals by means of the fast Fourier transform (FFT). The FFT is used at the core of a number of efficient computations such as bandpass filtering, convolutions and computing the signal power spectral density (PSD) [42].

Spectrum analysis plays an important role in condition monitoring applications related to rotation machinery which exhibits cyclo-stationary (or near cyclo-stationary) characteristics. Spectrum analysis allows for the identification and isolation of specific frequency components. These components may often be interpreted based on a physical understanding of the machine dynamics, hence offering important insight into the condition of the gearbox [1]. It is for instance known that damaged gears tend to result in increased signal energy at the gear mesh frequency and its sidebands [43].

Spectrum analysis is also relevant in the context of road condition monitoring, especially with regard to assessing ride quality. Humans tend to be especially sensitive to certain frequencies of excitation, since those frequencies may cause the human torso (or eyes, head, etc) to oscillate excessively [44]. Fourier analysis may also be used to interpret road conditions based on the spectral characteristics of the measured road profile. ISO codes specifies standardised road categories which could be used to classify road conditions [16, 17].

The Fourier Transform is a linear operator and depends on the wide-sense stationary character the analysed signal [45]. Spectral analysis is thus not well suited to deal with vibration signals which have been generated under time-varying operating conditions. Spectral analysis is also not well suited to represent and detect non-sinusoidal signal characteristics such as impulsive behaviour which might arise due to localized gear damage [46].

Parametric approaches may also be used to estimate the spectral characteristics of a signal. Two common parametric spectra estimating techniques include autoregressive (AR) spectra, and autoregressive-moving-average (ARMA) spectra [1].

1.2.2.5 Cepstrum analysis

Cepstrum analysis is a nonlinear signal processing technique, which was originally proposed as a better alternative to the autocorrelation function for detecting echo delay times [42, 47]. A number of different definitions are used for the cepstrum, but perhaps the most popular is the power cepstrum which is defined as the inverse Fourier transform of the logarithmic power spectrum [1, 47].

Cepstrum analysis has been implemented in three main areas of rotating machine diagnostics, including to collect families of uniformly spaced sidebands and harmonics; separate forcing functions (as measured on the gears) from the resulting measurements at various measuring points on the gearbox (as influenced by the transmission transfer function); and for identifying echoes, inverted echo pairs, and echo delay times [42].

The ability of cepstrum analysis is especially interesting for analysing localized gear faults. Uniform gear wear tend to increase the harmonics (spectrum) of the gear meshing frequency, especially the second harmonic. Sidebands around the harmonics tend to be narrowly spaced. Locals faults tend to result in more uniformly (flat) harmonics and sidebands. These harmonics and sidebands for localised faults may be difficult to interpret, especially if multiple gears are in mesh. Cepstrum analysis is capable of collecting all the harmonics and sidebands from individual families so that it becomes much easier to interpret [42].

Badaoui et al. [47] conducted research on the use of the power cepstrum for diagnosis of the gear system. An indicator of the condition of a gear was developed which was found to be little affected by the signal amplitude, the signal-to-noise ratio, or the sensor position.

Dalpiaz et al. [48] compared a number of different vibration analysis gear condition monitoring techniques, including time-frequency techniques, cyclostationarity analysis methods, cepstrum analysis and time-synchronous averaging. The power cepstrum was found insensitive to the crack evolution, while techniques related to the time-synchronous average, especially the residual signal was found to be more sensitive to the tooth damage.

1.2.2.6 Spectral kurtosis

Spectral kurtosis (SK) was introduced in the 1980s as a technique for detecting and localizing the presence of non-Gaussian signal characteristics [49]. Equivalently it may be used to detect frequency bands which are prone to impulsive characteristics. It was also shown that SK may be implemented as a filter to recover randomly occurring signal components which occur in the presence of stationary additive noise [50].

The properties of the SK was substantially simplified when it was redefined from being computed as the normalised fourth-order moment of the real part of the short-time Fourier transform, to a modified definition which is based on the normalised fourth-order moment of the magnitude of the short-time Fourier transform [51, 50].

Spectral kurtosis is a useful signal processing tool in the context of condition monitoring, since many rotating machine faults (e.g. tooth defects or bearing faults) tend to generate impulse responses which can successfully be detected by means of SK [42].

An interesting study on SK is presented by Barszcz and Randall [52], where SK was used on the difficult problem of detecting a tooth crack in the planetary gear in a wind turbine. The same fault which could not be detected by means of synchronous averaging until shortly before catastrophic failure, was shown to be detectable by means of SK as early as 8 weeks in advance. This is because the detected peaks caused by the tooth crack have a very short duration, occur seldom and have varying amplitudes. Barszcz and Randall concludes that the phase modulation induced by the elastic deformation of the gear train components impede the synchronous averaging process to such an extent that the high frequency pulses cannot be detected.

1.2.2.7 Time-frequency analysis

Frequency-domain analysis is inept at handling non-stationary waveform signals. A number of time-frequency techniques have subsequently been devised in an attempt to better analyse non-stationary signals by representing the signal energy (or power) as two dimensional functions of both time and frequency so that fault induced signal patterns may better be observed [1].

Some popular time-frequency transforms for condition monitoring of rotating machines include the short-time Fourier transform (STFT), or the spectrogram (the power of the STFT) [12, 53], the Wigner-Ville distribution and the Choi-Williams distributions [1, 54] and the S-transform as defined by Stockwell et al. [55] (not to be confused with the mathematical definition of the S-transform).

Essentially the STFT is implemented by dividing the waveform signal into windowed segments so that the Fourier transform may be applied to the individual signal segments. The segmentation procedure limits the time-frequency resolution of the Spectrogram so that it can only be applied to signals of which the characteristics vary slowly with time [1]. Bilinear transforms (such as Wigner-Ville distribution) are not based on signal segmentation and thus overcome some of the time-frequency resolution limitations of spectrogram. The Bilinear transforms do however suffer from interference terms [1]. Many additional studies have been conducted towards the goal of defining time-frequency distributions with good resolutions, while also experience minimal adverse effects of interference terms [53]. One such study is presented in [56].

Time-frequency transforms are also relevant to the monitoring of haul roads. Research was conducted by Hugo et al. [16] as well as by Ngwangwa et al. [17] where estimated road profiles are window segmented so that the power spectral densities of the individual windows could be computed and classified according to the ISO 8608 [57] road classification standards. Unlike traditional Time-frequency windowing where the window is a function of time, in the road monitoring context it rather is a function of distance. Similarly metrics which are concerned with ride comfort such as the ISO 2631 standard [21] can also be used to classify individual intervals of road. Since these metrics are based on a frequency-weighted acceleration energy, windowed Fourier transforms must also be computed.

Despite the potential of these techniques, their application in practice is still limited owing to the difficulty of quantitative interpretation based on time-frequency representations. Subsequently, the task of fault detection generally relies on interpretation by skilled experts, which not only introduces subjective factors in the fault diagnosis, but also limits real time condition monitoring [58].

1.2.2.8 Wavelets and wavelet packet analysis

During recent years Wavelet theory research have resulted in multiple powerful mathematical and signal processing tools [59]. Wavelets are not used to compute time-frequency representations of a signal, but rather time-scale representations. Wavelets are implemented by convolving different dilations and translations of a so-called mother wavelet with the signal of interest [1]. Unlike the STFT which has a constant time-frequency resolution, the time-scale resolution of the wavelet transform varies. At low scales (high frequencies) the wavelet enables a good time resolution at the expense of poor scale resolution, while at high scales (low frequencies) good scale resolution can be obtained at the expense of poor time resolution [1, 59]. Some of the applications for which Wavelets have been used in the context of condition monitoring include time-scale analysis, fault feature extraction, singularity detection, denoising and extraction of weak signals, vibration signal compression, system and parameter identification, etc [59].

Wavelet analysis have been used to extract features which could be used be used to detect gear damage (even in the presence of time-varying operating conditions) [11, 60, 61, 62, 63]), to detect and isolate electrical faults in an induction machine [64, 65], extract features which are related to bearing damage [66, 67], monitor the structural condition of airplanes [68], and even monitor the condition of a bridge based on the detection of singularities in the vibration response of the bridge when traversed by a vehicle [69].

Despite its wide range of application in literature, when compared to Fourier analysis wavelet analysis has still not achieved a standard status [59]. Wavelets are inherently complex to implement; there is for instance no general method to select the wavelet function for monitoring tasks, it may be difficult to determine the scale ranges of interest, and unlike the Fourier transform does the results from a wavelet transform not offer straightforward physical implications which may be used to extract useful information directly from the results of the wavelet transform [59].

1.2.2.9 Cyclostationarity

Cyclostationarity is a signal processing tool for analysing signals that exhibit periodic behaviour. Whereas classical Fourier analysis is well suited to analyse processes that are deterministic functions of time, cyclostationarity is well suited to analyse processes which exhibit statistical characteristics which vary periodically with time [70]. Cyclostationary processes have also been referred to as periodically correlated processes [70].

Hurd and Miamee [71] explain the concept of cyclostationarity (or periodically correlated sequences) as follow: "Periodically correlated (PC) random processes are random processes in which there exists a periodic rhythm in the structure that is generally more complicated than periodicity in the mean function." Examples of such periodic processes may arise in telecommunications, telemetry, radar and sonar due to modulation, sampling, multiplexing, and coding operations [70].

While a significant body of the precursory work in cyclostationarity arose in the context of communications, research on the application of cyclostationarity in mechanical signal analysis is now rapidly increasing [72]. One of the pioneers in the field, William A. Gardner, foresaw that the cyclostationary framework is appropriate for many physical phenomena that give rise to data with periodic statistical characteristics [72]. In fact it is now being realised that Fourier analysis is not well suited to explain most mechanical systems, because by definition, stationary signals are representative of physical phenomena that maintain a constant statistical behaviour in time. This behaviour is seldom realised in mechanical systems, even when the machines are subject to constant operating conditions.

Cyclostationarity in mechanical systems is evident in systems such the meshing of teeth in gears, the combustion of gas in internal combustion engines, the inversion of forces in reciprocating or cam mechanisms, the admission and exhaust of fluids in pumps and turbulence around fan blades where phenomena occur on a rhythmic basis [72].

1.2.2.10 Hilbert-Huang transform (HHT)

The Hilbert-Huang transform (HHT) was introduced in 1998 [73] as method for analysing nonlinear and non-stationary signals. The HHT is based on the sequential use of empirical mode decomposition (EMD), and the Hilbert transform. EMD is a signal processing technique which adapts according to the local characteristic time-scale of a signal. In essence the EMD method is used to decompose a potentially complex signal into a finite (generally small set) of intrinsic mode functions (IMFs). The nature of the IMFs is well suited to analysis by means of the Hilbert transform. The Hilbert transform may thus be used to extract instantaneous frequencies of the IMFs as function of time so that local signal characteristics may be identified.

Loutridis [74] explored the use of EMD to monitor the evolution of gear faults. It is found that the instantaneous frequency of the intrinsic mode functions may serve as a sensitive indicator of damage in the gear pair.

Yu et al. [75] investigate how gear tooth damage can be detected by comparing the HHT entropy from a signal generated by a damaged gearbox, to the HHT entropy of a healthy gearbox. The results from both a simulated and an experimental study indicates that the approach allows for the identification of gear damage, but it is not easy to classify the nature of the gear damage.

Rai and Mohanty [76] perform a study of the use of the HHT in the context of bearing fault diagnosis. In particular they argue that a simple FFT, rather than the Hilbert-Transform, of the intrinsic mode functions should be computed to avoid some of the complexities which are introduced by the time resolution of the Hilbert-Transform. They found that the FFT of the IMFs allowed for the efficient detection of bearing damage.

1.2.2.11 Signal normalisation with respect to fluctuating operating conditions

Time-varying operating conditions may induce signal distortions and ambiguities which adversely affect the ability of condition monitoring techniques to analyse the condition of a machine. A number of signal processing techniques have been proposed towards the goal of isolating or accounting for the effects of time-varying operating conditions on a vibration signal. The influence of time-varying operating conditions in the context of rotating machines have widely been investigated. Fluctuating torque tend to cause amplitude modulation of the vibration signal, while fluctuating rotational speeds induce frequency modulation [10]. Baydar and Ball [12] indicate that not only do the amplitudes and sideband content of frequency components change under varying load conditions, but that some non-fault related components may also appear or disappear completely.

Various order domain techniques exist which attempt to remove the effects of non-stationary speeds on the vibration signal. Computed order tracking resamples a time-domain signal to fixed increments of the shaft position. This resampling at least partially restores the wide-sense cyclo-stationary character of the vibration signal, hence rendering it possible to analyse the order domain spectrum [9]. Order domain averaging computes an average of the resampled signal over a number of revolutions in attempt to reduce noise and non-synchronous effects from sources other than the shaft of interest [10]. Other order tracking techniques, such as Vold-Kalman filters or Gabor order tracking, assume sinusoidal waveforms to extract order information from the time domain [9]. While these techniques often prove very valuable, they continue to suffer from certain limitations, such as sensitivity to jittery shaft speeds or fault induced signal modulation [9].

Wang and Heyns [9] investigate how a Vold-Kalman filter can be combined with computed order tracking to analyse vibration data which are subject to both frequency and load modulation. The Vold-Kalman filter is used to approximate the signal as the sum of a number of convolutions between sinusoidal carrier waves (of which the possibly varying frequency is related to specific orders of interest) and slow time-varying envelopes. The Vold-Kalman filter is used to extract selected sinusoidal components, before computed order tracking is used to resample the signal for better Fourier analysis. However, since the order tracking techniques assume sinusoidal base functions they are not well suited to analyze certain signal characteristics, such as localized impulses. For this reason Wang and Heyns propose that empirical mode decomposition can be included as an additional pre-processing stage so that the Vold-Kalman filter can subsequently be applied to the intrinsic mode functions rather than the raw signal.

As previously stated do order domain resampling techniques rely on a reference signal which can be used to resample the signal. Due to physical and financial constraints it is not always possible to install equipment (e.g. an optical encoder or tachometer) which are required to accurately measure the shaft angular position [77]. For this reason research [77, 78, 79] has been conducted on how the shaft angular position may directly be estimated from the vibration signal. Bonnardot et al. [77] consider a phase demodulation technique, where a shaft harmonic is selected and isolated by means of appropriate filtering. The Hilbert transform of the signal is used to represent the signal in its analytic from. This subsequently allows for the extraction of the phase angle of the identified frequency component. Urbanek et al.[78] investigate the relative performance of the phase-based methodology first proposed by Bonnardot et al.[77] with an amplitude-based method. In addition they also briefly refer to a methodology which allows for estimating the relative shaft speed by inspecting the local maxima in a spectrogram of the vibration signal. The proposed methodology avoids the need for expensive additional transducers to measure the shaft angular position, but is somewhat limited both in terms of the accuracy and the range of speeds to which it may be applied [77, 78]. Combet and Gelman [79] extend the algorithm proposed by Bonnardot et al. [77] by allowing for a method whereby the mesh harmonic which will be narrow-band demodulated and used to estimate the shaft phase, may automatically be selected based on the local-signal-to noise ratio. Further research on sensorless frequency/speed estimation via time scale and time-frequency analysis can be also found in works by Combet and Zimroz [80], Millioz and Martin [81], as well as Zimroz et al. [82, 83]. Interesting results have also been obtained by the research team led by Antoniadis by means of parametric modelling and by implementing the Teager-Kaiser operator [84, 85].

Stander and Heyns [43] propose a load demodulation normalization procedure which accounts for the influence of fluctuating loads. After the signal is resampled to account for fluctuating shaft speeds, the signal is subsequently divided by the load-modulating envelope. The envelope is estimated as the low pass filtered signal maxima. The low pass filter frequency bandwidth is optimized so as to ensure the best conformity between the statistical properties of vibration signals which are measured under different load conditions. A more consistent deterministic signal may subsequently be extracted by means of synchronous averaging.

Stander and Heyns [7] also propose a methodology to deal with transmission path phase distortions by performing synchronous averaging with regard to an instantaneous phase of the reference frequency of interest, rather than the actual shaft rotational position.

The effect of fluctuating operating conditions are not limited to rotating machines, but is also of consequence when computing RTRRMS (response-type road roughness measuring systems).

It was reported in the International Road Roughness Experiment [23] that RTRRMS (response type road roughness measuring systems) are generally sensitive to both the road profile and the vehicle operating speed. It is also noted in that the relationship between the RTRRMS metrics and the vehicle speed is not consistent, but rather depends on the type and roughness of the surface which is traversed. This is because the RTRRMS (such as the conventional IRI measure) reflects the presence of different wavebands and road signatures when operated over different surfaces at different speeds.

Rashid and Tsunokawa [86] considered the practical difficulty of estimating consistent IRI values on urban roads which are subject to traffic and where it hence is not possible to consistently maintain the specified speed of 80km/h. They proposed a number of speed normalising equations which could be used to relate the measured IRI readings at different speeds to the estimated IRI response at 80km/h. This approach generally improved the accuracy of their IRI values. The calibration strategy is however limited, since the calibration function fails to consider the unique waveband characteristics of the road on which the measurement is performed.

1.2.3 Empirical modelling

This section broadly refers to any models that are dependent for its construction or optimization on large sets of representative data. This discussion on empirical models includes the field of regression and classification, which is often collectively referred to as supervised learning, as well as to the estimation of probability density functions and the estimation of parameters for dynamic models.

Much research has been conducted on the use of classification algorithms (e.g. Neural networks, Support vector machines, Fuzzy logic systems, Decision trees, etc.) [1, 2, 11, 60, 61, 87, 88, 14] to interpret data and to subsequently classify the condition of rotating machines, or the expected time until failure.

Thompson and Visser [14] propose that a Neural Network (NN) may be used to classify vibration patterns that are representative of different road conditions. Hugo et al. [16] investigate the use of an inverse vehicle dynamics model to estimate the road profile from a measured vehicle response, while Ngwangwa et al. [17] propose that the road profile may be estimated by means of a NN, rather than an analytic model. Hugo et al. and Ngwangwa et al. subsequently assess the estimated road profiles according to the ISO 8608 [57] road classification standards. Emperical models may be trained to recognize complex patterns, which are beyond the capabilities of humans. Once trained a classification algorithm is cheap to implement. As such classification algorithms offer the very tempting potential of real time condition monitoring, automated diagnosis and estimated time to failures.

While empirical models techniques may potentially be very powerful, they are however also limited by the quality of the available data. In order to construct these machine condition classifiers data are required which must be representative of different fault conditions as experienced over different operating conditions. In the case of vehicle road classification it generally proved difficult and very costly to perform measurements or test in order to generate data which could be used to characterise vehicle components for the inverted vehicle model [16], or train neural networks to estimate the road characteristics[14, 16]. Since training data must usually be generated through the manual classification of fault historic data, or data obtained from destructive tests, these supervised classifiers have generally enjoyed limited success in industry [89].

There are however methods where empirical models may be implemented without the need for data which are representative of different fault conditions. It will be seen in this thesis how many of the advantages of empirical models may be leveraged in the absence of extensive and costly data sets.

1.2.3.1 Probability density functions (PDFs)

The general field of estimating PDFs is vast and may be approached from different perspectives, including classical (frequentists) statistics, Bayesian statistics, and the machine learning community. Some parametric density functions may be simple to implement, and when appropriate may be optimized with limited data. A chosen density function may however be a poor model of the distribution that generates the data, and may subsequently result in poor performance. Non-parametric approaches to density estimation make few assumptions regarding the form of the data, however these methods may quickly run into problems with data over fitting (especially for higher dimensional distributions) and large computational burdens [90].

Mixture models, such as Gaussian mixture models (GMM), offer a sort of middle ground between fully parametric and non-parametric density estimation techniques. A GMM comprises a weighted super positioning of individual multivariate Gaussian components [90]. The complexity of a GMM may be controlled by limiting the number of mixture components, or by constraining the covariance matrices. When sufficient training data are available GMMs may be used to estimate very complex density distributions [90, 91]. Mixture models are well suited to represent data which might be generated by different classes. The classes may be unknown and thus treated as latent variables [91].

PDFs may be used to model one class data (e.g. normal data from a gearbox) and then subsequently detecting samples (e.g. fault induced vibration response) which deviate from the baseline model. Timusk et al. [92] conduct an interesting study where a number of novelty detection algorithms, including an algorithm based on a GMM density distribution, are used to extract diagnostic robust features from data which were generated on an experimental gearbox test rig under time-varying operating conditions. A related problem is that of speaker change detection, where the problem is to determine the time instances in an audio stream when a voice or speaker change. In recent years GMM-based change detection methodologies have served as the dominant approach for speaker change detection, primarily due to GMM's good ability to detect various acoustic changes [93].

1.2.3.2 Regression

Regression is the mapping from a set of input vectors (independent variables) to an output space of one or more continuous (dependent) variables. Closed form analytic solutions exist for simple linear functions as optimised with least squared errors [94], while complex optimization strategies may be employed for powerful kernel based regression functions such as support vector machines [90, 91]. Different methodologies may be employed which aim to ensure good generalization of an optimized regression function. These methods may include regularization, Bayesian model selection and cross-validation [90, 91, 94]. Various regression function techniques have been investigated in the context of condition based maintenance. There are numerous good sources which discuss the unique advantages and challenges with regard to different regression function methodologies, two of the sources include the book by Bishop [90], and the book by Duda et al. [91] . A brief discussion is offered regarding some of the techniques which are of interest in this thesis.

Wang and Wong [38] use an autoregressive (AR) model to represent the healthy state of a machine response signal in order to compute a residual signal. An AR model is an example of linear regression which is applied to time-series modelling. AR models are well understood and have low computational demands, but may be too constrained to operate well under time-varying operating conditions.

Neural Networks (NN) are parametric models which allow for the nonlinear mapping of a set of input vectors to associated target values. The network is trained by iteratively adjusting a number of parameter values until the network outputs closely correspond to the example target values. The number of parameters (degrees of freedom) of the model determines its expressive power and it has been shown that even a three layer network (one hidden layer) may approximate any continuous function, provided that the hidden layer contains enough sigmoid nodes [91, 90]. NNs may be optimized by means of an efficient back propagation gradient descent optimization algorithm. This renders NNs well suited to the modelling of large data sets.

Mdlazi et al. [34] conduct an interesting study where the synchronous average of a gear vibration signal is estimated by means of machine learning regression functions, including a feed-forward neural network (FFNN), and a support vector machine. A reference signal that is synchronous with the shaft angular position is provided as input, and the vibration response is used as target output. It is found that the FFNN can generate robust estimates of the synchronous average. In one case study the neural network required only a quarter of the amount of data used by direct averaging to produce a synchronous average of similar quality.

Ngwangwa et al. [17] propose that a NN can be trained to model the inverse dynamics of a vehicle. This renders it possible to use NN regression to reconstruct a road profile based on a measure vehicle response.

Recently the development of practical Bayesian methods has opened some of the most active research directions in machine learning [90, 95]. In the Bayesian approach to model selection, all possible models are considered, but preference is given to simpler functions which tend to generalize well. Bayesian methods may be applied to parametric models, such as polynomial regression, or to non-parametric methods such as Gaussian Processes (GP).

A GP is a form of kernel machine, which implements the Bayesian framework for automatic model selection. A GP is implemented by placing a prior distribution directly over the space of functions [95]. Such a prior, which is inherent to the so called kernel functions, may typically specify that the true function is expected to be continuous and smooth. The probability that any function describes the true underlying function is thus related to both its simplicity and how well it describes the observed (training) data [96]. GPs are known to generalise very well, but are computationally expensive to implement [97, 98]. In practice GPs are generally limited to data sets of a few thousand examples at most [95].

1.2.3.3 Outlier, novelty and change detection

The AR filter based residual signal technique is conceptually related to a large group of methodologies which are concerned with outlier or novelty detection in data sets. Essentially these outlier or novelty detection methodologies are capable of modelling one class data (e.g. normal data) and then subsequently detecting samples (e.g. fault induced) which deviate from the baseline model. Some popular novelty or outlier detection methods include: rule based techniques, neural networks (including multilayer perceptrons, self organizing maps and auto-associative networks), support vector machines, Bayesian networks, nearest neighbour and distance techniques, as well as Gaussian Mixture Models (GMM) density based approaches [99, 100, 101, 102].

Timusk et al. [92] conducts an interesting study where a number of novelty detection algorithms are used in conjunction with various signal processing techniques to analyse data which were generated on an experimental gearbox test rig under time-varying operating conditions. The processing and novelty detection techniques are used to extract robust features which are shown capable of discerning between a number of specified fault conditions.

1.3 RESEARCH OBJECTIVES

This thesis is concerned with the development of a cost efficient diagnostic framework. In particular the objectives of the presented research is the development of a diagnostic framework that:

- extracts diagnostic information from machine response signals,
- extracts diagnostic information in such a manner that it is sensitive to the presence of machine

faults, yet also robust to time-varying operating conditions,

- is capable of fusing data from multiple sources in order to enhance the performance of the algorithm (such as considering contextual information from sensor measurements which are correlated with the time-varying operating conditions),
- is also capable of interpreting the vibration signal when limited, noisy or no measurements of the operating conditions are available,
- is not dependent on fault historic data, destructive tests, or extensive manual preparation of training data,
- allows for a simple intuitive representation of the extracted information, such that a non-expert, or a simple classification algorithm may interpret it.

Towards the above objectives is it investigated how empirical models (including PDFs and regression functions) may be used to model machine response signals and subsequently how these empirical models may be used to extract diagnostic information.

1.4 DOCUMENT LAYOUT

Chapter 2 presents an overview of the empirical signal modelling strategy, including the two subcategories of discrepancy analysis, and waveform reconstruction. A discussion and a summery on the implemented case studies conducted in this thesis is also included.

Chapters 3, 4, and 5 investigate discrepancy signal analysis with regard to condition monitoring of a gearbox.

Chapter 6 investigates neural network waveform reconstruction in the context of a alternator with end-winding damage, while chapters 7 and 8 investigates waveform reconstruction of a severity metric (envelope of the RMS vibration signal) at a constant speed for the monitoring of a haul road.

The conclusion along with suggestions for future research is presented in chapter 9.

Chapter 2

A methodology for cost effective condition monitoring

2.0.1 Overview

In this work it is investigated how diagnostic information may be extracted from vibration signals that have been generated under time-varying operating conditions. In particular it is considered how data driven models of machine responses may be constructed by means of statistical and machine learning techniques in order to use the models to isolate or account for the effects of the fluctuating operating conditions on the vibration signals. This in turn renders it possible to extract diagnostic information that is sensitive to machine faults, while also being more robust under time-varying operating conditions. The results may generally be presented in a simple and intuitive manner.

Two methodologies are proposed in this work, both of which make use of empirical models to account for the effects from fluctuating operating conditions on vibration signals. The first approach is referred to as discrepancy analysis, while the second approach is referred to as waveform reconstruction. Discrepancy analysis is applicable in situations where the operating conditions are unobserved (latent), while the waveform reconstruction methodology requires that the operating conditions are observed.

The discrepancy analysis and waveform reconstruction methodologies are investigated on a number of data sets. These data sets are obtained from simulations, experimental setups, and field measurements. The data sets are used to represent diverse applications; namely the condition monitoring of gearboxes, alternators, and haul roads. Central to all of these applications is the philosophy that it is possible to construct data driven models of the machine response signals. The data driven models do not require manually classified training data or fault histories. For this reason the proposed methodologies may generally be implemented as a cost effective signal processing stage where consistent diagnostic information is extracted from time-varying operations machine response signals.

The diagnostic information may be used to generate an intuitive visual interpretation of the condition of a machine, or to extract high quality features which may be implemented in a change detection algorithm, automated classification algorithm or may be used as events in hazard models. The decision support stages, which depends on failure mode analysis and economic information regarding the cost associated with different outcomes and maintenance methodologies fall outside the scope of this work.

Six case studies are presented in this thesis. The first three are concerned with discrepancy analysis, while the latter three are concerned with waveform reconstruction. An overview of the case studies are presented in table 2.1.

Table 2.1: Overview of case studies

Case study	Primary focus	Technique	Data
1	discrepancy analysis	weighted autoregressive models	gearbox
2	discrepancy analysis	Gaussian mixture model	gearbox
3	discrepancy analysis	autoencoder	gearbox
4	waveform reconstruction	Neural network	alternator
5	waveform reconstruction	Bayesian regression	haul roads
6	waveform reconstruction	Gaussian process	haul roads

2.0.1.1 Discrepancy analysis

Three discrepancy analysis case studies are presented (1-3). Discrepancy analysis is concerned with the comparison of a novel signal to a baseline signal. In order to compare the novel signal to the reference signal an empirical model is first constructed of the healthy signal. The baseline model is representative of the healthy vibration signal as experienced over different operating conditions. Discrepancy analysis is applicable to condition monitoring scenarios where the operating conditions are generally latent (unobserved). This baseline model is used to discern between patterns in a novel signal that are the result of time-varying operating conditions and those which are fault induced. Essentially the methodology allows for the transformation of a vibration signal into a discrepancy signal. The discrepancy transform exhibits low values at time instances where the vibration signal corresponds well to the reference vibration signal. Large discrepancy values are obtained at time instances where the vibration signal deviates from the baseline signal due to fault induced discrepancies. The discrepancy signal is smoother, has a better diagnostic information to noise ratio, and is less sensitive to time-varying operations than the vibration signal. The discrepancy signal may subsequently be interpreted with conventional signal processing techniques, including descriptive statistics, spectral analysis, cepstral analysis and synchronous averaging. Chapter 3 presents the first case study on discrepancy analysis. The approach followed to construct the discrepancy transform is inspired by the residual analysis as proposed by Wang and Wong [38], where an autoregressive (AR) model is used to represent the machine response of the healthy machine, in particular a gearbox. A discrepancy signal (which in this chapter is essentially defined as the residual signal), is obtained as the difference between a novel signal and the one step ahead predictions made by time-varying filter. The time-varying filter is obtained by adjusting the relative contribution from multiple AR models, where each AR model represents a specific range of operating conditions. The adaptive AR model is capable of accounting for latent operating conditions, such as torque fluctuations which often cannot directly be measured.

In case study 2 (chapter 4) it is proposed that discrepancy analysis may be performed by means of Gaussian mixture models (GMM). A GMM is used to represent patterns in signal segments as observed for a healthy gearbox as experienced over latent, time-varying operating conditions. The reference GMM density distribution is subsequently used to measure the extent to which short signal segments from a novel vibration measurement deviate from the healthy behaviour. In this case study it is also investigated how additional measurements (e.g. the shaft rotational speed) may be provided as input to the signal density model in order to render the model more capable of differentiating between time-varying operations and fault induced signal irregularities. Similar to case study 1 it is assumed that the shaft angular position is measured by means of a tachometer, which renders it possible to resample the signal in the order domain.

In case study 3 (chapter 5) an investigation is performed on data from a simulated gearbox. It is assumed that no additional information (e.g. the shaft angular position or the applied load) is available. Discrepancy analysis is again used to obtain a signal transform which is smoother than the vibration signal, while having better diagnostic information to noise ratio. In this section use is made of neural network regression and Hilbert transform enveloping to obtain a residual signal. The spectra of the discrepancy signal are subsequently computed. A method is also proposed for estimating the shaft angular velocity directly from the vibration signal. This renders it possible to approximately resample the discrepancy transform, so that superior diagnostic information may be extracted.

2.0.1.2 Waveform reconstruction

Waveform reconstruction is applicable to machine response signals where the operating conditions are observed. For this approach the data driven model represents the machine response signal as a function of instantaneous state space variables, including the machine state (e.g. angular position, GPS coordinates), and the operating conditions (e.g. speed, load, temperature). With all the causalities in the data adequately represented by the data driven model, it becomes possible to reconstruct the machine response under different conditions. It is for instance possible to reconstruct the machine response at a specified steady state operating condition (e.g. constant shaft angular speed) so that it subsequently becomes possible to apply spectral analysis. The machine response may also be reconstructed at a standardised operating condition so that it may be compared in a consistent manner with other measurements. Waveform reconstruction is investigated in case studies (4-6).

In case study 4 (chapter 6) an artificial neural network is used to model the vibration waveform as measured on an alternator. This renders it possible to reconstruct the vibration signal at specific shaft angular speeds for more consistent comparison. It also makes it possible to extract specific frequency components of interest. Compared to the other two waveform reconstruction case studies this case study has the benefit of having sufficient data, which is mostly noise free. In addition the model is only expected to perform well over a fixed range of input values. This case study thus makes it possible to use a single flexible function (namely a neural network) to accurately represent the waveform.

In case study 5 (chapter 7) an empirical model is used to represent a vehicle response-type road roughness measuring system (RTRRMS) as a function of different road intervals (e.g. GPS coordinates) and of speed. This subsequently renders it possible to reconstruct the machine responses at those different road intervals at a standardised speed in order to compare the RTRRMS values in a consistent manner. In contrast to the alternator case study, this data set is extremely noisy and the data that correspond to specific road intervals are very limited.

In case study 6 (chapter 8) the same data that were used in case study 5 are further explored. In this section a different waveform reconstruction method based on Gaussian processes is proposed for calibrating the vibration response (RTRRMS) from trucks to a consistent (calibrated) speed. The approach proposed in this section avoids the need to discretize the road into different intervals, simplifies the real time implementation of the algorithm, and generally offer better performance for road sections where only a single measurement is available.

Chapter 3

Residual signal and statistical analysis for condition monitoring of a gearbox

3.1 INTRODUCTION

In this chapter a form of discrepancy analysis is investigated. The approach followed to construct the discrepancy transform is inspired by the residual analysis as proposed by Wang and Wong [38], where an autoregressive (AR) model is used to represent the machine response of the healthy machine. A discrepancy signal (which is this chapter will be referred to as the residual signal), is obtained as the difference between a novel signal and the one step ahead predictions made by the AR filter.

The first key idea of this particular discrepancy analysis approach is that multiple AR models are used to represent the healthy machine response, so that each AR model represents a specific range of operating conditions. The multiple models are fused so that effectively a single, but time-varying AR model is obtained which adjusts to different instantaneous operating conditions. Even though the model is time-varying and adjusts according to the operating conditions, it remains constrained to the vibration response of a healthy machine. This renders it possible to obtain a sensitive residual signal which is more robust to fluctuating operating conditions. The adaptive AR model is capable of accounting for latent operating conditions, such as torque fluctuations which often cannot directly be measured.

The second key idea of this chapter is that the structure (magnitude and periodicity) of the residual signal contains valuable diagnostic information. In this chapter the structure of the residual signal is interpreted by realising that the residual signal exhibits second order stationary characteristics (i.e. the standard deviation of the signal varies cyclically). A statistical framework is presented which interprets the second order stationary character of the residual signal. The extracted diagnostic information is presented in a visually intuitive manner.

This chapter is structured as follows: section two investigates the nature of the gear casing waveform, and discusses how the adaptive time series model is used to generate a residual signal. The statistical framework which is subsequently used to interpret the residual signal is considered in detail. Section three investigates the proposed methodology on data which are simulated by means of a dynamic gear model. The final section discusses an experimental case study where the methodology is applied to a gearbox which is run to destruction over an accelerated life cycle.

3.2 THE GEAR CASING WAVEFORM

If it is assumed that a gear fault is present, such as a cracked or broken gear tooth, then it might be expected that the measured vibration waveform from the gear system will be distorted for the duration (or shortly after) that the faulty tooth has been in mesh. Let the measured gear casing waveform y_m at time instant m be described as a combination of the healthy waveform x_m, fault induced distortion f_m and some white measurement noise n_m. It is often most convenient to measure the vibration signal on the casing of the gearbox, but the analysis presented here is not restricted to this definition. It is thus assumed that the measured signal may be expressed as follow:

$$y_m = x_m + f_m + n_m \tag{3.2.1}$$

3.2.1 Modeling the healthy waveform

It is desirable to model the healthy waveform x_m, so that this component may be subtracted from the observed signal y_m in attempt to isolate the fault induced signal distortion f_m. It will now be discussed how the healthy waveform may be modelled by means of an adaptive time series which is based on Bayesian model selection.

Consider the AR process which predicts the one step ahead datum point $\hat{y}_m^{\rightarrow}$ from the p number of past observations $\{y_{m-i}; i = 1...p\}$.

$$\hat{y}_m^{\rightarrow} = \sum_{i=1}^{p} w_i y_{m-i} \tag{3.2.2}$$

The past observations may also be denoted in vector notation so that $\mathbf{y}^{\rightarrow}$ is a column vector of length p. The superscript $\{\rightarrow\}$ denotes that this vector will be used to make one step ahead predictions, in contrast to $\{\leftarrow\}$ which will later be used to denote one step back. The filter coefficients of the all-pole filter $\{w_i; i = 1...p\}$, contained in the column vector $\mathbf{w}$, are referred to as the AR parameters.

The number of parameters in $\mathbf{w}$, which is denoted by p, is referred to as the model order. In vector notation the model is denoted by $\hat{y}_m^{\rightarrow} = \mathbf{w}^{\mathrm{T}}\mathbf{y}^{\rightarrow}$.

Now consider that the AR coefficients $\mathbf{w}$ were trained on vibration data which are representative of the healthy (or normal) behaviour of the gear system. The filter coefficients will in some sense be constrained to the dynamics of the healthy gearbox. It is therefore expected that these coefficients may be applied to novel signals from damaged gearboxes, to obtain a waveform $\hat{y}_m$ which approximates the healthy waveform x_m, $\hat{y}_m \approx x_m$. An error term e_m accounts for the modeling errors which will result in a discrepency in this assumption.

$$x_m = \hat{y}_m^{\rightarrow} + e_m \tag{3.2.3}$$

A residual r_m is now defined as the difference between the observed waveform y_m and the one step ahead model prediction $\hat{y}_m^{\rightarrow}$, so that the residual is defined as:

$$r_m = y_m - \hat{y}_m^{\rightarrow} \tag{3.2.4}$$

Extending the AR model to deal with fluctuating operating conditions

The gear system may be subject to fluctuating operating conditions which may result in significant nonlinear gearbox dynamics. The standard AR model is simple to train and implement, but may not be sufficiently flexible to perform well under severe non linear conditions. For this reason it may desirable to linearize the structural response of the gear system around various operating conditions. This might be done by implementing a regression model which is constructed from a weighted ensemble of AR models, each of which is representative of the healthy gearbox, but for different operating conditions. The different AR models are obtained by training each model on a set of data which was measured under a different operating condition. The contribution from each AR model to the one step ahead prediction is weighted according to the statistical probability that the specific model is in fact the correct model for a specific time instant. This belief in any model is based on the iteratively updated performance of that particular model over a sliding window of recent data points. This ensures that the most relevant model contributes most to the one step ahead predictions.

Suppose that k AR models have been trained on k different data sets so that each model is representative of the healthy gearbox, but for a different loading condition. The column vector which corresponds to the coefficients of j'th model is denoted by $\{\mathbf{w}_j; j = 1...k\}$. A sliding window of q

residuals $\{r_{m-i}; i = 1 \ldots q\}$ are denoted by the column vector $\mathbf{r}_j$, where the subscript j indicates that these residuals were generated by the j'th AR model $\mathbf{w}_j$. The probability that the j'th model is the correct model may now be denoted by the posterior probability $P(\mathbf{w}_j|\mathbf{r}_j)$. The belief in each model at the datum point m is thus conditioned on the sliding window of residuals as predicted by that specific model.

The expected value for $\hat{y}_m^{\rightarrow}$ may now be predicted from the multiple AR models as follows:

$$\hat{y}_m^{\rightarrow} = \sum_{j=1}^{k} P(\mathbf{w}_j|\mathbf{r}_j)\mathbf{w}_j^{\mathrm{T}}\mathbf{y}^{\rightarrow} \tag{3.2.5}$$

By Bayes' law the posterior belief in each model is computed by:

$$P(\mathbf{w}_j|\mathbf{r}_j) = \frac{P(\mathbf{r}_j|\mathbf{w}_j)P(\mathbf{w}_j)}{P(\mathbf{r}_j)} \tag{3.2.6}$$

For simplicity the prior $P(\mathbf{w})$ over all the AR models are assumed to be uniform. The denominator (marginal likelihood) is a normalizing term which ensures that all the probabilities sum to unity.

The residuals are assumed to be independent identically and normally distributed $r_m \overset{\mathrm{iid}}{\sim} N(0, \sigma_r^2)$. The variance of the residuals σ_r^2 is estimated from the original fault free data. The likelihood function may thus be expressed as:

$$P(\mathbf{r}_j|\mathbf{w}_j) = (2\pi\sigma_r^2)^{-\frac{q}{2}} \prod_{i=1}^{q} \exp \frac{-(r_{m-i})^2}{2\sigma_r^2} \tag{3.2.7}$$

The independent Gaussian assumption allows for an efficient iterative updating of the likelihood term.

Minimizing the effect of smearing

The predicted waveform $\hat{y}_m^{\rightarrow}$ is subject to an effect called smearing, which implies that the time dependency of the AR model tends to smear out the signal distortion so that spurious residuals are observed. This effect is indicated in Figure 3.1, where fault induced deviations at time instants 3 and 4 give rise to spurious residuals at times 5 and 6.

The effect of smearing may be minimized by running the time series both forward and backward to obtain two different residual signals, each of which has smearing but in opposite directions. At those positions where a faulty tooth is in mesh, both residual signals will tend to have large values. At those positions where one residual is much larger than the other it will tend to be due to smearing.

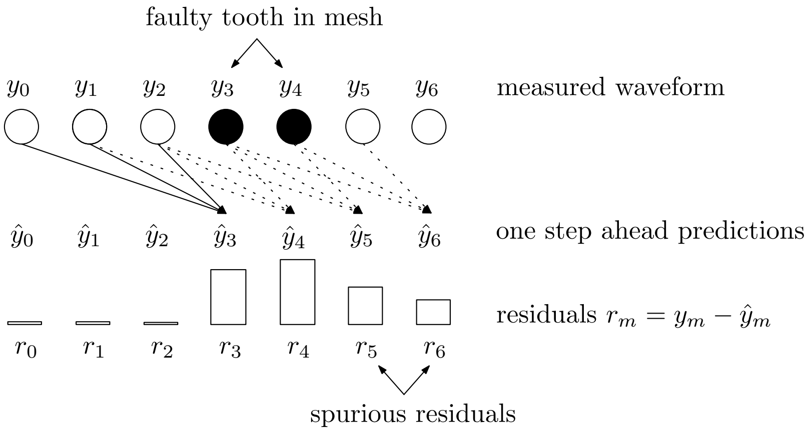


Figure 3.1: Temporal smearing resulting in spurious residuals

Let $\mathbf{y}^{\leftarrow}$ denote the column vector which consists of the p number of future observations $\{y_{m+i}; i = 1...p\}$ which are used to make the one step backward prediction $\hat{y}_m^{\leftarrow}$. The one step backward and one step ahead predictions at time instant m may now be compared:

$$\hat{y}_m^{\rightarrow} = \sum_{j=1}^{k} P(\mathbf{w}_j|\mathbf{r}_j)\mathbf{w}_j^T\mathbf{y}^{\rightarrow} \tag{3.2.8}$$

$$\hat{y}_m^{\leftarrow} = \sum_{j=1}^{k} P(\mathbf{w}_j|\mathbf{r}_j)\mathbf{w}_j^T\mathbf{y}^{\leftarrow} \tag{3.2.9}$$

At each time instance the residual with the smallest amplitude is selected:

$$r_m = \min(|y_m - \hat{y}_m^{\rightarrow}|, |y_m - \hat{y}_m^{\leftarrow}|) \tag{3.2.10}$$

3.2.2 Analysing the residual signal

A sequence of residuals have been estimated. It is now assumed that the modeling error **e** and white noise **n** have independent identical distributed zero mean normal distributions, with unknown but constant variances σ_e^2 and σ_n^2 which are independent of both the operating conditions and the damage in the system.

$$\begin{aligned} e_m &\overset{\text{iid}}{\sim} N(0, \sigma_e^2) \\ n_m &\overset{\text{iid}}{\sim} N(0, \sigma_n^2) \end{aligned} \tag{3.2.11}$$

The fault induced signal distortion f_m is related to the condition of the teeth in mesh at instant m. Let the part of the pinion which is in mesh at instant m be denoted by the pinion angular position α_m^p, and similarly let the part of the gear which is in mesh be denoted by α_m^g. The fault induced signal distortion f_m is thus jointly conditioned on α_m^p and α_m^g so that the conditional distribution is denoted by: $P(f|\alpha^p, \alpha^g)$.

Knowledge of the angular positions α^p and α^g are assumed via some transducer (such as a tachometer) on one of the shafts. The angular position variables can take on any values within the continuous interval $[0, 2\pi]$. To simplify the estimation of $P(f|\alpha^p, \alpha^g)$ the angular positions of both the pinion and the gear will be discretized into N intervals so that each interval has a width of $2\pi/N$. If N is sufficiently large then it may be assumed that the pinion or gear damage is uniform over that section. The N discretized pinion and N discretized gear angular positions are respectively denoted by $\{a_i^p; i = 1...N\}$ and $\{a_j^g; j = 1...N\}$. In vector notation the two $N \times 1$ vectors are denoted by $\mathbf{a}^p$ and $\mathbf{a}^g$. The likelihood of observing a fault induced signal irregularity at time instant m, is now conditional on which discrete pinion interval a_i^p is in mesh with which discrete gear interval a_j^g, so that the likelihood is expressed by $P(f_m|a_i^p, a_j^g)$.

Assume that the damage on each pinion interval $\{a_i^p, 1...N\}$ is characterised by an unkown variance $\{\eta_i^p; i = 1...N\}$, and similarly the damage on each gear interval is $\{a_j^g; j = 1...N\}$ characterised by the unknown variance $\{\eta_j^g; j = 1...N\}$. The unkown variances are collected in the two $N \times 1$ column vectors η^p and η^g.

The likelihood function of the fault induced signal distortion as conditioned on the unknown damage related variances is approximated with a zero mean normal distribution.

The components which contribute to the observed waveform y_m are illustrated in Figure 3.2.

From the definition of the observed waveform $y_m = x_m + f_m + n_m$, the estimate for the healthy waveform $x_m = \hat{y}_m + e_m$ and the definition of the residual $r_m = y_m - \hat{y}_m$ the residual may now be expressed as $r_m = f_m + e_m + n_m$. From the assumption that all the individual components of the residual signal have zero mean normal distributions, the likelihood term for the residual is obtained:

$$P(r_m|\eta_i^p, \eta_j^g, \sigma_e, \sigma_n) = N(0, [\eta_i^p]^2 + [\eta_j^g]^2 + \sigma_e^2 + \sigma_n^2) \quad (3.2.12)$$

Let Θ denote the set of all the unknown variances which must be estimated. The maximum likelihood values for the unknown parameters Θ_{ML} may now be found as the parameter values which optimize $P(\mathbf{r}|\Theta)$. The problem is still ill defined, since the constant contributions from σ_e^2 and σ_n^2 may freely

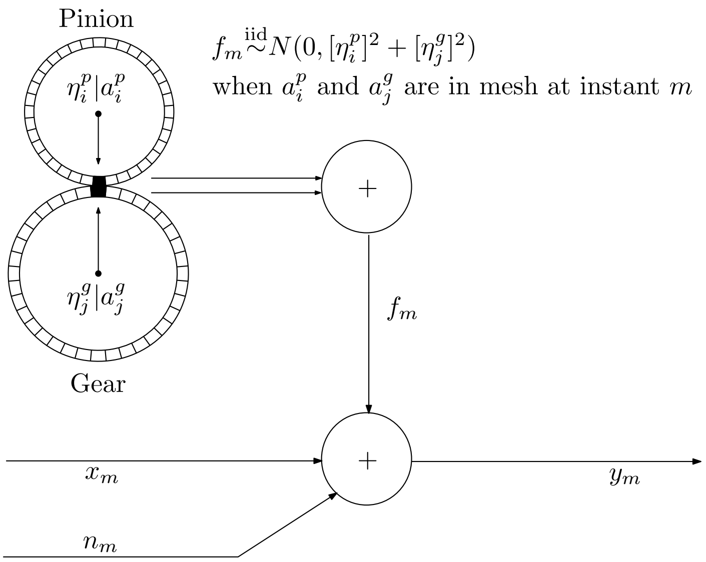


Figure 3.2: Noise process

be distributed between the different parameters.

The white measurement noise variance σ_n^2 and the model error variance σ_e^2 contain no diagnostic information and their cumulative effect may simply be lumped into one constant variance σ_c^2. The residual likelihood $P(r_m|\eta_i^p, \eta_j^g, \sigma_c)$ at each time instant m is now conditioned on three unknown variances. The pinion and gear rotate at different velocities (except when the gear ratio is unity), so that the different pinion angular intervals will mesh with the different gear angular intervals. Let the different combinations in which a_i^p and a_j^g can mesh be denoted by different states s. Let the state $s = i \times j$ denote when the pinion angular interval i is in mesh with the gear angular interval j. It is now possible to lump the variances for each state:

$$[\gamma_{ij}]^2 = [\eta_i^p]^2 + [\eta_j^g]^2 + [\sigma_c]^2 \tag{3.2.13}$$

The residual vector $\mathbf{r}$ can be rewritten as a number of smaller column vectors $\{\mathbf{r}_{ij}; i = 1...N, j = 1...N\}$, so that each subvector $\mathbf{r}_{ij}$ only contains residuals corresponding to that state $s = i \times j$. The maximum likelihood estimate for the variance $[\gamma_{ij}]^2$ which corresponds to each state s is simply:

$$[\gamma_{ij}]^2 = E([\mathbf{r}_{ij}]^T [\mathbf{r}_{ij}]) \tag{3.2.14}$$

One of two approaches may now be followed to estimate the combined noise variance $[\sigma_c]^2$. It may be estimated from the fault free data, in which case it may thus be used to track either general wear or localized defects on the gear system. Alternatively if it is more desirable to accurately monitor localized defects then it may be assumed that there must be at least one undamaged interval on the gear and one undamaged interval on the pinion so that there exists at least one state $s = i \times j$ where $[\eta_i^p]^2 + [\eta_j^g]^2 = 0$. This state will have the smallest residual variance and can then be used as the combined noise variance.

$$[\sigma_c]^2 = \min([\gamma_{ij}]^2; i = 1...N, j = 1...N) \tag{3.2.15}$$

At any state $s = i \times j$ there will now only be two unknown damage related variances $[\eta_i^p]^2$ and $[\eta_j^g]^2$:

$$[\gamma_{ij}]^2 - [\sigma_c]^2 = [\eta_i^p]^2 + [\eta_j^g]^2 \tag{3.2.16}$$

An over determined set of equations are then obtained which express the relationship between the lumped variances $[\gamma_{ij}]^2$ and the unknown fault related variances $[\eta_i^p]^2$ and $[\eta_j^g]^2$. If it is assumed that the certainties in the estimates of each of the state variances $[\gamma_{ij}]^2$ are approximately similar, then each of the fault related variances may be approximated by solving the set of equations in a least squares sense, which may be solved analytically. The fault induced standard deviations η^p and η^g, which indicate the expected magnitude of the fault induced distortion for each pinion and gear interval, may now be used to interpret the condition of the gear system.

3.3 STUDY ON SIMULATED DATA

The proposed methodology is first investigated on a simulated data which are representative of various operating and fault conditions.

3.3.1 Dynamic gear model

Much research has been conducted on the use of dynamic gearbox models to generate vibration signatures which are representative of various types of gear train damage [103, 104, 105]. In this section a very simple dynamic gear model is used to relate tooth faults to the gear casing vibration signature, which is typically where vibration measurements will be taken.

Since the assumed gear model is inherently simple and will only be used to generate vibration data for preliminary investigations it is decided to use a very simple gear mesh stiffness profile. Stander et al. [10] approximated the gear mesh stiffness profile as being sinusoidal, but some results by Jia et al. [106] indicate that it might be more appropriate to approximate it by a square function. The magnitude of the meshing stiffness k_m fluctuates by approximately 20% around the mean value, which is similar to the finite element profile presented by Jia et al.[106]

It is common to model the effects of tooth faults, such as a tooth root crack, by reducing the stiffness of the affected tooth [106]. The stiffness reduction ratio tends to be approximately linearly proportional to the extent of the damage such as the length of a tooth root crack [107]. For sake of brevity reference to a "5% tooth crack", or "5% damage" will hence forth be understood to signify a reduction of 5% in the tooth mesh stiffness of the individual tooth that is affected.

The model that is used to relate the influence of the time varying mesh stiffness on the gear housing is presented in Figure 3.3.

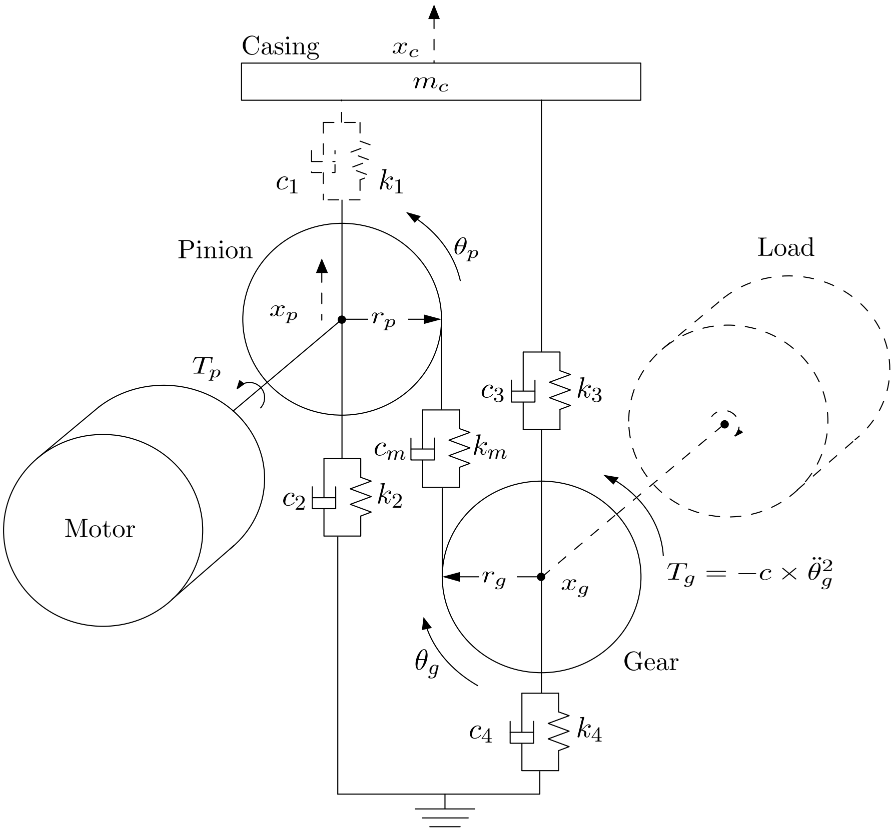


Figure 3.3: Dynamic one stage gear model

The stiffness coefficients k_1, k_2, k_3 and k_4 are assumed to satisfy Hooke's law by being linear in x. Standard viscous damping is assumed for all of the damping elements, with values which ensure that the system is highly underdamped (an actual damping to critical damping ratio of about 8%).

The gear is driven by the pinion, which in turn is subject to the pinion input torque T_p. The load on the gear is proportional to the second power of the gearwheel speed $T_g = K_s\dot{\theta}_2^2$ through the proportionality constant K_s. This allows the system to accelerate until equilibrium is reached. The proportionality constant K_s is selected at 0.4, so that the pinion rotates at about 240 rpm under an applied load of 500 Nm.

The equations of motion are presented in Equations (17) to (21)

$$\begin{aligned} m_c\ddot{x}_c + (c_1+c_3)\dot{x}_c - c_1\dot{x}_p - c_3\dot{x}_g & \\ +(k_1+k_3)x_c - k_1x_p - k_3x_g &= 0 \end{aligned} \tag{3.3.1}$$

$$\begin{aligned} m_p\ddot{x}_p - c_1\dot{x}_c + (c_1+c_2+c_m)\dot{x}_p - c_m\dot{x}_g + r_pc_m\dot{\theta}_p - r_gc_m\dot{\theta}_g & \\ -k_1x_c + (k_1+k_2+k_m)x_p - k_mx_g + r_pk_m\theta_p - r_gk_m\theta_g &= 0 \end{aligned} \tag{3.3.2}$$

$$\begin{aligned} m_g\ddot{x}_g - c_3\dot{x}_c - c_m\dot{x}_p + (c_m+c_3+c_4)\dot{x}_g - r_pc_m\dot{\theta}_p + r_gc_m\dot{\theta}_g & \\ -k_3x_c - k_mx_p + (k_m+k_3+k_4)x_g - r_pk_m\theta_p + r_gk_m\theta_g &= 0 \end{aligned} \tag{3.3.3}$$

$$\begin{aligned} j_p\ddot{\theta}_p + r_pc_m\dot{x}_p - r_pc_m\dot{x}_g + r_p^2c_m\dot{\theta}_p - r_gr_pc_m\dot{\theta}_g & \\ +r_pk_mx_p - r_pk_mx_g + r_p^2k_m\theta_p - r_pr_gk_m\theta_g &= T_p \end{aligned} \tag{3.3.4}$$

$$\begin{aligned} j_g\ddot{\theta}_g - r_gc_m\dot{x}_p + r_gc_m\dot{x}_g - r_gr_pc_m\dot{\theta}_p + r_g^2c_m\dot{\theta}_g & \\ -r_gk_mx_p + r_gk_mx_g - r_gr_pk_m\theta_p + r_g^2k_m\theta_g &= -T_g \end{aligned} \tag{3.3.5}$$

The equations of motion were written into state space form so that they could be solved with the standard ode45 differential equation solver in Matlab.

The model mass, damping and stiffness parameters are presented in Table 1.

Two loading scenarios are investigated, the first of which include a constant loading torque and the second where the torque fluctuates.

For the first scenario the driving load is set to a constant torque of $T_p = 500$ Nm. This simulation investigates how the vibration waveform of the casing changes as one tooth on the pinion regresses from being healthy to suffering from a 10% fault. The severity is incrementally increased by 1% so

Table 1: Model parameters

Parameter	Description	Value	Unit
N_p	Number of pinion teeth	22	
N_g	Number of gear teeth	33	
r_p	Pinion base circle radius	0.1	m
r_g	Gear wheel base circle radius	0.148	m
k_1	Structural stiffness	5×10^5	kN/m
k_2	Structural stiffness	2.5×10^5	kN/m
k_3	Structural stiffness	2.5×10^5	kN/m
k_4	Structural stiffness	5×10^5	kN/m
k_m	Meshing stiffness	$[5 - 0.5 \times \mathrm{sgn}(Np \times \theta_p)] \times 10^5$	kN/m
c_1	Structural damping	100	Ns/m
c_2	Structural damping	100	Ns/m
c_3	Structural damping	100	Ns/m
c_4	Structural damping	100	Ns/m
c_m	Meshing damping	100	Ns/m
m_c	Casing mass	0.6	kg
m_p	Pinion mass	0.6	kg
m_g	Gear mass	1.3	kg
j_p	Pinion moment of inertia	$1/2 \times m_p \times r_p^2$	$kg.m^2$
j_g	Gear moment of inertia	$1/2 \times m_g \times r_g^2$	$kg.m^2$

that 11 samples (from 0 to 10%) are obtained. The vibration waveform is presented in Figure 3.4, for (a) the healthy pinion (0% fault), and (b) where the one tooth on the pinion has a 10% fault. The first second of simulation, where the gear accelerates from rest up to the equilibrium velocity, is discarded.

The power spectral density (PSD) of the 11 vibration waveforms are subsequently presented in Figure 3.5. The pinion has a rotational frequency of 4.14 Hz, while the gear rotates at 2.8 Hz. Significant energy is observed at the resulting gear meshing frequency of about 91 Hz. The gear casing has a natural (undamped) frequency near 300 Hz, while the pinion and gear have natural frequencies around 520 and 684 Hz respectively. The tooth damage, which manifests itself as side bands around the gear meshing frequency, only becomes apparent by taking the log PSD. This is because the energy associated with the induced fault is much smaller than that of the general vibration energy in the signal.

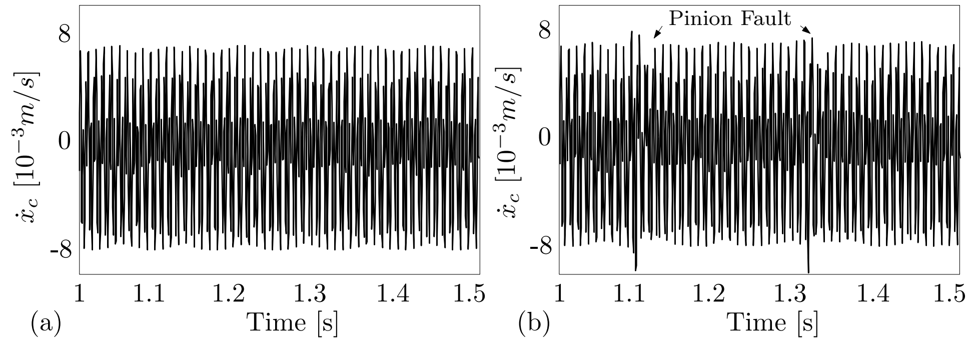


Figure 3.4: Gear casing velocity under steady state conditions (a) healthy, (b) 10% fault on one pinion tooth

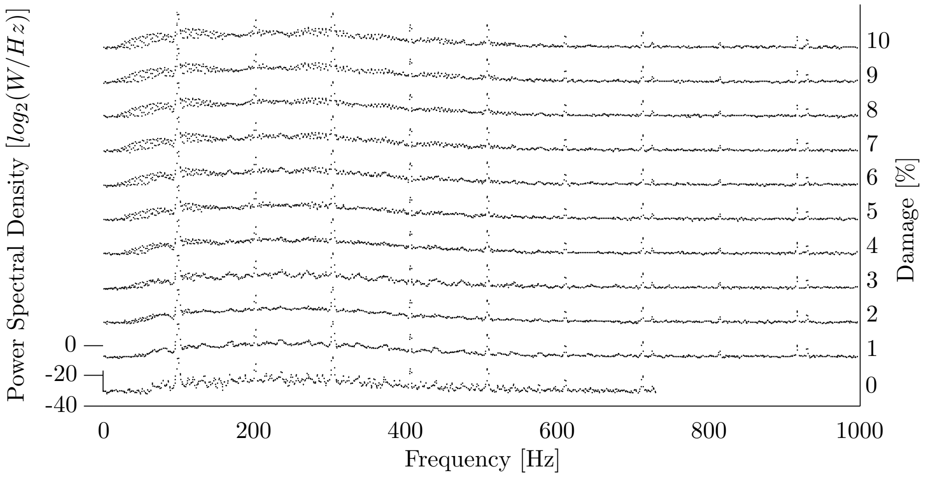


Figure 3.5: Log power spectral density of the constant loading conditions

The second simulation considers a time varying pinion driving load:

$$T_p = 500 \times [\tau + 0.1 \times \sin(10\pi t)] \text{ Nm} \tag{3.3.6}$$

The random variable τ is drawn from the uniform distribution $\tau \in [0.75, 1.25]$ and is sampled independently for each of the 11 damage settings. This simulates how both the speed and load of a gearbox which may be used for instance in a conveyer belt on in a crane may fluctuate significantly. A sinusoidal term in the driving torque accounts for possible rapidly changing loads. The driving load fluctuates at 5 Hz with 20% difference in amplitude around the mean torque of 500 Nm.

The second simulation investigates how the vibration waveform of the casing alters as simultaneously one tooth on the pinion and one tooth on the gear degrade from good health until 10% damage. The damage severity is again incrementally increased by 1% so that 11 samples (from 0 to 10%) are obtained.

In figure 3.6 (a) the waveform is observed where both the pinion and the gear are healthy. In (b) one tooth on the pinion and one tooth on the gear has each been damaged by 10%. Different values for τ cause a difference between the signal amplitudes observed in (a) and (b). The pinion completes slightly more than three rotations during the interval from 1 to 1.5 s, while the gear completes about two. The damage is not prominent in the signal and will not easily be detected with conventional time domain metrics such as crest factor or kurtosis.

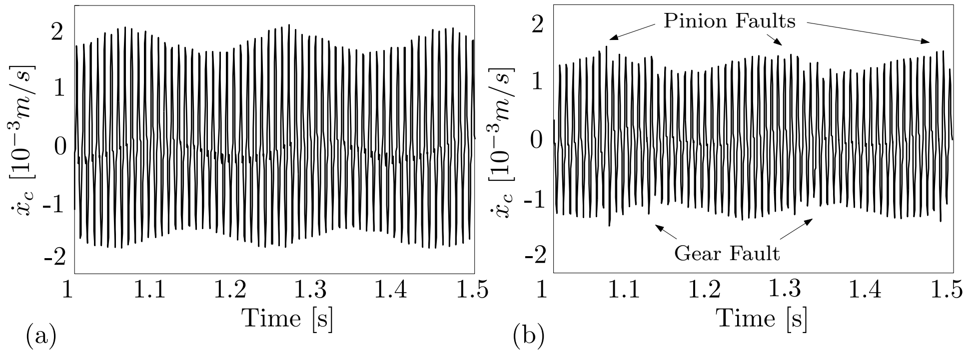


Figure 3.6: Gear casing vibration under steady state conditions (a) healthy, (b) 10% fault on one pinion tooth and on one gear tooth

The log PSDs of the samples as generated under the fluctuating loads are presented in figure 3.7. The sinusoidal variation of the driving load causes the signal energy to be smeared over multiple frequencies. Since τ differs for the various samples both the values of the frequencies and their associated energies differ. The fluctuating nature of the signals renders it very hard to detect the onset of damage.

3.3.2 Implementation

It is required to train AR models which are representative of different loading conditions. To do this a number of data sets are simulated, all of which are representative of the healthy gear system, but for different loading conditions. For each data set a different AR model is created. One of the advantages of AR models is that a closed form solution exists for estimating their coefficients in a least squared error sense (which in this case is equivalent to the maximum likelihood). The reader is referred to the book by Godsill and Rayner [94] for the details concerning different AR coefficient estimation techniques.

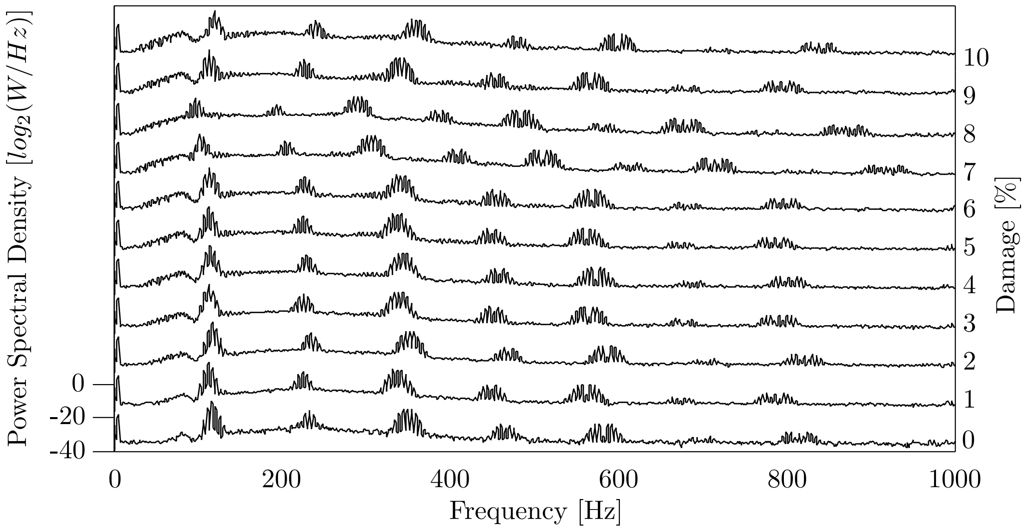


Figure 3.7: PSD Waterfall plot subject to both frequency shifts and spectral smearing due to fluctuating loads and operating speeds rendering it difficult to detect the propagating damage on both the pinion and gear

The dynamic model was sampled at 8.2 kHz. It was decided to down sample the signal to the minimum frequency that satisfies the Nyquist criterion. This frequency ensures that all the diagnostic information in the signal is considered, while the computational cost is minimized. It is deemed that all the harmonics which exceed about 800 Hz contain negligible energy (and hence diagnostic information). The signal is subsequently filtered with a 5th order low pass Butterworth filter which has a cutoff frequency of 1 kHz and a cutoff rate of 30dB/octave to prevent aliasing. The low pass filtered signal is then down sampled to 2 kHz. The AR models for the simulated data were each trained on 8s of data.

Each of the trained models had an order of 100. The model order selection is based on two criteria. A cross validation procedure was used to ensure that the models do not overfit the training data, which might result in poor generalization on novel data. Cross validation was performed by dividing the fault free data which is representative of different loading conditions into two main sets. The first set is referred to as the training data, while the second half is known as the validation data. AR models of different orders were trained on the first data set, while the evaluation data were used to gauge the performance of the trained model on independent data. The model order which results in the best performance on the evaluation data is expected to generalize well. However for the specific application of residual analysis it has here also been found that it is desirable to use a model order which is sufficiently smaller than the number of observations which corresponds to one shaft revolution. This is to ensure that periodic signal irregularities (on which the models have not been trained) do not regress to itself.

Figure 3.8 indicates the performance of AR models with different orders on the independent validation data, for (a) the constant load and (b) the fluctuating loads of the simulated data. At a sampling frequency of 2 kHz and an approximate pinion shaft speed of 4.14 Hz it is known that roughly 480 observations correspond to one revolution of the pinion.

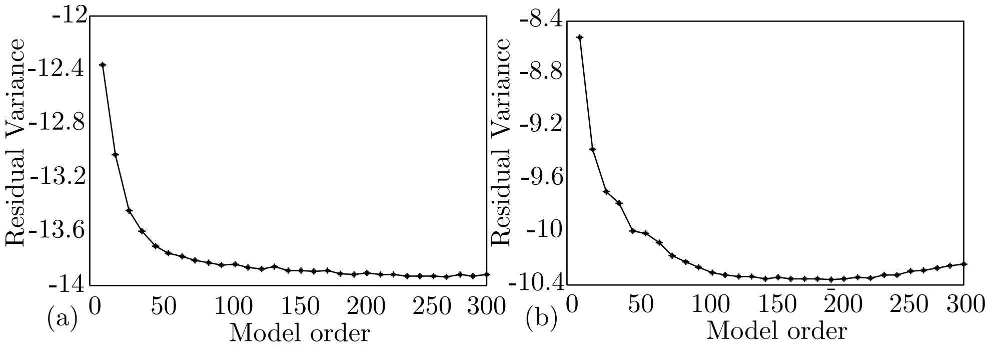


Figure 3.8: AR order selection

From figure 3.8 it seems that the simulated data favour very large model orders. This is because much training data are available so that over-fitting is less likely to occur and also because of the strong autocorrelation in the signal which is implied by its periodicity. However based on the criterion that the model order should be less than the number of observations per pinion revolution and to limit computational costs a model order of 100 is selected.

The number of AR models which must be trained for any gearbox will depend on the nature of the loading conditions and the non linear characteristics of the machine. A cross-validation approach was followed to estimate the number of AR models which were required to effectively model the simulated data. The cross-validation approach may be used to determine whether certain models are redundant or whether some loading scenarios are underrepresented. Since the simulated operating conditions are stochastic, 20 sets of data were sampled so that each data set contained 8 seconds of data which were down sampled to 2 kHz. The first 10 data sets were used to train 10 AR models. The performance of the ensemble of AR models was then evaluated on the remaining data sets by repeatedly ignoring only one AR model. Those AR modes which could be ignored with negligible effect on the model performance were discarded permanently. It was found that all the loading scenarios were well represented by a final ensemble of 6 AR models.

The residual signals of the steady state operating conditions were subsequently computed and used to estimate the lumped state variances $\gamma_{ij}^2; i = 1..N, j = 1..N$. Figure 3.9 illustrates the 1600 lumped variances (presented as a grid of 40x40 cells) for the (a) the fault free data, and (b) the 10% single tooth fault on the pinion. Each block in the matrix represents one state, so that each state represents when one pinion position meshes with one gear position. Dark blocks represent large lumped variances. It is clearly observed in (b) that large residuals are correlated with a specific angular interval on the pinion. This is indicative of a localized fault on the pinion.

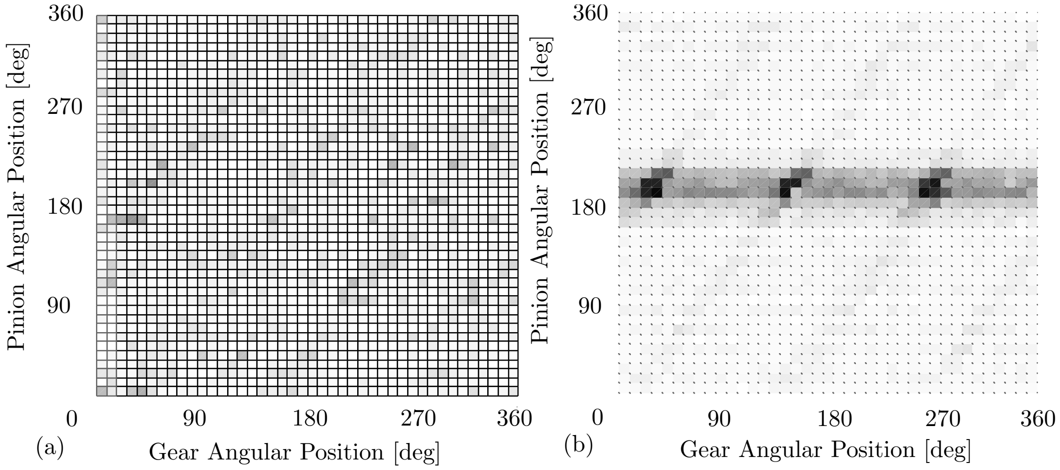


Figure 3.9: The lumped state variances as computed from the residuals (steady state load) associated with each of the 40 × 40 meshing states for (a) no damage and (b) 10% tooth crack on the pinion

Figure 3.10 (a) and (b) indicate the lumped variances for the second simulation (fluctuating loads) for (a) the fault free data, and (b) the two 10% tooth faults (one on both the pinion and the gear). It is clear from (b) that there exist two localized defects, one on each of the pinion and the gear.

3.3.3 Results

The equations which relate the lumped state variances $[\eta^{\mathbf{g}}]^2$ to the fault induced variances $[\eta^{\mathbf{p}}]^2$ and $[\eta^{\mathbf{g}}]^2$ are overdetermined and may be solved in a least squared sense. The fault induced standard deviations are then plotted as a function of angular position and extent of damage. The steady state loading conditions are presented in Figures 3.9 (a) and (b), while the fluctuating loads are presented in Figures 3.10 (a) and (b). The presence of localized faults are clearly observed.

Figures 3.11 (a) and (b) investigate the steady state loading conditions where one crack on the pinion is growing. The figures respectively indicate the standard deviations associated with (a) the gear and (b) the pinion as trended over the 11 samples. In Figure (b) the presence of a localized defect is clearly observed. The magnitude of the standard deviation linearly correlates well with the extent of

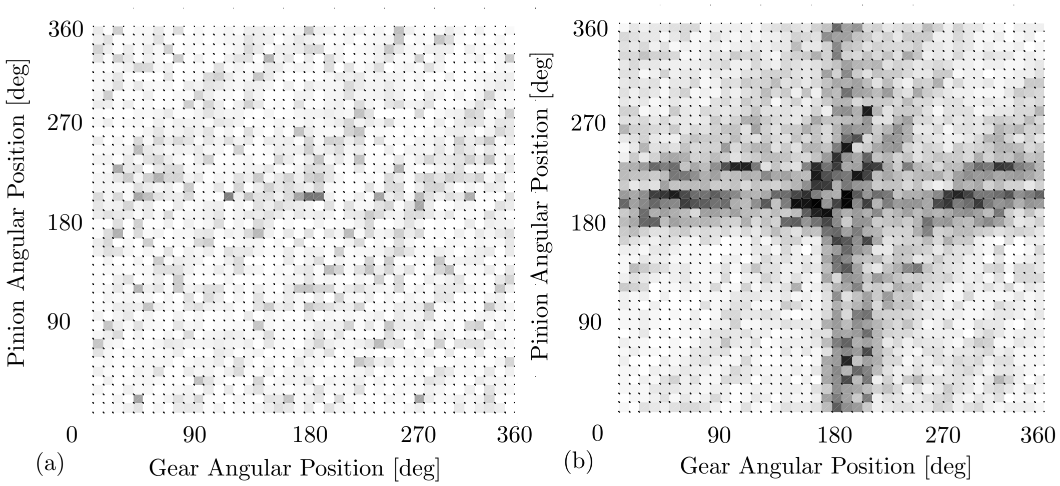


Figure 3.10: The lumped state variances as computed from the residuals (fluctuating loads) associated with each of the 40×40 meshing states for (a) no damage and (b) 10% tooth crack on both the pinion and gear

the damage on the pinion.

Figures 3.12 (a) and (b) respectively illustrate the fault induced standard deviations associated with (a) the gear and (b) the pinion for the fluctuating load conditions. Some noise occurs at positions with no damage, but the propagation of a localised defect on both the pinion and gear remains clear. Not only do the standard deviations grow near linearly over time (despite being generated under different loading conditions), but also do the magnitude between pinion and gear correlate well, confirming that the pinion and gear suffer from similar levels of damage. Given the limited amount of induced damage on the simulated signal these figures do appear to be both sensitive to small levels of damage and robust to fluctuating loads.

3.4 STUDY ON EXPERIMENTAL DATA

The proposed methodology is subsequently investigated on experimental data. The experimental data were generated on an accelerated gear life test that was conducted by Rudi Kroch in the Sasol Laboratory for Structural Mechanics at the University of Pretoria.

3.4.1 Test setup

A single stage Flender E20A gearbox (rated load of 20 Nm) with helical gears and a step up ratio of 1:1.96 was run until destruction on a mounted test bed. The gear life cycle was accelerated by connecting the monitored gearbox between two Flender E60A gearboxes, the first of which served to

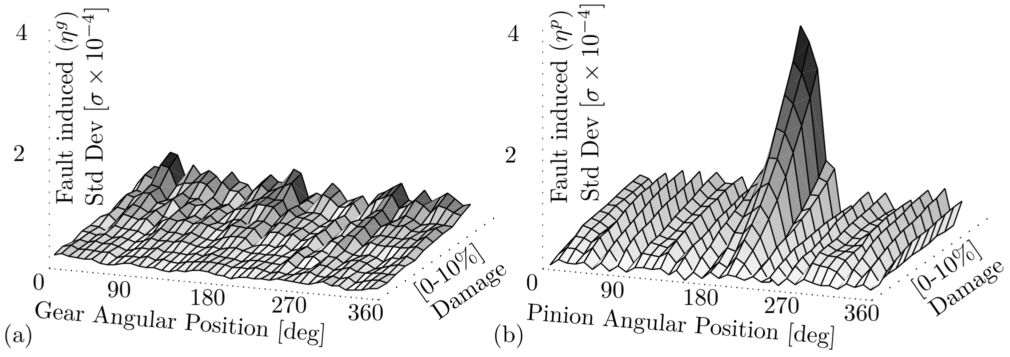


Figure 3.11: Estimated fault induced standard deviations (estimated under constant loads) for different (a) gear and (b) pinion angular positions as trended over different levels of damage

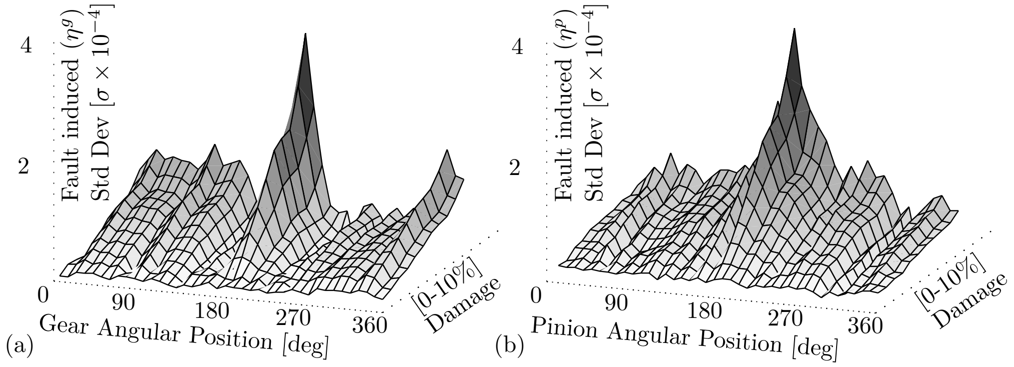


Figure 3.12: Estimated standard deviation (estimated under fluctuating loads) for different (a) gear and (b) pinion angular positions as trended over the life cycle of pinion

step down the shaft speed and the second to step the speed up. The E60A gearbox pair increased the torque over the E20A test gearbox by a ratio of 4.72. The face of the pinion in the test gearbox was also milled so that it had a tooth width of 4mm, which was slightly less than that of the gear. The step down gearbox was driven by a 5 kW three phase four pole WEG squirrel cage electric motor which maintained a constant driving speed of 1472 RPM. The step up gearbox was connected to a flywheel which in turn was coupled with a 5.5 kVA Mecc Alte Spa three phase alternator to apply the load. The vibration signature of the test gearbox was measured by means of a Polytec PDV-100 laser vibrometer, which measured the velocity, rather than the acceleration, of the gear casing. The readings from the laser vibrometer were passed through a junction box before being recorded at a sample rate of 50 kHz on the Somat eDAC-lite data acquisitioning system. A low pass filter at 22 kHz prevented aliasing. A proximity probe was mounted near the output shaft from the test gearbox. The proximity probe served as a tachometer by measuring the shaft speed with reference to a once

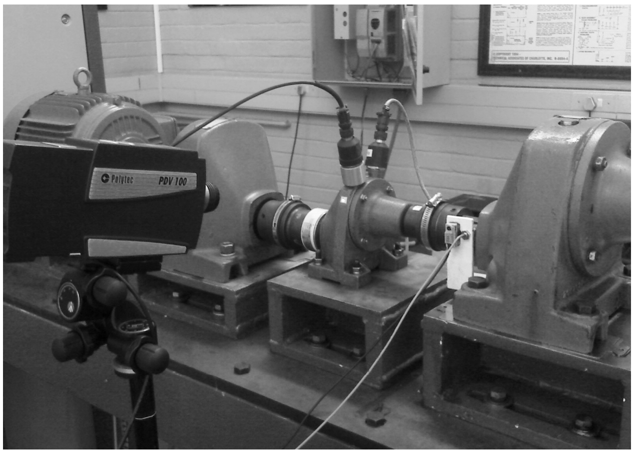


Figure 3.13: Test setup

per revolution pulse. The reading from the proximity probe was synchronized with that of the laser vibrometer on the eDAQ and was subsequently downloaded to a desktop computer. The test gearbox was subjected to a two hours long run in period after which the gearbox was monitored over a period of 20 hours by means of hourly measurements. Each measurement lasted for a period of 30 seconds (15×10^5 data points). The test gearbox suffered total failure during the 17'th hour so that by the time that the final measurement at 18 h was taken the pinion teeth had sheared off completely. This caused the gear to come to rest. The final measurement is subsequently not presented on the graphs. The schematic layout of the experimental set up is illustrated in Figure 3.14.

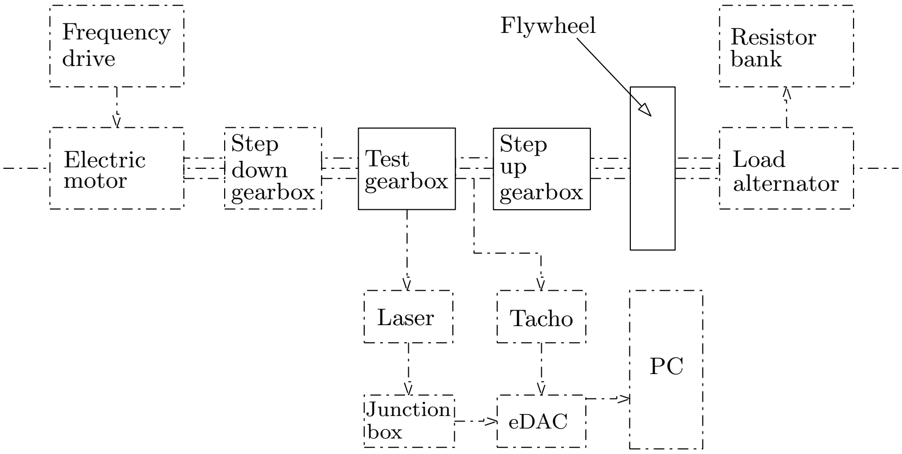


Figure 3.14: Test setup schematic layout

The fundamental frequencies that are expected in the waterfall plot, as computed from the gear ratios and from the tachometer readings, are presented in Table 2.

Table 2: Fundamental frequencies

Gearbox	Function	Shaft input	Teeth ratio	Shaft output	Mesh Freq
Flender E60A	Speed step down	24.5 Hz	18/85	5.19 Hz	441 Hz
Flender E20A	Test gearbox	5.19 Hz	43/22	10.12 Hz	223 Hz
Flender E60A	Speed step up	10.14 Hz	85/18	47.88 Hz	862 Hz

The first sample is taken to be representative of the healthy gearbox. Figures 3.15 (a) and (b) respectively represent the original gearbox response and also the final response at 17 h, just before total failure occured. Figure 3.15 (b) contains strong impulsive behaviour with a periodicity of 0.19s, which corresponds to the rotation frequency of the input (pinion) in the test gearbox. This is clearly indicative of a severely cracked or broken tooth on the pinion.

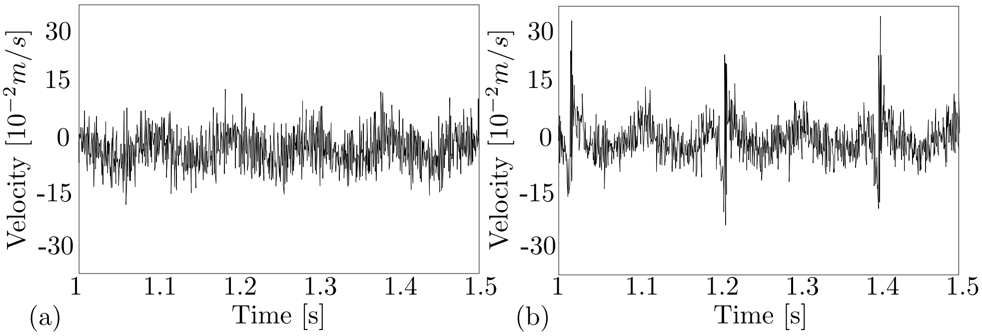


Figure 3.15: Gear casing velocity of (a) run in condition and (b) severely damaged condition

The power spectral density waterfall plot of the gear life cycle is shown in Figure 3.16. Owing to the constant loading conditions neither spectral smearing nor any amplitude modulation occurs. The spectrum of the healthy gearbox (at 0h) indicates energy at 10.14 Hz and 24.5 Hz, which correspond to the drive shaft frequencies for the E60 step up and step down gearboxes. The energy at 100 Hz probably corresponds to twice the AC line frequency, which tends to be more prominent in motors and alternators than the actual AC frequency component at 50Hz. A peak, similar to the energy observed at 100 Hz, is observed at 80 Hz. Two large components are identified at 200 Hz and 223 Hz. The 223 Hz corresponds to the test gearbox tooth meshing frequency, while the 200 Hz component corresponds with the 4th harmonic of the AC line current (50 Hz). The step down gearbox has a meshing frequency at 441 Hz and the step up gearbox has one at 862 Hz. Frequency modulation of the speed step up meshing frequency (441 Hz) and the test gearbox meshing frequency (223 Hz) may account for the component observed at near 664 Hz. A very large energy component is also observed at approximately 540 Hz. Its presence is not easily explained, but it might be caused by similar mechanisms to those of the 200 Hz component.

Upon inspection, after the test had been completed, both the step up and step down gearboxes were found to be in good condition. Since constant loading conditions were maintained it is reasonable to expect that all deviations in the system are related to the degradation of the test gearbox. The frequency component that corresponds to the test gearbox meshing frequency (223 Hz) shows a steady increase over the life of the gear. At around 13h energy suddenly appears in the frequency range 5 to 50 Hz, which is associated with the test gearbox shaft rotational speeds and their harmonics. Inspection of the time domain waveform indicates strong impulsive behaviour which corresponds to the rotational frequency (5.19 Hz) of the test gearbox input shaft. This is probably due to the presence of a severe pinion tooth fault. After the 17h the gearbox suffers from complete failure, with a number of gear teeth shearing off completely.

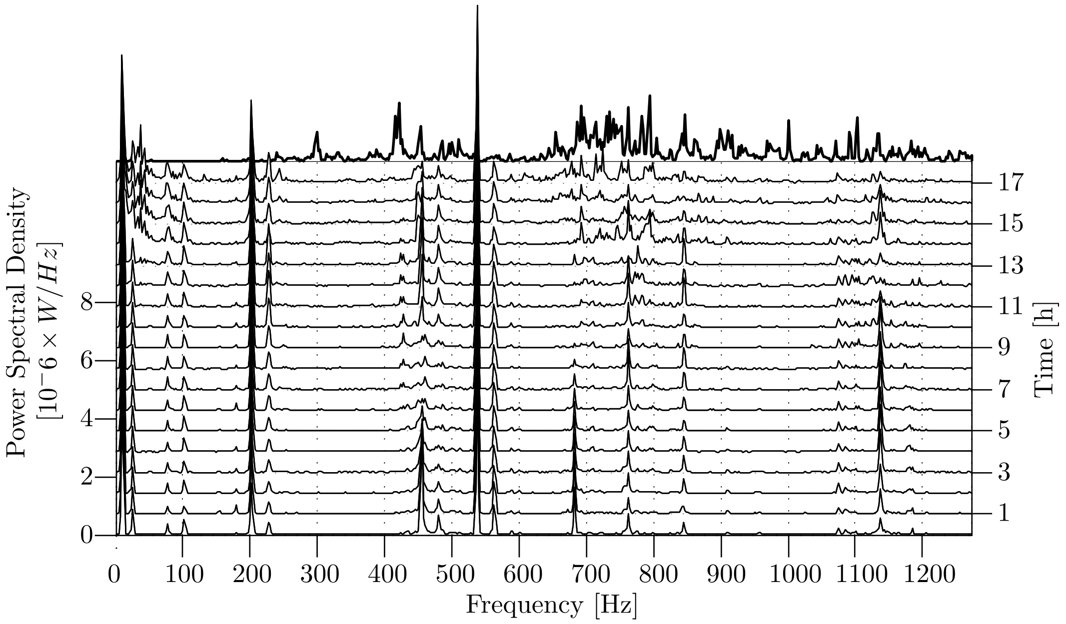


Figure 3.16: PSD Waterfall plot

3.4.2 Implementation

An inspection of the spectrogram indicates that the important frequency content falls below 1 kHz. A 5th order Butterworth filter with a low pass frequency of 1 kHz and a cutoff rate of 30 dB/octave is applied to the signal. The time series is down sampled with a ratio of 20:1 so that a new signal with sampling frequency of 2 kHz is obtained. The output shaft frequency of 10.12 Hz dictates that the AR model order should be less than 200 (2kHz/10) to prevent faults from regressing to itself. An AR model of order 150 is selected and subsequently trained on the assumed fault free data that were measured at 0 time. The performance of this AR model is evaluated on the data measured after one hour and good generalization is observed.

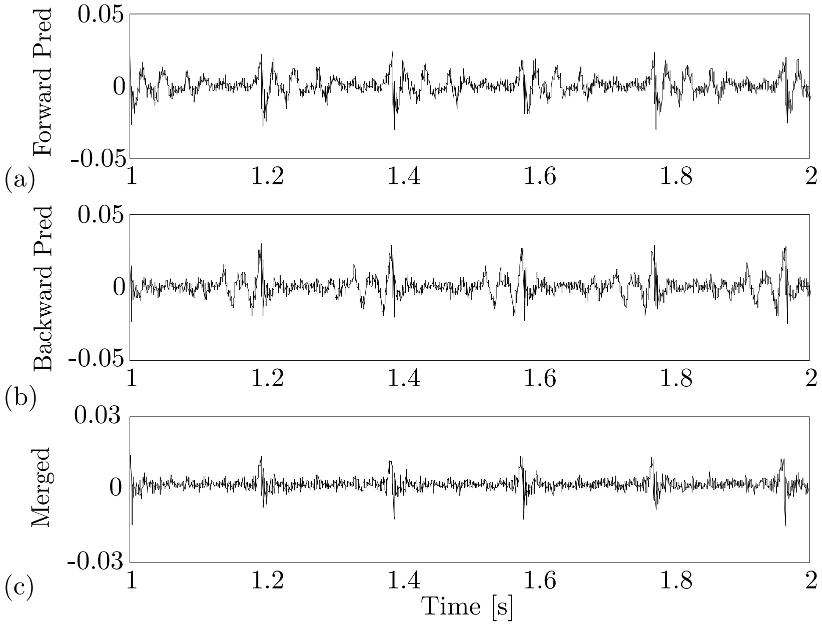


Figure 3.17: Residuals based on (a) forward and, (b) backward filtering, and (c) is obtained as the pointwise smallest. (c) is smearing free so that the signal irregularity is well localized

The gear angular position is estimated from the proximity probe which is synchronized with the laser vibrometer measurement. The proximity probe is mounted near the output shaft from the test gearbox so that a once per revolution pulse is produced. The intermediate gear angular positions, which correspond to the time vector of the vibration data, are estimated by means of piecewise cubic interpolation. The pinion angular positions are subsequently estimated from the known relationship between the gear and pinion. During each sample it is assumed that both the gear and pinion are at their zeroth positions at the instance that the first pulse is observed. This assumption allows for a consistant estimation of the gear angular position, which is always correlated with the proximity probe. However, the assumed zeroth position of the pinion will not be consistent over different samples. It will therefore be necessary to align the fault estimations from different samples based on their autocorrelations.

A subsample of six seconds of waveform data (12×10^3 data points) are used to estimate the lumped state variances. Six seconds are enough to ensure that the pinion-gear combination rotate through all the different meshing states. Effectively all the residuals are allocated to bins depending on which position of the pinion was in mesh with which position of the gear at the respective time instances.

The total variance is then computed for each bin based on all its corresponding residuals. The lumped variances are illustrated in Figures 3.18 (a) and (b) for respectively the healthy and the severely damaged gear system.

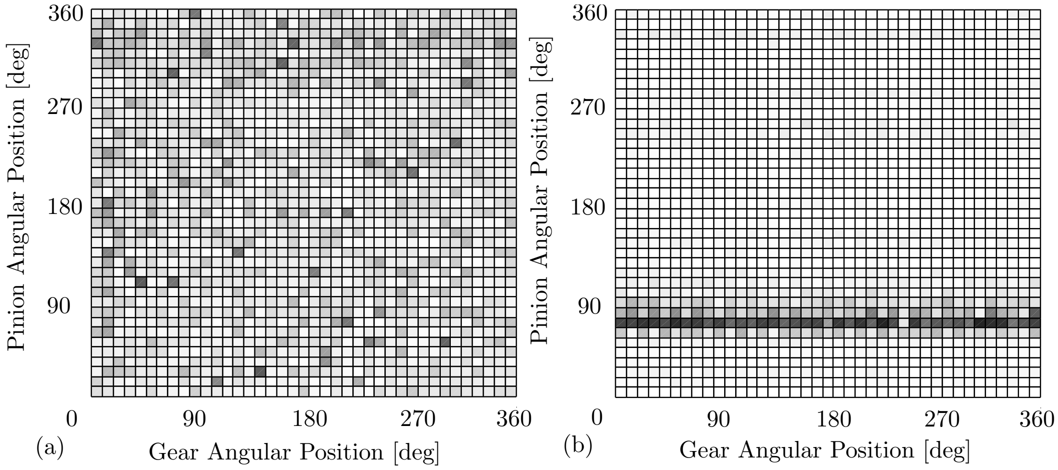


Figure 3.18: The lumped state variances associated with each of the 40 × 40 meshing states for (a) no damage (first sample) and (b) shortly before failure occurs (17th sample). A clear irregularity associated with a specific pinion angular position at approximately 80 degrees is observed, indicating the possible presence of a tooth crack.

Figure 3.18 (a) is representative of the healthy gearbox and illustrates the lack of any correlation between the noise and any specific pinion or gear angular positions. Figure 3.18 (b), which is representative of the damaged pinion, has a dark band at about 80 degrees on the pinion which indicates that the pinion is suffering from a significant localized defect such as a broken tooth.

3.4.3 Results

The lumped state variances are next used to estimate the fault induced variances. It is now possible to trend the fault induced standard deviations as a function of angular position and time. The results are displayed in Figures 3.4.3 (a) and (b) which respectively corresponds to the gear and pinion.

The final figures clearly indicate how, beginning at about the 8th hour, a signal irregularity which is strongly correlated with a specific pinion angular position begins to grow at a rapid rate until it reaches a peak at 17 h just before complete failure.

The fault induced standard deviations clearly indicate that it is the pinion and not the gear which is breaking. Not only is it possible to see evidence for damage much earlier than in the frequency domain, but it is now also evident that a localized section (such as a single tooth) on the pinion is

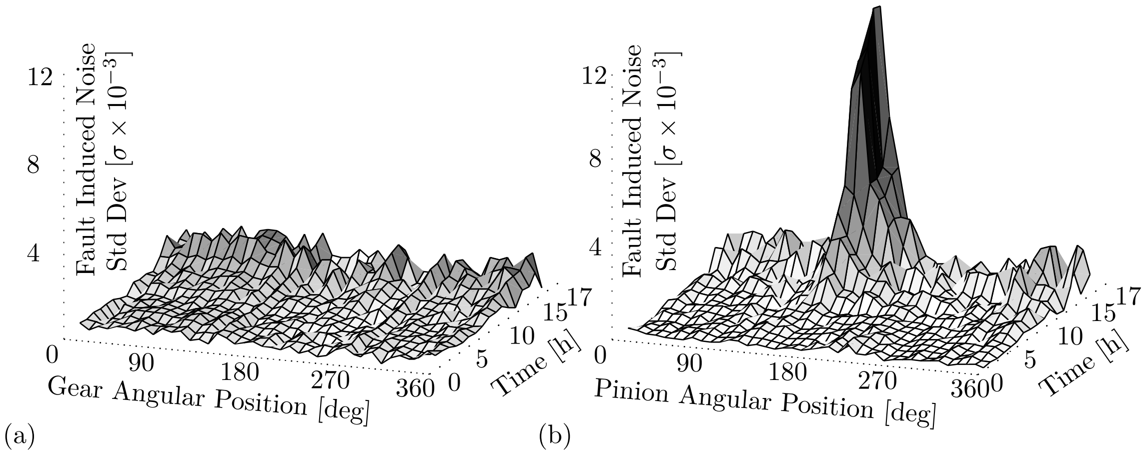


Figure 3.19: Estimated standard deviation for different (a) gear and (b) pinion angular positions as trended over the life cycle of pinion

suffering severe damage. This knowledge could be used to order the correct part in advance and also guide the maintenance procedure to only disassemble the required component.

3.5 CHAPTER CONCLUSION

This chapter proposed a methodology which can be used to remove the regular components from a vibration signal so that a residual (discrepancy) signal which is sensitive to the gear faults and robust to the operating conditions can be obtained. The residual signal was generated as the difference between the output from a time series regression model and the observed waveform. The regression model, which is comprised of the weighted average from an ensemble of AR models, is iteratively updated based on a Bayesian model selection framework. The time series regression was run both forwards and backwards to obtain two residual signals. The two signals were subsequently merged into a final residual signal by point-wise selecting the smaller of the two. The final residual signal was found much more capable of localising fault induced signal irregularities.

A statistical framework was also proposed which could be used to interpret the second order cyclostationary structure of the residual signal. This subsequently made it possible to localize the fault induced signal deviations, which in turn offered insight into the nature of the gear damage.

The proposed strategy was applied to two sets of simulated data. A simple dynamic gear model was used to approximately simulate the structural response of the gear casing under various loading and gear health conditions. Data were also obtained from an accelerated gear life test where a single stage gearbox was run until destruction under constant speed and loading. The results provided an intuitive understanding of the nature of the gear damage. From the simulation data it appears that the proposed methodology is not only sensitive to gear damage, but may also be used under significantly fluctuating operating conditions.

At this stage it may also be mentioned that the computed diagnostic metrics (lumped state variances) which serve as indication of the condition of the gears, may in themselves also be modelled as long term time series, e.g. as functions of days or weeks. This is not only true for the lumped state variances, but also for any other diagnostic metrics. This in turn renders it possible to better estimate the presence of long term tendencies (e.g. a gradual increase) in the diagnostic metric. This topic, along with optimal detection techniques, and may be combined with additional failure mode analysis, hazard models, and economic information (e.g. cost of false negatives and positives) in order to develop maintenance decision support strategies. These interesting and significant topics fall outside the scope of this research.

Chapter 4

Synchronous averaged Gaussian mixture model discrepancies

4.1 INTRODUCTION

In this section the author proposes a discrepancy analysis based gear monitoring technique, which combines ideas from residual analysis, novelty detection and synchronous averaging. A Gaussian mixture model (GMM) is used to represent patterns in signal segments as observed for a healthy gearbox as experienced over latent, time-varying operating conditions. The reference GMM density distribution is subsequently used to measure the extent to which short signal segments from a novel vibration measurement deviate from the healthy behaviour. A discrepancy signal is obtained, of which the magnitude and the periodicity (first order) are analysed for diagnostic information.

The proposed framework is illustrated on a simple analytical example which approximately represents vibration response signals from a gearbox which is subject to significant amplitude and phase modulation. The methodology is subsequently investigated on data which were generated on an experimental gear test rig. The test rig was subject to time-varying operating conditions. It is seen that GMM density distributions are potentially capable of detecting fault induced signal deviations in signals which are subject to fluctuating operating conditions. On the investigated data sets the negative log likelihood (NLL) synchronous averaging methodology outperforms conventional vibration signal synchronous averaging and spectral analysis.

It is occasionally possible to obtain measurements of some of the fluctuating operating conditions. This chapter indicates how those measurements may then serve as inputs to the GMM in order to better differentiate between changes caused by the fluctuating operating conditions and those caused by faults. It is however also seen (at least in the presented case studies) that good results may also be obtained even when the operating conditions are not directly provided as inputs to the GMM. In those scenarios the GMM is still capable of modelling the unobserved (latent) operating conditions. This is a valuable result, because it is not always possible or cost effective to measure all the parameters which are related to the operating conditions (e.g. applied load, angular speed, temperature, etc), so that it is almost always inevitable that some operating conditions will be latent. It is also occasionally desirable to avoid using additional transducers which may break and subsequently lead to false alarms.

4.2 METHODOLOGY

This chapter proposes a discrepancy analysis methodology which combines concepts from residual analysis, novelty detection and synchronous averaging. The methodology comprise four key steps as illustrated in figure 4.1.

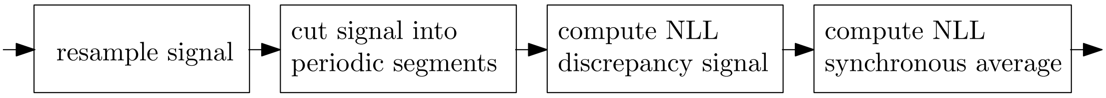


Figure 4.1: Diagram of the four key steps of the proposed NLL discrepancy signal synchronous averaging methodology.

1. The vibration signal is resampled at constant gear angular increments.
2. The resampled signal is segmented into block windows of fixed length. The signal segments are synchronous with specific shaft angular intervals. A window length may be chosen for instance to represent the (approximate) duration for which a gear-pinion tooth pair is in mesh.
3. The negative log likelihood (NLL) of each signal segment is computed. The NLL discrepancy signal is used as a measure of the extent to which individual signal segments deviate from the expected healthy behaviour.
4. All the NLL values which correspond to any particular gear (or pinion) angular interval are ensemble averaged. Gear angular intervals with small averaged NLL values correspond well to the expected behaviour, while large averaged NLL values indicate the possible presence of damage.

Novelty detection with Gaussian Mixture Models

The discrepancy analysis methodology implemented in this chapter makes use of Gaussian mixture model (GMM) density distributions to compute the negative log likelihoods (NLL) of short signal segments. A linear combination of Gaussians can give rise to complex multivariate probability densities.

Almost any continuous density function may be approximated to arbitrary precision by implementing sufficiently many Gaussian components with appropriately selected means, covariances and mixing coefficients [90, 91]. GMM density estimates may thus be well suited to represent the possibly non-linear covariances or multiple density modes which may arise in signal sequences which have been generated under time-varying operating conditions. Individual mixture components may represent various, unobserved (latent) operating conditions. GMM models are fairly simple to implement, given that sufficient data are available for optimisation by means of expectation maximisation, and given that the length of the vector of random variables (length of the signal segment) does not give rise to very large dimensionalities. A probability density distributions may be used to generate a metric of similarity based on the natural logarithm of the likelihood of the observed sample [108] [1]. In this chapter the negative natural logarithm of the likelihood (NLL) is used as a measure of discrepancy.

Dividing the signal into segments

The time domain vibration signal is resampled at constant increments of the shaft angular position. Signal resampling is implemented with piecewise-cubic interpolation and is based on a one pulse per shaft rotation tachometer reference signal. The resampled signal is subsequently cut into short signal segments by means of block windows.

The window length and the extent to which the windows overlap influence the resolution and sensitivity of the resulting NLL discrepancy signal. Longer window lengths offer more contextual information, however a long window length also increases the dimensionality of the reference density distribution. High dimensional density distributions as estimated with GMMs are more subject to local minima during optimisation and also to data overfitting, which may result in poor generalisation [90]. Long windows also result in increased smearing (reduced temporal resolution) of the NLL discrepancy signal.

Consider that the signal segments may be generated in a manner similar to the sliding window approach implemented by a conventional auto-regressive (AR) filter in time-series analysis. Signal segments are created by moving the window one datum point at a time. The AR model does not differentiate between vibration signal segments which are generated at different shaft angular intervals. The output from the AR model is thus completely based on the contextual information which is

---

[1] Please note that while probability and likelihood are related they are not interchangeable concepts. A likelihood value is obtained from a density distribution function when it is evaluated for a specific realisation of a continuous variable (or vector of variables). A probability is obtained by integrating a density distribution over a specified range of values which the random variable might take on. Probability is always limited to the interval [0-1], while likelihood is not.

contained in the previous number of signal datum points.

The NLL discrepancy signal implemented in this chapter was initially investigated using a similar sliding window approach to that employed in conventional time-series analysis. While this approach was seen to work reasonably well, it was later realised that a more specific density distribution could be constructed if the signal segments are generated in such a manner that it inherently reflects knowledge of the structure of the signal.

In this chapter it is assumed that only a single tachometer signal is available. If the tachometer is mounted on the gear shaft, then the gear angular position may be estimated in a consistent manner. Since the pinion generally rotates at a different frequency than the gear, it will not be known at the start of new measurements which pinion tooth is in mesh. If tachometers are mounted on both the pinion and the gear, or if it is only required to monitor the gear, then unique density models may be constructed for individual gear and pinion angular intervals.

In this chapter the signal segments are generated such that they correspond to the periodicity of the gear meshing frequency. The number of windows per rotation of the gear is set equal to the number of teeth on the gear. The windows do not overlap, but immediately succeed each other. The gear mesh frequency often significantly contributes to (or dominates) the overall energy in the vibration signal. Signal segments with periodicities which correspond to the gear mesh frequency are generally much more similar, compared to signal segments selected at random. These segments may subsequently be used to obtain a density distribution which is subject to less variance, and which may subsequently result in more sensitive NLL discrepancy values. Figure 4.2 illustrates the implemented windowing methodology.

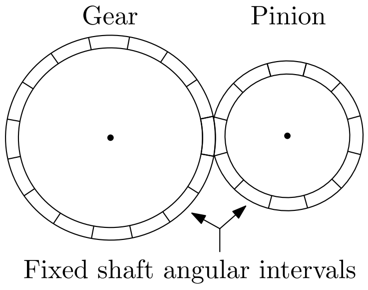


Figure 4.2: Representation of the fixed window intervals which correspond to approximate pinion-gear meshing periods.

Let the angular domain resampled vibration signal be denoted $\mathbf{p} = [p_1\ p_2\ ...\ p_N\ ...\ p_{rN}]^T$ where N is the number of resampled datum points per rotation of the shaft on which the tachometer is mounted

The vibration signal is truncated so that all the data before the first and after the last tacho pulses are discarded. Let r represent the integer number of complete shaft rotations during the measurement, so that rN is the length of the resampled vibration signal. Let T_g represent the number of teeth on the gear, and let T_p represent the number of teeth on the pinion.

If the tachometer is mounted on the gear shaft, then the window length L_w is obtained as the number of datum points per rotation of the gear, divided by the number of gear teeth $L_w = N/T_g$. Similarly if the tachometer is mounted on the pinion shaft then the window length is obtained as $L_w = N/T_p$. Let the first signal segment be represented by the column vector $\mathbf{s}_1 = [p_1 \ p_2 \ ... \ p_{L_w}]^T$ and the second segment by $\mathbf{s}_2 = [p_{L_w+1} \ p_{L_w+2} \ ... \ p_{2L_w}]^T$. This is repeated until rN/L_w signal segments have been obtained. The signal segments are collected in the matrix X_s, where $X_s = [\mathbf{s}_1 \ \mathbf{s}_2 \ ...\mathbf{s}_{rN/L_w}]$. Each column in X_s represents a pinion-gear meshing period.

A measure of discrepancy (NLL value) will be computed for each column in X_s. The resulting NLL discrepancy signal will have a much lower sampling rate than the original vibration signal. The lower sampling rate reduces the computational burden of the algorithm, without discarding diagnostic information. The NLL sampling frequency corresponds to one value for each pinion-gear mesh. The NLL discrepancy signal may thus be synchronous averaged with respect to either the gear or the pinion without the need for additional resampling.

Density estimation with Gaussian mixture models

A Gaussian mixture model (GMM) is optimised to represent the density distribution which corresponds to signal segments from the healthy gearbox. Alternative non-parametric methods, such as density estimation with Parzen windows, may also be considered. However, non-parametric approaches tend to be both memory and computationally intensive, while also being more prone to data overfitting, especially for high dimensional density distributions [91].

A GMM comprises a weighted super positioning of individual multivariate Gaussian components [90]. The complexity of a GMM may be controlled by limiting the number of mixture components, or by constraining the covariance matrices. When sufficient training data are available GMMs may be used to estimate very complex density distributions [90, 91]. Mixture models are for instance well suited to represent data which might be generated by different classes. The classes may be unknown and thus treated as latent variables [91]. A density distribution may thus represent vibration signal segments which have been generated under different operating conditions, but of which the operating conditions is unobserved (latent). A mixture model with K components is denoted as [90]:

$$p(\mathbf{g}) = \sum_{k=1}^{K} \pi_k N(\mathbf{g}|\mu_k, \Sigma_k), \tag{4.2.1}$$

where the mixing coefficients π_k must sum to 1, $\sum_{k=1}^{K} \pi_k = 1$. Given that $N(\mathbf{g}|\mu_k, \Sigma_k) \geq 0$ a sufficient condition that $p(\mathbf{g}) \geq 0$ is that $\pi_k \geq 0$ for all k. If each of the individual components are valid, and the aforementioned constraints are satisfied, then the resulting mixture model will also be a valid density function which integrates to 1 [90].

The model parameters are optimised in Matlab using expectation maximisation (EM) so as to maximise the likelihood of the observed data. A good reference on the topic of expectation maximisation is provided by Bishop [90]. The EM algorithm is initialised by centring the GMM components on randomly selected samples from the training set. The EM algorithm is not guaranteed to find the global optimum, so that it is expected that the performance from different optimisation runs will differ slightly. The EM optimisation step is thus repeated a number of times, and the solution which offers the best results is subsequently selected.

Conditional density distributions

Occasionally the operating conditions (or at least covariates which partially describe the operating conditions) may be measured. It may then be possible to use the knowledge of the operating conditions to refine the computed NLL values. This may be done by conditioning the outcome of the observed signal segment on the measured operating conditions. This should render the computed NLL values more robust to time-varying operating conditions, and subsequently more sensitive to fault induced signal outliers.

Consider for example that the shaft angular speed (as possibly estimated from the tachometer) may occasionally be proportional to the applied torque on a gearbox. The angular speed may subsequently co-vary (possibly in a non-linear manner) with the amplitude or phase modulation of a signal segment.

As previously defined the column vector $\mathbf{s}_i$ is used to represent a single signal segment. Let $\mathbf{c}_i$ represent the column vector which contains variables which are representative of the observed operating conditions which corresponds to the period during which the segment $\mathbf{s}_i$ is measured. The operating conditions vector $\mathbf{c}_i$ may for example contain one entry which represents the mean angular velocity and a different one which describes the applied load.

The likelihood of a signal segment $\mathbf{s}_i$ as conditioned on the vector describing the associated operating conditions $\mathbf{c}_i$, is equal to the joint likelihood of both $\mathbf{s}_i$ and $\mathbf{c}_i$, divided by the likelihood of $\mathbf{c}_i$, so that $P(\mathbf{s}_i|\mathbf{c}_i) = P(\mathbf{s}_i, \mathbf{c}_i)/P(\mathbf{c}_i)$ [90].

A GMM may first be used to estimate a joint density distribution for both the signal segments and the operating conditions as contained in the matrix X_{sc},

$$\mathbf{X_{sc}} = \begin{bmatrix} \mathbf{s}_1 & \mathbf{s}_2 & \dots & \mathbf{s}_{rN/L_w} \\ \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_{rN/L_w} \end{bmatrix}$$

A separate GMM may then be estimated for the operating conditions contained in matrix $X_c = [\mathbf{c}_1\ \mathbf{c}_2\ ...\mathbf{c}_{rN/L_w}]$ [2]. The conditional NLL value, y, for any signal segment may then be computed, $y_i = -\ln[P(\mathbf{s}_i, \mathbf{c}_i)/P(\mathbf{c}_i)]$.

Model selection

The more components a mixture model has the more expressive (more degrees of freedom) it becomes, theoretically allowing for a better representation of the density distribution of the data. Complex models are however also more prone to data overfitting, which may result in poor generalisation [90]. A model which does not generalise well may detect many false signal outliers, hence resulting in decreased performance.

There are a number of methods which may be considered for selecting the optimal number of mixture components. Some popular methodologies include: the Akaike information criterion, Bayes' information criterion, and Bayesian model selection. Since data of the gearbox in its healthy condition may generally be assumed to be fairly abundant, and due to its ease of implementation, this chapter resorts to model selection based on likelihood cross-validation as discussed by Smyth [109].

Cross-validation is a popular and simple method for measuring the ability of a model to generalise well [90] and Smyth [109] concludes that it is also well suited to selecting the number of components for a GMM, particularly when data and computational resources are relatively plentiful.

In essence the data which represent the healthy gearbox are divided into two sets. The first set is used to optimise the GMM by means of expectation maximisation. The second set is then used to measure the ability of the trained GMM to generalize well on novel data. This approach is used to investigate the appropriateness of different numbers of mixture components.

---

[2] The Gaussian distribution has the favourable characteristic that it is particularly easy to marginalise over variables. This renders it possible to avoid the necessity to train a second GMM model. The density distribution which only represents the operating conditions can be obtained from the full distribution which explain the joint signal segments and operating conditions distribution. This is done by simply discarding appropriate indices from the mean vector and covariance functions. The reader is referred to the book by Bishop [90] for additional reference to marginalisation in Gaussian distributions

Synchronous averaging

In this chapter a special case of the synchronous average is implemented. The NLL discrepancy signal $\mathbf{y}$ has a sampling frequency (samples per rotation of the gear) which is equal to the number of teeth on the gear. Similarly is the number of NLL datum points per rotation of the pinion equal to the number of teeth on the pinion. For this reason no further resampling of the NLL discrepancy signal will be required. The number of full shaft rotations during the measurement is represented by r. The gear synchronous average $\bar{\mathbf{y}}^g$ is computed as:

$$\bar{y}_i^g = \frac{1}{r}\sum_{n=1}^{r^g} y_{(n\times T_g)}, \tag{4.2.2}$$

where $\bar{y}_i^g$ represents the synchronous average associated with the ith tooth on the gear and r^g is the number of complete revolutions completed by the gear during the measurement. Similarly the pinion synchronous average $\bar{\mathbf{y}}^p$ is computed as:

$$\bar{y}_i^p = \frac{1}{r}\sum_{n=1}^{r^p} y_{(n\times T_p)}, \tag{4.2.3}$$

where $\bar{y}_i^p$ represents the synchronous average associated with the ith tooth on the pinion and r^p is the number of complete revolutions completed by the pinion during the measurement.

4.3 ANALYTIC INVESTIGATION

A simple analytical example is used to illustrate the underlying approach followed by the NLL discrepancy analysis technique. The analytic example is based on the gear model developed by McFadden [110], and which is also implemented by Combet and Gelman [79]. A vibration signal for a one stage gearbox is approximated by the following model:

$$p(t) = \sum_{m=0}^{M} P_h(1 + a_h(t))\cos(2\pi h f_m t + \theta_h(t) + \theta_h^0) + n(t) \tag{4.3.1}$$

where P_h is the amplitude of the hth mesh harmonic, f_m is the average mesh frequency for the stage considered, $a_h(t)$ and $\theta_h(t)$ are the amplitude and phase modulation functions of mesh harmonic h and θ_h^0 is the initial phase of harmonic h. The additive background noise $n(t)$ is assumed to be independently sampled from a zero mean normal distribution with a standard deviation of σ.

The mesh frequency f_m is proportional to the rotational speeds f_p and f_g of the pinion and gear, through the number of teeth T_p and T_g on the pinion and the gear.

$$f_m = T_p f_p = T_g f_g \tag{4.3.2}$$

The healthy vibration signal is approximated by 3 harmonic components, $M = 3$. The amplitudes P_h of all three components are simply set to 1. The signal is generated with 10k samples per revolution of the pinion.

The number of teeth on the pinion and the gear are respectively selected as $T_p = 20$, and $T_g = 43$. A measurement of 3s long is generated. During this time the gearbox is subject to a time-varying load. It is assumed that the pinion shaft speed is directly proportional to the applied torque. During the first 0.6s of the signal the gearbox is subject to a rapid and significant sinusoidal load fluctuation, during the last 2.4s the gearbox is subject to a smaller, more gradual sinusoidal load fluctuation. Due to the presence of the time-varying operating conditions, the signal experiences frequency, amplitude and phase modulation.

The pinion has a mean rotational frequency of $E(f_p) = 5$ Hz. Its instantaneous frequency $f_p(t)$ (due to the time-varying operating conditions) is indicated in table 4.1:

Table 4.1: Instantaneous pinion rotational frequency

time interval	pinion rotational frequency
first 0.6 s	$f_p(t) = 5 \times [1 + 0.5\sin(1.3\pi t)]$
last 2.4 s	$f_p(t) = 5 \times [1 + 0.3\sin(5 \times 1.3\pi t)]$

The amplitude and phase modulation are set proportional to the time-varying pinion shaft speed. The amplitude modulation a_h is similar for each of the tooth meshing harmonic components $h = 1,2,3$ as indicated in table 4.2.

Table 4.2: Instantaneous amplitude modulation

time interval	amplitude modulation function
first 0.6s	$a_h(t) = 0.5\sin(5 \times 1.25\pi t)$
last 2.4s	$a_h(t) = 0.3\sin(1.25\pi t)$

The phase modulation θ_h is also proportional to the instantaneous pinion frequency for each of the harmonics $h = 1,2,3$. However the phase modulation must also account for the wavelength of the harmonic. Short wavelengths must be phase modulated proportionally more than long wavelengths in order to ensure that the different phase components are not distorted and do not move relative to one another. The resulting phase modulation is presented in table 4.3

Table 4.3: Instantaneous phase modulation

time interval	phase modulation function
first 0.6s	$\theta_h = 0.2h \times 0.5\sin(5 \times 1.3\pi t)$
last 2.4s	$\theta_h = 0.2h \times 0.3\sin(1.3\pi t)$

The initial phase θ_h^0 of each of the harmonics is simply set to 0.

A single fault is simulated on the tenth pinion tooth. If the gear has a local defect such as a tooth root crack, then changes will occur in the vibration when the affected teeth are in mesh. McFadden [110] represents a localised gear crack by means of amplitude and phase modulation functions which are periodic with the gear rotational frequency. This assumption may result in a fairly deterministic fault component. The case study performed by Barszcz and Randall [52] however indicates that localised gear tooth cracks may give rise to high frequency (impulsive like) signal components, of which the amplitude may have a large variance. The experimental study performed in this chapter indicates similar fault behaviour.

The pinion tooth damage is simulated in this chapter by assuming that the tooth root crack gives rise to a relatively high frequency signal component. This is simulated by appending a tenth order harmonic to the signal, which obeys the same operating condition related amplitude and phase modulation rules as the first three (healthy gear)harmonic components. It is assumed that the high frequency component is only exited when the faulty tooth is in mesh, and then quickly dies out. This is done by amplitude modulating the 10th harmonic by an asymmetric saw-tooth waveform. The saw-tooth waveform is created in Matlab using the Tripuls function. The saw-tooth has a repetition frequency which corresponds to the rotational frequency of the pinion. The saw-tooth width is equal to the period that the 11th tooth is in mesh. In other words the 10th harmonic has zero energy at all positions, except for the approximate duration that the faulty tooth is in mesh. It will be seen that the high frequency 10th harmonic, which is subject to significant load induced amplitude and phase modulation will behave sufficiently stochastic so that conventional synchronous averaging will fail to detect it.

Figures 4.3 (a)-(d) investigate the nature of the simulated vibration waveform. Figure 4.3 (a) illustrates the noise-free, damaged vibration waveform, which is sampled at 10k datum points per revolution of the pinion. This results in 500 data points per signal segment. In Figure 4.3 (b) the noise

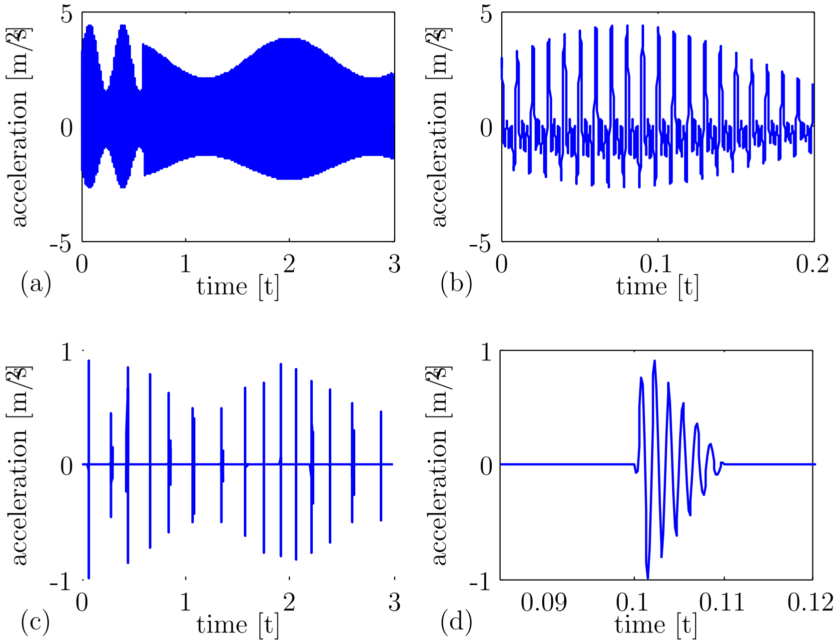


Figure 4.3: (a) Simulated damaged vibration waveform subject to amplitude and phase modulation, (b) enlarged view of the simulated vibration signal, (c) the saw tooth amplitude modulated 10th order harmonic component which represents the transient fault induced signal component, (d) an enlarged view of one of the fault induced signal transients indicates the saw-tooth nature of the envelope of the component.

free vibration signal is magnified, such that a single complete revolution of the pinion is visible. The fault induced signal distortion is seen at 0.1s. Figure 4.3 (c) shows the 10th harmonic (fault induced) signal component. The 10th harmonic is subject to the same load induced amplitude modulation experienced by the other 'healthy' harmonics, but is also transient due to its amplitude multiplication with the saw-tooth waveform. Figure 4.3 (d) presents an enlarged view of a single fault induced signal transient.

Figure 4.4 (a) illustrates the nature and extent of the amplitude and phase modulation present in signal segments which corresponds to a healthy tooth. The signal segments are presented without any additive noise. Each signal segment is 500 datum points long and corresponds to a tooth meshing period. Figure 4.4 (b) investigates the effect of the amplitude and phase modulation on signal segments which corresponds to the 11th (damaged) pinion tooth. Figure 4.4 (c) indicates the conventional vibration signal synchronous average. The high frequency component is smeared out due to the load induced phase modulation. This renders the damage on the 11th tooth near invisible in the conventional synchronous average. The synchronous average is computed on a noise free signal, so that the small

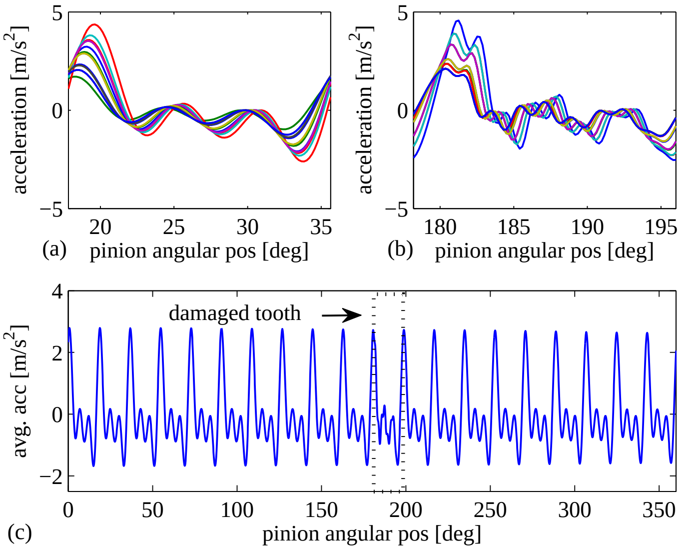


Figure 4.4: (a) Overlay of multiple signal segments from a healthy pinion tooth, indicating both amplitude and phase modulation, (b) overlay of multiple signal segments from the damaged pinion tooth, indicating how the higher frequency components significantly smears out (c) the resulting (noise free) synchronous average with respect to the pinion. The damage induced signal distortion on the 11th tooth (angular position 180-198 degrees) is almost invisible.

amount of damage on the 11th tooth may be observed. As soon as signal noise is added it becomes impossible to identify the damaged gear tooth.

A GMM density distribution is used to represent the signal segments which correspond to the fault free data. The vibration signal is sampled at 10k datum points per rotation of the pinion. Since there are 20 teeth on the pinion the signal segments have a length of 500 data points. The window length of 500 data points results in a GMM with many free parameters. In general it might be beneficial to consider down sampling the vibration signal so that a smaller GMM may be implemented. Alternatively it may also be considered to extract and model a subset of features (e.g. through principal component, or wavelet analysis). Lastly may it also be considered to constrain the covariance matrices, for instance by implementing diagonal covariance matrices.

Three approaches are investigated towards implementing the proposed strategy. The GMMs are implemented on vibration signals which were subject to random noise. A noise standard deviation of $\sigma = 0.25$ is implemented. This noise level renders it completely impossible to identify the faulty gear tooth by means of conventional synchronous averaging.

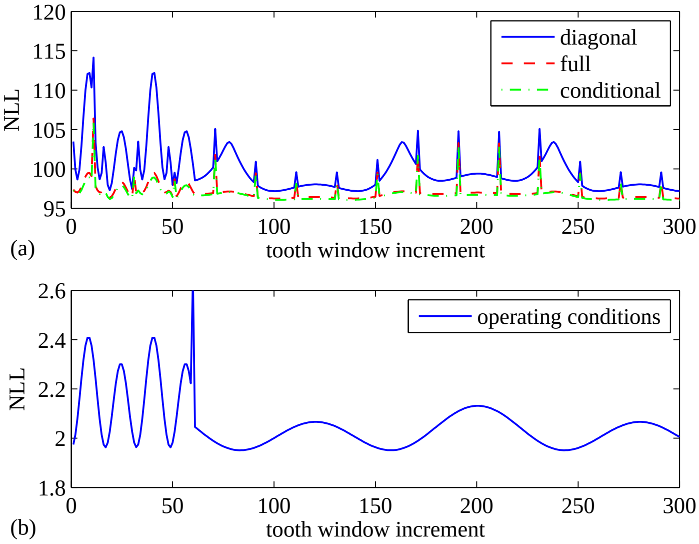


Figure 4.5: (a) Illustrates the NLL discrepancy signals as computed for the vibration signal, the continuous line represents NLL values $-\ln[P(\mathbf{s})]$ computed with a GMM which is limited to 2 components with diagonal covariance matrices, the dashed line represents a GMM with 2 components which have unconstrained covariance matrices, and the dash-dotted line (which closely corresponds to the dashed line) represents NLL values as conditioned on the pinion shaft angular velocity $-\ln[P(\mathbf{s_i},\mathbf{c_i})/P(\mathbf{c}_i)]$. (b) The NLL of the operating conditions $\ln[P(\mathbf{c}_i)]$.

The first approach is implemented by using a GMM with constrained covariance matrices. A diagonal covariance matrix is not capable of modelling the covariance between different datum points, but is also less prone to overfit data. GMMs with different numbers of mixture components are investigated. It is found that a single component with a constrained diagonal covariance is especially poor at discerning between time-varying operating conditions and fault induced signal outliers. Performance is significantly improved by implementing 2 components, but it is found that at least 10 components are required to obtain results which are comparable with a 1 component full rank covariance GMM model. The performance of a 2 components GMM with diagonal covariance matrices are indicated in figure 4.5 (a) with the solid (blue) line. The NLL values are seen to be sensitive to the operating conditions. The dimension of the input vector $\mathbf{x}$ is equal to the length of the vibration signal segment, nl 500.

The second approach implements a GMM with full rank covariance matrices. Full rank covariance matrices render it possible to model the interdependencies between different datum points in a signal segment. Even a single full rank covariance component performs quite well in this simple analytical example. The dashed line in figure 4.5 (a) indicates the NLL values as obtained with a 2 component GMM with full rank covariance matrices. The dimension of the input vector $\mathbf{x}$ is again equal to the length of the vibration signal segment, nl 500.

The third and final approach investigates the use of conditional NLL values. To compute any single conditional negative log likelihood value $-\ln[P(\mathbf{s}_i, \mathbf{c}_i)/P(\mathbf{c}_i)]$ it is required to estimate two GMM density distributions, one which represents the joint density distribution $P(\mathbf{s}_i, \mathbf{c}_i)$ for the signal segments $\mathbf{s}_i$, for $i = 1..rN$ and the operating condition variable(s) $\mathbf{c}_i$, for $i = 1..rN$, and a second density distribution $P(\mathbf{c}_i)$ which represents the operating conditions. Only a single operating condition variable is investigated, so that the vector $\mathbf{c}_i$ reduces to a variable c_i. The pinion angular velocity is used as operating condition variable c, since it is known to be indicative of the amplitude and phase modulation. Both density distributions are estimated with 2 component GMM models with full rank covariance matrices. The length of each input vector $\mathbf{x}$ is equal to the length of the vibration signal segment plus the operating condition dimention of length 1, so that the total length equals 501.

Figure 4.5 (b) indicates the computed NLL for the operating conditions. It is seen that the NLL values for the operating conditions which correspond to the first 0.6s generally have large values. This shows that the first 0.6s of operating conditions are not well representative of the typical operating conditions. A large spike is observed at the end of the 0.6s. This represents the transition from one operating condition to the other. The large spike indicates that this transitional operating condition is very uncommon.

The NLL discrepancy signal obtained for the GMM with the diagonally constrained covariance matrices is represented by the continuous line in figure 4.5 (a). The limited covariance matrix used by the GMM results in a NLL discrepancy signal which is sensitive to the operating conditions and subsequently not well suited for detecting the fault induced signal component.

The NLL values obtained for the full rank covariance GMM model is represented by the dashed line, and the dash-dotted line illustrates the conditional NLL values. The NLL discrepancy signal obtained for both the full rank covariance GMM model and the conditional NLL discrepancy signal are very similar. This indicates that the joint density distribution did not manage well to model the non-linear correlation between the signal segments and the pinion rotational frequency. In this scenario conditioning the NLL on the operating condition did not serve to render the NLL more robust to the time-varying operating conditions. However the NLL values obtained for the operating conditions (figure 4.5 (b)) may still be beneficial. Those sections of the vibration signal which were generated under operating conditions which are not well representative of the typical operating conditions may be discarded. For instance, figure 4.5 (b) may be used to determine that the first 0.6 seconds of signal should be discarded and not used to compute the NLL synchronous average.

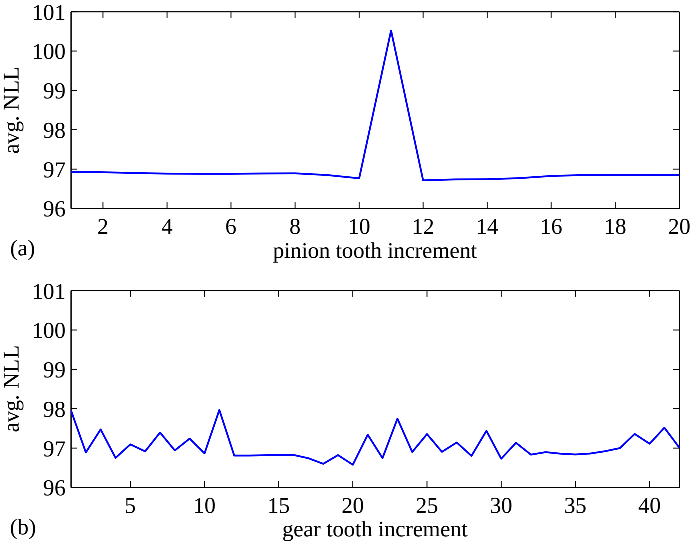


Figure 4.6: (a) The synchronous averages of the signal segment NLL values which correspond to the 20 teeth on the pinion, and (b) the synchronous averages which correspond to the 43 gear teeth.

Figures 4.6 (a) and (b) respectively indicate the NLL synchronous averages with respect to the pinion, and the gear. The NLL signals are generated with the unconditional full rank covariance GMM with 2 components. These synchronous averages are computed by using the whole (noisy) NLL signal, which includes the first 0.6s during which the gear was subject to significant fluctuating operating conditions. The pinion completed 14 whole rotations, while the gear only completed 6. A single averaged datum point corresponds to each gear mesh period. This renders it possible to synchronous average the NLL values with respect to both the pinion and the gear without the need to resample the NLL signal. A large average NLL value corresponds to the 11th pinion tooth, clearly indicating the presence of the fault induced signal irregularity.

In this simple analytical example, the NLL synchronous average was able to clearly detect tooth damage, which could not be detected by conventional synchronous averaging.

4.4 EXPERIMENTAL SETUP

The proposed NLL synchronous averaging methodology is subsequently investigated on experimental data which were recorded by Dr. C.J. Stander in the Sasol Laboratory for Structural Mechanics at the University of Pretoria [10].

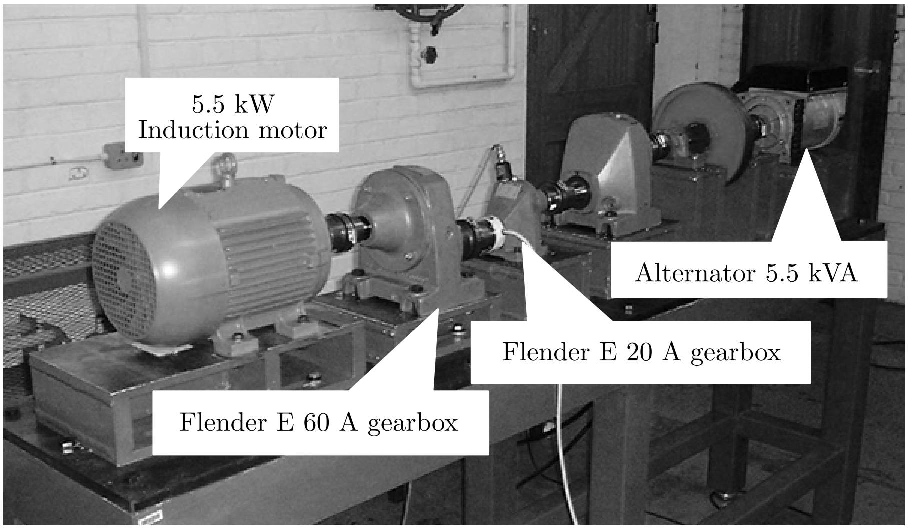


Figure 4.7: The experimental setup of the gear test rig.

Gear damage was investigated on a single stage Flender E20A gearbox (rated load of 20 Nm) with helical gears and a step up ratio of 1:1.96. A mounted test bed was designed to conduct accelerated gear life tests on the gearbox under time-varying load conditions. As such the monitored gearbox was mounted between two Flender E60A gearboxes, the first of which served to step the torque up and the second to step it down. The E60A gearbox pair increased the torque over the E20A test gearbox by a ratio of 1:4.72. The fundamental system frequencies are presented in table 4.4.

Table 4.4: Fundamental frequencies

Gearbox	Role	Shaft input	Teeth ratio	Shaft output	Mesh Freq.
Flender E60A	Speed step down	24.5 Hz	18/85	5.19 Hz	441 Hz
Flender E20A	Test gearbox, speed step up	5.19 Hz	43/22	10.13 Hz	223 Hz
Flender E60A	Speed step up	10.13 Hz	85/18	47.88 Hz	862 Hz

The step down gearbox was driven by a 5.5 kW three phase four pole WEG squirrel cage electric motor. The step up gearbox was connected to a flywheel which in turn was coupled with a 5.5 kVA Mecc Alte Spa three phase alternator to apply the load.

The instantaneous gear and pinion angular positions and velocities were estimated from a synchronising pulse which was measured by means of a proximity switch on the key of the E20A gear shaft. The gear casing acceleration response was measured in the vertical direction by means of a 10 V/g PCB integrated circuit piezoelectric industrial accelerometer. The measurements were taken with a

Siglab model 20-42 signal analyser at a sampling frequency of 51.2 kHz.

A number of time-varying loading conditions were investigated. The load conditions are summarized in Table 4.5.

Table 4.5: Load case specifications

Load case	Load function	Frequency	Minimum load	Maximum load
1	Sine	0.5 Hz	7.4 Nm	14.7 Nm
2	Sine	1 Hz	7.4 Nm	14.7 Nm
3	Square	0.5 Hz	7.4 Nm	14.7 Nm
4	Square	1 Hz	7.4 Nm	14.7 Nm
5	Chrip	0.1-2 Hz	7.4 Nm	14.7 Nm
6	Random	0.1-2 Hz	7.4 Nm	14.7 Nm

Four measurements for each load condition were recorded. The first measurement represented the healthy gearbox. Flank wear was subsequently induced on one of the gear teeth by progressively removing more of the gear tooth face. Each measurement was recorded for 30s.

Table 4.6: Seeded gear damage condition

Measurement	Fault severity
0	Good condition
1	100 μm tooth face removal
2	200 μm tooth face removal
3	300 μm tooth face removal

The subsequent sections investigate the vibration signal, the vibration based synchronous average, the spectrum analysis, the NLL discrepancy transform and the NLL synchronous average. The *magnitude* of the results vary slightly depending on the load scenario, however the *nature* of the results remain fairly consistent. For this reason one load scenario, namely the 0.5 Hz sinusoidal load scenario will consistently be used to visualise the application of the techniques.

Time domain waveform

The time domain waveform for load scenario 2, with the 0.5 Hz sinusoidal component, is illustrated in figures 4.8 (a)-(d), where (a) represents the healthy waveform, and (b)-(d) represent progressive damage. The sinusoidal component of the load induces periodic amplitude and phase modulation in the signal. As the damage increase the presence of signal impulses becomes apparent. In general the kurtosis values for the signals increase for the 2nd and especially the 3rd damage condition. The first damage condition can usually not be detected based on the kurtosis of the signal. The magnitude of the impulses, as observed for any one damage condition and any one loading scenario, tends to vary significantly.

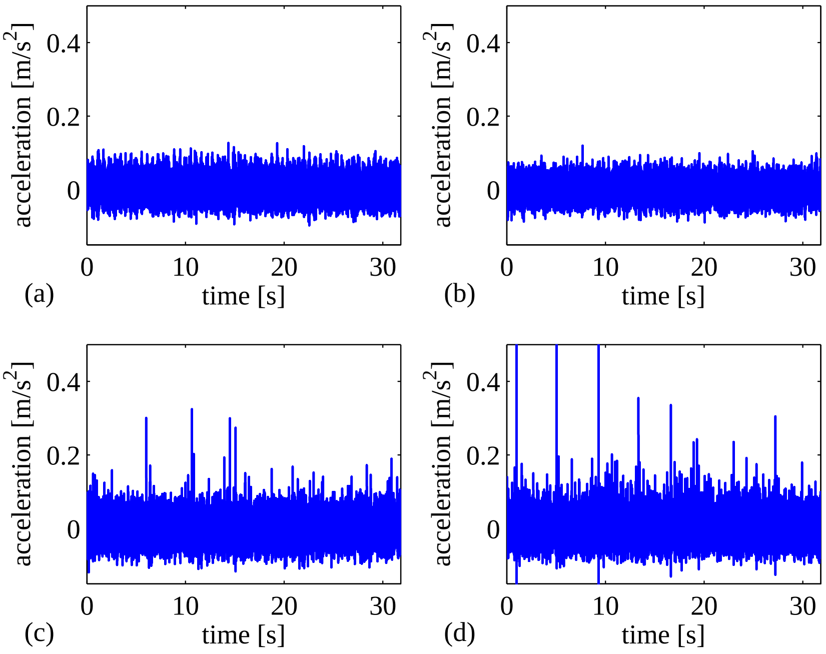


Figure 4.8: Vibration signal generated under 0.5Hz sinusoidal loading condition, for (a) the healthy gear, and (b)-(d) progressive stages of damage.

Power spectral density (PSD)

The order domain power spectral densities (PSD) of the vibration signals are investigated. Towards this goal the time domain signals are resampled to the order domain by means of piecewise cubic interpolation of the once per gear revolution impulses as measured with the proximity probe.

Before resampling is performed the signal is further low-pass filtered to avoid aliasing. The time-domain signal which has a sampling frequency of 51.2 kHz is resampled at 8600 datum points per revolution of the gear. The gear has an average rotational frequency of 5.19 Hz, so that the angular domain resampling of 8600 datum points approximately corresponds to a time domain sampling rate of 44.6 kHz. To avoid aliasing during the resampling period an eighth-order lowpass Chebyshev Type

I filter is used with a cut-off frequency of 35 kHz.

The natural logarithms for the order domain PSDs as computed for the 0.5 Hz sinusoidal time-varying load are presented in figures 4.9 (a)-(d), where (a) represents the gear in a good condition, and (b)-(d) represent progressive stages of damage. The first order corresponds to the gear rotational frequency of 5.19 Hz.

The PSD for each signal is computed as follows. The signal (of approximate length 1.4×10^6) is divided into 3, partially overlapping sections of 2^{19} datum points each. Each of the sections are Hanning windowed. Their FFTs (Fast Fourier Transforms) are then computed before the results are averaged. The FFTs are subsequently multiplied by their complex conjugates to obtain the PSD values. To compensate for the Hanning window the signal amplitude is corrected by factor 8/3. The PSD window length is selected to offer a compromise between spectral resolution, and the noise level of the PSD.

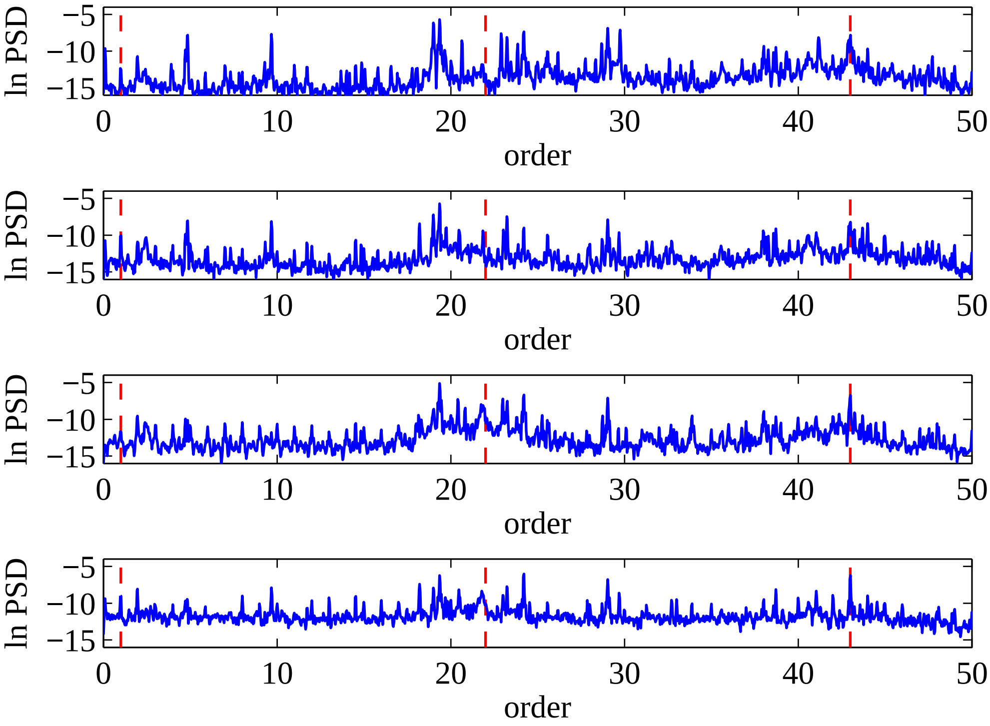


Figure 4.9: The order domain PSDs of the vibration signals as measured under the 0.5 Hz loading condition for (a) no damage, and (b) - (d) progressive stages of damage.

Significant energy is observed at approximately the 2nd order (10.14 Hz) and the 4.8th order (24.5

Hz), which respectively correspond to the drive shaft frequencies for the E60A step up and step down gearboxes. A prominent peak is observed at the 19.3th order (100 Hz). This peak corresponds to twice the 50 Hz AC line frequency. Generally the 100 Hz component tends to be more prominent in motors and alternators than the AC frequency component. Side bands due to amplitude modulation are observed around this 100 Hz frequency, this is especially noticeable in figures 4.9 (a) and (b). Large signal energy components are also identified at the following orders; the 28.9th (150 Hz), the 33.7th (175 Hz), the 38.5th (200 Hz), and the 43th (223 Hz). The 150 Hz and 200 Hz components may be harmonics which correspond to the excitation induced by the AC line current. The 175 Hz component might be due to amplitude modulation between the 200 Hz component and the 24.5 Hz E60A drive shaft frequency component. The 43th order component corresponds to the gear meshing frequency of the Flender E20A gearbox of interest.

While it is not indicated on the current axis, energy components which correspond to the E60A step down gearbox meshing frequency of 441 Hz and the E60A step up gearbox meshing frequency of 862 Hz may also be observed on the extended PSD.

The frequency components which are of interest with regard to the condition of the gear in the Flender F20A gearbox are magnified for better inspection in figures 4.10 (a)-(h).

Figures 4.10 (a), (c), (e) and (g) illustrate the signal energy around the first order component which correspond to the gear rotational frequency, while figures 4.10 (b), (d), (f) and (h) illustrate the signal energy associated with the 43 order (gear mesh frequency).

It is interesting to note the increase in the energy at the tooth-meshing frequency from a fault on one tooth. Damage to one tooth normally does not increase the tooth-mesh component since this component represents the averaged energy associated with the meshing of all of the teeth. This increased energy must be due to some nonlinear effects. It will be seen in vibration signal synchronous average (figure 4.11) that the signal energy does indeed slightly increase at sections other than just the damaged gear tooth. This might potentially be due to torsional dynamics induced by the shaft flywheel. Figures 4.10 (f) and (h) indicate increased energy at the tooth-meshing energy, as well as increased energy at the 42th and 44th orders. This is indicative of amplitude modulation between the 1st order (gear rotational frequency) and the 43th order (meshing frequency). This result is in line with the fact that the damage is on the gear and not the pinion. The energy associated with the 1st order does not consistently increase as the gear tooth damage progresses.

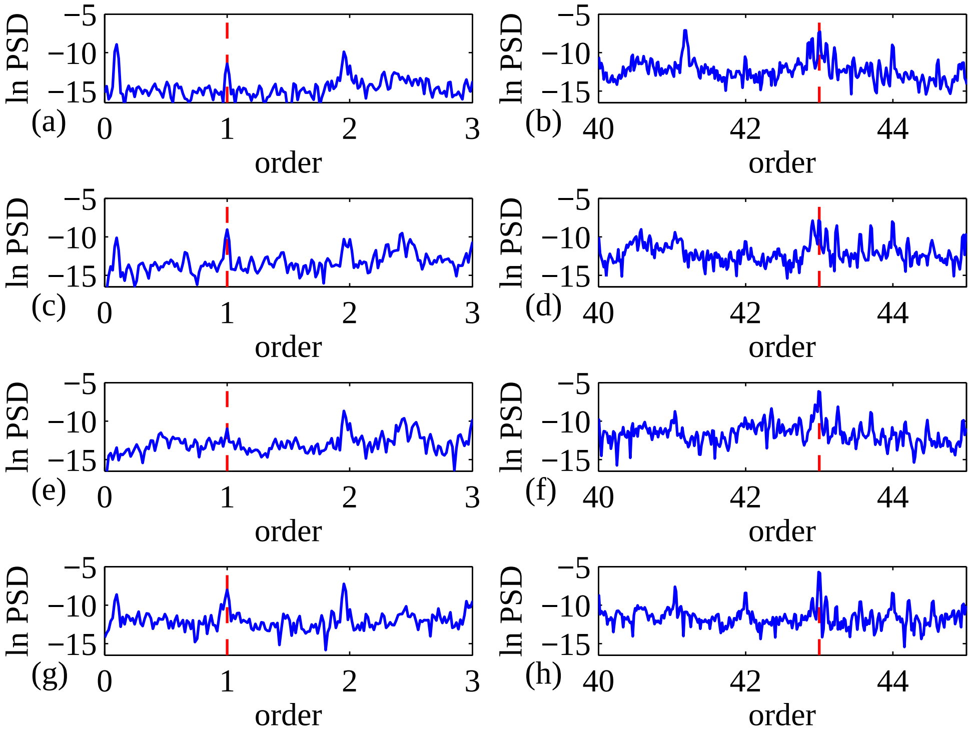


Figure 4.10: Close up of the natural logarithm of the order domain PSDs of the 0.5Hz sinusoidal time-varying load. Figures (a), (c), (e) and (g) magnifies the energy components in the vicinity of the 1st order (gear rotational frequency) for progressive stages of damage. Figures (b), (d), (f) and (h) magnifies the energy components in the vicinity of the 43st order (mesh frequency) for progressive stages of damage.

Synchronous averages of the resampled vibration signal

The vibration signal synchronous averages are computed, both for the gear and the pinion.

Each signal is approximately 30s long. With gear and pinion rotational frequencies of respectively 5.19 Hz and 10.12 Hz this results in 154 and 302 complete revolutions of respectively the gear and the pinion.

Figures 4.11 (a)-(h) illustrate the vibration signal synchronous averages for the 0.5 Hz sinusoidal time-varying loading condition. The figures (a), (c), (e) and (f) indicate the synchronous averages for progressive gear damage, while (b), (d), (f) and (h) are computed with respect to the pinion. Notice that the x-axis (angular position) indicated for the pinion is not consistent. This is because

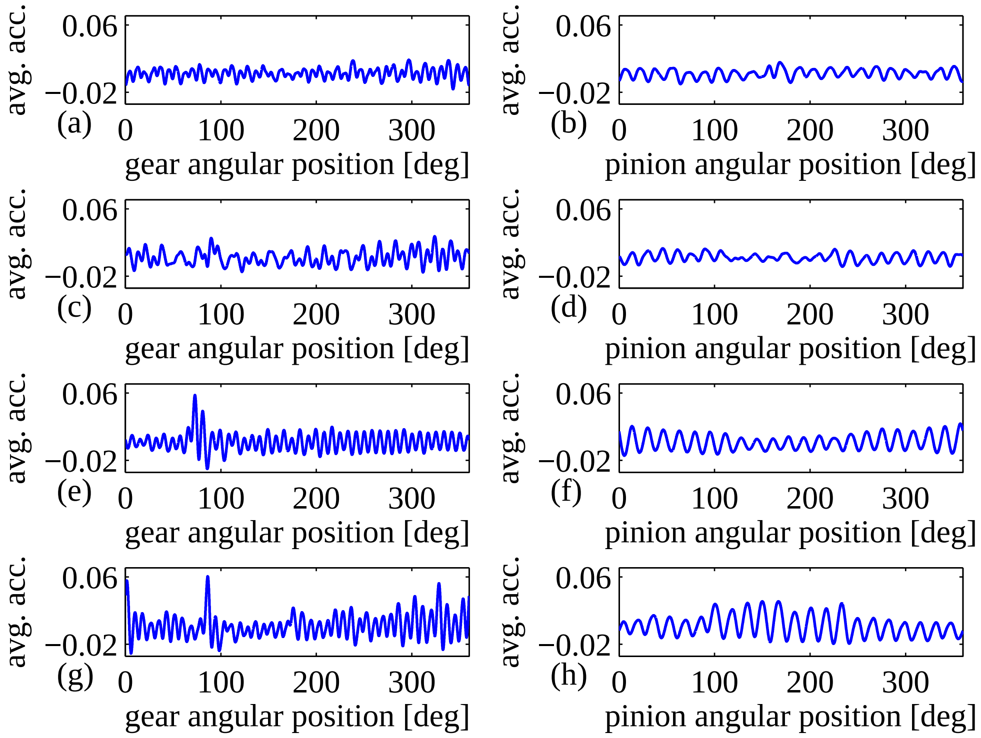


Figure 4.11: Vibration signal synchronous average for the 0.5 Hz fluctuating load. Progressive localised gear damage is indicated in (a) - (d), while (e) - (h) depict the synchronous averages of the fault-free pinion.

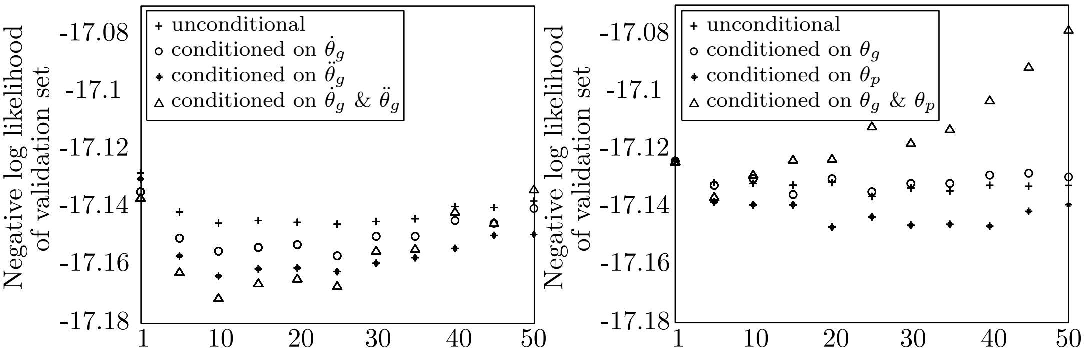
the tachometer is mounted on the gear shaft and not the pinion. The initial pinion angular phase is unknown at the beginning of each signal measurement. It is however possible to align the pinion synchronous averages by optimising their cross-correlation.

The initial gear damage (first stage) is not evident from the gear synchronous averages, however the more advanced second and third stages of damage is evident. The nature of the damage is however not readily evident. The pinion synchronous averages are also influenced by the presence of the gear tooth damage. This renders the pinion synchronous averages somewhat ambiguous, as it may appear that the pinion is subject to general wear.

NLL discrepancy measure implementation

The angular resampled frequency has 8600 datum points per revolution of the gear. The gear has 43 teeth, while the pinion has 22. A window length is selected so as to correspond to the gear meshing period. The window length is thus set equal to 200 datum points ($L_w = 8600/43$).

A reference density distribution is estimated by means of a GMM with full-rank covariance matrices. The cross-validated NLL is used as measure of generalisation.



(a) k number of mixture components (b) k number of mixture components

Figure 4.12: Model performance as estimated by cross-validation for different numbers of mixture components. Cross validation is performed for the estimated densities as conditioned (a) on gear speed and acceleration, (b) and on angular positions.

Figure 4.12 (a) investigates the performance of the density distribution (as estimated for different numbers of mixture components) on the validation set. The performance of the unconditioned model is compared to where the NLL is first conditioned on the gear angular speed $\dot{\theta}_g$, then the NLL is conditioned on the gear angular acceleration $\ddot{\theta}_g$, and lastly the NLL is simultaneously conditioned on both the gear angular speed and acceleration. Since the gear angular speed and the pinion angular speed are proportional it would be redundant to additionally condition the signal on the pinion angular speed and acceleration. The best performance (smallest negative log likelihood of the validation set) is observed when 10 mixture components are used and when the signal is conditioned on both the gear velocity and acceleration. When more than 10 components are used the performance on the validation set declines, indicating that the models are beginning to over-fit the training set.

Figure 4.12 (b) investigates the cross validation performance of the density models when the NLLs are conditioned first on the gear angular position θ_g, then on the pinion angular position θ_p, and lastly where it is simultaneously conditioned on both angular positions. It is observed that conditioning on the pinion angular position slightly improves the performance of the NLL estimate. The gear angular position has almost no effect on the NLL. This indicates that the healthy signal behaviour is fairly consistent for different gear angular positions. Joint conditioning on both the pinion and the gear angular positions resulted in reduced performance; this is probably due to model over-fitting.

It is interesting to note that the performance of the unconditional likelihood as presented in figure 4.12 (a) differs slightly from that of figure 4.12 (b). This reflects that for the EM algorithm the components are initialized by centring them on randomly selected samples from the training set. The EM algorithm is not guaranteed to find the global optimum, so that it is expected that the performance from different optimization runs will differ slightly.

Figure 4.12 suggest that best performance will be obtained by implementing a GMM with 10 components, and by conditioning the NLL on both the gear angular speed and acceleration. However, very similar NLL synchronous averaging results were obtained when a simpler model was used. For computational efficiency and to ensure proper generalization all subsequent results were obtained using an unconditional GMM with 2 full rank mixture components.

Figures 4.13 (a)-(d) illustrate the NLL discrepancy signal as computed for the 0.5 Hz sinusoidal loading condition. Figure 4.13 (a) illustrates the NLL values associated with the healthy gearbox, while (b)-(d) illustrate progressive stages of damage. The NLL discrepancy signal has a significantly lower sampling resolution (by a factor of 200) compared to the vibration signal. Each NLL datum point represents a single meshing period.

Figures 4.13 (a)-(d) illustrate the importance of analysing the natural logarithm of the likelihood, rather than the likelihood values themselves. The outliers must be scaled such that a few significant outliers do not completely dominate the discrepancy signal. The likelihood values of outliers in figures 4.13 (c) and (d) are so large (and subject to such variance) that the likelihood (compared to the log likelihood) synchronous averages would be very noisy and difficult to interpret. The NLL synchronous averages are also more linearly proportional to the magnitude of the fault induced damage.

The NLL synchronous averages for the 0.5 Hz sinusoidal fluctuating operating as computed with respect to the gear for progressive stages of damage are depicted in figures 4.14 (a), (c), (e) and (f), and as computed with respect to the pinion for progressive stages of wear are depicted in (b), (d), (f) and (h). The initial (first stage) damage on the gear results in very small NLL values. For this reason the y-axes on figures 4.14 (a)-(d) are magnified compared to the later stages of damage depicted in (e)-(h). The damage condition is fairly simple to interpret and the presence of damage of the gear is quite evident.

The results obtained for the other loading conditions proved conceptually very similar to the presented

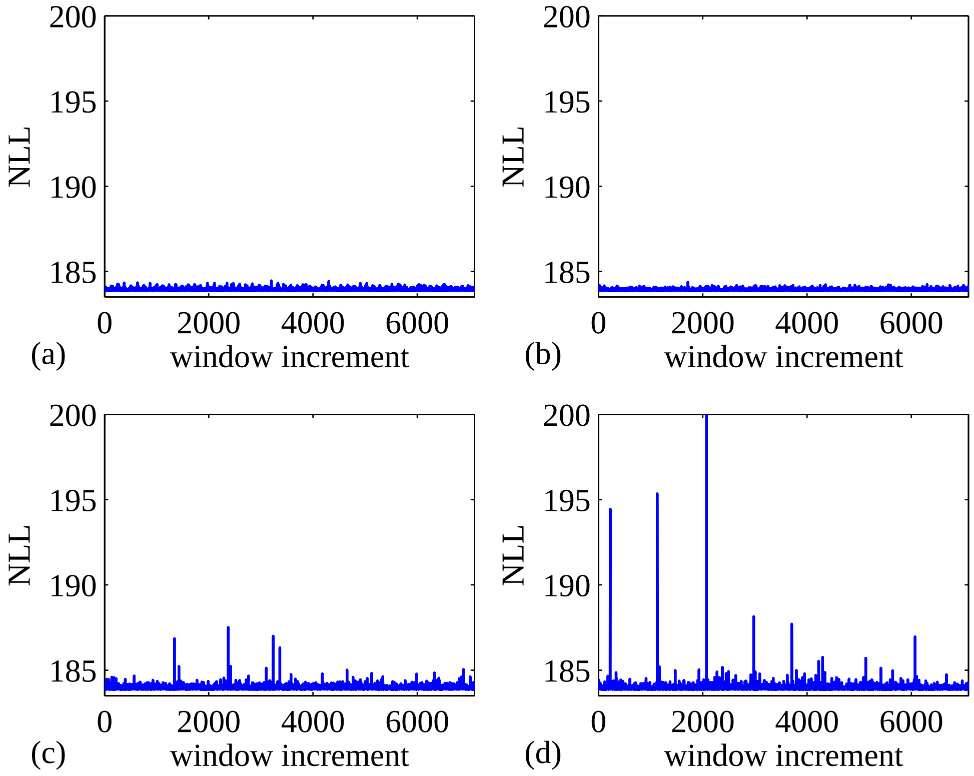


Figure 4.13: The NLL discrepancy signal associated with (a) the healthy gearbox, and (b)-(d) with increased stages of damage. The NLL discrepancy signal enhances the presence of fault induced signal outliers.

results for the 0.5 Hz sinusoidal loading condition. Future research should be conducted to investigate the methodology under greater time-varying operating conditions.

4.5 CHAPTER CONCLUSION

This chapter proposed a discrepancy analysis methodology for conducting gear condition monitoring. The method comprises two steps:

Firstly a NLL transform of a vibration signal is computed. The NLL transform is used to enhance diagnostic information in the signal, while suppressing the influence of signal components that are representative of normal operating conditions. The NLL is less sensitive to fluctuating load conditions, while also being smoother than the original signal.

Secondly the structure in the NLL discrepancy signal is analysed by means of synchronous averaging. The NLL synchronous average allows for an intuitive yet sensitive and robust, visual comparison of

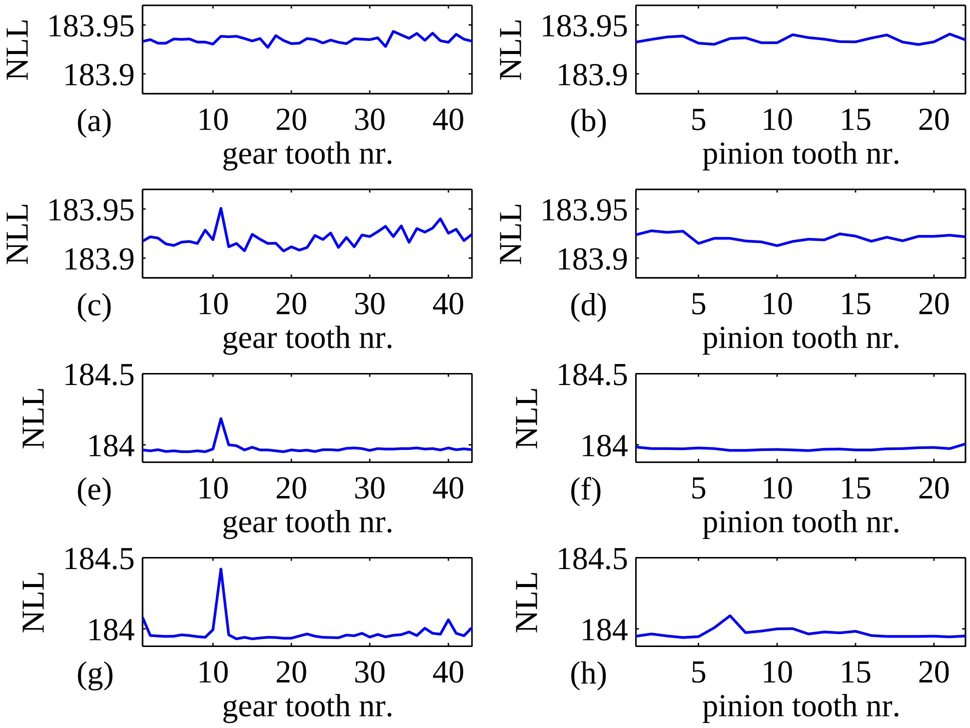


Figure 4.14: The NLL synchronous averages for the 0.5 Hz sinusoidal time-varying load condition, as computed with respect to the gear for progressive stages of damage (a), (c), (e) and (f), and as computed with respect to the pinion for progressive stages of wear (b), (d), (f) and (h). Please notice that the y-axes on (a)-(d) are scaled different to those of (e)-(h).

the novelty of signal patterns at different angular positions. The novelty of signal patterns can be used as an indication of the presence of damage.

A simple analytic example was used to illustrate how time-varying operating conditions (and especially phase modulation) may impede the ability of conventional vibration signal synchronous averaging to extract high frequency signal components. In this simple case study the NLL synchronous average proved significantly more sensitive to high frequency fault induced signal components.

Vibration data from an experimental gear test rig were subsequently used to further investigate the proposed methodology. An E20A Flender gearbox was subjected to a number of different time-varying loading conditions. Each of the loading conditions were used to generate vibration data for one of four stages of induced damage on a gear tooth. The NLL synchronous average proved simpler to interpret and more sensitive the induced fault compared to order domain spectral analysis and conventional synchronous averaging.

It is believed that the proposed methodology may be used to generate a simple, yet robust representation of the conditions of a gearbox. This representation may potentially be used to visualise the condition of a gearbox, or as a preprocessing stage to extract features which may subsequently be of use in a diagnostic change detection, or in a fault classification algorithm.

Chapter 5

Discrepancy analysis with sensorless signal resampling

5.1 INTRODUCTION

This short chapter concludes the discrepancy analysis methodology investigated in chapters 3 and 4.

This chapter is concerned with the condition monitoring of a rotating machine that is subject to load and speed fluctuations and where it is assumed that absolutely no shaft angular position measurements (tachometer, encoder, etc) are available. It is firstly proposed that a neural network may be used to remove the regular (possibly non-linear) components from the vibration signal so as to obtain a residual signal. The envelope of the residual signal is then computed by means of the Hilbert transform. The obtained transform is referred to as a discrepancy transform, since the discrepancy signal indicates the presence of fault induced signal distortions (discrepancies) which are not present in the baseline (healthy) signal. The discrepancy signal tends to be significantly simpler (smoother) than the original vibration waveform and thus tends to be more robust to slight inaccuracies during a resampling stage. The instantaneous shaft angular speed is subsequently estimated from the original vibration signal and used to approximately resample the discrepancy signal to account for frequency modulation.

This chapter uses the previously developed numerical gear model to illustrate how the combination of discrepancy analysis along with approximate signal resampling may be used to monitor the condition of a rotating machine which is subject to fluctuating operations when no shaft angular position measurement is available.

5.2 METHODOLOGY

The methodology implemented in this chapter is summarized in figure 5.1.

The first step is concerned with computing a residual signal by means of comparing a novel vibration signal to the reference neural network (NN) model. The reference model represents the vibration response from the gearbox while it is in a good condition. It is assumed that the baseline vibration response is representative of the different possible operating conditions. A non-linear NN regression model is used to represent the time-series data. The performance of the NN is compared to that of a linear AR model. Once the residual signal is available its analytic representation is obtained by means of the Hilbert transform. The analytic representation is used to compute the signal envelope, which is also referred to as a discrepancy signal. The discrepancy signal indicates the time instances and the extent to which the measured signal deviates from the baseline model.

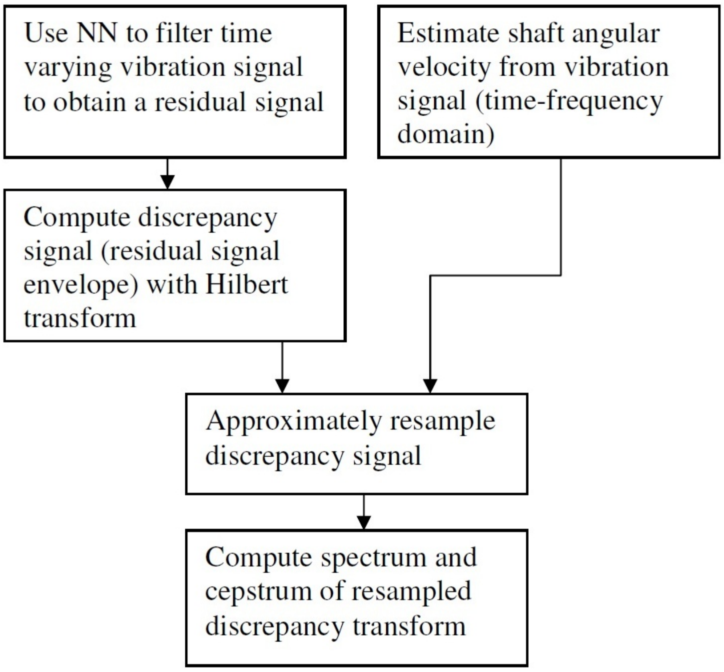


Figure 5.1: Flow chart of the proposed methodology

The information contained in the discrepancy transform is much simpler (lower bandwidth) compared to the original vibration waveform and is hence less sensitive to noise in the reference signal which will be used for resampling. The relative shaft rotational velocity may be estimated from the vibration waveform. In this chapter a simple clustering approach which considers the local maxima in the spectrogram is used to estimate the relative shaft speed. Depending on the application more advanced signal resampling methodologies may be desirable, however the aim of this chapter is to illustrate that the methodology may be implemented even with only approximate relative speed measurements. Lastly the spectrum and cepstrum of the discrepancy transform are computed and interpreted.

5.2.1 Time-series modelling

The residual signal r_t at time instant t is computed as the difference between the observed signal value x at time instant t and the one step ahead model prediction $\hat{x}_t$ so that $r_t = x_t - \hat{x}_t$.

5.2.1.1 Autoregressive (AR) filter

An AR model of order p is described by $\hat{x}_t = \sum_{k=1}^{p} a_k x_{t-k}$ so that the expected value $\hat{x}_t$ differs from the real waveform x_t by some white noise $b_0\sigma_t$. The AR spectral model, which is an all-pole model, may be shown to be good for modelling spectra with sharply defined peaks [94]. The AR coefficients are optimized based on a least squares error criterion.

5.2.1.2 Neural network regression

A simple feed forward neural network (NN) is also used to represent the time-series data. Similar to the AR model the NN is implemented so as to estimate the one step ahead prediction $\hat{x}_t$ based on the p most recent signal observations $\{x_{t-k}^{p}\}_{k=1}$. A NN architecture with sigmoid activation functions in the hidden layer and a linear activation function in the output layer is selected. To account for a possible constant offset a bias value of one is added to the inputs. It has been shown that when this architecture is employed with sufficiently many nodes in the hidden layer it is an universal approximator which is capable of representing any continuous function to arbitrary precision [90].

The training data set consists of n input vectors which correspond to the n target values (which are the one step ahead predictions). The neural network may be summarised by equation 5.2.1, where $g(\circ)$ represents the sigmoid activation function.

$$\hat{x}_t = (\sum_{j=1}^{p} w_j g(\sum_{k=1}^{p} w_{jk} x_{t-k} + w_j) + w_b) \tag{5.2.1}$$

The network is trained in Matlab by using backpropagation (efficient gradient descent) optimization whereby the weights of the networks are adjusted so as to minimize the sum squared error (SSE) between the predicted values $\hat{x}_t$ of the network and the target values.

5.2.2 Hilbert transform envelope

To compute the residual envelope the residual signal must first be represented in its analytic form z. The analytic form comprise of the real component r (the original residual signal) and the imaginary component $jH(r)$:

$$z = r + jH(r) \tag{5.2.2}$$

The imaginary part is computed by means of the Hilbert transform

$$H(r) = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{r(\tau)}{t-\tau}d\tau \tag{5.2.3}$$

and the residual envelope (or discrepancy signal) d is subsequently computed as the complex modulus:

$$d = \sqrt{r^2 - (jH(r))^2} \tag{5.2.4}$$

5.2.3 Sensorless relative speed estimation and resampling

The relative shaft speed is estimated directly from the vibration signal. As discussed in the literature study there are different methodologies which may be followed to estimate the shaft angular position (or velocity) directly from the vibration signal. The speed estimation method employed in this chapter is based on a time-frequency analysis. In essence the local maxima of a spectrogram (magnitude of the short-time Fourier transform) are computed. Each local maximum point is associated with a unique time $t*$, magnitude $m*$, and frequency $f*$ value.

Visual inspection is used to select a frequency component (the gear meshing frequency or one of its higher harmonics) which is synchronous with the frequency modulation. The selected frequency component needs to be free from other parasite (interfering) frequency components.

Beginning at a manually selected point a clustering algorithm is employed which groups together local maxima based on a nearest neighbour criterion. The distance δ_{ij} between neighbouring points i and j (local maxima) is based on the modulus of the scaled time difference Δt_{ij}, magnitude difference Δm_{ij}, and frequency difference Δf_{ij} between the two points:

$$\delta_{ij} = \sqrt{(\frac{\Delta t_{ij}}{a})^2 + (\frac{\Delta m_{ij}}{b})^2 + (\frac{\Delta f_{ij}}{c})^2} \tag{5.2.5}$$

In this case study the scaling factors a, b and c are selected so as to unit normalize the standard deviation for each of the clustering dimensions. The scaling factors will require different criteria for different problems.

Once all the maxima which are related to the frequency component of interest have been clustered they are low pass filtered and used to resample the vibration signal.

5.3 CASE STUDY

5.3.1 Gear model

The proposed methodology is briefly investigated on a data set which is simulated based on the same gear model as implemented in chapter 3. This dynamic gear model is inherently very simple and comprises a set of second order differential equations which relate the forcing function (applied torque), as well as the lumped mass, stiffness and friction coefficients to the vibration response of the gear casing. The simple gear model is preliminarily used to investigate the efficiency of proposed methodology. There are several assumptions which could be improved for gearbox dynamic oriented modelling as provided in [111] or recently [112] and [113]. Research conducted by Jia et al. [106] indicate that it is appropriate to approximate the gear mesh stiffness as a square function. The magnitude of the meshing stiffness is selected to fluctuate by approximately 20% around the mean value. It is common to model the effects of tooth faults, such as a tooth root crack, by reducing the stiffness of the affected tooth [106].

The stiffness coefficients k_1, k_2, k_3 and k_4 are assumed to satisfy Hooke's law by being linear in x. Standard viscous damping is assumed for all of the damping elements, with values which ensure that the system is highly underdamped. The gear is driven by the pinion, which in turn is subject to the pinion input torque T_p. The load on the gear is proportional to the second power of the gearwheel speed $T_g = K_s\dot{\Theta}_2^2$ through the proportionality constant K_s. This allows the system to accelerate until equilibrium is reached. The proportionality constant K_s is selected at 0.4, so that the pinion rotates at about 240 rpm under an applied load of 500 Nm.

A single gear fault is simulated by reducing the tooth mesh stiffness by 6% for the duration that the individual tooth that is affected is in mesh.

The differential equations are rewritten in state space notation and solved using Matlab's ode45 solver. The reader is referred to chapter 3 for the written out differential equations and parameter values. White noise is added to the signal such as to obtain a signal to noise ratio of 25 dB.

5.3.2 Operating conditions

Two loading conditions are investigated; namely a steady state operating condition (constant torque) and also a loading condition where the applied torque values with time. The time-varying load is chosen arbitrary, and simply aims to illustrate the effect on the signal of both applying a slow and a slightly faster time varying torque component. Since the speed is proportional to the applied torque,

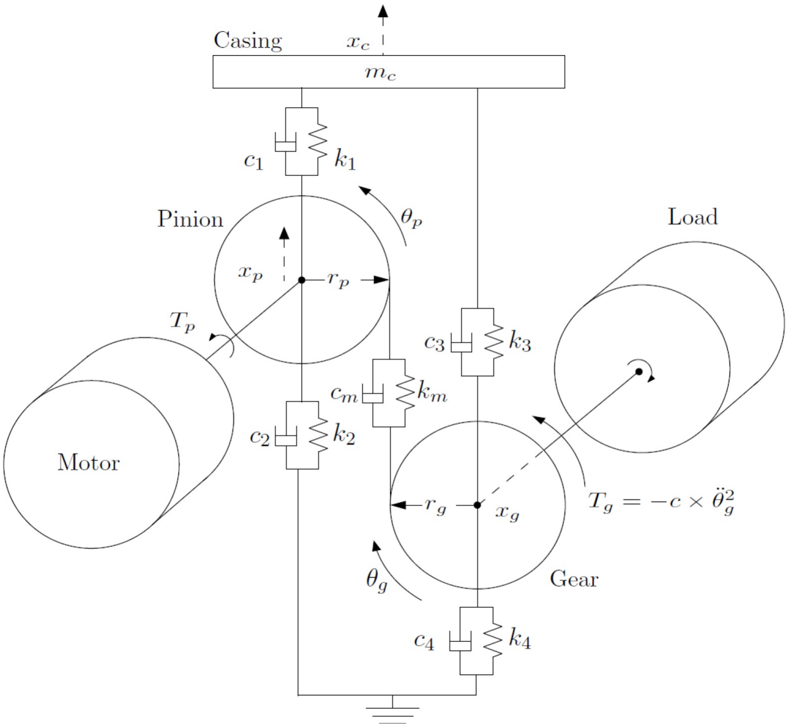


Figure 5.2: Gear model diagram

the time varying torque induces amplitude, frequency and phase modulation. The operating conditions are summarised in table 5.1.

Table 5.1: Instantaneous applied load

	Applied load
Steady state operations	$T_d(t) = 500$
Fluctuating operations	$T_d(t) = 500 \times [1 + 0.1\sin(2\pi t) + 0.03\sin(6\pi t)]$

The time domain vibration responses as simulated on the gearbox casing for fluctuating operating conditions are indicated in figures 5.3 (a) and (b). Amplitude modulation is clearly observed, and under closer zoom significant frequency modulation is also seen.

The power spectral densities (PSDs) of the gear casing vibration as respectively simulated under steady state operating conditions and fluctuating operating conditions are illustrated in figures 5.4 (a) and (b). The dotted red lines correspond to the gear meshing frequency (GMF) components at approximately 90 Hz. The GMFs along with their higher harmonics dominate the vibration signal.

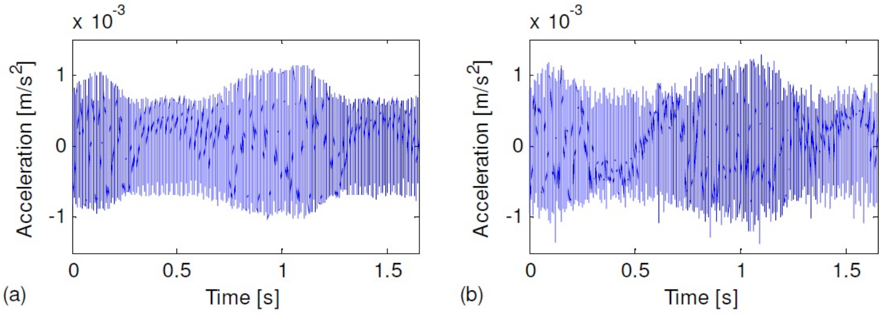


Figure 5.3: (a) time domain waveform of the healthy gearbox (fluctuating operating conditions), (b) time domain waveform of gear casing response for one damaged tooth (fluctuating operating conditions)

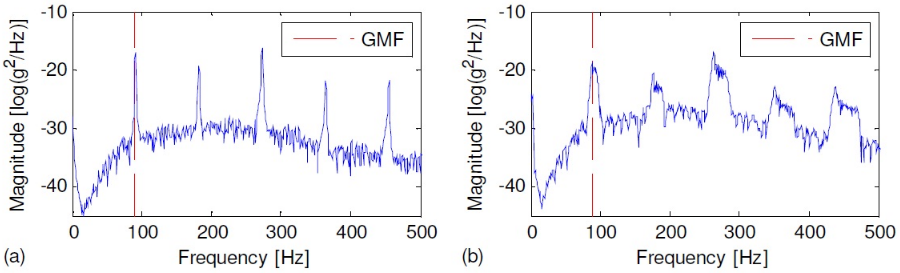


Figure 5.4: (a) PSD of gear casing response (steady state operations) (b) PSD of the gear casing (fluctuating operating conditions)

Figure 5.4 (b) also indicates how the higher order harmonics are increasingly prone to spectral smearing due to the frequency modulation.

5.4 IMPLEMENTATION AND RESULTS

5.4.1 Residual signals

Both the AR and the NN filters were implemented using model orders of 20. Separate models were trained for the stationary and the time-varying operating conditions.

The results for the stationary operating conditions are not illustrated, but both the AR and the NN residual signals clearly indicate the presence of the fault induced signal deviations.

The highly non-linear signals as simulated for the fluctuating operating conditions are next investigated. Figures 5.5 (a) and (b) illustrate the residuals computed for the damaged gearbox; respectively using the AR and NN filters. The AR model completely fails to detect the fault induced outliers, while the NN successfully detects them. Figures 5.5 (c) and (d) respectively illustrate the envelopes as computed for the obtained AR and NN residuals.

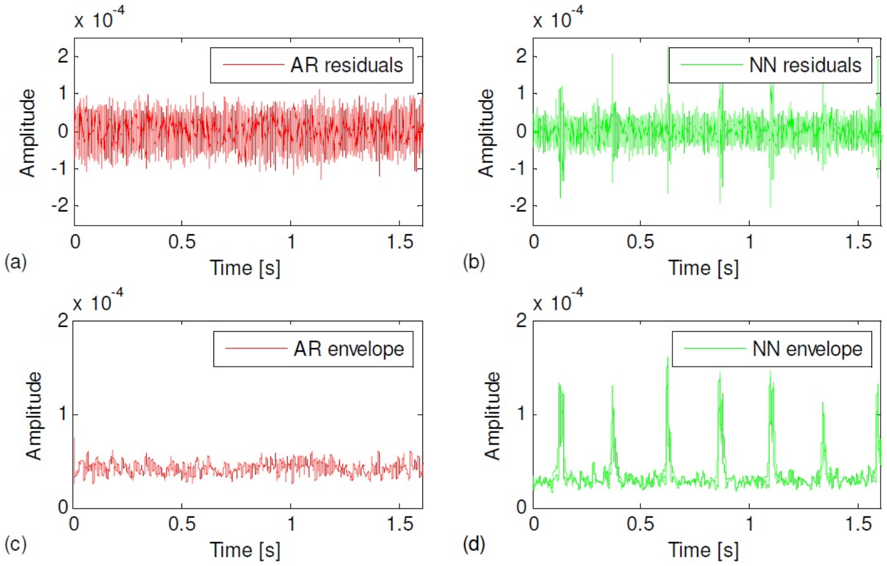


Figure 5.5: (a) AR residuals for damaged gearbox (fluctuating operating conditions) (b) NN residuals for damaged gearbox (fluctuating operating conditions) (c) envelope of AR residuals (d) envelope of NN residuals

It should be noted that the amplitudes of the impulses indicated in figure 5.5 are fairly consistent. In some scenarios it may be expected that the fault induced amplitudes may be more stochastic due to the physical mechanism of the damage, or be more load dependent. In those cases both envelope spectrum and cepstrum will be prone to greater inaccuracies.

5.4.2 Sensorless relative speed estimation

The relative speed is subsequently estimated based on the spectrogram of the vibration signal.

Figure 5.6 (a) illustrates a contour plot of the local maxima as estimated from the spectrogram of the vibration signal where the damaged gearbox was simulated under fluctuating operating conditions.

Figure 5.6 (b) indicates the local maxima which were clustered together based on the distances (modulus of the time, frequency and amplitude differences) between neighbouring points. Beginning at a manually selected point ([0s, 280Hz]), the clustering algorithm is used to group together all the local maxima which are related to the 3rd GMF harmonic.

Figure 5.6 (c) plots both the true values (continuous blue line) which correspond to the 3rd GMF

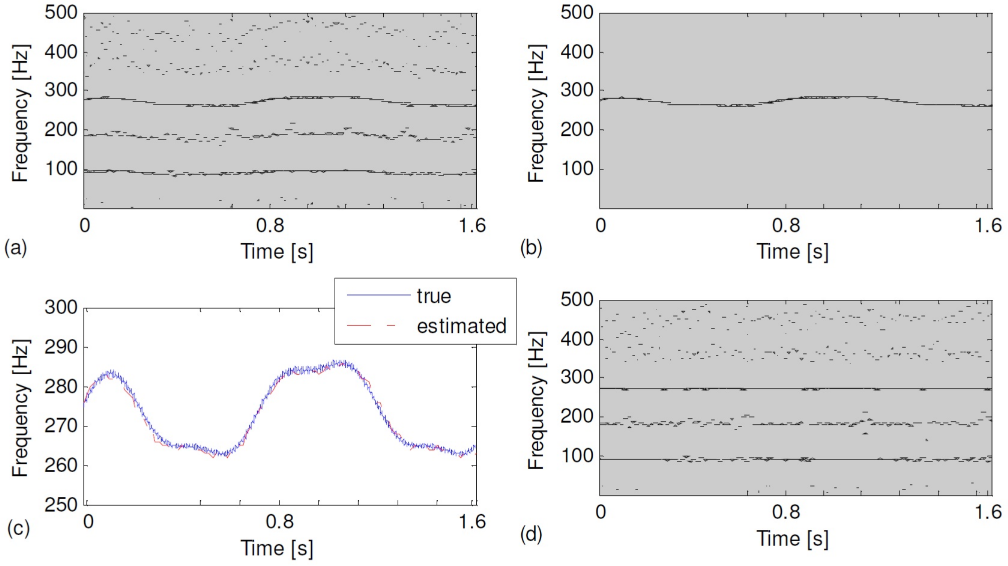


Figure 5.6: (a) contour plot of the local spectrogram maxima (b) the clustered local maxima which corresponds to the third harmonic of the GMF, (c) the estimated and true instantaneous values of third harmonic of the GMF, (d) contour plot of the local spectrogram maxima of the resampled vibration signal

harmonic (computed based on the true shaft angular velocity) as well as the values estimated based on the spectrogram (dashed red line). The estimated values are obtained by low-pass filtering the clustered local maxima. The estimated shaft angular speed corresponds fairly well to the global trends of the true angular speed, however it fails to detect the rapid speed fluctuations (shaft jitter).

Figure 5.6 (d) illustrates the contour plot of the local maxima for the resampled vibration signal. The low frequency components are now fairly well corrected for, however some of the higher order frequency components (e.g. at approximately 360 Hz and 450 Hz) are still significantly smeared.

Figures 5.8 (a) and (b) respectively illustrate the PSDs of the original vibration waveform and the PSD of the resampled vibration waveform. When these spectra are compared to the undamaged PSDs very little to no difference could be discerned in the amplitude of the gear meshing component (91Hz or 22nd pinion shaft order). No significant side bands could be detected around the GMF at intervals which correspond to the pinion shaft frequency.

Figures 5.8 (a) and (b) respectively illustrate the PSDs of the original discrepancy signal and the PSD of the resampled discrepancy signal of the damaged pinion. A clear difference in magnitude at the first pinion shaft order (4.1 Hz) component is observed when these PSDs are compared to their counterparts for the undamaged gearbox.

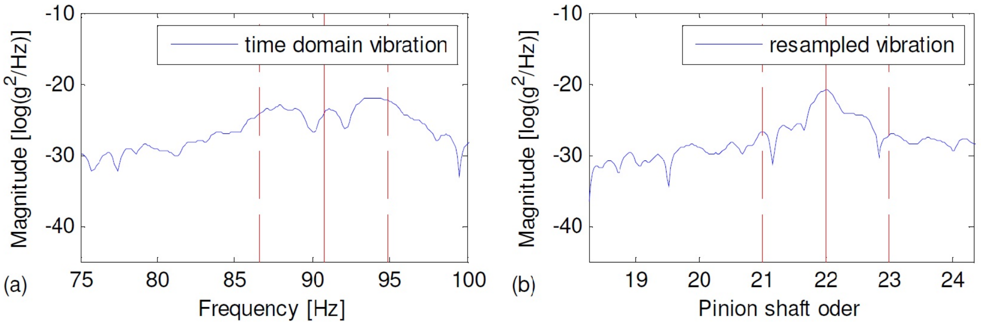


Figure 5.7: the PSDs for the damaged gearbox (fluctuating operations) for (a) the raw vibration waveform, (b) the resampled vibration waveform. The continuous red line indicates the meshing frequency while the dashed red lines correspond to the expected side bands which correspond to the pinion shaft rotational frequency

The discrepancy signal is impulsive in nature and thus not well suited to frequency analysis which assumes sinusoidal base functions. Spectral analysis of an impulsive waveform tends to give rise to a spectrum which exhibits multiple harmonics which are spaced at regular intervals. These intervals correspond to the periodicity of the impulses. Cepstrum analysis is efficient in collecting those families of harmonics and representing it in a concise manner.

Figures 5.9 (a) represents the real cepstrum of the NN discrepancy signal where the damaged gearbox was subjected to fluctuating operating conditions. Energy is observed at approximately 0.24s, which corresponds with the rotational period of the pinion. Figures 5.9 (b) represents the real cepstrum of the resampled NN discrepancy signal. The unit of the transform is referred to as derors (based on the cepstrum convention of frequency-quefrency, and order-deror). Significant energy is clearly observed at 1 deror, which corresponds with the rotational period of the pinion.

As an interesting side note it might be mentioned that smaller levels of pinion tooth damage generally tends to be more visible in the PSD of the resampled discrepancy signal than it is in its cepstrum. For this reason it is advised that both the spectra and the cepstra of resampled discrepancy signals should be monitored.

5.5 CHAPTER CONCLUSION

In this chapter it was proposed that discrepancy (envelope residual) analysis and approximate signal resampling may jointly be used to monitor the condition of rotating machines that are subject to fluctuating operations and where an accurate measurement of the shaft angular position is not available.

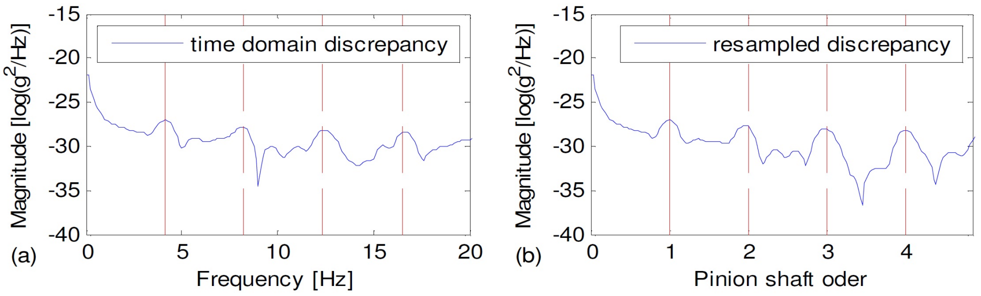


Figure 5.8: the PSDs for the damaged gearbox (fluctuating operations) for (a) the raw vibration waveform, (b) the resampled vibration waveform. The continuous red line indicates the meshing frequency while the dashed red lines correspond to the expected side bands which correspond to the pinion shaft rotational frequency

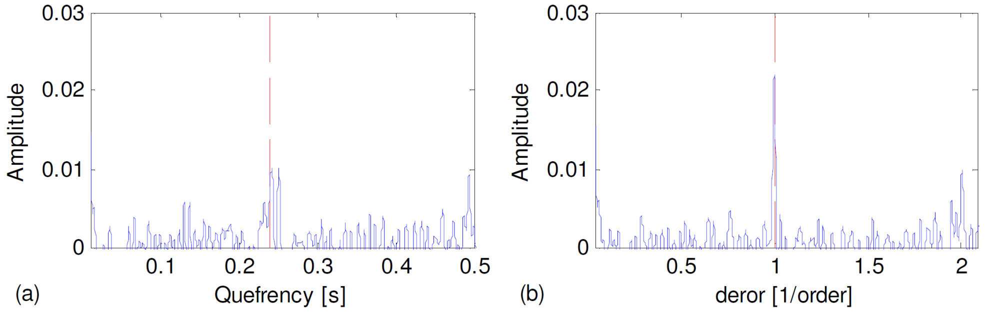


Figure 5.9: (a) cepstrum of the NN discrepancy signal (fluctuating operations) (b) cepstrum of the resampled NN discrepancy signal (referred to as deror)

The proposed methodology firstly employs a neural network to remove the regular components from the time-varying vibration waveform. The envelope of the residual signal is subsequently computed so as to obtain a discrepancy signal. It is argued that the discrepancy signal in generally more informative and is also simpler (smoother) than the original waveform. This renders the discrepancy signal less sensitive to noise in the shaft angular velocity reference signal. It subsequently becomes more viable to use a shaft speed reference speed that has been estimated from the vibration signal to resample the discrepancy signal. Positive results were obtained on data that were simulated by using a simple gear model.

Chapter 6

Neural network based waveform reconstruction for an alternator

6.1 INTRODUCTION

This chapter introduces the waveform reconstruction methodology. The waveform reconstruction methodology differs from the discrepancy transform in the sense that more information (sensors) is available so that the operating conditions are no longer latent. This renders it possible to construct regression models of the machine response waveform, and subsequently it becomes possible to manipulate the waveform, for instance by reconstructing it at a constant (standardised) steady-state operating condition. This reconstructed waveform may then be well suited to analysis by means of more conventional signal processing techniques.

The feed forward neural network (FFNN) based waveform reconstruction methodology proposed in this chapter is applied to vibration data that were measured on an automotive alternator laboratory test rig. Artificial damage was seeded in the alternator by scratching resin off from the outer isolation layers of the stator windings. Measurements were taken during the start up of the alternator, such that the alternator experienced significant speed fluctuations. It is demonstrated how the waveform reconstruction methodology may successfully be applied to the data so as to solve a number of potentially challenging signal decomposition and reconstruction problems.

In its simplest form the proposed methodology may be used in a manner similar to the work by Mdlazi et al. [34] namely to extract a specific deterministic signal component (e.g. the synchronous average). In the here proposed methodology, the FFNN may also be extended so as to model the complex interaction between time-varying operating conditions, and multiple (possibly asynchronous) signal components. Knowledge of the interaction between the operating conditions and the components may be used to reconstruct a waveform with desirable properties. The reconstructed waveform may for instance only contain specific signal components of interest (such as the gear mesh frequency and its harmonics), and may also be simulated at desirable operating conditions, such as a standardised steady state operating condition. The reconstructed waveform may subsequently be suitable for more conventional signal processing techniques, such as Fourier analysis. This chapter also briefly investigates the possibility of using the FFNN in a recursive manner to first model the first order cyclo-stationary signal components (means), and then subsequently the higher order cycle-stationary characteristics of the vibration signal (e.g. signal variance).

6.2 METHODOLOGY

It is proposed that a flexible regression function, as realized by means of a feed forward neural network (FFNN), may be used to approximate the deterministic and potentially also the stochastic components in a vibration signal which exhibits first or higher order cyclo-stationary characteristics. The FFNN may also be applied when a signal exhibits quasi-cyclo-stationary characteristics, meaning that the cyclo-stationary is only distorted due to the presence of time-varying operating conditions. Essentially the neural network is used to realise a function $f(\mathbf{x}_t)$ which approximates the magnitude of the vibration signal y_t at time instant t:

$$y_t = f(\mathbf{x}_t) + n_t \tag{6.2.1}$$

where $\mathbf{x}_t$ is the input vector which represents the time-varying operating conditions and wrapped phase angles at time instant t. A stochastic (noise) component which is assumed to be sampled from a Gaussian distribution is represented by n_t.

The input vector dictates the richness of the signal model. In the simplest case a FFNN may be used to model the deterministic component of the signal that is synchronous with a specific shaft angular position. This is realised by simply using the wrapped phase of the order of interest θ_t as the neural network input. The FFNN generates a parametric representation of the conventional synchronous average. The FFNN may also be extended to not only model a single order of interest, but rather to model m multiple (potentially asynchronised) order components $[\theta_t^1, \theta_t^2, ..., \theta_t^m]$. It has been found that the required complexity of the neural network architecture may occasionally significantly be reduced if higher order phase information (for instance the gear mesh frequency) is included in addition to the base order of interest (such as the shaft angular position).

The neural network may also be trained to account for the nonlinear effects (such as amplitude or phase modulation) which time-varying operating conditions may have on the vibration signal. This is done by appending the input vector with any n variables $[c_t^1, c_t^2, \ldots c_t^n]$ which represent the instantaneous operating conditions, such as the shaft angular speed or the applied torque.

Depending on the nature of the input vector the proposed neural network signal model may be employed towards a number of different goals. The uses explored in this chapter include:

- generating a parametric representation of the synchronous average
- modelling the vibration waveform as a function of different order components and operating conditions
- using the residuals to estimate the non-constant signal variance
- reconstructing the vibration waveform under steady state operating conditions which allows for subsequent analysis with conventional signal processing techniques such as Fourier analysis
- extracting and normalizing individual components of interest

6.2.1 Time synchronous averaging

The time synchronous average (TSA) is computed as the ensemble average of the time domain signal over a number of periods. This reduces noise and non-synchronous signal components, hence enhancing the signal characteristics of interest. The conventional TSA is defined as [1]:

$$\bar{y}_t = \frac{1}{N}\sum_{n=0}^{N-1} y(t+nT),\ 0 \leq t \leq T, \tag{6.2.2}$$

where y_t denotes the signal, T is the averaging period and N is the number of samples for averaging.

When the signal is subject to frequency modulation due to fluctuating operating speeds the signal must first be converted to the order domain, so that order domain synchronous averaging (ODSA) can be employed. Computed order tracking may be implemented by interpolating the signal to constant degrees of the unwrapped phase of the order of interest. It is typical to resample the signal with regards to the shaft angular position as estimated by means of a tachometer signal [43].

6.2.2 The feed forward neural network

The popular feed forward neural network (FFNN) is employed in this chapter as the regression function of choice. In particular a FFNN is implemented that comprises a single hidden layer with nodes which implement sigmoid activation functions, and an output layer with a single node which implements a linear activation function. The basic architecture of the FFNN is illustrated in figure 6.1. Such a neural network, if implemented with sufficiently many nodes in its hidden layer, has been shown capable of approximating any practical continuous function to arbitrary precision [90].

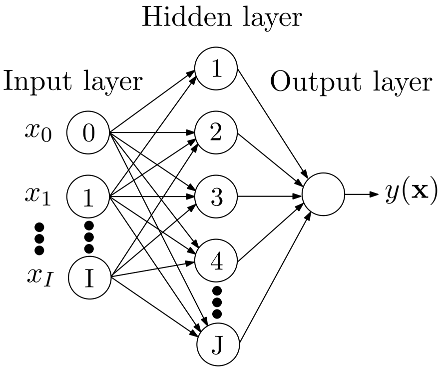


Figure 6.1: Architecture of a simple feed forward neural network.

It is envisaged that alternative regression function methodologies, including support vector regression and Gaussian processes may also be considered since these kernel based methods tend to generalise very well [90]. Gaussian processes may especially be of interest when limited data are available. The choice for implementing a FFNN in this study is based on its ease of implementation, relative compact parametric representation, and computational efficiency at being trained on large data sets by means of backpropagation optimization.

Let $f(\mathbf{x})$ be the nonlinear function represented by the neural network, so that $\mathbf{x}$ denotes the input vector of independent variables. The independent variables may include wrapped phase angles of interest, and variables representing time-varying operating conditions. A constant value of one is appended to the input vector to account for biases. As mentioned the neural network comprises an input layer which corresponds to an input vector of length I, a hidden layer with J nodes, and an output layer with a single node. The nodes in the hidden layer implements sigmoid activation functions, while the activation function in the output node is linear. Let the sigmoid activation function of the nodes in the hidden layer be denoted by $g(q) = \frac{2\times 1.716}{1+e^{-0.667\times q}} - 1.716$, where q is simply the sum of all the inputs to a hidden node. The sum of the outputs from the hidden layer nodes r is passed though the output layer with the single node with a linear activation function $r = f(r)$.

The complete FFNN may subsequently be summarized:

$$\bar{y}_t = f\left(\sum_{j=1}^{J} w_{kj}\cdot g\left(\sum_{i=1}^{I} w_{ji}x_i + w_{j0}\right) + w_{k0}\right) \tag{6.2.3}$$

where I is the length of the input vector, J is the number of nodes in the hidden layer, w_{j0} and w_{k0}

are the bias weights, w_{ji} is the weight that connects the $i'th$ input to the $j'th$ hidden node, and w_{kj} is the weight that connects the $j'th$ hidden node to the $k'th$ output layer. Since only one output layer is used to represent the magnitude of the signal, $k = 1$. The network is trained by means of the backpropagation algorithm which adjusts the weights of the networks to minimize the sum squared error (SSE) between the predicted values of the network $\hat{\mathbf{y}}$ and the target values $\mathbf{y}$.

The training set which consists of K input vectors is contained in the matrix $\mathbf{X} \equiv \{\mathbf{x}_k\}_{k=1}^{K}$ and the corresponding K target values are contained in the vector $\mathbf{y} \equiv \{y_k\}_{k=1}^{K}$.

Model complexity is empirically investigated by means of cross validation. The FFNN is implemented in Matlab by using the Neural Network toolbox.

6.3 SIMULATED CASE STUDY

A simple case study is briefly considered towards the aim of illustrating the ability of a neural network to extract a deterministic signal component from a noisy signal. A steady state vibration like signal is first investigated. A deterministic signal component is generated as the sum of a square wave of unit frequency and a smaller secondary square wave which has a relative frequency of 4. The signal is sampled at 100 samples for each cycle of the main (unit periodicity) square wave:

$$y_t = r(2\pi t) + 0.5 \times r(4 \times 2\pi t) + n_t \tag{6.3.1}$$

where the square wave is defined $r(\cdot) = \text{sgn}(\sin(\cdot))$. An additive noise process is represented by n_t, where n_t is independently sampled from a zero mean Gaussian distribution with a signal variance of 0.25, so that $n_t \sim \mathcal{N}(0, 0.25)$.

The conventional ODSA and the FFNN estimates are computed. The performance of each method is computed as the mean square error (MSE) between the estimated deterministic signal and the true deterministic signal. The relative performance of two different input vectors are investigated. The first input vector only consists of the unwrapped phase (in rectangular form) of the base periodicity, $\mathbf{x}_t = [\sin(2\pi t) \quad \cos(2\pi t)]$, while the second imput vector also contains phase information regarding the second (smaller) square wave which has a relative periodicity of 4 $\mathbf{x}_t = [\sin(2\pi t) \quad \cos(2\pi t) \quad \sin(4 \times 2\pi t) \quad \cos(4 \times 2\pi t)]$. Rectangular form coordinates are always employed, since it avoids sharp discontinuities in the FFNN input vector. The second input vector will be used to explore the assumption that high frequency components which to some extent are repeated during a shaft rotation, may be modelled in a more compact manner by appending its phase information to the FFNN input vector.

It is found that for limited data (typically between 1 to 5 rotations of the base frequency) the neural network slightly outperformed conventional ODSA. However in general (typically for 1-100 rotations) the ODSA and the neural network estimates compared well. As a rule it is found that the performance of the neural network was not very sensitive to the exact number of nodes used in the hidden layer. For limited data (typically 1 to 5 rotations) best results are obtained with few hidden nodes (approximately 5). Using only a few nodes allows for better generalisation and hence tends to be more robust to noise. More data favoured a more flexible FFNN architecture (more hidden nodes), which allowed for better approximation of the actual waveform. In general good results were obtained with 25 hidden nodes. The richer input vector (which also provided information regarding the wrapped phase of the square wave with a relative periodicity of 4) tended to produce better neural network estimates, even for networks with comparatively fewer nodes in the hidden layer.

A second scenario is also investigated which represents non-steady state operating conditions. A deterministic signal component is simulated where a time-varying component (which is non-synchronous with the shaft frequency) induces amplitude modulation in the signal:

$$y_t = v_t \times [r(2\pi t) + 0.5 \times r(4 \times 2\pi t))] + n_t, \tag{6.3.2}$$

where $v_t = 1 + 0.5 \times \sin(0.2 \times 2\pi t)$.

The time synchronous averages are once more approximated with the FFNN, using two different input vectors. The first input vector again contains the wrapped phase of the square wave of unit frequency, while the second input vector also contains the wrapped phase of the synchronous, but higher order square wave, with a relative frequency of 4. In addition, both input vectors now also contain an entry which is proportional to the component v_t, which is responsible for the amplitude modulation in the signal.

Figure 6.2 indicates the deterministic signal as estimated firstly by means of ODSA and secondly as estimated with a FFNN which contains 25 nodes in the hidden layer. The FFNN estimate performs well, while the ODSA fails to converge due to its inability to model the non-synchronous amplitude modulation.

Figures 6.3 (a) and (b) indicate how the performance of the synchronous averages and the FFNN estimates converge for increased numbers of cycles of the square wave with unit periodicity. Figure 6.3 (a) illustrates the results from the steady state waveform, while (b) presents results for the scenario with the time-varying amplitude modulated waveform. The synchronous averages are represented by

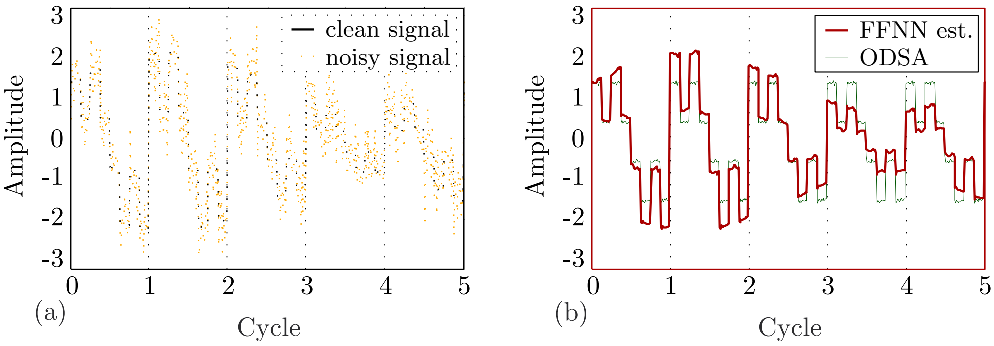


Figure 6.2: (a) The deterministic (noise free) signal and the noise contaminated signal (b) the ODSA fails to model the asynchronous load-induced amplitude modulation, while the neural network manages to represent it well.

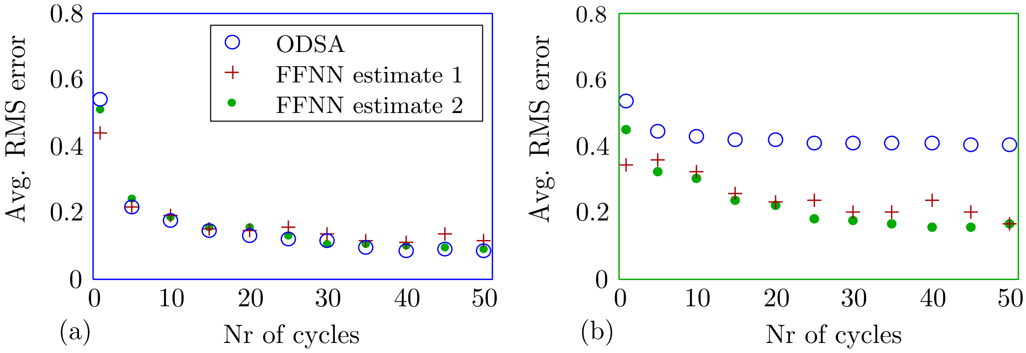


Figure 6.3: (a) Relative convergence (over increasing numbers of repititions or cycles) for the ODSA and the two FFNN estimates for the simulated steady state vibration like waveform (b) Relative convergence for the ODSA and the two FFNN estimates for the simulated amplitude modulated vibration like waveform.

the open circles, while two different NN estimates are represented by pluses and dots. The pluses are the FFNN estimates based on the more limited input vector, while the slightly better performing dots are the FFNN estimates based on the extended inputs which contains the phase information of the synchronous square wave with a periodicity of 4. The FFNN estimates clearly outperform conventional synchronous averaging for the non-steady state scenario.

6.4 EXPERIMENTAL SETUP

Wang and Heyns [9] constructed a laboratory-scale test rig of an automotive alternator. The experimental study was conducted towards the aim of investigating the combined use of Vold-Kalman filtering and Fourier analysis on vibration signals which were generated under significant speed fluctuations.

The automotive alternator implemented in the test rig was used to charge a 12 V car battery. A Saftronics variable speed induction motor (with a maximum speed of 1750 rpm) was used to drive the alternator. The induction motor was connected to the alternator by means of a V-belt. The motor in turn was controlled by a Saftronics DC controller (with a rated input of 220 V, 15 A and with a rated output of 180 V, 12 A). The alternator has 36 stator bars and 12 claws on the rotor. The experimental setup is depicted in figure 6.4 (a), while figure 6.4 (b) illustrates the stator and the exposed end windings.

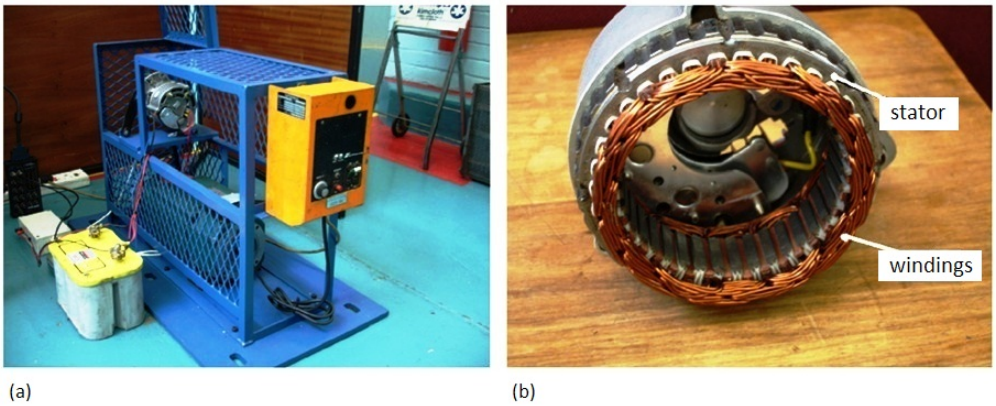


Figure 6.4: (a) The experimental setup and (b) the stator displaying the end windings.

The alternator response signal was measured by means of a PCB Piezotronics M353B18 accelerometer which has a sensitivity of 10.13 mV/g and a frequency range 1 Hz-10 kHz. The shaft rotational speed was measured with a Hengstler RI76TD shaft encoder which produced a single pulse per revolution. The data were recorded on a National Instruments board (NIBPCI-6024 E) at a sampling frequency of 10 kHz.

Artificial damage was seeded in the alternator by scratching some of the resin off from the outer isolation layers on the stator windings. This introduced an inter-turn short circuit between the windings. The accelerometer was mounted on the surface of the winding bars, so that the seeded fault could be expected to excite frequency orders which are multiples of 36 (number of winding bars). The 72nd order is of particular interest with regard to the presence of the turn-to-turn fault [9].

The vibration data were measured as the induction motor shaft rotational speed accelerated from approximately 6 rev/s up to slightly more than 23 rev/s. Figures 6.5 (a) and (b) respectively indicate the shaft rotational speed and acceleration as measured on a typical run.

The operating conditions for the two measurements differ slightly. Figures 6.6 (a) and (b) indicate

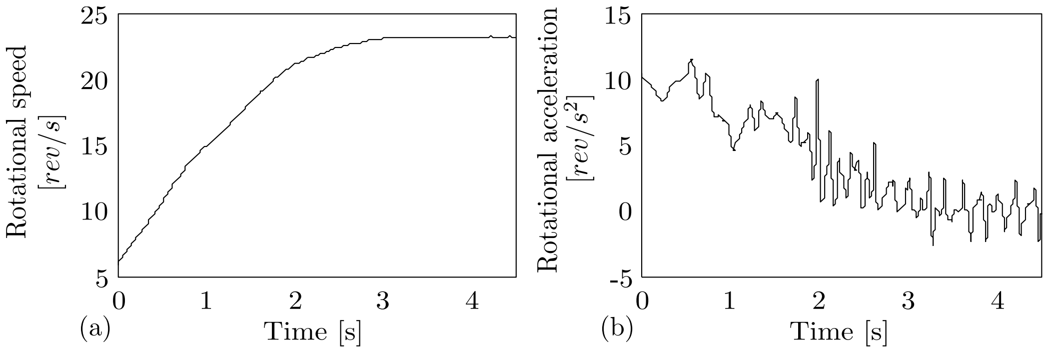


Figure 6.5: The alternator shaft rotational speed $[rev/s]$ and acceleration $[rev/s^2]$ as measured on a typical run.

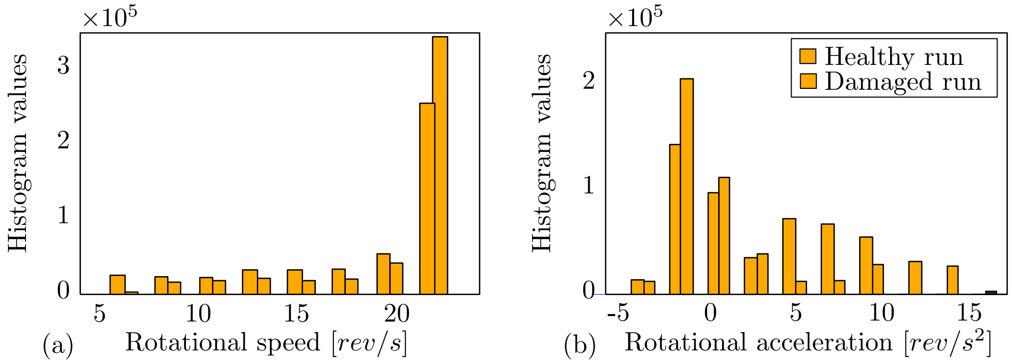


Figure 6.6: Histograms of the alternator shaft rotational speed $[rev/s]$ and acceleration $[rev/s^2]$ indicate differences in the operating conditions of the two runs.

with histograms the relative time which each of the vibration measurements were respectively subject to different shaft velocities and accelerations. While it is clear that the operating conditions influence the signal characteristics, it is at this stage still unclear how significant these effects would be on global characteristics of the signal (such the spectra of the signals) and how the differences in the operating conditions would influence a comparison between the two vibration signals.

The spectra of the raw vibration signals are illustrated in figures 6.7 (a) and (b), where (a) indicates the square root of the power spectral density (PSD) for the signal from the healthy alternator, and (b) depicts the square root of PSD for the signal from the damaged alternator. The square roots of the PSD are presented to allow for better scaling and visualization. Peak frequencies which correspond to the AC line current of 50 Hz are observed at 50 Hz, 100 Hz and 150 Hz. The second harmonic at 100 Hz is especially prominent.

Significant activity is also observed at approximately 833 Hz, which is related by a multiple of 36

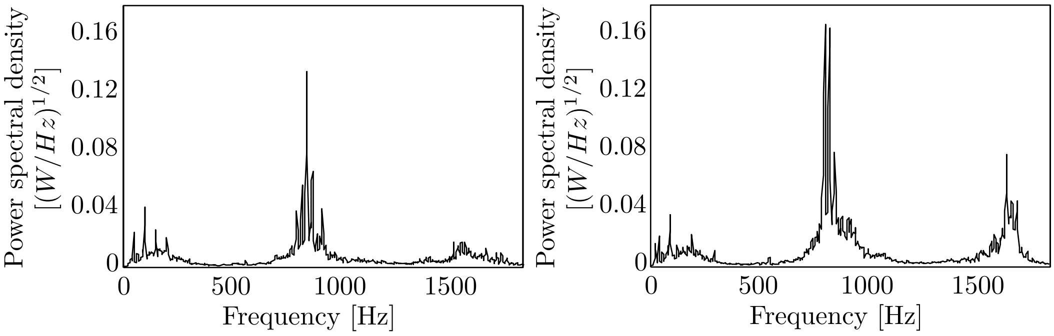


Figure 6.7: The square root of the power spectral densities as estimated from the raw vibration signals for (a) the healthy stator windings and (b) the seeded inter-turn short circuit.

to the shaft rotational speed of approximately 23 revolutions per second. The significant response at the 36th order is excited due to the corresponding number of stator bars. Aside from the general smearing which is the result of the varying shaft speed, prominent sidebands are observed around the 833 Hz component. The sidebands are induced by amplitude modulation and are spaced at near 23 Hz intervals around the centre frequency. This indicates that the amplitude modulating component is synchronous with the shaft angular position. Energy is also observed at the second harmonic of the 833 Hz component. The energy at around 1666 Hertz (or the 72nd order) is again subject to amplitude modulation which results in sidebands which are also spaced at 23 Hz intervals. Both the 833 Hz, and especially the 1666 Hz components are substantially greater in figure 6.7 (b) compared to 6.7 (a) due to the presence of the inter-turn short circuit.

6.5 IMPLEMENTATION AND RESULTS

The FFNN waveform reconstruction methodology is investigated on the experimental data. A number of different scenarios are considered, namely (i) where the FFNN is used to estimate the synchronous average, (ii) where the FFNN input vector is extended so that the operating conditions are also modelled, (iii) where the FFNN is used to model the non-constant signal variance, and (iv) where the FFNN is used to extract and normalize individual signal components of interest.

6.5.1 Synchronous averaging

The simplest implementation of the neural network waveform reconstruction methodology is to approximate the conventional synchronous average. This is done by approximating the waveform as a function of a number of wrapped phases of interest.

In principle the synchronous average with respect to the shaft angular position may be estimated by setting the FFNN input vector equal to the polar coordinates of the wrapped phase $\phi_t^s \in [0;2\pi]$ of the shaft (as measured with the tacho signal), so that $\mathbf{x}_t = [\sin(\phi_t^s) \quad \cos(\phi_t^s)]$. The target vector is the measured vibration signal y_t. Since the neural network approximates a continuous representation that express the waveform as a function of the wrapped shaft angular position it is not required to resample the signal.

If the waveform has high frequency content, then a very flexible neural network is required to accurately approximate the synchronous average. Investigations on the experimental data shows that a very flexible neural network (many unknown parameter values), combined with the fairly high signal noise, tends to result in poor performance during optimization of the FFNN. A significantly simpler network architecture may successfully be used if the input vector is appended with the wrapped phase information which corresponds to important high order signal components. A high order frequency component (such as the 36th order) may often be approximated by a single short waveform, which is repeated multiple (e.g. 36) times over the duration of a single shaft rotation. Without the additional higher frequency phase information, each of the repetitions of this simple waveform must be treated as unique. The higher order phase information allows the FFNN to recognize the repetitive nature of the waveform. The measured vibration waveform may thus be modelled in a more compact manner if the wrapped phase information ϕ_t^{36} is appended to the FFNN input vector so that $\mathbf{x}_t = [\sin(\phi_t^s) \quad \cos(\phi_t^s) \quad \sin(\phi_t^{36}) \quad \cos(\phi_t^{36})]$.

Table 6.1: Synchronous average performance for different modelling assumptions.

Implemented model	residual RMS
Constant mean value (standard deviation)	0.58
Order domain synchronous average	0.42
FFNN estimate with first order wrapped phase	0.57
FFNN estimate with first and 36th order wrapped phases	0.42

Table 6.1 indicates the performance of different synchronous averaging modelling strategies as evaluated on the data from the alternator in good condition (smaller is better). Performance is measured as the root mean square (RMS) of the residuals, where the residuals are computed as the differences between the models and the measured signal. In the first modelling scenario the signal is simply approximated by its (constant) mean value. No dependence on any wrapped phases are thus considered. The RMS of the residuals are equivalent to the standard deviation of the signal. The second model

is computed as the order domain synchronous average. The third model only considers the wrapped phase (ϕ_t^s) of the shaft angular position, while the fourth model implements a FFNN input vector which contains both the wrapped phase information from the 1st and 36th orders ((ϕ_t^s) and (ϕ_t^{36})). The importance of appending the wrapped phase information from the 36th order is clearly observed. If the 36th order is omitted the FFNN fails to extract the high frequency deterministic content in the signal. This results in poor performance which is essentially equal to the standard deviation of the signal. By appending the 36th order phase information to the input vector the FFNN becomes capable of estimating a synchronous average with a near identical performance to that of the conventional synchronous average.

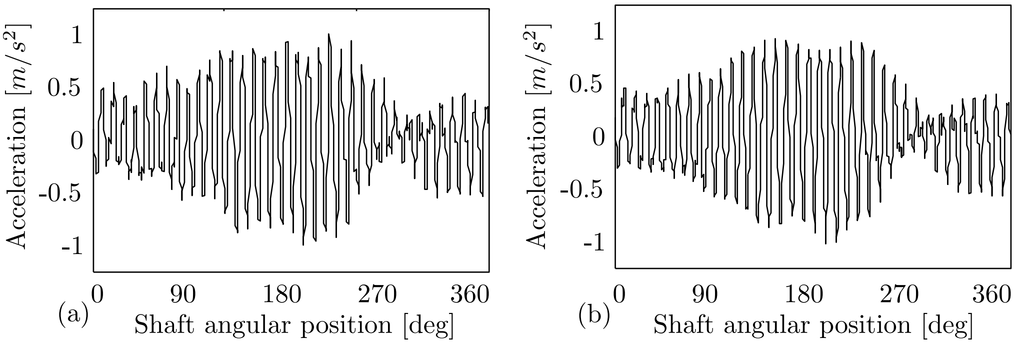


Figure 6.8: Synchronous averages as computed (a) with the conventional method of signal resampling (ODSA) and (b) as estimated from the original signal by means of the neural network with the input vector which constains information of the wrapped phases from both the 1st and the 36th orders.

Figure 6.8 (a) illustrates the conventional synchronous average as estimated from a order domain resampled signal, while (b) is the neural network model estimation. The implemented neural network had 25 hidden nodes, and the input vector consistent of the wrapped phase information of the 1st and the 36th orders. Unlike the conventional synchronous average where each phase increment is estimated independently, the parametric FFNN imposes functional continuity. This renders the FFNN smoother than the ODSA. This smoothness may result in better generalization and better robustness to noise.

6.5.2 Extended modelling of the deterministic signal components

It is subsequently investigated how the FFNN input vector may be extended so that the non-linear waveform characteristics as induced by the time-varying operating conditions may also be modelled. It is desirable to implement a FFNN architecture which is sufficiently flexible so as to allow for accurate representation of the vibration waveform. However neural networks with fewer nodes in the hidden layer tend to result in better generalization, while also reducing the computational burden of training. Similarly may excessive operating condition representative variables in the input vector also result in a too-flexible network which is prone to data over-fitting which will result in poor estimates.

The generalization performance of the neural network is investigated based on its ability to represent a set of independent data. In this study an independent validation data set is generated by removing from the training data set all the time instances where the shaft angular speed corresponds to the interval between 21 to 22 rev/s. Those instances are then used as the validation set. Performance is measured as the RMS difference between the estimated (interpolated) waveform and the validation set. Since the maximum shaft angular speed exceeds 23 rev/s, the validation set offers a measure of indication of the ability of the FFNN to interpolate well.

Various FFNN input vectors and numbers of hidden nodes were investigated. In general it is found that the waveform estimate is not overly sensitive with regard to the number of nodes in the hidden layer. Good performance is typically obtained for 20 to 100 hidden nodes. All subsequent results presented here are obtained with 25 nodes in the hidden layer. The input vector significantly influence the FFNN performance. It is very important to provide the wrapped phase information which is synchronous with both the 1st and 36th orders (as measured with respect to the shaft angular position). It is found that appending the shaft angular acceleration to the FFNN input vector results in slightly lower interpolation performance due to overfitting. This form of overfitting is referred to as the curse of dimensionality, and is explained by the increasing sparsity of data in higher dimensions [90]. As expected the speed is a very important independent variable. It is also found that appending an additional wrapped phase at 50 Hz allows the model to capture the dynamic effects induced by the AC line current at 50 Hz. It is interesting to note that the 50 Hz phase is asynchronous with the time-varying phase of the shaft. The 50 Hz phase information also enables the FFNN to model the vibration content at the higher harmonics of the 50 Hz component, including at 100 Hz and 150 Hz.

Once the desired network architecture was determined the neural network was subsequently optimized with all the vibration data (including the validation set). The optimized model is subsequently evaluated at the exact same wrapped phase information, shaft speeds and accelerations as the original signal to allow for a visual comparison in the time domain.

Figure 6.9 (a) illustrates the original vibration signal, while (b) indicates the reconstructed waveform of the deterministic components in the signal. The deterministic signal as illustrated in figure 6.9 (b) tends to have smaller minima and maxima peak values compared to the raw signal. The root mean square (RMS) values as computed for the raw vibration signals (healthy and damaged) tend

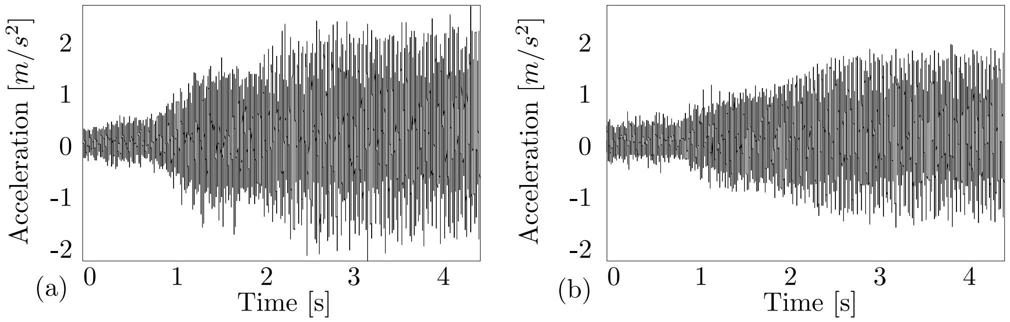


Figure 6.9: Time domain representations of (a) the measured raw signal which corresponds to the healthy stator windings and (b) the reconstructed (deterministic) signal as simulated with the healthy FFNN model at the exact same operating conditions as that of the raw signal.

to be approximately 10% greater than the RMS values for the corresponding deterministic signals. A spectral analysis of the residuals (the differences between the raw signal and the deterministic components) however indicate that the residuals are fairly uniformly distributed over the frequency axes similar to typical white noise.

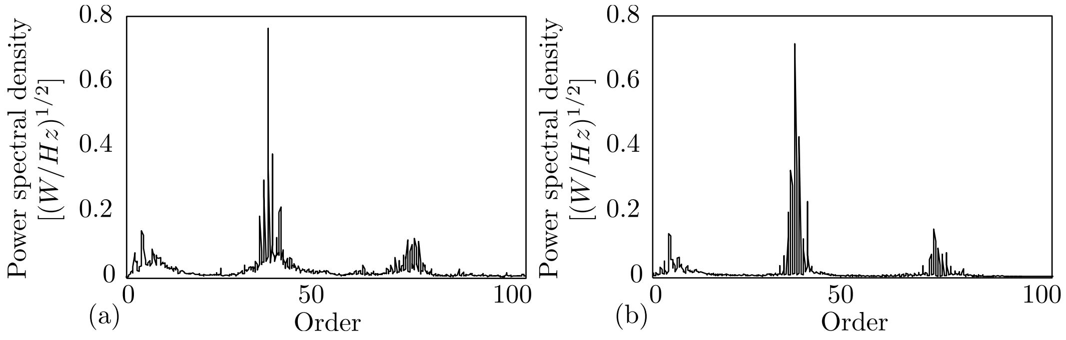


Figure 6.10: The square root of the PSD for the resampled order domain representations of (a) the measured raw signal which corresponds to the healthy stator windings and (b) the reconstructed (deterministic) signal as simulated with the healthy FFNN model at the same operating conditions as (a).

Figure 6.10 compares the spectra (square root of the PSD) as computed on the resampled (order domain) signals for (a) the measured signal and (b) the reconstructed signal. Generally the FFNN reconstructed signal corresponds well to the original signal. The reconstructed signal does however tend to be less smeared with more definite peaks, especially around the dominating 36th order. This lack of smearing on the reconstructed waveform is due to the modelling assumptions imposed by the input vector. The sidebands around the 36th order is evenly spaced at intervals of one, indicating modulation between the base frequency (shaft rotational frequency) and the 36th order.

A neural network is similarly estimated for the vibration signal which is representative of the damaged alternator end windings. The two FFNN models may be used to reconstruct the waveforms at standardised operating conditions, such as a specific (constant) speed. This renders it possible to compare the two waveforms in a more consistent manner.

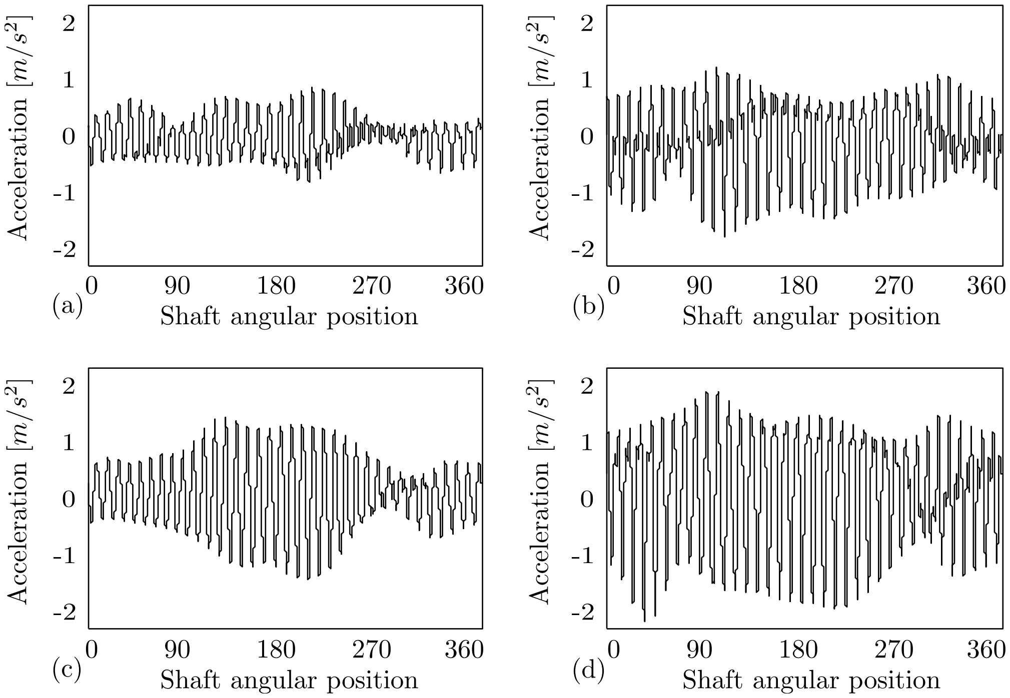


Figure 6.11: (a) The waveform from the healthy alternator reconstructed a consistent shaft rotational speed of 22 rev/s (b) waveform from the damaged alternator reconstructed a consistent shaft rotational speed of 22 rev/s (c) waveform from the healthy alternator reconstructed at a consistent shaft rotational speed of 23 rev/s (d) waveform from the damaged alternator reconstructed a consistent shaft rotational speed of rev/s

Figures 6.11 (a)-(b) depict the vibration waveforms for (a) the healthy and (b) the damaged alternator as reconstructed with a shaft rotational speed of 22 rev/s, while figures (c) and (d) respectively illustrate the reconstructed waveforms for the healthy and damaged alternator as reconstructed with a shaft rotational speed of 23 rev/s. It is clearly observed that the speed significantly influences the signal characteristics.

6.5.3 Modelling the signal variance

Often it is not only the deterministic signal components, but also the stochastic signal components which might be of interest. Randall [42] discusses that a signal may have higher order cyclo-stationary characteristics if its *n*th order statistics are periodic. A second-order cyclo-stationary signal for instance has a periodic variance. Similar to the deterministic signal components can non-stationary operating conditions also influence the characteristics of the stochastic signal components.

Conceptually it is possible to use the same neural network regression methodology to estimate certain stochastic characteristics of a vibration signal. In particular the signal variance is investigated here.

Consider that if a random variable z has a mean (or expected) value of $\mu = E[z]$ then the variance σ^2 of z is given by $\sigma^2 = E[(z-u)^2]$. This definition may be extended to the vibration waveform y_t, where the neural network has already been used to approximate the expected (least square error estimate) $f(\mathbf{x}_t)$ of the waveform. The signal variance as a function of different phase components and operating conditions may subsequently be represented as $\sigma_y^2(\mathbf{x}) = E[y_t - f(\mathbf{x}_t)]^2$. A neural network may be used to approximate $h(\mathbf{x}) \approx \sigma_y^2(\mathbf{x})$ by setting the square of the residual as the target output for the neural network.

The process of estimating higher order signal moments is however more difficult than estimating the expected signal value (first order moment). The estimated higher order moments are unstable in the presence of outliers, and this problem tends to become more severe for ever higher orders [114]. The risk of model overfitting thus becomes more prominent for the estimation of the higher order moments.

A number of techniques, such as discarding signal outliers, selected noise addition, and employing regularization during network training [90], are employed to render the neural network estimates more robust so that the signal variance may be estimated as a function of the different phase components, and shaft angular velocities.

Similar to the validation strategy during the previous section performance is again measured on an independent validation set which was obtained by extracting all the samples which were generated under a shaft rotational speed of 21-22 rev/s from the training set.

The best performance (which also proved most stable with respect to local minimas during optimization) is obtained by simply modelling the signal variance as a function of only the shaft angular speed.

Figure 6.12 (a) illustrates a random extract of residuals which were obtained by subtracting the recon-

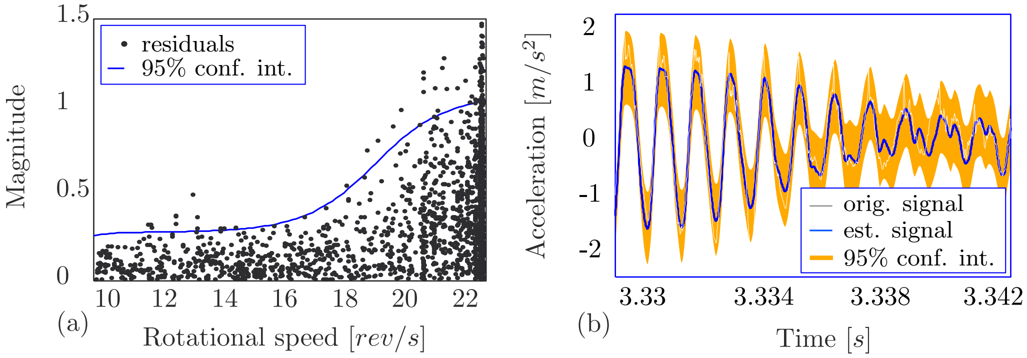


Figure 6.12: (a) Signal residuals plotted against the shaft angular speed, as well as the estimated 95% confidence interval based on the FFNN signal variance estimate (b) a short extract from a vibration signal superimposed on the corresponding (estimated) deterministic component as well as the 95% confidence interval

structed waveform from the signal as measured on the healthy alternator. A clear trend is observed, namely that the residuals tend to be greater for higher shaft angular velocities. Figure 6.12 (a) also indicates the 95% confidence interval (CI) as based on the neural network estimated signal variance 95%, where CI $= f(\mathbf{x}) \pm 2 \times \sqrt{h(v)}$, where $\mathbf{x}$ is the input vector to the FFNN used to estimate the deterministic signal component, and v is the shaft angular speed used as input to the FFNN used to represent the signal variance. The signal variance neural network $h(v)$ is implemented using only 5 nodes in the hidden layer.

The 95% confidence interval provides insight regarding the accuracy of the estimated deterministic signal, and may be used to help quantify signal changes due to the onset of damage.

6.5.4 Extracting individual componets of interest

Occasionally it may be of interest to isolate specific signal components. In the experimental case study the signal components which are synchronous with the 36th order are of particular interest as indication of the inter-turn winding fault. These components are isolated by using a neural network input which only contains the wrapped phase of the 36th order and the shaft rotational speed.

Figure 6.13 (a) indicates the waveform from the healthy alternator as reconstructed at a consistent shaft speed of 22.6 rev/s, and where the input vector contains phase information regarding the 1st and 36th orders. Figure 6.13 (b) indicates the waveform from the healthy alternator as reconstructed at a consistent shaft speed of 22.6 rev/s, and where the input vector only contains phase information regarding the 36th orders. The amplitude modulation observed in figure 6.13 (a) is averaged out in

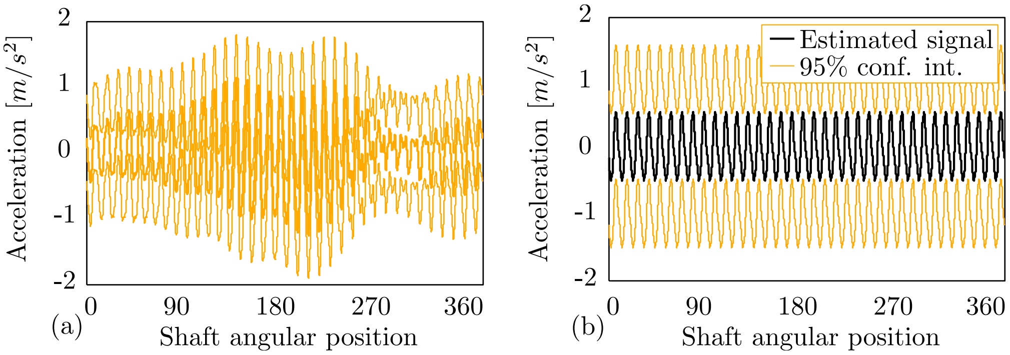


Figure 6.13: FFNN estimate reconstructed at a consistent speed of 22.6 rev/s for (a) where the 1st and 36th phase information is appended to the FFNN input vector, and (b) where only the 36th order is appended.

figure 6.13 (b).

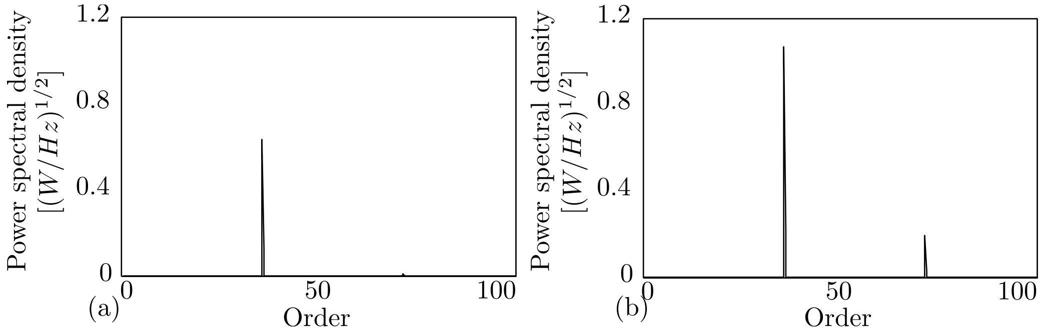


Figure 6.14: Square roots of the PSDs for (a) the healthy and (b) the broken waveforms where the signal components synchronous with the 36th order have been isolated and simulated at a constant speed of 22.6 rev/s.

Figures 6.14 (a) and (b) illustrate the PSD of the waveform as reconstructed at a consistent shaft speed of 22.6 rev/s and where only the 36th order phase information and the shaft angular speed were appended to the input vector for (a) the healthy signal and (b) the damaged alternator signal. Figure 6.14 (b) indicates the presence of energy at the 72nd order. The energy at the second harmonic occurs because the extracted components which are synchronous with the 36th order are not confined to a sinusoidal base function.

6.6 CHAPTER CONCLUSION

A waveform reconstruction methodology was proposed which implements neural network regression to model a vibration signal as a non-linear function of wrapped phases (such as the shaft angular position and its higher orders) along with the instantaneous operating conditions. The proposed methodology is simple to implement, and is not restricted to simple base functions such as sinusoids.

In this chapter an investigation was performed into how the neural network may be used to model the vibration response from an alternator which was subject to significant speed variation. It was shown how the estimated model may be used to reconstruct a waveform with desirable properties such as steady state operating conditions, or how to extract individual components of interest.

Chapter 7

Bayesian parameter estimation for monitoring a haul road

7.1 INTRODUCTION

In this chapter an empirical model is used to represent a vehicle response-type road roughness measuring system (RTRRMS) as a function of different road intervals (e.g. GPS coordinates) and of speed. This subsequently renders it possible to reconstruct the machine responses at those different road intervals at a standardised speed in order to compare the RTRRMS values in a consistent manner. In contrast to the alternator case study, this data set is extremely noisy and the data that correspond to specific road intervals are very limited. For these reasons a simple model is preferred which will generalize better using a limited amount of available data. However, adjacent road sections may exhibit very different RTRRMS values (e.g. due to a localized pothole). If a single model was used to model the entire road, then the model would have to be very flexible in order to accommodate the sharp jumps/ discontinuities of the RTRRMS values. This section proposes that it is possible to allow for these discontinuities in the measured RTRRMS values by discretizing the road into multiple sections and by fitting unique regression functions to each of the road intervals. However information is shared across the different road intervals, to partially make up for the limited data within each road interval. This is done by means of a Bayesian prior. In this chapter it is investigated how the Bayesian framework allows for the inclusion of priors which render it possible to generate regression functions which are more robust, given the limited and noisy data.

7.2 METHODOLOGY

A low cost road monitoring methodology based on Bayesian parameter estimation and regression is presented. In essence the methodology aims to fit unique regression functions to the vehicle response measurements which correspond to individual road intervals. By representing the road as a series (or grid) of individual road intervals it allows the interpolation functions to reflect the unique spectral characteristics of each specific road interval. The regression functions may subsequently be used to reconstruct the vehicle response at a standardised speed to allow for more consistent comparison.

A RTRRMS is sensitive to the characteristics of the vehicle on which it is measured. Since haul roads are typically traversed by many different vehicle types, it is proposed to only instrument and consider the responses from a subset of similar (representative) vehicles.

One of the main challenges associated with fitting unique interpolation functions to each of the road intervals is the scarcity of data. This scarcity of data may be driven by three causes; namely that some roads may be less frequented, road intervals do not have uniform spectral characteristics and as such will result in data scatter (perceived as noise), and thirdly some sections may always be traversed at lower or higher speeds than the standardised speed so that the calibrated RTRRMS estimates will depend on extrapolation rather than interpolation. In an attempt to negate these problems use is made of Bayesian regression. In the Bayesian formalism both a data driven likelihood and a prior probability are considered during parameter estimation. This prior term allows experience from other sections of road to be fused with the observations made at a specific interval of road so as to obtain more robust regression functions. It is even possible to estimate an updated regression function for a road interval where only one observation has been made. The Bayesian approach also allows for an automated model selection criterion which selects the complexity of the implemented interpolation function based on the quantity and quality of observed data.

7.2.1 Response-type road roughness measuring systems (RTRRMS)

A large number of different RTRRMS metrics have been investigated in the literature. Some of the metrics are concerned with ride comfort (human response to vibration) such as the ISO 2631 standard [21]. This metric estimates the average power absorbed by the driver and is computed based on a frequency weighted acceleration RMS value as measured on the driver seat. Other metrics may be more concerned with the vehicle-road interaction forces, such as the metric proposed by Quinn and Wilson [22] which aims to represent the dynamic tyre forces (and subsequently the road quality) based on the measured tyre pressures. The international road roughness index (IRI) was proposed as a standardised metric indication of road roughness based on a large scale study financed by the World Bank [23]. The IRI is based on the ratio of the accumulated suspension motion of a standard vehicle to the distance travelled by the vehicle during the measurement. The IRI may be estimated directly from a known road profile by simulating the response of a standardised quarter car vehicle at a constant speed of 80km/h. The IRI has widely been adapted as a standardised measure of road roughness.

It is intended that the here proposed methodology should be relevant to most RTRRMS metrics. An experimental study is conducted where this idea is investigated on a single RTRRMS metric, namely the RMS acceleration as measured on the rear shaft axle of a utility vehicle (UV). The acceleration RMS is selected as a simple indication of the severity of the vibration (and thus the load) to which the vehicle is subjected.

As with other RTRRMS metrics the RMS response is not only sensitive to the road profile and the vehicle operating conditions, but also to the characteristics of the vehicle on which it is measured. For this reason it is advised that a single representative type of vehicle should be selected for any one haul road environment. The measurements obtained from these vehicles could then be compared in a consistent manner.

The merit of different RTRRMS metrics should be investigated further with regard to the demands of large vehicles and the additional maintenance criteria which are typical to haul road environments. It is therefore not the intention of this chapter to necessarily promote the implemented RMS value. It anticipated that a study which investigates the correlation between a metric based on tyre pressures, or strain gauge measurements on specific vehicle components, and maintenance cost would be extremely interesting.

7.2.2 Bayesian parameter estimation

This section reviews the basics of Bayesian linear regression and indicates how it is implemented in the proposed methodology. The reader is referred to [90] for a comprehensive treatment on Bayesian parameter estimation.

Suppose that a grid has been defined which divides the road into various intervals. The vehicle response may be separated into segments associated with these different intervals based on a reference signal such as coordinates estimated by a global positioning system (GPS), reflective markers on the ground, or a combination of a push button and signal cross correlation alignment. The grid should be chosen such that the road condition over each section may be assumed to be fairly homogeneous.

A unique regression function (which explains the measured RTRRMS values as a function of speed) is fitted to each of the road intervals. These regression functions are estimated based on both the data driven likelihood and a parameter prior. The prior serves as indication of the typical nature of these interpolation functions. As such the prior allows for more robust interpolation functions, especially if only limited and noisy data are available.

Let the ith measured RMS value that is associated with a specific road interval j as measured at a mean speed of v_i^j over that interval be denoted by y_i^j. Some variability in the vehicle response as measured over any road interval at a given speed is expected, since the road profile spectral characteristics are not perfectly uniform. This variability is accounted for by a noise term n_i, which is assumed to have a zero mean independent identically distributed Gaussian distribution $n \sim N(0, \sigma_n^2)$ with a constant variance σ_n^2.

An operating condition vector $\mathbf{x}$ is defined and implemented so that the ith vector $\mathbf{x}_i^j$ corresponds to each RMS datum point y_i^j as measured on the jth road interval. This operating condition vector determines the flexibility of the implemented regression function. The regression function in turn reflects the expected influence of operating conditions on the RTRRMS values. In this chapter three basic polynomial regression functions of different degrees of freedom are investigated for Bayesian model selection, namely $\mathbf{x} = [1]$, $\mathbf{x} = [1\ v]^\top$ and $\mathbf{x} = [1\ v\ v^2]^\top$ with v representing the vehicle speed. The three vectors respectively assume (1) speed independence, (2) a linear dependence on speed, and (3) a quadratic dependence on speed. The operating condition vector may easily be appended to also investigate higher order polynomial functions, or to also consider the influence of other variables such as vehicle acceleration. Bayesian model selection is used to select the most appropriate among the proposed interpolation functions.

An observation y_i^j may now be described as the sum of the specific road interval interpolation function as evaluated for the corresponding operating condition vector $f(\mathbf{x}_i^j)$ and the noise term n_i.

$$y_i^j = f(\mathbf{x}_i^j) + n_i \tag{7.2.1}$$

If the operating condition vector is sufficiently expressive it may be justified to assume that the interpolation function can be approximated as having a linear dependence on $\mathbf{x}$. This linear dependence is expressed by the parameter vector $\mathbf{w}^j$:

$$f(\mathbf{x}^j) \approx \{\mathbf{x}^j\}^\top \mathbf{w}^j \tag{7.2.2}$$

Any discrepancies in this assumption is absorbed by the noise term.

Let all the RMS measurements which correspond to a specific road interval be denoted by the vector $\mathbf{y}^j$, and let all the associated operating condition vectors be contained in the matrix $\mathbf{X}^j$. The likelihood of the data given the model as represented by the parameter values $\mathbf{w}^j$ may now be denoted $p(\mathbf{y}^j|\mathbf{X}^j, \mathbf{w}^j)$. Due to the independent noise assumption the joint likelihood for the K RMS observations which corresponds to any road interval, may simply be computed as the product of the individual datum point likelihoods:

$$p(\mathbf{y}^j|\mathbf{X}^j,\mathbf{w}^j) = \prod_{i=1}^{K} p(y_i^j|\mathbf{x}_i^j,\mathbf{w}^j) \tag{7.2.3}$$

$$= N(\{X^j\}^\top \mathbf{w}^j, \sigma_n^2 \mathbf{I}) \tag{7.2.4}$$

The parameter estimates which optimize equation 7.2.4 is equivalent to the least square error (LSE) solution . It is assumed that there are a number of road intervals where sufficiently many measurements are available so that the LSE solution will be fairly good. Such intervals may automatically be selected based on certain criteria. The parameter values for those LSE estimates (as estimated at road intervals with good data) are used as guideline to what the typical parameter values are to be at other road intervals where sufficient data may not be available. A multivariate Gaussian distribution is estimated from the LSE solutions for the reference road intervals. This distribution is subsequently used as the prior distribution $p(\mathbf{w})$. Let the prior mean be denoted by the vector $\mathbf{m}_0$, and let the covariance matrix be denoted as $\mathbf{S}_0$, so that the prior may be expressed as:

$$p(\mathbf{w}) = N(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0) \tag{7.2.5}$$

By Bayes' theorem the prior and the data driven likelihood may now be used to obtain a posterior distribution over the parameter values:

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}},$$

$$p(\mathbf{w}^j|\mathbf{y}^j,\mathbf{X}^j) = \frac{p(\mathbf{y}^j|\mathbf{X}^j,\mathbf{w}^j)p(\mathbf{w}^{\mathbf{j}})}{p(\mathbf{y}^j|\mathbf{X}^j)}, \tag{7.2.6}$$

where the marginal likelihood $p(\mathbf{y}|\mathbf{X})$ serves to normalize the posterior. It may be shown that the posterior distribution is also a Gaussian distribution [90]:

$$p(\mathbf{w}^j|\mathbf{y}^j, X^j) = N(\mathbf{w}^j|\mathbf{m}_n^j, \mathbf{S}_n^j) \tag{7.2.7}$$

where the posterior mean $\mathbf{m}_n$ and covariance $\mathbf{S}_n$ for road section j are given by:

$$\mathbf{m}_n^j = \mathbf{S}_n(\mathbf{S}_0^{-1}\mathbf{m}_0 + \sigma_n^{-2}\{\mathbf{X}^j\}^\top \mathbf{X}^j) \tag{7.2.8}$$

$$\mathbf{S}_n^j = (\mathbf{S}_0^{-1} + \sigma_n^{-2}\{\mathbf{X}^j\}^\top \mathbf{X}^j)^{-1} \tag{7.2.9}$$

The likelihood of observing a RMS value y_* at an operating condition $\mathbf{x}_*^j$ while traversing road section j may be estimated from the updated likelihood function which is again of Gaussian form [90]:

$$p(y_*|\mathbf{y}^j, \mathbf{X}^j, \mathbf{m}_0, \mathbf{S}_0, \sigma_n) = N(y_*|\mathbf{w}_n^\top \mathbf{x}_*^j, \{\tau_*^j\}^2) \tag{7.2.10}$$

The variance $\{\tau_*^j\}^2$ of the predictive distribution indicates the uncertainty in a prediction at operating condition $\mathbf{x}_*^j$ and is given by:

$$\{\tau_*^j\}^2 = \sigma_n^2 + \{\mathbf{x}_*^j\}^\top \mathbf{S}_n \mathbf{x}_*^j \tag{7.2.11}$$

7.3 EXPERIMENTAL SETUP

7.3.1 Measurements

An accelerometer was attached to the rear wheel shaft. The vehicle response was measured with a DC coupled Crossbow accelerometer and recorded on a Somat eDAC-lite data acquisitioning system at a sampling frequency of 400 Hz. The integrity of the DC coupled Crossbow accelerometer was tested on a high frequency actuator and found to be very accurate over the frequency range of interest. An infrared eye and a push button were also connected to the eDAC to provide reference signals. A proximity sensor on the drive shaft was used to measure its rotational speed.

The vehicle speed was calculated from the measured drive shaft speed and from knowledge of the shaft-wheel rotation ratio and the effective rolling radius of the wheel. The vehicle's speed was integrated over time to obtain an estimated relative position vector. The acceleration signal was subsequently resampled to be a function of position rather than time.

Figure 7.1 depicts in (a) the measured raw vibration signal and in (b) the corresponding RMS severity metric. The severity metric is computed as the RMS acceleration over 1 metre intervals. The interval windows are selected so as to overlap the previous interval by a half metre. The RMS severity metric thus has a frequency of two datum points per metre.

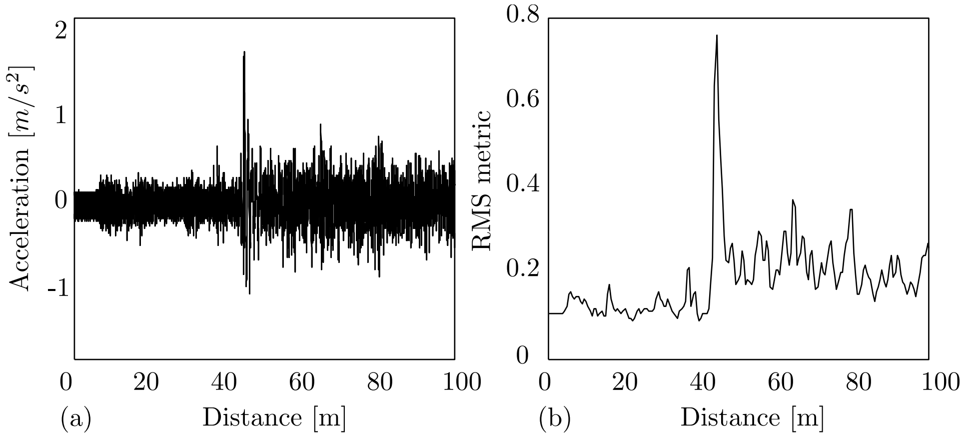


Figure 7.1: The severity metric is derived from the acceleration signal by computing its RMS over a 1 metre sliding window with 0.5 metre overlap

7.3.2 Road profiles and signal alignment

The response of the UV was measured both above and below ground. Two above ground road profiles with favourable characteristics were available on site. The UV response over these profiles were measured with the intention investigating the interpolation procedure over simple road profiles. Figure 7.2 (a) indicates how an impulse excitation was created by laying down a row of concrete blocks on a level surface. Figure 7.2 (b) illustrates the second profile which comprised two closely situated humps of similar dimension. The two humps were situated in a low wavelength ditch.

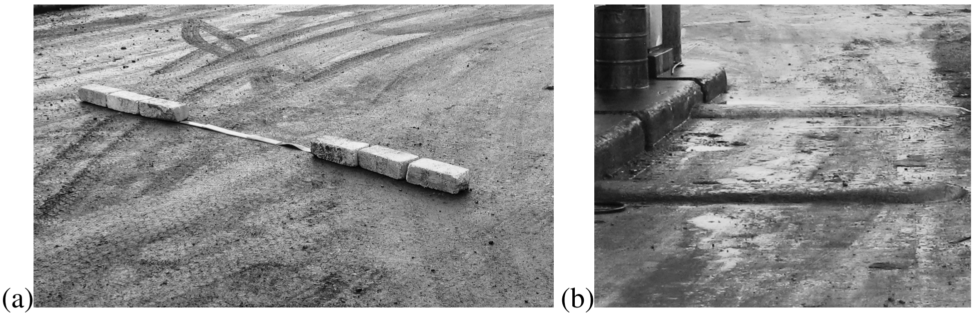


Figure 7.2: The known road profiles (a) impulse like excitation (b) double humps situated in ditch

The above ground road profiles were recorded by measuring the vertical distance between the road surface and a reference cable which was tensioned above the road. The vehicle's position relative to the road profiles was estimated by means of an infrared eye which detected reflective tape strips on the ground which was situated at the start and end of the runs.

Measurements were also performed underground. The UV response was measured over two 730 m long sections of underground road, which were identified as being representative of typical underground conditions. The UV was driven thrice over each of the two road sections. Due to the underground working conditions it was not possible either to measure the road surfaces, nor to use reflective tape to indicate the start and ends of each run. For this reason the vehicle responses were aligned by optimizing their cross-correlations.

7.3.3 Parameter prior

It has been explained that the Bayesian prior is used to improve the estimated parameter values for the regression functions and is especially important for road intervals which are subject to constrained data. The prior is based on general parameter values (and covariance's) as observed at other road intervals.

In this case study the measurements from one of the underground profiles is used to construct the prior. The road profile is 730 m long and was traversed three times. The RMS severity metric was computed with 0.5 m overlap so that there are a total of 1460 intervals. Using the standard least square error (LSE) criterion an interpolation function is estimated for each of the 1 metre intervals. The standard LSE solution is efficiently solved using the closed form analytic solution [90].

None of the road intervals have been traversed especially many times, so that a subset of representative road segments (where the parameter values are expected to be more accurate) will not be identified. Rather all the 1460 estimates (as based on the LSE criterion) will be used to estimate the prior over the parameter values. While it is expected that some of the LSE estimates may be very crude, it is still anticipated that the estimated distribution over all of the LSE estimates will tend toward a good representation of the true parameter ranges as associated with road intervals of different spectral characteristics.

The functional relationship between the vehicle speed and response is not known a priori. The process of computing the LSE solution for each road interval is therefore repeated for different modelling assumptions. In this chapter three models are investigated which respectively assume that the vehicle response as measured over a specific section of road is either: (i) independent of speed, (ii) linearly dependent on speed or (iii) a quadratic function of speed. An individual prior over the parameter values is estimated for each of the possible modelling assumptions.

Each of the priors are approximated as a multivariate Gaussian density. This multivariate density indicates the expected parameter values, their variances, along with the covariance which exist between the parameter values. The prior is thus fully characterized by its mean $\mathbf{m}_0$ and covariance $\mathbf{S}_0$.

7.3.4 Bayesian model selection

Once the prior density distribution for each of the interpolation functions has been estimated, it remains to be determined which of the interpolation assumptions are most appropriate for the given data set. It is also required to estimate the observation noise variance σ_n^2. If the observation noise is assumed very small relative to the variance of the parameter values as expressed by the prior, then each data observation will be much trusted and thus significantly influence the posterior. Conversely, if the noise variance is assumed very large then many observations will be required to significantly influence the posterior parameter estimate.

The problem of both selecting the most appropriate interpolation function as well as the optimal noise variance may simultaneously be solved by means of Bayesian model selection. Let each of the three interpolation functions be paired with a range of values for the noise variance. Each pair represents a potential model. The most appropriate model is the one that optimizes the posterior probability as denoted in equation 7.2.6. The posterior probabilities are based on all the data as measured over the first underground road profile of 730 m in length.

The data driven likelihoods, the model priors, and the subsequent posterior probabilities for each of the possible models are computed. The results for the second and third order interpolation functions as paired with different values for the noise standard deviation are presented as graphs. For sake of brevity the results for the speed independent assumption is not presented, all evidence however clearly points to a speed dependence in the data.

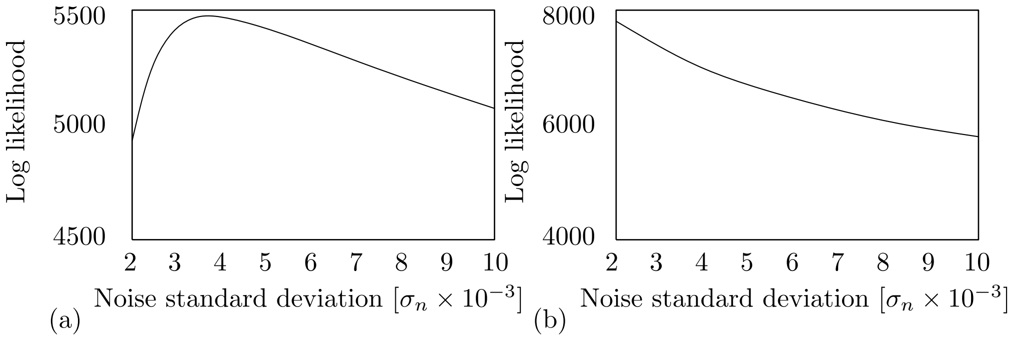


Figure 7.3: Logarithm of the likelihood term with respect to different noise standard deviations $[\sigma_n]$ as computed for (a) the first order and (b) the second order functions of speed

Figure 7.3 depicts the log likelihoods as computed over all the observations for all the road intervals for (a) the linear dependence on speed and (b) the quadratic dependence on speed. The log likelihood for the linear model reaches a maximum at a noise standard deviation of approximately $\sigma_n = 3.5 \times 10^{-3}$. The log likelihood for the quadratic term outperforms that of the linear model, it never reaches a maximum but rather tends towards infinity as the noise standard deviation approaches zero. This reflects the fact that the quadratic interpolation functions are sufficiently flexible to perfectly model the three observations per interval.

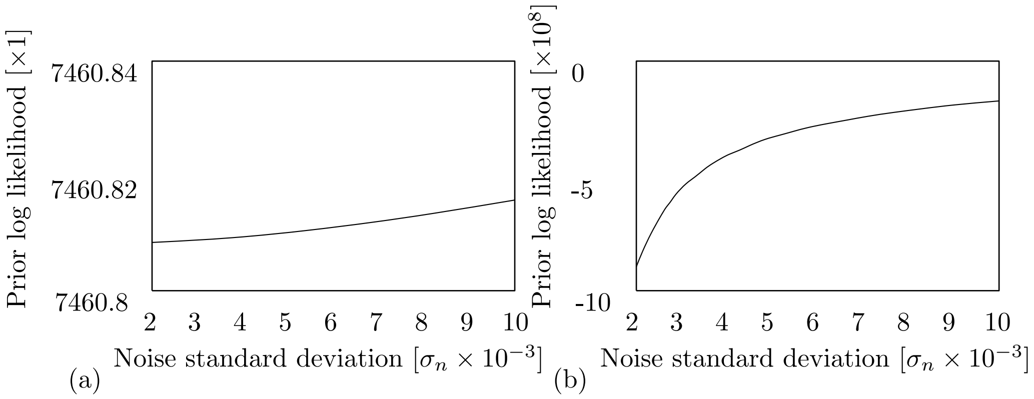


Figure 7.4: Logarithm of the prior probabilities with respect to different noise standard deviations $[\sigma_n]$ as computed for (a) the first order and (b) the second order functions of speed

The prior probabilities of the model parameters are computed and illustrated in figure 7.4 for (a) the linear dependence on speed and (b) the quadratic dependence on speed. A larger noise standard deviation resulted in parameter estimates which better corresponded to the prior means, hence resulting in a greater prior probability. Consider that the quadratic dependence requires the estimation of an additional 1460 parameter values. This additional flexibility significantly penalized the quadratic term, especially when the observation noise is assumed small. The simpler linear dependence (which spread its density distribution over fewer parameter) resulted in significantly higher parameter prior probabilities.

A log posterior model probability is computed as the sum of the associated log likelihood, the log prior probability, and the negative log marginal probability. The marginal probability expresses the probability of observing the data given all of the assumed possible models. The marginal probability reduces to a constant value which acts as a normalizing term to all of the posteriors. Figure 7.5 ignores the marginal term and subsequently compares the unnormalized posterior probabilities for both (a) the linearly dependencies with different noise standard deviations and (b) the quadratic dependencies as computed with different noise standard deviations. The model with the maximum posterior probability corresponds to a linear dependence on speed, with a noise standard deviation of approx-

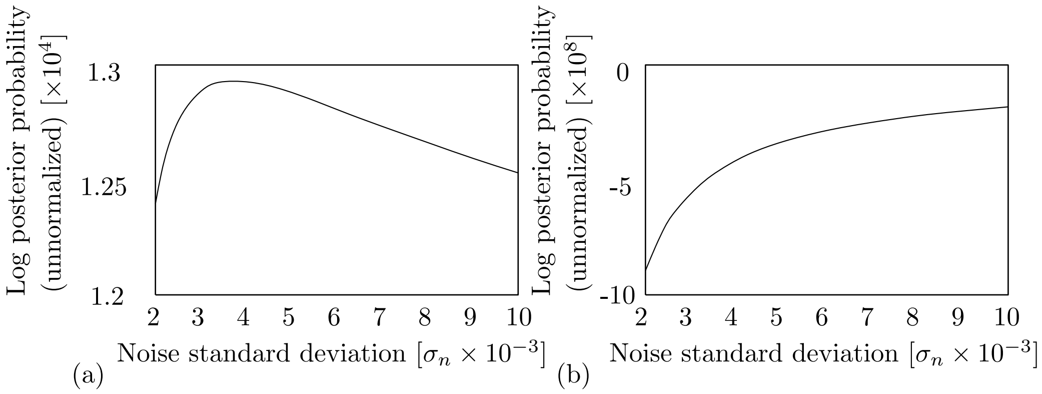


Figure 7.5: The logarithm of the unnormalized posterior probabilities with respect to different noise standard deviations [σ_n] as computed for (a) the first order and (b) the second order functions of speed

imately 3.7×10^{-3}. If the posteriors are normalized with respect to the different speed dependencies (independent, linear or quadratic) then the posterior of the linear assumption is virtually 1, with the other two model posteriors effectively equalling 0.

The linear dependence on speed assumption clearly proves to the most appropriate this data set. It is important to note that these results do not prove the optimality of a linear assumption for other RMS data sets. It is very possible that higher model orders would be favoured if more data per interval are available.

The results presented in figures 7.3, 7.4 and 7.5 are based on the assumption that a single type of regression function should be fitted to all of the road intervals. This is by no means a requirement. The most appropriate type of regression function for any individual road interval may automatically be updated as more measurements become available and the posterior begins to favour more accurate interpolation functions.

7.3.5 Iterative updating

The Bayesian approach allows for the iterative updating of the posterior function with each data observation. After an observation for a section has been made and the posterior has been computed the posterior will be used as the prior for the next observation [90]. The solution is mathematically equivalent to the solution obtained by considering all the data simultaneously, however it avoids the requirement of storing any data.

Figure 7.6 illustrates the growing confidence in the interpolation function as more measurements be-

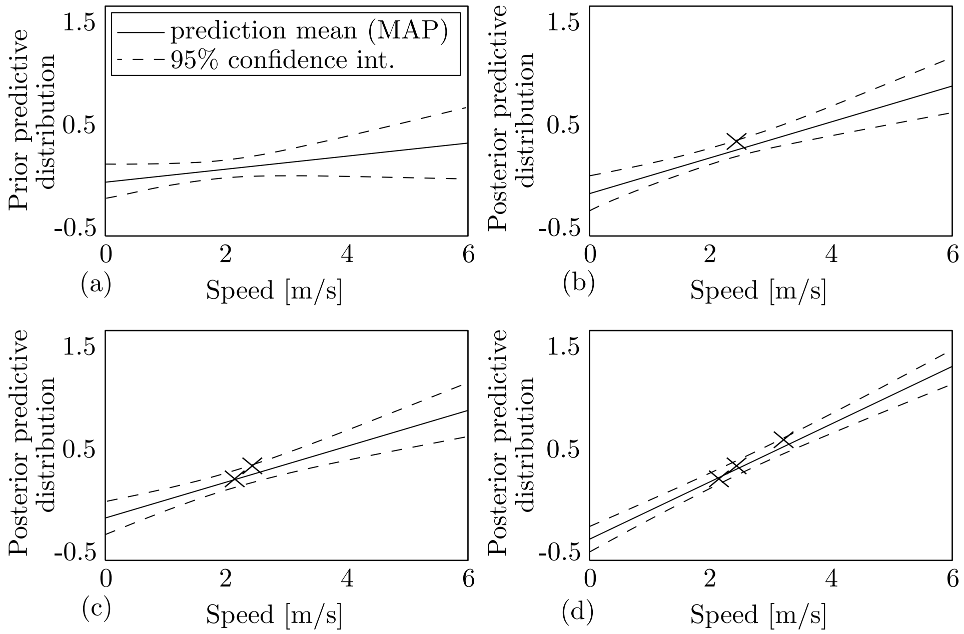


Figure 7.6: Updating of the estimated regression function as more measurements become available, starting from (a) the prior, (b) one observation, (c) two observations, (d) three observations

come available. The prior function (linear assumption) is illustrated in figure 7.6 (a). Two dashed lines are used to indicate the two standard deviation bounds (95% confidence interval). The prior confidence bounds are fairly weak, thus indicating that the regression function is still uncertain. Figure 7.6 (b), (c) and (d) depicts how the intercept and slope adapt as more measurements become available. The confidence over the interpolated values increases accordingly.

7.4 RESULTS

The priors and the observation noise variance have been estimated from data which are representative of typical underground road conditions as measured by a specific UV. The ability of the proposed regression function approach to estimate speed standardised RMS values is subsequently investigated on the two above ground roads of which the profiles have been measured. The UV was driven over the concrete blocks at four different speeds, resulting in four measurements; m_1 to m_4. Figure 7.7 (a) depicts a level road with a sudden impulse like excitation due to the concrete blocks. The vehicle response is underdamped so that the energy takes some time to dissipate from the system. All of the measurements, including the underground runs, were conducted at speeds which vary between 1 and 4 m/s. These slow speeds were either mandated by the poor road conditions, or safety precautions. A standardised speed of 3 m/s was selected as the speed which was most adept to interpolation. From figures 7.7 (b) and (c) it may be seen that the RMS severity significantly depends on the vehicle speed. The 3 m/s standardised severity metric is presented in figure 7.7 (d). The solid line (m_1) indicates the estimated RMS value at 3 m/s (also referred to as the maximum a posteriori (MAP) solution), while the two standard deviations dashed lines indicate the 95% confidence bounds. It is seen that the RMS values which represents the vehicle passing over the concrete blocks as measured at higher speeds (more than 3 m/s) are almost double the magnitude of the measurement made at a low speed (about 1.5 m/s). Those RMS values which were measured over the intervals of smooth road were near speed independent. Figure 7.7 (d) indicates that the localized regression functions were able to account for the different dependencies on speed for different road intervals.

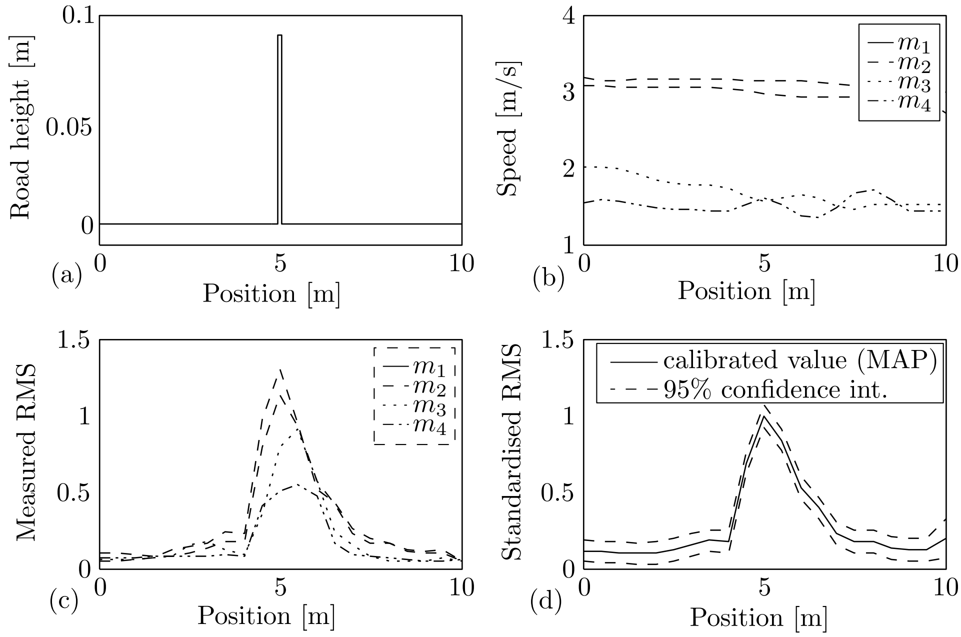


Figure 7.7: Impulse excitation profile simulated with concrete blocks with severity metric standardised to vehicle speed of 3 m/s (four measurements m_1 to m_4).

The response of the vehicle was subsequently measured over the low wavelength ditch which contained two significant humps. The measured road profile is depicted in figure 7.8 (a). The long wavelength character of most of the profile did not excite the vehicle. The speed dependence of the severity metric over the two humps is clearly illustrated in figures 7.8 (b) and (c) by the dotted line (m_2) which shows how the driver decelerated for the first hump, but traversed the second hump at fairly high speed. As expected the measured severity metric (m_2) was much smaller for the first hump compared to the second hump. The severity metric which has been standardised to a vehicle speed of 3 m/s indicates that both humps are of similar severity.

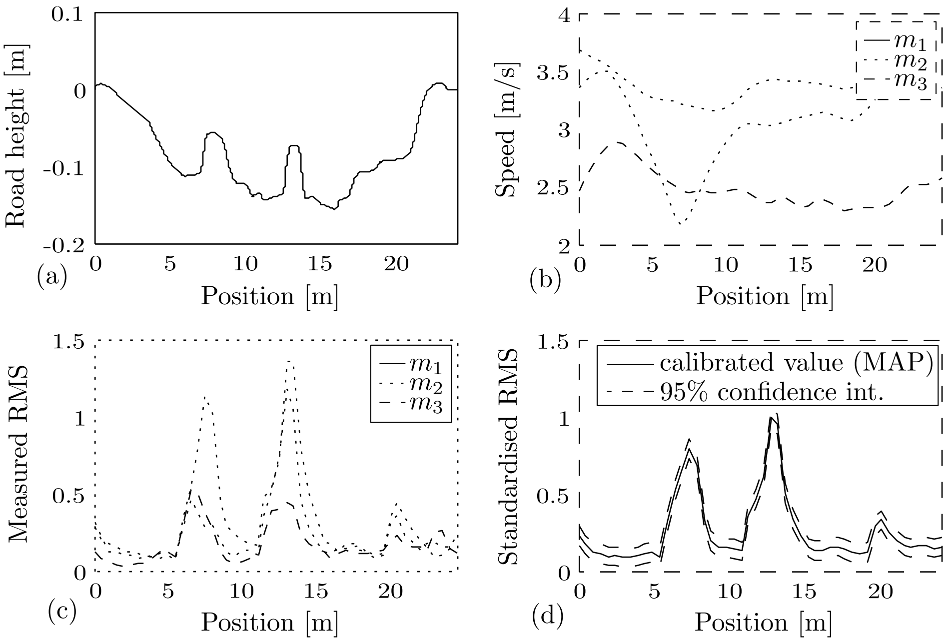


Figure 7.8: Double hump road profile with RMS metric standardised to vehicle speed of 3 m/s.

Underground measurements over two 730 m long road profiles were conducted. The runs from the one road profile were used to generate the parameter priors. Figure 7.9 indicates the RMS values for the second road profile as estimated at a consistent speed of 3 m/s. A number of small peaks which are indicative of relatively poor sections of road are identified. The maximum standardised RMS value is 0.4, which is less than half of the severity of the concrete blocks.

The correlation between the regression function parameters and the condition of different road surfaces is briefly explored in figures 7.10 (a) and (b). The estimated RMS values at 3 m/s (for the underground measurement) is presented on the x-axis of the figures. A linear speed dependence is assumed, so that each road interval has two estimated parameter values including the offset parameter, and the slope parameter. Only the slope parameter is considered here. The y-axis indicates the value of the slope parameter (dependence on speed). The LSE slope estimates are presented in 7.10 (a), while the Bayesian (MAP) estimates are shown in 7.10 (b).

It is observed that the slopes of the interpolation functions generally increase for more severe roads. Figure 7.10 (b) illustrate how the inclusion of the prior bias the results closer to the general trend. A number of the LSE slope estimates are negative. This indicates that those LSE calibration functions would predict lower RMS values at higher speeds. Although this is not theoretically impossible, it does seem highly unlikely. The prior renders the Bayesian estimates more robust, thereby avoiding negative slope estimates.

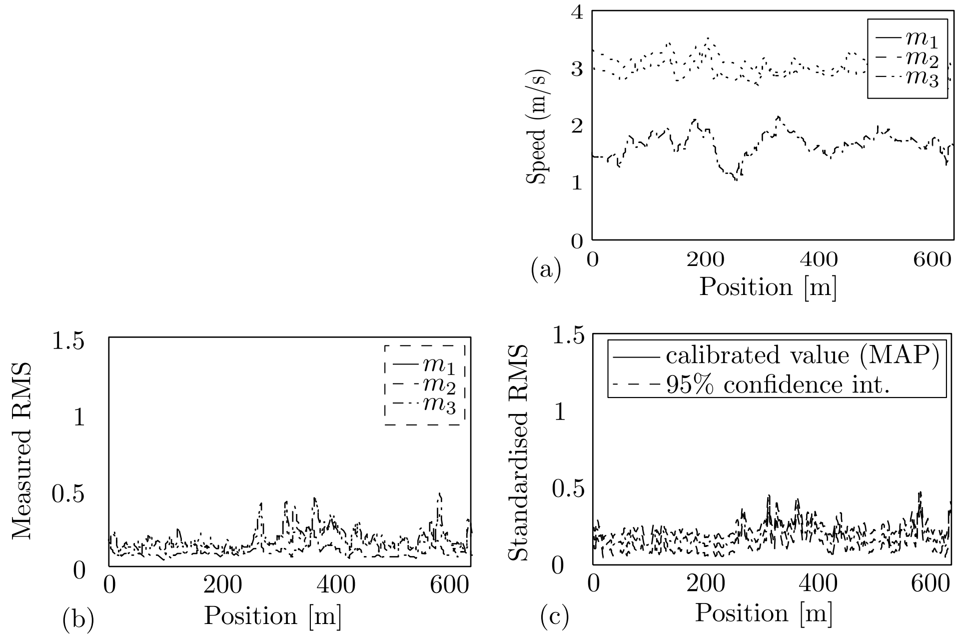


Figure 7.9: Evaluation of underground road

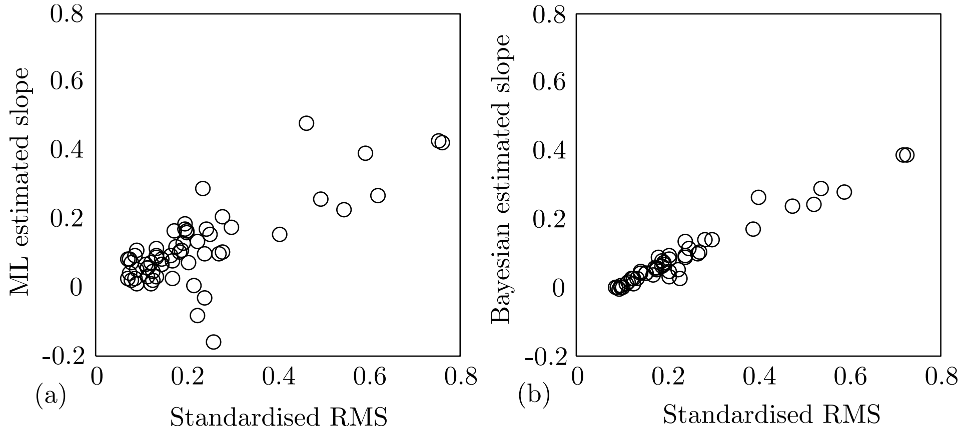


Figure 7.10: (a) Maximum likelihood parameter (slope) estimates compared to (b) Bayesian paramter (slope) estimates

7.5 CHAPTER CONCLUSION

A waveform reconstruction methodology was proposed that is concerned with the cost-effective monitoring of the condition of haul roads. Unique regression functions are estimated for individual road intervals, so as to reflect the spectral properties of that interval. The regression functions are estimated by means of Bayesian parameter estimation. This allows for the incorporation of a strong prior over the range of expected parameter values. The prior reflects the experience which was gained from other road intervals.

The methodology was investigated on experimental data that were obtained by measuring the response from a utility vehicle in a coal mine environment. The RMS acceleration response as measured on the vehicle's rear axle was used as RTRRMS metric. The Bayesian framework rendered it possible to estimate a calibrated RMS value for the different sections of the road in robust manner. While the current study focussed on calibrating the vehicle response with respect to speed, it is anticipated that the same methodology may be implemented to also account for other possibly significant influences on the RTRRMS, such as vehicle acceleration and payload.

It is believed that the proposed waveform reconstruction methodology may be implemented as part of a generic haul road monitoring strategy.

Chapter 8

Vehicle response normalisation with Gaussian Processes

8.1 INTRODUCTION

Similar to chapter 7 this chapter indicates how it is possible to calibrate the haul truck response to a consistent speed. The approach followed in this chapter implements a single flexible regression function to jointly represent all the relevant vehicle responses as measured at all relevant speeds over the complete range of road spectral characteristics. Since there is no longer a need to construct unique training functions for individual sections of road as was done in chapter 7, the calibration function now only has to be optimised once. This renders it possible to calibrate data even when a section of road is only traversed once, or when noisy and limited GPS measurements are available. Unlike the Bayesian prior approach that biases all the vehicle responses towards the mean, the Gaussian Process approach proposed in this section allows for better calibration of road conditions that might have been under represented in the training data.

8.2 METRIC CALIBRATION FOR CONSISTENT ROAD QUALITY EVALUATION

Severity metric

Road quality may be interpreted as one, or a combination, of multiple criteria [15]. Some criteria which may be considered include how safe the road is, how much wear it causes to a vehicle, and how comfortable it is to drive on. Toward this goal it is reasonable that various types of vehicle response measurements may be used to assess road quality. Since the vehicle response is dependent on the operating conditions, such as the vehicle speed, it is required to calibrate the proposed metrics to a standardised operating condition which will allow for consistent road quality comparison.

The proposed calibration framework discussed in this chapter is intended to be generic, and not specific to a selected metric. However, for the presented case study a single metric which is deemed relevant to the continuous monitoring of haul roads is investigated.

An accelerometer is attached to the wheel axle. The position on the wheel axle, which is an unsprung mass, is selected to ensure good transmissibility. The envelope of the acceleration signal is subsequently computed. This is done by taking the point-wise absolute value of the acceleration signal and then using a low pass filter to remove the high frequency content. The obtained metric is instantaneous and reflect the presence of individual humps. The metric is representative of the magnitude of the acceleration, and thus the force, which is transmitted to the vehicle. The metric is also easy to implement, since it only requires the installation of a single accelerometer.

Dynamic calibration function

The focus of this chapter is the implementation of a dynamic calibration function which adjust itself to the instantaneous nature of the road which the vehicle traverses so that more robust calibration may be performed.

The implemented dynamic calibration function is based on the assumption that if a vehicle traverses a section of road at a measured speed v_i so that vehicle response r_i is observed, then the combination of these two measurements contains information regarding the underlying condition of that section of road. These two variables may thus be used in some non-linear manner to adjust the calibration function to better represent the specific section of road. The adjusted calibration function may subsequently be used to estimate the expected vehicle response r_j if the vehicle were to traverse the same section of road at a different (standardised) speed v_j. The process where an individual datum point from the continuous measurement is calibrated is presented in figure 8.1.

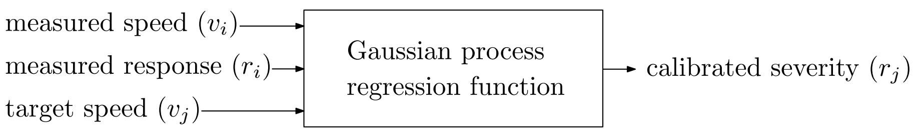


Figure 8.1: Calibrating a single datum point from the continuous severity measurement to a specified target speed.

Nonlinear regression techniques may be used to learn the proposed calibration functions from training data. It will be shown how the required training data may be obtained in a cost effective manner, without the need for manual classification or profiling of the road surface.

Various regression techniques such as polynomial, spline or Neural Network (NN) regression analysis may be implemented. The results indicated in this chapter are based on Gaussian process regression, since this is the technique which offered the best performance on our data set. Rasmussen and Williams [97] examines a number of case studies where Gaussian process regression generalizes well and occasionally significantly outperforms other regression techniques, especially when limited training data are available. An empirical study by Wang et al. [98] similarly found one of the main strengths of the Gaussian models to be their ability to generalize well from small data sets. For sake of brevity, and since the aim of this chaper is not to compare different regression methodologies, our results from the polynomial and NN regression approaches are not included in this chapter. Essentially it was found that the fairly small data set that is subject to significant noise was well suited to analysis by Gaussian processes. While Gaussian processes regression tends to offer good generalization, it does have the downside of being computationally intensive on larger data sets [90]. For this reason it might be preferable to implement neural network regression if more training data are available.

8.3 GAUSSIAN PROCESS REGRESSION

Gaussian process (GP) regression may be used to model nonlinear functional mappings from an input space to a target space. It is based on Bayesian statistics, and incorporates a non parametric approach which differs substantially from parametric techniques like NNs. By definition the GP represents an infinite collection of random variables, any finite subset of which have joint Gaussian distributions [95]. A GP may be understood as describing a probability distribution over functions. The Gaussian process favours functions that are both smooth and which explains the training data well. It is this smoothing characteristic of the function that allows it to generalise well.

Let the vector $\mathbf{x}_n$ denote a single position in the input space. The set of training input vectors $\mathbf{X}_N \equiv \{\mathbf{x}^n\}_{n=1}^{N}$ corresponds to the target vector $\mathbf{y}_N \equiv \{y_n\}_{n=1}^{N}$. A mean function $m(\mathbf{x}_n) = \mathbb{E}[y(\mathbf{x}_n)]$ is used to describe the expected value at any point of the input space, before any training data have been considered. During the preprocessing stage the data may be zero mean unit normalised so that the mean function may be set to zero, $m(\mathbf{x}) = 0$ for all values of $\mathbf{x}$.

Consider that the covariance between two function values at fixed positions $\mathbf{x}_a$ and $\mathbf{x}_b$ may be denoted by $\text{cov}(\mathbf{x}_a, \mathbf{x}_b) = \mathbb{E}[((y(\mathbf{x}_a) - m(\mathbf{x}_a))(y(\mathbf{x}_b) - m(\mathbf{x}_b))]$. Given a zero mean function the covariance reduces to $\text{cov}(\mathbf{x}_a, \mathbf{x}_b) = \mathbb{E}[(y(\mathbf{x}_a)(y(\mathbf{x}_b))]$. Consider that knowledge of this covariance function renders it possible to infer one function value, given knowledge of the other. The GP extends this idea by making use of a kernel function to explain the covariance as a function of the input space $k(\mathbf{x}_a, \mathbf{x}_b) \approx \text{cov}(y_a, y_b)$. This renders it possible to infer the value y_* of a function at a novel position $\mathbf{x}_*$ in the input space.

By implementing any kernel function some implicit assumptions about the function space are made. It is possible to construct new kernel functions by combining simpler kernel functions [90]. This makes it possible to impose a prior over the function space with certain desirable properties. The kernel function k^c that is implemented in this chapter comprises of the addition of three simple kernel functions, namely the squared exponential with automatic relevance detection k', a linear kernel with automatic relevance detection k'', and a kernel function k''' which assumes that the signal is subject to independent identically distributed (iid) Gaussian noise. The implemented kernel may thus be expressed as:

$$k^c(\mathbf{x}_a,\mathbf{x}_b) = k'(\mathbf{x}_a,\mathbf{x}_b) + k''(\mathbf{x}_a,\mathbf{x}_b) + k'''(\mathbf{x}_a,\mathbf{x}_b) \tag{8.3.1}$$

The squared exponential (SE) is perhaps one of the most popular kernel functions. It makes the assumption that function values which lie closely together in the feature space are likely to be more similar, with near unity covariances for variables that have closely corresponding feature inputs $k'(\mathbf{x}_a,\mathbf{x}_b) = \sigma_f \times \exp(-(\mathbf{x}_a - \mathbf{x}_b)^\top \mathrm{M}_1(\mathbf{x}_a - \mathbf{x}_b))$. M_1 is a diagonal matrix, with positive automatic relevance detection (ARD) parameters $\mathrm{M}_1 = \mathrm{diag}(\ell_1)$ where ℓ_1 is a vector of length D so that it corresponds to the dimension of the input space. The ARD parameters are also referred to as the characteristic length-scale parameters and determines the rate at which the function varies in the corresponding direction of the input space. The shorter the length scale parameter is for a specific feature component, the faster the function tends to vary for any variation of that feature. A short length scale thus corresponds to high relevance. The function variance σ_f^2 is related to the overall variance of the function.

The linear covariance function with automatic relevance detection (ARD) is parameterized by $k''(\mathbf{x}_a,\mathbf{x}_b) = \mathbf{x}_a^\top \mathrm{M}_2 \mathbf{x}_b$. M_2 represents the diagonal matrix $\mathrm{M}_2 = \mathrm{diag}(\ell_2)$.

The final constituent of the covariance function is the noise variance $k'''(\mathbf{x}_a,\mathbf{x}_b) = \sigma_n \times \delta(a,b)$ which accounts for the possible presence of zero mean Gaussian noise. The noise variance is represented by σ_n and $\delta(a,b)$ is the Kronecker delta function, which is 1 if and only if $a = b$ and zero otherwise.

The matrices containing the ARD parameters $\{\mathbf{M}_1,\mathbf{M}_2\}$, the signal variance $\{\sigma_f^2\}$ and the noise variance $\{\sigma_n^2\}$ are all unknown and must be inferred from the training data. These parameters are collectively referred to as the hyperparameters of the GP and denoted by θ. By means of Bayes' rule it may be shown that the maximum a posteriori hyperparameter values $\theta_{\mathbf{MP}}$ may be found by maximising the marginal likelihood $p(\mathbf{y}_N|\mathbf{X}_N,\theta_{MP})$. This is equivalent to minimising the negative log marginal likelihood:

$$\log \mathrm{P}(\mathbf{y}_N|\mathbf{X}_N,\theta) = -\frac{1}{2}\mathbf{y}_N^\top(\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N))^{-1}\mathbf{y}_N - \frac{1}{2}\log|\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N)| - \frac{2}{n}\log 2\pi \tag{8.3.2}$$

where $\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N)$ is the $N \times N$ covariance matrix between all pairs of training inputs $\mathbf{X}_N$ and is computed with equation 8.3.1. The joint distribution of the observed target values and the function values at the novel locations is obtained as:

$$\begin{bmatrix} \mathbf{y}_N \\ \mathbf{y}_* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N)+\sigma_n^2 I & \mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_*) \\ \mathbf{K}^c(\mathbf{X}_*,\mathbf{X}_N) & \mathbf{K}^c(\mathbf{X}_*,\mathbf{X}_*) \end{bmatrix}\right) \tag{8.3.3}$$

The conditional distributions subsequently reduces to [95]

$$P(\mathbf{y}_*|\mathbf{X}_N,\mathbf{y}_N,\mathbf{X}_*) \sim N(\bar{\mathbf{y}}_*,\mathrm{cov}(\mathbf{y}_*)) \tag{8.3.4}$$

So that MAP regression outputs $\bar{\mathbf{y}}_*$ at the new locations $\mathbf{X}_*$ may subsequently be inferred as

$$\bar{\mathbf{y}}_* = \mathbf{K}^c(\mathbf{X}_*,\mathbf{X}_N)[\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N)+\sigma_n^2 I]^{-1}\mathbf{y}_N \tag{8.3.5}$$

$$\mathrm{cov}(\mathbf{y}_*) = \mathbf{K}^c(\mathbf{X}_*,\mathbf{X}_*) - \mathbf{K}^c(\mathbf{X}_*,\mathbf{X}_N)[\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_N)+\sigma_n^2 I]^{-1}\mathbf{K}^c(\mathbf{X}_N,\mathbf{X}_*) \tag{8.3.6}$$

The maximum a posteriori (MAP) estimates $\bar{\mathbf{y}}_*$ will be used as calibrated severity metrics.

Gaussian Process regression is implemented in this chapter based on the code developed by Rasmussen and Nickisch [115].

8.4 EXPERIMENTAL INVESTIGATION

Vehicle response data were measured on a utility vehicle which is used in an under ground coal mine environment.

Figure 8.2 (a) illustrates the position of an aluminium block on the wheel axle where the accelerometer was mounted. The vehicle response was measured with a DC coupled Crossbow accelerometer and recorded on a Somat eDAC-lite data acquisitioning system at a sampling frequency of 400 Hz. The integrity of the DC coupled Crossbow accelerometer was tested on a high frequency actuator and found to be very accurate over the 0-40 Hz frequency of range of interest.

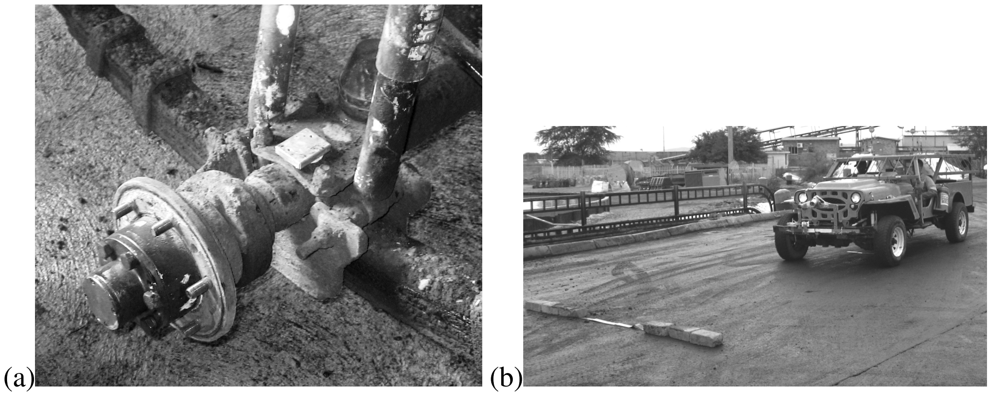
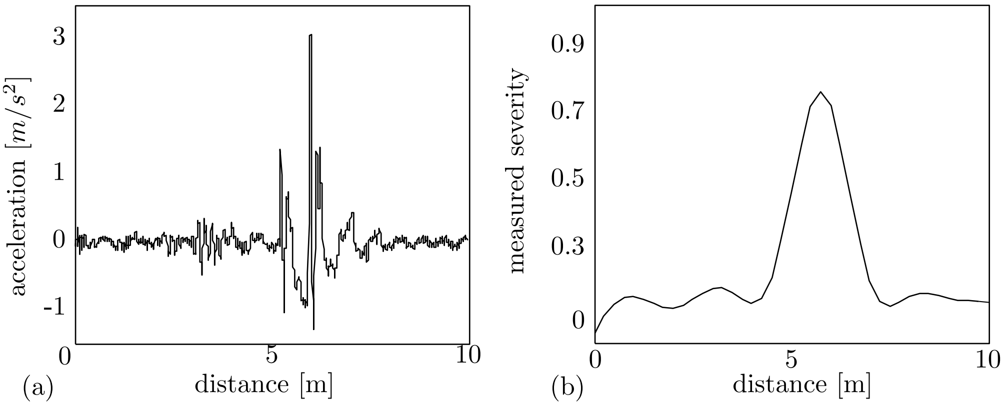

(a) (b)

Figure 8.2: (a) Placement of accelerometer on wheel axle. (b) Measuring the vehicle response over impulse like excitation.

Figures 8.3 (a) and (b) illustrate how the severity metric is obtained by rectifying the acceleration signal and low pass filtering it at a cut-off frequency of 2 Hz.



(a) (b)

Figure 8.3: The Figure (a) raw acceleration signal, (b) envelope of acceleration signal to serve as severity metric

A proximity sensor on the drive shaft was used to measure the rotational speed of the shaft. The vehicle speed is calculated from the measured drive shaft speed, the shaft-wheel rotation ratio and the rolling radius of the wheel. The vehicle speed was subsequently integrated over time to obtain an estimated vehicle position for each datum point. The time domain severity metric was interpolated according to the estimated vehicle position so that the severity metric may be visualized as a function of position, rather than time.

An infrared eye that detected the presence of reflective tape on the ground, and a human operated push button were also connected to the eDAC. These reference signals helped to position the vehicle relative to the investigated road profile. With the severity metric being a function of distance it was possible to align the measurements that were measured at different speeds.

8.4.1 Generating training data and optimising hyperparameters

The calibration function is inferred from training data. The training data represent how the severity metric varies when the utility vehicle traverses the same position on a road but at different speeds.

A simple track is used to generate training data. The track consists of a 10 meter long smooth and level surface, which halfway is intersected by a single row of 100 mm high and 200 mm wide concrete blocks. The concrete blocks serve to excite the vehicle in an impulse like fashion.

The acceleration signals are converted to the severity measure resampled to a function of distance rather than time. The various measurements are then aligned.

A typical training point is created in the following manner. The instantaneous speed v_i^w and severity r_i^w, as measured at a specific position w of the road, is taken as the first two entries of an input vector $\mathbf{x}_{ij}^w$. The instantaneous speed v_j^w from a second measurement (at the same road position) is added as the third and final entry of the input vector so that $\mathbf{x}_{ij}^w = [v_i^w \ r_i^w \ v_j^w]$. The measured severity from the second run at position w is used as target $y_{ij}^m = r_j^w$. This process is illustrated in figure 8.4 where a typical training point is created from two severity measurements which were obtained as the vehicle traversed the concrete hump at different speeds.

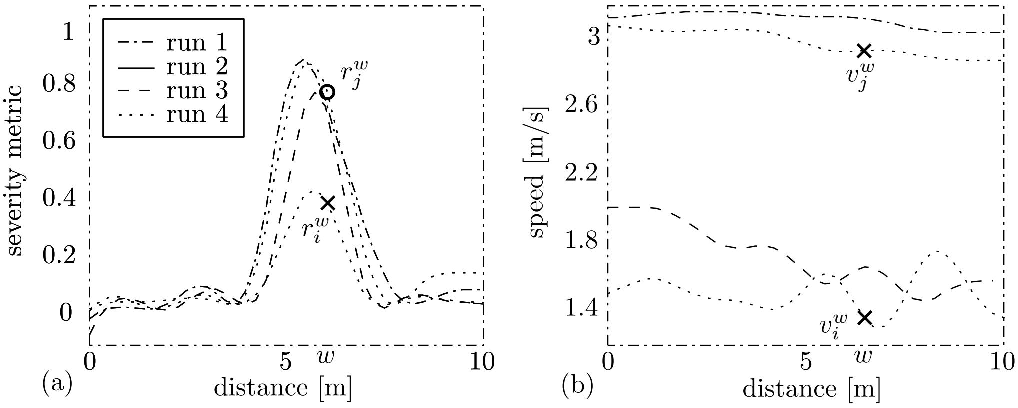


Figure 8.4: A single training point is constructed with input vector entries indicated by the crosses in (a) and (b) and a target value indicated by the circle in (a). The training point is created from two measurements at the exact same position on the track so that the underlying (latent) road condition for those two measurements is the same.

The process of creating training points is now repeated for different combinations of measurements.

Figures 8.4 (a) and (b) indicate the variation between four severity measurements as generated under different speeds. The first two entries of the input can be taken from any one of the four measurements $i = 1..4$ and again be paired up with any one of the four measurements $j = 1..4$ to generate the third entry of the input vector as well as the target. A total of 16 training points are thus created to represent a single position in road. It is important that some training points exist where a measurement regress to itself, so that the calibration function can learn that the severity should not be altered if $v_i^w = v_j^w$. These 16 datum points are now representative of a single road condition at position w, even though that condition is not actually specified. Similarly can 16 additional points be created for every other position on the road. In this chapter the 16 training points are generated at half meter intervals over the 10 meters of track $w = 1..21$ so that a total of $N = 336$ training points are obtained.

The training points are clustered in the 336×3 input matrix $\mathbf{X}_N$ and the associated target vector $\mathbf{Y}_N$.

$$\mathbf{X}_N = \begin{bmatrix} v_1^1 & r_1^1 & v_1^1 \\ v_1^1 & r_1^1 & v_2^1 \\ \vdots & \vdots & \vdots \\ v_2^1 & r_2^1 & v_1^1 \\ v_2^1 & r_2^1 & v_2^1 \\ \vdots & \vdots & \vdots \\ v_4^1 & r_4^1 & v_3^1 \\ v_4^1 & r_4^1 & v_4^1 \\ \vdots & \vdots & \vdots \\ v_1^2 & r_1^2 & v_1^2 \\ v_1^2 & r_1^2 & v_2^2 \\ \vdots & \vdots & \vdots \\ v_4^{21} & r_4^{21} & v_3^{21} \\ v_4^{21} & r_4^{21} & v_4^{21} \end{bmatrix}, \qquad \mathbf{Y}_N = \begin{bmatrix} r_1^1 \\ r_2^1 \\ \vdots \\ r_1^1 \\ r_2^1 \\ \vdots \\ r_3^1 \\ r_4^1 \\ \vdots \\ r_1^2 \\ r_2^2 \\ \vdots \\ r_3^{21} \\ r_4^{21} \end{bmatrix}$$

The Gaussian process hyperperameters are now chosen such as to minimize the negative log marginal likelihood of the training data. The implemented code as developed by Rasmussen and Nickisch [115] makes use of the Polack-Ribiere flavour of conjugate gradients to compute search directions, and a line search using quadratic and cubic polynomial approximations. The Wolfe-Powell stopping criterion is used together with the slope ratio method for guessing initial step sizes.

8.4.2 Function interpretation

The characteristics of the calibration function are subsequently investigated.

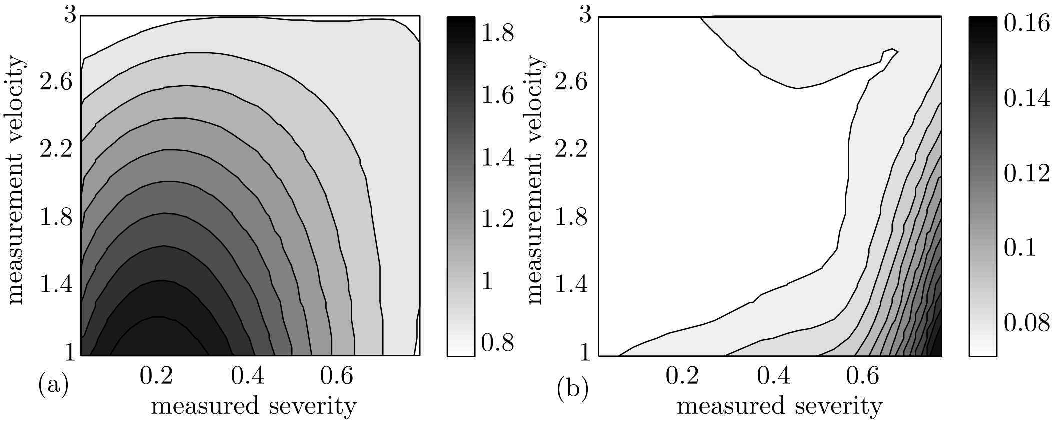


Figure 8.5: (a) Illustrates the calibration ratio (the severity as calibrated to a consistent speed of 2.5 m/s divided by the measured severity). (b) The GP's uncertainty (one standard deviation) in the accuracy of the calibration function under different conditions.

A contour plot of the calibration ratio is presented in figure 8.5 (a). The calibration ratio is obtained as the severity estimate from the Gaussian process y^* divided by the measured severity. The contour plot is consistently defined for a target speed of 2.5 m/s. The y-axis indicates the measurement velocity and the x-axis the magnitude of the measured severity.

Figure 8.5 (b) depicts the estimated standard deviation associated with any prediction. Dark contours indicate areas where the Gaussian process has observed little or no training data, so that it has little confidence in its own predictions. Bright areas indicate those areas where the GP is expected to make good predictions.

In figure 8.5 (a) a dark circular patch is observed in the vicinity where the measured severity has a magnitude of about 0.2, and the measured velocity is close to 1. From the colour bar it is seen that the adjustment ratio is 1.8. In other words, if the utility vehicle traverses a section of road at a slow speed of 1 m/s and it experience a moderate vehicle response severity of 0.2, then the GP predicts that the utility vehicle will experience a response severity which is 1.8 times greater if it were to travel over that section of road at 2.5 m/s.

If the vehicle travels over a section of road at 1 m/s and the measured severity is small, then the adjustment ratio will be close to 1. This corresponds to the utility vehicle travelling over a smooth road, where the vehicle response is approximately consistent, regardless of the vehicle speed.

When the measured speed as indicated on the y-axis exceeds the target speed of 2.5 m/s the adjustment ratio tends to drop beneath 1 to indicate that it is expected that the vehicle will respond less violently if it is driven slower over such a section of road.

It might be expected that a very large severity measurement such as 0.6 at a low speed should have a very large adjustment ratio, but figure 8.5 (a) indicates a near unity adjustment ratio. This is explained by noticing in figure 8.5 (b) the large uncertainty of the GP. The limited amount of training data did not contain any examples of such severe road conditions. In the absence of training examples the GP predicts the most likely severity value, which is simply the mean of all the target values.

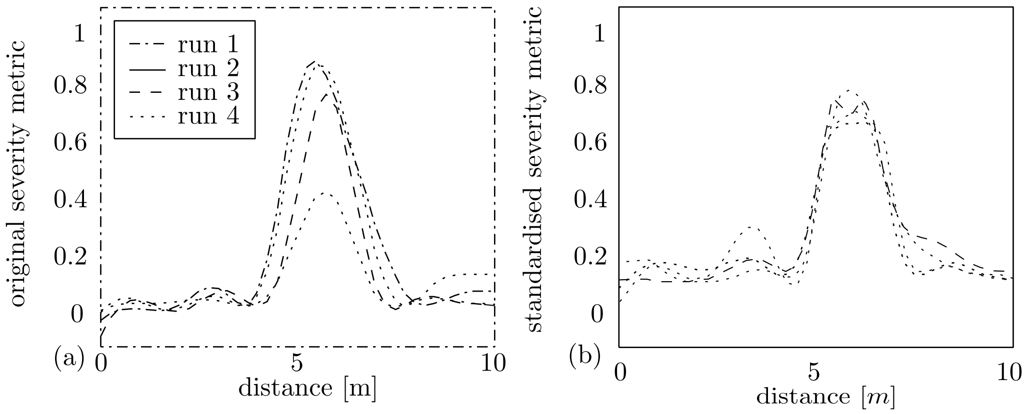


Figure 8.6: Standardising the training data to a consistent speed of 2.5 m/s.

Figures 8.6 (a) and (b) illustrate the results where the calibration function standardise each of the four training measurements to a consistent speed of 2.5 m/s. The severities as estimated from the four measurements are now much more consistent. The peak from run 4 which was measured at a speed of approximately 1.4 m/s now corresponds fairly well with the peak from run 1 which was measured at close to 3.2 m/s. It is however also seen that the supposedly near zero excitation just before the hump of run 2 is wrongly amplified. Better training data should avoid this problem.

It is thus seen that the calibration function makes intuitive sense within the range of training data. The training data represent both smooth road surfaces, as well as a road surface with a significant excitation. The training data also represents a speed range from about 1.4 to 3.2 m/s which is the typical range which was measured in underground operating conditions. It is thus expected that the calibration function is fairly representative of the conditions under which it will be expected to perform calibration. However, it is also anticipated that better performance might be obtained if training data were collected over a slightly larger range of speeds, and especially over a greater variety of excitations.

8.4.3 Model evaluation and results

The ability of the calibration function to generalise well on novel data is subsequently investigated. The utility vehicle was repeatedly driven over an almost 700 m track of underground road. The human operated push button was used to indicate the start and end of each run. Measurements were taken when the vehicle drove in the forward direction, as well when the vehicle returned. The forward and return runs did not follow the same path, so that it is more consistent to split the forward and return runs into two sets of data. Each set of data contained three runs, which are representative of the vehicle response as measured at different, slightly fluctuating, speeds.

The envelopes of the acceleration signals are again computed to serve as measured road severity. The estimated vehicle speed is subsequently used to convert the measured severity to a function of distance, rather than time.

The calibration function is evaluated on both the forward and return data sets. Two methods are used to investigate the validity of the calibration function. The first method will calibrate the severity from one run so as to resemble the speed of the other run. The second method will compare the standardised severity measurements.

8.4.3.1 Forward run

The data set which contains the three runs as measured in the forward direction is first investigated.

It has been explained that the input vector is defined as $x = [v_i \ r_i \ v_j]$ and the target value as $y = [r_j]$. The first two entries in the input vector (v_i and r_i) contain information that indicates the underlying road condition, while the third entry v_j determines the speed at which it is desirable to know the expected or calibrated severity r_j. The calibration function is applied pointwise to the measured severity. Although it is ultimately the aim to standardise all the severity measurements to a constant speed to allow for consistent road quality comparison, the calibration function is not limited in this fashion. The target speed v_j may be set equal to the measured (possibly fluctuating) speed of any other run. Each of the three forward runs were performed over nearly the same road profile. It is therefore expected that one run would more closely resemble another run, if it is speed calibrated to the measured speed of that second run.

Table 8.1 investigates how the characteristics of one severity measurement is affected if it is speed calibrated to the speed of a second run. The first column indicates which run is calibrated, and the second column indicates the speed of which run is used as target speed. The third column is the velocity difference between the two runs. The fourth, fifth, and six columns respectively indicate the

Table 8.1: Comparison of the three forward runs by calibrating one run to the speed of the other. The three runs had respective means speeds of 3.23 m/s, 3.15m/s and 2.08m/s.

From	To	Delta speed	$O_{rms} \times 10^{-2}$	C_{rms} 10^{-2}	T_{rms} 10^{-2}	Org. error	Cal. error
run 1	run 2	-0.08	16.0	15.1	15.6	2.1 %	3.5 %
run 2	run 1	0.08	15.6	15.7	16.0	2.2 %	1.8 %
run 1	run 3	-1.15	16.0	9.5	11.4	40.5 %	16.7 %
run 3	run 1	1.15	11.4	17.4	16.0	28.8 %	8.9 %
run 2	run 3	-1.07	15.6	9.4	11.4	37.6 %	16.8 %
run 3	run 2	1.07	11.4	17.0	15.6	27.3 %	8.9 %

root mean square (rms) of the measured severity of the original run O_{rms}, the rms value of its point wise calibrated severity C_{rms}, and the rms value of the severity metric of the run which served as target speed T_{rms}.

The original error (Org. error) is defined as

$$\text{Orginal error} = \frac{O_{rms} - T_{rms}}{T_{rms}} \tag{8.4.1}$$

while the calibrated error (Cal. error) is defined as

$$\text{Calibrated error} = \frac{C_{rms} - T_{rms}}{T_{rms}} \tag{8.4.2}$$

The calibrated error generally is much smaller than the original error.

Figures 8.7 (a) and (b) investigate the statistics of the measured and pointwise calibrated severity measures. The probability distributions are estimated by means of Parzen-window density estimation [90] as implemented in the Matlab ksdensity function. Normal distribution kernels and a consistent window length of 0.4 are used. The abscissa indicates the magnitude of the severity, while the ordinate indicates the relative frequency that any severity occured.

Figure 8.7 (a) indicates the process where the severity measurement of run 3 is calibrated to the measured speed of run 1. From table 8.1 it is seen that the measured speed from run 1 is approximately 1.15 m/s faster than the speed of run 3. It is therefore expected that the calibrated severity will be greater than the measured severity. The dashed line in figure 8.7 (a) which represents the original

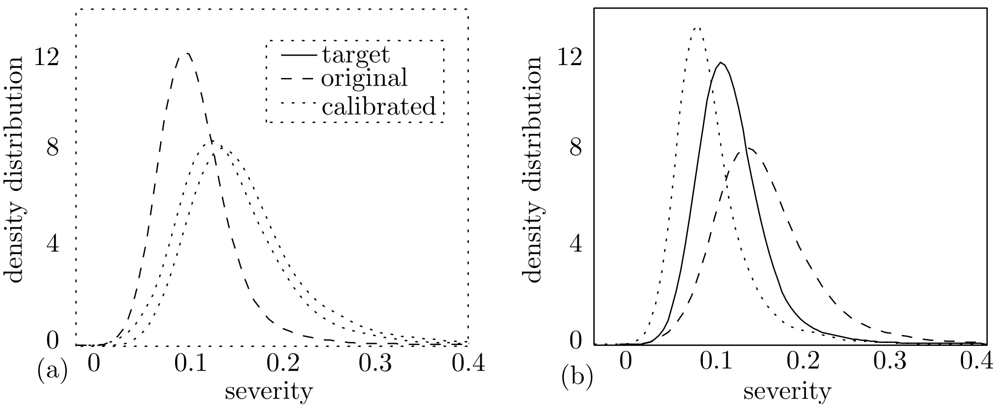


Figure 8.7: Probability distributions of the continuous severity measurements, (a) where each of the datum point from the severity measurement of run 3 have been calibrated to the speed of run 1, and (b) and the reverse where the continuous severity measurement from run 1 is pointwise calibrated to the speed of run 3 .

measurement (run 1) indicates a mode (peak) at roughly 0.1, which is significantly less than the measurement of run 2. The calibrated severity from run 3 corresponds much closer to the measured response of run 1.

Figure 8.7 (b) indicates the process where run 1 is calibrated to the lower speed of run 3. In this circumstance the calibration function over compensated by reducing the severity a bit too much. From table 8.1 it is seen that the calibrated rms error is quite large (16.7 %), but it is still smaller than the original error (40.5 %).

The calibration function is subsequently investigated in the context which it is intended for, namely to standardise the severity metric to allow for consistent road quality evaluation. Figure 8.8 (a) depicts the original severity measurements, while figure 8.8 (b) shows the severities as calibrated to a consistent speed of 2.5 m/s. In figure 8.8 (a) the grey area specify the interval between the minimum and maximum measured severity measurements, and in 8.8 (b) it denotes the interval between the minimum and maximum calibrated severities. The dashed lines represent the means. The grey areas provide an indication of the variation between severity measurements obtained under different speeds.

The calibrated severities are presented in figure 8.8 (b). Some variation is still observed. This may partly indicate that the calibration function needs to be refined, but may also be due to the utility vehicle following slightly different routes over the same track introducing some unavoidable measurement variations. In general the calibrated road severities are much more consistent, so that any one of the calibrated runs may serve as support for maintenance decisions.

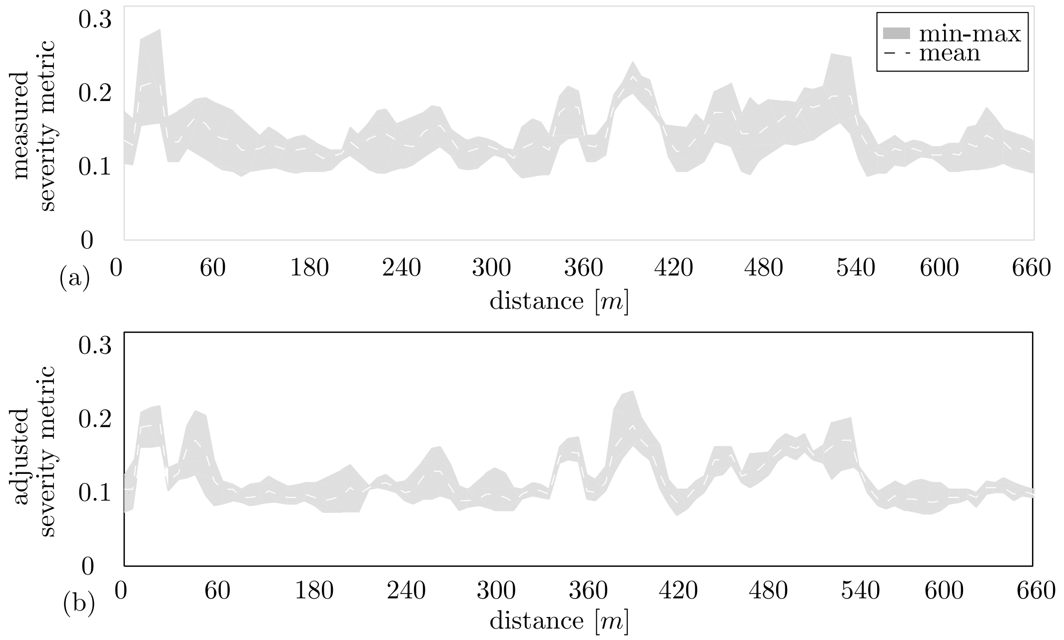


Figure 8.8: Comparison of consistency (a) between measured severities (b) and between the standardised severities which are calibrated to consistent speed of 2.5 m/s

8.4.3.2 Road section 2

Results, similar to those produced with the forwards runs, are presented for the reverse runs.

Table 8.2: Comparison of the three reverse runs by calibrating one run to the speed of the other. The three runs had respective mean speeds of 2.81 m/s, 2.92 m/s and 1.56 m/s.

From	To	Delta speed	O_{rms} 10^{-2}	E_{rms} 10^{-2}	T_{rms} 10^{-2}	Org. error	Cal. error
run 1	run 2	0.11	13.8	14.8	14.6	5.5 %	1.8 %
run 2	run 1	-0.11	14.6	13.6	13.8	5.1 %	1.9 %
run 1	run 3	-1.25	13.8	7.9	7.4	88.0 %	7.8 %
run 3	run 1	1.25	7.4	11.4	13.8	46.8 %	17.7 %
run 2	run 3	-1.36	14.6	7.5	7.4	97.6 %	2.3 %
run 3	run 2	1.36	7.4	11.9	14.6	49.4 %	18.0 %

Table 8.2 again investigates how the characteristics of one severity measurement is affected if it is speed calibrated to the speed of a second run. The magnitude of the mean speed difference between run 2 and run 3 is 1.36 m/s, which is the largest difference observed between any two runs, including those in the forward direction. The calibrated error is seen to be significantly smaller compared to the

original errors.

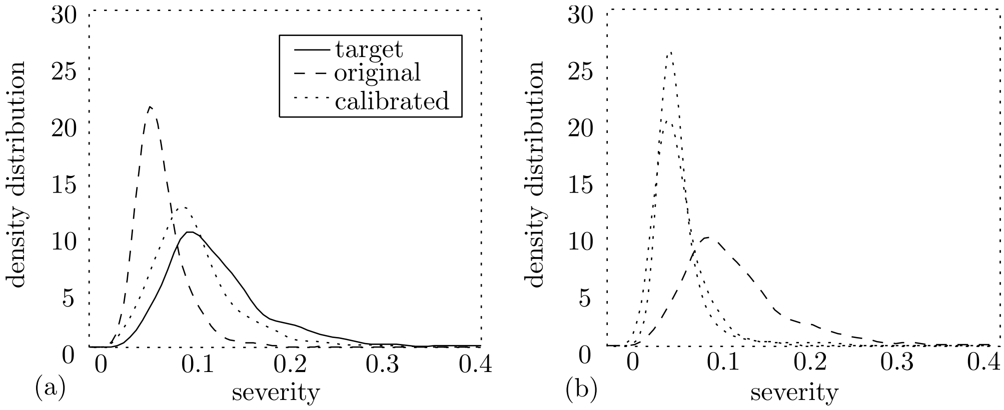


Figure 8.9: Probability distributions of the magnitude of the severities, (a) where the severity measurement of run 3 is calibrated to the speed of run 2, (b) and the reverse where the severity measurement from run 2 is calibrated to the speed of run 3 .

The probability distributions of runs 2 and 3 are subsequently also investigated in figures 8.9 (a) and (b). Figure 8.9 (a) illustrates the process where the severity measurement of run 3 is calibrated to the measured speed of run 2. In figure 8.9 (b) the reverse is presented, so that the severity measurement of run 3 is calibrated to the measured speed of run 2. The modes of the calibrated severity measures correspond much better with the target distributions, so that it is clear that both runs were subject to road profiles with similar characteristics.

The last figures again compare the variability (depicted by the grey patches) associated with the measured severities (figure 8.10 (a)) with those of the calibrates severities (figure 8.10 (b)). The speed calibrated responses are very similar. Each of the three calibrated severity measures provides a good and consistent estimate of the underlying road condition.

8.5 CHAPTER CONCLUSION

This chapter proposed a waveform reconstruction methodology that performs vehicle response calibration towards the aim of consistent haul road monitoring. Gaussian Processes, which offers a robust, statistical and non parametric approach to perform inference is used to infer a calibration function from training data.

The methodology was investigated on experimental data that were obtained by measuring the response from a utility vehicle in a coal mine environment. The implemented methodology requires no prior knowledge of the vehicle, nor does it require costly tests to characterise either the vehicle or

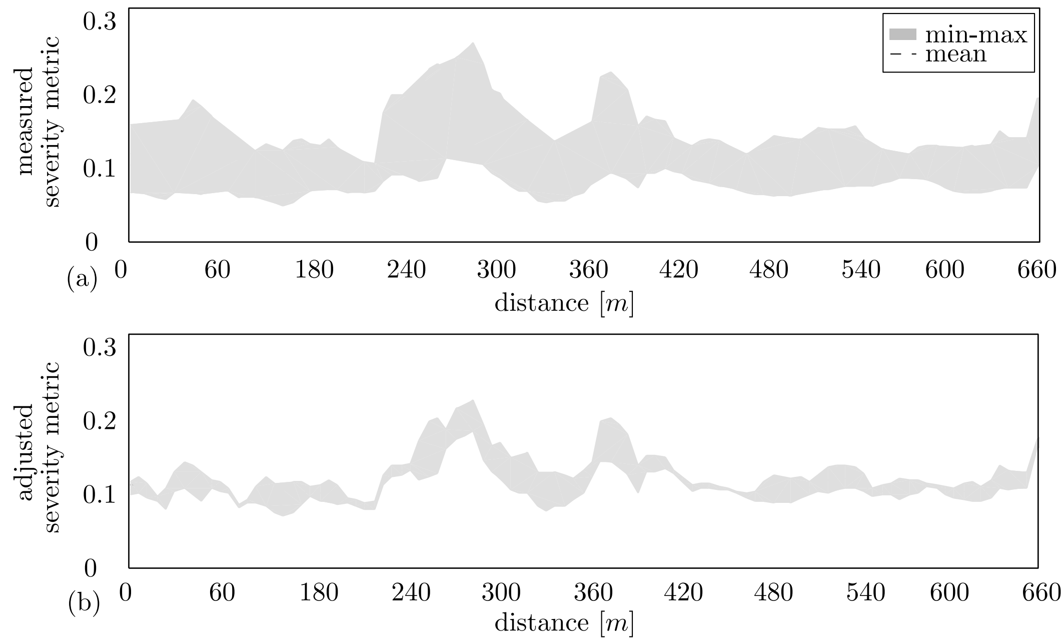


Figure 8.10: Comparison of consistency between (a) the measured severity (b) and the standardised severity which has been calibrated to a consistent speed of 2.5 m/s

individual components on it. Compared to the un-calibrated severity measurement, the standardised results are significantly more consistent. It thus appears that the proposed waveform reconstruction methodology may be used as a cost effective framework to monitor haul roads by means of vehicle response measurements.

Chapter 9

Conclusions and future research

It was explained in the introduction that a condition based monitoring system typically comprises three key aspects, namely data acquisitioning, data processing, and maintenance decision support. The focus of this thesis aligns with the second stage of data processing and interpretation. In particular the objectives of the proposed research were presented in the introduction as the development of a diagnostic framework that:

- extracts diagnostic information from machine response signals,
- extracts diagnostic information in such a manner that it is sensitive to the presence of machine faults, yet also robust in time-varying operating conditions,
- is capable of fusing data from multiple sources in order to enhance the performance of the algorithm (such as considering contextual information from sensor measurements that are correlated with the time-varying operating conditions),
- is also capable of interpreting the vibration signal when limited, noisy or no measurements of the operating conditions are available,
- is not dependent on fault historic data, destructive tests, or extensive manual preparation of training data,
- allows for a simple intuitive representation of the extracted information, such that a non-expert, or a simple classification algorithm may interpret it.

By example of six case studies on gears, alternator end windings and haul roads it was shown how machine learning techniques may be used to extract diagnostic information from machine response signals. The machine learning techniques were capable of making use of additional sensor signals, such as angular positions, coordinates and instantaneous speeds in order to separate the effects imposed on the signal due to time-varying operating conditions and those signal characteristics that were related to damage in the mechanical systems. The case studies also illustrated how it is possible to extract diagnostic information without the need for destructive tests. By effectively constraining the machine learning problems and by making use of appropriate priors the proposed solutions also work for limited and noisy data sets. The extracted diagnostic information is generally intuitive to interpret and may serve as input to automated decision rules for fault detection. Two main approaches, namely *discrepancy analysis* and *waveform reconstruction*, were proposed whereby machine learning techniques can be used to extract diagnostic information from machine response signals that are subject to time-varying operating conditions.

The discrepancy analysis approach was investigated in the context of three gear condition monitoring case studies. Discrepancy analysis essentially entails the construction of a data driven reference model that represents the vibration response of the healthy machine or system as measured for all relevant time-varying operating conditions. This reference model is subsequently used to transform a novel machine response signal into a discrepancy signal. The discrepancy transform is generally robust to time-varying operating conditions and tends to be sensitive to fault induced signal irregularities. The magnitude and periodicity of the discrepancy signal may subsequently be analysed and used to interpret to the condition of the machine. Discrepancy analysis is especially relevant to situations where the operating conditions are unknown or noisy.

Waveform reconstruction was investigated in the context of three case studies, including alternator end-winding and haul road condition monitoring. The waveform reconstruction approach is appropriate for situations where information regarding the instantaneous state space variables (e.g. shaft angular position, haul truck GPS coordinates, speed, load, etc) is available. In this framework regression models are used to approximate the machine response as a function of the state space variables. The optimized regression functions may subsequently be used to reconstruct the machine response waveforms to have certain desirable characteristics. The machine response waveform may for instance be reconstructed at a standardised steady state operating condition to allow for further analysis with conventional signal processing techniques.

There are a number of promising avenues for future research on the methods discussed in this thesis. Some suggestions with respect to discrepancy analysis include the incorporation of pre-processing techniques that are more specific to different types of machine faults. Consider for instance that vibration signal distortions that result because of gear faults are generally more deterministic, while signal distortions that are the result of defective bearings are generally more stochastic in nature. A pre-processing stage, such as self adaptive noise cancellation, may be used to decompose a vibration signal into the deterministic and stochastic components. Discrepancy analysis may subsequently separately be applied to the different components, so that two discrepancy transforms can be obtained. One discrepancy transform will be particularly sensitive to gear faults, while the other will be particularly sensitive to bearing faults. Up to now discrepancy analysis has only been applied to the original time-domain waveform. It may however be more effective to implement discrepancy analysis in the time-frequency, or in the wavelet domains. This then opens many exciting possibilities for applying Hidden Markov Models or Graphical Models. It is anticipated that significant increases in sensitivity may be obtained with these models.

There are also a number of interesting possibilities for future research on waveform reconstruction. One example is with regards to the neural network regression methodology that was implemented on the alternator end windings vibration signal. In that problem it is assumed that the signal is essentially deterministic in nature with superimposed normally distributed noise. This restrictive model may not be optimal for many vibration signals. Current research is underway where use is made of Conditional Mixture Models (where a neural network with multiple outputs is used to approximate the adaptive means and variances in a Gaussian mixture model). The Conditional Mixture Model renders it possible to model a vibration signal even when some hidden states are unobserved or when the signal noise is inherently non-Gaussian or multi-modal.

Lastly it is also believed that it would be of practical value to pursue research that aims to effectively integrate the findings of this thesis within the general framework of a condition monitoring system. It is especially believed that research should be conducted on how the extracted diagnostic information may be incorporated in decision support methodologies. This line of research should focus on aspects such as quickest detection methods, evaluation of proportional hazard models, and economic considerations with regard to the respective costs of false alarms, or false negatives that result in unexpected breakdowns.

Bibliography

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, “A review on machinery diagnostics and prognostics implementing condition-based maintenance,” *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] P. D. Samuel and D. J. Pines, “A review of vibration-based techniques for helicopter transmission diagnostics,” *Journal of sound and vibration*, vol. 282, no. 1-2, pp. 475–508, 2005.

[3] A. Heng, T. Zhang, S., and J. A. C. C., Mathew, “Rotating machinery prognostics: State of the art, challenges and opportunities,” *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.

[4] P. S. Heyns, “Tool condition monitoring using vibration measurements - a review,” *Insight - Non-Destructive Testing and Condition Monitoring*, vol. 49, pp. 447–450, 2007.

[5] S. Burguet and O. Lénard, “Assessment of an anomaly detector for jet engine health monitoring,” *International Journal of Rotating Machinery*, vol. 2011, pp. 1–11, 2011.

[6] B. L. Eggers, P. S. Heyns, and C. J. Stander, “Using computed order tracking to detect gear condition aboard a dragline,” *The Journal of The Southern African Institute of Mining and Metallurgy*, vol. 107, pp. 115–122, 2007.

[7] C. J. Stander and P. S. Heyns, “Transmission path phase compensation for gear monitoring under fluctuating load conditions,” *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1511–1522, 2006.

[8] P. D. McFadden and M. M. Toozhy, “Application of synchronous averaging to vibration monitoring of rolling element bearings,” *Mechanical Systems and Signal Processing*, vol. 14, pp. 891–906, 2000.

[9] K. S. Wang and P. S. Heyns, “Application of computed order tracking, vold-kalman filtering and emd in rotating machine vibration,” *Mechanical Systems and Signal Processing*, vol. 25, pp. 416–430, 2010.

[10] C. J. Stander and P. S. Heyns, “Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions,” *Mechanical systems and signal processing*, vol. 19, no. 4, pp. 817–835, 2005.

[11] A. Belsak and J. Flasker, “Wavelet analysis for gear crack identification,” *Engineering Failure Analysis*, vol. 16, no. 6, pp. 1983–1990, 2009.

[12] N. Baydar and A. Ball, “Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum,” *Mechanical Systems and Signal Processing*, vol. 14, no. 6, pp. 907–921, 2000.

[13] S. Burningham and N. Stankevich, “Why road maintenance is important and how to get it done,” The World Bank, Washington DC, TRN-4, Tech. Rep., 2005.

[14] R. Thompson, A. Visser, R. Miller, and T. Lowe, “Development of real-time mine road maintenance management system using haul truck and road vibration signature analysis,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1819, no. 1, pp. 305–312, 2003.

[15] A. Ruotoistenmäki and T. Seppälä, “Road condition rating based on factor analysis of road condition measurements,” *Transport Policy*, vol. 14, no. 5, pp. 410–420, 2007.

[16] D. Hugo, P. S. Heyns, R. J. Thompson, and A. T. Visser, “Haul road defect identification using measured truck response,” *Journal of Terramechanics*, vol. 45, no. 3, pp. 79–88, 2008.

[17] H. M. Ngwangwa, P. S. Heyns, F. J. J. Labuschagne, and G. K. Kululanga, “Reconstruction of road defects and road roughness classification using vehicle responses with artificial neural networks simulation,” *Journal of Terramechanics*, vol. 47, no. 2, pp. 97–111, 2010.

[18] C. Cempel, “Diagnostically oriented measures of vibroaccoustical processes,” *Journal of Sound and Vibration*, vol. 73, pp. 547–561, 1980.

[19] S. Wu, M. J. Zuo, and A. Parey, “Simulation of spur gear dynamics and estimation of fault growth,” *Journal of Sound and Vibration*, vol. 317, no. 3-5, pp. 608–624, 2008.

[20] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition," *Applied Acoustics*, vol. 53, pp. 211–226, 1998.

[21] *ISO 2631-1, Mechanical evaluation of mechanical vibration and shock - evaluation of human exposure to whole body vibration - part 1: general requirements, 1997.*, International Organization for Standardization Std.

[22] B. E. Quinn and C. C. Wilson, "Can dynamic tire forces be used as a criterion of pavement conditions?" *Joint Highway Research Project, Indiana Department of Transportation and Purdue University, West Lafayette, Indiana*, vol. 32, pp. 1–33, 1963.

[23] M. W. Sayers, T. D. Gillespie, and C. A. V. Queiroz, "The international road roughness experiment, establishing correlation and a calibration standard for measurements," *World Bank Technical Paper*, vol. 45, pp. 1–105, 1986.

[24] T. D. Gillespie, "Everything you always wanted to know about the IRI, but were afraid to ask!" in *Road Profile Users Group Meeting, Lincoln, Nebraska*, 1992.

[25] O. Kropac and P. Mucka, "Be careful when using the international roughness index as an indicator of road unevenness," *Journal of sound and vibration*, vol. 287, no. 4-5, pp. 989–1003, 2005.

[26] J. D. Smith, *Gear noise and vibration, second edition*. Marcel Dekker, 2003.

[27] I. Khemili and M. Chouchane, "Detection of rolling element bearing defects by adaptive filtering," *European Journal of Mechanics A/Solids*, vol. 24, pp. 293–303, 2004.

[28] M. M. Ettefagh and M. H. Sadeghi, "Gear fault diagnosis via non-stationary adaptive martin distance," *Scientia Iranica*, vol. 18, no. 1, pp. 59 – 65, 2011.

[29] T. D. Popescu, "Blind separation of vibration signals and source change detection-application to machine monitoring," *Applied Mathematical Modelling*, vol. 34, no. 11, pp. 3408–3421, 2010.

[30] G. Gelle, M. Colas, and G. Delaunay, "Blind sources separation applied to rotating machines monitoring by acoustical and vibrations analysis," *Mechanical Systems and Signal Processing*, vol. 14, no. 3, pp. 427–442, 2000.

[31] R. B. Randall, N. Sawalhi, and M. Coats, “A comparison of methods for separation of deterministic and random signals,” *The International Journal of Condition Monitoring*, vol. 1, pp. 11–19, 2011.

[32] S. Braun, “The synchronous (time domain) average revisited,” *Mechanical Systems and Signal Processing*, vol. 25, no. 4, pp. 1087–1102, 2011.

[33] E. Bechhoefer and M. Kingsley, “A review of time synchronous average algorithms,” in *Annual Conference of the Prognostics and Health Management Society*, 2009.

[34] L. Mdlazi, H. Stander, C. J., and T. P. S., Marwala, “Data reduction in gear vibration analysis,” in *Proceedings of the 4th International Conference on Condition Monitoring*, 2007, pp. 1415–1425.

[35] R. M. Stewart, “Some useful data analysis techniques for gearbox diagnostics,” in *Report MHM/R/10/77*. Machine Health Monitoring Group, Institute of Sound and Vibration Research, University of Southampton, 1977.

[36] D. Lin, M. Wiseman, D. Banjevic, and A. K. S. Jardine, “An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure,” *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 993–1007, 2004.

[37] A. J. Miller, “A new wavelet basis for the decomposition of gear motion error signals and its application to gearbox diagnostics,” in *MSc Thesis*. The Graduate School, The Pennsylvania State University, 1999.

[38] W. Wang and A. K. Wong, “Autoregressive model-based gear fault diagnosis,” *Journal of vibration and acoustics*, vol. 124, pp. 172–179, 2002.

[39] J. Antoni and R. B. Randall, “Differential diagnosis of gear and bearing faults,” *Journal of Vibration and Acoustics, Transactions of the ASME*, vol. 124, pp. 165–177, 2002.

[40] R. Zimroz and W. Bartelmus, “Application of adaptive filtering for weak impulsive signal recovery for bearings local damage detection in complex mining mechanical systems working under condition of varying load,” in *Solid State Phenomena*, 2012.

[41] R. Makowski and R. Zimroz, “Adaptive bearings vibration modelling for diagnosis,” in *Lecture Notes in Comp. Sci. (in Lecture Notes in Artificial Intellig.) LNAI 6943*, 2011.

[42] R. B. Randall, *Vibration-based condition monitoring*. Wiley, 2011.

[43] C. J. Stander, P. S. Heyns, and W. Schoombie, “Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions,” *Mechanical systems and signal processing*, vol. 16, no. 6, pp. 1005–1024, 2002.

[44] P. E. Uys, P. S. Els, and M. J. Thoresson, “Criteria for handling measurement,” *Journal of terramechanics*, vol. 43, no. 1, pp. 43–67, 2006.

[45] D. F. Elliot, *Handbook of digital signal processing*. Academic press limited, 1987.

[46] J. D. Smith, *Vibration measurement and analysis*. Butterworths, 1989.

[47] M. E. Badaoui, F. Guillet, and J. Danière, “New applications of the real cepstrum to gear signals, including definition of a robust fault indicator,” *Mechanical systems and signal processing*, vol. 18, no. 5, pp. 1031–1046, 2004.

[48] G. Dalpiaz, A. Rivola, and R. Rubini, “Effectiveness and sensitivity of vibration processing techniques for local fault detection in gears,” *Mechanical Systems and Signal Processing*, vol. 14, no. 3, pp. 387–412, 2000.

[49] R. Dwyer, “Detection of non-gaussian signals by frequency domain kurtosis estimation,” in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP’83.*, 8, Ed., 1983, pp. 607–610.

[50] J. Antoni, “The spectral kurtosis: a useful tool for characterising non-stationary signals,” *Mechanical Systems and Signal Processing*, vol. 20, no. 2, pp. 282–307, 2006.

[51] C. Ottonello and S. Pagnan, “Modified frequency domain kurtosis for signal processing,” *Electronics Letters*, vol. 30, no. 14, pp. 1117–1118, 1994.

[52] T. Barszcz and R. B. Randall, “Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine,” *Mechanical Systems and Signal Processing*, vol. 23, no. 4, pp. 1352–1365, 2009.

[53] I. Yesilyurt, “Fault detection and location in gears by the smoothed instantaneous power spectrum distribution,” *NDT & E International*, vol. 36, no. 7, pp. 535–542, 2003.

[54] W. Bartelmus, “Diagnostic information on gearbox condition for mechatronic systems,” *Transactions of the Institute of Measurement and Control*, vol. 25, no. 5, p. 451, 2003.

[55] R. G. Stockwell, L. Mansinha, and L. R. P., "Localization of the complex spectrum: the s transform," *IEEE Transactions on Signal Processing*, vol. 44, pp. 998–1001, 1996.

[56] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, and W. Jones, "Smoothed pseudo-wigner distribution, choi-williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison," *Signal Processing*, vol. 43, no. 2, pp. 149–168, 1995.

[57] *ISO 8608: 1995(E). Mechanical vibration - road surface profiles - reporting of measured data; 1995.*, International Organization for Standardization Std.

[58] B. Li, P. Zhang, D. Liu, S. Mi, and P. Liu, "Classification of time-frequency representations based on two-direction 2dlda for gear fault diagnosis," *Applied Soft Computing*, vol. 11, no. 8, pp. 5299–5305, 2011.

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

[60] G. G. Yen and K. C. Lin, "Wavelet packet feature extraction for vibration monitoring," *Industrial Electronics, IEEE Transactions on*, vol. 47, no. 3, pp. 650–667, 2000.

[61] J. E. Lopez, R. R. Tenney, and J. C. Deckert, "Fault detection and identification using real-time wavelet feature extraction," in *Time-Frequency and Time-Scale Analysis, 1994., Proceedings of the IEEE-SP International Symposium on.* IEEE, 1994, pp. 217–220.

[62] D. Boulahbal, M. F. Golnaraghi, and F. Ismail, "Amplitude and phase wavelet maps for the detection of cracks in geared systems," *Mechanical systems and signal processing*, vol. 13, no. 3, pp. 423–436, 1999.

[63] X. Wang, V. Makis, and M. Yang, "A wavelet approach to fault diagnosis of a gearbox under varying load conditions," *Journal of Sound and Vibration*, vol. 329, no. 9, pp. 1570–1585, 2010.

[64] G. K. Singh and S. A. K. S. Ahmed, "Vibration signal analysis using wavelet transform for isolation and identification of electrical faults in induction machine," *Electric Power Systems Research*, vol. 68, no. 2, pp. 119–136, 2003.

[65] T. W. S. Chow and S. Hai, "Induction machine fault diagnostic analysis with wavelet technique," *Industrial Electronics, IEEE Transactions on*, vol. 51, no. 3, pp. 558–565, 2004.

[66] S. Seker and E. Ayaz, "Feature extraction related to bearing damage in electric motors by wavelet analysis," *Journal of the Franklin institute*, vol. 340, no. 2, pp. 125–134, 2003.

[67] V. Purushotham, S. Narayanan, and S. A. N. Prasad, "Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden markov model based fault recognition," *NDT & E International*, vol. 38, no. 8, pp. 654–664, 2005.

[68] C. Smith, C. M. Akujuobi, P. Hamory, and K. Kloesel, "An approach to vibration analysis using wavelets in an application of aircraft health monitoring," *Mechanical systems and signal processing*, vol. 21, no. 3, pp. 1255–1272, 2007.

[69] D. Hester and A. González, "A wavelet-based damage detection algorithm based on bridge acceleration response to a vehicle," *Mechanical Systems and Signal Processing*, vol. 28, pp. 145–166, 2011.

[70] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, pp. 639–697, 2006.

[71] A. L. Hurd and A. Miamee, *Periodically correlated random sequences, spectral theory and practice*. John Wiley & Sons, Inc, 2007.

[72] J. Antoni, "Cyclostationarity by examples," *Mechanical Systems and Signal Processing*, vol. 23, pp. 987–1036, 2009.

[73] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N. C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, vol. 454, no. 1971, p. 903, 1998.

[74] S. J. Loutridis, "Damage detection in gear systems using empirical mode decomposition," *Engineering Structures*, vol. 26, no. 12, pp. 1833–1841, 2004.

[75] D. Yu, Y. Yang, and J. Cheng, "Application of time-frequency entropy method based on hilbert-huang transform to gear fault diagnosis," *Measurement*, vol. 40, no. 9-10, pp. 823–830, 2007.

[76] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using fft of intrinsic mode functions in hilbert-huang transform," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2607–2615, 2007.

[77] F. Bonnardot, M. El Badaoui, R. B. Randall, J. Danière, and F. Guillet, “Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation),” *Mechanical Systems and Signal Processing*, vol. 19, pp. 766–785, 2005.

[78] J. Urbanek, T. Barszcz, N. Sawalhi, and R. B. Randall, “Comparison of amplitude-based and phase-based methods for speed tracking in application to wind turbines,” *Metrology and measurement systems*, vol. 18, pp. 295–304, 2011.

[79] F. Combet and L. Gelman, “An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor,” *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2590–2606, 2007.

[80] F. Combet and R. Zimroz, “A new method for the estimation of the instantaneous speed relative fluctuation in a vibration signal based on the short time scale transform,” *Mechanical Systems and Signal Processing*, vol. 23, pp. 1382–1390, 2009.

[81] F. Millioz and N. Martin, “Time-frequency segmentation for engine speed monitoring,” in *Pattern Recognition in Acoustics and Vibration, Proc. Of Thirteen International Congress on Sound and Vibration, Stratford upon Avon*, 2010.

[82] R. Zimroz, F. Millioz, and N. Martin, “A procedure of vibration analysis from planetary gearbox under non-stationary cyclic operations for instantaneous frequency estimation in time-frequency domain,” in *Proceedings of Condition Monitoring, Stratford upon Avon, UK*, 2010.

[83] R. Zimroz, T. Barszcz, J. Urbanek, W. Bartelmus, F. Millioz, and N. Martin, “Measurement of instantaneous shaft speed by advanced vibration signal processing - application to wind turbine gearbox,” *Metrology and Measurement Systems*, vol. 18, pp. 701–712, 2011.

[84] K. Gryllias and I. Antoniadis, “Application of the energy operator separation algorithm (eosa) for the instantaneous amplitude and frequency calculation of nonlinear dynamic systems response,” in *Proceedings of the ASME International Design Engineering Conference and Computers and Information in Engineering Conference IDECT/CIE, San Diego, USA.*, 2009.

[85] Y. Christos, K. Gryllias, and I. Antoniadis, “Instantaneous frequency in rotating machinery using a harmonic signal decomposition (hard) parametric method.” in *Proceedings of the ASME International Design Engineering Conference and Computers and Information in Engineering Conference IDECT/CIE, San Diego, USA.*, 2009.

[86] M. M. Rashid and K. Tsunokawa, "Potential bias of response type road roughness measuring systems: causes and remedial measures," *Open Transportation Journal*, vol. 2, pp. 65–73, 2008.

[87] A. Widodo and B. S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.

[88] L. Wang, J. Chu, and J. Wu, "Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process," *International Journal of Production Economics*, vol. 107, no. 1, pp. 151–163, 2007.

[89] R. B. Randall, "The application of fault simulation to machine diagnostics and prognostics," *ICSV16, Krakow, Poland*, vol. 5, no. 9, pp. 81–89, 2009.

[90] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.

[91] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification, second edition*. A Wiley-Interscience Public, 2001.

[92] M. Timusk, M. Lipsett, and C. K. Mechefske, "Fault detection using transient machine signals," *Mechanical Systems and Signal Processing*, vol. 22, no. 7, pp. 1724–1749, 2008.

[93] A. S. Malegaonkar, A. M. Ariyaeeinia, and P. Sivakumaran, "Efficient speaker change detection using adapted gaussian mixture models," in *IEEE Transactions on Audio, Speech, and Languadge Processing*, vol. 15, no. 6, 2007, pp. 1859–1869.

[94] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*. Springer-Verlag London Limited, 1998.

[95] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. the MIT Press, 2005.

[96] D. J. C. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

[97] C. E. Rasmussen, "Evaluation of gaussian processes and other methods for non-linear regression," Ph.D. dissertation, Graduate Department of Computer Science, University of Toronto, 1996.

[98] J. M. Wang, D. J. Fleet, and A. Hertzmann, "Gaussian process dynamical models for human motion," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 30, 2008, pp. 283–298.

[99] D. Miljkovic, "Review of novelty detection methods," in *Information and Communication Technology, Electronics and Microelectronics, MIPRO Proceedings of the 33rd International Convention*, 2010.

[100] S. Hawkins, H. He, G. Williams, and R. Baxter, "Outlier detection using replicator neural networks," in *Data Warehousing and Knowledge Discovery*, ser. Lecture Notes in Computer Science, Y. Kambayashi, W. Winiwarter, and M. Arikawa, Eds. Springer Berlin Heidelberg, 2002, vol. 2454, pp. 113–123.

[101] V. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artificial Intelligence Review*, vol. 22, pp. 85–126, 2004, 10.1023/B:AIRE.0000045502.10941.a9. [Online]. Available: http://dx.doi.org/10.1023/B:AIRE.0000045502.10941.a9

[102] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," in *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, ser. SIGMOD '01. New York, NY, USA: ACM, 2001, pp. 37–46. [Online]. Available: http://doi.acm.org/10.1145/375663.375668

[103] F. K. Choy, V. Polyshchuk, J. J. Zakrajsek, R. F. Handschuh, and D. P. Townsend, "Analysis of the effects of surface pitting and wear on the vibration of a gear transmission system," *Tribology International*, vol. 29, no. 1, pp. 77–83, 1996.

[104] C. D. Begg, T. Merdes, C. Byington, and K. Maynard, "Dynamics modeling for mechanical fault diagnostics and prognostics," in *Maintenance and Reliability Conf.* Citeseer, 1999.

[105] N. Sawalhi and 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," *Mechanical Systems and Signal Processing*, vol. 22, no. 8, pp. 1924–1951, 2008.

[106] S. Jia, I. Howard, and J. Wang, "The dynamic modeling of multiple pairs of spur gears in mesh, including friction and geometrical errors," *International Journal of Rotating Machinery*, vol. 9, no. 6, pp. 437–442, 2003.

[107] F. Chaari, T. Fakhfakh, and M. Haddar, “Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness,” *European Journal of Mechanics-A/Solids*, vol. 28, no. 3, pp. 461–468, 2009.

[108] N. H. Pontoppidan and J. Larsen, “Unsupervised condition change detection in large diesel engines,” in *Neural Networks for Signal Processing, IEEE XIII Workshop on*, 2003, pp. 565–574.

[109] P. Smyth, “Model selection for probabilistic clustering using cross-validated likelihood,” *Statistics and Computing*, vol. 9, pp. 63–72, 2000.

[110] P. D. McFadden, “Determining the location of a fatigue crack in a gear from the phase of the change in the meshing vibration,” *Mechanical Systems and Signal Processing*, vol. 2, no. 4, pp. 403–409, 1988.

[111] W. Bartelmus, “Mathematical modelling and computer simulations as an aid to gearbox diagnostics,” *Mechanical Systems and Signal Processing*, vol. 15, pp. 855–871, 2001.

[112] W. Bartelmus, F. Chaari, R. Zimroz, and M. Haddar, “Modelling of gearbox dynamics under time varying non-stationary operation for distributed fault detection and diagnosis.” *European Journal of Mechanics - A/Solids*, vol. 29, pp. 637–646, 2010.

[113] M. T. Khabou, N. Bouchaala, F. Chaari, T. Fakhfakh, and M. Haddar, “Study of a spur gear dynamic behavior in transient regime,” *Mechanical Systems and Signal Processing*, vol. 8, pp. 3089–3101, 2011.

[114] M. Welling, “Robust higher order statistics,” in *the Tenth International Workshop on Artificial Intelligence and Statistics*, 2005, pp. 405–412.

[115] C. E. Rasmussen and H. Nickisch, “Gaussian processes for machine learning (gpml) toolbox,” *The Journal of Machine Learning Research*, vol. 9999, pp. 3011–3015, 2010.