

Statistical gear health analysis which is robust to fluctuating loads and operating speeds

T Heyns^{a,*}, SJ Godsill^b, JP de Villiers^a, PS Heyns^c

^a*Department of Electronic and Computer Engineering, University of Pretoria; Defence Peace Safety and Security (DPSS) division of the Council for Scientific and Industrial Research (CSIR) of South Africa*

^b*Signal Processing and Communications Laboratory, Electronic Engineering, Cambridge*

^c*Dynamic Systems Group, Department of Mechanical and Aeronautical Engineering, University of Pretoria*

Abstract

Condition-based maintenance is concerned with the collection and interpretation of data to support maintenance decisions. The non-intrusive nature of vibration data enables the monitoring of enclosed systems such as gearboxes. It remains a significant challenge to analyze vibration data that are generated under fluctuating operating conditions. This is especially true for situations where relatively little prior knowledge regarding the specific gearbox is available. It is therefore investigated how an adaptive time series model, which is based on Bayesian model selection, may be used to remove the non fault related components in the structural response of a gear assembly to obtain a residual signal which is robust to fluctuating operating conditions. A statistical framework is subsequently proposed which may be used to interpret the structure of the residual signal in order to facilitate an intuitive understanding of the condition of the gear system. The proposed methodology is investigated on both simulated and experimental data from a single stage gearbox.

Keywords: Gear maintenance, Residual analysis, Statistical gear model, Bayesian model selection

1. Introduction

Condition-based maintenance (CBM) is concerned with the collection and interpretation of data to support maintenance decisions. Vibration analysis is especially applicable to the monitoring of enclosed systems, such as gearboxes, due to its non-intrusive nature [1]. A significant challenge remains with the monitoring of gears under fluctuating operating conditions [2]. Fluctuating loads tend to cause amplitude modulation of the vibration signal, while fluctuating rotational speeds induce frequency modulation [2]. Bayer and Ball [3] 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. It may subsequently become difficult to differentiate between load-dependent and fault-dependent characteristics in the time or spectral domains [1].

The inability of spectral domain analysis to interpret non-stationary waveforms may to some extent be overcome by using time-frequency transforms (including the short-time Fourier transform, Wigner-Ville, and Choi-Williams distributions [3, 4]) or time-scale transforms (including Wavelets and Wavelet packet analysis [5, 6, 7]) to extract features which could be used to detect gear damage in the presence of fluctuating loading conditions. However

*Corresponding author

Email address: theoheyne@gmail.com (T Heyns)

manual interpretation of these transforms tends to be difficult and requires the expertise of highly trained and experienced personnel.

Automatic pattern recognition algorithms, such as Neural networks [5, 6, 7], Support vector machines [8], and Fuzzy logic systems [9], avoid the need for manual interpretation of the data. However due their dependence on expensive training data - which need to be representative of various fault and operating conditions for the specific machine - it is often not possible to implement these supervised learning techniques in practice [10].

Various order domain techniques exist which attempt to remove the effects of non stationary speeds on the vibration signal. Computed order tracking resamples a signal to fixed increments of the shaft position, rendering it possible to analyse the order domain spectrum [11]. 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 [2]. Other order tracking techniques, such as Vold-Kalman filters or Gabor order tracking, assume sinusoidal waveforms to extract order information from the time domain [11]. 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 [11].

Residual signal analysis is a technique which attempts to isolate or extract the diagnostic information from the vibration waveform. A residual signal is obtained by removing the non-fault related vibration components from the signal so that only those parts of the signal which are indicative of true faults remain. A number of different approaches of obtaining residual signals have been investigated in literature. Stewart [12] computed a residual signal (RS) by eliminating the gear meshing harmonics and its adjacent sidebands from the signal average spectrum before converting the signal back to the angle domain. He subsequently defined a non-dimensional parameter (FM4) which was found to be especially sensitive as a general fault detector. Lin et al. [13] investigate a residual signal based on the work by Miller [14], 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. The residuals are shown to be sensitive to propagating gear faults. Wang and Wong [15] propose using a autoregressive (AR) model of the signal of the gear in interest to represent its healthy state. A residual signal is subsequently obtained as the difference between a future observed signal and the one step ahead prediction made by an autoregressive filter. The residual signal is shown to be an effective tool in the detection and diagnosis of gear faults.

This paper investigates two novel ideas. The first is to use an adaptive time series model (based on Bayesian model selection) to obtain a sensitive residual signal which is more robust to fluctuating operating conditions. The second idea is that the structure of the residual signal may also contain valuable diagnostic information. The proposed method will be seen to have some analogy to performing a synchronous averaging of the residual signal. This paper presents a statistical framework which interprets the structure of the residual signal with the aim of offering an intuitive interpretation of the condition of a gearbox which is robust to fluctuating operating loads or speeds.

The remainder of the paper 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.

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 gear, but the analysis presented here is not restricted to this definition. It is thus assumed that the measured signal may be expressed as follow:

$$y_m = x_m + f_m + n_m \quad (1)$$

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 $\{y_{m-i}; i = 1 \dots p\}$.

$$\hat{y}_m^{\rightarrow} = \sum_{i=1}^p w_i y_{m-i} \quad (2)$$

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 $\{w_i; i = 1 \dots p\}$, 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 $\hat{y}_m^{\rightarrow} = \mathbf{w}^T \mathbf{y}^{\rightarrow}$.

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 discrepancy in this assumption.

$$x_m = \hat{y}_m^{\rightarrow} + e_m \quad (3)$$

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:

$$r_m = y_m - \hat{y}_m^{\rightarrow} \quad (4)$$

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 $\{\mathbf{w}_j; j = 1 \dots k\}$. A sliding window of q residuals $\{r_{m-i}; i = 1 \dots q\}$ 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:

$$\hat{y}_m^{\rightarrow} = \sum_{j=1}^k P(\mathbf{w}_j|\mathbf{r}_j) \mathbf{w}_j^T \mathbf{y}^{\rightarrow} \quad (5)$$

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

$$P(\mathbf{w}_j|\mathbf{r}_j) = \frac{P(\mathbf{r}_j|\mathbf{w}_j)P(\mathbf{w}_j)}{P(\mathbf{r}_j)} \quad (6)$$

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 $r_m \stackrel{\text{iid}}{\sim} N(0, \sigma_r^2)$. The variance of the residuals σ_r^2 is estimated from the original fault free data. The likelihood function may thus be expressed as:

$$P(\mathbf{r}_j|\mathbf{w}_j) = (2\pi\sigma_r^2)^{-\frac{q}{2}} \prod_{i=1}^q \exp\left\{-\frac{(r_{m-i})^2}{2\sigma_r^2}\right\} \quad (7)$$

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 1, where fault induced deviations at time instants 3 and 4 give rise to spurious residuals at times 5 and 6.

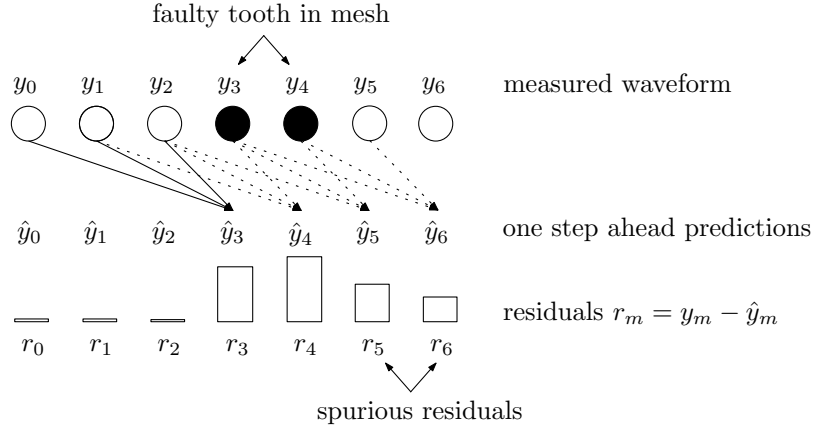


Figure 1: Temporal smearing resulting in spurious residuals

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.

Let \mathbf{y}^{\leftarrow} denote the column vector which consists of the p number of future observations $\{y_{m+i}; i = 1 \dots p\}$ 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:

$$\hat{y}_m^{\rightarrow} = \sum_{j=1}^k P(\mathbf{w}_j | \mathbf{r}_j) \mathbf{w}_j^T \mathbf{y}^{\rightarrow} \quad (8)$$

$$\hat{y}_m^{\leftarrow} = \sum_{j=1}^k P(\mathbf{w}_j | \mathbf{r}_j) \mathbf{w}_j^T \mathbf{y}^{\leftarrow} \quad (9)$$

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

$$r_m = \min(|y_m - \hat{y}_m^{\rightarrow}|, |y_m - \hat{y}_m^{\leftarrow}|) \quad (10)$$

2.2. Analysing the residual signal

A sequence of residuals have been estimated. It is now assumed that the modeling error \mathbf{e} and white noise \mathbf{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.

$$\begin{aligned}
 e_m &\stackrel{\text{iid}}{\sim} N(0, \sigma_e^2) \\
 n_m &\stackrel{\text{iid}}{\sim} N(0, \sigma_n^2)
 \end{aligned} \quad (11)$$

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

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 2.

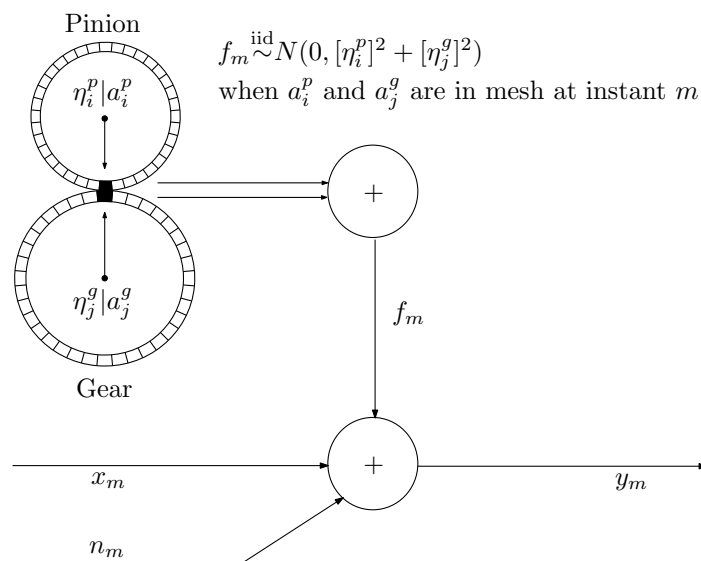


Figure 2: Noise process

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:

$$P(r_m|\eta_i^p, \eta_j^g, \sigma_e, \sigma_n) = N(0, [\eta_i^p]^2 + [\eta_j^g]^2 + \sigma_e^2 + \sigma_n^2) \quad (12)$$

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 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:

$$[\gamma_{ij}]^2 = [\eta_i^p]^2 + [\eta_j^g]^2 + [\sigma_c]^2 \quad (13)$$

The residual vector \mathbf{r} can be rewritten as a number of smaller column vectors $\{\mathbf{r}_{ij}; i = 1 \dots N, j = 1 \dots N\}$, 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:

$$[\gamma_{ij}]^2 = E([\mathbf{r}_{ij}]^T [\mathbf{r}_{ij}]) \quad (14)$$

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.

$$[\sigma_c]^2 = \min([\gamma_{ij}]^2; i = 1 \dots N, j = 1 \dots N) \quad (15)$$

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

$$[\gamma_{ij}]^2 - [\sigma_c]^2 = [\eta_i^p]^2 + [\eta_j^g]^2 \quad (16)$$

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. Study on simulated data

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

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 [16, 17, 18]. 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. [2] approximated the gear mesh stiffness profile as being sinusoidal, but some results by Jia et al. [19] 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.[19]

It is common to model the effects of tooth faults, such as a tooth root crack, by reducing the stiffness of the affected tooth [19]. 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 [20]. 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.

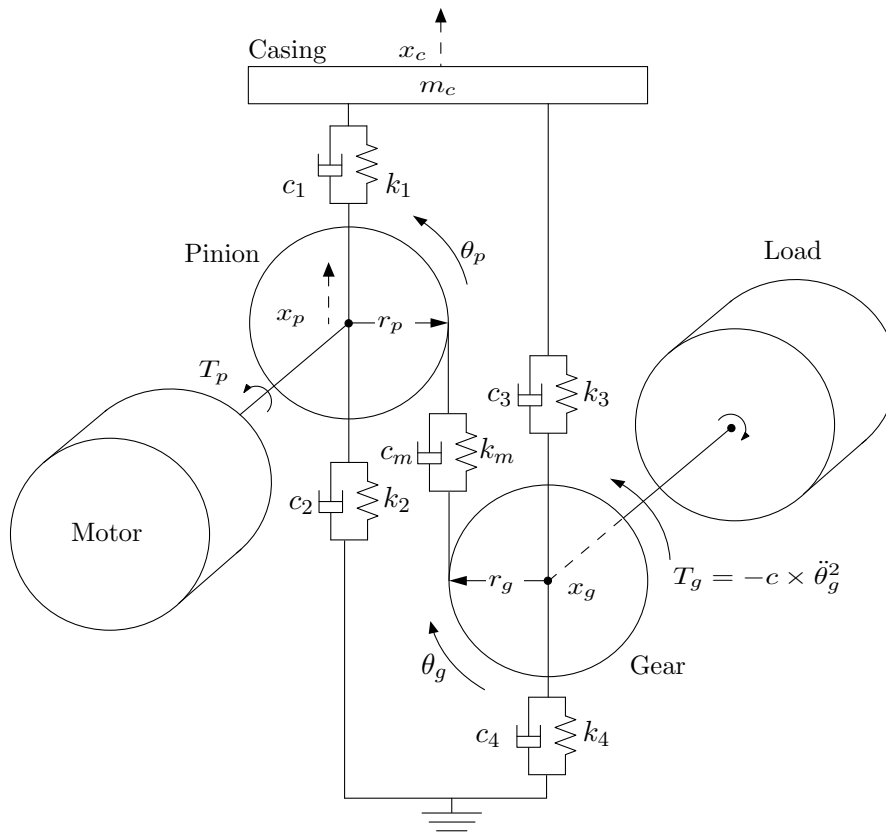


Figure 3: Dynamic one stage gear model

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 $T_g = K_s \dot{\theta}_g^2$ 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)

$$\begin{aligned} m_c \ddot{x}_c + (c_1 + c_3) \dot{x}_c - c_1 \dot{x}_p - c_3 \dot{x}_g \\ + (k_1 + k_3) x_c - k_1 x_p - k_3 x_g = 0 \end{aligned} \quad (17)$$

$$\begin{aligned} m_p \ddot{x}_p - c_1 \dot{x}_c + (c_1 + c_2 + c_m) \dot{x}_p - c_m \dot{x}_g + r_p c_m \dot{\theta}_p - r_g c_m \dot{\theta}_g \\ - k_1 x_c + (k_1 + k_2 + k_m) x_p - k_m x_g + r_p k_m \theta_p - r_g k_m \theta_g = 0 \end{aligned} \quad (18)$$

$$\begin{aligned} m_g \ddot{x}_g - c_3 \dot{x}_c - c_m \dot{x}_p + (c_m + c_3 + c_4) \dot{x}_g - r_p c_m \dot{\theta}_p + r_g c_m \dot{\theta}_g \\ - k_3 x_c - k_m x_p + (k_m + k_3 + k_4) x_g - r_p k_m \theta_p + r_g k_m \theta_g = 0 \end{aligned} \quad (19)$$

$$\begin{aligned} j_p \ddot{\theta}_p + r_p c_m \dot{x}_p - r_p c_m \dot{x}_g + r_p^2 c_m \dot{\theta}_p - r_g r_p c_m \dot{\theta}_g \\ + r_p k_m x_p - r_p k_m x_g + r_p^2 k_m \theta_p - r_p r_g k_m \theta_g = T_p \end{aligned} \quad (20)$$

$$\begin{aligned} j_g \ddot{\theta}_g - r_g c_m \dot{x}_p + r_g c_m \dot{x}_g - r_g r_p c_m \dot{\theta}_p + r_g^2 c_m \dot{\theta}_g \\ - r_g k_m x_p + r_g k_m x_g - r_g r_p k_m \theta_p + r_g^2 k_m \theta_g = -T_g \end{aligned} \quad (21)$$

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.

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	$[5 - 0.5 \times \text{sgn}(N_p \times \theta_p)] \times 10^5$	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	$1/2 \times m_g \times r_g^2$	$kg.m^2$

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 that 11 samples (from 0 to 10%) are obtained. The vibration waveform is presented in Figure 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.

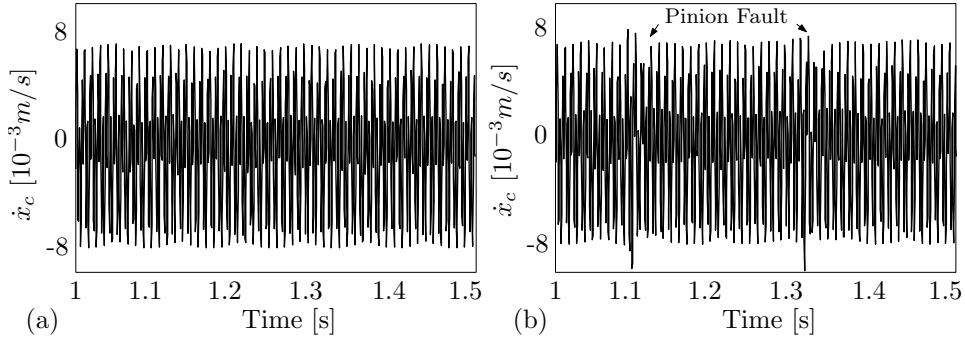


Figure 4: Gear casing velocity under steady state conditions (a) healthy, (b) 10% fault on one pinion tooth

The power spectral density (PSD) of the 11 vibration waveforms are subsequently presented in Figure 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.

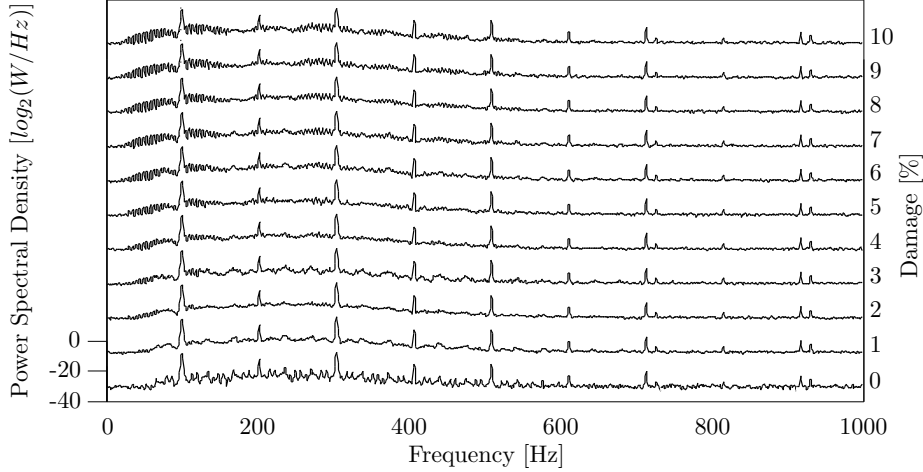


Figure 5: Log power spectral density of the constant loading conditions

The second simulation considers a time varying pinion driving load:

$$T_p = 500 \times [\tau + 0.1 \times \sin(10\pi t)] \text{ Nm} \quad (22)$$

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 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 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.

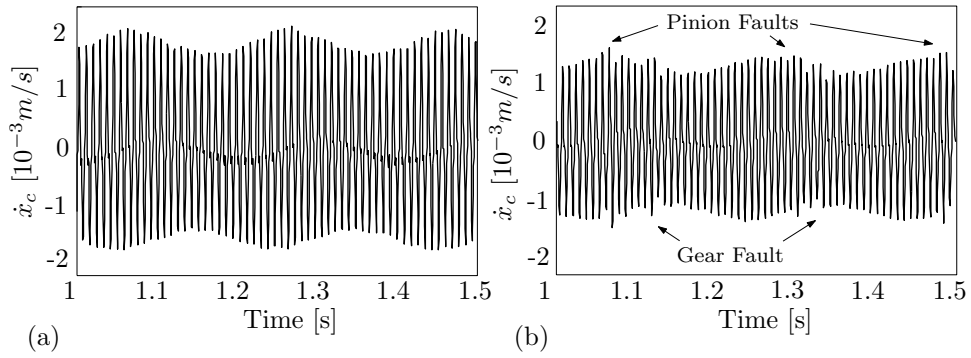


Figure 6: Gear casing vibration under steady state conditions (a) healthy, (b) 10% fault on one pinion tooth and on one gear tooth

The log PSDs of the samples as generated under the fluctuating loads are presented in figure 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.

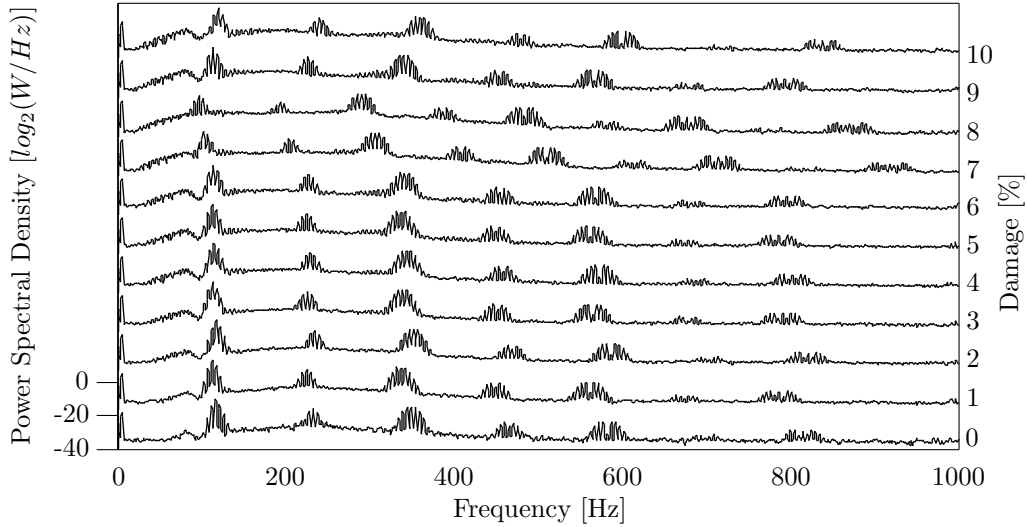


Figure 7: PSD Waterfall plot subject to both frequency shifts and spectral smearing due to fluctuating loads and operating speeds rendering it difficult to detect the propagating damage on both the pinion and gear

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 [21] for the details concerning different AR coefficient estimation techniques.

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 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.

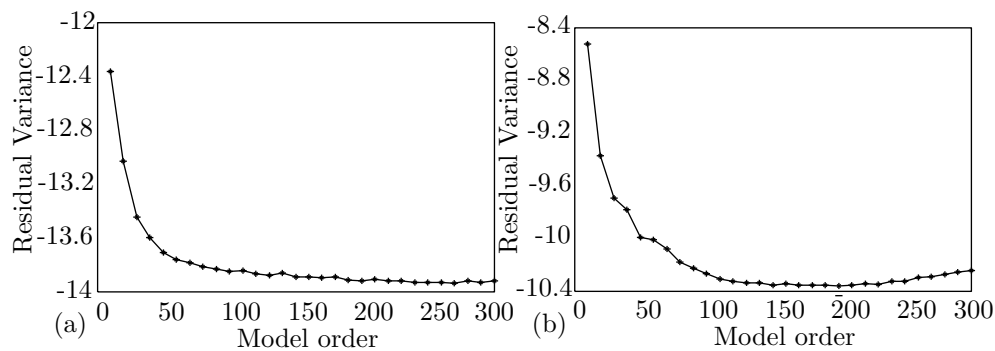


Figure 8: AR order selection

From figure 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 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.

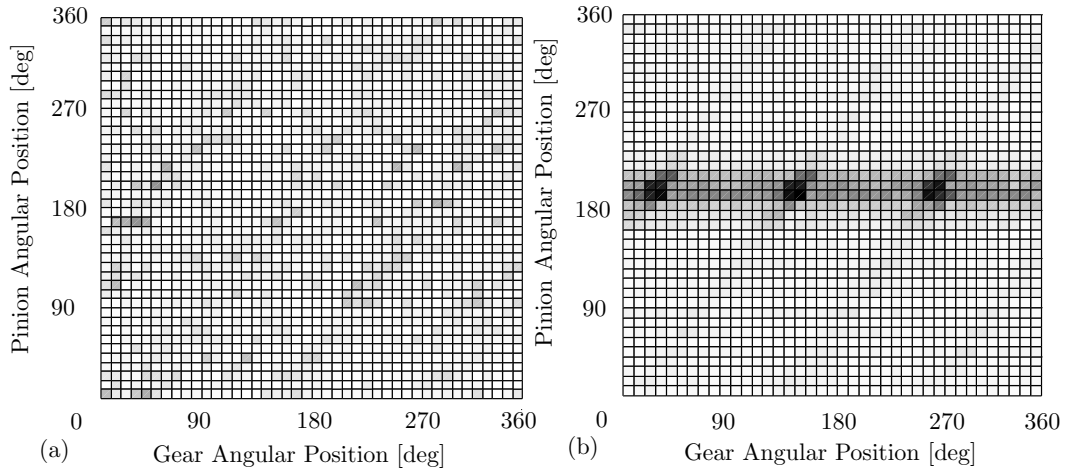


Figure 9: The lumped state variances as computed from the residuals (steady state load) associated with each of the 40×40 meshing states for (a) no damage and (b) 10% tooth crack on the pinion

Figure 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.

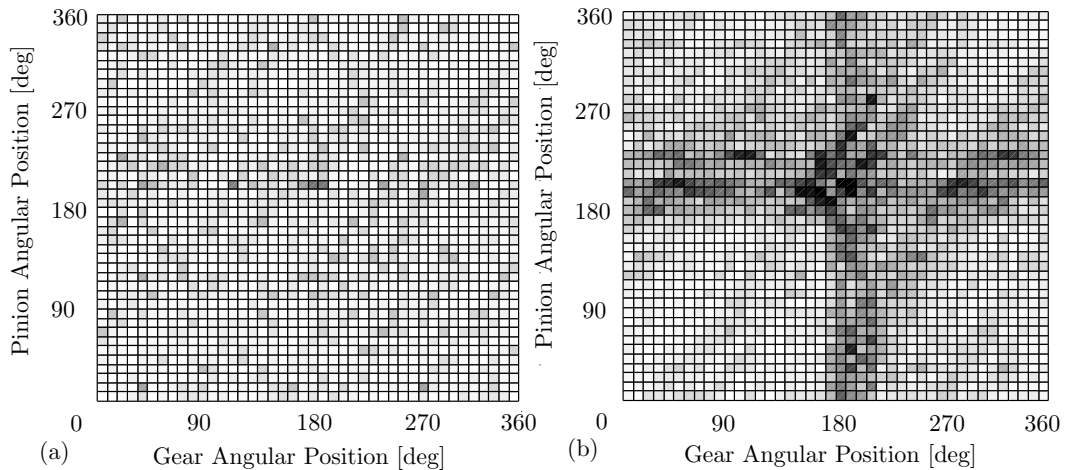


Figure 10: The lumped state variances as computed from the residuals (fluctuating loads) associated with each of the 40×40 meshing states for (a) no damage and (b) 10% tooth crack on both the pinion and gear

3.3. Results

The equations which relate the lumped state variances $[\eta^g]^2$ to the fault induced variances $[\eta^p]^2$ and $[\eta^g]^2$ 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 9 (a) and (b), while the fluctuating loads are presented in figures 10 (a) and (b). The presence of localized faults are clearly observed.

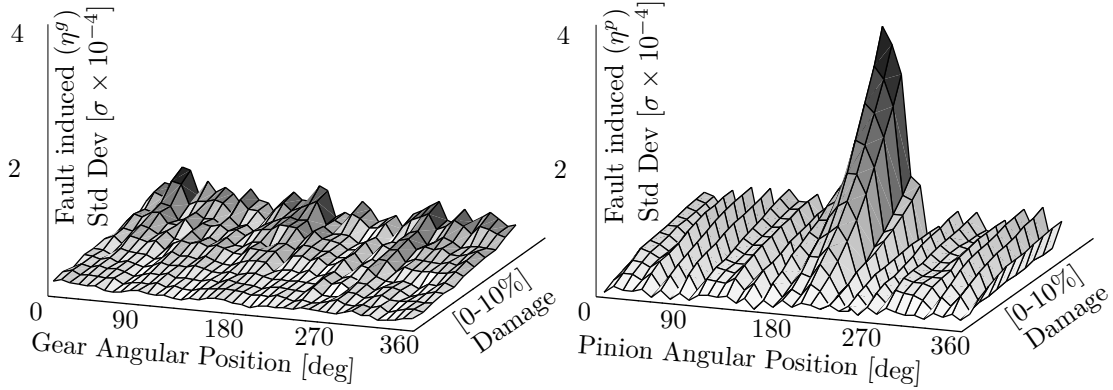


Figure 11: Estimated fault induced standard deviations (estimated under constant loads) for different (a) gear and (b) pinion angular positions as trended over different levels of damage

Figures 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 the damage on the pinion.

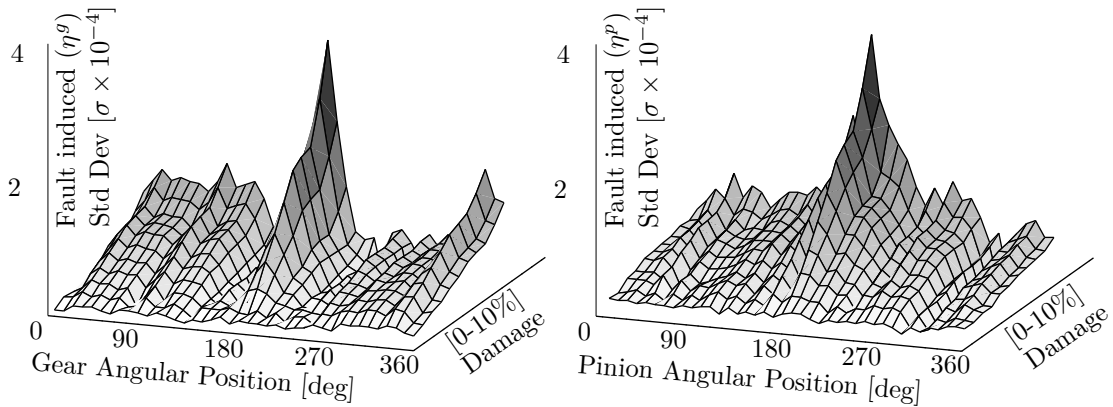


Figure 12: Estimated standard deviation (estimated under fluctuating loads) for different (a) gear and (b) pinion angular positions as trended over the life cycle of pinion

Figures 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.

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 which was performed in the Sasol Laboratory for Structural Mechanics at the University of Pretoria.

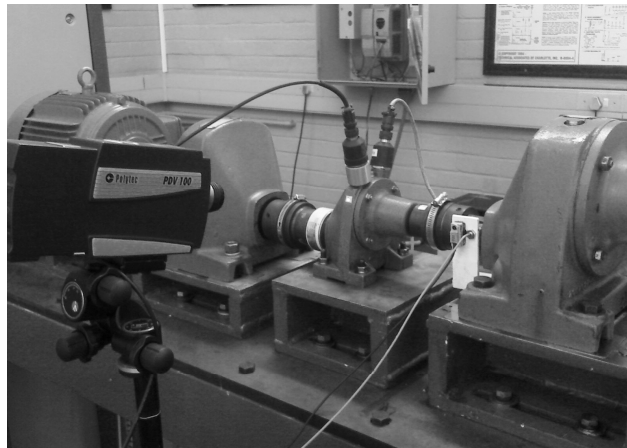


Figure 13: Test setup

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 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 eDAQ-lite data acquisition 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 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 17th 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 14.

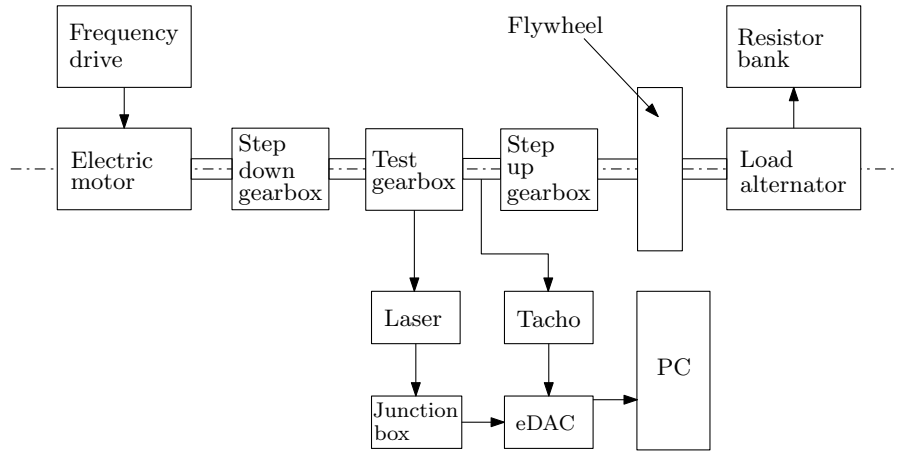


Figure 14: Test setup schematic layout

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 15 (a) and (b) respectively represent the original gearbox response and also the final response at 17 h, just before total failure occurred. Figure 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.

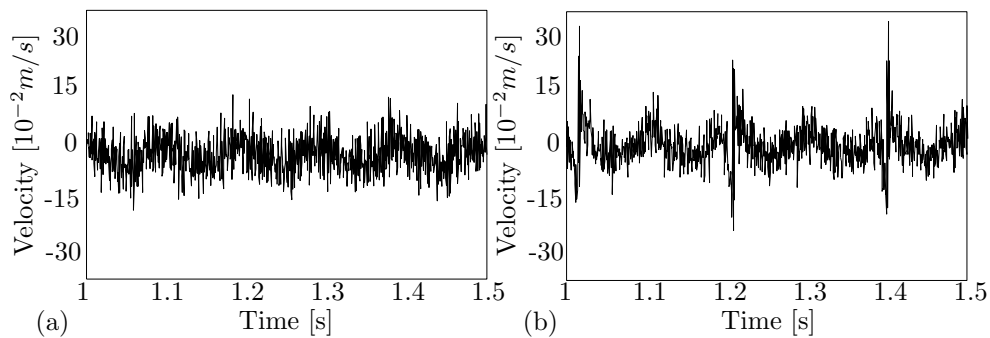


Figure 15: Gear casing velocity of (a) run in condition and (b) severely damaged condition

The power spectral density waterfall plot of the gear life cycle is shown in Figure 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. The 200 Hz component is not well understood, but it might be the result of structural resonance, or it might originate from either the motor or the alternator. 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.

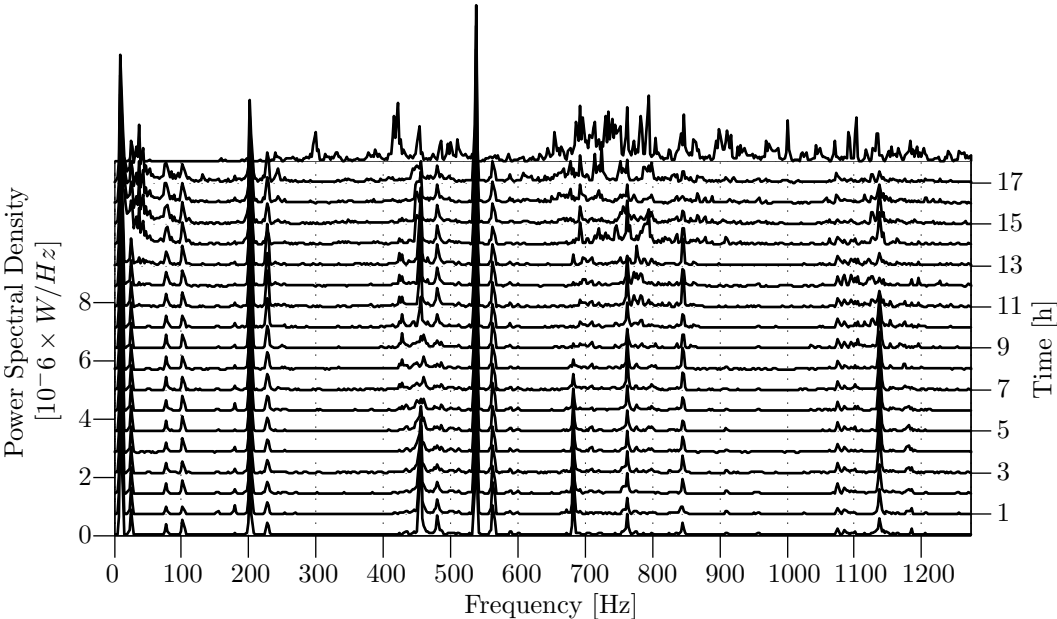


Figure 16: PSD Waterfall plot

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 ($2\text{kHz}/10$) 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.

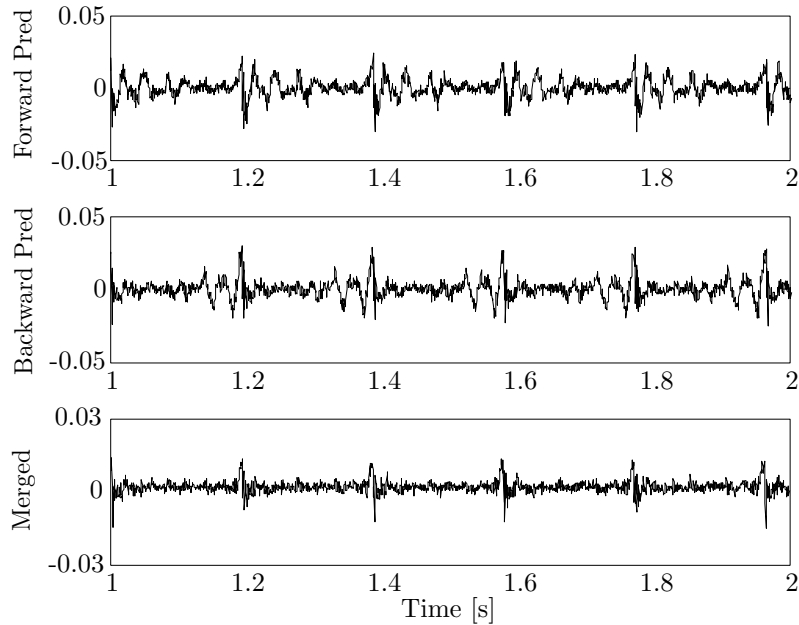


Figure 17: Residuals based on (a) forward and, (b) backward filtering, and (c) is obtained as the pointwise smallest. (c) is smearing free so that the signal irregularity is well localized

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 constant 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 18 (a) and (b) for respectively the healthy and the severely damaged gear system.

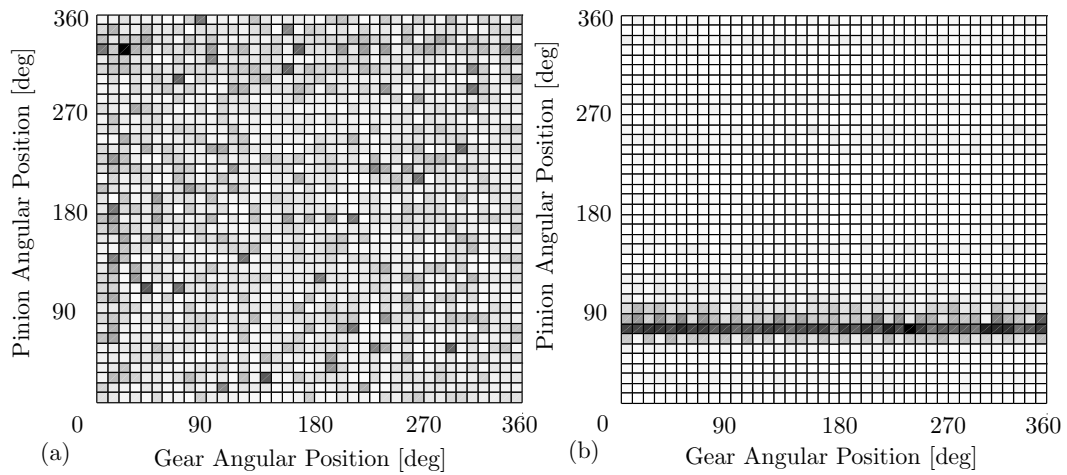


Figure 18: The lumped state variances associated with each of the 40×40 meshing states for (a) no damage (first sample) and (b) shortly before failure occurs (17th sample). A clear irregularity associated with a specific pinion angular position at approximately 80 degrees is observed, indicating the possible presence of a tooth crack.

Figure 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 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.

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 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.

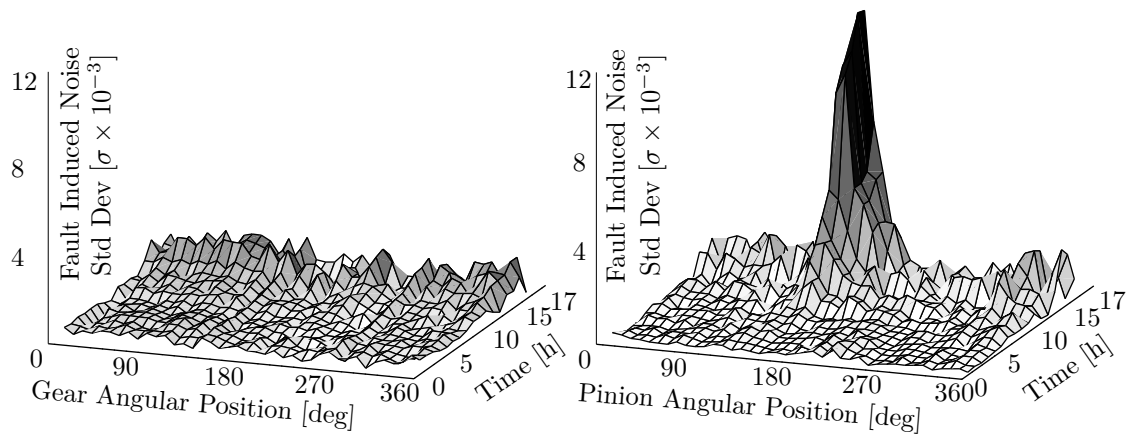


Figure 19: Estimated standard deviation for different (a) gear and (b) pinion angular positions as trended over the life cycle of pinion

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 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.

5. Conclusion

This paper proposed a methodology which can be used to remove the regular components from a vibration signal so that a residual 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 assembly 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 pointwise 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 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.

Future work will investigate the proposed methodology on a range of more representative data, however based on this initial investigation it is believed that the proposed methodology offers an intuitive and cost effective approach which can be used to visualize the condition of a gearbox and serve to support maintenance decisions.

Acknowledgement

The authors gratefully acknowledge Mr Rudi Kroch who conducted the accelerated gear tests at the University of Pretoria.

References

- [1] A.K.S. Jardine, D. Lin, D. Banjevic, A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, Elsevier 20 (7) (2006) 1483-1510.
- [2] C.J. Stander, P.S. Heyns, Instantaneous angular speed monitoring of gearboxes under non-cyclic stationary load conditions, *Mechanical Systems and Signal Processing* 19 (4) (2005) 817-835
- [3] N. Baydar, A. Ball, Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum, *Mechanical Systems and Signal Processing* 14 (6) (2000) 907-921

- [4] W. Bartelmus, Diagnostic information on gearbox condition for mechatronic systems, *Transactions of the Institute of Measurement & Control* 25 (5) (2003) 451
- [5] A. Belsak, J. Flasker, Wavelet analysis for gear crack identification, *Engineering Failure Analysis* 16 (6) (2009) 1983-1990
- [6] G.G.Yen, K.C. Lin, Wavelet packet feature extraction for vibration monitoring, *IEEE Transactions on Industrial Electronics* 47 (3) (2000) 650-667
- [7] J.E. Lopez, R.R. Tenney, J.C. Deckert, Fault detection and identification using real-time wavelet feature extraction, *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis* (1994) 217-220
- [8] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, *Mechanical Systems and Signal Processing* 21 (6) (2007) 2560-2574
- [9] L. Wang, J. Chu, J. Wu, Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process, *International Journal of Production Economics* 107 (1) (2007) 151-163
- [10] R.B. Randall, The application of fault simulation to machine diagnostics and prognostics, *ICSV16, Krakow, Poland* 5 (9) (2009) 81-89
- [11] K.S. Wang, P.S. Heyns, Application of computed order tracking, VoldKalmanfiltering and EMD in rotating machine vibration, *Mechanical Systems and Signal Processing* 25 (2011) 416-430
- [12] R.M. Stewart, Some useful data analysis techniques for gearbox diagnostics, *Proceedings of Meeting on Applications of Time Series Analysis, ISVR, Southampton, UK* (1977) 18.1-18.19
- [13] D. Lin, M. Wiseman, D. Banjevic, A.K.S. Jardine, An approach to signal processing and condition-based maintenance for gearboxes subject to tooth failure, *Mechanical Systems and Signal Processing* 18 (5) (2004) 993-1007
- [14] A.J. Miller, A new wavelet basis for the decomposition of gear motion error signals and its application to gearbox diagnostics, *MSc Thesis, The Graduate School, The Pennsylvania State University* (1999)
- [15] W. Wang, A.K. Wong, Autoregressive model-based gear fault diagnosis, *Journal of Vibration and Acoustics, Transactions of the ASME* 124 (2) (2002) 172-179
- [16] F.K. Choy, V. Polyshchuk, J.J. Zakrajsek, R.F. Handschuh, D.P. Townsend, Analysis of the effects of surface pitting and wear on the vibrations of a gear transmission system, *Tribology International* 29 (1) (1996) 77-83
- [17] C.D. Begg, T. Merdes, C. Byington, K. Maynard, Dynamics modeling for mechanical fault diagnostics and prognostics, *Mechanical System Modeling for Failure Diagnosis and Prognosis, Maintenance and Reliability Conference, Gatlinburg, Tennessee* (1999)
- [18] N. Sawalhi, R.B. Randall, Simulating gear and bearing interactions in the presence of faults Part I. The combined gear bearing dynamic model and the simulation of localised bearing faults, *Mechanical Systems and Signal Processing* 22 (8) (2008) 1924-1951
- [19] S. Jia, I. Howard, J. Wang, The dynamic modeling of multiple pairs of spur gears in mesh including friction and geometrical errors, *International Journal of Rotating Machinery* 9 (6) (2003) 437-442
- [20] F. Chaari, T. Fakhfakh, M. Haddar, Analytical modelling of spur gear tooth crack and influence on gearmesh stiffness, *European Journal of Mechanics and Solids* 28 (3) (2009) 461-468
- [21] S.J. Godsill, P.J.W. Rayner, *Digital audio restoration - a statistical model based approach*, Springer-Verlag London (1998)