A SYNCHRONOUS FILTER FOR VIBRATION MONITORING USING COMPUTATIONAL INTELLIGENCE

L. MDLAZI, C.J. STANDER and **P.S. HEYNS**

Dynamic Systems Group

Department of Mechanical and Aeronautical Engineering

University of Pretoria

Pretoria, 0002, South Africa

E-mail: stephan.heyns@up.ac.za

AND

T. MARWALA

School of Electrical and Information Engineering,

University of the Witwatersrand,

Private Bag 3, Wits, 2050, South Africa

E-mail: t.marwala@ee.wits.ac.za

ABSTRACT

Calculating the time domain average of a gear vibration signal by direct averaging using digital computers requires large amounts of vibration data. This requirement for large amounts of gear vibration data makes it difficult to develop online ~~gearbox condition monitoring systems that utilise~~systems for monitoring gearbox condition, that utilize time domain averaging calculated by direct averaging, to enhance diagnostic capability. This ~~paper~~study presents a novel approach ~~for~~to estimating the time domain average of a gear vibration signal, using less data than would otherwise be used when calculating the time domain average~~is calculated~~ by direct averaging. Artificial neural networks and support vector machines are used ~~to estimate~~for estimating the time domain average of a gear vibration signal. Two models are presented. Model 1 uses a feed-forward network configuration to map the ~~inputs to the output~~input data to the output, using a section of the input data. The ~~inputs are rotation synchronised~~input data comprises rotation-synchronized gear vibration signals and the output is the time domain average signal. Model 2 is ~~utilised~~utilized in two stages. In the first stage, a feed-forward network configuration estimates the instantaneous average of the gear vibration using a small number of gear rotations. The ~~outputs~~output data from the first stage ~~are used as inputs to~~is used as input into the second stage, where a second feed-forward network configuration is used ~~to estimate~~for estimating the time domain average of the entire vibration signal. When Model 1 is used, the results indicate that the amount of gear vibration data~~that is~~ required to calculate the time domain average is reduced to 25 percent of the amount of data~~that is~~ required when calculating the time domain average by direct averaging. When Model 2 is used, the amount of vibration data~~that needs~~ to be stored in the data

acquisition system is reduced to less than 20 percent of the vibration data that would otherwise be stored when calculating the time domain average by direct averaging.

Keywords: *Artificial neural networks, ~~Time~~time domain averaging, ~~Multi-layer~~multilayer perceptron, ~~Radial~~radial basis function, ~~Support~~support vector machines, gear vibration signal <<delete if not necessary>>*

1. INTRODUCTION

Gears and gearbox systems are vital components in many industrial applications. An unexpected failure of a gearbox in a production environment ~~can~~may lead to unplanned maintenance, ~~which decreases~~decreasing the availability of the production ~~plant. Vibration based~~plant <<suggest rephrasing: leading to downtime at the production plant>>. Vibration-based analysis techniques have been widely used ~~to monitor~~for monitoring the condition of gearboxes. The underlying premise of vibration monitoring is that changes in the mechanical condition result in changes in the measured vibration. Vibration spectrum analysis is ~~utilised to analyse~~utilized to analyze structural response ~~measurements since it is difficult to detect~~measurements, as the small changes in the structural response ~~related~~that relate to component deterioration are difficult to detect through conventional time domain analysis. The interaction of various components in a rotating machine may generate excitation forces at frequencies that overlap with the frequencies of the forces generated by the other components ~~with~~in the system. Conventional vibration spectrum analysis does not attenuate the frequencies produced by other ~~system components, which~~components of the system, and in many applications these other frequencies mask the change in the structural response ~~due to~~arising from the deterioration of the monitored component.

~~To overcome this problem in gearboxes, time domain averaging is used. To calculate the time domain average (TDA) using a digital computer, a~~Time domain averaging is

used to overcome this problem in gearboxes. A rotational signal from a sensor on the input shaft or any other suitable location on the gearbox is ~~used.~~used for calculating the time domain average (TDA) using a digital computer. The rotational signal is used either to control the sampling of the total vibration signal or to determine accurately~~determine~~ the rotation period of an individual gear and ~~separate~~single out its vibration component. When the signal is used to control the sampling of the total vibration signal, the vibration and the rotational signals are sampled simultaneously at a ~~fixed clock~~fixed-clock frequency. The period of the monitored gear rotation can be estimated from the rotational signal and the transmission gear ratio. Once the period of the monitored gear rotation has been obtained, the TDA signal of the rotation of the monitored gear is obtained by directly averaging the segments of the total vibration that have been cut out consecutively to correspond ~~to~~with the calculated period [1]. When sufficient averages are taken, all the vibration from the gearbox, which is asynchronous with the vibration of the gear of interest, is attenuated. The resulting ~~time synchronously~~time-synchronously averaged signal indicates the vibration produced during one rotation of the monitored gear. The TDA signal can be related to the meshing stiffness of the monitored gear. Variations in the meshing stiffness indicate wear ~~and or~~and/or incipient local defects~~that are~~ related to a variation in ~~gear teeth~~the stiffness of gear teeth [1, 2, 3].

~~The topic of time domain averaging data has been investigated by many researchers. Different~~Many researchers have investigated the topic of data from time domain averaging, and proposed different models for calculating the TDA~~ have been proposed~~. These models address the diagnostic capability of time domain averaging and the inaccuracies that may occur during the time domain averaging process. The

models do not address the requirement for large amounts of data when the TDA signal is calculated by direct averaging using digital computers. Consequently, this paper presents a methodology ~~that reduces~~for reducing the amount of vibration data~~that is~~ required to calculate the TDA of a gear vibration by means of direct averaging. Experiments were conducted on a gearbox test rig over the entire life of the gear, operating under constant and fluctuating load conditions. ~~The~~Also investigated was the use of artificial neural networks (ANNs) and support vector machines (SVMs) in the development of a synchronous filter for time domain averaging to reduce the amount of gear vibration data required to estimate the TDA~~ was investigated~~.

As early as 1968, Trimble [2] defined the time domain averaging process as a comb filter. Braun [3] in 1975 ~~analysed~~analyzed the extraction of periodically repeating signals from noise coherent averaging. He ~~considered~~regarded the averaging process as a filtering process and conducted most of his analysis in the frequency domain. Subsequently Braun and Seth [4] proposed techniques for the analysis of repetitive mechanism signatures.

McFadden [5] showed that the comb filter model for time domain averaging does not correctly describe the extraction of periodic waveforms from additive noise, because it assumes a knowledge of the signal over an infinite time ~~and~~but the result it produces is not exactly periodic. ~~He then~~Then he presented a revised model which requires a finite number of samples of the signal and ~~produces~~does produce a periodic result. In a later publication McFadden [6] presented a higher-order interpolation technique for the time domain averaging of gear vibration by digital computers.

Hongxing et al. [1] ~~analyse~~analyzed why the direct averaging approach is not ~~sufficiently effective.~~effective enough. These authors observed the influence of the periodic cutting error on the direct averaging results. They suggested an improved algorithm for direct averaging to prevent the influence of periodic cutting error.

Artificial neural networks have been widely applied in the automated detection and diagnosis of machine conditions [7]. ~~Traditional neural network approaches have limitations on generalisation~~The traditional approaches to neural networks allow only limited generalization, giving rise to models that can over fit the training data. This? <<limited generalisation? over-fitting? training data? Meaning not ~~This~~clear>> is caused by the error functions~~that are~~ used in the ~~optimisation~~optimization of artificial neural networks. The error functions seek to ~~minimise~~minimize the difference between the target output and the network output, therefore ~~over fitting~~over-fitting the data. In recent years support vector machines, based on statistical learning theory, have been increasingly ~~theory have found increasing use.~~used. The difference between artificial neural networks and support vector machines is the approach ~~in which~~that each method ~~minimises~~takes to minimizing the risk. In support vector machines, the structural risk ~~minimisation, which minimises an upper bound~~is used for minimizing an upper limit on the expected risk, ~~while~~whereas in neural networks the empirical risk ~~minimisation that minimises~~is used for minimizing the error ~~on~~in the training data~~, is used~~ [7,8,9].

In this ~~work it is envisaged~~study, the aim was to train ANNs and SVMs to estimate the time domain average of a large number of ~~rotation synchronised~~rotation-synchronized gear vibration signals, using less data than would otherwise be ~~used~~

~~when~~needed for calculating the time domain average by direct averaging. The ~~artificial neural networks that are used are multi-layer~~ANNs used are multilayer perceptron (MLP) networks and radial basis function (RBF) networks. In neural networks, the~~reduction of the~~ input space is ~~achieved~~reduced by transforming the input data space into a lower dimensional space or by trimming ~~off~~the redundant features ~~within~~inside the input space. T~~ransformation of t~~he input space ~~to~~is transformed into a lower dimensional space~~isachieved~~ by using procedures such as principal component analysis (PCA) [10]. Engineering judgement and procedures such as automatic relevance determination (ARD) [11, 12, 13] are used ~~to prune~~for pruning the input space. Mdlazi et al. [14] compared the performance of ARD to PCA, focusing on the practical implementation issues of the two input-selection schemes by using practical vibration examples. In the context of this ~~work,~~study, the interest is in the time domain representation of the gear vibration data and it is undesirable to lo~~o~~se any of the underlying dynamics ~~with~~in the input vibration data, which ~~could~~might be the case when the input space is pruned or transformed. ~~This results in~~Hence the requirement ~~of efficiently mapping~~that the input gear vibration signal has to be efficiently mapped to the time domain average of the gear vibration signal.

Two different filter models are considered. Model 1 uses a ~~feed forward~~feed-forward network configuration to map the input space to the target. Model 2 ~~is utilised in two stages. In the first stage, it uses a feed forward~~has two stages. The first stage uses a feed-forward network configuration to estimate the instantaneous time domain average of the gear vibration, after a small number of gear rotations. The output data <<note: as input / output are already plural, English does not have outputs or inputs>>

from the first stage ~~are used as inputs to~~is used as the input data into the second stage, where a second ~~feed forward~~feed-forward network configuration is used ~~to predict~~for predicting the time domain average of the entire vibration signal. The results indicate that the amount of gear vibration data~~that is~~ required to calculate the direct time domain average using Model 1 can be reduced to 25 percent of the amount of data~~that is~~ used when calculating the TDA by direct averaging. The amount of gear vibration data that needs to be stored in the data acquisition system when Model 2 is used, can be reduced to less than 20 percent of the vibration data that would otherwise have to be stored in the data acquisition system when calculating the TDA by direct averaging.

Two parameters were selected to verify ~~that~~whether the TDA estimate from the models would retain the diagnostic capability of the original time domain averaging process. These parameters are the kurtosis for impulses and the peak value for overall vibration. ~~A comparison of~~The computational time is also ~~done~~compared to determine the suitability of the proposed models in ~~real time~~real-time analysis.

2. VIBRATION DATA

The data used in this study was obtained from the accelerated gear life test rig developed by Stander and Heyns [15] for their work on~~condition~~ monitoring the condition of gearboxes under fluctuating load conditions. This experimental set-up consists of three Flender Himmel Motor helical gearboxes, driven by a 5.5 kW three-phase ~~four~~ four-pole WEG squirrel cage electric motor. A 5.5 kVA Mecc alte spa three-phase alternator was used for applying the load. The gear test rig was designed to conduct accelerated tests of gear life~~tests~~ on the Flender E20A gearbox under

varying load conditions. Two additional Flender E60A gearboxes were incorporated into the design in order to increase the torque applied to the smaller Flender E20A gearbox. The rated load of the gears in the Flender E20A gearbox was 20 Nm. The ~~Direct Current~~direct current (DC) fields of the alternator were powered by an external DC supply~~in order~~ to control the load~~that was~~ applied to the gears. Figure 1 illustrates the ~~accelerated life~~accelerated-life gear test rig.

A Hengstler R176T01 1024ED 4A20KF shaft encoder was used, which produced 1024 pulses per revolution. The reference point for time domain averaging ~~is~~was measured as a single pulse from the shaft encoder. Acceleration measurements were taken in the vertical direction with a 10 V/g ~~PCB integrated~~PCB-integrated circuit piezoelectric industrial accelerometer and a Siglab model 20-42 signal ~~analyser~~analyzer [15]. Table 1 lists the specifications for the loading conditions ~~that were used.~~used in this study.

~~(Table~~[Table 1 should be placed in this vicinity of the final ~~document)~~document]

3. INPUT FEATURES

The acceleration signal from the gearbox and the pulse signal from the shaft encoder~~pulse signals~~ were measured from the gearbox and shaft respectively. The acceleration signal was sampled at a frequency of 51.2 kHz to ensure ~~a true data representation.~~true representation of the data. The measured acceleration signal was ~~low pass~~low-pass filtered using an ~~eighth order low~~ eighth-order low-pass Butterworth filter with a cut-off frequency of 300 Hz. This was done to improve the amplitude ~~digitisation~~digitization range for the portion of the signal under

~~consideration,~~consideration: in this case, the fundamental gear mesh frequency at 224 Hz and its sidebands. The filtered acceleration signal was synchronized with the pulse signal from the ~~synchronised with the shaft encoder pulse signal,~~shaft encoder, using consecutive pulses as the start and end of the ~~rotation synchronised~~rotation-synchronized signals. The number of sample points per gear rotation differs because the shaft speed was not ~~stationery~~stationary [16]. The gear rotation signals were ~~resampled~~sampled again at 8192 points so that an equal ~~amount~~number of samples ~~are~~would be obtained for each shaft rotation. There were 165 gear rotations per test. ~~Synchronising~~Synchronizing the measured acceleration signal with the shaft encoder signal and ~~re~~sampling each signal ~~to~~again at 8192 points resulted in an input space of 165 × 8192 per test. For convenience, an input space of 160 × 8192 was selected.

4. ARTIFICIAL NEURAL NETWORKS AND SUPPORT VECTOR MACHINES

Artificial neural networks (ANNs) and support vector machines (SVMs) may be viewed as ~~parameterised~~the parameterized non-linear mapping of input data to the output data. Learning algorithms are viewed as methods for finding the parameter values that look probable in the light of the data. The supervised learning process ~~occurs~~takes place by training the neural networks or support vector machines ~~through supervised learning~~. Supervised learning is ~~the case where~~a process in which the input data set $\boldsymbol{X}$ and the output data set $\boldsymbol{Y}$ are both known, and neural networks or support vector machines are used to approximate the functional mapping between the two data sets. ~~In this paper Multi-layer~~This study considers the multilayer perceptron (MLP) and the radial basis function (RBF) neural networks ~~are considered~~.

4.1. MULTI-LAYER PECEPTRON

A ~~2-layer~~two-layer MLP architecture was ~~used. This selection was made because of the Universal~~selected because the universal approximation theorem [17]~~, which~~ states that a two-layered architecture is adequate for the multi~~-~~layer perceptron. The MLP provides a distributed representation ~~with respect to~~of the input space due to ~~cross coupling~~the cross-coupling between hidden units. The output of a ~~2-layer~~two-layer perceptron can be ~~written as follows:~~expressed as the following equation:

$$y_k = f_{outer}\left(\sum_{j=0}^{M} w_{kj}^{(2)} f_{inner}\left(\sum_{i=0}^{d} w_{kj}^{(1)} x_i\right) + w_{k0}^{(2)}\right). \tag{1}$$

where f_{outer} and f_{inner} are activation functions, $w_{ji}^{(1)}$ denotes a weight in the first layer, going from input i to hidden unit j, $w_{k0}^{(2)}$ is the bias for the hidden unit k and $w_{kj}^{(2)}$ denotes a weight in the second layer. f_{inner} was selected as hyperbolic tangent function "tanh" and f_{outer} was selected as a linear function. The hyperbolic tangent function maps the interval [-∞,∞] onto the interval [-1,1] and the linear activation function maps the interval [-∞,∞] onto the interval [-∞,∞]. The maximum-likelihood approach was used ~~to train~~for training the MLP network. ~~In maximum-likelihood training, optimisation methods are used~~Maximum-likelihood training uses optimization methods to identify a set of parameter values that ~~maximises the~~maximizes ~~ability of the network~~the network's ability to predict the output whenever it is presented with input data, by identifying the weights and biases of Equation (1). A cost function, which is a mathematical representation of the overall objective of the problem, is ~~chosen in order to use the optimisation technique. In this work, the~~selected so that the optimization technique can be used. The overall objective in this study was to identify a set of neural network ~~weights, that map the~~

~~rotation synchronised~~weights which would map the rotation-synchronized gear vibration signals to the TDA of the gear ~~signal using fewer rotation synchronised~~signal, and would use fewer rotation-synchronized gear vibration signals than ~~would otherwise be used to calculate~~calculating the TDA by direct ~~averaging. The sum-of-squares-of-error and the weight decay~~averaging would use. The sum-of-squares-of-error and the weight-decay regularization were used as a cost function [10, 18]. ~~—~~The weight decay is a regularizer ~~and it penalises~~that penalizes large weights and ensures that the mapping function is smooth, avoiding an over-fitted mapping between the ~~inputs and the outputs~~input data and the output data [18]. In this ~~work a regularisation~~study, a regularization coefficient of 1.5 was found~~to be~~ most suitable. ~~Before minimisation of the~~The ~~cost function is performed, the~~ network architecture ~~needs to~~should be constructed by choosing the number of hidden units, *M*, before minimizing the cost function. If ~~*M* is~~the *M* selected is too small, the neural network will not be flexible enough and will ~~give poor generalisation of the data because of~~generalize the data poorly because of a high bias. If ~~*M*~~the *M* selected is too large, the neural network will be too flexible ~~resulting in poor generalisation due to over-fitting caused by high variance~~and will also generalize the data poorly, but in this case because of the over-fitting from a variance that is too high [19]. The weights w_i and biases in the hidden layers ~~are~~were varied using ~~optimisation~~optimization methods until the cost function ~~is minimised.~~was minimized. The Scaled Conjugate Gradient (SCG) method was used because of its observed computational efficiency [17, 20]. It was determined empirically that a ~~2-layer~~two-layer MLP network with ~~5~~five hidden units ~~is most suitable for~~was best suited to this application.

4.2. RADIAL BASIS FUNCTIONS

~~While the~~The MLP network approximates functions by combining hyperbolic tangent activation functions, whereas the RBF network approximates functions by a combination of radial basis functions (and a usually linear output layer) [21]. The RBF neural networks provide a smooth interpolating function for which the number of basis functions is determined by the complexity of the mapping to be represented, rather than by the data set. The RBF neural network mapping is given by

$$y_k(\mathbf{x}) = \sum_{j=1}^{M} \omega_{kj}\phi_j(\mathbf{x}) + \omega_{k0} \tag{2}$$

where ω_{k0} are the biases, ω_{kj} are the output layer weights, $\mathbf{x}$ is the d-dimensional input vector and $\phi_j(\cdot)$ is the jth basis function. ~~In this work the~~The thin plate-spline basis function was used in this study [10]. The~~training of the~~ radial basis function ~~takes place~~is trained in two stages. In the first stage the input data set $\mathbf{x}^n$ alone is used to determine the basis function parameters. After the first training stage, the basis functions are kept fixed and the second layer of weights ~~are found~~is determined in the second training phase. Since the basis functions are considered fixed, the network is equivalent to a ~~single-layer~~single-layer network that can be ~~optimised by minimisation of~~optimized by minimizing a suitable error function. The sum-of-s~~um-of-s~~quare error function is also used to train RBF [18]. The error function is a quadratic function of the weights and its minimum can therefore be found in terms of the solution of a set of linear equations. For regression the basis function parameters are found by treating the basis function ~~centres~~centers and widths, along with the second-~~second-~~layer weights, as adaptive parameters to be determined by ~~the minimisation of~~minimizing an error function. In this ~~work~~study it was determined

empirically that an RBF network with ~~5~~five basis functions ~~is most suitable.~~was the most suitable network.

4.3. SUPPORT VECTOR MACHINES

The foundations of SVMs ~~have been~~were developed by Vapnik [9,18] and have gained popularity in recent years. The SVM formulation embodies the Structural Risk ~~Minimisation~~Minimization (SRM) principle, which has been shown to be superior to the traditional Empirical Risk ~~Minimisation~~Minimization (ERM) principle employed by conventional neural networks [8, 9, 18]. SRM ~~minimises an upper bound~~minimizes an upper limit on the expected risk, as opposed to ~~ERM that minimises~~the ERM that minimizes the error on the training data. It is this difference that ~~equips SVMs with~~gives SVMs a greater ability to ~~generalise.~~generalize. When SVMs are applied to regression problems, loss functions are used that include a distance measure~~ are used~~. Possible loss functions are the quadratic loss function, the Laplacian loss function, the Huber loss function and the ε-insensitive loss function. [9]. The ε-insensitive loss function was selected for this ~~work,~~study, because unlike the quadratic and Huber cost functions~~,~~ where all the data points will be support vectors, the support vector solution can be sparse. The ε-insensitive loss function is defined by

$$L_{\in}(y)=\begin{cases}0 & \text{for} \quad |f(\mathrm{x})-y|<\varepsilon \\ |f(\mathrm{x})-y|-\varepsilon & \text{Otherwise}\end{cases}. \tag{3}$$

In non-linear regression, non-linear mapping is used to map the data to a higher dimensional feature space where linear regression is performed. The kernel approach is employed to address the curse <<curve? course? curse sounds odd in this context>>

of dimensionality. The non-linear support vector regression solution, using the ε-insensitive loss function, is given by

$$\max_{\alpha,\alpha^*} W\left(\alpha,\alpha^*\right) = \max_{\alpha,\alpha^*} \sum_{i=1}^{l}\sum_{j=1}^{l} \alpha_i^*\left(y_i - \in\right) - \alpha_i\left(y_i + \in\right) - \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\left(\alpha_i^* - \alpha_i\right)\left(\alpha_j^* - \alpha_j\right)K\left(x_i, x_j\right) \tag{4}$$

with constraints,

$$0 \le \alpha_i, \alpha_i^* \le C, \qquad i = 1,\ldots,l$$
$$\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0. \tag{5}$$

Solving Equation (3) with the constraints in Equation (5) determines the Lagrange multipliers, α and α^* and the regression function is given by

$$f\left(\mathrm{x}\right) = \sum_{i=1}^{l}\left(\bar{\alpha}_i - \bar{\alpha}_i^*\right)K\left(\mathrm{x}_i, \mathrm{x}\right) + \bar{b} \tag{6}$$

where

$$\langle \bar{w}, x\rangle = \sum_{SVs}\left(\alpha_i - \alpha_i^*\right)K\left(x_i, x_j\right)$$
$$\bar{b} = -\frac{1}{2}\sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(K\left(x_i, x_r\right) + K\left(x_i, x_s\right)\right). \tag{7}$$

Different kernels were investigated for mapping the data to a higher dimensional feature space where linear regression ~~is~~was performed. The exponential radial basis function kernel [10] with an order of 10 was found most suitable for this application.

5. ANN AND SVM MODELS FOR TIME DOMAIN AVERAGING

The main aim of this ~~work~~study was to reduce the amount of vibration data that ~~needs~~would have to be collected and stored in data acquisition equipment before computing the TDA of a gear vibration signal~~ is computed~~. This requires the efficient mapping of the input space to the output space, using less data than would ~~otherwise be used in~~be used when calculating direct time domain averaging using a digital computer. Figure 2 ~~shows~~is a flow diagram of the proposed procedure for time domain averaging~~ procedure~~.

Two different models are proposed. Model 1 maps the input space to the target using simple ~~feed forward~~feed-forward network configuration. The size of the input space is reduced to find the optimal number of input vectors that can be used to~~correctly~~ predict the target vector correctly. When the ANNs and SVMs are properly trained, Model 1 is capable of mapping the input space to the target, using less data than would otherwise be used when calculating the TDA by direct averaging. It was determined empirically that 40 ~~rotation synchronised~~rotation-synchronized gear vibration signals were suitable for predicting the TDA with Model 1.~~Using 40 rotation synchronised gear vibration signals~~ ~~means that~~Consequently the amount of data~~that is~~ required for ~~to calculate~~calculating the TDA is reduced to 25 percent when using 40 rotation-synchronized gear vibration signals, since 160 ~~rotation synchronised~~rotation-synchronized gear vibration signals were used ~~to~~in calculating the TDA by direct averaging. Figure 3 shows a schematic diagram ~~for~~of Model 1.

Model 2 estimates the TDA of the input space in small sequential steps, analogous to taking a running average of the input space. This model consists of a number of ~~feed forward~~feed-forward networks similar to the network~~that is~~ used in Model 1. ~~In~~

~~Model 2,~~By contrast, instead of using the network to estimate the TDA of the ~~whole~~entire input data, ~~it~~Model 2 is ~~used to~~ first used sequentially to estimate the average of subsections of the input data. The output~~s~~ of the first stage ~~are used as inputs to~~is used as input into the second network that estimates the TDA of the entire input data. The ~~feed forward~~feed-forward networks are trained offline to reduce computation time. In Model 2, data can be discarded immediately after ~~it~~use. ~~has been used.~~ This means that the model does not require large amounts of data to be stored ~~within the data logger although all the data needs to~~in the data logger, even though all the data should be collected. In this ~~work, 10 rotation synchronised~~study, 10 rotation-synchronized gear vibration signals were found most suitable for estimating the instantaneous average in the first stage of estimation. ~~This resulted in~~As a result, the amount of vibration data ~~that needs to~~required be stored in the data logger ~~being~~was reduced to less than 20 percent of the amount of data ~~that needs~~required to be stored in the data logger when calculating the TDA by direct averaging. Figure 4 shows a schematic diagram of Model 2.

6. ESTIMATION RESULTS

Model 1 and Model 2 were used ~~to estimate~~for estimating the TDA of the gear vibration signal ~~to make a comparison between~~so that the TDA estimated by the proposed models ~~and~~could be compared with the TDA calculated by direct averaging. The comparison was made in~~both~~ the time and the frequency domains, using data from tests conducted under constant and varying load conditions. ~~To~~For quantifying the accuracy of the TDA estimated by the models, the 'fit' parameter η_{sim} [22] is defined by

$$\eta_{sim} = 100 \frac{\sum_{k=1}^{N} |e(k)|}{\sum_{k=1}^{N} |y_{desired}(k)|} \qquad [\%] \quad , \tag{8}$$

where *e(k)* is the simulation accuracy defined by Equation (9) and *N* is the number of data points~~is~~ used.

$$e(k) = y_{desired}(k) - y_{achieved}(k). \quad . \tag{9}$$

$y_{desired}$ is the TDA signal calculated by direct averaging and $y_{achieved}$ is the TDA signal estimated by the models. The 'fit' parameter η_{sim} is attractive ~~in that~~since it gives a single value for each simulation, therefore it can be used to compare the performance of the different formulations over the entire life of the gear. A high value ~~of~~for the 'fit' parameter η_{sim} implies a bad fit ~~while a small~~whereas a low value implies a good fit. ~~It~~Through trial and error it was established~~by trial and error~~ that $\eta_{sim} = 40\%$ ~~is a suitably~~ is a suitable upper cut-off ~~of the~~point for simulation accuracy.

Two data sets of dimensions 16×160×8192 ~~each,~~each were used. These data sets were taken from 16 tests throughout the life of the gear for constant and varying load conditions respectively~~, were used~~. Each test data set consisted of 160 ~~rotation synchronised~~rotation-synchronized gear vibration signals with 8192 sample points. For Model 1, the first 40 rotation-~~rotation synchronised~~synchronized signals from the first test were used for training the ANNs and SVMs, giving a training set of dimensions ~~1×40×8192, while the~~1×40×8192. The rest of the data of dimensions 15×160×8192 was used as validation data. For Model 2, the whole data set from the

first test ~~were~~was used for training the ANNs and SVMs, giving a training set of dimensions ~~1×160×8192, while the~~1×160×8192. The rest of the data of dimensions 15×160×8192 was used as validation data. To improve ~~on~~the computational efficiency, the test and validation data sets were resampled to 1024 points per gear rotation for estimation with ANNs, resulting in data sets of 1×40×1024 and (15×40×1024) respectively. For estimation with SVMs the test and validation data sets were resampled to 256 points per gear rotation, resulting in data sets of 1×40×256 and (15×40×256) respectively.

6.1. ESTIMATION WITH MODEL 1

Model 1 with MLP, RBF and SVMs was trained with 40 rotations of the ~~synchronised~~synchronized gear vibration signals that ~~were~~had been used to estimate the time domain average, which can be obtained after 160 gear revolutions. The data was obtained from the test conducted under constant and varying load conditions. Figure 5 shows the estimation results obtained when Model 1 with an MLP network was simulated using unseen validation sets of 40 input signals during the ~~running in~~running-in stages of gear life. The dotted line is the TDA estimated by Model 1 with an MLP network, and the solid line is the TDA calculated by direct averaging. The first plot in Figure 5 is the time domain representation of the ~~results, while~~results; the second plot is the frequency domain representation. ~~It is observed that~~As can be seen, both the time and frequency domain representations of the plots in Figure 5 are almost exact fits. This shows that Model 1 with MLP networks retains the time and frequency domain properties of the original time domain averaging process when using data from the tests conducted under constant loading conditions~~ is used~~. Similar performances were obtained throughout the life of the gear, even though there were

significant changes in the vibration signatures as the condition of the monitored gear deteriorated. The changes in the vibration signatures were ~~caused by~~due to the changes in the meshing stiffness caused by cracks in the gear teeth. The good performance of Model 1 under constant load~~ing~~ conditions is due to the mapping and ~~generalisation~~generalization capabilities of ANNs and SVMs.

Figure 6 shows the estimation results obtained for Model 1 with an MLP network trained and simulated with test data conducted under varying load conditions. It ~~is observed that the obtained~~can be seen that these results are not as accurate as the results obtained using data from tests conducted under constant load~~ing~~ conditions. The poorer performance of Model 1 under varying load~~ing~~ conditions can be ~~contributed~~attributed to the fact that varying load conditions introduce amplitude modulation in the measured gear vibration signal. In this case, the load is not synchronous with the rotation of the shaft and consequently the amplitude modulation is averaged out if a sufficient number of ~~synchronised~~synchronized gear vibration signals are used during the TDA calculation. The effect of varying load conditions therefore increases the degree of non-linearity between the input~~s~~ and the output of the ANN.

Similar results were obtained when the TDA was estimated using Model 1 with RBF networks and SVMs, under both constant and varying load~~ing~~ conditions. ~~This is due~~These similar results can be ascribed to the fact that MLP networks, RBF networks and SVMs ~~generalised~~generalized well to the data~~that was~~ used.

Figures 7 shows the simulation accuracy η_{sim} plotted against the gear life for Model 1 for tests conducted under constant load~~ing~~ conditions. It ~~is observed that the performance~~can be seen that Model 1 with RBF network and Model 1 with SVMs ~~is~~performed similarly under the same~~under~~ constant load~~ing~~ conditions. Their performance ~~is~~was slightly better than the performance of Model 1 with MLP networks. The performances of all three formulations are acceptable because η_{sim} is less than the cut-off value for all the formulations.

Figure 8 shows the simulation accuracy η_{sim} plotted against the gear life for Model 1 for tests conducted under varying load~~ing~~ conditions. Measurements were taken at three different stages of the gear ~~life, the running in stage, the intermediate wear stage and the wear out~~life: the running-in stage, the intermediate-wear stage and the wear-out stage.

It ~~is observed~~can be seen that for Model 1, SVMs perform best for varying load conditions and MLP ~~gives the worst performance. This is because of~~performs worst. The reason for the variation in performance is the structural risk ~~minimisation~~minimization used in SVMs, which is said to ~~generalise~~generalize better than the empirical risk ~~minimisation~~minimization used in ANNs [8, 9].

When Model 1 is used, ~~a data reduction~~the amount of data is reduced to 25 percent of the ~~rotation synchronised~~rotation-synchronized gear vibration signals that would otherwise be used to calculate the TDA by direct ~~averaging is achieved. It is further observed~~averaging. It can also be seen that both the time domain and frequency domain simulation results are very close fits to the TDA calculated by conventional

direct averaging using digital computers throughout gear life. This performance is attributed to the good ~~generalisation~~generalization properties of ANNs and SVMs [8; 10].

6.2 ESTIMATION WITH MODEL 2

Figure 9 shows the estimation from Model 2 with MLP ~~feed-forward~~feed-forward networks superimposed on the TDA of the gear vibration signals, calculated using direct averaging. This plot indicates that in both the frequency domain and the time domain representation, Model 2 with ~~a MLP feed-forward~~an MLP feed-forward network can correctly estimate the TDA for 160 gear rotations over the entire life of the gear. ~~This is because~~The reason is that in Model 2 the network is exposed to the entire data set, therefore it trains and simulates more effectively.

Figure 10 shows the results obtained when Model 2 with an MLP network is simulated with unseen gear rotations for tests conducted under varying load conditions. This plot shows that Model 2 with ~~a MLP feed-forward networks~~an MLP feed-forward network can correctly estimate the TDA for 160 gear rotations over the entire life of the gear. Similar estimation results were obtained for Model 2 with RBF networks and SVMs.

The simulation accuracy η_{sim} , was also assessed for Model 2. Figure~~s~~ 11 shows the simulation accuracy η_{sim} plotted against the gear life for Model 2 for tests conducted under constant load~~ing~~ conditions. It ~~is observed that the performance~~can be seen that Model 2 with RBF network ~~is the same~~performs similarly under constant load~~ing~~ conditions. The performances of all three formulations are acceptable since η_{sim} is less

than the selected cut-off value of η_{sim} = 40 percent. ~~This performance is because~~These performances are acceptable because, when simulating with Model 2, the ANNs and SVMs are exposed to the entire data set, therefore~~the~~ Model 2 simulates more efficiently ~~as opposed to~~than Model 1~~in~~ which sees only a section of the data.

Figures 12 shows the simulation accuracy η_{sim} plotted against gear life for Model 2 for tests conducted under varying load conditions. Measurements were taken at three different stages of the gear life, the ~~running-in stage, the intermediate-wear stage and the wear-out~~running-in stage, the intermediate-wear stage and the wear-out stage.

~~For Model 2 the~~The performance of the formulation is ~~the same.~~similar for Model 2. The good performance of Model 2 under varying loads ~~is~~can be attributed to the fact that Model 2 uses the whole data set during ~~simulation as opposed to Model 1 that~~simulation, whereas Model 1 uses only a section of the data set. ~~This is due to the fact that~~The reason is although Model 2 only uses small sections of the ~~inputs~~input data at a time, it still uses the whole data set for estimation ~~as opposed to~~whereas Model 1~~that~~ uses only ~~a section~~part of the data set for simulation. This allows Model 2 to train and simulate more efficiently since it is exposed to all the underlying dynamics ~~with~~in the data set.

7. ASSESSING THE DIAGNOSTIC CAPABILITY

In addition to simply considering the goodness of fit of the data, the diagnostic capabilities of the TDA estimated by the models was assessed using two parameters. The first parameter is the peak value of the vibration X_{max} during a given interval T. This parameter is used ~~to monitor~~for monitoring the overall magnitude of the

vibration to distinguish between acceptable and unsatisfactory vibration states. The second parameter is the kurtosis, ~~the kurtosis is~~namely the fourth statistical moment of the vibration signal. The kurtosis of a signal is useful for detecting the presence of an impulse ~~within~~inside the signal [23]. The peak value of the vibration X_{max} and the kurtosis of the TDA calculated using direct ~~averaging is~~averaging, were compared to the TDA from Model 1 and Model 2 with all three formulations.

7.1 ASSESSING DIAGNOSTIC CAPABILITY UNDER CONSTANT LOADS

Figure 13 and Figure 14 are the plots of X_{max} and kurtosis calculated from the TDA estimated by the developed models, superimposed on the X_{max} and kurtosis calculated from the TDA ~~calculated~~obtained by using direct averaging for constant load~~ing~~ conditions throughout gear life for the three formulations.

~~From~~ Figure 13 and Figure 14 ~~it is observed that for all three formulations~~indicate that the kurtosis is an exact fit for all three formulations. This implies that the TDA predicted by Model 1 can be used to monitor the presence of impulses in the measured gear vibration signal. It ~~is also observed~~can also be seen that the kurtosis is very high during the early stages of gear life. This is characteristic of the ~~running in~~running-in stages of the gear ~~life in~~life, during which the vibration signature tends to be random ~~in nature~~. A similar trend is observed during the ~~wear-out~~wear-out stage in which strong impulses are caused by the reduction in stiffness in a cracked gear tooth or broken teeth. However, only the peak values obtained from Model 1 with MLP and SVM are close fits and can be used to monitor the amplitude of the overall vibration. ~~The~~Model 1 with RBF achieved an unsatisfactory performance ~~of the Model 1 with RBF is due to the fact that~~because the RBF network selected in this simulation was

not ~~optimal,~~optimal; consequently it did not ~~generalise~~generalize well to changes in the measured vibration as gear failure progressed.

~~In~~ Figure 14 ~~it is observed~~indicates that the peak values obtained using Model 2 are close to those calculated using direct averaging for all three formulations. This implies that Model 2 can be used ~~to monitor~~for monitoring the amplitude of the overall vibration. ~~The~~Model 2 has a superior performance~~of Model 2~~ on the peak values ~~is attributed to the fact that~~because Model 2 uses the entire gear vibration during ~~estimation,~~estimation; therefore the network is exposed to all the underlying dynamics ~~with~~in the measured vibration.

7.2 ASSESSING DIAGNOSTIC CAPABILITY UNDER VARYING LOADS

The X_{max} and the kurtosis values were also calculated from the estimation results obtained ~~using~~from the test data measured under random load~~ing~~ conditions. Figure 15 and Figure 16 show X_{max} and the kurtosis values calculated from the TDA estimated by Model 1 and Model 2 respectively for MLP, RBF and SVMs.

The X_{max} and kurtosis are only plotted for the ~~running in, intermediate wear and wear out stages of the gear life. In Figure 15 it is observed~~running-in, intermediate-wear and wear-out stages of gear life. It can be seen from Figure 15 that the kurtosis is not a good fit. This implies that Model 1 cannot be used ~~to monitor~~for monitoring the presence of impulses in the gear vibration under varying load conditions. The peak values ~~for~~ obtained by using Model 1 fit well. This implies that the peak values of the TDA obtained using Model 1 can be used ~~to monitor~~for monitoring the overall vibration of the gear signal.

Figure 16 shows the results obtained using Model 2 for MLP, RBF and SVMs during the ~~running in, intermediate wear and wear out~~running-in, intermediate-wear and wear-out stages of gear life. It ~~is observed~~can be seen that for Model 2 the kurtosis and the peak values ~~are not far off~~do not differ widely from those obtained from the TDA calculated by direct averaging. This implies that the TDA estimated by Model 2 can be used ~~to monitor~~for monitoring impulses and~~the~~ overall gear vibration. Table 2 and Table 3 ~~present a summary of~~summarize the properties of Model 1 and Model 2 for the three formulations.

(Table 2 about here)

(Table 3 about here)

8. ESTIMATION TIME

To put the proposed methods into perspective, ~~a comparison of~~the computation times ~~between~~were compared for the existing time domain averaging method and the proposed ~~models was done. A~~models, using a Pentium 4 computer with a 1.60 GHz processor~~ was used~~. The computation times are ~~presented~~listed in Table 4.

(Table 4 about here)

It is ~~observed that the required pre-processing time for Model 1 is less than the required pre-processing time for~~clear that Model 1 requires less pre-processing time than does the TDA calculated by direct averaging. ~~This is because~~The reason is that Model 1 uses 25 percent of the vibration data ~~as opposed to~~but the original TDA ~~process, which uses the all~~process uses all of the vibration data. The required pre-processing time for Model 2 is equal to the pre-processing time for the TDA calculated by direct ~~averaging. This is~~averaging, because both models use the same

amount of vibration data. When Model 1 and Model 2 are used, RBF and MLP give the best performance in terms of simulating time and SVMs give~~s~~ the poorest performance. The models are trained off-line therefore the training time does not influence the simulating time in ~~a real time~~real-time applications. When the models are used with MLP and RBF networks, they perform much better in terms of simulation time, than the original TDA calculated by direct ~~averaging in terms of simulation time. It is however, observed that~~averaging. However, when the models are used with ~~SVMs, the~~SVMs their performance is very poor. The poor performance ~~in~~with SVMs is due to the training ~~problem, which~~problem. This is a quadratic ~~optimisation~~optimization problem with 2N variables, where N is the number of data training points. ~~Therefore it requires longer~~Longer training times ~~of~~are needed because more data is used in the training process. This is much slower than the MLP and RBF neural networks in which only the weights and biases or the basis ~~centres~~centers are obtained by ~~minimising~~minimizing the error functions.

9. CONCLUSION

In this paper a new approach ~~using~~is presented to the use of artificial neural networks and support vector machines to reduce the ~~data that is~~amount of data required to calculate the time domain average of a gear vibration signal~~ is presented~~. Two models are proposed. Model 1 uses ~~feed forward~~feed-forward network architecture to map input space to the target. ~~Using Model 1 a~~A data reduction ~~to~~of 25 percent was achieved by using Model 1. Model 2 operates in two steps. ~~In the~~The first step~~it~~ uses a small section of the input data to predict the instantaneous time domain average of the gear vibration. Then the input data can~~then~~ be deleted from the memory immediately after use. The output of the first step is used as input ~~to a second feed forward~~into a

second feed-forward network to predict the time domain average of the entire vibration. Model 2 was found to be ~~very effective~~excellent at predicting the time domain average ~~with~~but had the disadvantage of requiring more operations to execute. ~~Using Model 2 a reduction to~~When using Model 2, less than 20 percent of the data ~~that needs~~needed to be stored ~~with~~in the data logger at any ~~instance was achieved.~~given time. Furthermore, the suitability of the developed models for diagnostic purposes ~~is~~was assessed. It ~~is observed~~was seen that for measurements taken from tests carried out under constant load conditions, the performances of the Model 1 and Model 2 ~~are~~were similar over the entire life of the gear. For the test data from tests carried out under varying load ~~conditions Model 2 performs~~conditions, Model 2 performed better than Model 1 over the entire life of the gear. The superior performance of Model 2 ~~is because~~can be attributed to the fact that Model 2 uses the whole data set for training and ~~simulation as opposed to Model 1, which~~simulation, whereas Model 1 uses only a section of the data set. Using the whole data set during training and simulation exposes the formulations in the model to all the transient effects ~~within the data~~in the data, resulting in a more accurate estimate of the time domain average. The performance of Model 1 ~~strongly depends on the generalisation~~relies heavily on the generalization capabilities of the formulation ~~that is used. It is also observed~~used. It can also be seen that the proposed models ~~in~~generally ~~general, perform poorer~~perform more poorly under varying load conditions.

REFERENCES

[1] L. Hongxing, Z. Hongfu, J. Chengyu and Q. Liangheng, 2000 *Mechan~~n~~ical Systems and Signal Processing* 14(2), pp. 279-285, An improved algorithm for direct time domain averaging.

[2] C.R. Trimble, 1968 *Hewlett-Packard Journal* 19(8), pp. 2-7. What is signal averaging?

[3] S. Braun, 1975 *Acustica* (32), pp. 69-77, The extraction of periodic waveforms by time domain averaging.

[4] S. Braun and B. Seth, 1980 *Journal of Sound and Vibration* 70(4), pp. 513-526 Analysis of repetitive mechanism signatures.

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

[6] P. D. McFadden, 1989 *Mechanical Systems and Signal Processing* (3), pp. 87-97, Interpolation techniques for time domain averaging of gear vibration.

[7] B. Samanta, 2004 *Mechanical Systems and Signal Processing* (18), pp. 625-644, Gear fault detection using artificial neural networks and support vector machines.

[8] S.R. Gunn, 1998 *Technical report*, University of Southampton, Department of ~~electrical and computer science,~~Electrical and Computer Science, UK, Support vector machines for classification and regression.

[9] V.N. Vapnik, 1999 *IEEE Transactions on ~~neural networks~~* ~~10, pp~~*Neural Networks* 10, pp. 988-1000, An overview of learning theory.

[10] C.M. Bishop, 1995 *Neural ~~Networks for Pattern Recognition~~*~~, Clarendon Press, Oxford.~~*networks for pattern recognition*, Oxford: Clarendon Press.

[11] D. J. C. MacKay, 1994 *ASHRAE Transactions* 100(2), pp. 1053-1062, Bayesian non-linear model~~l~~ing for the energy prediction competition.

[12] R. M. Neal, 1996 *Bayesian* ~~*Learning for Neural Networks.*~~*learning for neural networks.* New York, NY: ~~Springer-Verlag.~~Springer-Verlag.

[13] R. M Neal, 1998 *Neural* ~~*Networks and Machine Learning.*~~*networks and machine learning.* New York, NY: Springer-Verlag. Assessing relevance determination methods using DELVE.

[14] L. Mdlazi, T. Marwala, C. Stander, C. Scheffer and P.S. Heyns, 2003 *Proceedings of the 21st International Modal Analysis Conference*, San Antonio, pp. 37-42. Principal component analysis and automatic relevance determination for damage identification in structures.

[15] C.J. Stander and P.S Heyns, 2002 Proceedings of the 15th International Congress on Condition Monitoring and Diagnostic Engineering Management,~~,~~ Birmingham UK, 2-4 September 2002, pp. 220-230. Instantaneous shaft speed monitoring of gearboxes under fluctuating load conditions.

~~[17]~~[16] C.J. Stander and P.S Heyns, 2002 *Mechanical Systems and Signal Processing* 16(6) pp. 1005-1024, Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions.

~~[18]~~[17] S. Haykin,~~,~~ 1999 *Neural networks*, 2nd edition, New Jersey, USA: Prentice-Hall Inc~~, New Jersey, USA~~.

[18] V.N. Vapnik,~~,~~ 1995 *The nature of statistical learning theory*,~~Springer-Verlag,~~ New York, USA: Springer-Verlag.

[19] S. Geman, E. Bienenstock, and R. Doursat, 1992 *Neural Computation* 4, pp. 1-58, Neural networks and the bias/variance ~~dilemma,~~dilemma.

[20] T. Marwala, 2001 *Fault identification using neural networks and vibration data*, Ph.D. ~~Thesis,~~thesis, University of Cambridge, UK.

[21] M. J. D. Powell, 1987 Radial basis functions for multivariable interpolation: a review. In. Mason, J. C. and Cox M. G., *Algorithms for approximation*, pp. 143-167. Oxford: Clarendon Press.

[22] A.D. Raath, 1992 *Structural dynamic response reconstruction in the time domain*, ~~PhD~~Ph.D. thesis, Department of Mechanical and Aeronautical Engineering, University of Pretoria.

[23] M. P. Norton, 1989 *Fundamentals of noise and vibration analysis for engineers*, New York: Cambridge University Press.

TABLE 1

Load specification

Load case	Load function	Frequency	Minimum load	Maximum load
1	Constant	0 Hz	10.7 Nm	10.7 Nm
2	Random	2 - 5.12 Hz	7.4 Nm	14.7 Nm

TABLE 2

Summary of properties for Model 1 with MLP, RBF and SVMs

	Strength	Weaknesses	Ideal application
~~**MLP**~~	~~Good generalisation under constant load conditions~~ ~~Good generalisation under varying load conditions~~	~~Depends on training and generalisation of selected network~~	~~Monitoring of overall vibration and impulses in a gear vibration under constant load conditions~~ ~~Monitoring of peak values under varying load conditions~~
MLP	Good generalization under constant load conditions Good generalization under varying load conditions	Depends on training and generalization of selected network	Monitoring of overall vibration and impulses in a gear vibration under constant load conditions Monitoring of peak values under varying load conditions
~~**RBF**~~	~~Good generalisation under constant load conditions~~	~~Poor generalisation under varying load conditions~~ ~~Depends on generalisation of selected network~~	~~Monitoring of overall vibration and impulses constant loading conditions~~ ~~Monitoring of peak values under varying load conditions~~
RBF	Good generalization under constant load conditions	Poor generalization under varying load conditions Depends on generalization of selected network	Monitoring of overall vibration and impulses constant loading conditions Monitoring of peak values under varying load conditions
~~**SVM**~~	~~Good generalisation under constant load conditions~~ ~~Good generalisation under varying load conditions~~	~~Depends on generalisation of SVM~~	~~Monitoring of overall vibration and impulses under constant loading conditions~~ ~~Monitoring of peak values under varying load conditions~~

SVM	Good generalization under constant load conditions Good generalization under varying load conditions	Depends on generalization of SVM	Monitoring of overall vibration and impulses under constant loading conditions Monitoring of peak values under varying load conditions

TABLE 3

Summary of properties for Model 2 with MLP, RBF and SVMs

	Strength	**Weaknesses**	**Ideal application**
~~**MLP**~~	~~Good generalisation under constant load conditions~~ ~~Good generalisation under varying load conditions~~	~~Looses diagnostic capability for overall vibration under varying loads~~ ~~Depends on generalisation of selected network~~	~~Monitoring of overall vibration and impulses in under both constant and varying load conditions~~
MLP	Good generalization under constant load conditions Good generalization under varying load conditions	Looses diagnostic capability for overall vibration under varying loads Depends on generalization of	Monitoring of overall vibration and impulses under both constant and varying load conditions

		selected network		
~~**RBF**~~	~~Good generalisation under constant load conditions~~ ~~Good generalisation under varying load conditions~~	~~Poor generalisation under varying load conditions~~ ~~Looses diagnostic capability for overall vibration under varying loads~~ ~~Depends on generalisation of selected network~~	~~Monitoring of overall vibration and impulses in under both constant and varying load conditions~~	
RBF	Good generalization under constant load conditions Good generalization under varying load conditions	Poor generalization under varying load conditions Looses diagnostic capability for overall vibration under varying loads Depends on generalization of selected network	Monitoring of overall vibration and impulses under both constant and varying load conditions	
~~**SVM**~~	~~Good generalisation under constant load conditions~~ ~~Good generalisation under varying load conditions~~	~~Depends on generalisation of SVM~~	~~Monitoring of overall vibration and impulses in under both constant and varying load conditions~~	
SVM	Good generalization under constant load conditions Good generalization under varying load conditions	Depends on generalization of SVM	Monitoring of overall vibration and impulses under both constant and varying load conditions	

TABLE 4

Computation time

	TDA	**MLP**	**RBF**	**SVM**
Model 1 Pre-processing time [s]	1.011	0.703	0.703	0.703
Model 1 Training time [s]	-	22.24	2.219	497.0
~~Model 1 simulating time [s]~~	~~0.75~~	~~0.016~~	~~0.047~~	~~5.500~~
Model 1 Simulating time [s]	0.75	0.016	0.047	5.500
Model 2 Pre-processing time [s]	1.011	1.011	1.011	1.011
Model 2 Training time [s]	-	1.14	1.015	963.8
~~Model 2 simulating time[s]~~	~~-~~	~~0.08~~	~~0.078~~	~~83.76~~
Model 2 Simulating time[s]	-	0.08	0.078	83.76

<<INSERT FIG 1>>

Figure 1 Accelerated gear life test rig

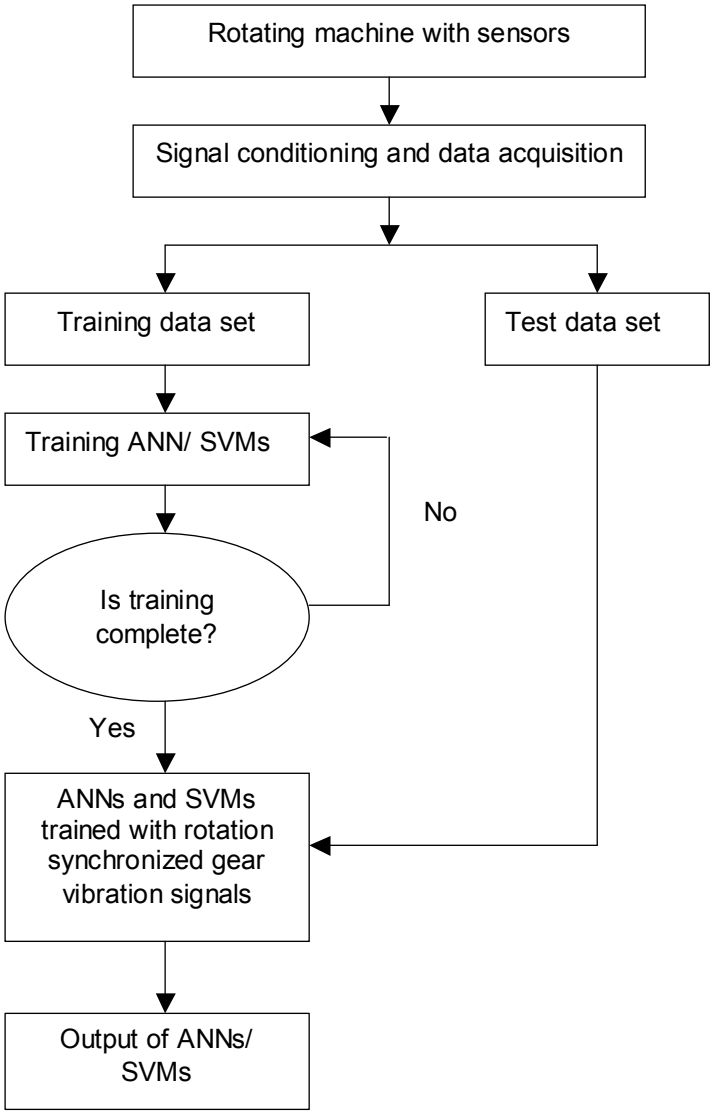

Figure 2 Flow chart of proposed time domain procedure

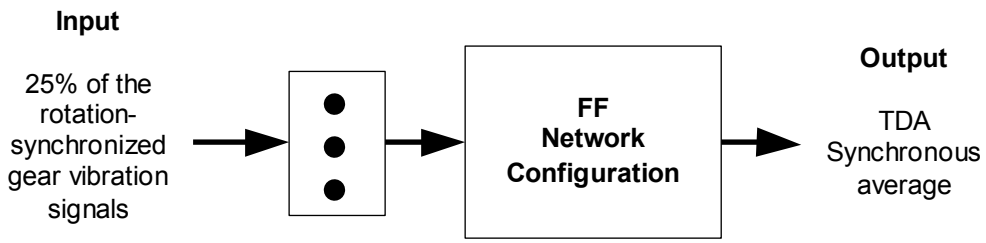

Figure 3 Schematic diagram of Model 1

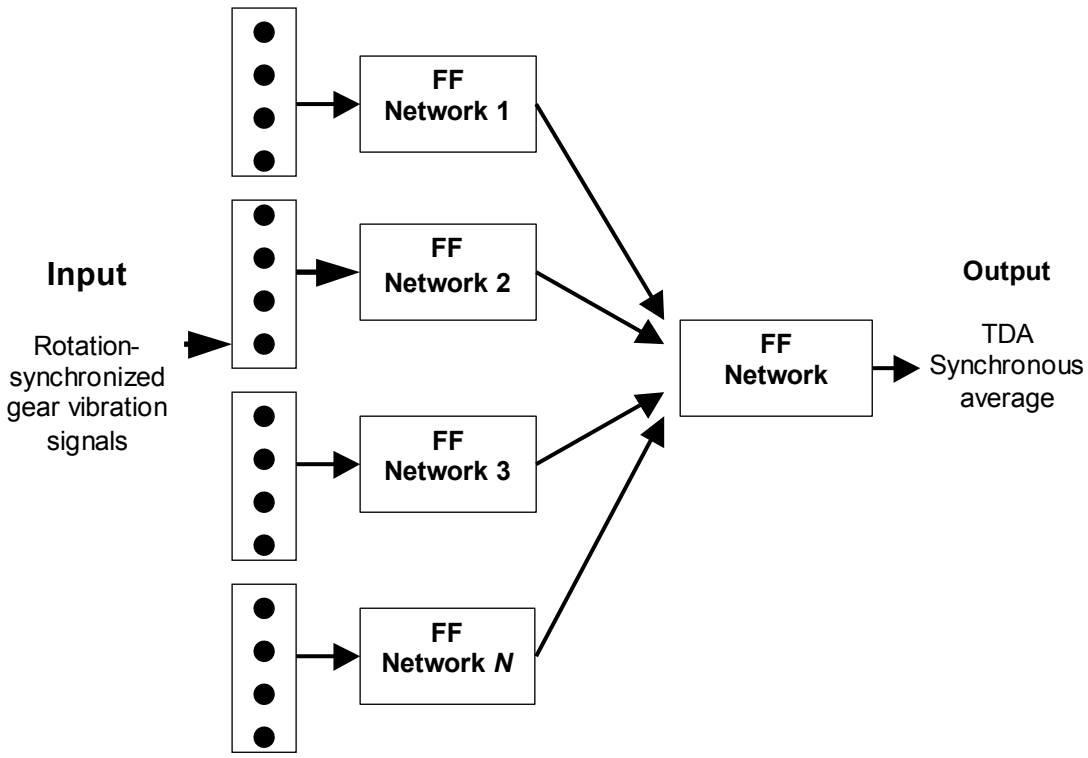

Figure 4 Schematic diagram of Model 2

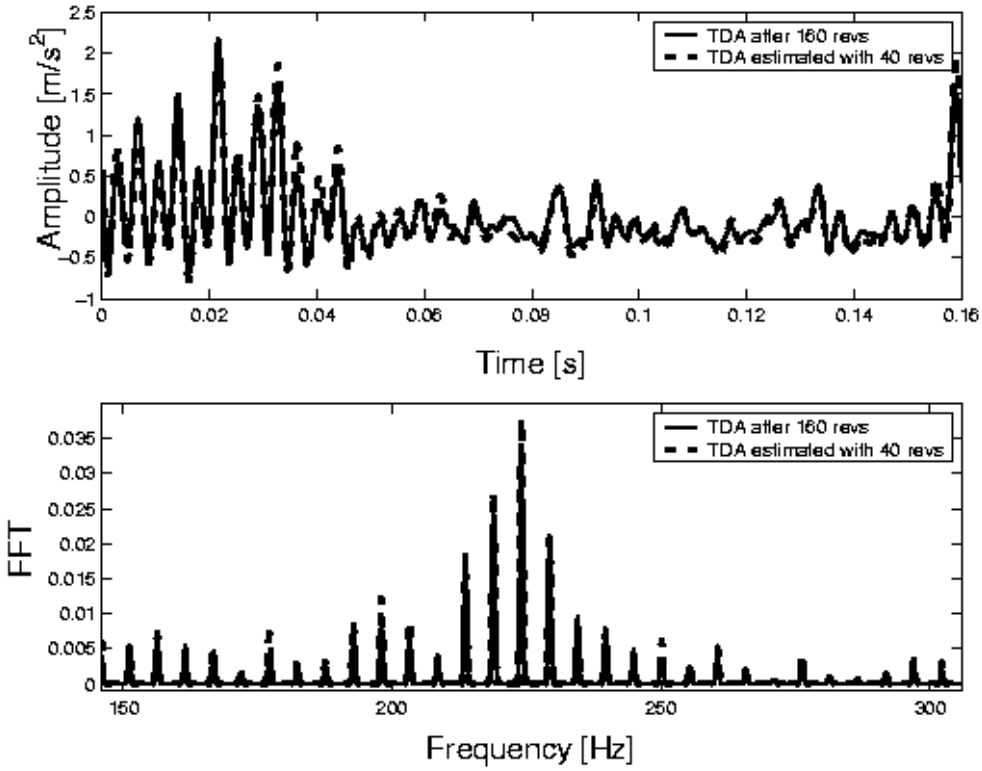

Figure 5 Model 1 with MLP estimation using data from a test conducted under constant load~~ing~~ conditions

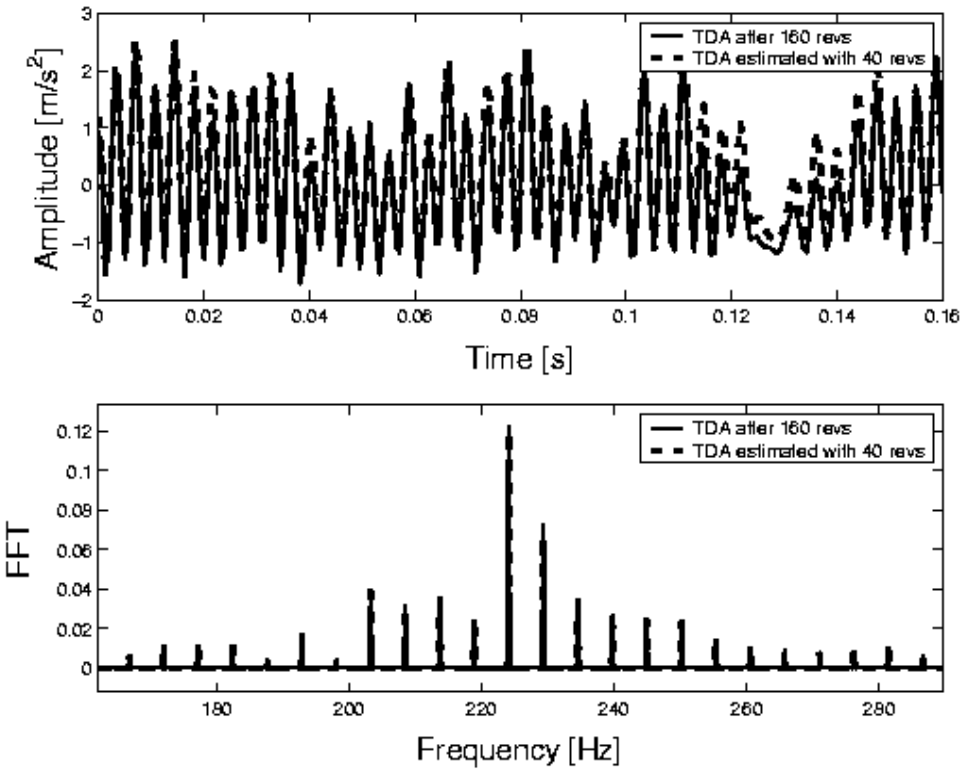

Figure 6 Model 1 with MLP estimation using data from a test conducted under varying load~~ing~~ conditions.

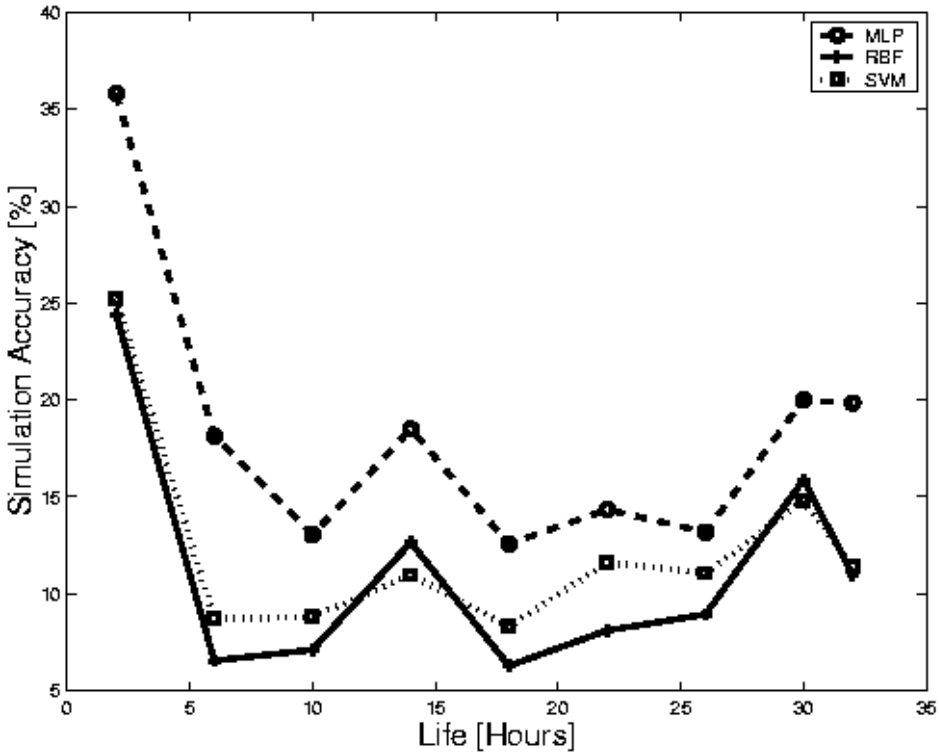

Figure 7 Model 1 ~~simulation~~Simulation accuracy η_{sim} vs. gear life under constant load~~ing~~ conditions

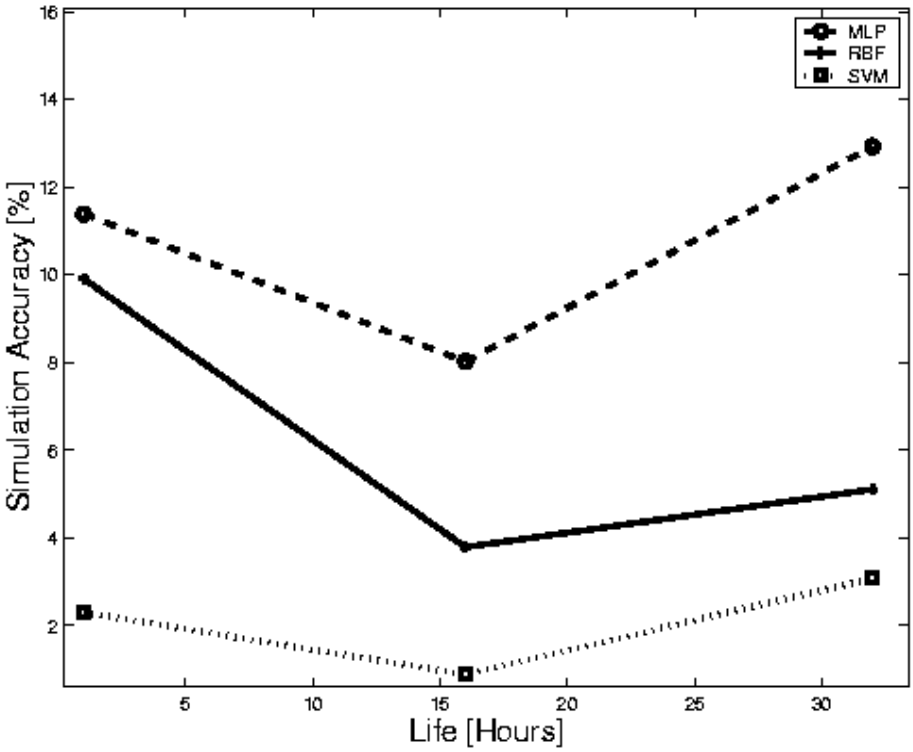

Figure 8 Model 1 ~~simulation~~Simulation accuracy η_{sim} vs. gear life under varying load~~ing~~ conditions

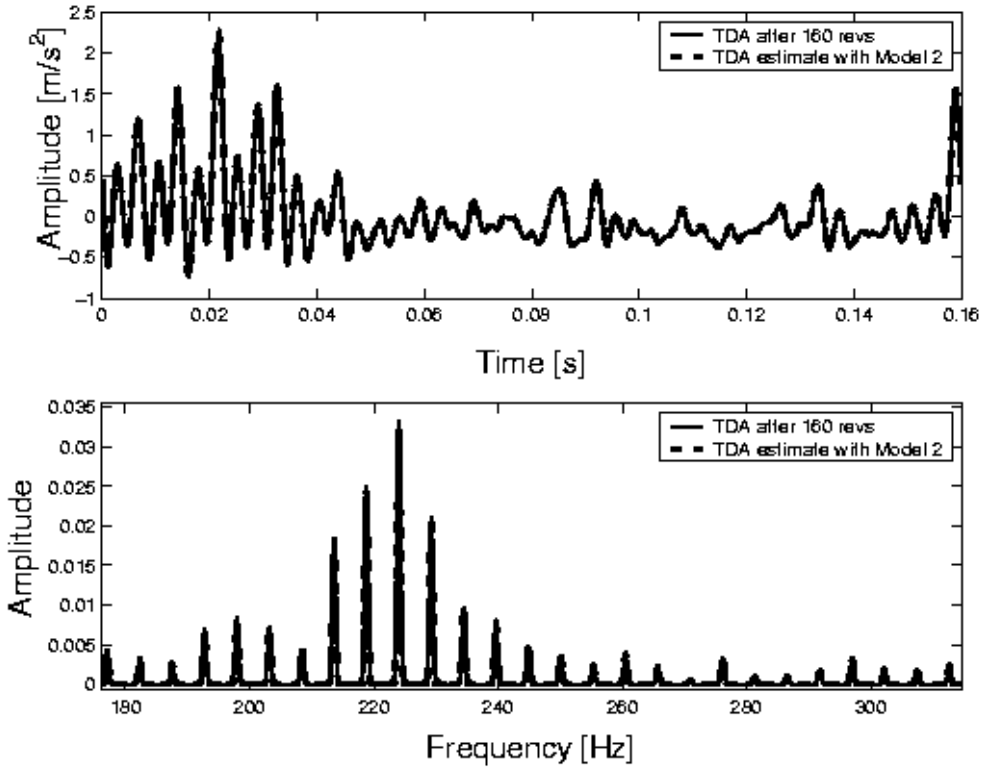

Figure 9 Model 2 ~~estimation~~Estimation results for simulation with a validation set measured from a test conducted under constant load~~ing~~ conditions

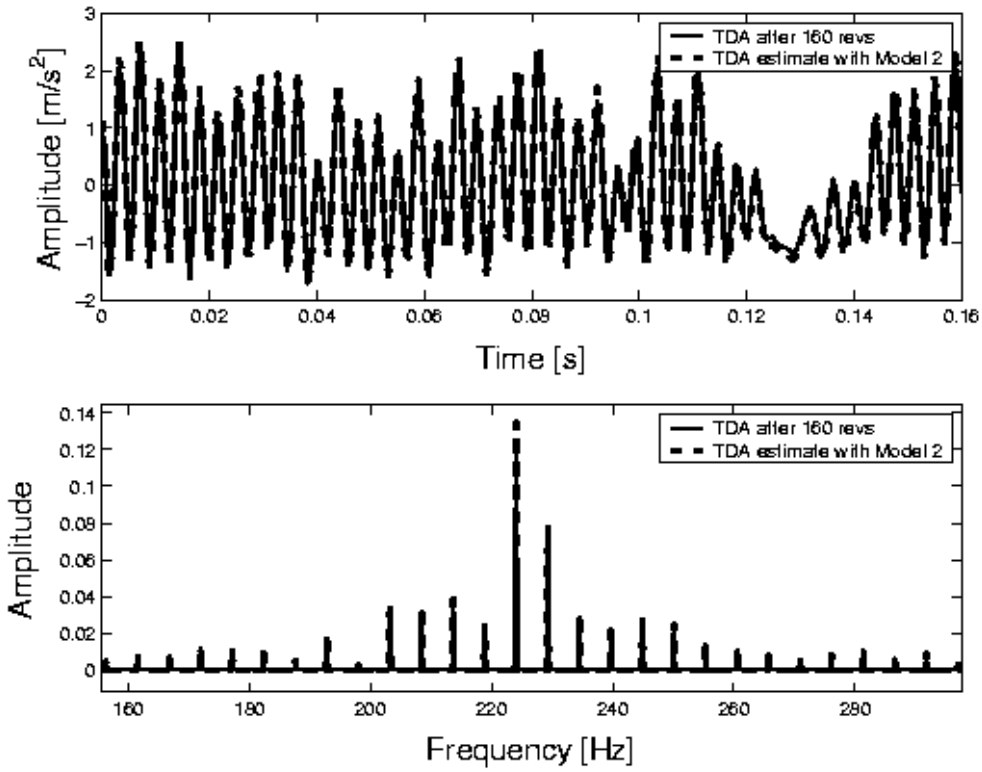

Figure 10 Model 2 ~~estimation~~Estimation results for simulation with a validation set measured from a test conducted under varying load conditions

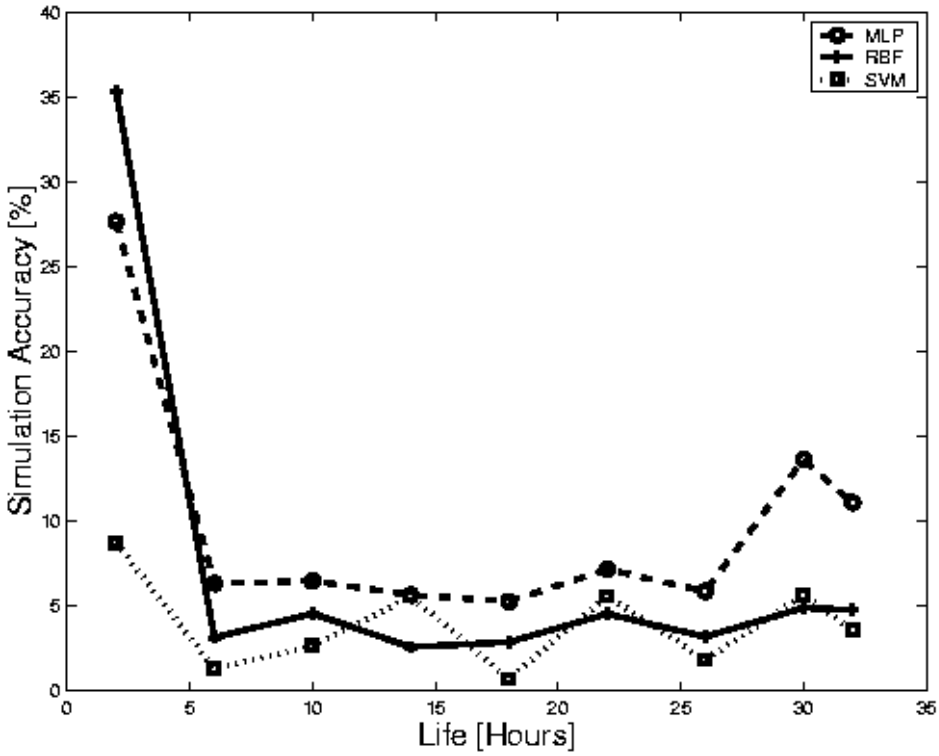

Figure 11 Model 2 ~~simulation~~Simulation accuracy η_{sim} vs. gear life under constant load~~ing~~ conditions

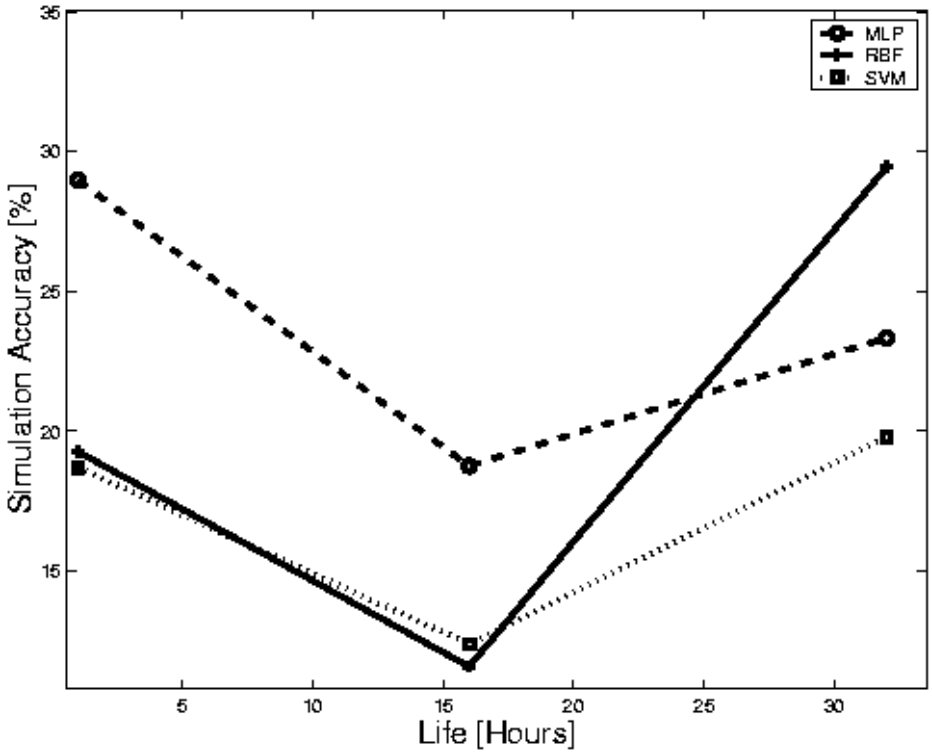

Figure 12 Model 2 ~~simulation~~Simulation accuracy η_{sim} vs. gear life under varying load conditions

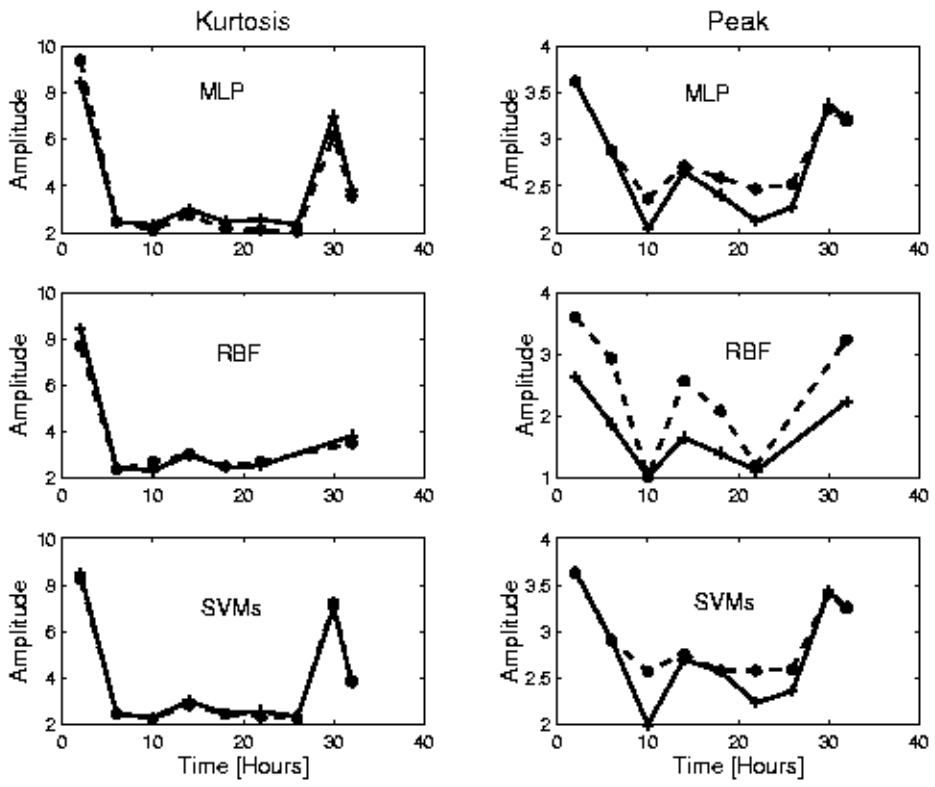

Figure 13 Comparison of kurtosis and peak values for the TDA calculated by direct averaging (+) and the TDA predicted by Model 1 (*) with MLP, RBF and SVMs

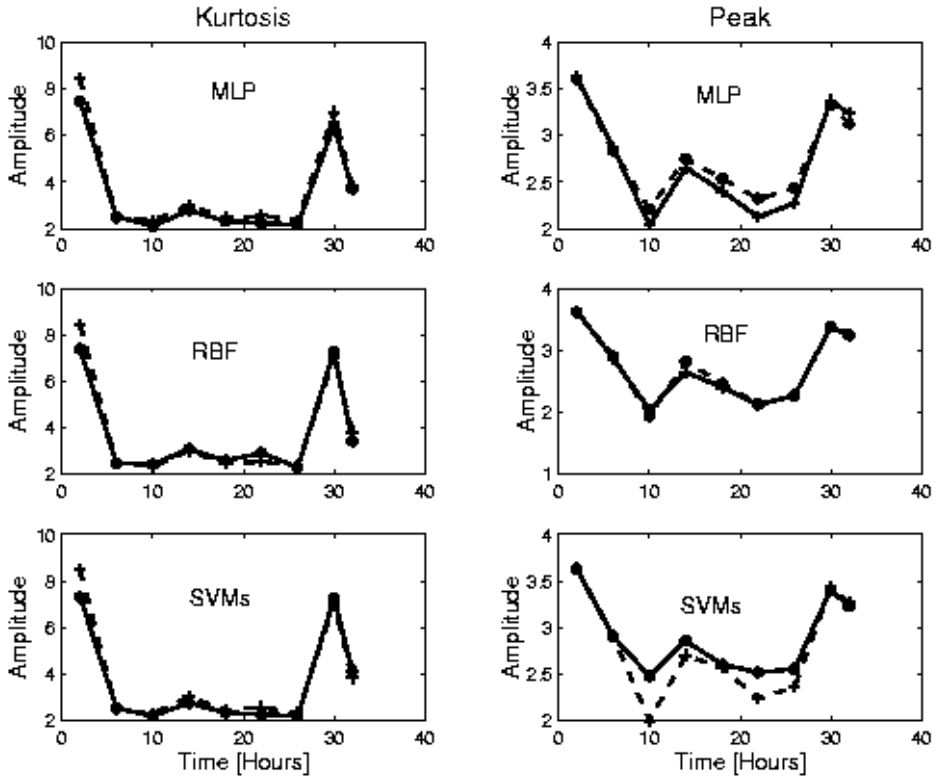

Figure 14 Comparison of kurtosis and peak values for the TDA calculated by direct averaging (+) and the TDA predicted by Model 2 (*) with MLP, RBF and SVMs.

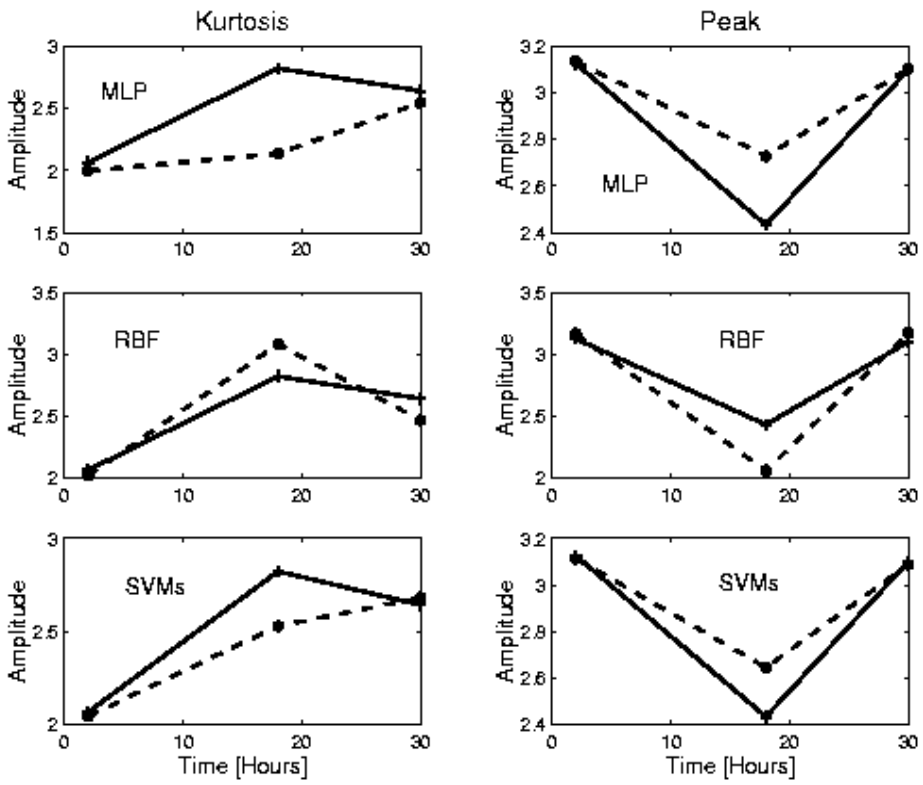

Figure 15 Comparison of kurtosis and peak values for the TDA calculated by direct averaging (solid line) and the TDA predicted by Model 1 (dotted line) with MLP, RBF and SVMs.

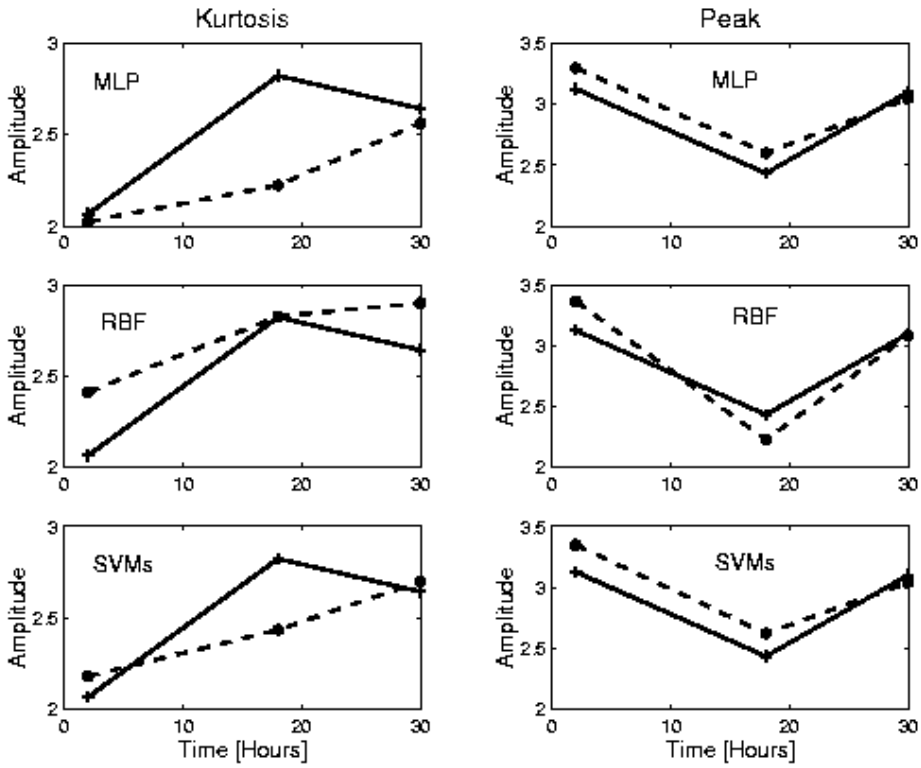

Figure 16 Comparison of kurtosis and peak values for the TDA calculated by direct averaging (solid line) and the TDA predicted by Model 1 (dotted line) with MLP, RBF and SVMs.