Journal of Sound and Vibration 327 (2009) 623–646

Contents lists available at ScienceDirect

Journal of Sound and Vibration

journal homepage: www.elsevier.com/locate/jsvi

Fault detection in variable speed machinery: Statistical parameterization

Jordan McBain, Markus Timusk *

School of Engineering, Laurentian University, Sudbury, Ontario, Canada P3E 2C6

ARTICLE INFO

Article history:
Received 10 January 2009
Received in revised form
10 July 2009
Accepted 17 July 2009
Handling Editor: K. Shin
Available online 18 August 2009

ABSTRACT

Variable speed machinery presents a particular challenge to automated condition-monitoring systems; changes in speed have a strong relation to the vibration response collected by accelerometers—the effect of which may mask fault conditions in standard condition monitoring techniques. In order to account for the effects of this measurable variable, the vibration response will be segmented into speed bins with a small range of speed. The mean and covariance matrix for the feature vectors in each speed bin will be computed in order to derive a statistical novelty boundary for that bin. Each component of these statistical parameters can then be interpolated or regressed in order to derive boundaries for speed segments where no training data is available. A comparison of the use of a statistical decision boundary and support vector boundaries, whose inputs have been centralized and whitened with these statistical parameters, will reveal a stronger classification approach. These methods were validated on data gathered from an experimental gearbox and motor apparatus operating at variable speeds; the results indicate a high degree of separability between data from healthy and faulted states—providing exceptional classification error.

1. Introduction

Rotating machinery is prevalent in most industrial applications; increasingly it is operated under varying duty—a condition offering many benefits including better process control, more efficient power usage, etc. For example, on demand ventilation systems using variable frequency drive controlled fans have provided the underground mining industry with a means of satisfying operational requirements at drastically reduced costs.

The transient nature of such machinery can have a troubling effect on attempts at online condition monitoring. Increase in speed and load can greatly distort the vibratory response of rotating machinery to the extent that traditional condition monitoring techniques, those which do not discriminate between modal parameters, may break down. Under the assumption that the vibratory parameters of a machine, operating with variable speed, will be strongly related to its speed, the authors will explore a statistical parameterization approach first suggested in the context of structural health monitoring [1] and will employ a new approach combining data whitening and support vector techniques to improve classification results.

There are certain non-stationarities present in signals from steadily operating machinery. The most well known of these is the modulation of bearing speeds due to changes in ball bearing sizes when subject to deflection in the loaded zone.

* Corresponding author. Tel.: +1705 675 1151x2243.
E-mail address: mtimusk@laurentian.ca (M. Timusk).

0022-460X/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jsv.2009.07.025

Similarly, steadily operating gears experience modulations in speed due to changes in apparent stiffness as the number of teeth in mesh vary.

When variable speed and load are introduced the issues become much more exacerbated. As speed varies, the machinery will move through regions of damping and resonance; the effects on condition monitoring become significant when health begins to degrade [2]. Changing speed will smear the spectral characteristics of faulty bearings by adding a frequency modulation component to the already amplitude-modulated signal. Variances in load will further modulate the amplitudes of the impulsive events and will increase the variance of the slippage effect with a resulting increase in the spectrum's variance and attenuation [3]. Vibrations from gear deflections will be similarly impacted by varying speeds and loads; characteristic spectral components associated with various gear faults will be altered by resultant modulation effects.

The condition-monitoring approach employed to manage these issues is one of many that have appeared in the literature for condition monitoring of systems with changing environmental variables. In the context of variable speed/load machinery, there are three predominant techniques: frequency-domain variants, artificial intelligence (AI) techniques, and parameter normalization.

The work in [4] investigated the use of frequency processing techniques such as the Wigner–Ville Distribution (WVD) and its qualities in the context of variable speed machinery, which may be limited by confounding spurious energy leakages in various parts of the WVD's spectrum. Similar work can be found in [5,6] in the monitoring of gear drives. Kar [7] examined the use of a scaled window convolved with a transient load signal, in a technique termed multi-resolution Fourier transform, in order to gain better resolution on critical frequency ranges. In [8], empirical mode decomposition was employed to account for the transient nature of large rotating machinery. Additionally, [9] developed order spectrum analysis which removes the effects of varying speed by re-sampling a vibration signal's spectrum by the speed of the machinery generating that signal; an application of an analogous method can be found in [10] where the instantaneous angular spectrum was used to monitor torsional vibrations in a diesel engine. Order spectrum analysis may only addresses the skew of frequency responses from variable speed and not the changes in amplitude/power from resonance or the consequent complexities from signal modulation effects. The instantaneous power spectrum was used in [11] for the detection of gear deterioration under varying load conditions. Mechefske examined the effects of variable speed machinery on spectral approximation techniques in [12]. Adaptive filters were trained in [13,14] to process occasional changes in damage signals—an approach that might be extended to variable duty applications.

AI techniques consist of neural networks used as means of nonlinear principle component analysis (NL/PCA) to detect unmeasured environmental parameters in [1] whose limitations are subject to the assumptions of NLPCA (i.e. Gaussian distributions in some augmented space); "large-normal condition models" consists of techniques where a classifier is trained over all ranges of operation but inevitably are dependent on the detection of a fault being invariant on changes in operating modes (a poor assumption) [15]; and, by varying a classifier depending on changes in the mode of operation as demonstrated by the kinematics of the machinery at the time (which may not be sufficiently discriminating to distinguish all effects of changes of mode) [16–19].

Parameter normalization techniques found in [20,21] focus on reducing the effects of modal changes according to measurable changes in those modes. Principal component analysis and support vector regression were used to filter out temperature effects on structural monitoring in [22]. In [23], the authors normalize machinery vibration data first with order spectrum analysis to remove speed change effects (which may not be adequate to address machinery resonances), apply a load-normalizing technique (whose sophistication in addressing all spectral aspects of changes in load is at best dubious), followed by a pseudo-Wigner–Ville distribution from which statistical parameters are computed for novelty detection with statistical process control.

2. Statistical parameterization

In [15,24], Worden examined how a structure's vibration measurements varied with modal parameters such as temperature. He conducted this effort based on the reasonable assumption that changes in the vibratory parameters of a structure have some deterministic relation to changes in the temperature of the structure. He assumed by determining this relation and accounting for it in a fault detection scheme, that he could eliminate the poor classification results commonly associated with systems impacted by environmental factors. Worden presented good results when evaluating this method on condition monitoring of the theorized vibration of modeled structures; this work will retrace his steps in an effort to evaluate how well his technique works in the context of real variable-speed machinery.

In the context of variable-speed machinery, elements of the vibratory response are assumed to have a strong relation to the speed of the given machinery. As a consequence, it follows that the statistical characteristics of the distribution of whatever parameters are chosen to represent the vibratory response should be strongly related to the machine's speed. Fig. 1 demonstrates how this assumption might appear in 2-dimensions (i.e. with two parameters representing the vibratory response). The means of the distributions representing various speeds (10, 20, and 30 Hz in the case of the figure) are presumed to increase with increasing speed. The variances of these distributions should also experience a similar relation with speed possibly due to the system's resonance response.

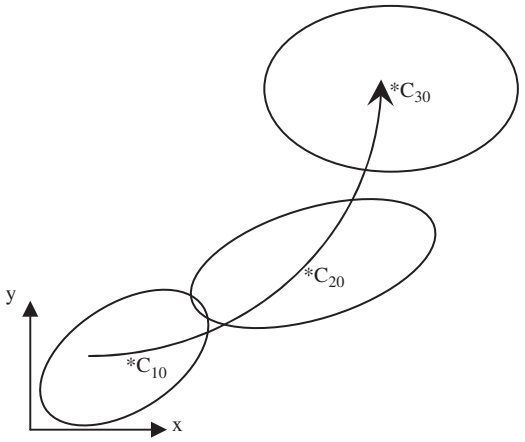


Fig. 1. Theorized changes in vibration parameter distribution with various speeds (10, 20, and 30 Hz) for the 2-dimensional case (means and variances presumed to increase with seed).

If these assumptions are valid, then each of the statistical parameters (i.e. each component of the mean vector and each component of the variances) could be interpolated or regressed in order to determine those of speeds for which data was not collected. In the case of regression, a model for each of these components could be retained rather than retaining each statistical characteristic for each distribution. In practice, it will be demonstrated that the model generated by regression provides too poor a fit for the changes in statistical parameters with speed (which are not at all smooth in the case of rotating machinery).

A segment of the vibration signal from an accelerometer is first parameterized and then grouped into speed bins based on the machine's average speed in each segment. Once this is accomplished, the statistical distribution of each speed bin must be estimated. If the data are normally distributed, with n parameterized segments, the estimate of the mean ($\hat{\mu}$) and covariance matrix ($\hat{\Sigma}$) are given by maximum-likelihood estimation [25]:

$$\hat{\mu} = \frac{1}{n}\sum_{k=1}^{n} x_k \tag{1}$$

$$\hat{\Sigma} = \frac{1}{n}\sum_{k=1}^{n}(x_k - \hat{\mu})(x_k - \hat{\mu})^t \tag{2}$$

which is to say, that the mean and covariance of the sample population are assumed to approximate the population's true values. This assumption warrants further investigation since the quantity of data to characterize each speed bin's distribution is often limited; as the dimensionality of the space increases, the amount of data to accurately describe each bin's distribution will also increase. In pattern recognition this is known as the curse of dimensionality. This is a demanding problem since classification often involves a precarious trade off with the desire to increase the problem dimensionality, in order to obtain better classification results, and a desire to reduce the amount of data necessary to describe the volume of the space. Worden's statistical parameterization, in a sense, is doubly cursed by this symptom since the number of modal parameters could increase (e.g. one could examine the distribution of data not only for speed but also for load). In this case, not only would one need sufficient data to characterize the dimensionality of the space chosen, one would also need to do so for all combinations of speeds and loads.

3. Novelty detection

The purpose of novelty detection is to describe a normal state with representative data, presumably available in abundance, in order to discriminate it against faulted states with exemplars which may not have been available during training. A host of novelty detection techniques are available (see [26,27]); two will be employed in this research—discordance tests and support vector data descriptors.

3.1. Discordance test

Worden used a discordance test to discern if a test object fell outside the distribution of the healthy data for a given modal instance. The novelty boundary in the case of uni-variate data is given by the radius, R, from mean μ, such that 95% of the healthy data lies within the interval:

$$P(|x - \mu| \leq R) = 0.95 \tag{3}$$

In the case of normally distributed data with standard deviation σ, this condition reduces to

$$\frac{|x-\mu|}{\sigma} \leq 1.96 \tag{4}$$

such that the novelty boundary is given by the extremes of

$$-1.96\sigma + \mu \leq x \leq 1.96\sigma + \mu \tag{5}$$

as demonstrated in Fig. 2.

Multivariate data requires greater effort to derive the appropriate boundary. In the event that the distributions of each of the components of the parameterized vibratory response are independent, the combined probability distribution is simply the product of the independent uni-variate ones. The joint probability of multiple independent variables is well known, by fundamental laws of probability, to be the product of the probability of each random variable (i.e. $P(A \cap B) = P(A) * P(B)$). If each component of the d-dimensional parameterized vibratory response has normal probability distribution $p_{x_i}(x_i) = N(\mu_i, \sigma_i^2)$, the joint probability distribution is given by [25]:

$$\begin{aligned} p(\vec{x}) &= \prod_{i=1}^{d} p(x_i) = \prod_{i=1}^{d} \frac{1}{\sqrt{2\pi}\sigma_i} e^{-(1/2)(x_i-\mu/\sigma_i)^2} \\ &= \frac{1}{\sqrt{(2\pi)^d \prod_{i=1}^{d} \sigma_i}} e^{-(1/2)\sum_{i=1}^{d}(x_i-\mu/\sigma_i)^2} = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} e^{-(1/2)(\vec{x}-\vec{\mu})^t \Sigma^{-1}(\vec{x}-\vec{\mu})} \end{aligned} \tag{6}$$

Inherent in this analysis is the assumption that the components of the parameterized vibratory response are in fact independent. This assumption is a poor one in the context of variable speed machinery as components of a parameterized vibration response are often dependent. Many of the spectral components of the vibratory response of machinery are intrinsically related as demonstrated in [28,29]. For instance, the spectral characteristics of bearings are well known to consist of impulses induced by faults on the inner or outer races, whose amplitudes are amplified by the load born by the bearing as it transits through the race's fault. The resultant signal is amplitude modulated; amplitude modulated signals consist of three primary spectral components, the frequency component from the shaft speed with two identical sidebands representing the frequency response of the interaction of the bearings with the fault. Other more simple examples can be found when examining harmonics of shaft speed generated by gear meshing and conditions such as unbalance or oil whirl/whip. An assumption of a normal distribution of multivariate data rests on an underlying assumption of the independence of the components of the data in use [25]. Worden's assumption of normally distributed data may have worked reasonably well in the context of monitoring the frequency response of structural vibrations [24]; it may, however, prove a poor one in the context of variable speed machinery because of the complex interrelations of the frequency components of machinery's vibratory response.

This assumption will nevertheless be undertaken since the alternative is to attempt to define the relation between these components and then generate a conditional multivariate probability distribution (see [30]). While much effort has been

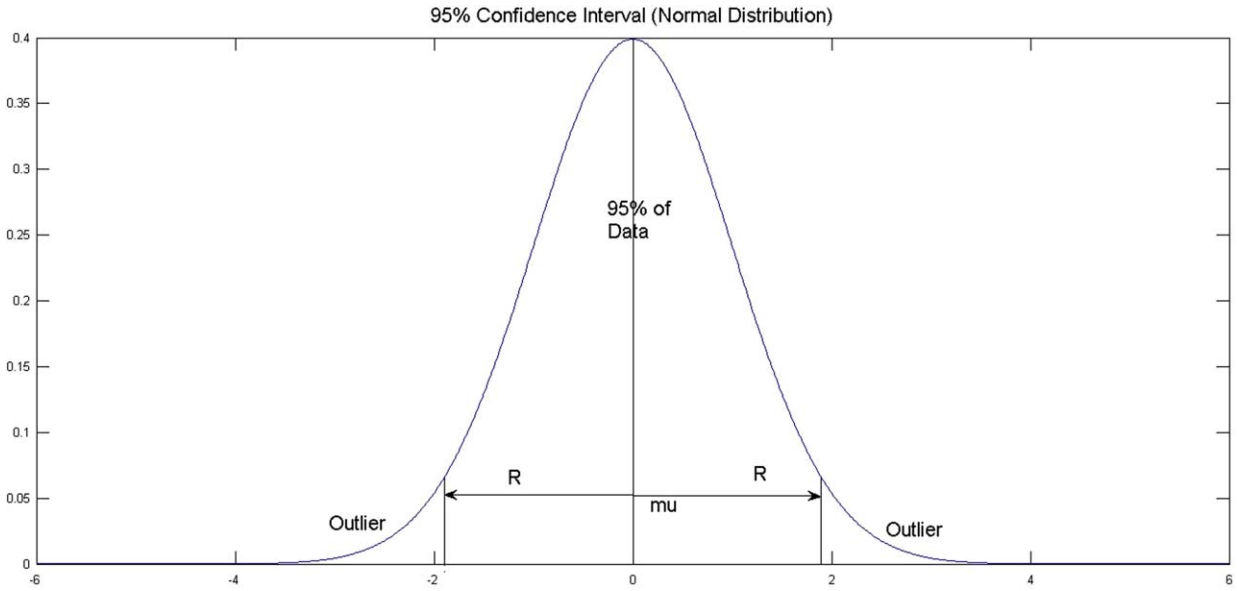


Fig. 2. Boundary for uni-variate normal distribution.

made to model these relations (see, for instance, [3]), the task is a complex one and has not been completed to the extent necessary for application. The signal-to-noise ratio in rotating machinery is simply too poor for any such relation to be of practical value at present.

Fig. 4 demonstrates the joint probability distributions of the two uni-variate distributions in Fig. 3 (the first distribution having $\mu = 5$ and $\sigma = 4$ and the second having $\mu = 10$ and $\sigma = 8$). The joint distribution has $\vec{\mu} = [5\ \ 10]$ and $\Sigma = \left[\begin{smallmatrix} 4 & 0 \\ 0 & 8 \end{smallmatrix}\right]$. The level curves, shown in Fig. 5, of the distribution in Fig. 4 are determined by the Mahalanobis squared distance given by (the exponential in Eq. (6))

$$r^2 = (\vec{x} - \vec{\mu})^t \Sigma^{-1} (\vec{x} - \vec{\mu}) \tag{7}$$

The Mahalanobis distance of the joint distribution of Fig. 4, reduces to

$$\begin{aligned} r^2 &= \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \right)^t \Sigma^{-1} \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \right) \\ &= \begin{bmatrix} x_1 - 5 & x_2 - 10 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}^{-1} \begin{bmatrix} x_1 - 5 \\ x_2 - 10 \end{bmatrix} \\ &= \begin{bmatrix} x_1 - 5 & x_2 - 10 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{8} \end{bmatrix} \begin{bmatrix} x_1 - 5 \\ x_2 - 10 \end{bmatrix} \\ &= \begin{bmatrix} \frac{x_1 - 5}{4} & \frac{x_2 - 10}{8} \end{bmatrix} \begin{bmatrix} x_1 - 5 \\ x_2 - 10 \end{bmatrix} \\ &= \frac{(x_1 - 5)^2}{4} + \frac{(x_2 - 10)^2}{8} \end{aligned} \tag{8}$$

which is the equation of the level curves/ellipses of Fig. 5 centered at $\vec{\mu} = [5\ \ 10]$. In practice, the covariance matrix Σ will not be diagonal and the equation of the level ellipses (or ellipsoids in higher dimensional spaces) will involve cross terms

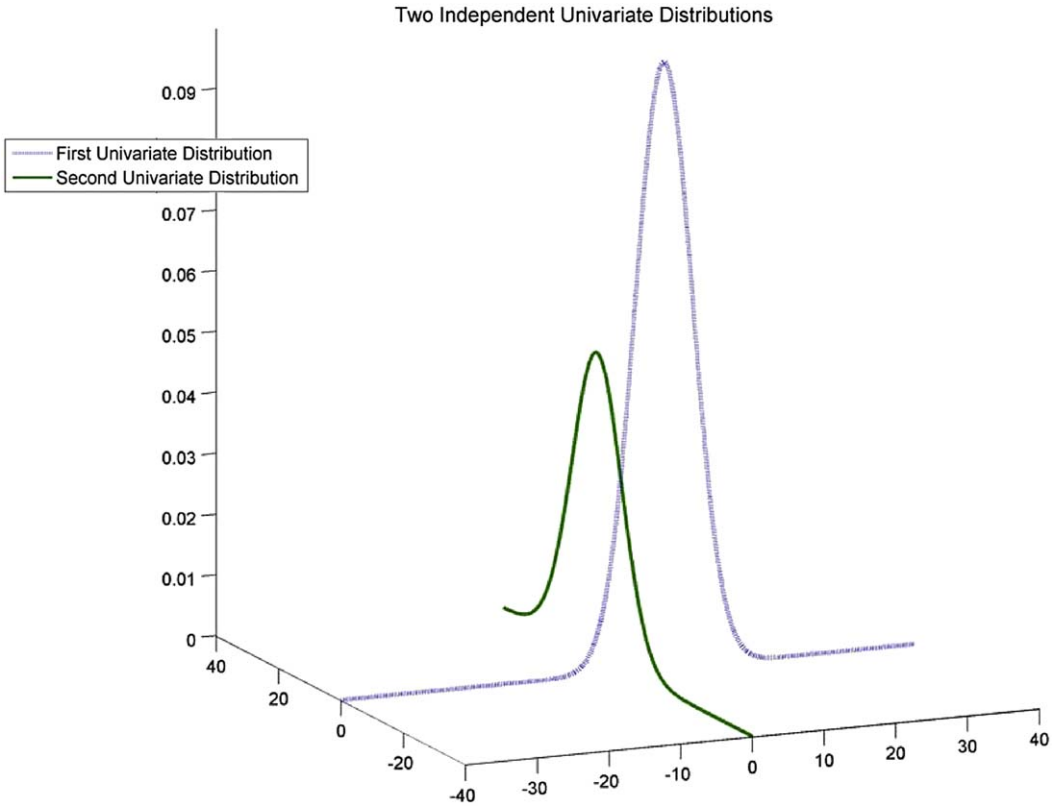


Fig. 3. Two independent uni-variate distributions.

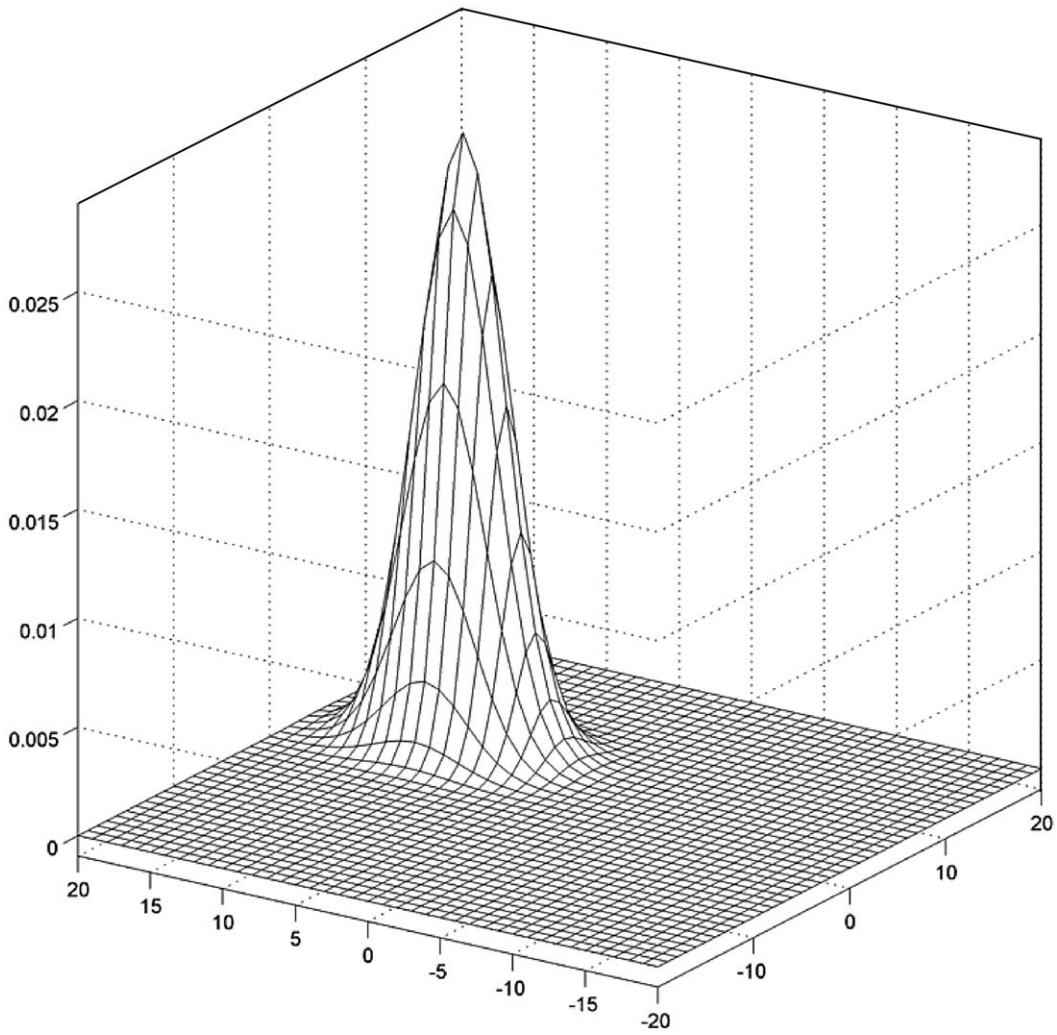


Fig. 4. Joint distribution of two independent uni-variate distributions.

that will rotate and skew the axes of the ellipsoids such that their principal axes no longer are in alignment with the coordinate system's axes (and principle component analysis (PCA) will be necessary to determine the orientation of the ellipsoids). This task, however, is not relevant at this stage as the primary focus is the derivation of the Mahalanobis squared distance that encapsulates a certain percentage of the healthy data to act as a novelty boundary.

The analysis surrounding Eq. (4) is more difficult in the multivariate case since there are no tables available for cumulative multivariate normal distributions. One could employ a numerical-method approach to find the boundary containing a certain quantile of the data (see the Monte-Carlo approach in [31]) or one could employ the approximation that the Mahalanobis squared distance fits a chi-squared distribution [32]. In the latter case, the Mahalanobis squared distance, k, in d dimensions, containing the quantile, $0 \leq q < 1$, is given by

$$k = \text{chi2inv}(q, d) \tag{9}$$

3.2. Whitened data for support vector description

An alternative approach to using quantiles of a statistical distribution is to centralize and whiten the data given the modal parameters and their relation with the data's distributions. Once the data has zero mean and unit covariance, any of a variety of artificial intelligence techniques for pattern recognition could be used to detect outliers in test patterns presented to the detector.

Centralizing the data, by subtracting the distribution's mean from each of its data points, will ensure it has zero mean. The task of ensuring uniform covariance, by whitening the data, is more complex. In the case of the data analyzed in Eq. (8)

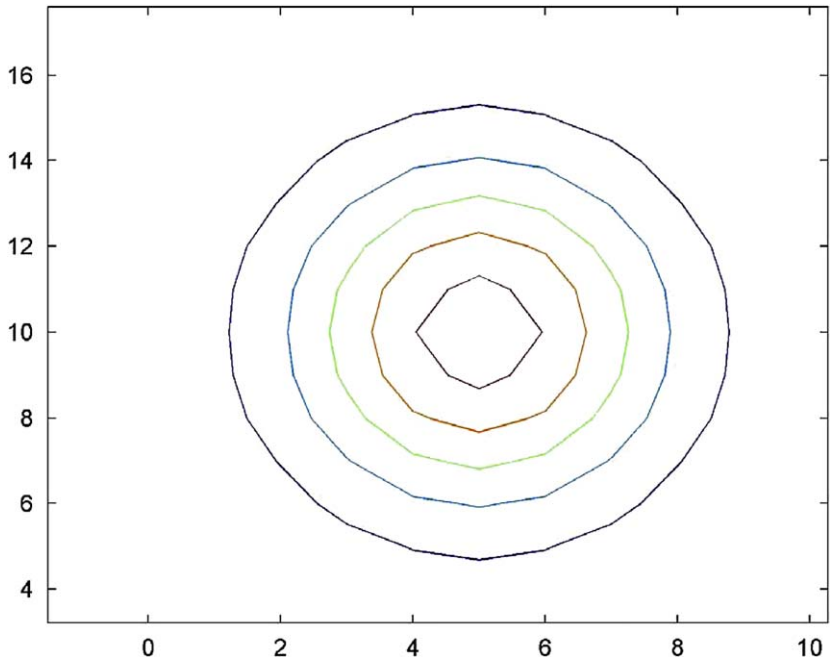


Fig. 5. Level curves of the surface in Fig. 4.

(for the distribution in Fig. 4), whitening a data point can simply be achieved by multiplying by the inverse of the covariance matrix since the principle components of the ellipses are aligned with the axes of the coordinate system. In the case where the ellipsoids are skewed, i.e. the covariance matrix contains cross terms, one must first find the ellipsoids' principle components and then find the matrix A_w that will reduce them to unit length. Duda [25] describes the whitening matrix as

$$A_W = \Phi \Lambda^{-1/2} \tag{10}$$

where Φ is the matrix whose columns are the orthonormal eigenvectors of Σ, and Λ is the diagonal matrix of the corresponding eigenvalues.

Once the data has zero mean and unit variance, a support vector classifier can then be trained on them in order to provide a more robust boundary description between data describing healthy and faulted machinery states. Tax's support vector data descriptor (SVDD) as described in [33] will be employed to this end. The SVDD maps data into a high dimension kernel space and attempts to fit a sphere of minimal radius around it; a minimal-radius sphere will ensure a tight boundary around the healthy data. Mapping into a higher dimensional space is achieved via the "kernel trick" where inner products in the sphere's calculation are replaced with a kernel function as outlined in [33]. Once mapped back into the initial space, the SVDD's spherical boundary is able to enclose complex non-spherical, non-convex regions as demonstrated in the two related diagrams in Fig. 6. The left diagram in Fig. 6 demonstrates the SVDD's complex decision boundary and the right diagram demonstrates the corresponding novelty scores for various points of this boundary. The tightness of fit demonstrated in the left hand of Fig. 6 is affected by the choice of kernel parameter; in this research, the Gaussian kernel was employed which has one kernel parameter known as sigma—a high sigma produces a very generalized decision surface while a small sigma (in the range of 1–6), produces a tighter fit as shown.

With zero mean and unit variance data, it is hoped that the SVDD may provide better classification results than the discordance tests since the assumption of normally distributed and independent data has been determined to be a poor one. The normalization process depends, in part, on these same assumptions; however, it is thought that the SVDD may be better able to accommodate the skewed results from normalizing data lacking a Gaussian distribution.

4. Experimental methodology

4.1. Apparatus

The data employed in this research were generated on a Spectraquest® gear dynamics simulator as shown in Fig. 7. This simulator was configured using a computer controlled variable frequency drive to subject a gearbox to a duty cycle of variable load and speed. The apparatus consists of a two-stage parallel reduction gearbox driven by a 3 hp variable speed

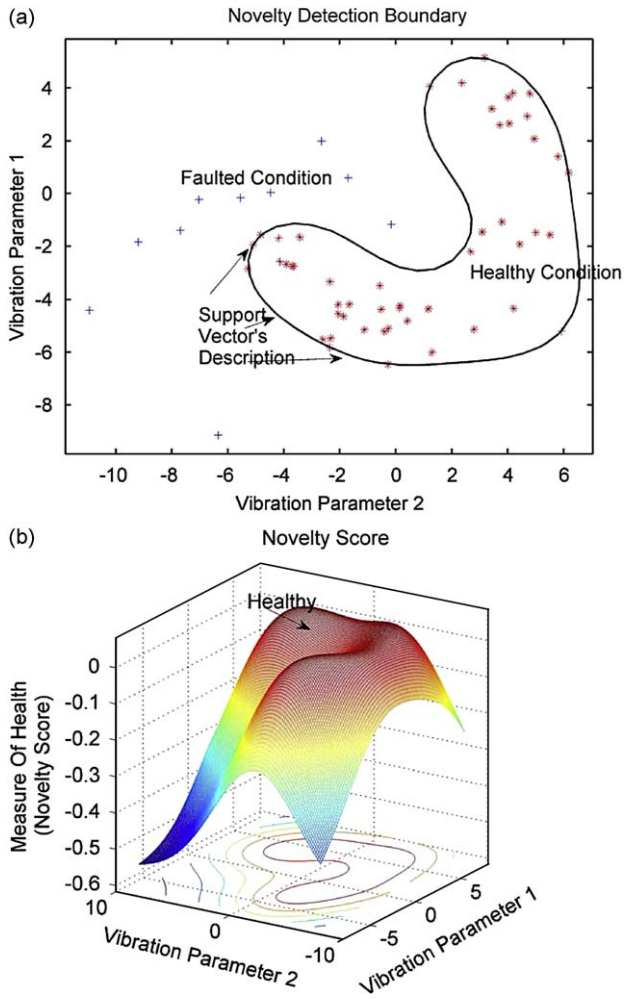


Fig. 6. SVDD's boundary (in part generated with [33]).

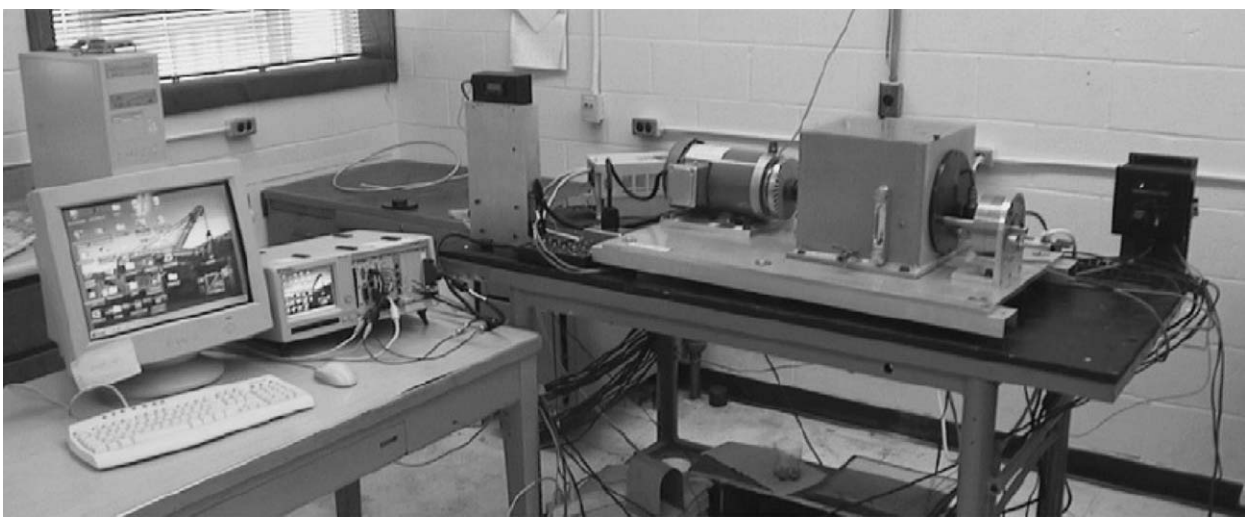


Fig. 7. Experimental apparatus with data acquisition system.

induction motor and loaded with a magnetic particle brake. The data acquisition system is displayed in Fig. 7. A National Instruments PXI data acquisition system was used to control the motor's speed and particle brake's loading as well as to monitor the signal of several industrial ceramic shear ICP accelerometers from IMI sensors. These sensors had a reported sensitivity of 10.02 mV/(m/s^2) with a claimed error of $\pm 5\%$ over a range of 0.5–6500 Hz. These accelerometers' vibration measurements were taken in radial directions both horizontally and vertically on both the gearbox and motor bearings. All measurement channels were sampled synchronously at 4 kHz.

The general process for selecting an appropriate sampling frequency is suggested in [34]. The simulator's motor has a maximum rotational speed of 57.5 Hz which can be reduced by the reduction gearbox to as low as 16 Hz—serving as the range of potential shaft frequencies. The faults under consideration will appear at multiplicative frequencies of shaft speed. Rotor faults are generally synchronous and will not exceed the shaft frequency ranges. Gear faults generally appear at gear mesh frequencies given by the ratio of the output to input gears multiplied by the shaft's frequency. In the case of the Spectraquest simulator, the highest gear mesh frequency of concern is approximately 800 Hz (for a gear ratio of 48–80 at a shaft frequency of 16 Hz). Bearing faults produce faults at characteristics frequencies (see [29])—with an ultimate signal containing sidebands multiplexed around a given shaft frequency with a bandwidth that can be quite large. To be safe, it was assumed that frequencies as high as 2 kHz could be relevant; this "natural system frequency" was then multiplied by 2 under Nyquist's rule to obtain the appropriate sampling frequency.

The gearbox and motor were instrumented with two accelerometers, measuring vibration in the vertical and horizontal planes. The motor accelerometers measured vibrations in the motor housing nearest to the shaft bearings nearest the output. The gearbox bearings were mounted on machined aluminum blocks which were in turn affixed to a metal plate which housed the main shaft bearings.

4.2. Faults

Motor and gearbox faults were analyzed in the course of these experiments. Four motors were alternated through the test bench—a healthy motor, one with a combination of bearing faults in the front motor bearing, one with broken rotor bars, and one with rotor unbalance. Motors were located on pins to ensure repeatability between tests by precisely controlling alignment.

Gearbox faults consisted of scenarios with a missing tooth, a chipped pinion tooth, an outer race bearing fault, a bearing faulted on the inner and outer races, and fault-free conditions. Repeatability was ensured by the apparatus' ability to accurately re-position components between subsequent tests.

4.3. Segmentation

In order to segment the vibration data into regions of "steady" speed and load, the vibration data were grouped according to a specified number of shaft rotations (Fig. 8). It was thought that selecting data over only a few shaft rotations would provide a minimum segment length to ensure both signal coherency and stability of modal parameters. Equipment operating at high speeds will provide a sufficient data segment very quickly and equipment operating at low speeds will take longer to generate an acceptable signal length for processing as demonstrated in Fig. 9.

The sensor data segments defined by this number of shaft rotations were then windowed in order to minimize the introduction of high frequency components associated with quickly switching on and off the signal with a rectangular window. A Gaussian window was employed with 70% overlap to ensure interesting events would not be overlooked. Fig. 9 shows the effects on the window characteristics of both a steady segment of operation (top figure) and that of a segment with drastically reducing speed (bottom figure) as well as the consequential effect on the width of the window (with 10 rotations in each window guaranteed).

To validate the assumption that collecting sensor data over a pre-specified number of shaft rotations (i.e. 10) produces segments of reasonably stable modal parameters consider Fig. 10. Here, as expected, when shaft speed is small, it takes a longer time to collect data of sufficient length and the segment has a wider range of speeds. When the rotor system has sufficient speed, data segments are collected quickly and the standard deviation of a segment is reasonably small depending on the slope of the speed curve. Where the standard deviation is larger, one may place less confidence on that segment's effect on the classifier.

Gathering data over multiple operating sessions with speed profiles similar to those in Fig. 10 produces a histogram as demonstrated in Fig. 11. This figure has 40 equally sized speed segments and was collected over 61 files of 12 s length (12 min total). It demonstrates that higher speeds have a greater number of objects for training for the given speed profile. The large predominant peak is due to a long period spent at a constant speed (as shown in Fig. 10). The distribution of data is primarily dependent on the shape of the speed curve; in general, however, one can expect that data collected at higher speeds will be more populous since it takes much less time to collect data over a fixed number of shaft rotations.

The tradeoff in the case of speed buckets should be reasonably clear. The greater the number of buckets, the greater will be the resolution of the data buckets and the lesser will be the number of elements in each bucket. Fig. 11 shows the effect of varying the number of bins and its effect on bin size and number of elements in each bin. For a fixed amount of data, bin size will have to be balanced against the dimensionality of the space used due to the curse of dimensionality.

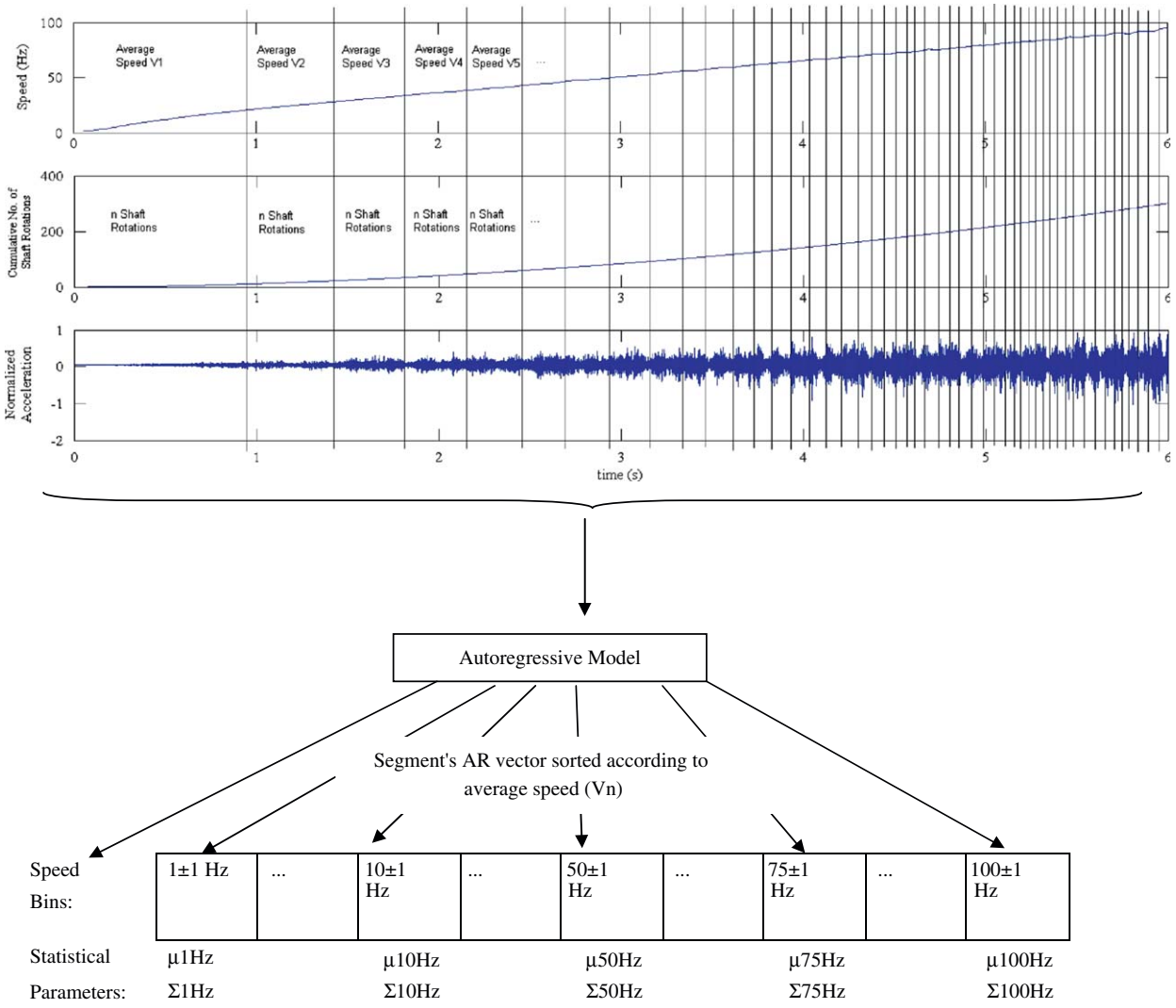


Fig. 8. Theorized signal segmentation (based on constant no. of shaft rotations) (segment overlap not shown).

4.4. Feature extraction

There are a number of parameters that could be employed to describe a segment's vibration response. One could evaluate time-domain features such as crest factor, average power, kurtosis, impulse factor, etc. or one could evaluate frequency-domain characteristics given by similar parameters in the frequency domain. Given the success of the frequency-domain parameters for fault detection [16,24], an autoregressive model of a vibration segment will be employed.

An autoregressive model (AR) is a model of a statistical process generated by regressing previous values of that statistical process with itself. AR models can be used for prediction of the future values of a process or they may be used to provide a simplified representation of the peaks and troughs of a system's frequency response. The spectrum of a normalized power series, $X(z)$, can be modeled in the z-domain according to

$$X(z) = \frac{k}{1 - \sum_{i=0}^{D-1} w_i z^{-i-1}} \tag{11}$$

when a linear combiner with D delays is used for the model. Recursive least squares (RLS) is often used as the learning algorithm for the weights of this model [35].

A higher order AR model will be able to approximate a system with a greater number of spectral peaks. For instance, Fig. 12 illustrates the effect of higher orders of AR models; an AR100 has a far greater number of spectral components than an AR25. This trend is illustrative of the tradeoffs involved with the curse of dimensionality; a larger AR model will provide

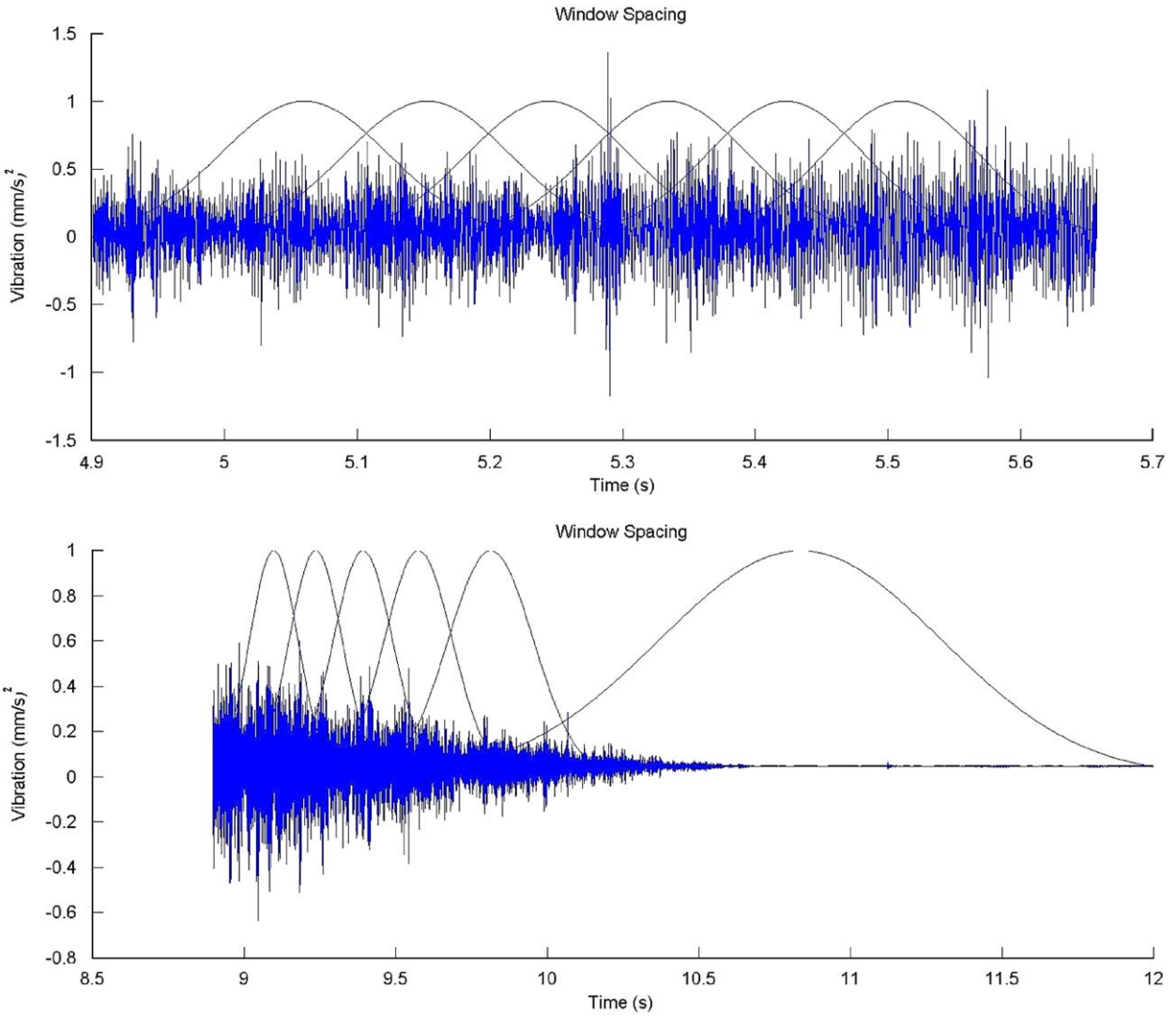


Fig. 9. Window parameter selection (top: steady speed segment—bottom: slowing segment).

more distinguishing components (and thus, potentially, a greater ability to improve classification) but has a much larger volume of space requiring a greater number of data points to fully describe a given class.

5. Classification results

Exploring the technique in 3-dimensions will provide intuition into how the technique works, which can then be extended into higher dimensions with the hope of achieving better classification results. As previously mentioned, statistical parameterization is doubly subject to the curse of dimensionality—a limitation which should be borne in mind when selecting feature space dimensionality.

Figs. 13–15 demonstrate the scatter of healthy data from various speed bins; generally the groupings are quite separable (even when comparing consecutive speed bins in Fig. 14). The distribution of each bin tends to be elongated with what could be a Gaussian distribution. It seems that irrespective of the prior analysis (i.e. requiring independence of vibratory components in order for data to follow a multivariate normal distribution) the underlying assumption of a multivariate Gaussian distribution may not be such a poor one.

Fig. 16 demonstrates that the data for a given speed bin from each class are quite separable. It seems possible then, that if one could fit a normal distribution to the healthy data, that a discordance test might prove a suitable boundary for separating faulted data from healthy data. With these initial intuitive checks undertaken, it seems a worthwhile investment to investigate statistical parameterization for variable speed fault detection.

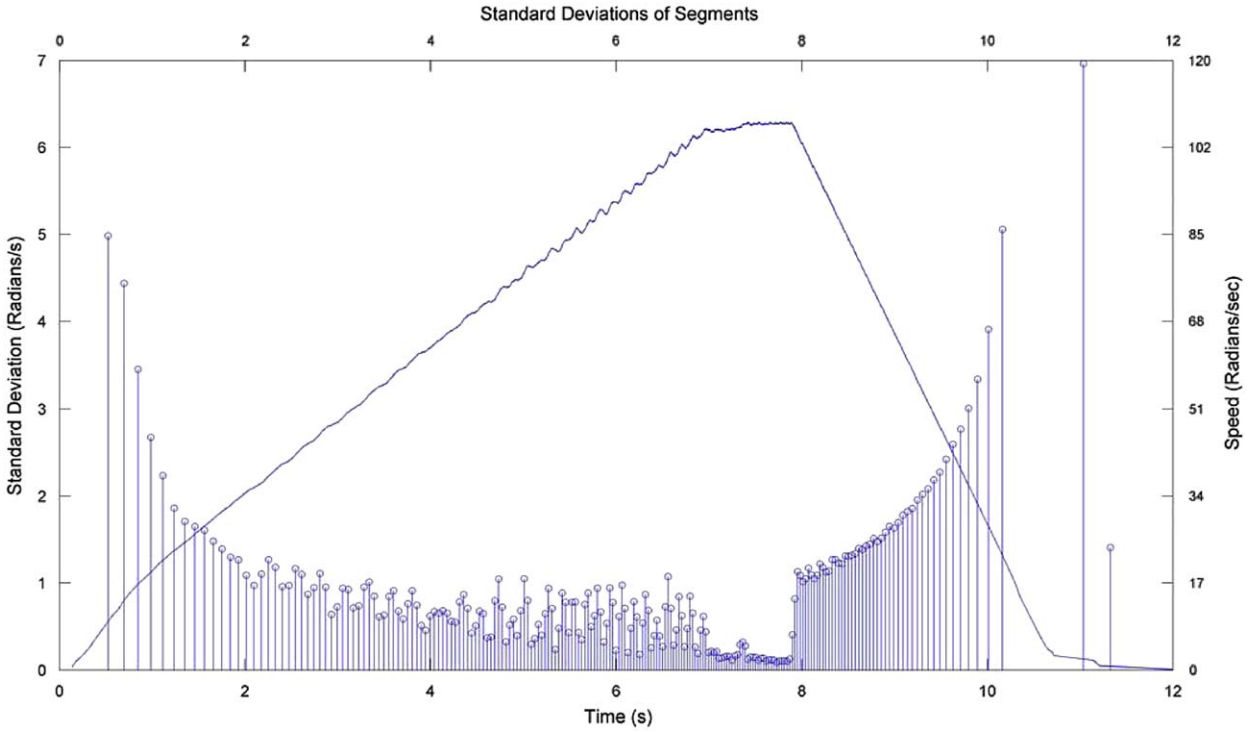


Fig. 10. Standard deviations of segments (standard deviation as stems with typical speed profile superimposed).

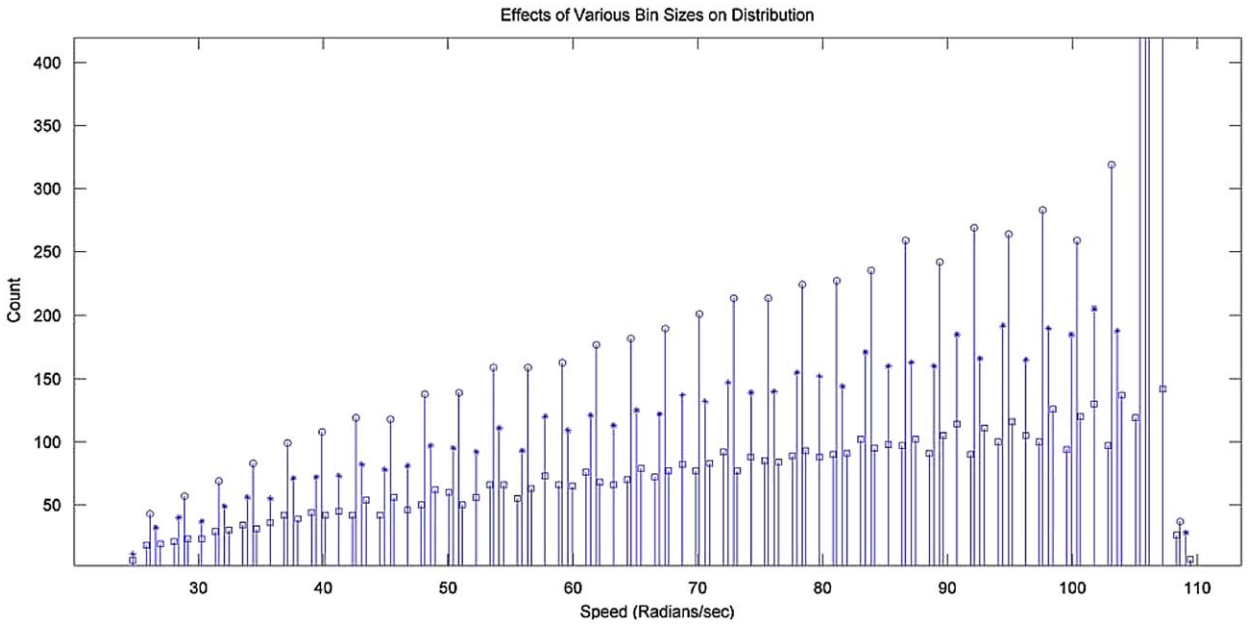


Fig. 11. Effects of various bin sizes on distribution.

5.1. Regression

In the case of regression, the first step is to fit each of the components of the statistical parameters to a polynomial model. Fig. 17 demonstrates the three components of the mean vectors for each of the speed bins for 100 speed bins (left figure) and 50 speed bins (right figure). Generally, the trends of the mean's components are smooth and appear not to be affected by the number of speed bins employed. In [24], the mean vector components were far smoother than those shown

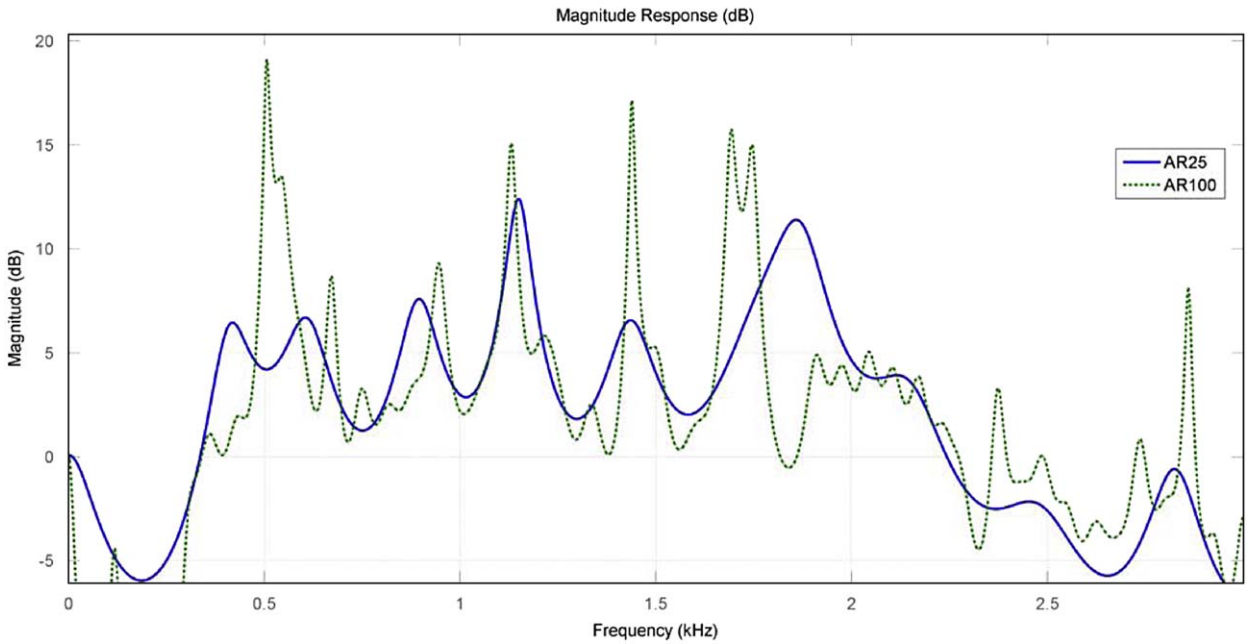


Fig. 12. Various orders of AR models.

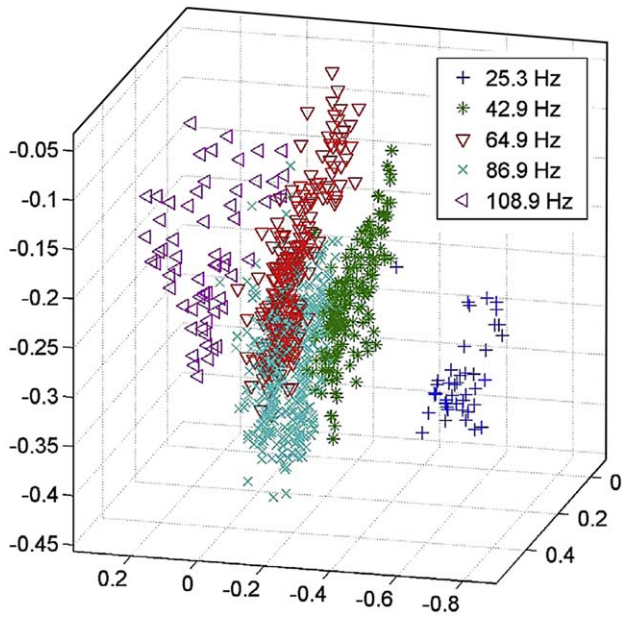


Fig. 13. Scatter plot of data from various bins.

in Fig. 17. When a fourth order polynomial is fit to each of these components (Fig. 18), the regression model can be seen to track the actual data reasonably well until the higher speed ranges. At this point there is some undesirable deviation from the true values; this problem is not solved by varying the order of the polynomial fit.

The problem becomes even worse when comparing the components of the covariance matrix. In Fig. 19, the variances (i.e. diagonal entries of the covariance matrix) are plotted for each of their speed bins. The variances are wild and could not reasonably be expected to satisfy a simple polynomial model. These variances are likely connected to the machine's resonance response; in resonance, there are wideband excitations in the machinery and as such, one would expect more chaotic frequency responses—ultimately producing irregular changes in statistical parameters of the feature vectors. Machinery operating in resonance bands will also suffer from a diminished signal-to-noise ratio. This notion supports the widely accepted practice that resonance bands are avoided during condition monitoring.

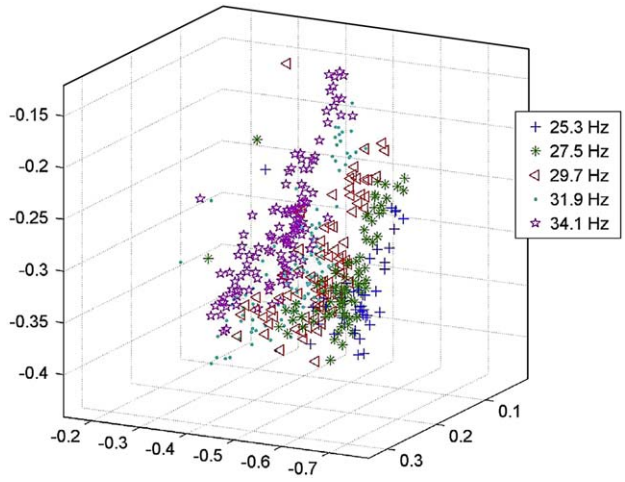


Fig. 14. Scatter plot of data from consecutive bins.

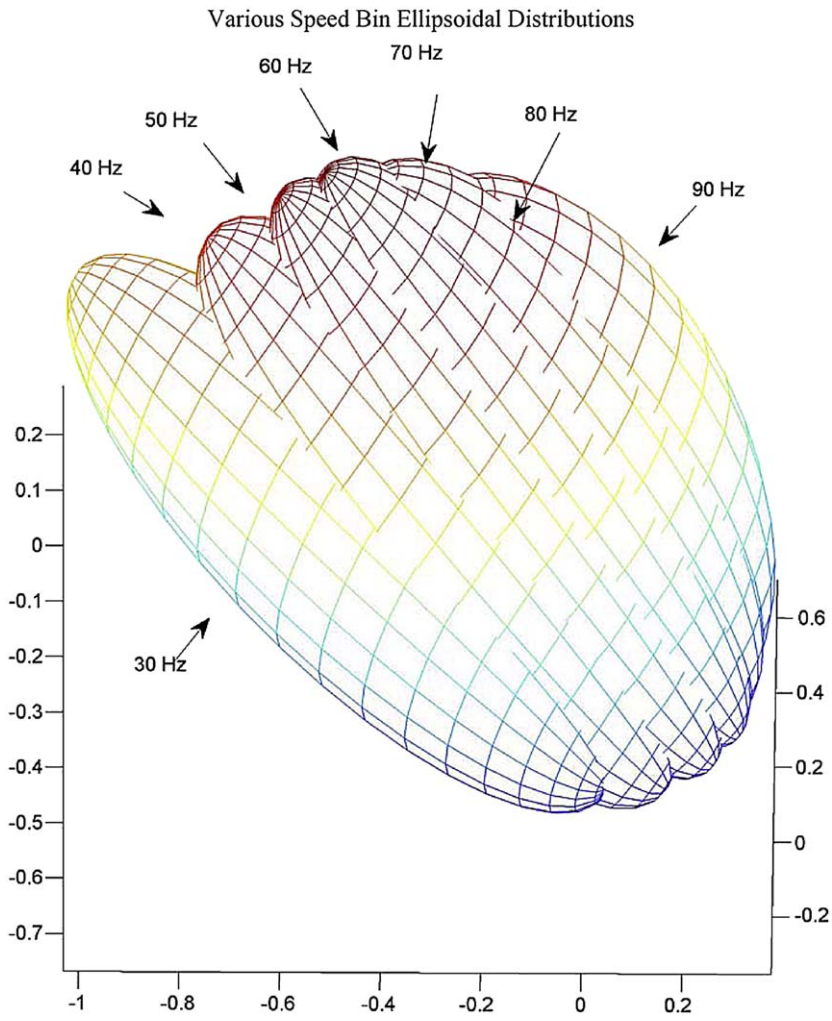


Fig. 15. Various speed bin ellipsoidal distributions.

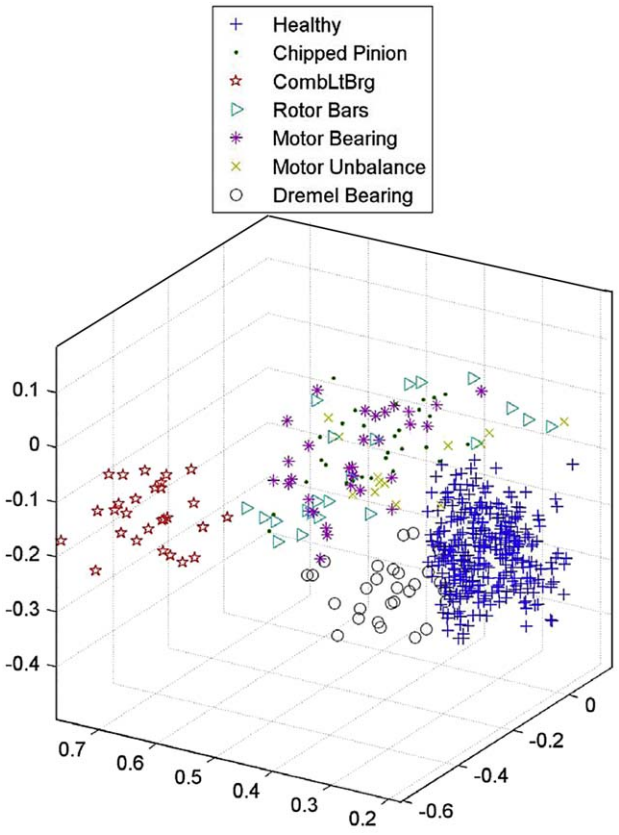


Fig. 16. Healthy and faulted data clusters (Bin 75.9 Hz).

The case against regression as a means of parameterizing the statistical elements has all but been made. With the additional concern, that through regression, one might derive a covariance matrix which is not positive semi-definite (i.e. one not satisfying $x^t \sum x \geq 0$), regression becomes an undesirable choice. This problem was originally identified in [24]; while Worden, developed a method to ensure that a derived covariance matrix would be positive semi-definite, the irregular nature of the statistical parameters compounded with the complex method of ensuring positive semi-definiteness makes regression an unnecessarily demanding approach at best.

The classification results shown in Fig. 20 support the validity of the above speculation. Attempting to regress statistical parameters provides very poor classification. These poor results are attributable to the unsteady nature of each statistical parameter's trending over various speeds. When a sufficiently high dimensional space (i.e. AR10) is used, some speed bins generate good results; however, the poor fit of the regression model eventually degrades performance for higher speed bins. Moving into dimensions higher than an AR10 model does not produce better classification results in this context as the primary constraint is the suitability of fit of the regression model. Future work might find fruitful grounds in attempting to fit a piece-wise linear model of these components (since they tend to be very jagged) or employing more complex regression techniques.

5.2. Interpolation

As Worden discovered the better approach, by far, is to store each mean vector and each covariance matrix for each modal bin and to interpolate between each of these statistical parameters for bins that are ill conditioned or under sampled. Interpolated data will not result in derived covariance matrices lacking semi-positive definiteness. The classification results are also very agreeable as shown in Fig. 21 and Table 1; this can be attributed to the faulted and healthy data being nearly completely separable (as demonstrated in Figs. 13–15). While this is likely more a product of using a simulator rather than real industrial equipment, the results are very encouraging.

Table 1 emphasizes the tradeoff between a high acceptance rate of the healthy data and a low acceptance rate of the faulted data. The greater the amount of healthy data enclosed by the statistical novelty boundary, the greater will be the error on the faulted data.

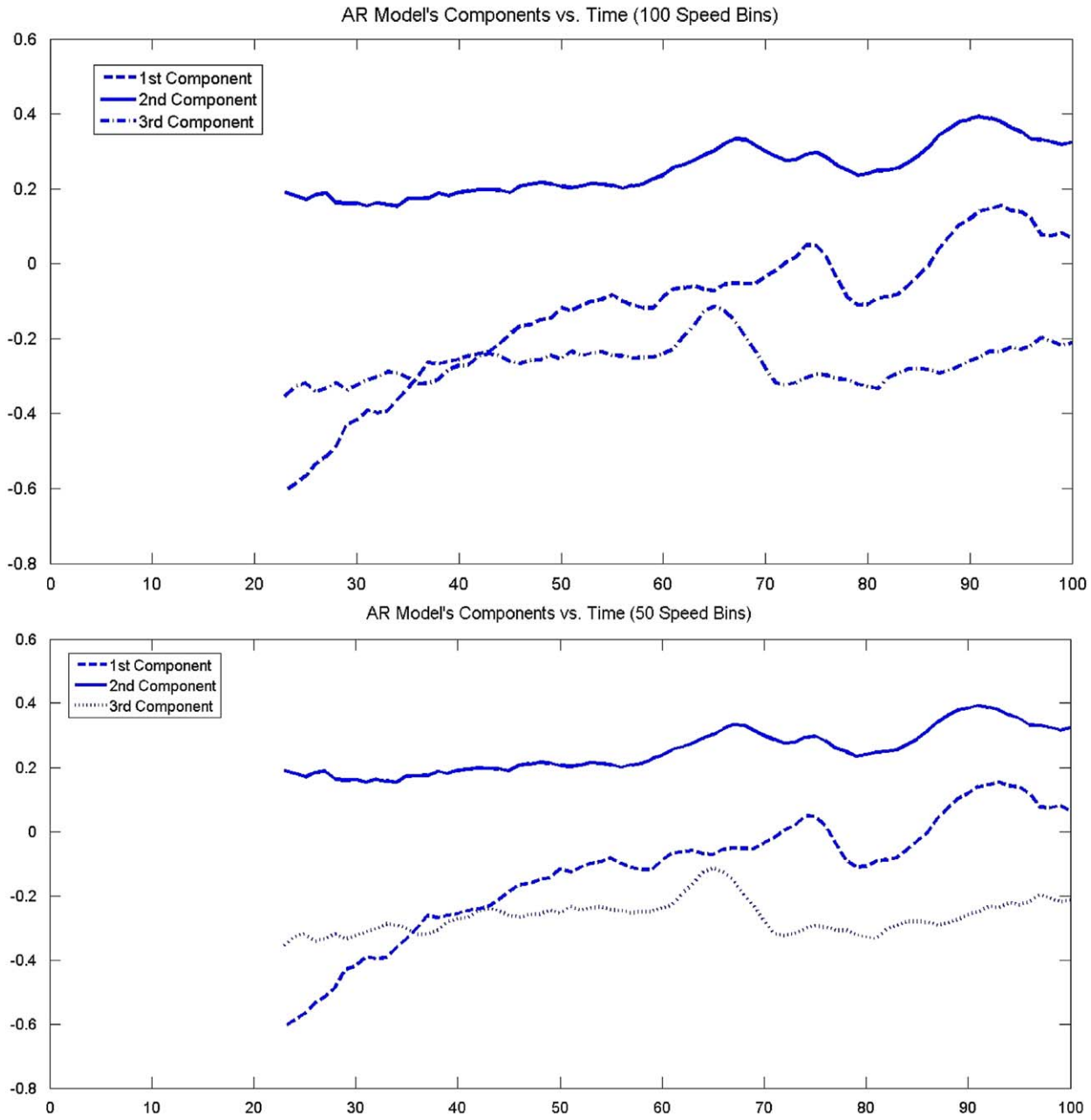


Fig. 17. AR model's components vs. time (100 speed bins—left; 50 speed bins—right).

Fig. 21 demonstrates that an AR model order as low as 3 or 5 could be sufficient for classification purposes. It may be the case that higher order models could provide better classification results; however, the curse of dimensionality has severely restricted the exploration of such models for the fixed amount of data available in these experiments.

Fig. 21 shows the classification results of storing the covariance matrices and mean vectors for each bin and recalling them at the time of classification; it does not demonstrate the effect of interpolating the statistical parameters when they are absent for any given bin. Our previous experience with the irregular nature of the statistical parameters, as shown in Figs. 18 and 19, tends to indicate that interpolation should not occur over too great a range. What speed range could reasonably be interpolated over very much depends on how irregular the changes are in each of the statistical parameters. The safest approach is to interpolate over only as many bins as necessary.

If bin 20 (centered at 42.9 Hz) of the left-hand side of Fig. 21 is omitted, the quality of classification results remains quite good for an interpolated version of it. The true mean of bin 20 is $[-0.2605 \;\; 0.1761 \;\; -0.2600]^t$ with an interpolated

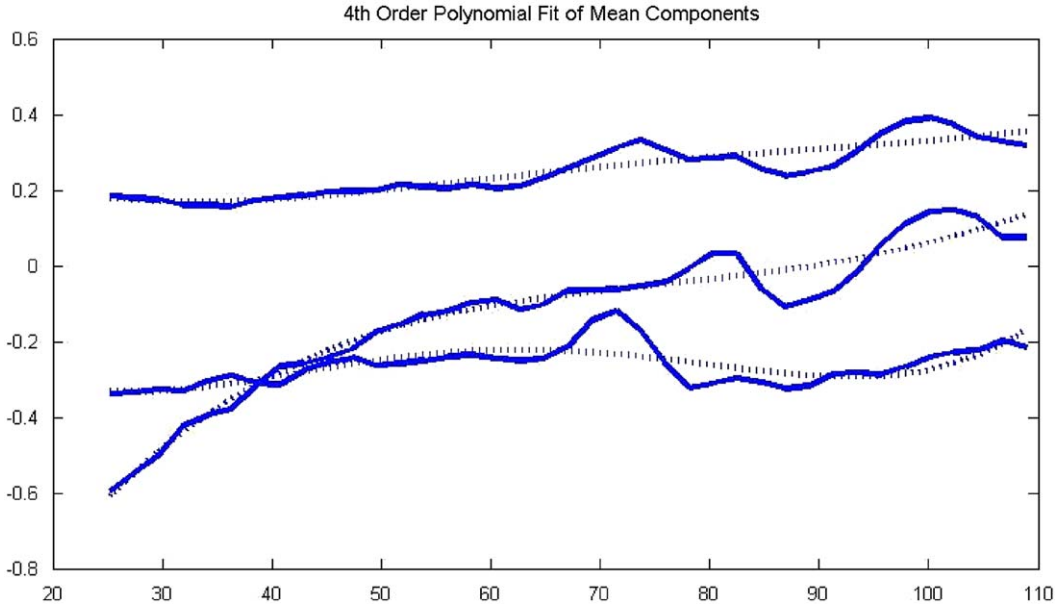


Fig. 18. Polynomial fit of mean components.

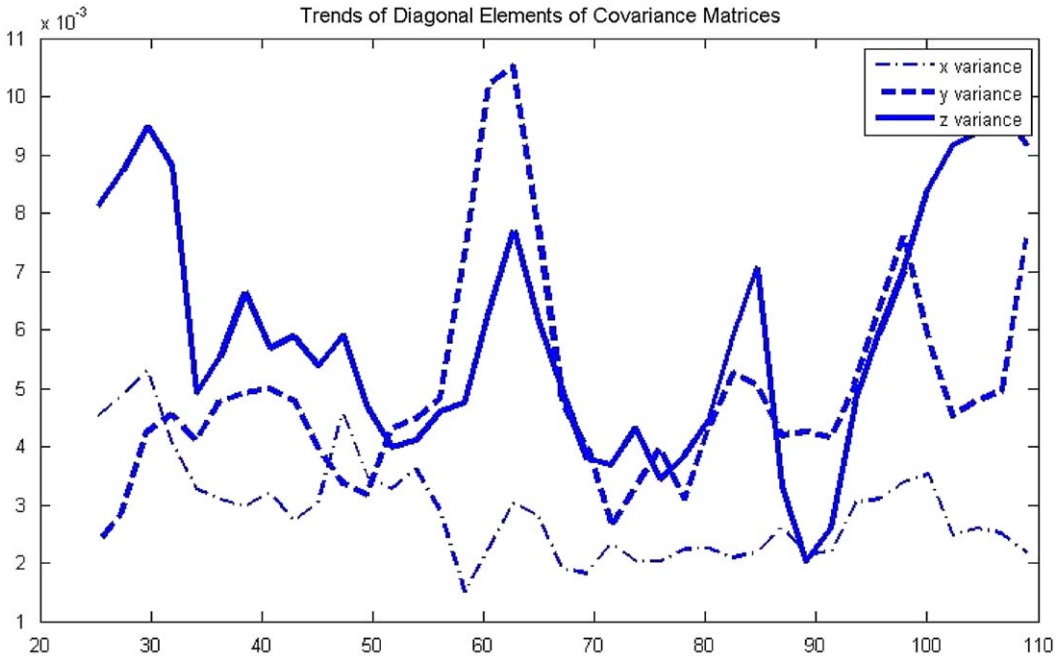


Fig. 19. Tending of variances vs. speed bin.

mean of $[-0.2599\ \ 0.1770\ \ -0.2686]^t$ with true and interpolated covariance matrices given respectively by

$$\sum_{\text{true}} = \begin{bmatrix} 0.0052 & -0.0017 & -0.0033 \\ -0.0017 & 0.0025 & 0.0010 \\ -0.0033 & 0.0010 & 0.0035 \end{bmatrix} \quad \sum_{\text{interp}} = \begin{bmatrix} 0.0051 & -0.0017 & -0.0034 \\ -0.0017 & 0.0027 & 0.0010 \\ -0.0034 & 0.0010 & 0.0037 \end{bmatrix}$$

which are very similar. The difference between the classification error from the known distribution and the interpolated distribution is given in Table 2 for each fault. The error from the interpolated parameters is essentially the same as the error computed from the known distribution.

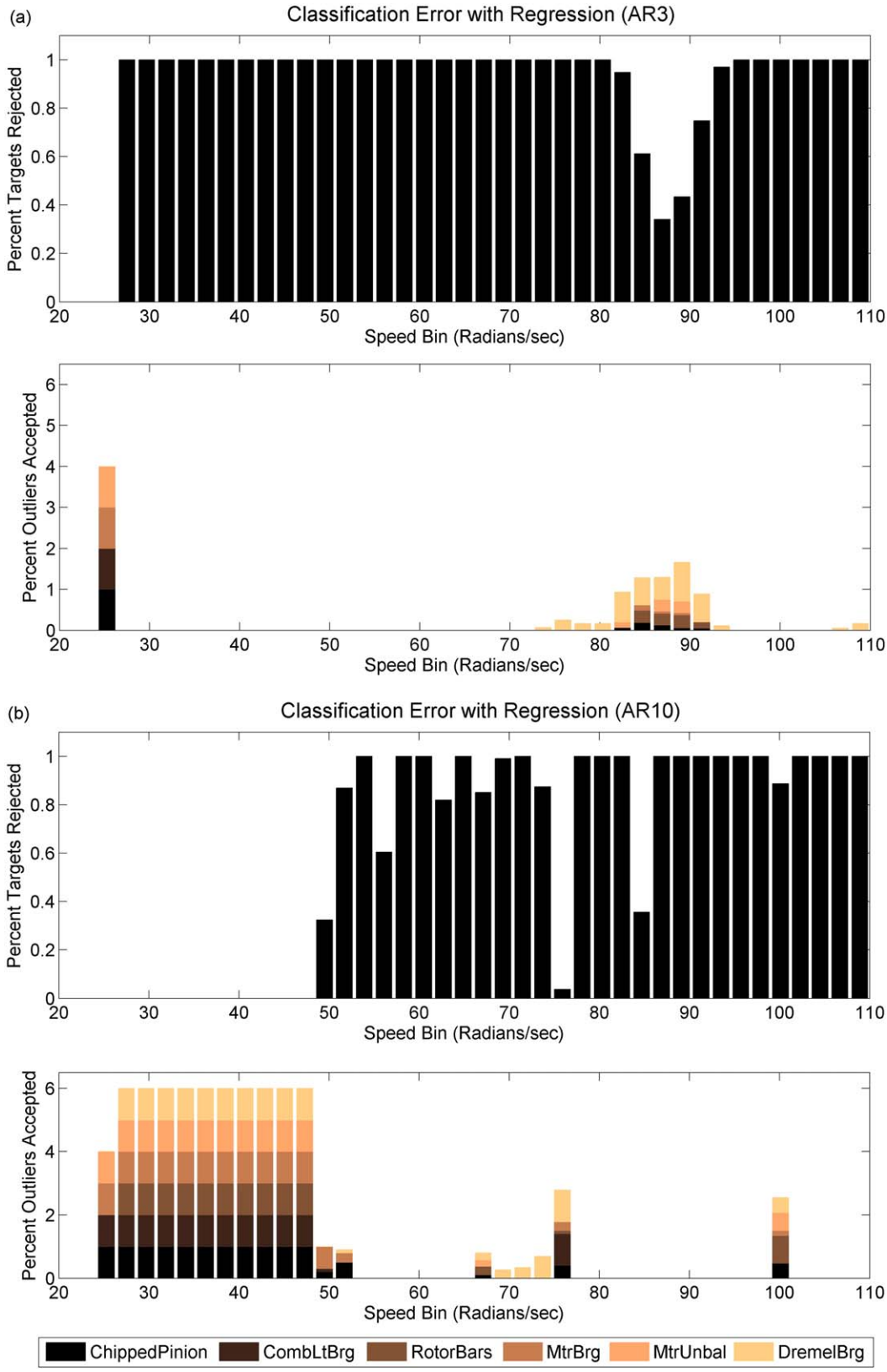
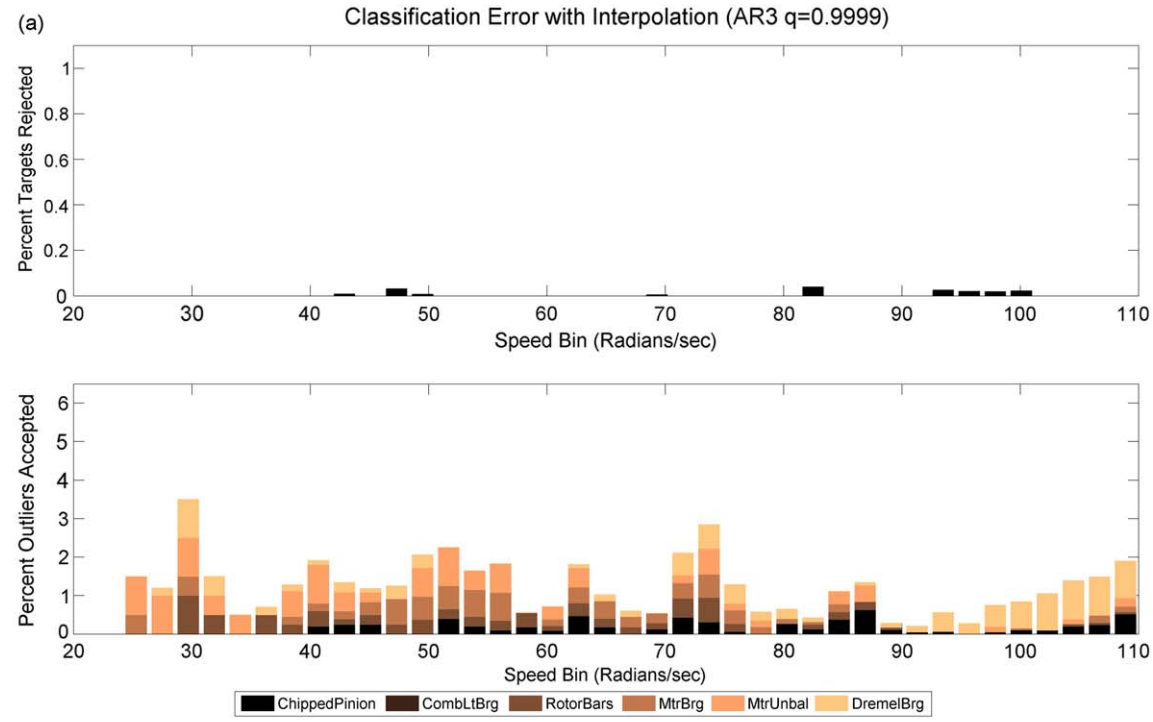


Fig. 20. Classification results with regression.



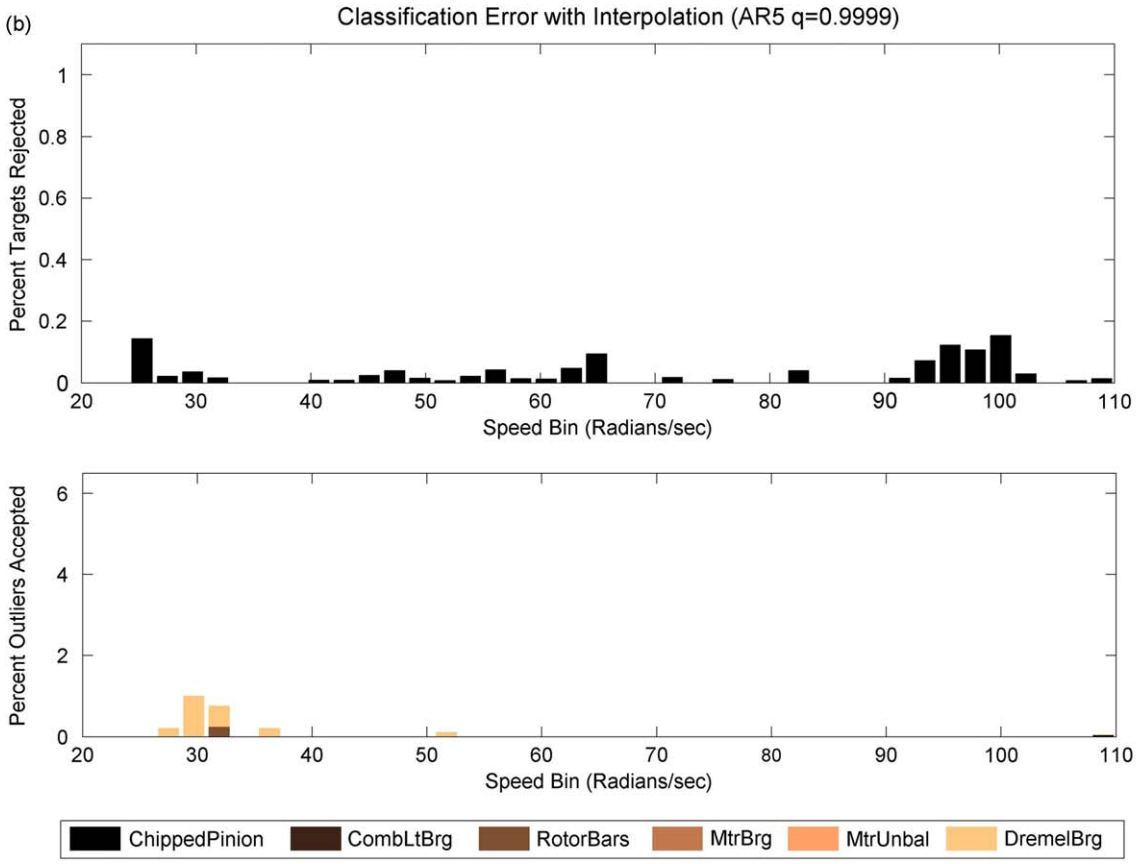


Fig. 21. Classification error with interpolation (no missing bins).

Table 1
Classification errors with interpolation (various quantiles as novelty boundaries).

Fault class	Quantile				
	0.9999	**0.98**	**0.97**	**0.96**	**0.95**
Healthy	0.004	0.095	0.123	0.149	0.175
Chipped pinion	0.156	0.015	0.014	0.009	0.009
CombLtBrg	0	0	0	0	0
Rotor bars	0.203	0.082	0.070	0.055	0.055
Motor bearing	0.233	0.043	0.035	0.030	0.028
Motor unbalance	0.325	0.141	0.119	0.113	0.106
Dremel bearing	0.311	0.133	0.115	0.102	0.096

Table 2
Precision interpolation over 1 bin.

Fault class	Interpolated	Known distribution
Healthy	0	0.009
Chipped pinion	0.250	0.250
CombLtBrg	0	0
Rotor bars	0.142	0.142
Motor bearing	0.200	0.200
Motor unbalance	0.500	0.500
Dremel bearing	0.250	0.250

Table 3
Probability of misclassification of interpolation over 4 missing consecutive bins.

Fault class	Interpolated				Known distribution			
	Bin 19	**Bin 20**	**Bin 21**	**Bin 22**	**Bin 19**	**Bin 20**	**Bin 21**	**Bin 22**
Healthy	0	0.009	0	0	0	0.009	0	0.031
Chipped pinion	0.2	0.25	0.25	0	0.2	0.25	0.25	0
CombLtBrg	0	0	0	0	0	0	0	0
Rotor bars	0.4	0.428	0.375	0.125	0.4	0.142	0.25	0.25
Motor bearing	0.2	0.2	0.555	0.666	0.2	0.2	0.333	0.666
Motor unbalance	1	0.5	0.5	0.333	1	0.5	0.25	0
Dremel bearing	0.333	0.375	0.2	0.333	0.111	0.25	0.1	0.333

Leaving out bins 19 through 22 produces only slightly different results between interpolation of statistical parameters and employing their known distributions (as shown in Table 3). As the number of "missing" consecutive bins increases beyond this small number, the error from interpolation increases to unacceptable levels. These data are collected from the horizontal gearbox accelerometer; the faults related to the gearbox have the same probability of misclassification between the interpolated values and known true distribution—it is the faults on components that are spatially distant from the measured accelerometer which suffer from the worst classification error under this interpolation.

The ability of an accelerometer to classify a fault from a component not spatially near that accelerometer may seem somewhat dubious. However, other acoustic classification contexts (such as passive sonar classification in the maritime environment) demonstrate that it is entirely possible for a transducer to receive vibration data from a distant source having travelled through noisy and disagreeable mediums with differing mechanical impedances. Ultimately, the characteristics of this well-travelled signal may be used in classifying the source of that signal (e.g. an in-board pump, motor, engine, etc. on a distant ship). Comparing such an environment to a machine, rigidly connected together whose components have a high bulk modulus of elasticity, should suggest that these results are at least possible. If true, this may help to explain difficulties experienced by technicians searching for a fault after it has been detected by a condition-monitoring system; if the technician operates under the assumption that the fault will be located on the machinery component onto which the accelerometer is mounted, he or she may be mislead. Future work might examine combining classifiers across different transducers in order to improve classification results and possibly help localize a fault.

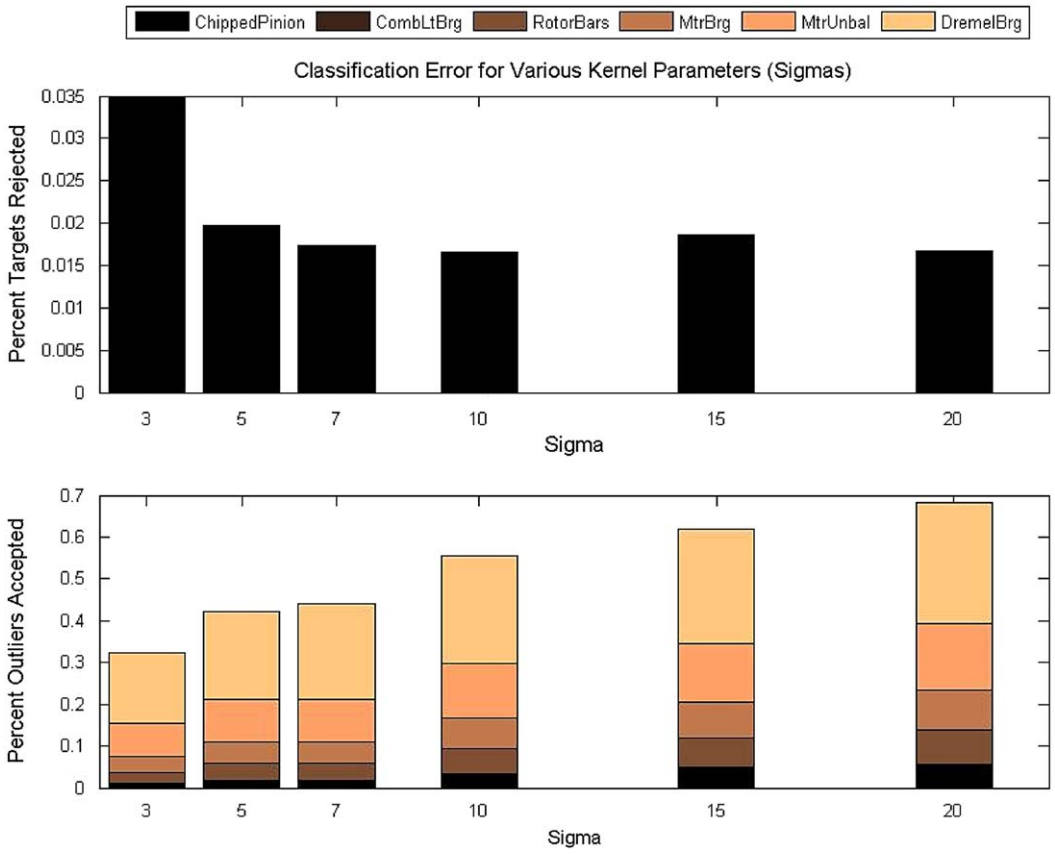


Fig. 22. Classification error for various sigmas.

Table 4
Classification error for various sigmas (from Fig. 21).

Fault	Sigma					
	3	**5**	**7**	**10**	**15**	**20**
Healthy	0.035	0.020	0.017	0.016	0.019	0.017
Chipped pinion	0.011	0.019	0.017	0.032	0.048	0.054
CombLtBrg	0.000	0.000	0.000	0.000	0.000	0.000
Rotor bars	0.028	0.041	0.042	0.061	0.072	0.083
Motor bearing	0.036	0.051	0.051	0.074	0.084	0.098
Motor unbalance	0.081	0.100	0.103	0.132	0.143	0.156
Dremel bearing	0.169	0.210	0.227	0.257	0.272	0.290

5.3. Whitening for support vector descriptor

These procedures can be exploited further by whitening the data for each speed bin such that support vectors may be used to describe the novelty boundary of the healthy data set for all speed ranges. Each classification object for each fault in each speed bin is shifted by that speed bin's healthy mean and contracted by the healthy bin's whitening matrix; the effect is to normalize the data for each speed bin such that the healthy data is centered at the origin with unit variance and the fault data is reduced to a position such that it "orbits" the normalized healthy data in a manner commensurate with its former relative positioning. The mean and whitening matrices for each bin must still be stored—although, the novelty-detection approach is reduced to one classifier (rather than a classifier for each speed bin).

The classification results prove to be quite good. Fig. 22 demonstrates the effect of various kernel parameters employed. The tighter the bound around the healthy data (i.e. the smaller the choice of sigma), the greater is the classification error on that data—while the faulted data enjoys a much lower error for such tight boundaries. While this trend is clear, the choice of sigma has only a marginal effect on the classification error of the healthy data; a small sigma will produce only a slightly

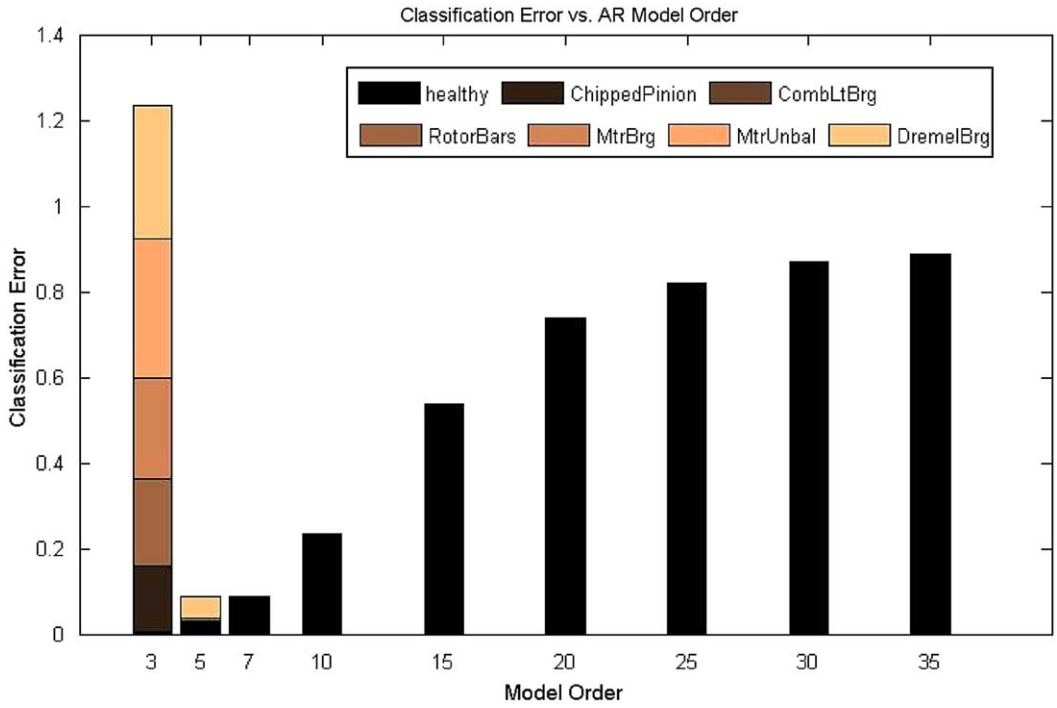


Fig. 23. Average bin classification error vs. AR model order (the effects of the curse of dimensionality).

larger percentage of targets rejected; however, it will also result in an agreeable decrease in the classification error on faulted data.

The classifier remains able to accurately classify faults that are not spatially connected to the machinery component on which the accelerometer is located. In fact, the gear box bearing sensor, whose classification results are displayed in Fig. 22, has poor performance only on the Dremel bearing fault (located on output end of the motor shaft). This quality is more evident in Table 4.

Table 4 compared to Table 1 elucidates the superiority of kernel whitening over the statistical discordance test. In Table 1, it is clear that not only does discordance-test classification produce marginal results, it also demonstrates a very frustrating tradeoff between the quality of classification of healthy and faulted modes; these problems are eliminated with the much more flexible boundary produced by support vector descriptions as shown in Table 4. The previous speculation that the SVDD might be better able to handle the skewed distribution of a non-Gaussian data spread has proven correct; trying to fit an ellipsoid around a class is too rigid a boundary for data whose distribution is not a perfect match for a Gaussian one.

At first glance, the results achieved with this approach may seem too good to be true. For healthy data and outlier data from faults located close to the measured accelerometer, the error is generally less than 2 percent. Separating healthy and faulted data according to the speed of the machine at the time of collection appears to eliminate overlapping classes that usually occur when one trains a classifier over all speed ranges without first conditioning it for speed. In [36], Tax explored training classification error on a variable speed submersible pump with the SVDD and was able to obtain classification error close to 8%. It seems reasonable that attempting to first discriminate classification objects for speed, followed by the application of the SVDD, could produce superior results.

5.4. *Curse of dimensionality*

Fig. 23 demonstrates, for a fixed amount of data, that an increase in the dimensionality of the space results in unacceptable classification errors on the healthy/target data; the quantity of the data appears to be insufficient to describe the appropriate distribution in high dimensional spaces.

6. Conclusions

This research explored the approach outlined by Worden in [24] in the context of variable speed machinery for the detection of incipient faults. The vibratory parameters representative of a machine's vibratory response for various speeds were found to cluster together tightly in an *N*-dimensional space. When these parameters were examined for small

sections of speed, it was found that vibratory parameters representing faults were highly separable from those representing healthy machine states.

Worden's method of determining the means and covariance matrices for small values of modal parameters (i.e. speed bins) and then employing a statistical discordance test to classify healthy data from faulted data was examined. The discordance test consisted of an ellipsoidal novelty boundary containing a certain quantile of healthy data. An improvement to [24] was found in the use of Tax's SVDD on whitened data. The robust boundary description of the SVDD then proved a far superior means of separating healthy and faulted data; it was able to give a tighter fit on the data once centered and normalized—providing better classification results.

Worden's original approach suffered from the assumption of a Gaussian distribution as well as from being doubly cursed by dimensionality. The SVDD may have eliminated some of the ill effects of the Gaussian assumption; however, it is still subject to the curse of dimensionality. As the number of modal parameters increases (e.g. speed, load, temperature, etc.), the approach may break down in scenarios with all but an abundance of data. Future work will eliminate the underlying assumption of Gaussian distributions and will mitigate the effect of the curse of dimensionality. Future work will also involve testing of the presented methods on other datasets to further validate the approach.

Acknowledgments

The authors would like to thank the Center for Excellence in Mining Innovation (CEMI) in Sudbury, ON for their generous financial support of the work. The support of Dr. Chris Mechefske is also much appreciated as it facilitated the rapid generation of the presented results.

References

[1] H. Sohn, K. Worden, C.R. Farrar, Novelty detection in a changing environment: regression and interpolation approaches, *Journal of Sound and Vibration* 258 (4) (2002) 741–761.
[2] J.R. Stack, T.G. Habetler, R.G. Harley, Effects of machine speed on the development and detection of rolling element bearing faults, *IEEE Power Electronics Letters* 1 (2003) 19–21.
[3] R.B. Randall, J. Antoni, S. Chobsaard, The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals, *Mechanical Systems and Signal Processing* 15 (2001) 945–962.
[4] F.K. Choy, S. Huang, J.J. Zakrajsek, R.F. Handschuh, D.P. Townsend, Vibration signature analysis of a faulted gear transmission system, *Journal of Propulsion and Power* 12 (1996) 289–295.
[5] Y.Y. Ivanov, G. Meltzer, Fault detection in gear drives with non-stationary rotational speed—part I: the time-frequency approach, *Mechanical Systems and Signal Processing* 17 (2003) 1033–1047.
[6] Y.Y. Ivanov, G. Meltzer, Fault detection in gear drives with non-stationary rotational speed—part II: the time-frequency approach, *Mechanical Systems and Signal Processing* 17 (2003) 273–283.
[7] C. Kar, A.R. Mohanty, Vibration and current transient monitoring for gearbox fault detection using multiresolution Fourier transform, *Journal of Sound and Vibration* 311 (2008) 109–132.
[8] W. Yang, P.J. Tavner, Empirical mode decomposition an adaptive approach for interpreting shaft vibratory signals of large rotating machinery, *Journal of Sound and Vibration* 321 (2009) 1144–1170.
[9] P. Saavedra, C.J. Vicuna, Online monitoring system for mining shovels focused on vibration analysis, *Canadian Institute of Mining* 2 (2007) 101.
[10] P. Charles, J.K. Sinha, F. Gu, L. Lidstone, A.D. Ball, Detecting the crankshaft torsional vibration of diesel engines for combustion related diagnosis, *Journal of Sound and Vibration* 321 (2009) 1171–1185.
[11] N. Baydar, A. Ball, Detection of gear deterioration under varying load conditions by using the instantaneous power spectrum, *Mechanical Systems and Signal Processing* 14 (2000) 907–921.
[12] C.K. Mechefske, L. Liu, Fault detection and diagnosis in variable speed machines, *International Journal of Condition Monitoring and Diagnostic Engineering Management* 5 (2002) 29–40.
[13] N. Roy, R. Ganguli, Filter design using radial basis function neural network and genetic algorithm for improved operational health monitoring, *Applied Soft Computing Journal* 6 (2006) 154–169.
[14] N. Roy, R. Ganguli, Helicopter rotor blade frequency evolution with damage growth and signal processing, *Journal of Sound and Vibration* 283 (2005) 821–851.
[15] K. Worden, C. Surace, Some aspects of novelty detection methods, *Proceedings of the Third International Conference on Modern Practice in Stress and Vibration*, Dublin (1997) 89.
[16] M.A. Timusk, M.G. Lipsett, C.K. Mechefske, Automated duty cycle classification for online monitoring systems, *2007 Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, DETC2007, v 1 PART A, 2008, pp. 563–573.
[17] M. Timusk, M. Lipsett, C.K. Mechefske, Fault detection using transient machine signals, *Mechanical Systems and Signal Processing* 22 (2008) 1724.
[18] M.A. Timusk, A Unified Method for Anomaly Detection in Unsteady Systems, Ph.D. Thesis, University of Queens, 2006.
[19] M.A. Timusk, C.K. Mechefske, M. Lipsett, A pragmatic framework for on-line neural network based detection of machinery, *International Journal of Condition Monitoring and Diagnostic Engineering Management* 8 (2005) 35–41.
[20] Y.Q. Ni, J.M. Ko, X.G. Hua, R. Zimroz, Variability of measured modal frequencies of a cable-stayed bridge under different wind conditions, *Smart Structures and Systems* 3 (2007) 341.
[21] H. Sohn, M. Dzwonczyk, E.G. Straser, A.S. Kiremidjian, K.H. Law, T. Meng, An experimental study of temperature effect on modal parameters of the Alamosa Canyon Bridge, *Earthquake Engineering and Structural Dynamics* 28 (1999) 879–897.
[22] X.G. Hua, Y.Q. Ni, J.M. Ko, K.Y. Wong, Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique, *Journal of Computing and Civil Engineering* 21 (2007) 122–135.
[23] C.J. Stander, P.S. Heyns, W. Schoombie, Using vibration monitoring for local fault detection on gears operating under fluctuating load conditions, *Mechanical Systems and Signal Processing* 16 (2002) 1005–1024.
[24] K. Worden, H. Sohn, C.R. Farrar, Novelty detection in a changing environment: regression and interpolation approaches, *Journal of Sound and Vibration* 258 (2002) 741–761.
[25] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, NY, 2001.
[26] M. Markou, S. Singh, Novelty detection: a review—part 1: statistical approaches, *Signal Processing* 83 (2003) 2481–2497.

[27] Y.Q. Ni, X.G. Hua, K.Q. Fan, J.M. Ko, Correlating modal properties with temperature using long-term monitoring data and support vector machine technique, *Engineering Structures* 27 (2005) 1762–1773.
[28] R.B. Randall, New method of modeling gear faults, *American Society of Mechanical Engineering Papers* (1981) 9.
[29] R.B. Randall, State of the art in monitoring rotating machinery—part 1, *Journal of Sound and Vibration* 38 (2004) 14–21.
[30] P.Z. Peebles, *Probability Random Variables and Random Signal Principles*, fourth ed., McGraw-Hill Higher Education, New York, NY, 2001.
[31] K. Worden, G. Manson, N.R.J. Fieller, S. Curtis, Damage detection using multivariate statistics. Part I: outlier analysis, *Proceedings of the 23rd International Conference on Noise and Vibration Engineering*, *ISMA* (1998) 189.
[32] P. Filzmoser, R.G. Garrett, C. Reimann, Multivariate outlier detection in exploration geochemistry, *Computers and Geosciences* 31 (2005) 579.
[33] D.M.J. Tax, R.P.W. Duin, Support vector domain description, *Pattern Recognition Letters* 20 (1999) 1191–1199.
[34] R.C. Eisenmann, *Machinery Malfunction Diagnosis and Correction: Vibration Analysis and Troubleshooting for the Process Industries*, Prentice Hall PTR, Upper Saddle River, New Jersey, 1998.
[35] J.C. Principe, N.R. Euliano, W.C. Lefebvre, *Neural and Adaptive Systems*, Wiley, Toronto, ON, 2000.
[36] D.M.J. Tax, One-class Classification: Concept Learning in the Absence of Counter-examples, Ph.D. Thesis, University of Delft of Technology, 2001.