A Low Frequency Uni-variate Model for the Effective Diagnosis and Prognosis of Bearing Signals Based Upon High Frequency Data

Jamie L. Godwin[1], Peter Matthews[1]

[1] *School of engineering and computing sciences, Science Laboratories, Durham University, South Road, Durham, DH1 3LE*

j.l.godwin@durham.ac.uk
p.c.matthews@durham.ac.uk

Abstract

Prognosis of rotating machinery is of vital importance to ensure ever increasing demands of availability, reduced maintenance expenditure and increased useful life are met. However, the prognosis of bearings typically employs techniques in the frequency or time-frequency domain due to the high frequency nature of the data involved (typically >20 KHz). This data quickly becomes unmanageable in practice and often has inferior prognostic horizons in comparison to those techniques which are based upon low frequency data analysis.

This paper presents a novel methodology based upon the computation of the deviation from the empirically derived cumulative density function (CDF) of bearing data. For this purpose, the non-parametric, two sample, uni-variate Kolmogorov-Smirnov test is employed for the analysis. In particular, this paper focuses on mitigating the requirement of a-priori knowledge for bearing prognosis.

Initially, assumptions regarding the underlying structure of high frequency bearing data are explored on publically available data, and found to deviate from what would be expected.

Exploiting this, we use the non-parametric two-sample uni-variate Kolmogorov-Smirnov test to define normal operational behaviour, whilst mitigating the requirement for a-priori knowledge. This reduces the computational complexity of the system whilst having the prospect to reduce the inherent noise within the high frequency bearing signal.

Strong trends of degradation which can be used to derive prognostic maintenance conditions are observed, with sound statistical analysis performed. In particular, statistically significant degradation is found to occur 75 hours before failure occurred (representing identification at 54.2% of bearing life). Both the Kolmogorov-Smirnov D statistic and p -value are employed as health metrics to which degradation can be inferred from. A series of 4 experiments is presented, showing the versatility of the described technique and cases where the technique cannot be employed.

The technique is validated on a failed bearing and then verified on an independent, healthy bearing, and is shown to correctly identify the bearing of question in each case, enabling the prioritisation of maintenance actions which can be used to assist in reducing overall maintenance expenditure.

1. Introduction

With the continually reducing cost of data storage and acquisition, prognosis of critical assets is cheaper than ever. However, the effective exploitation of all this data is not trivial. With more data comes more noise, more conflicting signals, the need for new analytical techniques and the ability to process this data in real time.

As an example, storing data sampled at 20 KHz (20,480 samples per second) requires 13.5GB of data per day, equating to almost 2 billion data points. This makes the identification of degradation within the data difficult, both in automated analysis and also for human operators who can be overloaded by the quantity of data.

Although large quantities of data are collected for analysis, only a subset of this data refers to degraded or failed conditions; in some instances, even for common fault modes, less than 0.1% of the collected data can be used in analysis (Verma & Kusiak, 2011). As such, the use of cutting edge data-mining techniques for these issues is limited. However, this can be exploited through the use of statistical techniques to exploit the known normal behaviour of the data which has been collected.

Data has been identified as a key enabler of next generation maintenance methodologies - such as E-Maintenance

Jamie. L. Godwin et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

(Levrat et al., 2008) - due to the benefit of 5 key points (Hameed et al., 2009):

1. The ability to avoid premature breakdowns
2. Reducing the cost of maintenance
3. Enabling remote diagnosis
4. Increasing production through effective maintenance scheduling
5. Design refinement due to better quality analysis

In this work, a robust uni-variate model for the effective diagnosis and prognosis of bearings is presented. Publically available data collected by the IMS centre and made available by NASA (Lee et al., 2007) is employed to derive a sound statistical time based feature which can be used to determine asset condition. By exploiting normal operational behaviour characterised by the distribution of high frequency data, deviation from expected behaviour can be identified by empirical analysis of the cumulative density function (CDF) of the data. For this purpose, the non-parametric uni-variate Kolmogorov-Smirnov test is used to quantify the deviation from the known behaviour state to the degraded state, whilst quantifying statistically the likelihood of degradation being present.

This overcomes the current limitations of statistical pattern recognition techniques employed in prognostics and health management by empirically defining the CDF and measuring deviations from this. This allows for non-normally distributed data to be effectively analysed without the necessity to ―pre-whiten" data or use one-way statistical transforms on the data.

The paper is organised as follows. Section 1 has introduced the motivation for this research, with Section 2 discussing the related literature. The dataset employed is described in Section 3. Following this, the analytical model is presented in Section 4, with experimental design in Section 5. Results are presented in Section 6 with discussions and conclusions following in Section 7and 8 respectively.

2. RELATED WORK

As previously stated, data-mining techniques are often ineffective in practice due to the large bias in favour of the majority class – typically normal operational behaviour – which reduces the incentive for machine learning algorithms to truly encapsulate failure behaviour. This occurs as in a dataset with 0.1% failure data, the system can achieve a classification accuracy of 99.9% by merely returning the default case (Godwin & Matthews, 2014).

Many algorithms have been proposed to remove the inherent bias in unbalanced datasets (such as in the realm of prognosis). These fall into two main categories, namely under-sampling and over-sampling. Under-sampling removes data from the majority class to remove the bias, whereas over-sampling adds data to the minority class. As such, these techniques will often either reduce the information content in the data, or create synthetic data which needs to be validated and verified. For a full review of data balancing techniques, please refer to Baydar et al., 2001.

It should be noted that these techniques often require labelled data (Baydar et at., 2001). In practice, this is often not available (as failures are yet to occur), or it is too costly to manually label high frequency data. As such, analysis of high frequency data should be performed by statistical techniques which can exploit the high frequency nature of the data to increase the statistical power of the results.

High frequency data is often employed for bearing prognosis due to the ability to extract time, time-frequency and frequency domain features. This enables the use of many different techniques to assist in the diagnostic and prognostic process.

Amongst the most commonly used techniques for bearing diagnosis and prognosis is that of the fast Fourier transform (FFT) (Rai & Mohanty, 2007). This is a frequency domain signal that can be used to detect degradation and identify failure modes. Work done by (Zappalà et al., 2013) uses sideband analysis of key harmonic frequencies in order to monitor the degradation of components over time. As sideband analysis utilises specific harmonic frequencies, the relationship between the harmonic and the immediate sideband frequencies can be analysed as degradation occurs. As such, the technique can be applied where traditional frequency domain techniques are not as powerful (such as in non-stationary signal analysis), for instance, in wind turbine gearbox analysis (Zappalà et al., 2012).

Various other techniques for frequency domain analysis have been explored for rotating machinery such as gearboxes and bearings. Typically, these involve the use of the power spectrum (Ho & Randall, 2000) or Cepstrum analysis (van der Merwe & Hoffman, 2002).

The most commonly utilised domain for frequency analysis is that of the time-frequency domain. Within this, the use of the wavelet transform (Raffiee et al., 2010) is prevalent. Due to the ability to combine frequency domain information in conjunction with time domain data (Raffiee et al., 2010), many strong prognostic signatures can be identified in these techniques.

The wavelet transform is employed due to its ability to remove noise from the data. As various wavelet functions exist (known as mother wavelets), different signatures and artefacts from high frequency data can be discovered and used for diagnostic and prognostic analysis (Lin & Zuo (2003), Peng & Chu (2004), Jardine et al., 2006).

Recently, the use of time synchronous averaging (TSA) has become more prevalent in the literature for prognosis of high frequency data such as bearings and gearboxes (Bechhoefer et al., 2013). This technique is a hybrid time-frequency technique which employs a tachometer in order to deduce the current orientation of the rotating component. This enables further information to be gathered in the prognostic process, such as the identification of specific bearing roller elements which have degraded or if a specific gear tooth has degradation. Derivations of TSA exist which do not require a tachometer (Bechhoefer et al., 2009); however, these often simply estimate the tachometer signal. For a review of TSA techniques as applied to health assessment, please refer to the extensive review undertaken by (Bechhoefer et al., 2009).

Within the time-domain, often statistical features are extracted from the signal. Commonly in the literature, skewness and kurtosis are employed for diagnosis and prognosis (Heng & Nor, 1998 and Tandon, 1994). Skewness is the third standardised moment and represents the asymmetry of an underlying distribution, whereas Kurtosis is the fourth standardised moment and represents the peaked-ness of the underlying distribution.

In practice, due to the high frequency of the data, it is often assumed that the data is normally distributed due to the central limit theorem. As the behaviour of the normal distribution is well understood, we can exploit a-priori knowledge for prognosis. Typically, for a healthy bearing or gear, little to no skewness will exist in the data, and the peaked-ness of the data will typically be 3. However, these features are not reliable for a variety of reasons. When used in uni-variate models, it is possible for the underlying distribution of the data to change due to factors such as degradation, without effecting the skewness and kurtosis of the distribution. As such, the use of these features without additional context (additional features, a-priori knowledge or otherwise) should be avoided.

It should also be noted that typically accelerometer data is employed for analysis in all three commonly used domains. However, the use of acoustic emission (AE) sensor data is becoming more widespread due to potentially increased sensitivity (Bechhoefer et al., 2009) in a variety of methods.

Other time domain features can be used for diagnosis and prognosis. Amongst the most reliable time domain feature is that of oil analysis through the use of oil debris monitoring systems (Feng et al., 2012). These systems are able to monitor the particulate level in parts per million (PPM) in the oil of an asset in order to infer information regarding degradation or potential future failure modes (Feng et al., 2012). These systems are used extensively within the wind industry for monitoring of the gearbox, which is of critical importance (Stephens, 1974). However, these sensors are currently prohibitively expensive for practical use in non-mission-critical scenarios.

As the use of skewness and kurtosis requires making assumptions regarding the underlying distribution of the data, and may not accurately reflect the true change in condition, new techniques are needed. A robust uni-variate nonparametric approach to mitigate these issues can be derived by employing empirical statistical techniques. To demonstrate this, publically available data is employed.

3. DATASET DESCRIPTION

For the following series of experiments, publically available data was employed for transparency. The data was collected by the centre for intelligent maintenance systems (IMS), with the support of the Rexnord Corporation, and made available by NASA (Lee et al., 2007).

Four bearings (force lubricated) were installed onto a shaft which was kept at a constant 2000 RPM by an AC motor. A 6000 lbs radial load was applied via a spring mechanism to the shaft. Rexnord ZA-2115 double row bearings were used, with data collection performed by a National Instruments DAQ 6062E. The accelerometers used in the experiment were PCB 353B33 High Sensitivity Quartz ICP accelerometers. Data was sampled at 20 KHz, equating to 20,480 samples per second. Data was sampled every 10 minutes until oil debris monitoring equipment reached a particulate count which indicated bearing failure. At this point the data collection was deemed complete, and the bearings were removed for inspection. All bearings exceeded their design life expectation. Vibration data pertaining to acceleration was collected during rotational operation, and is measured in G.

4. MODEL DEVELOPMENT

Due to the cases which exist when employing skewness or kurtosis in time series analysis for prognosis, new prognostic features must be developed. In order to ensure that new features do not suffer from the same pitfalls of skewness and kurtosis, 3 factors must be taken into consideration.

Firstly, the technique should be nonparametric. As such, little to no assumptions regarding the underlying data is required. This would enable the technique to work as effectively on normally distributed data as data which is not ordinarily normally distributed, as is often the case in practice for prognostic applications. Secondly, the technique should be robust to noise. Noise is inherent in all real-world signals, and as such, techniques should be robust to this. By identifying data which may potentially be anomalous, this can be disregarded or exploited for further prognosis.

Finally, the technique should accurately respond to changes in the condition of the asset. Skewness and kurtosis have the potential to remain constant whilst degradation occurs. Whilst this may seem trivial, cases such as this should always be checked to ensure that degradation is always observed.

As such, in this work, we propose the use of the two-sample Kolmogorov-Smirnov test (Stephens, 1974) for the diagnosis and prognosis of bearing condition. This is a non-parametric uni-variate technique which can be employed to compare a sample with a given distribution to quantify and signify significant deviations.

The two-sample test statistic quantifies the distance between two cumulative density functions (empirically derived or otherwise). This enables the test statistic to be used as a prognostic health index by fixing one sample to a known state of normal operation behaviour. Thus, it is expected that should degradation occur the distribution of the underlying data will change accordingly. Differing levels of statistical significance can be employed to identify inspection, maintenance and replacement thresholds, with a prognostic time series derived by plotting the changes of the statistic over time.

The Kolmogorov-Smirnov test can be defined as follows (Stephens, 1974):

$$D_{n,n'} = \sup_{\mathrm{x}} |F_{1,n}(x) - F_{2,n'}(x)| \quad (1)$$

Where $\sup_x$ refers to the supremum of set x, and $F_{1,n}$ and $F_{2,n'}$ refer to the empirical distribution function, defined as:

$$\mathrm{F(x)} = \frac{1}{n}\sum_{i=1}^{n} I_{X_i \le x} \quad (2)$$

Where *I* refers to the indicator function, defined as:

$$I_{\mathrm{X_i} \le x} = \begin{cases} 1 \; if \; X_i \le x \\ 0 \; otherwise \end{cases} \quad (3)$$

As such, the test statistic D (as in Eq. 1) represents the maximum difference between the empirically defined distribution $\mathrm{F_1}$ and $\mathrm{F_2}$.

Thus, for a given behaviour, it is possible to accurately measure the deviation from this behaviour and determine its statistical significance. This enables the creation of a health metric as described in the following Section.

5. EXPERIMENTAL SETUP

In order to determine deviations from a known state, a-priori knowledge of the know state must be utilised within the model. Previous work which utilises the Kolmogorov-Smirnov test pre-whitens the data (Cong et al, 2011). Pre-whitening of the data ensures that the data is effectively white noise mixed with the transient signal of the bearing. As such, it is possible to employ a one sample Kolmogorov-Smirnov test for the purposes of bearing degradation assessment by sampling against a Gaussian distribution.

Whilst this removes the need for a-priori knowledge as the effective sample from which degradation is measured, it also infers assumptions regarding the underlying data.

For instance, with regards to the NASA bearing dataset, normality testing was performed via the highly sensitive Anderson-Darling test (Anderson & Darling, 1954). This is a one sample non-parametric test with higher power than the Kolmogorov-Smirnov test, and is computed by:

$$A = -n - \frac{1}{n}\sum_{i=1}^{n}[2i-1][\ln(p_{(i)}) + \ln(1-p_{(n-1+i)}] \quad (4)$$

Where $p_{(i)} = \Phi([x_i - \bar{x})]/s)$ where Φ refers to the CDF of the normal distribution, and $\bar{x}, s$ refer to the mean and standard deviation of the data (respectively).

Within the 2nd set of NASA bearing data, 4 bearings across 984 files were assessed for normality. Of the 3936 normality assessments, 16 samples ($< 0.5\%$) of the bearing data were normally distributed ($p < .05$). As such, given the large sample size (20,480) of each sample, we can infer that the underlying structure of the data is not normal. This is expected; however, as previous work pre-whitens the data, it may be the case that pre-whitening of the data synthetically manipulates the data to ensure normality. Whilst this is effective, it is also computationally intensive, and has the ability to swamp or mask the true bearing signal (Bendre, 1989) and increase noise within the signal.

By replacing the normal distribution reference sample with a known behaviour, we remove the computational intensity, reduce the number of assumptions regarding the underlying data and also reduce the noise within the signal.

In order to explore the use of the Kolmogorov-Smirnov test for the diagnosis and prognosis of bearing faults, three experiments were performed, with an additional experiment utilising the one sample Anderson-Darling test for comparison.

In the first experiment, the Anderson-Darling test is used to quantify the deviation of the data from the normal distribution. This experiment explores the relationship between the normal distribution and the degradation of the bearing. It is expected that as the bearing degrades, the deviation will increase, and can be used to quantify the current level of degradation on the bearing. The second experiment employs the Kolmogorov-Smirnov test without the use of a-priori knowledge. In this case, each data sample is tested against the previous sample to quantify the degradation which has occurred in the previous 10 minutes. Significant degradation of the bearing which occurs between samples are expected to be revealed by this test. The third experiment employs a-priori knowledge to fix a sample point from normal behaviour within a bearing, from which all samples are then measured against. Although this requires the use of a-priori knowledge (in the form of normal operational behaviour), the authors believe this trade off is practical due to normal operational behaviour relating to the majority class. In order to validate the approach, in this experiment, data from a single bearing is employed (2nd test, bearing 1). As this bearing is known to fail, this experiment is intended to prove the Kolmogorov-Smirnov test as a viable time domain feature for diagnosis and prognosis. In the final experiment, data from a healthy bearing is employed as the sample for the Kolmogorov-Smirnov test. This mitigates the practical issues which occur in the third experiment (namely, use of data sampled from a bearing which failed which may not be available in practice) to increase the viability of the approach. As many bearings are subjected to identical conditions (for instance, in a production facility or wind turbine), by utilising known normal behaviour of a single bearing, the approach can systematically be applied to all of the assets in the facility individually.

6. RESULTS

In the first experiment, the Anderson-Darling test is employed as a non-parametric one sample statistical test to measure deviation from the normal distribution. As degradation is expected to cause deviations from this distribution in mean value, standard deviation, skewness, and kurtosis, this test should perform well. However, as can be seen in Figure 1, this is not the case.

Figure 1 (a) presents a healthy bearing and a failed bearing over time (Bearings 1 & 2 from the 2nd set of test data (Lee et al., 2007)) as measured by the p-value of the Anderson-Darling test statistic. Although the healthy bearing line remains stable, the test only identifies a single peak on the failed bearing. Although this is over 46 hours before failure, no progressive trend is observed. As degradation is often an exponential phenomenon, the log plot of Figure 1 (a) is taken and presented in Figure 1(b). This is the natural transformation of exponential data. Although degradation phenomena is observed much earlier due to this transformation (at over 67 hours before failure), there are many inconsistencies with the trend; for instance, degradation seems to decrease and increase over many cycles. Although this does provide insight into the underlying characteristics of the bearing, it violates the prognostic principles metrics must adhere to set out in section 4. The second experiment employs the two-sample non-parametric uni-variate Kolmogorov-Smirnov test to quantify degradation based upon the empirical CDF of the data. Each data sample is compared to the previous collected data sample to determine significance which may imply degradation has occurred.

Figure 2 presents the Kolmogorov-Smirnov D statistic for both the same healthy and failed bearing as in the previous experiment. As can be seen in Figure 2(a), both time series appear to be highly correlated. A Pearson product-moment correlation coefficient was computed to assess the relationship between the healthy bearing, and the failed bearing, and were found to be highly correlated ($r = .97$). It is interesting to note that the peak which has been highlighted in Figure 2(a) is identified in both bearings, and may be due to external factors which occurred during the data collection process. Figure 2(b) presents the log-transform of Figure 2(a). Again, it is difficult to separate the healthy bearing from the failed bearing as no obvious signatures are apparent. Figure 2(c) shows the p-value of the Kolmogorov-Smirnov test for each bearing. It can be seen that this is limited in its use for diagnosis and prognosis, due to many false positives in early life and many false negatives when degradation has occurred. The third experiment exploits these results by fixing the sample to a constant behaviour, from which deviations are then computed. Although this requires a-priori knowledge, this can be taken from OEM documentation. As in this case, it is essential that the fixed points contain no degraded behaviour, the point from which the sample is fixed directly correlates to the quality of the metric which is derived. As such, we exploit historical data in conjunction with OEM documentation and traditional reliability analysis to determine normal behaviour. As each bearing has a design life of 1 million revolutions and the experimental setup ran the bearings at 2000 RPM, we can easily determine from the time elapsed, a percentage of expected useful life. Due to the existence of infant mortality due to manufacturing defects as commonly presented by the so-called ―bathtub curve" (Leemis , 1995) we can then define a point or a set of points which are likely to correspond to normal operational behaviour. For simplicity, data taken from 10-15% of asset life was utilised in this experiment. The first 10% of asset life is not taken into consideration due to the possibility of manufacturing defects or potential infant mortality.

Figure 3 shows the same healthy bearing and same failed bearing when a fixed sample is chosen for the two-sample Kolmogorov-Smirnov test. In practice, we would not retrospectively analyse the first 15% of bearing life, however, for completeness, this has been left in Figure 3. As can be seen in Figure 3(a), for the failed bearing, a strong prognostic signature is detected when employing the D statistic from the Kolmogorov-Smirnov test. Exponential degradation is present, and can be identified as early as 75 hours prior to failure. Initially, a linear trend is found to occur, this is followed by healing phenomena, which afterwards reverts to exponential degradation. Figure 3(b) depicts the logarithmic transform of same experiment, with the artefacts mentioned above highlighted. It should be noted that the same artefacts as in experiment two are observed at the beginning of the time series, which is of interest. The healthy bearing is found to be consistently healthier than the failed bearing, which is promising. Similarly, the D-value remains stable during operation, with exponential degradation occurring at the end of life. This shows the potential of the Kolmogorov-Smirnov test as a prognostic index for bearing health assessment.

The D statistic is employed due to its many features which are complementary for reliability engineering analysis, and

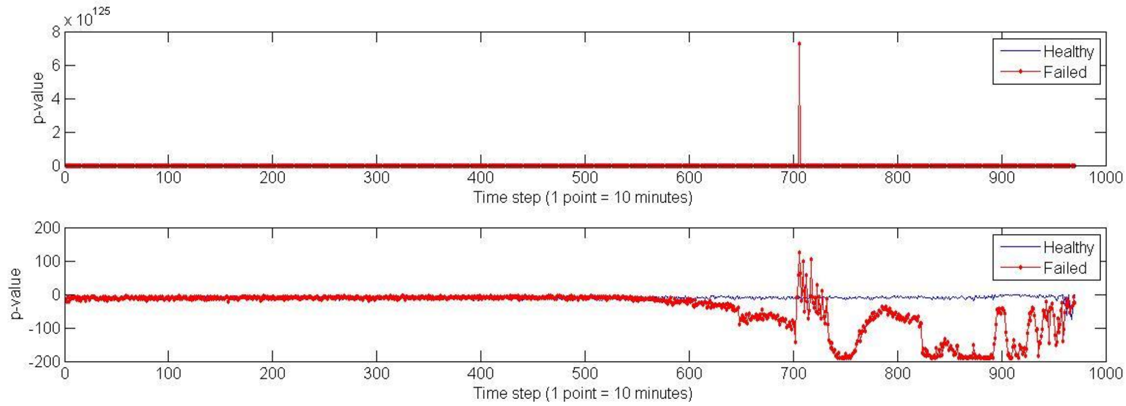

Figure 1. Anderson-Darling test for degradation, showing (a - top) raw values, and (b - below) the logarithmic transform.

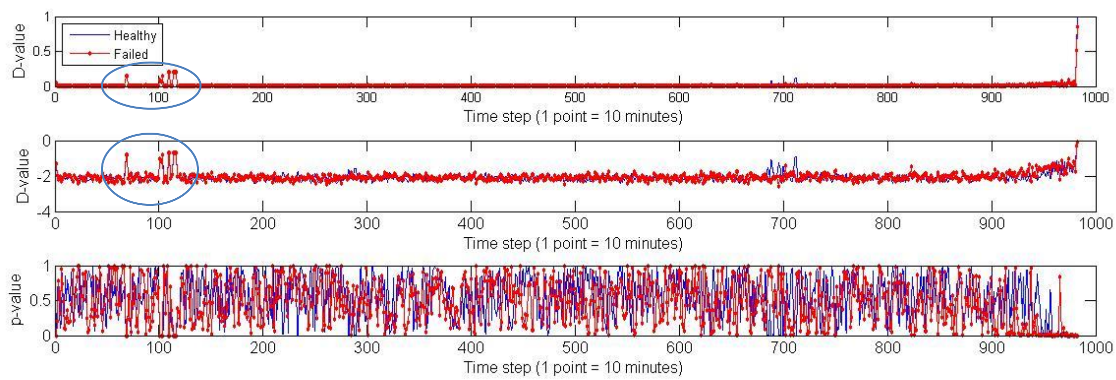

Figure 2. Two sample, transition based, Kolmogorov-Smirrnov showing (a - top) raw D-statistic, (b - centre) the logarithmic transform and (c - below) the associated significance (p-value).

prognostics in general. For instance, the D statistic is bounded between 0 (no difference in the distributions) and 1 (maximum difference in the distributions). As such, it is expected to increases as degradation occurs (as in Figure 3). This bounding also provides a simple means to estimate the percentage of useful life used.

Figure 3(b) shows the log-transform of Figure (A). This then presents the degradation which occurs as a linear phenomenon. This then enables further statistical analysis, such as regression analysis to perform remaining useful life (RUL) estimation for some given condition (D-value). In addition to the D-value being employed, the p-value of the test allows a natural extension of this analysis. If we are to check significant deviations ($p < .05$), the first consistent (repeated 3 times or more) significance is found 73 hours prior to failure, and remains significant until failure (on the failed bearing). For the healthy bearing, consistent significant deviations are found 17 hours prior to the end of the test, which may refer to the initial stages of degradation on the bearing. As such, the use of various p-values can be seen as an effective means for identifying inspection of maintenance activities for decision making within enterprise.

In the final experiment, the fixed sample in the Kolmogorov-Smirnov test was derived as in the previous experiment, however, from an independent bearing which did not fail (Bearing 3, test 2 (Lee et al., 2007)). This experiment explores the versatility and generalisability of the technique. If the bearings are subjected to similar conditions, then normal behaviour of each bearing should be similar. As such, regardless of the bearing used to fix the first sample, the deviation from this should correlate highly to the results achieved in experiment 3. Figure 4 shows the healthy bearing and failed bearing when the fixed sample used for the analysis is from an independent bearing. As expected, this is similar to the results achieved in experiment 3. A Pearson product-moment correlation coefficient was computed to assess the relationship between the D-statistic of the failed bearing taken from experiment 3, and the D-value taken from the failed bearing in experiment 4. These were found to be highly correlated ($r = .86$). Similarly, a further Pearson product-moment correlation coefficient was computed to assess the same relationship for the healthy bearing. This was again found to be highly correlated ($r = .97$). This shows the effectiveness of the technique when applied to new bearings which are expected to operate in similar

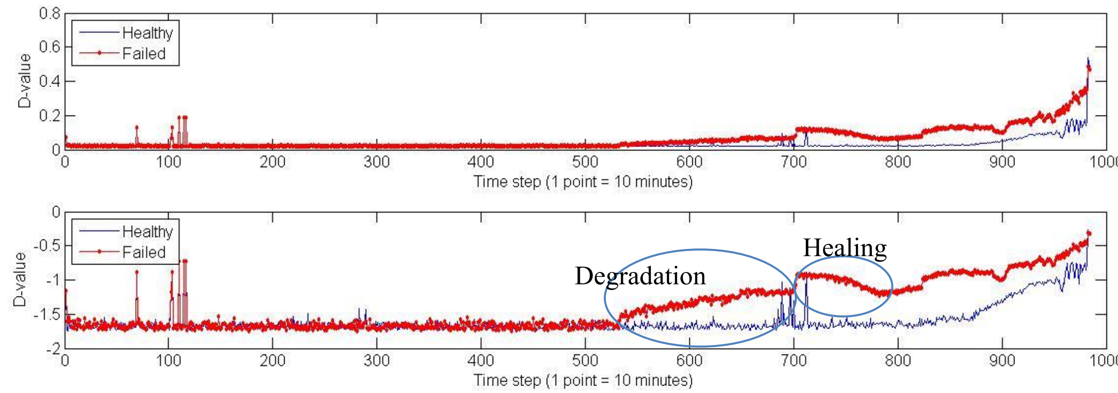

Figure 3. Two-sample, fixed Kolmogorov-Smirnov test, showing (a - top) raw *D*-statistic and (b - below) the log transform.

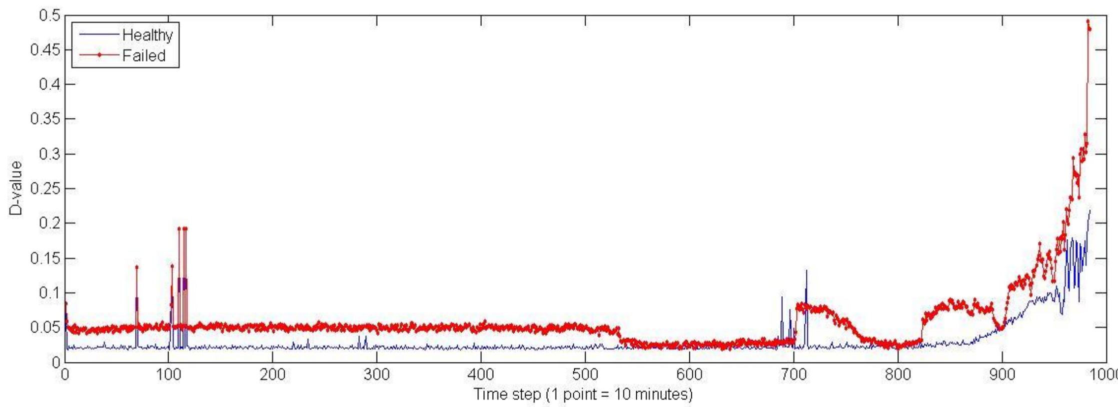

Figure 4. Independent verification of experiment 3 (Figure 3(a)) showing raw *D*-value.

conditions to those which the fixed sample was derived from.

With regards to the significance of the p-values derived from the final experiment in relation to the prognostic horizon, the sensitivity of the technique hinders the benefit gained. As in this case, a 6000 lbs radial load was applied to the shaft, this affects each bearing in a different way. As such, the underlying distributions are inherently different, and thus differ significantly. This then makes each observation appear to be significantly different. However, it is still possible to use the degree of significance as a means for prognosis, as the p-value continues to decrease in proportion to the degradation apparent in the bearing.

7. Discussion

In the first experiment, the Anderson-Darling test was used as a one-sample test in order to mitigate the necessity of a-priori knowledge. However, in this case, the data is not normally distributed and as such, this technique is not effective. In other systems where high frequency data is normally distributed, this may be more sensitive than the Kolmogorov-Smirnov test, and as such, should be used initially.

The Anderson-Darling test is used in the initial analysis over the Shapiro-Wilk test due to the high frequency nature of the data involved. The Shapiro-Wilk test is highly sensitive for large sample sizes, and as such, rejects the null hypothesis often.

As both the Anderson-Darling and Shapiro-Wilk tests are one-sample, they cannot be utilised to empirically derive the CDF of the underlying data, and as such, if the data is not normally distributed, cannot be used to identify deviations specifically from the distribution of the data in question.

It is interesting to note that the artefacts at the start of the time series which can be observed in figures 2 through 4 do not occur in figure 1. This is likely due to the insensitivity of this test due to the underlying distribution of the data. The cause of these artefacts is currently unknown; as similar artefacts are observed throughout both bearings it has been inferred that this is due to the experimental setup and external factors associated with this. The artefacts in figures 2 through 4 for the healthy bearing at approximately time step 700 are unexplained. This could potentially be due to the development of degradation on the failed bearing (from time step 550 as per figure 3) causing particulates in the oil which were transferred to this bearing and ultimately resulted in degradation on the healthy bearing.

The reduction in D-value observed in figure 4 should also be noted. This is an artefact caused by employing a different bearing (with slightly different manufacturer tolerances and defects) in a different bearing position in the experimental setup as a reference. This was undertaken as a proof of concept and I practice, as each bearing will behave in a unique way, historical data pertaining to the bearing in question should be employed.

With regards to fixing the data representing normal behaviour for the two-sample Kolmogorov-Smirnov test, it is essential that no degradation is incorporated into this sample. This is difficult to determine a-priori.

One solution to this would be to use robust outlier analytical techniques to derive a sound subset across the full life of one bearing. As the operational behaviour of the bearing would dictate degradation to be outlying, this would effectively be removed.

In practice, the use accelerometer data is not ideal for robust analysis due to the limited sensitivity of the data collection equipment. If robust techniques such as Median Absolute Deviation (MAD) are used to remove outliers, significant parts of the distribution tails are removed. This limits the effectiveness of the two-sample Kolmogorov-Smirnov test due to the resultant effect on the empirical CDF, which inherently increases the noise within the derived prognostic. The authors recommend not using robust outlier removal in conjunction with accelerometer data, as by their definition, outliers are inherently beneficial for prognosis.

In the case where acoustic emissions (AE) sensors are employed, due to increased sensitivity, the use of robust outlier techniques can potentially be employed effectively.

8. CONCLUSION

This paper has shown the viability of the use the two-sample uni-variate Kolmogorov-Smirnov test as a means to derive low-frequency time-domain prognostic signatures from high frequency data. The versatility of the technique is explored with publically available data (Lee et al., 2007).

Strong prognostic signatures are found for both bearings on which analysis was performed as early as 54.2% of the bearings life (for the failed bearing), and 89.6% of bearing life (for a bearing which ultimately did not fail).

By empirically deriving the CDF function of the data, external conditions are inherently considered and taken into account by the prognostic system. Although this requires a-priori knowledge (historical high frequency data), should this not be available, the empirical function could be approximated by establishing the underlying distribution and using the exact CDF of the chosen distribution.

Although the technique is versatile, it cannot be applied to non-stationary techniques; the transient nature of the signal would almost certainly ensure that statistically significant deviations from the pre-defined normal behaviour are consistently observed whilst no degradation is present: this would violate the prognostic principles laid out previously. For the purposes of this work stationary is defined as a lack of temporal dependency of the marginal distribution (i.e., the distribution of the bearing values does not change with time).

Future work will look to extend this analysis to non-stationary signals for wind turbine gearbox analysis by normalising for loading transitions. The signal can be broken into a series of stationary signals with transient periods which can be identified by correlating the data with the onboard SCADA system.

ACKNOWLEDGEMENTS

This research is funded through an EPSRC Industrial CASE award in collaboration with 5G Technology Ltd

BIOGRAPHIES

Jamie L. Godwin is working towards the degree of Doctor of Philosophy within the department of Engineering and Computing sciences at the University of Durham. He has researched areas such as SCADA data analysis and robust multivariate prognostic techniques. His current doctoral research focuses on metrics for maintenance effectiveness, SCADA data analysis and robust multivariate statistical measures for prognosis.

Peter Matthews is a Lecturer in Design Informatics at the School of Engineering and Computing Sciences. He is the author or co-author of numerous books, technical papers and EU patents. His current research interests are centred around industrial data analysis involving collecting and analysing data obtained either from production process monitoring (e.g., SCADA logs) or service life data (e.g. maintenance logs) and utilising Monte Carlo simulations, Evolutionary Algorithms, Bayesian Belief Networks to form future design and operation decisions.

REFERENCES

Verma, A. and Kusiak A., (2011). *Predictive Analysis of Wind Turbine Faults: A Data Mining Approach*. in Proceedings of the 2011 Industrial Engineering Research Conference, Reno, Nevada, May 19-23,1 – 9.

Levrat, E., Iung, B., Crespo, (2008). E-*maintenance:f review and conceptual framework*. Production Planning & Control 19 (4), 408429.

Hameed, Z., Hong, Y., Cho, Y., Ahn S., and Song, C., (2009). *Condition monitoring and fault detection of wind* *turbines and related algorithms: A review*. Renewable and Sustainable energy reviews, 13(1), 1–39.

Lee, J., Qiu, H., Yu, G., Lin, J., and Rexnord Technical Services (2007). *'Bearing Data Set'*, IMS, University of Cincinnati. NASA Ames Prognostics Data Repository, [http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/], NASA Ames, Moffett Field, CA.

Godwin, J. L., & Matthews, P. C., (2014a) *Robust Statistical Methods for Rapid Data Labelling.* In V. Bhatnagar, "Data Mining and Analysis in the Engineering Field". IGI Global : Hershey, PA, USA.

Baydar, N., Chen, Q., Ball, A. and Kruger, U. (2001). *Detection of incipient tooth defect in helical gears using multivariate statistics*. Mechanical Systems and Signal Processing, 15 (2), 303—321.

Rai, V. K., & Mohanty, A. R. (2007). *Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform*. Mechanical Systems and Signal Processing, 21(6), 2607-2615.

Zappalà, D., Tavner, P., Crabtree, C., & Sheng, S. (2013, January). *Sideband Algorithm for Automatic Wind Turbine Gearbox Fault Detection and Diagnosis*. In European Wind Energy Association (EWEA) Conference, Vienna, Austria.

Zappalà, D., Tavner, P. J., & Crabtree, C. J. (2012). *Gear fault detection automation using WindCon frequency tracking*. In Proceedings European Wind Energy Conference.

Ho, D., Randall, R.B., (2000) *Optimisation of bearing diagnostic techniques using simulated and actual bearing fault signals*, Mechanical Systems and Signal Processing 14 (5) pp. 763–788

N.T. van der Merwe, A.J. Hoffman, (2002). *A modified cepstrum analysis applied to vibrational signals*, in: Proceedings of 14th International Conference on Digital Signal Processing (DSP2002), vol. 2, Santorini, Greece, 2002, pp. 873–876

Raffiee, J., Raffiee, M., Tse, P., 2010. *Application of mother wavelet functions for automatic gear and bearing fault diagnosis*. Expert Systems with Applications 37 (6), 4568 - 4579.

Peng, Z., Chu, F., (2004). *Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography.* Mechanical Systems and Signal Processing 18 (2), 199 - 221.

Lin, Zuo, M., (2003). *Gearbox fault diagnosis using adaptive wavelet filter*. Mechanical Systems and Signal Processing 17 (6), 12591269.

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

Bechhoefer, E., Qu, Y., Zhu, J. and He, D. (2013) *Signal Processing Techniques to Improve an Acoustic Emissions Sensor*. In proceedings of the annual conference of the PHMsociety October 14 – 17, New Orleans, LA, USA. 4(3). pp. 18.

Bechhoefer, E., & Kingsley, M. (2009). *A review of time synchronous average algorithms*. In Annual conference of the prognostics and health management society.

Heng, R.B.W., Nor. M.J.M., (1998) *Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition.* Applied Acoustics, 53 (1-3), pp. 211–226

Tandon, N.. (1994) *Comparison of some vibration parameters for the condition monitoring of rolling element bearings*. Measurement: Journal of the International Measurement Confederation, 12 (3), pp. 285–289

Feng, Y., Qiu, Y., Crabtree, C., Long, H. & Tavner P., (2012) *Monitoring wind turbine gearboxes*. Wind Energy. Vol 16(5). pp. 728 – 740.

Crabtree, C.J. (2010). *Survey of Commercially Available Condition Monitoring Systems for Wind Turbines*, SuperGen Wind.

Stephens, M. A. (1974). *EDF Statistics for Goodness of Fit and Some Comparisons*. Journal of the American Statistical Association (American Statistical Association) 69 (347): 730–737

Cong, F., Chen, J., & Pan, Y. (2011). *Kolmogorov-Smirnov test for rolling bearing performance degradation assessment and prognosis*. Journal of Vibration and Control, 17(9), 1337-1347.

Anderson, T. W., & Darling, D. A. (1954). *A test of goodness of fit*. Journal of the American Statistical Association, 49(268), 765-769.

Bendre, S. M. (1989). *Masking and swamping effects on tests for multiple outliers in normal sample*. Communications in Statistics-Theory and Methods, 18(2), 697-710.

Leemis, L. M. (1995). *Reliability: probabilistic models and statistical methods*. Prentice-Hall, Inc..