Estimation of remaining useful life of ball bearings using data driven methodologies

Conference Paper · June 2012

DOI: 10.1109/ICPHM.2012.6299548

4 authors, including:

Hyunseok Oh
Gwangju Institute of Science and Technology
37 PUBLICATIONS **1,655** CITATIONS
SEE PROFILE

Arvind Vasan
University of Maryland, College Park
18 PUBLICATIONS **759** CITATIONS
SEE PROFILE

Michael Pecht
University of Maryland, College Park
1,583 PUBLICATIONS **53,700** CITATIONS
SEE PROFILE

Estimation of Remaining Useful Life of Ball Bearings using Data Driven Methodologies

Edwin Sutrisno, Hyunseok Oh, Arvind Sai Sarathi Vasan, and Michael Pecht
Center for Advanced Life Cycle Engineering (CALCE)
University of Maryland, College Park
pecht@calce.umd.edu

***Abstract*— This paper describes the three methodologies used by CALCE in their winning entry for the IEEE 2012 PHM Data Challenge competition. An experimental data set from seventeen ball bearings was provided by the FEMTO-ST Institute. The data set consisted of data from six bearings for algorithm training and data from eleven bearings for testing. The authors developed prognostic algorithms based on the data from the training bearings to estimate the remaining useful life of the test bearings. Three methodologies are presented in this paper. Result accuracies of the winning methodology are presented.**

Keywords—PHM Data Challenge, prognostics, RUL, bearings

I. INTRODUCTION

This paper presents the methodologies developed by a team from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland for the IEEE 2012 PHM Data Challenge competition held by the IEEE Reliability Society and the FEMTO-ST Institute. The goal of the competition was to provide the best estimate of remaining useful life of ball bearings under experimental loading conditions. The experimental data set was provided by the FEMTO-ST Institute. The data set consisted of six training sets obtained from run-to-failure experiments and eleven test sets showing truncated experimental data. A detailed description of the experimental setup can be found on the FEMTO-ST Institute's website [1].

The experiments from which the data set was derived involved three different loading conditions. Condition 1 had seven ball bearings operated at 1800 rpm with 4000 N radial load. Complete run-to-failure data for algorithm training were provided for two of the bearings, and truncated data for algorithm testing were provided for the other five bearings. Condition 2 featured seven ball bearings operated at 1650 rpm with 4200 N radial load. Of the seven bearings, complete run-to-failure data for training were provided for two, and data for testing were provided for five of the bearings. Condition 3 featured three bearings operated at 1500 rpm with 5000 N radial load. Data from two of the bearings were provided for training and data from the other bearing was provided for testing.

Two accelerometers were mounted on the bearing housing to measure vibration in the vertical and horizontal directions. Data sampling was conducted at 10 s intervals at a 25.6 kHz sampling rate and 0.1 s duration; hence, each observation contained 2560 points.

Three methodologies were developed to estimate the remaining useful life (RUL) of the 11 test bearings. The feature extraction, classification, and prediction techniques of the methodologies are presented in the following sections. Advantages and challenges of each methodology are briefly discussed in the comparison section.

II. RUL ESTIMATION METHODOLOGIES

A. Method 1: Moving Average Spectral Kurtosis and Bayesian Monte Carlo

Features were extracted from vibration signals of the bearings in the time and frequency domains [1]. The time domain features included root mean square (rms), peak, crest factor, and kurtosis of the vibration signals. The frequency domain features included the magnitude at bearing defect frequencies. In a preliminary analysis of the vibration signals of the six training bearings, we found that neither the time domain features nor the frequency domain feature exhibited a consistent trend of bearing degradation. Therefore, we developed a signal processing method that took advantage of both time and frequency domain features.

The vibration signal from each 0.1 s measurement was bandpass-filtered with a frequency band of 500 Hz. Twenty-four bandpass-filtered vibration signals were extracted. They covered a frequency range of 0–12 kHz of the vibration signals. A moving average filter was applied to the time series kurtosis to identify trends over time. When a vibration signal follows a Gaussian distribution that indicates a healthy condition of the bearing, the kurtosis is 3.0 [2]. A kurtosis value of 3.0 was observed at the beginning of the experiment as seen in Figure 1.

From a visual inspection of the time series kurtosis after applying the moving average filter, an increasing trend could be identified. In order to quantify the applicability of a trend for prognostics, we calculated a correlation coefficient between the kurtosis and time. Based on the assumption that a feature that monotonically increases over time is the ideal degradation signal, Spearman's rank correlation coefficient was used to assess how strong the monotonic relationship was between the kurtosis and the time duration of the experiment. The correlation coefficient for the training bearing 2_2 (i.e., bearing

978-1-4673-0358-3/12/$31.00 ©2012 IEEE

2, condition 2) was 0.829, which implied a strong monotonic relationship between the two variables.

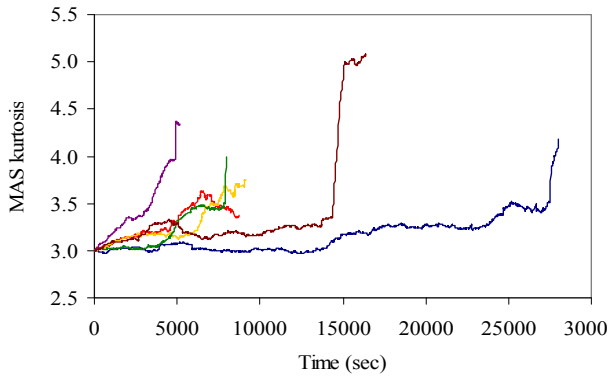

Figure 1. Moving-average spectral kurtosis of six training bearings: frequency range of 5.5 to 6.0 kHz, moving average of 100 data points.

A correlation coefficient was calculated for each of the vibration features: rms, peak, crest factor, and kurtosis. Based on the correlation analysis using the extracted features from the six training bearings, the kurtosis extracted from the bandpass-filtered vibration signals with a frequency range of 5.5 to 6.0 kHz was identified as a best feature for bearing prognostics. The window size of the moving average for the time-series of kurtosis was 100 data points.

A linear or exponential model was used to describe trends in the degradation signals of bearings [3][4]. In order to understand the relevance of using the models for bearing prognostics, a regression analysis was conducted for the moving-averaged spectral (MAS) kurtosis. We compared the trend of the MAS kurtosis of each training bearing with the three different types of degradation signal models: $y=at+b$, $y=a\ exp(bt)$, and $y=a\ exp(bt^2)$, where a and b are the model constants, y is the magnitude of the degradation signal, and t is time. The exponential model, $y=a\ exp(bt^2)$, showed the best fit among the three degradation signal models for the six training bearings. Table I shows the goodness of fit of the exponential model. The R-square of four out of the six bearings exceeded the value of 0.800. However, the R-squares of the other two bearings 1_2 and 3_2 were 0.432 and 0.565, respectively. We anticipated that the exponential model would not provide high prediction accuracy for those two bearings compared to the other four. The results from the regression analysis showed the limitations of using the exponential model for future analyses. More efficient features and/or a more efficient degradation model should be investigated in the future. At the time of competition, the MAS kurtosis was the best feature we were able to extract in terms of its correlation coefficient, and the exponential model was showing the best fit. Therefore, we estimated remaining useful life of the eleven test bearings based on the MAS kurtosis and the exponential model.

TABLE I. GOODNESS OF FIT OF $y=a\ exp(bt^2)$ TO MOVING AVERAGE SPECTRAL KURTOSIS OF SIX TRAINING BEARINGS

	Bearing 1_1	Bearing 1_2	Bearing 2_1	Bearing 2_2	Bearing 3_1	Bearing 3_2
R -square	0.800	0.432	0.879	0.885	0.919	0.565

The constants of the exponential model, a and b, were sequentially updated using a Bayesian Monte Carlo method as new observations of the MAS kurtosis became available. Details about the method can be found in references [5][6][7]. The remaining useful life of each test bearing was estimated by finding the moment when the magnitude from the exponential model exceeded a predefined threshold. In this analysis, the failure threshold was selected to be 4 based on the observation of the kurtosis of the six training bearings shown in Figure 1. A different value could be used for a threshold, since a threshold is typically defined based on engineering judgments, industrial standards, or published studies.

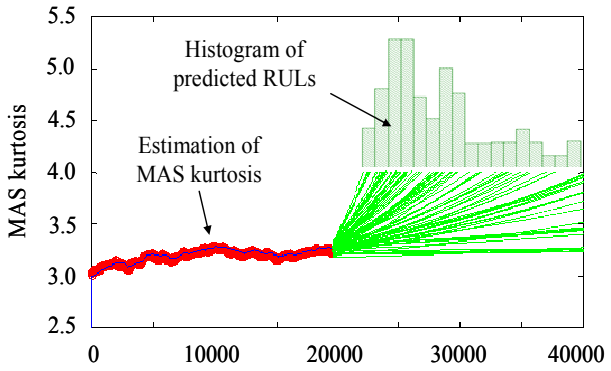

Figure 2. Remaining useful life estimation using MAS kurtosis and the Bayesian Monte Carlo method: test bearing 2_5.

Figure 2 shows a histogram of the failure time predictions of test bearing 2_5. The remaining useful life (RUL) of bearing 2_5 estimated at 20,000 seconds was 3,580 seconds. The actual remaining life of bearing 2_5 was 3090 seconds. The error between the estimated and true remaining life for test bearing 2_5 was -15.9%. The negative value in the error means that life was overestimated. The error was calculated using Equation (1).

$$Error(\%) = \frac{ActualRUL - EstimatedRUL}{ActualRUL} \times 100\% \qquad (1)$$

The MAS kurtosis and Bayesian Monte Carlo method provided distributions of the RUL of test bearings. A conservative approach was taken by choosing the RUL value from the left end of the generated distribution, e.g. 5 percentile.

B. *Method 2: Soft Computing Model with Support Vector Regressor*

This section provides the remaining useful life (RUL) estimation method using a least squares-support vector regressor. Through successive refinement of vibration signals from the bearings, features were extracted for failure prognosis. An overview of the approach is schematically represented in Figure 3.

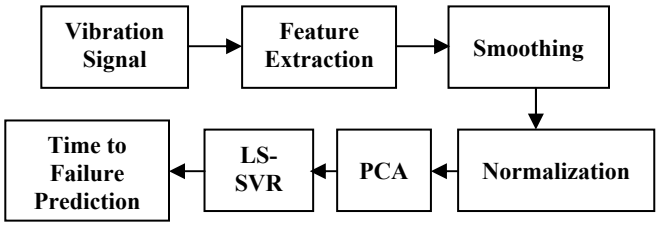

Figure 3. Overview of the prognostic process using a soft computing technique.

The first activity involved feature extraction from the bearing's vibration signal. Seventeen features were extracted from each horizontal and vertical vibration signals. The first five features were extracted using a higher-order crossing method [8]. Ten other features were extracted using a wavelet transformation. Among the ten features, five features corresponded to the energy in the approximate coefficients at the first five levels of decomposition. Similarly, the five features corresponded to the energy in the detailed coefficients at the first five decomposition levels. Additional two features were the cumulative signal energy up to a time instant [9] and the peak of the vibration signal.

Once all the thirty four features (seventeen for horizontal and seventeen for vertical vibration signals) were extracted, they were smoothed using a moving average filter to suppress noise in measurements. Normalization of features was performed after smoothing to bring the features within the same scale, thereby reducing bias due to features of large dynamic range.

The next activity was variable reduction. Large number of features could potentially overfit the data and reduce the overall performance of the soft computing model [10]. In this method, principal components analysis (PCA) was used for variable reduction. PCA reduces a system of *p*-features into *k*-principal components using a linear transformation while maintaining most of the variability from the feature set. PCA was applied to the feature set extracted from the training bearings and the first three principle components accounted for more than 99.5% of the data variability. Hence, we concluded that a reduction in the feature set from thirty four features to three principle components was reasonable. Figure 4 shows an illustrative time-series plot of the first three principal components for bearing 2_2.

For RUL estimation a least squares-support vector regressor (LS-SVR) technique was used [11]. This prognostic method consisted of training and testing modes.

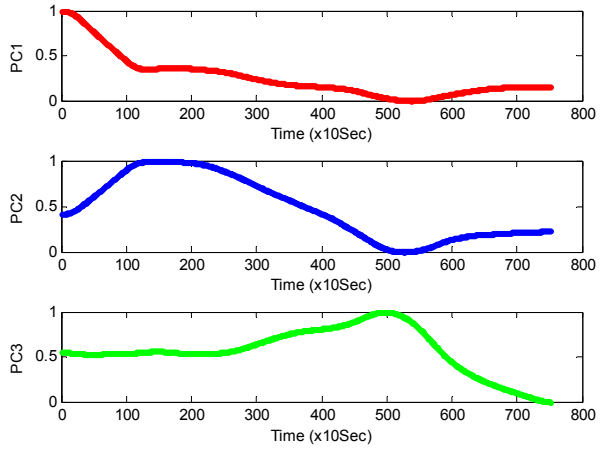

Figure 4. Smoothed and normalized time-series plot of the first 3 principal components for bearing 2_2.

In the training mode of the LS-SVR, the three principal components were given as input and time to failure as target (Figure 5). Before training the LS-SVR, the input data were randomly permuted so the algorithm could learn the underlying function mapping of input states to the desired output in a static way, not dynamic. If learning was conducted in a dynamic fashion, the performance of the LS-SVR would be time dependent, which was undesirable.

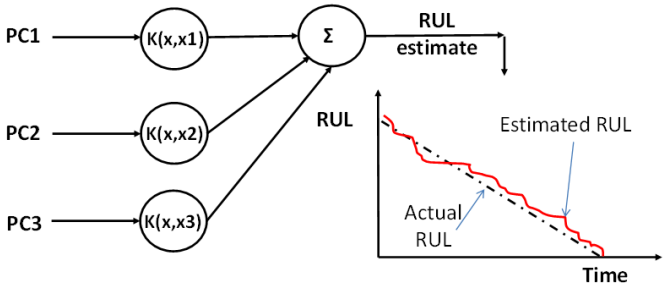

Figure 5. Bearing RUL estimation illustration using LS-SVR.

In the testing mode, test bearings features were extracted from the vibration signals and the corresponding three principal components were computed using the transformation matrix obtained during the training mode. The resulting principal components were given as input to the trained LS-SVR to generate estimates of the remaining useful life.

C. *Method 3: Vibration Frequency Signature Anomaly Detection and Survival Time Ratio*

This section describes a methodology for estimating the remaining useful life (RUL) of the test bearings using an anomaly detection, degradation feature extrapolation, and survival time ratios.

An anomaly was detected when there was a change in the frequency of the peak vibration in the frequency spectrum. The frequency spectrum was generated using the Fast Fourier Transform (FFT) algorithm. An example of an anomaly detection is shown in Figure 6 for bearing 1_3, where a new

range of peak vertical vibration frequencies of around 5300 Hz begins appearing at observation #822. Additional changes in the frequency signature can be observed from the peak horizontal vibration frequencies from observations #822 to #1247, where frequencies that initially existed near 2500 Hz disappeared. Each number of observations on the plot is equivalent to a 10 s duration.

A possible physical explanation for the change in the bearing frequency signature is the initiation of crack, spall, or other surface defects within the bearing that caused the structure to vibrate with new natural frequencies and modes. This hypothesis, however, could not be confirmed because no information on failure analysis was provided in the competition.

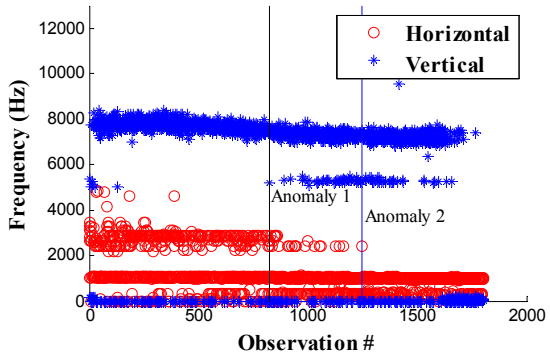

Figure 6. Anomaly detection for bearing 1_3 at observation #822 marked by the vertical line.

Based on the inspection of the training data, we assumed that the bearings exhibited several stages of degradation before reaching failure, as suggested by multiple anomaly detection times. Some bearings showed a degradation stage with gradually increasing features, while others did not show any trend, suddenly experienced a high increase in the features, and then failed shortly after.

The prognostic feature used in tracking degradation was defined as the average of the five highest absolute acceleration values measured in each observation. Averaging was done to reduce the effect of noise. Figure 7 shows an example of how this feature was defined for bearing 1_1 at observation #2000. The highest five points were selected out of the 2560 points of horizontal acceleration. The same approach was used to define the feature for the vertical acceleration data. The equation to define the feature is the following:

$$f = \frac{1}{5}\sum_{i=1}^{5} acc_i \qquad (2)$$

where *acc* is the acceleration and *i* is the index of the data point after being sorted in an ascending order according to the acceleration value.

Plot of the horizontal feature for bearing 1_1 is shown in Figure 8. A gradual increase in the horizontal feature was observed after an anomaly was detected at observation #1218. At this moment the bearing entered the first anomalous region. At observation #2747 the bearing entered a second anomalous region where the feature increased abruptly. The degradation trend of bearing 1_1 was shared by test bearing 1_3. This was the only case where a consistent increasing feature was found in a test bearing (Figure 10).

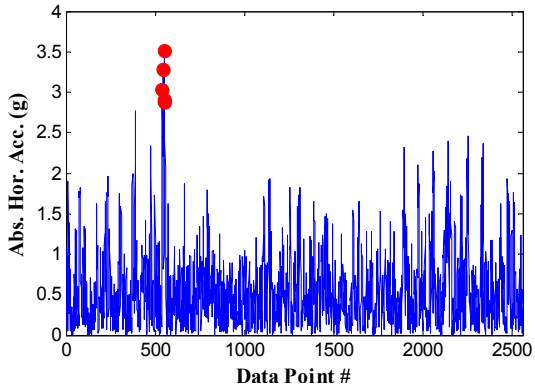

Figure 7. Absolute horizontal accelerations for bearing 1_1 at observation #2000. Plot shows 2560 data points with circles marking the 5 largest values.

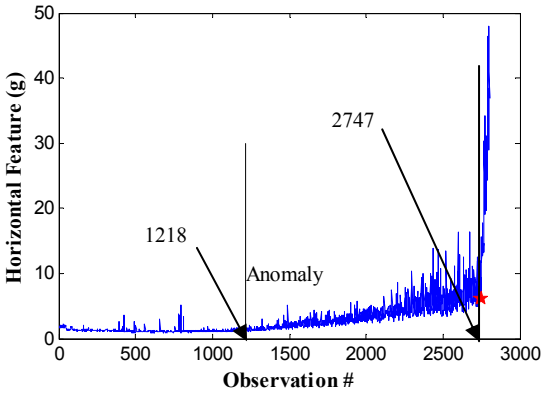

Figure 8. Plot of horizontal feature for bearing 1_1. Anomaly detection was marked by vertical line. Star marks the onset of a second anomalous region where the feature abruptly increased.

To predict failure in test bearing 1_3, a two-step approach was taken. The first step was to estimate the time when the bearing entered a second anomalous region. The second step was to estimate how long the second anomalous region lasted before the bearing failed. For the trend shown in Figure 8, an exponential curve fit was taken from the first anomaly detection time until the end of the first anomalous region. The fitted value at the end of the first anomaly divided by the value at anomaly detection gave 5.47 (Figure 9). The time for bearing 1_1 to fail after entering the second anomalous region was 560 s. The duration of the second anomaly divided by the first anomaly gave the anomaly duration ratio:

$$AnomalyRatio = \frac{560s}{15290s} = 0.0366 \quad \text{for Bearing 1\_1}$$

The plot of the horizontal feature for bearing 1_3 is shown on Figure 10, along with the time of anomaly detection and the

exponential curve fit. The future value of the horizontal feature was predicted by a curve extrapolation to reach a threshold of 5.99 at observation #2115. This threshold was obtained by multiplying the value of the horizontal feature at the time of anomaly detection by the 5.47 factor obtained previously.

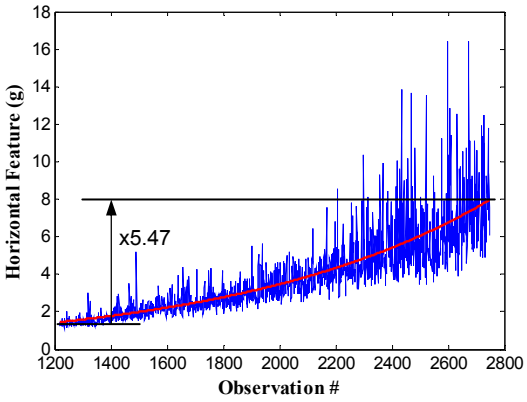

Figure 9. Plot of horizontal feature for bearing 1_1 from the time of anomaly to the end of the first degradation mode. Exponential curve fit is shown.

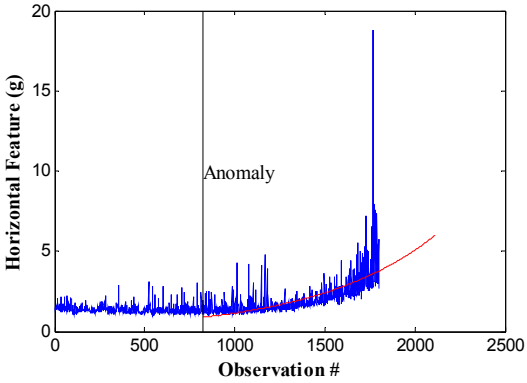

Figure 10. Plot of horizontal feature for bearing 1_3 with an exponential curve fit extrapolation. The feature is predicted to reach the threshold at observation #2115.

Anomaly detection at observation #822 is assumed to be the beginning of anomaly 1 (Figure 10). By subtracting 822 from 2115, we obtained the duration of the anomaly 1 as 1293 observations, or 12,930 seconds. The duration of anomaly 2 was estimated by multiplying the duration of anomaly 1 by the anomaly ratio obtained from bearing 1_1, as shown by the following equation:

$$2^{nd}\ Ano.\ Duration = 0.0366 \times 1^{st}\ Ano.\ Duration$$

which computed to be 478 seconds, rounded to the nearest second. The time to failure from the point of anomaly detection was calculated by adding both anomaly durations together:

$$Time\ from\ 1^{st}\ anomaly\ to\ failure = 12{,}930 + 478 = 13{,}404\ \text{s}$$

Since bearing 1_3 survived for 9800 seconds into the experiment, the RUL estimation for bearing 1_3 was:

$$RUL\ of\ Bearing\ 1\_3 = 13404 - 9800 = 3604\ \text{s}$$

With respect to the actual RUL provided by the competition organizer, this estimate had an error of 37%. Bearing 1_3 was the only test bearing that had an increasing trend and an exponential curve fit could be applied to extrapolate the feature to a threshold line.

In other cases, the bearing features did not exhibit an increasing trend over time. This was the case for training bearing 2_2. Figure 11 shows the plots of anomaly detection for bearing 2_2 and the corresponding response of the horizontal and vertical vibration features. Vertical lines in the plot divide the plot into a healthy region and three anomalous regions. In test bearing cases where the vibration features did not exhibit an increasing trend, RUL estimation was done by computing the ratios of the durations of anomalous regions. Referring to bearing 2_2 as an example (Figure 11), the region for Anomaly 1 lasted from observations #181 to #415, i.e., 2340 seconds. The region for Anomaly 2 lasted from observations #416 to #752, i.e., 3360 seconds. Lastly, the region for Anomaly 3 lasted from observations #753 to failure, i.e., 440 seconds. In most of the test data, the Anomaly 2 and Anomaly 3 regions were not known; therefore, the durations of Anomaly 2 and Anomaly 3 had to be estimated based on the duration of Anomaly 1. Taking the ratios of the anomaly durations for training bearing 2_2, we obtain the relationships:

$$Ano2\ Duration = 1.434\ Ano1\ Duration$$

$$Ano3\ Duration = 0.187\ Ano1\ Duration$$

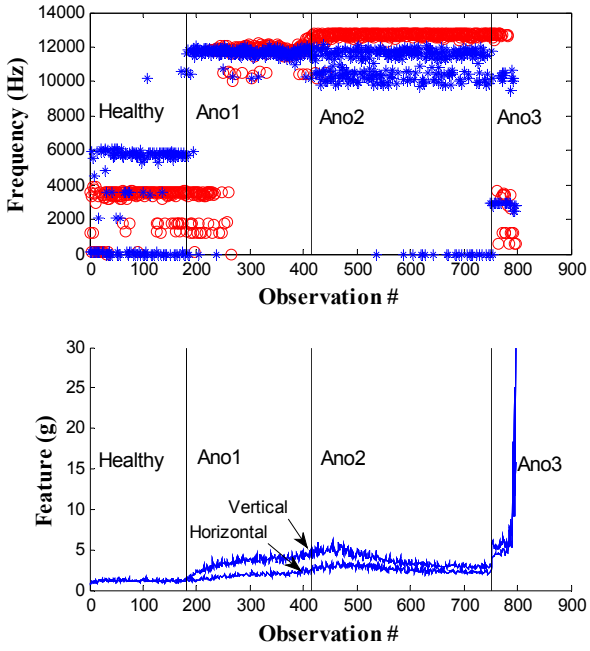

Figure 11. Plots of anomaly detection (top) and features (bottom) for training bearing 2_2. Vertical lines show where anomalies were detected in the frequency analysis.

The method of computing the ratios of anomaly durations was used when there was no increasing trend in the test data provided. RUL estimation using this method did not give a

distribution of failure time due to the small amount of training data. A summary of the RUL estimation errors calculated using Method 3 is shown on Table II. For each test bearing, two RUL values were generated using the two training bearings provided for each loading condition. The average of the two RUL values was taken. An exception was made for bearing 1_3 because it had a unique trend that resembled bearing 1_1. Another exception was when one of the RULs was calculated to be a negative number. In that case, it was omitted from the averaging calculation.

TABLE II. ERRORS IN RUL ESTIMATION USING METHOD 3

Condition 1		Condition 2		Condition 3	
Bearing	**Error (%)**	**Bearing**	**Error (%)**	**Bearing**	**Error (%)**
B1_3*	37	B2_3	64	B3_3	90
B1_4	80	B2_4	10		
B1_5	9	B2_5	-440		
B1_6	-5	B2_6	49		
B1_7	-2	B2_7	-317		

D. Comparison of RUL estimation methods

The three methods presented above used different features for tracking degradation and different approaches for estimating RUL. In this section we discuss the strengths and challenges in using each method.

In method 1, spectral kurtosis of the acceleration signal was found as the feature that had the most monotonic trend of increasing over time. This feature was useful in tracking the accumulation of damage with time. However, the exponential model the Bayesian Monte Carlo simulation was not ideal to capture the abrupt increase behavior near the end of life. This behavior introduced a tendency of overestimation of RUL. One of the advantages of this method is its ability to generate a distribution of RUL from which an operator may take decisions based on risk level.

In method 2, a monotonic trend was not required of the features because the soft computing technique was able to identify patterns among the features. The six training bearings exhibited large variability in failure times (1hr to 7 hrs). Due to the limited training samples and their large variability the LS-SVR over-estimated the RUL of some of the test bearings. If more training bearings were provided for each loading condition, the algorithm accuracy would have performed with better accuracy.

In method 3, estimation of RUL was based on making comparisons on durations of degradation stages between the training and test bearings, except for bearing 1_3. Anomalies were detected by analyzing the changes in frequency signatures. In some bearings however, the anomalies were not clearly identified due to no change in frequencies or noisy patterns in the frequencies. Once anomalies were identified, the RUL estimation process was straight forward and required a minimal computing complexity. Further improvement of this method may come from adding more training bearings where RUL estimate can be obtained as a distribution instead of a single average value.

III. CONCLUSIONS

Three methods for estimating the remaining useful life of ball bearings have been presented. The limited amount of training data introduced high uncertainties in results of all three methods, but the third method was found to be the most accurate overall, which led to our team winning the 2012 data challenge competition. Conservative measures were taken in the estimation process to favor an early prediction over a late one. Future work may include adding a confidence interval to provide the operator with a choice of risk level in estimating RUL.

Despite the multiple challenges in analyzing the data—including limited training samples, no information about failure modes, no fixed failure threshold, and a wide range of failure times—the authors were able to devise a novel method for bearing prognostics which, in four out of eleven cases, was able to predict within a 10% error margin.

ACKNOWLEDGMENT

The authors would like to thank the more than 100 companies and organizations that support research activities at the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland annually. Special thanks go to the members of the Prognostic and Health Management Consortium who directly support the prognostics works at CALCE.

REFERENCES

[1] FEMTO-ST, "IEEE PHM 2012 Data Challenge," online website, last accessed on May 31, 2012. <http://www.femto-st.fr/en/Research-departments/AS2M/Research-groups/PHM/IEEE-PHM-2012-Data-challenge.php>

[2] N. Tandon and A. Choundhury, "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," Tribology International, vol. 32, no. 8, pp. 469–480, 1999.

[3] Y. Shao and K. Nezu, "Prognosis of remaining bearing life using neural networks," Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, pp. 217–230, 2000.

[4] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: A neural network approach," IEEE Transactions on Industrial Electronics, vol. 51, pp. 694–700, 2004.

[5] W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster-Shafer theory and the Bayesian Monte Carlo method," Journal of Power Sources, vol. 196, pp. 10314–10321, 2011.

[6] N. Gebraeel, M. Lawley, R. Li, and J. Ryan, "Residual-life distributions from component degradation signals: A Bayesian approach," IIE Transactions, vol. 37, pp. 543–557, 2005.

[7] F. Cadini, E. Zio, and D. Avram, "Model-based Monte Carlo state estimation for condition-based component replacement," Reliability engineering and System Safety, vol. 94, pp. 752–758, 2009.

[8] B. Kedem, "Spectral analysis and discrimination by zero-crossings," Proceedings of the IEEE, vol. 74, pp. 1477-1493, 1986.

[9] Z-L. Gaing, "Wavelet-based neural network for power disturbance recognition and classification," IEEE Transaction on Power Delivery, vol. 19, no. 4, pp. 1560-1568, 2004.

[10] P. Bonissone And K. Goebel, " When will it break? A hybrid soft computing model to predict time-to-break margins in paper machines," Proceedings of SPIE 47th Annual meeting , International Symposium on Optical Science and Technology, vol. 4787, pp. 53-64, 2002.

[11] J. A. K. Suykens and J. Vandewalle, "Recurrent least squares support vector machines," IEEE Transactions on Circuits and Systems-I, vol. 47, no. 7, pp. 1109-1114, 2000.