Aalborg Universitet

Predicting the Remaining Useful Life of Rolling Element Bearings

Jantunen, Erkki; Hooghoudt, Jan Otto; Yi, Yang; McKay, Mark

Predicting the Remaining Useful Life of Rolling Element Bearings

Erkki Jantunen
Industrial Systems
VTT Technical Research Centre
of Finland Ltd.
Espoo, Finland
erkki.jantunen@vtt.fi

Jan-Otto Hooghoudt
Department of Civil
Engineering
Aalborg University
Aalborg, Denmark
joh@civil.aau.dk

Yi Yang
Department of Civil
Engineering
Aalborg University
Aalborg, Denmark
yy@civil.aau.dk

Mark McKay
Industrial Systems
VTT Technical Research Centre
of Finland Ltd.
Espoo, Finland
mark.mckay@vtt.fi

Abstract—**Condition monitoring of rolling element bearings is of vital importance in order to keep the industrial wheels running. In wind industry this is especially important due to the challenges in practical maintenance. The paper presents an attempt to improve the capability of prediction of remaining useful life of rolling bearings. The approach is based on the understanding of the wear of bearings i.e. wear modelling is briefly discussed. A simulation model has been built to produce vibration data of the monitoring of rolling bearings taking into account typical vibration excitations in addition to the wear. The simulation model is used to develop signal analysis methods and means of prognosis of the remaining useful life. One complete example of the above described process is shown and discussed in the paper.**

Keywords—rolling element bearing, condition monitoring, vibration measurements, signal analysis, diagnosis, prognosis, remaining useful life

I. Introduction

Rolling element bearings (REBs) are the most vulnerable mechanical components in rotary machines. Failures of REBs decrease the availability of relevant equipment, which may cause economic loss. To deal with this issue, researchers around the world have proposed many techniques to predict the evolution of damage; have constructed tools to predict the remaining useful life (RUL), and finally have developed preventive strategies for operation and maintenance (O&M). The primary objective of this paper is to review some commonly used diagnosis and prognosis approaches for REBs, to propose some tailor-made evaluation procedures especially suitable for engineering applications, and to do a case study to validate the procedures.

A. Literature Review

The diagnosis and prognosis of REBs can be done from three perspectives, i.e. physics-based method (Method 1), data-driven method (Method 2) and the combined physics-based and data-driven method (Method 3). This paper will focus on the application of Method 3.

A physical model (also commonly known as degradation model) refers to a model that can physically characterize the real deterioration of the mechanical components in a qualitative or quantitative manner. An example is for instance the Paris-Erdogan law, which has become rather popular due to its good performance in characterizing the crack propagation. The Paris-Erdogan model is based upon a 'direct' damage indicator, i.e. damage size. For a REB, the damage indictor may, however, not be necessarily the crack size.

A general form of damage could be defined here for the extension of Paris-Erdogan law, e.g. for spalling areas or pitting. Some researchers have proposed a modified Paris-Erdogan [1, 2]. In principle, a physical model should be constructed based upon mechanics analyses and the measurement-and inspection information either from the real in-service REBs or from full-scale laboratory tests. A physical model based upon the direct damage indicator has advantages over data-driven methods, as for instance it can, once it has been derived, straightforwardly explain the failure mechanism and quantify the damage evolution. Unfortunately, commonly such a model is difficult to put into practice, as for instance, the real damage sizes of REBs during their lifetime are rarely known or available in practice. What is even worse, the results from full-scale laboratory tests are commonly not applicable to engineering practices. Therefore, efforts should be put in selecting some useful indirect damage indicators. Often, vibration signals have been successfully put into practice for diagnosis and prognosis for many years. As a forerunner of the application of vibration signals, Gebraeel developed a general degradation model which follows an exponentially varying law based upon the observations of deterioration of REBs [3, 4].

You and Meng applied the degradation model proposed by Gebraeel for predictive maintenance scheduling for mechanical components [5]. Following the principles proposed by Gebraeel, various researchers adopted this degradation model to predict RUL of mechanical components. For instance, Li et al. adopted the main ideas proposed by Gebraeel but modified the probabilistic distribution of the model parameter to a Gamma distribution to fit the vibration signals of their cases. With the updated model parameters, they estimated the probabilistic distribution of the RUL [6]. Si et al. first constructed a general stochastic process-based degradation model and presented a degradation path-dependent approach for adaptive RUL estimation via real-time condition monitoring data [7].

Over the last decade, data-driven methods have got large attention, as they allow for RUL prediction without the necessity to exactly understand the failure mechanisms. Commonly used data-driven methods include: Artificial

Neural Networks (ANN), Hidden Markov Models (HMM), Genetic Algorithms (GA), Support Vector Machines (SVM), as well as many others. We mention that the HMM has for instance been combined with degradation modelling in engineering applications; i.e. in [8] Prakash, Narasimhan and Pandey adopted the degradation mode proposed by Gebraeel and generated virtual degradation vibration signals as testing data for pre-trained HMM to do prognosis and maintenance for bearings and predict the RUL of low-speed bearings.

B. Scope of Work

This paper aims to propose procedures to do diagnosis and prognosis for the industrial operators. Therefore, based upon the basic concept of the degradation model proposed by Gebraeel, a degradation model is constructed by performing simplified trend analysis and curve fitting. The results will be further validated by statistical analyses on (high-pass filtered) vibration signals as well as expert judgement.

The outline of this paper is as follows: Section 1 presents a short literature review and a brief comparison between commonly used approaches. In Section 2 a brief presentation of wear of REBs and an illustration of typical wear processes of REBs are given. In Section 3 Simulation model the presentation of dynamic analysis model used to generate raw vibration signals is given. Section 5, Signal analysis and fault diagnosis, presents the simplified short-term trend analysis and the statistical analysis of high frequency vibration components performing diagnosis of the unhealthy bearing and prediction of RUL based upon trend analysis and expert judgement. Section 6 Discussion summarizes some of the weaknesses and strengths of the presented approach. Section 7 presents a short conclusion.

II. WEAR OF ROLLING ELEMENT BEARINGS

Bearing wear typically begins with normal cyclic loading, mainly vibration. This results in a phenomenon called fretting, where small amplitude, localized and tangential motion occurs between any of the three pairs of surfaces in relative motion to one another. With the lubricant forced away from the contact area, local and permanent changes begin to take place. These changes include, loose oxidized and relatively hard particles, exposed and hard surface asperities and surface fatigue. As a result, the surfaces will experience micro pitting called false brinelling.

The resulting failure processes can be divided into two categories: surface and sub-surface failures, based on their location. Since the material is at least somewhat fatigued and continues to be perturbed during bearing rotation, material failures tend to occur. Having hardened surfaces to protect against wear, faults cracks and other material failures typically initiate in the sub-surface region. However, failures located either on the surface or very close to it, are also common. In addition to surface and sub-surface rolling contact fatigue initiated failures, all regular failure mechanism are also relevant in bearing operation. These include for example: real brinelling, corrosion and manufacturing errors, among others [9].

The wear process in bearings, which eventually leads to failure, can be divided into five stages [10]. In the initial

stage the formation of dents occurs. This is commonly increased by poor operating conditions, such as inappropriate lubrication and contamination. The dents are then further reshaped and evolve into defects. Next, the defects grow in size and propagate. This is followed by spalling of the material and damage growth. Finally, the process ends in bearing failure. The whole process is depicted in Fig. 1.
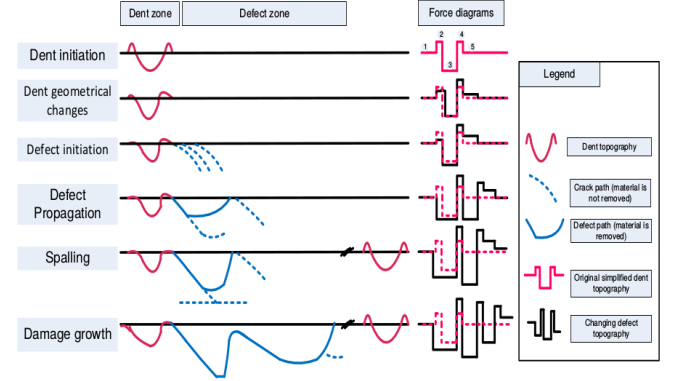

Fig. 1. The evolution of bearing wear, from [10]

III. SIMULATION MODEL

A simulation model has been created in order to be able to produce vibration acceleration data for testing and development of different kind of diagnosis and prognosis tools for the detection of bearing faults and the prediction of the remaining useful life. The simulation model can handle unbalance, misalignment, cage, outer, and inner race faults and can also introduce noise into the measuring signals.

A. Unbalance

Unbalance is usually considered as very easy to simulate since only a sinusoidal signal needs to be introduced at the running speed. Naturally, the height of the amplitude of the signal depends on the amount of unbalance there is. It is possible that some other excitation takes place at the running speed and possibly at a different phase. In this paper the applied model simply covers the basic unbalance and no harmonic components that might be caused by for instance the blades although easy to introduce, are not considered.

It should be noted that the unbalance acts as an exciter for misalignment and bearing faults i.e. if there would be no unbalance or any other excitation in the rotating system, there would not be any loading on the bearings and consequently no wear. Also, misalignment gets the original excitation from other root causes.

B. Misalignment

If unbalance is the most common fault type in rotating machines then misalignment is the next most common. The problem with misalignment is that it increases the loading of bearings and thus can be the root cause of bearing failures. When condition monitoring is carried out based on vibration measurements, it is commonly known that the harmonics of the rotation speed in the vibration spectrum reveal the presence of misalignment. What is not so widely realized is that there are actually not really harmonic components acting as loads due to misalignment, but that it is a feature of the Fourier Transform to show somehow "twisted" signals i.e. signals that are not purely sinusoidal with these

harmonics of the rotational speed [11]. In the simulator the misalignment fault is created by flattening the originally sinusoidal signal on one side based on the idea that in case of misalignment the bearing is pressed on one side and thus the vibration is not symmetric. The end result is naturally that we see these harmonics of the rotational speed in the spectrum just as expected.

C. Bearing faults

In the previous section the wear of rolling element bearings has been described. From the simulation point of view it is important to take into account the first natural frequency of the bearing structure in question i.e. outer inner race etc. The impact that is the result of the rolling element hitting the fault i.e. geometrical change of the race way excites the bearing vibration at the first natural frequency. The natural frequency is a function of the bearing geometry and the larger the bearing, the lower the frequency. In case of small bearings that are used in laboratory tests, it might be over 10 kHz and in case of large bearings that can be found in wind turbines, it might well below 1 kHz. Typically, vibration measurement equipment are tuned for envelope detection at frequencies from 1 kHz to 5 kHz.

When the bearing fault is tiny, it is typically very sharp at edges and consequently causes a very clearly seen impact. However, when it increases in size the impact is not sharp anymore and can actually consist of a number of impacts. The simulation of the bearing takes into account the size of the bearing and the phase of the development of the fault. Detailed data of the bearing that has been simulated is shown in Table I. However, the current version of the simulator does not fully take into account the parameters that have been indicated with grey background. Instead, these are assumed to be typical representative values. Naturally, there is a lot of variation in the development of bearing faults and thus the simulator has some features built into it.

TABLE I.   PARAMETERS OF THE BEARING IN THE SIMULATION

Rotor mass	0.6 kg
Rotor speed	2000 rpm
Damping	200 Ns/m
Inner Race Diameter (D_i)	49.912 mm
Outer race diameter (D_o)	80.088 mm
Pitch diameter (D)	65 mm
Ball diameter (d)	15.081 mm
Radius of curvature of the outer ring raceway (r_o)	8.01 mm
Radius of curvature of the inner ring raceway (r_i)	7.665 mm
Number of balls (Z)	8
Contact angle	0º
Radial clearence	0.005 mm

D. Noise

When vibration measurements are made there is always noise present. It can come from the same machines or other machines, measuring electronics etc. but the problem is that the noise always makes it more difficult to detect the signals that indicate wear. In order to make the simulation

correspond to reality additional noise is introduced to the simulated measuring signal. Naturally, the type and amount of noise can be tuned. One special use for the noise is in the testing of cheap sensor and measuring solutions. The noise level can be tuned according to the technical specification of the measuring equipment and the cheap solutions can be tested against more sophisticated equipment in order to see how early the indication of bearing wear can be seen with each of these measuring solutions.

Virtual vibration simulation of a bearing was done to generate the time series of vibration signals. The bearing used in [12] was selected for this case study. The duration of this simulation is about 6.5 years. The parameters are listed in Table 1. The average load amplitude is 20 N with ±10% deviation. The average running speed is 2000 rpm with ±10% deviation. A total of 99 data files were collected, with each data file representing 0.2 seconds of vibration signal. .

The time dependent development of the bearing fault was introduced based on the experience gained from laboratory tests with a set of four bearings that were run to failure in accelerated bearing tests [13]. The idea was to introduce such a development of the fault that the indirect condition monitoring methods would show similar development with simulated data and laboratory data. Unfortunately, even though bearings are widely monitored in the industry not a lot of data has been made available that could be used to study the whole life span of a bearing. The natural reason for this is that the lifetime typically is very long thus making it challenging to keep the records available for the whole lifetime. However, the simulated wear development can be claimed to resemble that of real bearings especially when the end of the lifetime is studied i.e. when there are indicators revealing the presence of a fault.

In Figs. 2 and 3 two examples of simulation are shown. The time data presented in figure 2 shows the bearing in good condition without any indication of an upcoming failure, whereas in figure 3 the time data is taken from the period when the bearing is very close to collapsing totally. This wear can be seen as spikes in figure 3.
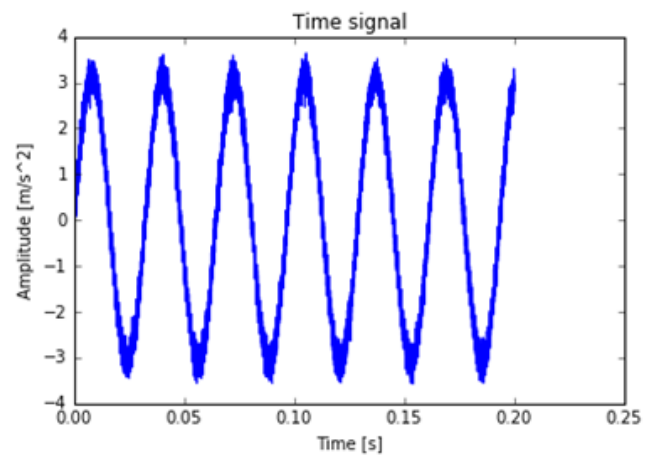

Fig. 2. Time domain vibration of a faultless bearing

IV. SIGNAL ANALYSIS AND FAULT DIAGNOSIS

When signal analysis is carried out in order to detect the presence of a fault, it is logically important that the chosen signal analysis techniques are able to separate noise and

other influencing factors from the features that reveal the presence of faults. Clearly in the case of REB faults it is of highest importance that the used technique can detect the influence of the impacts described earlier in this paper. One principal problem in doing this is that the impacts do not take place at constant intervals due the variation between sliding and rolling in a rolling element bearing. This basic challenge has been the reason why envelope detection has become so popular. The Hilbert transform that is carried out in connection with envelope detection simply helps in handling this issue. The small differences in the interval of the impacts become meaningless and also the type of the fault i.e. outer or inner race etc. can be defined through the use of spectrum analysis in the case of the time domain that has been passed through the Hilbert transform.
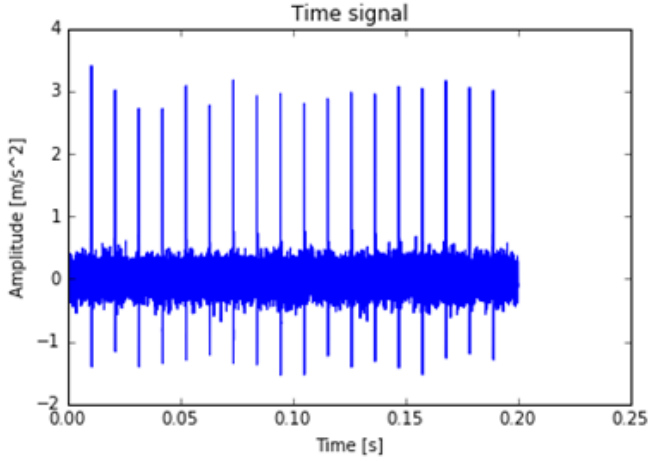

Fig. 3. Time domain vibration of a faulty bearing

A. Peak envelope

As mentioned in the introduction, a combination of physical modelling (degradation model) and statistical analysis will be used to do diagnosis and prognosis for a specific case as briefly presented in Section IV.

An optimal selection of damage indicators is very important to construct a degradation model. Generally, there are many vibration features to choose from. Possibly, some damage indicators might only predict the faults under well-controlled testing conditions in laboratory. According to engineering experience, the difference of consecutive vibration signals is used as a damage indicator and a threshold value of two is applied. Pre-processing of the damage indicator time series is briefly described below:

- Envelope analysis [12, 14] of the damage indicator time series;
- Simplified trend analysis for the envelope of the damage indicator time series to predict the short-term trend with consideration of 95% confidential interval, by using the Matlab Curve Fitting toolbox [15].

A sensitivity study was performed to determine the optimal number of data points used for envelope analysis. According to the results, 2000 data points could be an optimal option. In order to give a higher resolution plots, the envelope of the damage indicator was truncated (i.e. 1st to 58th data sets were truncated), as shown in Figure 4. It was noticed that the envelope fluctuated around a horizontal line

over the initial 70~80 percent of time duration; an increase in envelope amplitude was observed from around the sample point No. 6.3×10^5 (around 2106 days from the start of operation); afterwards, the envelope increased significantly and finally went down which might mean the bearing failed. In this case study efforts were given to long-term trend prediction, however, the results of curve fitting may be only more applicable to a shorter period of time ahead of the failure. The envelope time series was fitted to an exponential law, i.e.

$$y = ae^{bt} \tag{1}$$

The means of a and b are 0.6027 and 1.848×10^{-6}, respectively. The 95% confidential intervals of a and b are [0.6, 0.6055] and [1.84×10^{-6}, 1.857×10^{-6}], respectively.

The fitted curve, as well as its 95% confidential interval, is shown in Fig. 5 and Fig. 6. According to the envelope curve, the first passage to the threshold seemed to be around the sample point No. 1.6×10^5 (around 535 days from the start of operation). Expert judgement shall be adopted together with the trend analysis results to determine whether it is necessary to conduct NDT inspections to prevent catastrophic outcome. According to the fitted trend curve, the first passage was around the sample point No. 6.4×10^5 (around 2136 days from the start of operation)
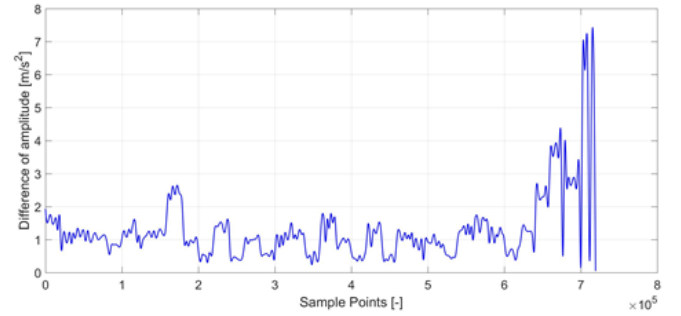
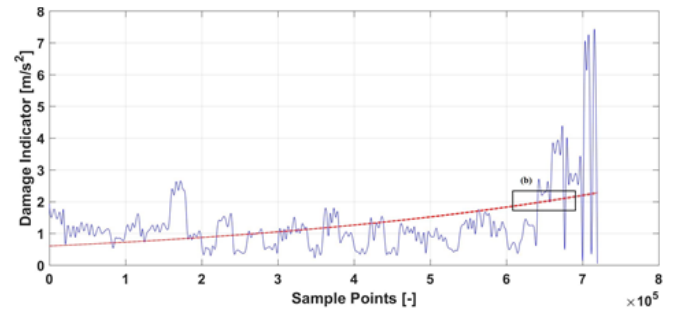

Fig. 4. Envelope of the damage indicator

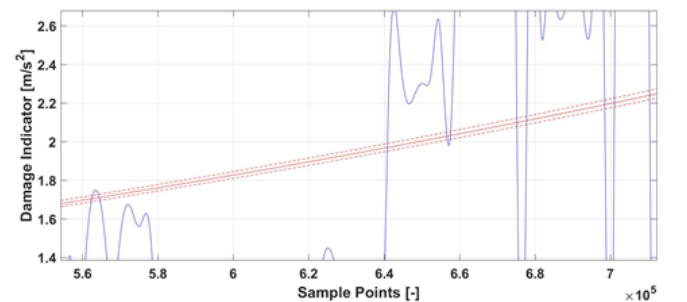Fig. 5. Fitted exponential trend. Enlarged view of the black rectangular box is shown in Fig. 6.

Fig. 6. Enlarged view of the trend

B. Kurtosis of High Frequencies

In this section, the presence of a bearing fault is diagnosed by monitoring the kurtosis of the high frequency vibration signal. We start, however, by analysing the vibration signal for four samples in slightly more depth. One of the samples relates to the healthy state of the bearing (Sample 7) and the other three samples relate to the end of the life of the bearing (Samples 89, 91 and 95). We note that the end of the life of the bearing occurred for Sample 96.

In Fig. 7 the raw vibration data for these four samples is plotted. Clearly, as previous displayed in Fig. 3, a faulty bearing commonly displays high frequency spikes. Therefore, we study the high frequency spectrum of the bearing in slightly more detail here. For each sample, the high frequency signal is obtained by subtracting the low frequency signal from the full signal. The low frequency signal is computed by averaging for each time observation t the vibration signal over the fifty observations before and after this observation and at time t itself (so over 101 observations). In Fig. 8 the high frequency signal is plotted for the same samples as plotted in Fig. 7.
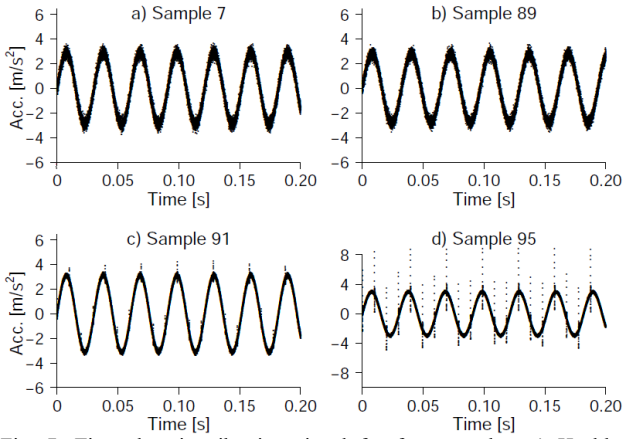

Fig. 7. Time domain vibration signal for four samples: a) Healthy bearing, b-d) bearing close to end of life. The end of the life of the bearing occurred for Sample 96.
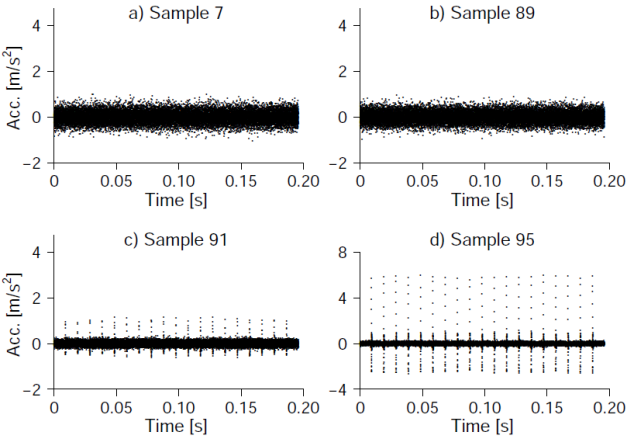

Fig. 8. High frequency vibration signal for the four samples

We have analyzed the high frequency signal further by aid of QQ-plots and by plotting the Log-kurtosis for all the samples. In Fig. 9 QQ-plots are shown for the high frequency signals of the four samples, while in Fig. 10 the Log-kurtosis of the high frequency signal is plotted for each of ninety-six samples.

Fig. 9 shows that for the Samples 7 and 89 the high frequency signal is normally distributed around its mean of zero, while for samples 91 and 95 the high frequency signal displays many more observations in the tail of the distribution than expected for a Normal distribution. Further, from Fig. 10 we observe that the Log-kurtosis of the Samples 91 to 96 show excessively large values (all above 2), while also the samples 80, 86 and 90, show values larger than any of the first 79 observations. This implies that the kurtosis is possibly not only a good indicator for extensive bearing damage, but also for initial bearing damage.
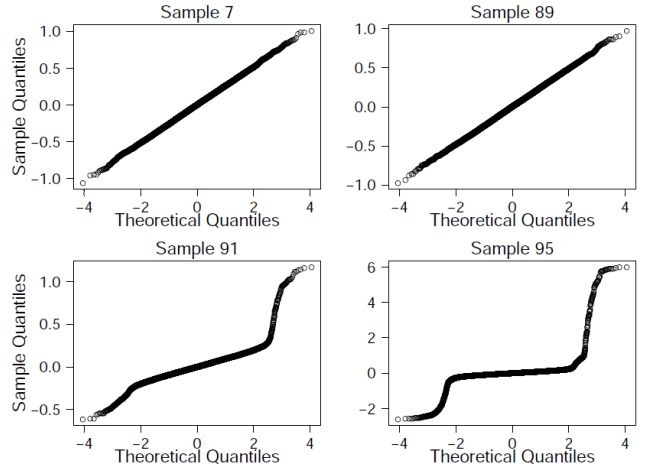

Fig. 9. QQ-plots of the high frequency vibration signal for the four samples
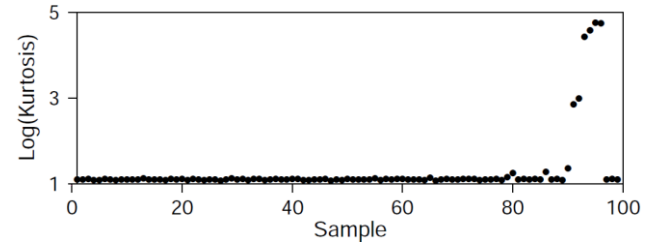

Fig. 10. Logarithm of the kurtosis of the high frequency vibration signal plotted for each sample.

An easily to be implemented automated strategy for bearing monitoring could therefore be implementing a Control chart (see e.g. [16]) for the kurtosis of the high frequency vibration signal, as implemented below in Fig. 11. In this case we simply computed the standard deviation of the Log-kurtosis over the first forty samples and constructed the upper control limit as three standard deviations added to the mean (UCL= 1.13).
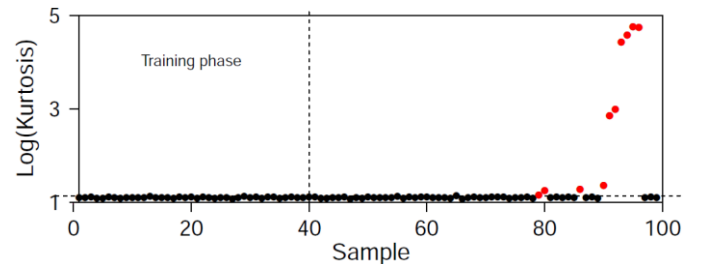

Fig. 11. Control chart of the Log-kurtosis of the high frequency vibration signal. Points in red are above the upper control limit (horizontal dashed line).

V. Discussion

In this study signal processing was done from two perspectives as presented in Section IV. First, simplified envelope analyses were performed to remove the effect of noise on the estimation of trend. Second, high frequency band-pass vibration signals were extracted for the statistical analysis of the kurtosis of this signal. According to the results of simplified trend analysis together with engineering judgement, the first passage of the signal to the threshold occurred at the 67th data set, while the start of the failure actually started from dataset 65 onwards. However, this rather precise observation could be due to coincidence as the increase in the envelope seems to be just a glitch, as not repeated for a long time until close to collapse (failure of the bearing happened at the 96th dataset). Further, the exponential fit is mainly influenced by the final fast, but very noisy, increase after 6e5 sample points in Fig. 5. Therefore the fit is probably not too reliable for the times before the strong increase in amplitude. Similar, but probably slightly more robust, results were gained by the high frequency analysis of the kurtosis. However, clearly, both signal analyses methods are able to detect the bearing fault and so the results seem promising.

In the close future more simulated data sets of failed bearings will be generated in order to increase statistics and predictive power. By the high frequency kurtosis analysis a clear and easy to implement method for RUL prediction is envisaged. That is, from a sufficient amount of failure simulation datasets the plan is to deduct the probability that the bearing fails after time t when: a first out-of-control-observation has been observed; a second out-of-control-observation has been observed; etc.

Finally, the main motivation for the development of the simulator is to provide aid in the development of remaining useful life prediction techniques. The researchers plan to further develop the simulator in order to make it mimic more closely the development of bearing faults and thus provide a tool that can easily provide a large amount of data for the development of more reliable and effective tools for the prediction of remaining useful life of rolling element bearings. Naturally, in the end all these findings need to be verified with real data.

VI. Conclusion

In order to be able to follow the Condition-Based Maintenance strategy it is necessary to know the condition of machinery. This is especially important in the case of wind turbines due to the challenges in carrying out practical maintenance work. In this paper, an approach to further develop the techniques for predicting the remaining useful life of rolling element bearings has been presented. The approach is based on the understanding of wear of bearings, through the simulation of vibration acceleration. The simulator provides the opportunity to get sufficient amount of condition monitoring data from various phases of the bearing life. This data can be used for the testing and development of sophisticated signal analysis techniques together with various diagnostic approaches and prognosis methods for the prediction of the remaining useful life. The complete procedure has been presented in this paper with one typical rolling element bearing. The example clearly shows how complicated and challenging the prediction of the remaining useful life of a rolling element bearing is. The

results in the presented case are promising and meaningful but still far from optimal. It is expected that in the future the developed simulation model will also be further tested against real condition monitoring data from industry.

References

[1] L. Wang and R.X. Gao, (2005), "Condition Monitoring and Control for Intelligent Manufacturing", Springer, ISBN 978-1-846628-268-3

[2] B. Zhang, C. Sconyers and C. Byington C, "A Probabilistic Fault Detection Approach: Application to Bearing Fault Detection", IEEE Transactions on Industrial Electronics, Vol. 58 No. 5, 2011.

[3] Z. Nagi Gebraeel. "Real-time Degradation Modelling and Residual Life Prediction for Component Maintenance and Replacement", Ph.D. Dissertation. School of Industrial Engineering. Purdue University. 2003.

[4] Z. Nagi Gebraeel, M. A. Lawley, R. Li and K.R. Jennifer. "Residual-life Distributions from Component Degradation Signals: A Bayesian Approach", IIE Transactions 37, 543-557. 2005.

[5] M-Y. You and G. Meng. "A Modularized Framework for Predictive Maintenance Scheduling", Journal of Risk and Reliability. 226(4) 380-391. 2011.

[6] B. Li, Z. Gao, Z.J. Shen, X.F. Chen and Z.J. He. "A Gamma Bayesian Exponential Model for Computing and Updating Residual Life Distribution of Bearings", Journal of Mechanical Engineering Science. 227(11) 2620-2633. 2013.

[7] X.S. Si, W.B. Wang, M.Y. Chen, C.H. Hu and D.H. Zhou. "A Degradation Path-dependent Approach for Remaining Useful Life Estimation with An Exact and Closed-form Solution", European Journal of Operational Research. 226(2013) 53-66. 2013.

[8] G. Prakash, S. Narasimhan and M. D. Pandey, "Condition-based Maintenance of Low Speed Rolling Element Bearings Using Hidden Markov Model". International Journal of Prognosis and Health Management. 2017.

[9] W.T. Becker and R.J. Shipley. "Failure analysis and prevention, 2002", Vol. 11. USA: ASM International

[10] I. El-Thalji and E. Jantunen, "Fault analysis of the wear fault development in rolling bearings," Engineering Failure Analysis, 57 (2015), pp. 470-482

[11] E. Jantunen, I. El-Thaji, D. Baglee, and T.L. Lagö. "Problems with using Fast Fourier Transform for rotating equipment: Is it time for an update?", 27th International Congress of Condition Monitoring and Diagnostic Engineering, COMADEM 2014, Sep 2014, Brisbane, Australia

[12] J. Liu, Y.M. Shao, T.C. Lim. "Vibration Analysis of Ball Bearings with A Localized Defect Applying Piecewise Response Function". Mechanism and Machine Theory. Vol. 56, 2012, pp. 156-169.

[13] E. Jantunen. "How to Diagnose the Wear of Rolling Element Bearings Based on Indirect Condition Monitoring Methods". International Journal of COMADEM. Vol. 9, Issue 3, 2006, pp. 24-38.

[14] L.H. Wang, R.X. Gao, "Condition Monitoring and Control for Intelligent Manufacturing", Springer-Verlad London Limited, 2006, pp. 155-157.

[15] Mathworks, Matlab Curve Fitting User Manual.

[16] NIST/SEMATECH e-Handbook of Statistical Methods.