International Journal of Advancements in Mechanical and Aeronautical Engineering – IJAMAE
Volume 1 : Issue 3 [ISSN 2372 –4153]
Publication Date : 30 September, 2014

Condition monitoring of slow rotating bearings based on Bayesian linear regression

[S.A. Aye, P.S. Heyns and C.J.H. Thiart]

***Abstract*— Preventive maintenance and run to failure techniques of bearing condition monitoring could be quite expensive. The failure could equally be catastrophic thereby leading to the damage of several other components of the machinery. Many bearing studies are done at constant loading and speed and constant conditions. However, this study is carried at varying loads and speed conditions. This paper therefore presents a condition monitoring methodology based on Bayesian linear regression technique which is at varying load and speed condition.**

***Keywords*— condition monitoring, Bayesian linear regression, slow rotating bearings**

I. Introduction

Zaidan et al [1] investigated fatigue induced crack-growth using Bayesian approaches to illustrate the use of data-driven prognostics to deliver benefits to the industry. Zonta et al. [2] presented a damage detection procedure based on Bayesian analysis of data recorded by permanent monitoring systems as applied to condition assessment of Precast Reinforced Concrete (PRC) bridges. Given the prior distribution, the method assigned posterior probability to each scenario as well as updated probability distributions to each parameter. The effectiveness of this method was illustrated as applied to a short span PRC Bridge instrumented with a number of fiber-optic long gauge-length strain sensors. Zuffranieri and Robinson [3] applied Bayesian medical monitoring concepts based on using real-time performance-related data to make statistical predictions about a patient's future health. Zhang et al. [4] proposed a Bayesian approach for estimating the failure rate of power transformers. This was accomplished by using likelihood function which is constructed based on available condition monitoring information to update the assumed probabilistic model of transformer failure rate. Heyns et al. [5] proposed a methodology based on Bayesian regression to isolate the effect of varying vehicle speed on the measured vehicle response metric.

Most condition monitoring techniques for bearings do not take into cognisance the varying load and speed. This paper therefore presents a condition monitoring methodology based on Bayesian linear regression technique which is independent of load and speed changes. Also, there are number of studies that applied Bayesian approach to bearing condition monitoring. However, most of these studies have essentially applied it for classification purposes. Therefore, this study deviates from previous studies by using the Bayesian approach for predictive purpose. Specifically, the study uses a Bayesian linear regression with Gaussian distribution assumption to determine the relationship between bearing vibration signal and speed at various loading conditions for prediction purposes.

S.A. Aye/Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa

P.S. Heyns/Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa

C.J.H. Thiart/Department of Mechanical and Aeronautical Engineering, University of Pretoria, South Africa

II. Bayesian Linear Regression methodology

The parametric approach focuses on the use of probability distributions having specific functional forms governed by a small number of adaptive parameters, such as the mean and variance whose values are to be determined from the data set. The probability distributions include beta (binomial) and Dirichlet (multinomial) distributions for discrete random variables and the Gaussian distribution and Gaussian mixture distribution for continuous variables. In this study the data is continuous hence the Gaussian distribution and Gaussian mixture distributions are considered [6].

The Gaussian, also known as the normal distribution, is a widely used model for the distribution of continuous variables [6]. For the case of a single real-valued variable, , the Gaussian distribution is defined by:

$$N(x|\mu,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{1/2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad (1)$$

which is governed by two parameters: μ, called the mean, and σ^2, called the variance.

The reciprocal of the variance is called the precision and is written as:

$$\tau=1/\sigma^2 \quad (2)$$

In this study vibration signal is extracted at different operating conditions (speeds, angles and loading conditions). A regression function, which measures the bearing vibration as a function of the different operating conditions is fitted. The regression function is estimated based on both the data driven likelihood and a parameter prior. The prior serves as indication of the classic nature of these interpolation functions. As such the prior allows for more vigorous interpolation functions, particularly if only limited and noisy data are available.

Let the original Kurtosis value which is associated with a specific loading condition, j as measured at a mean speed of v_i^j over that load be denoted by y_i^j. Some variability in the bearing response as measured over any loading condition at a given speed is expected. This variability is accounted for by including an error term e_i, which is assumed to have a zero mean independent identically distributed Gaussian distribution $e \sim N(0, \sigma_e^2)$ with a constant variance, σ_e^2.

An operating condition vector, x, is defined and implemented so that one vector x_i^j corresponds to each Kurtosis datum point y_i^j . This operating condition vector controls the flexibility of the implemented regression function. The regression function in turn reveals the expected influence of operating condition (speed) on the vibrations values. Four simple regression functions are investigated, namely: linear, quadratic, exponential and double log. The operating condition vector may easily be extended to also investigate higher order polynomial functions, or to also consider the influence of other variables such as bearing acceleration, angular rotation, etc. Bayesian model selection is used to select the most appropriate among the proposed interpolation functions.

An observation y_i^j is described as the sum of the specific loading condition interpolation function as evaluated for the corresponding operating condition vector $f(x_i^j)$ and the noise term e_i .

$$y_i^j = f(x_i^j) + e_i \qquad (3)$$

The interpolation function can be approximated as having a linear dependency on x if the operating condition vector is sufficiently expressive. It may be justified to assume that this linear dependency is expressed by the parameter vector w^j :

$$f(x^j) = \{x^j\}^T w^j \qquad (4)$$

Any discrepancies in this assumption are absorbed by the noise term.

Let all the Kurtosis measurements which correspond to a specific loading condition be denoted by the vector y^j , and let all the associated operating condition vectors be contained in the matrix X^j . The likelihood of the data given the model as represented by the parameter values w^j is denoted as $p(y_i^j | x_i^j, w^j)$. Due to the independent noise assumption the joint likelihood for the, *K*, Kurtosis observations which correspond to any loading condition, may simply be computed as the product of the individual datum point likelihoods:

$$p(y^j | X^j, w^j) = \prod_{i=1}^{K} p(y_i^j | x_i^j, w^j) \qquad (5)$$

$$= N(\{X^j\}^T w^j, \sigma_e^2 I) \qquad (6)$$

The parameter estimates which optimize equation (6) is equivalent to the least square error (LSE) solution. It is assumed that there are a number of loading conditions where sufficiently many measurements are available so that the LSE solution will be fairly good. The parameter values for those LSE estimates (as estimated at loading conditions with good data) are used as guideline to what the typical parameter values are to be at other loading conditions where sufficient data may not be available.

A multivariate Gaussian distribution is estimated from the LSE solutions for the reference loading condition.This distribution is subsequently used as the prior distribution p(w). Let the prior mean be denoted by the vector μ_o, and let the covariance matrix be denoted as Σ_o, so that the prior may be expressed as:

$$p(w) = N(w | \mu_o, \Sigma_o) \qquad (7)$$

According to Bayes' theorem the prior and the data driven likelihood is used to obtain a posterior distribution over the parameter values:

$$posterior = \frac{likelihood \times prior}{marginal\ likelihood}$$

$$p(w^j | y^j, X^j) = \frac{p(y^j | X^j, w^j)(p(w^j))}{p(y^j | X^j)} \qquad (8)$$

where the marginal likelihood $p(y|X)$ serves to normalize the posterior. Prior probability is the probability available before the observation. However, posterior probability is the probability obtained after the observation. The likelihood function expresses how probable the observed data set is for settings of the parameter vector. It may be shown that the posterior distribution is also a Gaussian distribution [6]:

$$p(w^j | y^j, X^j) = N(w^j | \mu_e^j, \Sigma_e^j) \qquad (9)$$

where the posterior mean μ_e and covariance Σ_e for loading condition j is given by:

$$\mu_e^j = \Sigma_e(\Sigma_o^{-1}\mu_o + \tau_e \{X^j\}^T X^j) \qquad (10)$$

$$\Sigma_e^j = (\Sigma_o^{-1} + \tau_e \{X^j\}^T X^j)^{-1} \qquad (11)$$

The likelihood of observing a kurtosis value y_* at different speeds x_*^j may be estimated from the updated likelihood function which is again of Gaussian form [6]:

$$p(y_* | y^j, X^j, \mu_e, \Sigma_e, \sigma_e) = N(y_* | w_*^T x_*^j, \{\sigma_*^j\}^2) \qquad (12)$$

The variance $\{\sigma_*^j\}^2$ of the predictive distribution indicates the uncertainty in a prediction at operating condition x_*^j and is given by:

$$\{\sigma_*^j\}^2 = \sigma_e^{-2} + \{x_*^j\}^T \Sigma_e x_*^j \qquad (13)$$

Bayesian model selection is used to select the most appropriate among the proposed interpolation functions.Bayesian model comparison criterion was proposed by Spiegelhalter et al [7] based on the principle shown in equation 14:

Deviance Information Criterion (DIC) = (goodness of fit) + (complexity) (14)

The deviance ($D(\theta)$) gives the measure of the goodness of fit as shown in equation 15:

$$D(\theta) = -2\log L(data \mid \theta) \qquad (15)$$

Where $\log L$ is the log likelihood and θ is the parameters of the model.

Complexity (pD) measured by estimating the effective number of parameters as shown in equation 16:

$$pD = E_{\theta|y}[\theta] - D(E_{\theta|y}[\theta]) \qquad (16)$$
$$= \overline{D} - D(\overline{\theta})$$

i.e. posterior mean deviance minus deviance evaluated at the posterior mean of the parameters. The DIC is then defined analogously to Akaike Information Criteria (AIC) as equation 17:

$$DIC = D(\bar{\theta}) + 2\,pD \qquad (17)$$
$$= \overline{D} + pD$$

Models with smaller DIC are the best models as they are better supported by the data.

Once, the best model has been selected, the next step would be to predict and/or project the value of kurtosis given the parameters of the model and the specified operating condition (in this case speed).

The posterior predictive distribution is either the replication of y given the model (usuallyrepresented as y^{rep}), or the prediction of a new and unobserved y (usually represented as y^{new} or y'), given the model. This is the likelihood of the replicated or predicted data, averaged over the posterior distribution $p(\theta \mid y)$. These are given by equations 18 and 19:

$$p(y^{rep} \mid y) = \int p(y^{new} \mid \theta) p(\theta \mid y) d\theta \qquad (18)$$

or

$$p(y^{new} \mid y) = \int p(y^{new} \mid \theta) p(\theta \mid y) d\theta \qquad (19)$$

If y has missing values, then the missing ys can be estimated with the posterior predictive distribution [8] as y^{new} from within the model. For the linear regression example, the integralfor prediction is given in equation 20:

$$p(y^{new} \mid y) = \int p(y^{new} \mid \beta, \sigma^2) p(\beta, \sigma^2 \mid y) d\beta d\sigma^2 \qquad (20)$$

The posterior predictive distribution is estimated by equation 21:

$$y^{new} \sim N(\mu, \sigma^2) \qquad (21)$$

Where $\mu = X\beta$ and μ is the conditional mean, while σ^2 is the residual variance.

III. Experimental setup

An experimental setup is used in this research to collect acoustic emission signals from slow rotating bearing. The setup captures the essential physics involved in real life applications as well as practically feasible and fit into the laboratory and not to be too expensive. Damage is introduced to the bearing outer race. The test rig is designed to simulate early stage of bearing defects. The pictorial view of the setup is shown below in Figs. 1. The slow rotating bearing test setup comprises of 36 components: the top end plate, sonic shaker, flat spacer to fit sonic male, load cell, flat spacer male to fit load cell, cylindrical bush, test bearing shaft, dummy bearing shaft, test bearing v belt pulley, servo motor pulley, dummy bearing pulley, bearing inner race, bearing outer race, spacer for boom arm, boom arm for loads, bottom end plate, pillars, servo motor, servo motor support bracket, servo motor v belt, test bearing v belt, base plate bolts, servo motor support bracket bolt, servo motor bolt, support for the setup, test bearing block, speed controller, cables, data loggers, coupling, and acoustic emission sensors.The test rig which consists of all the component parts listed above subdivided into the servomotor unit, the slowly rotating bearing unit, the dummy bearing unit, and a dynamic loading unit. The test setup is designed to be able test any type of slow rotating bearing. Again because of cost limitations, only one type of slow rotating bearing would be tested.

The experimental test setup shown in Fig. 1 below is used in this research to collect the acoustic emission signals for two purposes: to study the vibration and acoustic emission signatures generated by incipient bearing faults, and for the substantiation of the methods to be established. The system is driven by an AC servo motor with the speed range between 0 and 1000 rpm. The shaft rotation speed is controlled by a speed controller. A tachometer is used for shaft rotational speed measurement. Hence, the shaft rotational speed is read from the speed controller, which can be further confirmed by applying the FT on the signals from the tachometer. A dummy bearing is used between the servomotor and the test bearing so as to prevent direct loading and damage to the servo motor. Accelerometers and acoustic emission sensors are mounted on the housing of the tested bearing to measure the vibration and acoustic signals.

The slow rotating bearing is loaded at various equivalent dynamic load of between 0 kN and up to 2 kN load using the sonic shaker. The speed controller controls the rotational speed of the bearing.

The test is conducted at several dynamic loadings of a maximum of between 2 kN. Sonic actuator is used for inducing the dynamic loads. Usually dynamic loads are applied both axially and radially. However, it is possible to find equivalent dynamic loads in only one direction. Hence, application of equivalent dynamic loading is to be done in the axial directions.

The test is conducted at moment loadings of a maximum of between 1 kNm.

Fig. 1a: The experimental test set up

Fig. 1b: Dynamic loading sine wave input shown on K7500 Servocontroller for -1.2200V

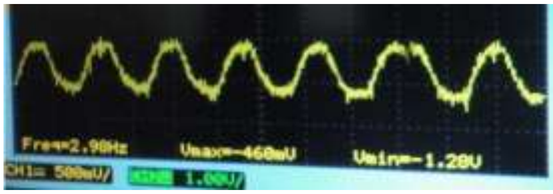

Fig. 1c: Dynamic loading sine wave output shown on Oscilloscope for -1.2800V

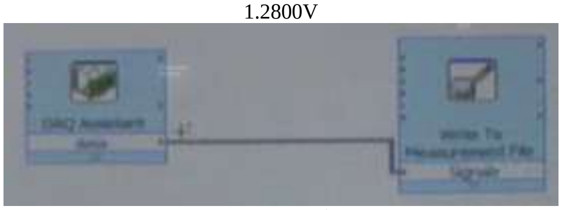

Fig. 1d: Labview interphase for data acquisition

IV. Data acquisition and Signal Processing

Data acquisition is the process of obtaining information on the condition of the health of equipment through method of continuous sampling at selected time intervals. Sensors are used in the collection of data in analog form. The obtained data is then converted into digital format for further processing and feature extraction with the help of special software. Signal detection procedure for bearing condition monitoring is important part for predictive maintenance of machinery. Selection of desired features and their relevant features plays a vital role for both diagnostic and prognostic purposes. The accurate forecast of bearing impended faults can lead to proper planning and replacement in order to avoid disastrous failures of the whole machinery.

The main purpose of data acquisition of signals is to quantify the changes in machinery conditions. Whenever, mechanical distress occurs it appears in the form of relative motion on entire components in the form of vibration or acoustic emission etc. Thus with the help of appropriate sensors this deviation from the norm is measured and processed in order to build either the fault diagnostic models or prognosis models. This research work is dedicated to the prognosis of slowly rotating bearings under time variant conditions. Usually, bearings are chosen based on their life calculation with respect to their industrial use and operational factors.

The AE data acquisition system consisted of piezoelectric-type AE transducers, amplifiers, an A/D card/data logger, and the computer. The signal output from the preamplifier will be connected directly to a commercial acquisition card that occupies one of the ISA slots within a Pentium host PC.

Signal processing software for Acoustic Emission

The National Instruments Lab View software will be used for collection of acoustic emission data. The function for capturing time domain and pre selected sampling time and interval will be used. The rest of the processing and analysis would be performed through Matlab programs for signal processing and analysis.

V. Results and Discussion

The observed kurtosis and speed at different loading conditions is shown in Fig. 2. The kurtosis of a signal is very useful for detecting the presence of an impulse within the signal. It is generally used for detecting discrete impulsive faults in rolling element bearings. Good bearing have a kurtosis value of approximately 3, and bearings with impulsive faults tend to have values greater than 3[9].

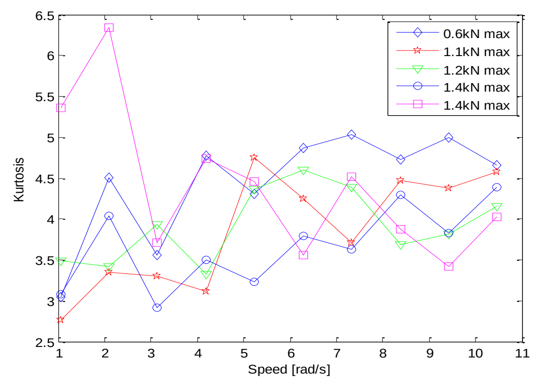

Fig. 2: Observed kurtosis and speed at different loading conditions

Four Bayesian models were specified namely, linear, quadratic, exponential (log linear) and double log. The Bayesian models were estimated using 11000 MCMC simulations. The initial 1000 were used as burn-ins to mitigate the start-up effect.

VI. Model Selection

One easy way to compare Bayesian models is by using the deviance information criterion (DIC) statistic. This statistic is intended to be a measure of model complexity. The model with the smallest DIC is taken to be the best model [7]. Hence, to determine the best model for the data, the DIC was used for model selection. The DIC values for the various models at various loading conditions are tabulated in Table 1. It can be clearly seen that the highest DIC values are for the linear and quadratic models. Similarly, the lowest values are for Exponential and Double log models. After convergence, the DIC values obtained for the linear, quadratic, exponential and log-log models are 18.689, 17.365, -25.282 and -28.998 respectively for the first loading condition. The double log has the smallest DIC value.The double log is consistently the best model for all the loading conditions with the exception the 1.4kN loading thus providing evidence that the double log model better describes the data. Therefore further inferences will be solely based on the double log model.

Table 1: DIC values of the various models

Model	DIC@ 0.6kN	DIC@ 1.1kN	DIC@ 1.2kN	DIC@ 1.4kN	DIC@ 1.4kN2
Linear	18.69	17.89	15.73	14.22	26.14
Exponential	-25.28	-26.06	-28.19	-27.87	-21.15
Quadratic	17.37	18.83	16.09	16.00	27.40
Double Log	-29.00	-26.97	-29.11	-26.33	-22.07

VII. Convergence

Convergence of the parameters can be monitored using the time series plots, autocorrelation plots and Gelman-Rubin statistic. The history plots for the sample of 10000 iterations are displayed in Fig. 3. The figures show that the current model has converged and therefore valid inferences can be made from the results.

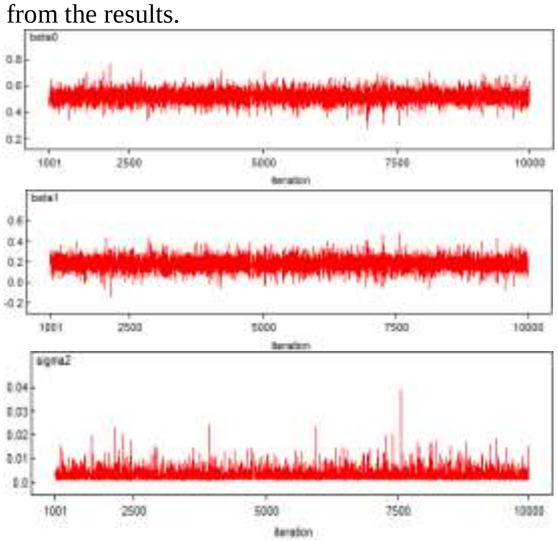

Fig. 3: Trace plot for 0.6kN loading condition

VIII. Posterior distribution and estimates

The posterior density of the parameters of the model are plotted in Fig. 4 while Table 2 reports the posterior mean, standard deviation, MC error and the 95% credible interval for the parameters (beta0, beta1 and sigma2) at the 0.6kN loading condition.The accuracy of the posterior estimates can be determined using the MC error for each parameter. The rule of thumb is that the posterior estimate is accurate if the MC error is less than 5 % of its standard deviation. The MC error for each parameter is indeed less than 5 percent of its standard deviation, thus confirming the accuracy of the posterior estimates.

At 0.6kN maximum loading condition, if the speed increases by 1%, then kurtosis increases by 0.18%. The significance of the effect of speed on kurtosis can be determined by using the standard deviation (analogous to the standard error in frequentist approach). If the mean is greater than twice the standard deviation, then the effect of speed on kurtosis is statistically significant. Alternatively, the t-statistics can be used and this is calculated as the ratio of the mean to the standard deviation. This computed t-value is then compared to the critical value of t obtained from the table of student t distribution. For a two tail-test as in this case: If 1.645<t<1.960 it implies that the effect of speed on kurtosis is significant at 10% level; If 1.960<t<2.576 it implies that the effect of speed on kurtosis is significant at 5% level; If t>= 2.576 it implies that the effect of speed on kurtosis is significant at 1% level. The computed t-value for beta1 at loading condition 0.6kN is 3.18 which implies that the effect of speed on kurtosis is positive and significant at 1% level.

The value of the intercept (beta0) is equal to the value of kurtosis when speed is zero. In this case, the value of kurtosis will be 0.52 at zero speed. The computed t-value for beta0 at loading condition 0.6kN is 12.41 which imply that the intercept is significant at 1% level. All the posterior means are contained in the credible interval covering 95 per cent of the posterior mass.

Table 2: Parameter estimates at loading condition 0.6kN

node	mean	Sd	MC error	2.50%	97.50%
beta0	0.5211	0.04199	4.89E-04	0.4368	0.6058
beta1	0.1804	0.05668	6.86E-04	0.06615	0.2929
sigma2	0.0030	0.00197	2.43E-05	0.00101	0.0081

Fig. 4 is the posterior densities (distributions) for the parameters (beta0, beta1 and sigma2) at the 0.6kN loading condition.

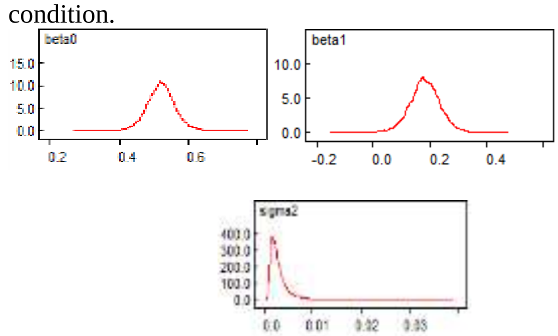

Fig. 4: Posterior densities of the parameter estimates at 0.6Kn

IX. Model prediction

Reasons for making predictions about unknown quantity (eg kurtosis) may include: to “fill in” missing data; replicate datasets in order to check the adequacy of the model; and/or make predictions about the future. Obtaining the appropriate full predictive distribution of the dependent variable could be challenging. One needs to account for three components: uncertainty about the expected future value of kurtosis, the inevitable sampling distribution of kurtosis around its expectation, and the uncertainty about the size of that error, as well as the correlations between these components.

Fortunately, it is so trivial to obtain such predictive distributions using MCMC that it can be dealt with very briefly. The prediction for kurtosis given the estimated and the observed values of speed are shown in Fig. 5 alongside the 95% confidence band. The fit of the model is good as the fitted line falls within the lower and upper bounds. The observed data points are clearly fall on the fitted line as illustrated in Fig. 6. This again confirms the predictive power of the model. The boxplot is also shown in Fig. 7.

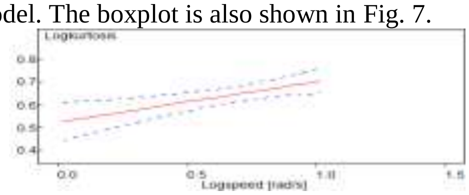

Fig. 5: Model fit at 0.6kN

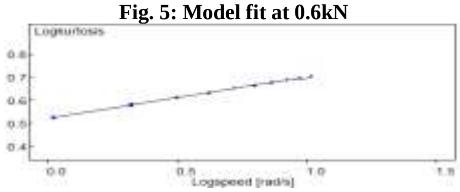

Fig. 6: Model fit at 0.6kN

Scatter plot with the rhombus shaped data points.

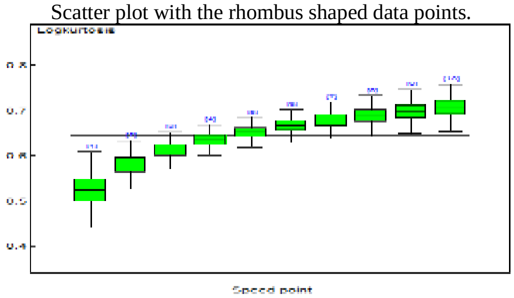

Fig. 7: Model fit at 0.6kN

In order to see the adequacy of the model for future projections, the three new values of speed were added and the unknown values of kurtosis were projected for those new data points. The projected results are shown in Figs. 8, 9 and 10. Again the adequency of the model for future predictions is established.

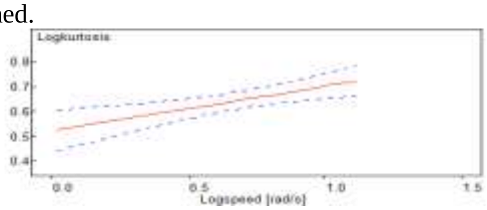

Fig. 8: Model projection at 0.6kN

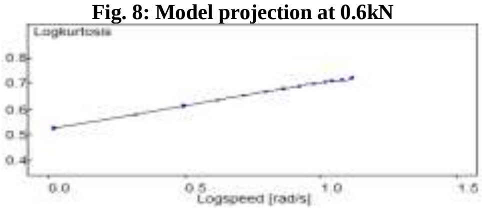

Fig. 9: Model projection at 0.6kN

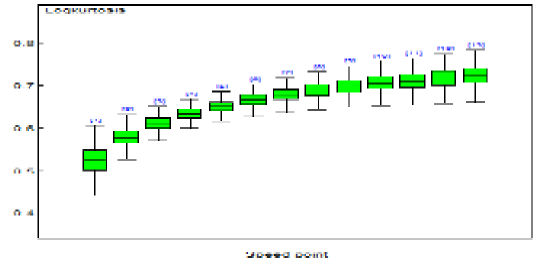

Fig. 10: Boxplot for model projection

X. Conclusion

Preventive maintenance and run to failure techniques of bearing condition monitoring could be quite expensive. The failure could equally be catastrophic thereby leading to the damage of several other components of the machinery. Many bearing studies are done at constant loading and speed. However, this study is carried at varying loads and speed conditions using Bayesian linear regression. This paper therefore presents a condition monitoring methodology based on Bayesian linear regression technique under varying load and speed conditions. Four models namely: linear, quadratic, log linear and double log were fitted to the data. The double log was selected as the best predictive model based on DIC. The results from the model show that speed has significant effect on kurtosis at different loading conditions. The estimated parameters of the model and the observed speed were used to make predictions into the future. The model was able to forecast the current kurtosis values as well as the future values accurately.

References

[1] M. A. Zaidan, A. R. Mills, and R. F. Harrison, "Towards enhanced prognostics with advanced data-driven modelling," Proceedings of the eighth international conference on condition monitoring and machinery failure prevention technologies, Cardiff, 20-22 June, 2011.

[2] D. Zonta, M. Pozzi and P. Zanon, "Bayesian approach to condition monitoring of PRC bridges," Key Engineering Materials, 347, 2007, pp. 227–232.

[3] J. V. Zuffranieri and D. G. Robinson, "A Bayesian approach for health monitoring of critical systems," Proceedings of IEEE Aerospace Conference, Big SKY, MT, 2005, pp. 3641 – 3646.

[4] Z. Zhang, Y. Jiang and J. D. McCalley, Bayesian analysis of power transformer failure rate based on condition monitoring information. State University, research sponsored by Power Systems Engineering Research Center, USA, 2006. http://home.engineering.iastate.edu/~jdm/WebConfPapers/BayesianXfmrNAPS.pdf; Accessed12 March 2012.

[5] T. Heyns, P. S. Heyns and J. P. de Villiers, "A method for real-time condition monitoring of haul roads based on Bayesian parameter estimation," Journal of Terramechanics, 49, 2012a, pp. 103–113.

[6] C. M. Bishop, "Pattern Recognition and Machine Learning Springer," Science and Business Media, New York, USA, 2006.

[7] D. J. Spiegelhalter, N. G. Best, B. P. Carlin and A. van der Linde, "Bayesian measures of model complexity and fit", J.R.Statist. Soc.B., 64, Part 4, 2002, pp. 583-639

[8] H. Jeffreys, "Theory of Probability," Third edition. Oxford University Press, Oxford, England, 1961.

[9] M. P. Norton and D. G. Karczub, "Fundamentals of Noise and Vibration Analysis for Engineers," 2nd Edition, Cambridge University Press, UK, 2003.