IMEKO 22nd TC3, 12th TC5 and 3rd TC22 International Conferences
3 to 5 February, 2014, Cape Town, Republic of South Africa

# SLOW ROTATING BEARING CONDITION ASSESSMENT BASED ON BAYESIAN GAUSSIAN MIXTURE REGRESSION

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

Department of Mechanical & Aeronautical Engineering, University of Pretoria, Pretoria, South Africa,
*sylvester.aye@up.ac.za; or Sylvester.a.aye@gmail.com

**Abstract:** This paper presents the condition monitoring of slowly rotating bearing using experimental data from acoustic emission signal. The condition monitoring methodology is based on a nonlinear parametric Bayesian technique, Gaussian Mixture Regression which is expected to accurately diagnose bearing damage under fluctuating load and speed conditions. The proposed model has the ability to model high dimensional or multi-modal data and retains the flexibility of nonparametric approach. Therefore, the Gaussian Mixture regression method is applied to the condition monitoring of slowly rotating bearing in this study. Results show that the GMR approach is an appropriate, powerful, cost effective and an easy-to-tune regression technique for monitoring and predicting slowly rotating bearing damage under fluctuating speed and loading conditions.

**Keywords:** condition monitoring, Gaussian mixture regression, slow rotating bearings, damage.

## 1. INTRODUCTION

The condition monitoring of bearing has been the subject of extensive research among engineers over the years. This is because bearings are one of the most important components in the vast majority of machines and strong demands are made upon their carrying capacity and reliability. Failure to diagnose damage early can be costly in terms of repair costs, downtimes, and requirements of productivity, profit growth and safety (Stevens, 2011). Vibration analysis has been used in the condition monitoring of bearings for over half a century.

There are several approaches for automatic fault detection and diagnosis which includes fuzzy logic (Liu et al., 1996), evolutionary algorithms (Zhang et al., 2005) and artificial neural networks (Wang and Hope, 2004) among others. Despite their successes, these techniques suffer certain limitations with respect to damage prediction. It is often difficult to obtain a multiclass data covering all aspects of the bearing symptoms that can appear very different and confusing (Chen et al., 2008). Hence, this study uses a class of Bayesian technique called Gaussian Mixture Regression (GMR) model which is robust to high dimensional data to predict slow rotating bearing damage. Instead of modelling the regression function directly, the joint density of the data is modelled using a Gaussian Mixture Model (GMM). GMM allows one to deal with multi-modality in the data and to derive explicit conditional distributions for inference, in the form of GMR (Sung, 2008). The approach is thus expected to be more cost effective than the standard diagnostic techniques for condition monitoring of slow rotating bearings since it avoids the rigor and cost of acquiring more training while at the same time limiting over-fitting.

The GMR approach has been applied recently in a number of mechanical systems damage detection (Marwala et al, 2006; Rasmussen, 2000; Sung, 2008; Calinon, 2009; Wu et al., 2009; Marwala, 2012; Wang et al, 2013). However, its application to monitoring and predicting slow rotating bearings damage is unique to this study. This study therefore presents the case of GMR application to the condition monitoring of slowly rotating bearing using experimental data from acoustic emission signal.

## 2. THE GAUSSIAN MIXTURE REGRESSION METHODOLOGY

The standard single Gaussian distribution has some important analytical properties. However, it suffers from significant limitations when it comes to modelling real data. If a dataset forms more than one dominant clump, the simple Gaussian distribution is unable to capture the structure whereas a linear super position of two or more Gaussians can give a better characterization of the dataset. Such linear characterisation formed by taking linear combination of more basic distributions such as Gaussians, can be formulated as probabilistic models known as mixture distribution (Bishop, 2006).

In this study a Gaussian mixture regression (GMR) is used to predict slow rotating bearing damage extracted from the acoustic emission signal data. Short extracts from the reference signal are extracted. The joint density distributions of these samples are then estimated by means of a Gaussian Mixture Model (GMM). Subsequently, the parameters of the GMM are used to fit a GMR which describes bearing a function of operating conditions such as speed, load, etc. The estimates of the response model parameters are intuitive and robust signal representations which can be interpreted for maintenance support (Heyns et al., 2012).

Assume $X$ represent the vector of the explanatory variables (e.g. operation conditions such as speed, time, load etc). The explanatory variables are those variables which may have impact on the signal characteristics, but which is generally independent of the bearing condition. $Y$ is the vector of the response or dependent variables (e.g. bearing kurtosis extracted from acoustic emission signal), . $x$ is the input training data $(x \in X)$ and $y$ is the output data $(y \in Y)$. For the given $x$ and $y$, the joint probability density is given as (Sung, 2008, Wang et al., 2013).

$$f_{X,Y}(x,y) = \sum_{j=1}^{K} \pi_j \varphi(x,y;\mu_j,\Sigma_j) \quad (1)$$

where $\sum_{j=1}^{K} \pi_j = 1,\ \mu_j = \begin{bmatrix} \mu_{jx} \\ \mu_{jy} \end{bmatrix},\ \Sigma_j = \begin{bmatrix} \Sigma_{jXX} & \Sigma_{jXY} \\ \Sigma_{jYX} & \Sigma_{jYY} \end{bmatrix}$

Here $\varphi(x,y;\mu_j,\Sigma)$ denotes the probability density function (pdf) of the multivariate Gaussian Mixture Model (GMM). Equation (1) shows that the relationship between the explanatory variables and the response variable can be can be described by several GMM models. The parameters of equation (1) include the number of the mixture components, $K$, the priors $\pi_j$, the mean value $\mu_j$, and the variance of each Gaussian component $\Sigma_j$, which are represented as $\theta = (\theta_1, \theta_2, ..., \theta_K)$ with $\theta_j = (\pi_j, \mu_j, \Sigma_j)$ and the constraint $\sum_{j=1}^{K} \pi_j = 1.$

As noted by (Sung, 2008) each Gaussian component can be partitioned and the joint density can be rewritten as

$$f_{X,Y}(x,y) = \sum_{j=1}^{K} \pi_j \varphi(y|x;m_j(x),\sigma_j^2)\varphi(x;\mu_{jX},\Sigma_{jX}) \quad (2)$$

Then marginal probability density of $X$ is

$$f_X(x) = \int f_{X,Y}(x,y)dy = \sum_{j=1}^{K} \pi_j \varphi(x;\mu_{jX},\Sigma_{jX}) \quad (3)$$

The conditional pdf of $(Y|X)$ can deduced by combining equation (1) and (3)

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \sum_{j=1}^{K} w_j(x)\varphi(y;m_j(x),\sigma_j^2) \quad (4)$$

with the mixing weight

$$w_j(x) = \frac{\pi_j \varphi(x;\mu_{jX},\Sigma_{jX})}{\sum_{j=1}^{K} \pi_j \varphi(x;\mu_{jX},\Sigma_{jX})} \quad (5)$$

From equation (4), the regression function for the prediction given a new input is

$$m(x) = E[Y|X = x] = \sum_{j=1}^{K} w_j(x)m_j(x) \quad (6)$$

and the conditional variance function is

$$v(x) = Var[Y|X = x] = \sum_{j=1}^{K} w_j(x)(m_j(x^2) + \sigma_j^2) - (\sum_{j=1}^{K} w_j(x)(m_j(x))^2 \quad (7)$$

where

$$m_j(x) = \mu_{jY} + \Sigma_{jYX} \sum_{jX}^{-1} (x - \mu_{jX}) \quad (8)$$

and

$$\sigma_j^2 = \Sigma_{jYY} - \Sigma_{jYX}\Sigma_{jX}^{-1}\Sigma_{jXY} \quad (9)$$

$m(x)$ in equation (6) is the Gaussian Mixture Regression (GMR) model of index $K$, simply abbreviated as GMR(K) or $m(x;K)$. Although the regression function $m(x)$ from the joint mixture Gaussian density is of the form of a kernel estimator commonly used in nonparametric models, the weight function $w_j(x)$ is not determined by local structure of the data but by the components of a global GMM. Thus the GMR is a global parametric model with nonparametric flexibility (Sung, 2008, Wang et al., 2013).

A major task in fitting the GMR is the estimation of the parameters, $\theta$, of GMM for the joint density $f_{X,Y}$. This can be achieved by maximizing the log likelihood function $L(\theta_k)$ denoted as (Guo and Lyu, 2000)

$$L(\theta_k) = \ln \prod_{i=1}^{N} p(x_i, y_i) = \sum_{i=1}^{N} \ln \sum_{j=1}^{K} \pi_j \varphi(x,y;\mu_j,\Sigma_j) \quad (10)$$

For the given training data, the parameters $\theta$ (comprising the means, covariances and missing coefficients) of a GMM is learnt by maximizing equation (10) using the Expectation Maximization (EM) algorithm in the iterative means. There are some advantages of using EM algorithm. The EM algorithm is simple to implement and understand, avoids the calculation and storage of derivatives, it is usually faster to converge than general purpose algorithms and can also be extended to deal with data sets where some points have missing values (Nabney, 2002).

The EM algorithm includes two steps:

1. E step (expectation step):

Calculate the posterior probability according to

$$p(j|X) = \frac{\pi_j \varphi(X,\mu_j\Sigma_j)}{p(X,\theta)} \quad \text{with } j = 1,2,...k \quad (11)$$

2. M step (maximum step):

$$\pi_j^{New} = \frac{1}{N}\sum_{j=1}^{N} \frac{\pi_j^{old}\varphi(X,\mu_j,\Sigma_j)}{} = \frac{1}{N}\sum_{i=1}^{N} p(j|x_j) \quad (12)$$

$$\mu_j = \frac{1}{\pi_j N}\sum_{i=1}^{N} p(j|x_i)x_i \quad (13)$$

$$\Sigma_j = \frac{1}{\pi_j N}\sum_{i=1}^{N} p(j|x_i)[(x_i - m_j)(x_i - m_j)^T] \quad (14)$$

It is convenient to recast the maximising problem in the equivalent form of minimising the negative log likelihood of the data set (Nabney, 2002):

$$E = -L = -\sum_{j=1}^{K} \log p(y|x) \quad (15)$$

The two steps are iterated until the model converges to a local minimum (Callinon, 2009 and Guo, 2000). The entire data set is divided into training and test sets. The training set is used in estimating the parameters of the GMM while the test set is kept for prediction of bearing damage. The results obtained may be highly sensitive to the number of mixing components used. The more components a mixture model has the more expressive and flexible it becomes. A sufficiently expressive model may be optimized so as to accurately represent the reference signal. However, models which are too expressive may over fit the training data. This may result in poor generalization and subsequently impair the ability of the model to discern between normal signal components and fault related outliers (Bishop, 2006, Heyns et al., 2011). Different numbers of mixing components (K) are fitted and the best is selected. There are several model selection criteria such as the root mean square error (RMSE), the leave-one-out cross validation technique (CV), the percentage prediction error (PE), the Bayesian Information Criterion (BIC), Bayesian model selection, and the Akaike Information Criterion (AIC) among others. In this study the percentage prediction error (PE) was used to select the best number of mixing components. The PE is given by equation (16):

$$PE = \left( \frac{K_a - K_p}{K_a} \right) \times 100\% \tag{16}$$

Given the test set, the GMR models can be obtained using the parameters of the GMM which outputs a smooth generalized version of the data encoded in GMM and associated constraints expressed by covariance matrices (Callinon, 2009). This smooth generalized version of the data is the bearing failure prediction. The schematic chart of the GMR showing the modelling procedures as described is shown in Figure 1.

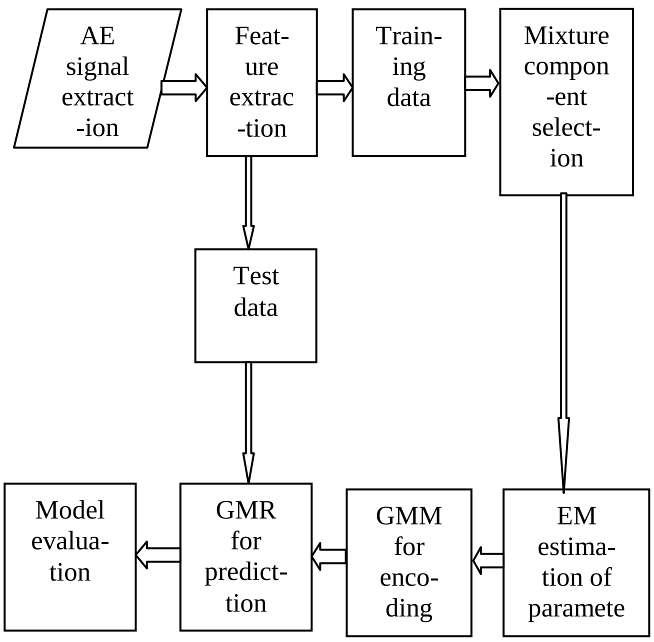


Figure 1: GMR procedure

## 3. RESULTS AND DISCUSSION

The performance of the GMR model is evaluated using Prediction error (PE) as estimated from the leave-one-out cross-validation (cv) technique. Cross-validation is a model validation technique for assessing the performance of the model robustness i.e. how the results from the statistical analysis (GPR model) can be generalised to an independent data set. It is used in settings where the goal is prediction as in this study in order to estimate how accurately a predictive model will perform in practice. Cross-validation is implemented by partitioning the original bearing data two segments of equal length, performing the analysis on one segment (called the training set) and validating the analysis on the other segment (called the testing or validation set). Each model is retrained three times using the EM algorithm. The prediction error as computed from different number of mixing of components based on the validation set is presented in figure 2. The best performing model is one with the minimum PE. The slope of figure 2 remains almost constant until when the number of mixture components, K, is 6. Thus, the GMR (6) is selected as the best model for the bearing data. When more than 6 components are used the performance on the validation set reduces indicating that the models are beginning to over-fit the training set. Subsequently, the GMR prediction analysis are performed by conditioning the slow rotating bearing acoustic emission signal extracts on speed and load using 6 mixing components.

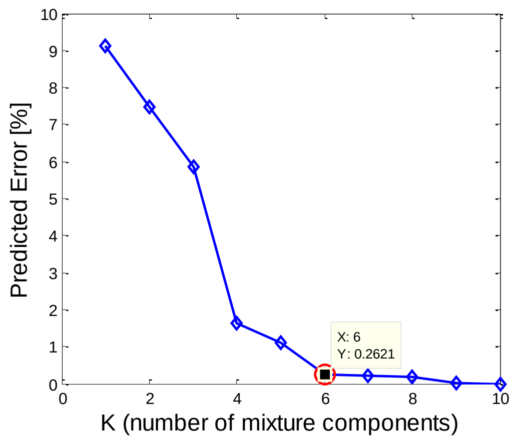


Figure 2: Model performance for different number of mixing components’

The encoding results of the GMM for the four loading conditions are illustrated in figure 3 in which the speed is plotted against kurtosis values. From the figure it can be seen that the position of each GMM model are arranged automatically to adapt to the spatial distribution of the training samples. For all loading conditions, four out of the six Gaussian components are large while 2 are relatively small. The GMMs with 6 Gaussian components are found to efficiently encode the slowly rotating bearing data preparatory to implementing the GMR.

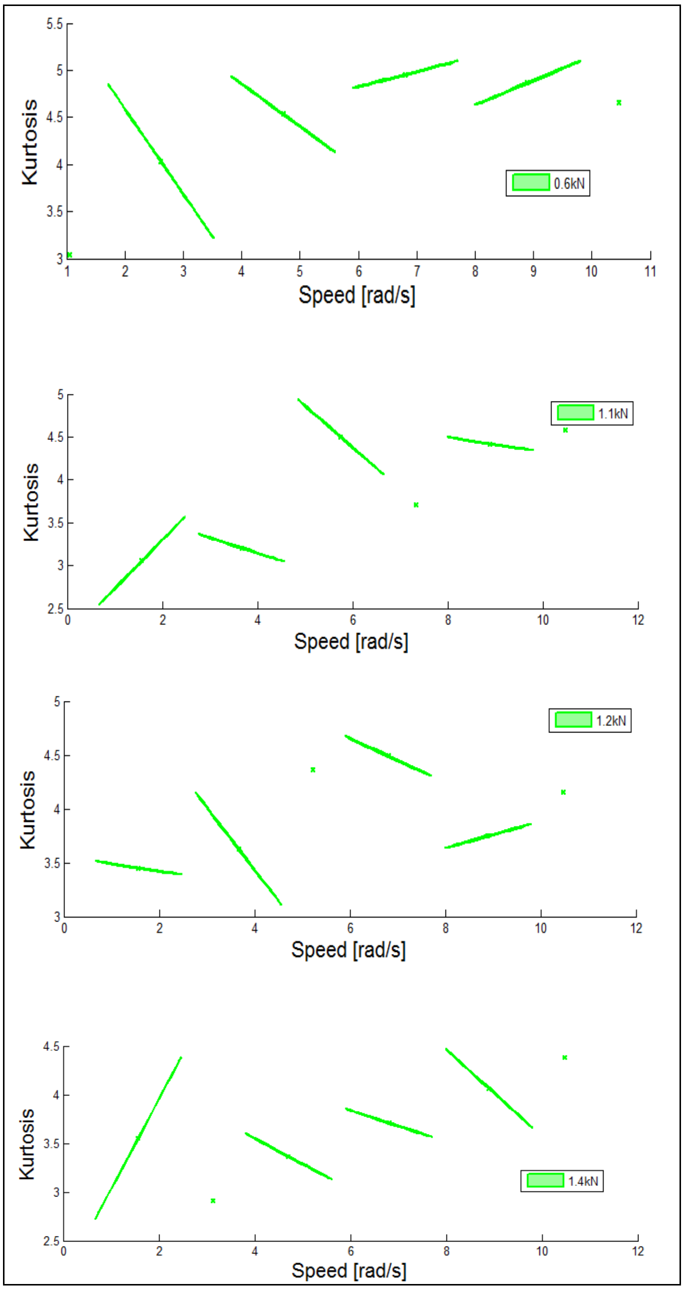


Figure 3: Encoding results of GMM

The effectiveness of the GMR to detect incipient damage was further investigated. Figure 4 is a plot of the GMR predicted damage kurtosis values of slowly rotating bearing for 0.6kN loading condition. There is a near perfect fit between the predicted and actual damage values as shown in figure 4. Figure 5 is the GMR posterior density plot of the slowly rotating bearing for 0.6kN loading condition. The kernal density of the actual and the predicted again shows a near perfect fit. From the posterior density plot it can be seen that the predicated kurtosis has a mean value of 4.7 just as the actual. This shows that truely the bearing is damaged since the predicted mean kurtosis value is above 3.

Figures 6, 8 and 10 are also the plots of the GMR predicted damage kurtosis values of slowly rotating bearing for 1.1kN, 1.2kN and 1.4kN loading conditions respectively. Again, there is a near perfect fit between the predicted and actual damage values as shown in figure 6, 8 and 10. Figures 7, 9 and 11 are the GMR posterior density plots of the slowly rotating bearing for the 1.1kN, 1.2kN and 1.4kN loading conditions. The kernal densities of the actual and the predicted values again shows a near perfect fit. From the posterior density plots it can be seen that the predicated kurtosis have mean values of 4, 3.8 and 3.6 just as the actual for the 1.1kN, 1.2kN and 1.4kN loading conditions respectively. This shows that truely the bearing is damaged since the predicted mean kurtosis values are all above 3.

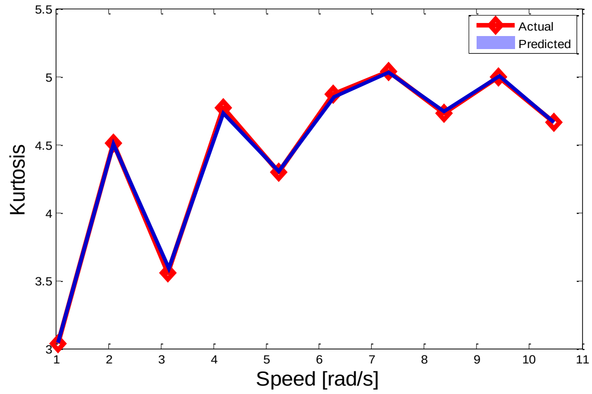


Figure 4: GMR posterior mean prediction at 0.6kN loading condition

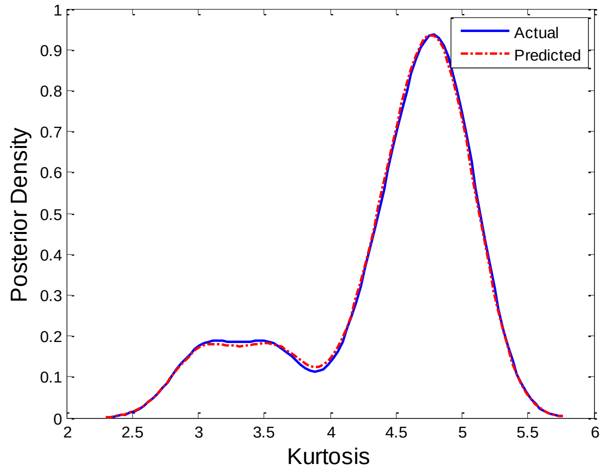


Figure 5: GMR posterior distribution (kernel density) prediction of at 0.6kN loading condition

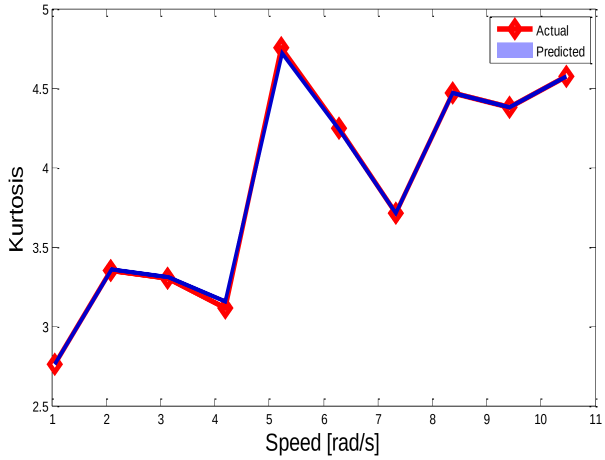

Figure 6: GMR mean prediction at 1.1kN loading condition

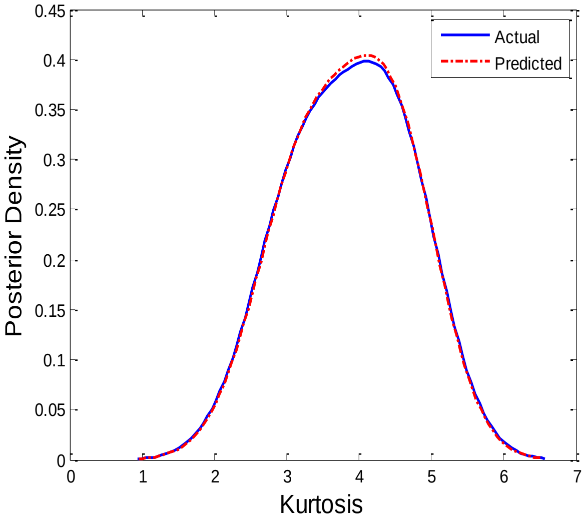

Figure 7: GMR posterior distribution (kernel density) prediction at 1.1kN loading condition

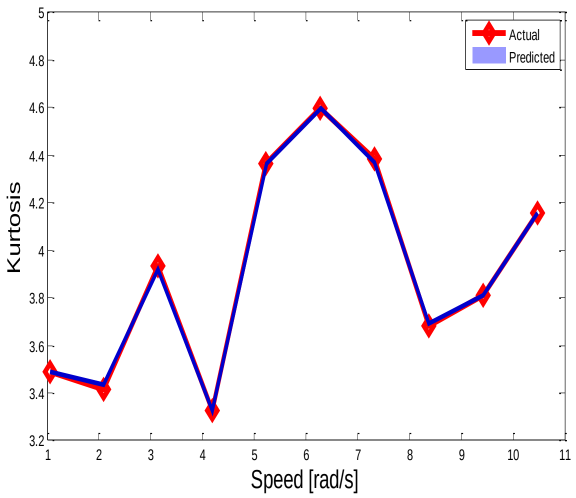

Figure 8: GMR mean prediction at 1.2kN loading condition

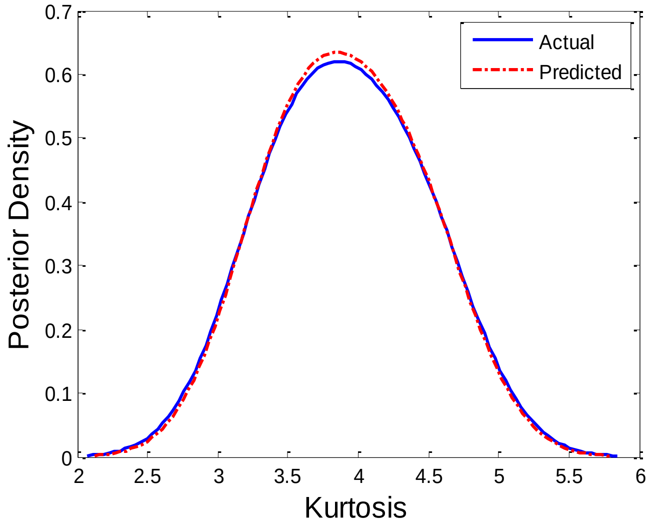

Figure 9: GMR posterior distribution (kernel density) at 1.2kN loading condition

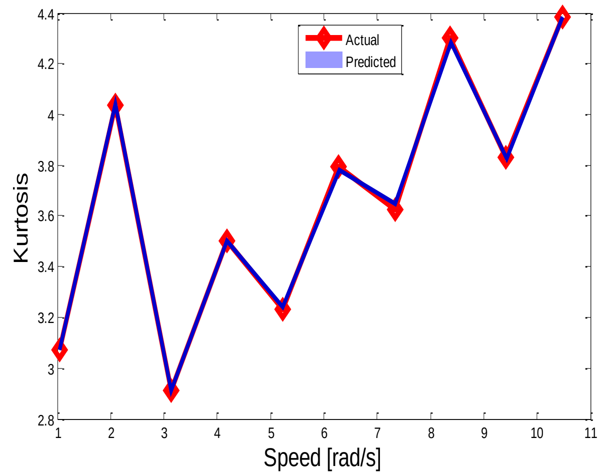

Figure 10: GMR mean prediction at 1.4kN loading condition

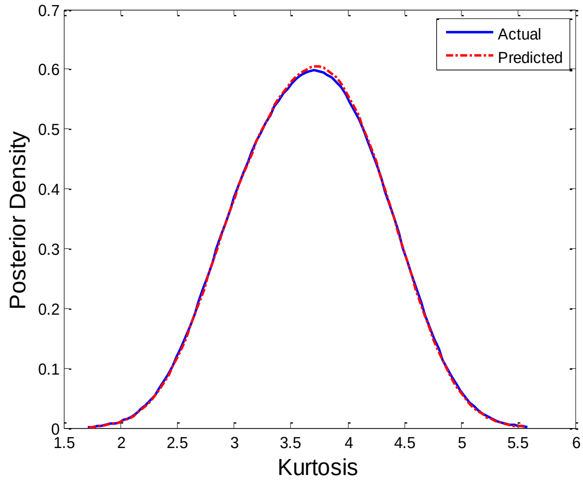

Figure 11: GMR posterior distribution (kernel density) at 1.4kN loading condition for GMR

Prediction Error was employed in other to verify the accuracy of the GMR to predict the slowly rotating bearing damage. The prediction error was computed using equation

16. The predicted data for the 4 loading conditions were computed at various speeds ranging from 1-10 rad/s and average PE computed for each loading condition. The Prediction error was subsequently computed and the bar chart in figure 12 obtained. The PE values were 0.24%, 0.26%, 0.16%, and 0.20% for the various loading conditions at 0.6kN, 1.1kN, 1.2kN and 1.4kN. The prediction error values are very small showing that the GMR predicted damage very accurately.

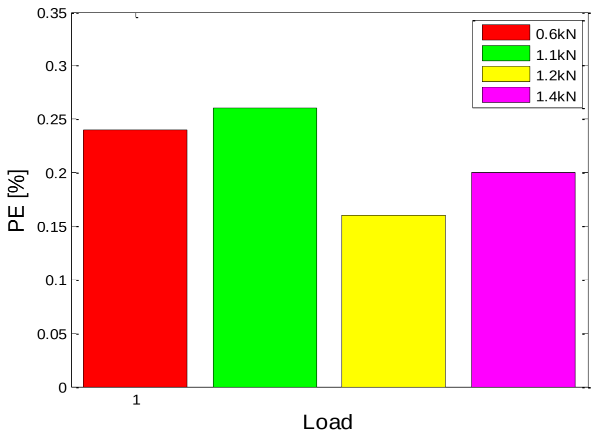

Figure 12: Gaussian mixture prediction error under different loading conditions

Overall based on the GMR posterior mean and posterior density compared the original signal extract, it is obvious that the proposed methodology can accurately diagnose slow rotating bearing damage under fluctuating load and speed conditions.

## 4. CONCLUSIONS

The majority rotating-machine failures are frequently related to bearing failures. Appropriate condition monitoring on bearings is therefore important to lessen the duration of equipment down-times and related costs. In this study the Gaussian Mixture Regression (GMR) model is applied to the condition monitoring of slowly rotating bearing using features extracted from acoustic emission signal. The GMR builds a global parametric model so that the fluctuation of the extracted features is modelled as a variance of the related GMM model. Unlike the traditional condition monitoring techniques, the GMR models multi-modal data more effectively without requiring huge acquisition of training data and also avoids over fitting the model. The proposed GMR model was used to predict the damage of slow rotating bearing under fluctuating load and speed operating conditions. The performance of the GMR models was evaluated using Prediction error (PE). A GMR with six mixing components was selected for further analysis. The effectiveness of the GMR to detect incipient damage was investigated and it was demonstrated that the GMR predicted damage kurtosis values of slowly rotating bearing perfectly for all loading conditions. The proposed methodology accurately diagnoses slow rotating bearing damage under fluctuating load and speed conditions. Hence, the GMR approach is found to be a powerful, cost effective and an easy-to-tune regression technique for monitoring and predicting slowly rotating bearing damage.

## 5. REFERENCES

[1] Stevens, (2011). “Bearing Failures & Their Causes”, David Stevens IEng-Bearing failures - Introduction to bearing failures, http://www.vibanalysis.co.uk/technical/contents.html; Accessed on 2011-10-22

[2] Liu T.I., Singonahalli J.H., Iyer N.R., (1996). “Detection of roller bearing defects using expert system and fuzzy logic”. Mechanical Systems and Signal Processing, 10(5): 595-614.

[3] Zhang L., Jack L.B., Nandi A.K. (2005). “Fault detection using genetic programming”. Mechanical Systems and Signal Processing, 19:271-89.

[4] Wang L., Hope A.D. (2004). “Bearing fault diagnosis using multi-layer neural networks”. Insight: Non-Destructive Testing and Condition Monitoring, 46(8): 595-614.

[5] Chen S.L., Craig M., Wood R.J.K., Wang L., Callan R., Powrie H.E.G. (2008). “Bearing Condition Monitoring Using Multiple Sensors and Integrated Data Fusion Techniques”, Paper of VIRM08.

[6] Sung HG (2008). “Gaussian mixture regression and classification”. Dissertation, Rice University.

[7] Marwala T., Mahola U. and Nelwamondo F. V. (2006) Hidden Markov Models and Gaussian Mixture Models for Bearing Fault Detection Using Fractals, International Joint Conference on Neural Networks Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada, July 16-21, 2006

[8] Rasmussen C.E., (2000). The Infinite Gaussian Mixture Model, Advances in Neural Information Processing Systems 12

[9] Calinon S., (2009). Robot Programming by Demonstration: A Probabilistic Approach. 1st Edition, École Polytechnique Fédérale de Lausanne (EPFL) press, Switzerland.

[10] Wang L., Hope A.D. (2004). Bearing fault diagnosis using multi-layer neural networks. Insight: Non-Destructive Testing and Condition Monitoring, 46(8): 595-614.

[11] Wu, S. Li, Y.F., Zhang J. (2009). Probabilistic Cluster Signature for Modelling Motion Classes. IEER/RSJ International Conference on Intelligent Robots and Systems, October 11-15, 2009, St. Louis, USA

[12] Marwala T., (2012). Gaussian Mixture Models and Hidden Markov Models for Condition Monitoring, Condition Monitoring Using Computational Intelligence Methods 2012, pp 111-130

[13] Wang G, Qian L., Guo Z. (2013). Continuous tool wear prediction based on Gaussian mixture regression model. International Journal of Advanced Manufacturing Technology, June 2013, volume 66, issue 9-12, pp 1921-1929.

[14] Bishop, C.M. (2006). “Pattern Recognition and Machine Learning”. Springer Science and Business Media, New York, USA.

[15] Heyns, T., Heyns, P.S. and Villiers, J.P. (2012). “Combining synchronous averaging with a Gaussian mixture model novelty detection scheme for vibration-based condition monitoring of a gearbox”. Mechanical Systems and Signal Processing, 32, 200-215.

[16] Guo, P. and Lyu, M. (2000) Software quality prediction using mixture models with EM algorithm. The first Asia-Pacific Conference on Quality Software, Hong Kong, China, 69-78.

[17] Nabney, I.T. (2002). NETLAB algorithms for pattern recognition. Springer Publication, Great Britain.