

# Acoustic emission-based prognostics of slow rotating bearing using Bayesian techniques under dependent and independent samples

S.A. Aye<sup>a</sup> and P.S. Heyns

Department of Mechanical & Aeronautical Engineering, University of Pretoria, 0002,  
Pretoria, South Africa

<sup>a</sup>[Sylvester.Aye@up.ac.za](mailto:Sylvester.Aye@up.ac.za)

## Abstract

This study develops a novel degradation assessment index (DAI) from acoustic emission signal obtained from slow rotating bearings and integrates same into alternative Bayesian methods for the prediction of remaining useful life (RUL). The DAI is obtained by the integration of polynomial kernel principal component analysis (PKPCA), Gaussian mixture model (GMM) and exponentially weighted moving average (EWMA). The DAI is then used as inputs in several Bayesian regression models such as the multi-layer perceptron (MLP), radial basis function (RBF), Bayesian linear regression (BLR), Gaussian mixture regression (GMR) and the Gaussian process regression (GPR) for RUL prediction. The combination of the DAI with the GPR model, otherwise, known as the DAI-GPR gave the best prediction with the least error. The findings show that the GPR model is suitable and effective in the prediction of RUL of slow rotating bearings and robust to varying operating conditions. Further, the findings are also robust when the training and tests sets are obtained from dependent and independent samples. Therefore, the GPR model is found useful for monitoring the condition of machines in order to implement effective preventive rather than reactive maintenance, thereby maximizing safety and asset availability.

Keywords: degradation assessment index, diagnostics, prognostics, remaining using life, Bayesian models

## Introduction

The act of predicting or forecasting a fault before it occurs (prior event analysis) is called prognosis. The prediction of defects refers to the determination of imminence of the fault and an estimation of how soon a defect will likely occur. Once the current health condition is defined, the next task is to predict the change in component health as a function of remaining useful life (RUL) based on anticipated future missions (Camcia *et al.*, 2012). Jardine *et al.* (2006) defined the RUL as a conditional random variable of the time left before observing a failure given the current machine age and condition and past operation profile.

According to Marble and Morton (2005), bearing prognosis is the key to maximizing safety and asset availability while minimizing logistical costs, by allowing maintenance to be proactive rather than reactive. However, Jardine *et al.* (2006) noted that although prognosis is much more efficient than diagnosis to achieve zero-downtime performance, diagnosis is required when fault prediction of prognosis fails and a fault occurs. Therefore, both diagnosis and prognosis is very important aspects that need to be pursued concurrently. Prognosis ensures that maintenance is carried out at the most appropriate time after damage detection without impairing the safety requirements since this is vital for effective operation and

management (Camcia *et al.*, 2012). The traditional approach of detecting bearing damage and failure (for example manual inspection of defect size after every machine operation) is labour intensive and forces machinery to shut down, thus causing tremendous time, productivity and capital loss (Bolander *et al.*, 2009; Camcia *et al.*, 2012). Therefore, it would be highly beneficial to be able to predict expected remaining bearing life with a large degree of certainty. Recent advances in sensor technology and computational intelligence have made real-time bearing prognosis feasible. Whereas there is large number of studies on bearing diagnosis, the extension to prognosis is limited. Hence, this study integrates both diagnosis and prognosis (RUL prediction) aspects of condition monitoring of slow rotating bearing using whole life bearing data from a lab experiment.

There is no doubt that prognosis is surrounded by uncertainties arising from a variety of sources such as the current age of the asset or mechanical system, the health information or observed condition monitoring, measurement noise, process noise, modelling uncertainty and the environment in which the system is operated (Saxena, 2010; Si, *et al.*, 2011), which makes the process inherently stochastic. Therefore, the behaviour observed from a particular run may not exhibit the true nature of prediction trajectories. It is therefore expected that a prognosis algorithm should provide information about the confidence around the prediction (Saxena *et al.*, 2009). Bayesian techniques which are mainly statistical are gaining widespread application in damage detection and remaining useful life of bearings due to their ability to handle uncertainties as opposed to traditional statistical methods (Nabney, 2002). This feature is useful for risk analysis and maintenance decision making (Si, *et al.*, 2011). Bayesian methods have ability to not only obtain point estimates for the variables of interest but also their probability distributions and permit the researcher to characterise the uncertainty about the parameter values using confidence intervals (Hippert, and Taylor 2010).

Prediction of RUL is often difficulty as the results depend on the models used in obtaining them. Therefore, it is important to evaluate the predictions from alternative models and choose the best based on an objective criterion. A model is deemed superior if it effectively minimizes the one-step-ahead (or multi-step-ahead) prediction errors by producing a lower prediction error than its competitors. Against this background, this study evaluates the performance of alternative Bayesian methods for slow rotating bearing fault prognosis based on acoustic emission data obtained from a run-to-failure experiment.

A newly developed degradation assessment index (DAI) was used as an input in several Bayesian regression models such as the multi-layer perceptron (MLP), radial basis function (RBF), Bayesian linear regression (BLR), Gaussian mixture regression (GMR) and the Gaussian process regression (GPR) for RUL prediction. The DAI incorporated all the advantages of the various extracted features (kurtosis, peaktopeak, RMS, skewness, crest factor) capitalizing on the strengths of each, and thereby becoming more sensitive and robust in prognosis, while at the same time reducing the number of dimensions for condition monitoring (Malhi and Gao, 2004). The mean absolute percentage error (MAPE) and root mean square error (RMSE) were used in evaluating the performance of the models and hence selecting the best performing model for the prediction of slow rotating bearing remaining useful life.

A number of studies have employed at least one of these models for prognostics. Examples include Nabney, 2002; Gebraeel *et al.*, 2004; Rasmussen and Williams, 2006; Skabar, 2007; Chen and Ren, 2009; Saxena *et al.*, 2009; Hippert, and Taylor, 2010; Wang and Wang, 2012; Hong and Zhou, 2012a, 2012b; Liu *et al.*, 2012; Calinon, 2009; Chatzis *et al.* 2012 amongst others. This study is by no means the first to evaluate the performance of alternative models for RUL prediction of mechanical and allied systems. For instance, Hong and Zhou (2012b) evaluated the performance of GPR and wavelet neural network (WNN) for prediction of bearing RUL and found GPR to show more excellent features than WNN with

GPR predicting faster and having more stable prediction as well as lower prediction error in general than WNN. Goebel *et al.* (2008) compared the performance of GPR, RVM and NN-based approach for prognostics of aerospace rotating equipment. GPR seemed to have performed better than NN and RVM especially for the algorithm with specific damage estimates given the small GPR error though with late predictions than RVM. Saha *et al.* (2009) evaluated the performance of particle filter-based, autoregressive integrated moving average (ARIMA) and extended kalman filter (EKF) models and found that the particle filter framework has significant advantages over ARIMA and EKF for predicting the remaining useful life of batteries. An *et al.* (2012) compared the performance of the particle filter (PF), the overall Bayesian method (OBM), and the recursive Bayesian method (RBM) and found that the performance of PF and OBM differ depending on the stage of the damage state and hence should be used as complementary models. Chatziz *et al.* (2012) compared the performance of the standard GMR, Dirichlet process GMR and GPR and found the first two to be more suited for robot prognosis than GPR. An *et al.* (2013) compared NN and GPR under different levels of noise and found that GPR under a no noise case (perfect data), GPR show exact result and outperform NN under small noise while under large noise NN outperform GPR.

This study contributes to literature on prognostics by evaluating the performance of the MLP, RBF, BLR, GMR and GPR models using the same data set. There is no known study that has evaluated the performance of these set of models using the same data set. Further, this study also makes contribution to prognostics literature by evaluating the performance of these models under a leave-out-one cross validation approach that is based on two types of samples or data set namely dependent and independent samples. This dependent and independent samples scenario is important because it has been theoretically argued that cross validations based on training and test sets which are from the same sample (dependent) may break down and lead to overfitting in nonparametric methods since the errors may be positively correlated (Opsomer *et al.*, 2001; Arlot, 2010).

## **Methodology**

This section describes the different models used in developing the proposed approach to prognosis of slow rotating bearings. The entire condition monitoring process is implemented in a unified framework (figure 1). Generally the process involves four main steps: (1) Obtaining the degradation assessment index; (2) Using the obtained DAI as input into MLP, RBF, BLR, GMR and GPR respectively; (3) The various models are then used for the remaining useful life prediction of slow rotating bearings. (4) The five obtained models are evaluated to find out which of them gave the best prediction.

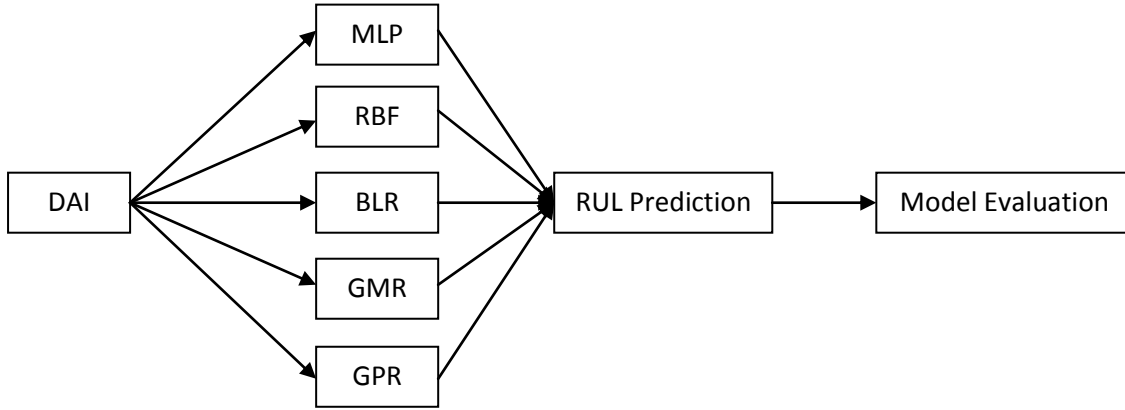


Figure 1: Framework for DAI integrated approach to bearing prognostics

Key: DAI-degradation assessment index; MLP-multi-layer perceptron; RBF-radial basis function; BLR-Bayesian linear regression; GMR-Gaussian mixture regression; GPR-Gaussian process regression; RUL-remaining useful life

### ***A degradation assessment index***

Providing a quantifying degradation indication for the assessment of machine performance is the vital for prognosis. The degradation assessment index is obtained by the combination of the polynomial kernel principal component analysis (PKPCA), the Gaussian mixture model (GMM), and weighted moving average (EWMA). The PKPCA is a widely used technique for dimensionality reduction and feature extraction in machine learning (Schölkopf *et al.*, 1998; 1999; Lee 2004). Subsequent to feature extraction by PKPCA from the high-dimensional statistics, the nonlinear (multimodal) features in low-dimensional data space can still be preserved. The polynomial kernel principal component henceforth, PKPC, which were extracted are then used as inputs in the GMM which is an outstanding technique of complex data description, with benefits of high-performance computation and robustness. The GMM describes complex data distribution that often occurs in acoustic emission data by outputting the negative log likelihoods (NLL) utilizing numerous Gaussian components. The reliability and sensitivity of the NLL to the bearings slight degradation was improved by employing the exponentially weighted moving average (EWMA) statistic as an improved quantification index for prognostics of slow rotating bearing. The proposed EWMA is a kind of infinite impulse response filter applying exponentially decreasing weighting factors. Each older datum points weighting never reaches zero, decreasing exponentially. The resulting quantification index is named degradation assessment index (DAI). Therefore, in this study the DAI developed is used as a bearing degradation index in prognostics of slow rotating bearings.

### ***Multi-layer perceptron regression***

The DAI is used as input into multi-layer perceptron network (MLP) for the prognostics of slow rotating bearing in order to determine its remaining useful life (RUL). MLP is one of the generally utilized architecture for empirical usage of neural networks. It more often than not comprise of basically two layers of adaptive weights. There is a complete linkage connecting the inputs to the hidden units, as well as another connecting the hidden units to the output units (Nabney, 2002).

The MLP is a mathematical function which has been parameterized by a set of numerical weights  $w_1, w_2, \dots, w_n$ , which we shall represented jointly by a vector of weight

$w$ . The Bayesian technique entails inference of the posterior distribution of weights,  $p(w|D)$ , given data  $D$ . It considers a functional probability distribution over the weighting space. The outputting prediction resulting from the input vector  $x$  is then gotten by implementing a weighted sum of the predictions over all possible weight vectors, where the weighting coefficient for a particular weight vector is dependent on the posterior weight distribution. The predicted value is given as (Skabar, 2007):

$$\hat{y}^n = \int f(x^n, w)p(w|D)dw \quad (1)$$

where  $f(x^n, w)$  is the MLP output, and  $\hat{y}^n$  is the predicted value.

The probability density function,  $p(w|D)$ , can be approximated using the fact that  $p(w|D) \propto p(D|w)p(w)$ , where  $p(D|w)$  and  $p(w)$  are known respectively as the likelihood and prior.

The prior weight distribution,  $p(w)$ , is the weight distribution before the observation of any data reflecting the prior knowledge of the MLP complexity. To obtain a smooth function for the reduction of the risk of over fitting,  $p(w)$  is assumed to be Gaussian with zero mean and inverse square variance  $\alpha$  which gives:

$$p(w) = \left(\frac{\alpha}{2\pi}\right)^{m/2} \exp\left(-\frac{\alpha}{2} \sum_{i=1}^m w_i^2\right), \quad (2)$$

Where  $m$  is the number of weights in the MLP. Because  $\alpha$  controls the value of other parameters (ie the weights) it is referred to as a hyperparameter.

Since the prior depends on  $\alpha$ , the modification of equation (1) with the inclusion of the posterior distribution over parameters of the hyperparameters gives the predicted value as:

$$\hat{y}^n = \int \int f(x^n, w)p(w, \alpha|D)dw d\alpha \quad (3)$$

where

$$p(w, \alpha|D) \propto p(D|w)p(w|\alpha)p(\alpha) \quad (4)$$

### ***Radial basis function regression***

Similarly, the DAI is used as input into radial basis function (RBF) regression for the prognostics of slow rotating bearing in order to determine its remaining useful life (RUL). The RBF is used for non-linear modelling. The RBF has many advantages. One of which is that it has a two phase training process which is significantly quicker than MLP. In the first phase, the parameters of the basis functions are set to model the unconditional data density. In the second stage of training, the weights in the output layer are determined. Secondly, it is possible to assign an interpretation to the hidden units and also to determine the intrinsic degrees of freedom of the network (Nabney, 2002). The RBF network mapping could be written in the following form as shown in equation 5.

$$y_k(x) = \sum_{j=1}^M w_{kj} \phi_j(x) + w_{k0} \quad (5)$$

where  $\phi_j$  are the basis functions,  $w_{kj}$  are the output layer weights.

The bias weights can be absorbed into the summation by including an extra basis function  $\phi_0$  whose activation is constant value 1. This leads to equation 6.

$$y_k(x) = \sum_{j=0}^M w_{kj} \phi_j(x) \quad (6)$$

Two Bayesian approaches have been found to be effective in practice to neural networks namely: Gaussian approximation to the posterior weight distribution in the weight space (known as Laplace approximation) often coupled with use of the evidence procedure for optimal hyperparameter estimation; secondly the Monte Carlo techniques, particularly the hybrid Monte Carlo (Nabney, 2002).

### ***Bayesian linear regression***

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 (Bishop, 2006). The Gaussian, also known as the normal distribution, is a widely used model for the distribution of continuous variables (Bishop, 2006). For the case of a single real-valued variable,  $x$ , the Gaussian distribution is given as:

$$N(x|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \quad (7)$$

where  $\mu$ , known as the mean, and  $\sigma^2$ , is known as the variance. The reciprocal variance is referred to as precision is defined by:

$$\tau = 1/\sigma^2 \quad (8)$$

In this study acoustic emission signal is extracted at different operational conditions (speeds, and dynamic loading conditions). A regression function, which measures the bearing vibration as a function of the different operating conditions is fitted. The regression function is approximated based on the parameter prior and the data driven likelihood. The prior indicates the characteristic nature of the functions of interpolation. As such the prior allows for more vigorous interpolation functions, particularly if only noisy and limited data are obtainable.

An observation  $y_i^j$  is given as the summation of the specific loading condition function as computed for the equivalent operational condition vector  $f(x_i^j)$  and the noise term  $e_i$ .

$$y_i^j = f(x_i^j) + e_i \quad (9)$$

The function of interpolation could be taken to an approximate linear dependency on  $x$  if the operating conditional vector is adequately expressive it may be necessary to make the assumption. The linearly dependent function is given by the parameter vector  $w^j$  :

$$f(x^j) = \{x^j\}^T w^j \quad (10)$$

For the LSE solutions for the reference loading condition, a multivariate Gaussian distribution is approximated. This distribution is consequently utilized as the prior distribution  $p(w)$ . Let the prior mean be taken as vector  $\mu_o$ , and let the covariance matrix be taken as  $\Sigma_o$ . Hence, the prior is given as:

$$p(w) = N(w|\mu_o, \Sigma_o) \quad (11)$$

Based on Bayes' theorem, the prior and the data determined likelihood are utilized in obtaining a posterior distribution for the values of parameters:

$$\text{Posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}} \quad (12)$$

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

where the posterior is normalized by the marginal likelihood  $p(y|X)$ . Prior probability is the probability available before the observation. However, posterior probability is the probability obtained after the observation. The likelihood function shows the possibility of the data set observed for the settings of the vector of parameters. The posterior distribution could equally be demonstrated to be a Gaussian distribution (Bishop, 2006):

$$p(w^j|y^j, X^j) = N(w^j|\mu_e^j, \Sigma_e^j) \quad (13)$$

where the posterior mean  $\mu_e$  and covariance  $\Sigma_e$  for loading condition  $j$  is given by:

$$\mu_e^j = \Sigma_e(\Sigma_o^{-1}\mu_o + \tau_e \{X^j\}^T X^j) \quad (14)$$

$$\Sigma_e^j = (\Sigma_o^{-1} + \tau_e \{X^j\}^T X^j)^{-1} \quad (15)$$

The likelihood of the observation of a DAI value  $y_*$  at an operational condition  $x_*^j$  while traversing bearing time interval  $j$  may be obtained from the recomputed likelihood function and is a type of a Gaussian (Bishop, 2006):

$$p(y_*|y^j, X^j, \mu_o, \Sigma_o, \sigma_e) = N(y_*|w_e^T x_*^j, \{\sigma_*^j\}^2) \quad (16)$$

The variance  $\{\sigma_*^j\}^2$  of the predictive distribution is indicative of the uncertainty in the prediction at an operational condition  $x_*^j$  defined as:

$$\{\sigma_*^j\}^2 = \sigma_e^{-2} + \{x_*^j\}^T \Sigma_n x_*^j \quad (17)$$

### ***Gaussian mixture regression***

In spite of the vital analytical properties of the standard single Gaussian distribution it has some considerable limits in real data modelling. If a dataset forms more than one dominant clump, the basic Gaussian distribution is incapable of capturing the structure whilst the linear superposition of two or more Gaussians can give an improved description 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 remaining useful life from the DAI. 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. the DAI developed from extracted bearing features obtained 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 (Wang et al., 2013).

$$f_{x,y}(x, y) = \sum_{j=1}^K \pi_j \varphi(x, y; \mu_j, \Sigma_j) \quad (18)$$

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

The probability density function (pdf) of the multivariate GMM is denoted by  $\varphi(x, y; \mu_j, \Sigma)$ . Equation (18) shows that the relationship between the explanatory variables and the response variable can be described by several GMM models. The parameters of equation (18) 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

$$\theta = (\theta_1, \theta_2, \dots, \theta_K) \text{ with } \theta_j = (\pi_j, \mu_j, \Sigma_j) \text{ and the constraint } \sum_{j=1}^K \pi_j = 1.$$

As noted by each Gaussian component can be partitioned and the joint density can be rewritten as

$$f_{x,y}(x, y) = \sum_{j=1}^K \pi_j \varphi(y|x; m_j(x), \sigma_j^2) \varphi(x; \mu_{jX}, \Sigma_{jX}) \quad (19)$$

Then marginal probability density of  $X$  is



$$f_X(x) = \int f_{X,Y}(x,y)dy = \sum_{j=1}^K \pi_j \varphi(x; \mu_{jX}, \Sigma_{jX}) \quad (20)$$

The conditional pdf of  $(Y|X)$  can deduced by combining equation (19) and (20)

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \sum_{j=1}^K w_j(x) \varphi(y; m_j(x), \sigma_j^2) \quad (21)$$

with the mixing weight

$$w_j(x) = \frac{\pi_j \varphi(x; \mu_{jX}, \Sigma_{jX})}{\sum_{j=1}^K \pi_j \varphi(x; \mu_{jX}, \Sigma_{jX})} \quad (22)$$

From equation (22), the regression function for the prediction given a new input is

$$m(x) = E[Y|X = x] = \sum_{j=1}^K w_j(x) m_j(x) \quad (23)$$

and the conditional variance function is

$$v(x) = \text{Var}[Y|X = x] = \sum_{j=1}^K w_j(x) (m_j(x)^2 + \sigma_j^2) - \left( \sum_{j=1}^K w_j(x) m_j(x) \right)^2 \quad (24)$$

Where

$$m_j(x) = \mu_{jY} + \Sigma_{jYX} \Sigma_{jX}^{-1} (x - \mu_{jX}) \quad (25)$$

and

$$\sigma_j^2 = \Sigma_{jYY} - \Sigma_{jYX} \Sigma_{jX}^{-1} \Sigma_{jXY} \quad (26)$$

$m(x)$  in equation (23) is the 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 (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

$$L(\theta_k) = \ln \prod_{i=1}^N p(x_i, y_i) = \sum_{i=1}^N \ln \sum_{j=1}^K \pi_j \varphi(x, y; \mu_j, \Sigma_j) \quad (27)$$

For the given training data, the parameters  $\theta$  (comprising the means, covariances and missing coefficients) of a GMM is learnt by maximizing equation (27) 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

$$p(j|X) = \frac{\pi_j \varphi(X, \mu_j, \Sigma_j)}{p(X, \theta)} \quad \text{with } j = 1, 2, \dots, k \quad (28)$$

2. M step (maximum step):

$$\pi_j^{New} = \frac{1}{N} \sum_{j=1}^N \frac{\pi_j^{old} \varphi(X, \mu_j, \Sigma_j)}{p(X, \theta)} = \frac{1}{N} \sum_{i=1}^N p(j|x_i) \quad (29)$$

$$\mu_j = \frac{1}{\pi_j N} \sum_{i=1}^N p(j|x_i) x_i \quad (30)$$

$$\Sigma_j = \frac{1}{\pi_j N} \sum_{i=1}^N p(j|x_i) [(x_i - \mu_j)(x_i - \mu_j)^T] \quad (31)$$

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

$$E = -L = -\sum_{j=1}^K \log p(y|x) \quad (32)$$

The two steps are iterated until the model converges to a local minimum (Callinon, 2009). 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 and RUL. 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). 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 leave-one-out cross validation technique (CV), was used to select the best number of mixing components. Given the test set, the GMR models can be obtained using the parameters of the GMM which has an output of a smoothed general description of the GMM encoded data and linked constraints given by matrices of the covariance (Callinon, 2009). This general smoothed description of the data is the prediction of the failure of the bearing.

## Gaussian process regression

The use of Gaussian process regression (GPR) for prognosis (prediction of the remaining useful life (RUL)) of slow rotating bearings based on the DAI is possible. Gaussian processes (GP) are a recent development in non-linear nonparametric modelling. In GP, the parametric model is dispensed and instead a prior probability distribution is defined over functions directly (Bishop, 2006). A nonlinear functional mapping from an inputting space to a target space is achieved by the use of GP modelling. The GP is defined as an infinite collection of arbitrary variables of which any of the fixed subsets has joint Gaussian distributions. The Gaussian method is favourable to smooth functions and those that properly explain the training data. The smooth attribute of the function leads to its plausible generalisations (Heyns *et al.*, 2012).

To motivate the GP viewpoint, let the vector  $\mathbf{x}_n$  represent the DAI in the input space. The training set of inputting vectors  $\mathbf{X}_N \equiv \{\mathbf{x}^n\}_{n=1}^N$  corresponds to the targeted vector  $\mathbf{y}_N \equiv \{y^n\}_{n=1}^N$ . For prognostics as in this study,  $x$  is the time period while  $y$  is a novel degradation assessment index for monitoring the health state of slow rotating bearings. A Gaussian process  $f(\mathbf{x})$  can be fully described by its mean and covariance (or kernel) function (Rasmussen and Williams, 2006). These functions are specified separately, and consist of a specification of a functional form as well as a set of parameters called hyper parameters.

The mean function describes the value of the function expected at any point of the inputting space, before the consideration of any trained data. The mean function can be defined as:

$$m(\mathbf{x}) = E(f(\mathbf{x})) \quad (33)$$

In supervised learning, the idea of likeness linking the various data points is vital. It is an essential similarity assumption that the points of inputs  $\mathbf{x}$  which are in proximity's target values  $y$  are expected to be similar. Hence, training points which are close to a test points prediction should be insightful. In the Gaussian process viewpoint, the covariance function depicts the nearness or similarity (Rasmussen and William, 2006). The covariance function betwixt two functional values evaluated at fixed points  $\mathbf{x}$  and  $\mathbf{x}'$  is given as

$$k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \quad (34)$$

The covariance function enables the inference value of a function given the knowledge of the other. Thus, the covariance function  $k(\mathbf{x}, \mathbf{x}')$  can be interpreted as the measure of the distance between the input points  $\mathbf{x}$  and  $\mathbf{x}'$ . The Gaussian process can then be written as:

$$f(\mathbf{x}) \sim GP[m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')] \quad (35)$$

The basic GPR consists of a simple zero mean and squared exponential covariance functions. The zero mean function is given as:

$$m(\mathbf{x}) = 0 \quad (36)$$

for every value of  $\mathbf{x}$ . One of the generally used kernel functions is the squared exponential (SE). It assumes that the function values in close proximity in the feature space are probably

going to be similar, with close to unity covariance for variables that have feature inputs that are close. The squared exponential covariance function with automatic relevance detection is given as (Rasmussen and Williams, 2006):

$$k_{SEard}(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}')^T M (\mathbf{x} - \mathbf{x}')\right) \quad (37)$$

where  $M$  matrix is diagonal with positive ARD parameters,  $M = \text{diag}(l)$  and  $l$  is length  $D$  vector corresponding to the input space dimension. The characteristic length-scale parameters, also known as the ARD parameters, determine the rate of variation of the function in the direction of the corresponding inputting space. A function tends to vary faster for any variation of its component feature for its shorter length scale parameter for a specific feature component. A short length scale thus corresponds to high relevance.  $\sigma_f^2$  is the signal variance linked to the general function variance.

The free parameters (i.e. hyperparameters) in the covariance function can be compressed into a matrix denoted by  $\varphi$ . The values of these hyperparameters are all unknown and inference is made from the data being the trained. It can be shown by utilization of the Bayes' rule that the maximum a posteriori hyperparameter values  $\varphi$  can be obtained by maximising the marginal likelihood  $p(\mathbf{y}|\mathbf{X}, \varphi)$  which is the same as minimising the negative log marginal likelihood (Rasmussen and Williams, 2006):

$$\log P(\mathbf{y}|\mathbf{X}, \varphi) = -\frac{1}{2} \mathbf{y}^T (\mathbf{K}(\mathbf{X}, \mathbf{X}))^{-1} \mathbf{y} - \frac{1}{2} \log |\mathbf{K}(\mathbf{X}, \mathbf{X})| - \frac{n}{2} \log 2\pi \quad (38)$$

where  $\mathbf{K}(\mathbf{X}, \mathbf{X})$  is the  $N \times N$  covariance matrix between all pairs of training inputs and is computed with equation (5) or (6). It is important to note however, that the application in this study used an  $N \times 1$  matrix of time points.

Given a set of training points, one can derive the posterior distribution over functions by imposing a restriction on prior joint distribution. Once a posterior distribution is derived, it can be used to estimate predictive values for the test data points (Saxena *et al.*, 2009). Denoting  $\mathbf{X}$  as the training inputs and  $\mathbf{X}_*$  as the test inputs, prediction of  $\mathbf{y}_*$  at the new locations  $\mathbf{X}_*$  may be inferred by conditioning the joint distribution on the observed target values. For the basic GPR with zero mean, the following equations describe the predictive distribution (Rasmussen and Williams, 2006):

Prior:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_* \end{bmatrix} \sim N\left(0, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right) \quad (39)$$

Posterior:

$$P(\mathbf{y}_*|\mathbf{X}, \mathbf{y}, \mathbf{X}_*) \sim N(\bar{\mathbf{y}}_*, \text{cov}(\mathbf{y}_*)) \quad (40)$$

where

$$\bar{y}_* = E[y|\mathbf{X}, \mathbf{y}, \mathbf{X}_*] = \mathbf{K}(\mathbf{X}_*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{y} \quad (41)$$

$$\text{cov}(\mathbf{y}_*) = \mathbf{K}(\mathbf{X}_*, \mathbf{X}_*) - \mathbf{K}(\mathbf{X}_*, \mathbf{X})\mathbf{K}(\mathbf{X}, \mathbf{X})^{-1}\mathbf{K}(\mathbf{X}, \mathbf{X}_*) \quad (42)$$

The maximum a posteriori (MAP) estimates  $\bar{y}_*$  can then be used as slow rotating bearing remaining useful lifemetrics.

### Model evaluation

The models would be evaluated using MAPE and RMSE. The MAPE and RMSE between the predicted and the original DAI were calculated using equations 43 and 44 respectively.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{DAI_i - \hat{DAI}_i}{DAI_i} \right| \times 100 \quad (43)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (DAI_i - \hat{DAI}_i)^2}{n}} \quad (44)$$

where  $DAI_i$  is the actual value of the degradation assessment index for the  $i$ th observation which is in this case the time point,  $\hat{DAI}_i$  is the predicted value of degradation assessment index,  $n$  is the number of observations.

The leave-one-out cross validation technique (CV) was used in selecting the test set for validating the predictions from each model. Two approaches were considered. The first is based on dependent sample. In this approach the bearing data set was divided into equal samples of training and test sets. The training set is the “seen” because it was used in training the parameters of the model while test set is the “unseen” as it was never fed into the model during training. However, it is been argued that when the training and test samples are dependent, the two errors may be positively correlated; resulting in a breakdown of the cross validation selection approach and can equally lead to model overfitting (Opsomer *et al.*, 2001; Arlot, 2010). Therefore, the second approach is based on independent samples whereby, two different sets of bearings are trained together and hence used as the training set while a third bearing data is used as the test set.

### Experimental setup

The experimental setup is used in this research to collect acoustic emission signals from slow rotating bearing. The test setup is designed to be able test any type of slow rotating bearing. The experimental test setup shown in figure 2 below and used in this research to collect the acoustic emission signals for two purposes: to study the vibration and acoustic

emission signatures generated by bearing faults. The speed controller controls the rotational speed of the bearing. The system is driven by an AC servo motor with the speed set at 70 rpm, 80 rpm and 100 rpm for bearings 1, 2, and 3 respectively. The shaft rotation speed is controlled by a speed controller. Acoustic emission sensor is mounted on the housing of the test bearing for measurement of acoustic emission signals. In all the bearings damage occurred on the outer race. Life test is performed on three slow rotating bearings until failure. Ground metal debris was introduced gradually into bearing 1 to hasten damage. Bearings 2 and 3 were simply not lubricated from start to ending to hasten bearing degradation. The slow rotating bearing is loaded at various dynamic loads using the Zonic shaker. Bearing 1 was loaded at maximum and minimum dynamic loads of 1.6kN and 1.0kN respectively. Bearing 2 was loaded at maximum and minimum dynamic loads of 1.8kN and 1.4kN respectively. Bearing 3 was loaded at maximum and minimum dynamic loads of 2.0kN and 1.7kN respectively.

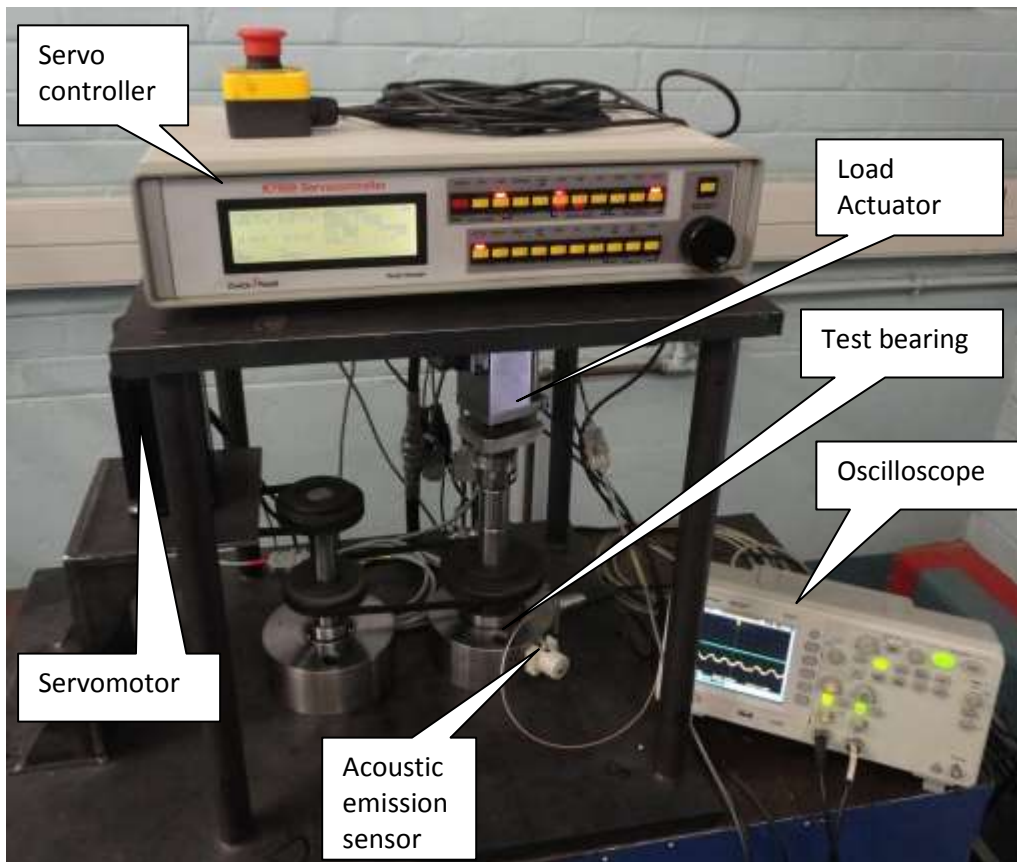


Figure 2: Setup

An acoustic emission sensor was used in the collection of data in analog form. The AE transducers are mounted on the outside surface of the outer race and on the bearing housing. The sampling rate of the AE measurement was set at 200 kHz, and the sample time as 1 second. The data was collected for a duration of 1 second after every 20 minutes until failure for bearings 1, 2 and 3 respectively. The obtained data was then converted into digital format for further processing and feature extraction with the help of dedicated 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 impending bearing faults can lead to proper planning and replacement in order to avoid disastrous failures of the whole machinery.

The major components of the slow rotating bearing test setup are the sonic shaker, load cell, test bearing, servo motor, speed controller, data loggers and acoustic emission sensors.

The AE data acquisition system consisted of piezoelectric-type AE transducers, amplifiers, an A/D card/data logger, and the computer. Broadband piezoelectric AE transducers connected to a 40dB gain pre amplifier for measurement was employed. The output signal from the preamplifier was connected to a National Instrument (NI) data acquisition card which was put into the ISA slot of a computer. The National Instruments Lab View software was used for collection of acoustic emission data. The function for capturing time domain and pre selected sampling time and interval was used. The rest of the processing and analysis was performed through Matlab programs for signal processing and analysis.

## **Results and discussions**

### **Prediction based on dependent samples**

The predictions in this section are based on dependent observations whereby the training and the test sets are obtained using a leave-one-out cross validation technique (CV) that involves the division of the bearing data set into equal samples of training and test sets.

The health state of a bearing is divided into three during its whole life namely, healthy or normal state, slightly degradation state and failure state. There is no need for remaining useful life when a bearing is in its healthy state. When the computed features are above their incipient damage threshold values then it is considered that slight degradation has set in. The prediction model is then used in the prediction of the future value of the degradation assessment index.

In this investigation healthy bearings are run until they are failed. A degradation assessment index is developed to assess the degradation of the slow rotating bearing. The DAI is then used in the several regression models, namely, the multi-layer perceptron (MLP), radial basis function (RBF), Bayesian linear regression (BLR), Gaussian mixture regression (GMR) and the Gaussian process regression (GPR) models for prediction of bearing damage, RUL and failure at a future instant of time.

Secondly, the mean absolute percentage error (MAPE) and root mean square error (RMSE) were used in model evaluation to select the best performing model. Thirdly, the best performing model from the resulting novel methodologies is recommended and used for the prediction of slow rotating bearing remaining useful life.

### ***RUL using multi-layer perceptron (MLP) regression***

Multi-layer perceptron is one of the most frequently used feedforward artificial neural networks which make use of a supervised learning algorithm. Essentially, it has three layers which include the input layer, pattern (hidden) layer and output layer (Şengüler *et al.*, 2010).

Dimensionality of the feature vectors were reduced to 2 PKPCs from 5 bearing extracted features using polynomial kernel principal components analysis (PKPCA) which subsequently fed into the GMM to obtain the degradation assessment index (DAI). The MLP neural network was trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE between the predicted and the actual DAI are shown in figures 3, 4 and 5 for bearings 1, 2 and 3 respectively. The MLP neural

network approach was used to monitor the trend of the incipient bearing damage and RUL of bearings 1, 2 and 3 are shown at the top right hand corner of figures 6, 7 and 8 respectively.

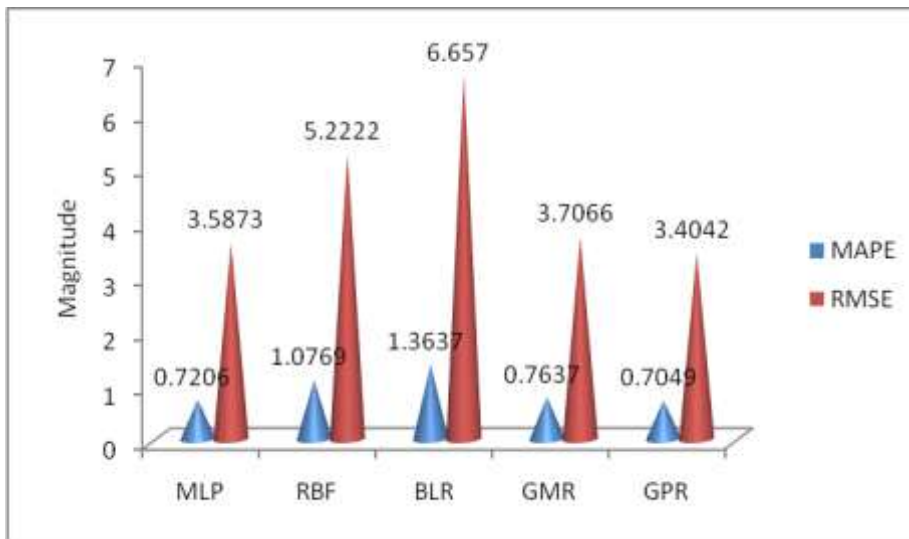


Figure 3: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 1 based on the dependent samples

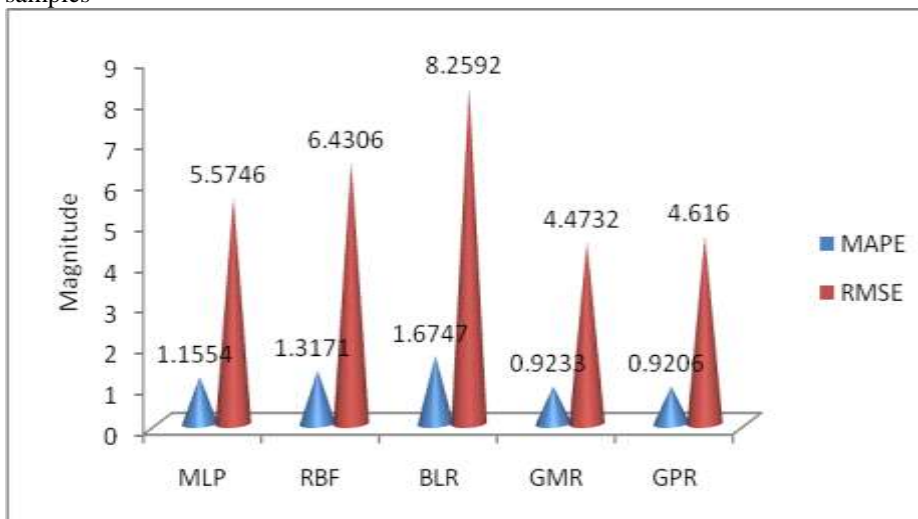


Figure 4: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 2 based on the dependent samples

### ***RUL using radial basis function (RBF) regression***

The RBF uses local hyper-sphere surfaces (non-linear mapping) to separate the classes in the input space as a response to cluster, rather than the global hyper-planes (lines) used in MLP networks (Al-Raheem and Abdul-Karem, 2010).

The degradation assessment index (DAI) was used as input into the RBF. The RBF was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI and shown in Figures 3, 4 and 5 for Bearings 1, 2 and 3 respectively. The RBF predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as shown at the top right hand corner of Figures 6, 7 and 8 respectively.



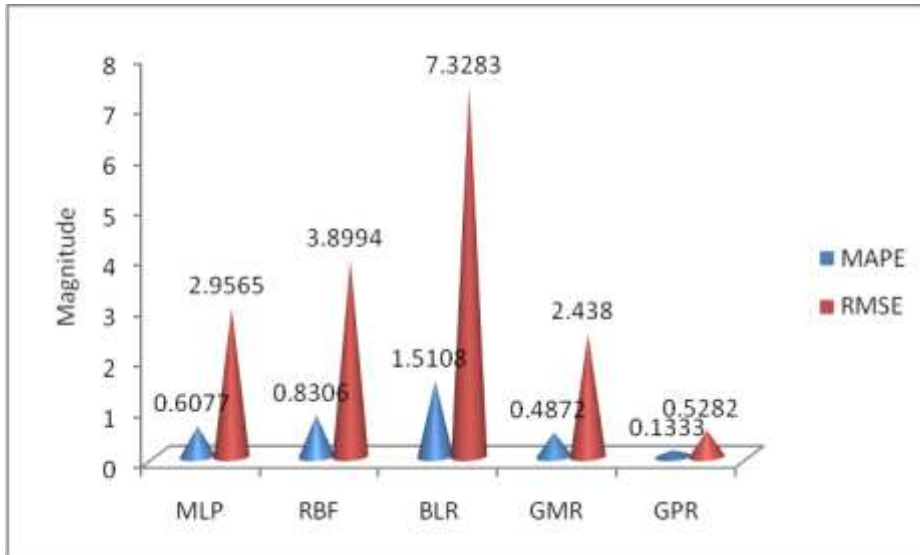


Figure 5: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for Bearing 3 based on the dependent samples

### ***RUL using Bayesian linear regression (BLR)***

Similarly, the degradation assessment index (DAI) was used as input into the BLR. The BLR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI and shown in Figures 3, 4 and 5 for Bearings 1, 2 and 3 respectively. The BLR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as shown in the middle left hand of Figures 6, 7 and 8 respectively.

### ***RUL using Gaussian mixture regression (GMR)***

Furthermore, the degradation assessment index (DAI) was used as input into the GMR. The GMR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI and shown in figures 3, 4 and 5 for bearings 1, 2 and 3 respectively. The GMR predictions of the incipient bearing damage and RUL of bearings 1, 2 and 3 were subsequently plotted as shown in the middle right hand of figures 6, 7 and 8 respectively.

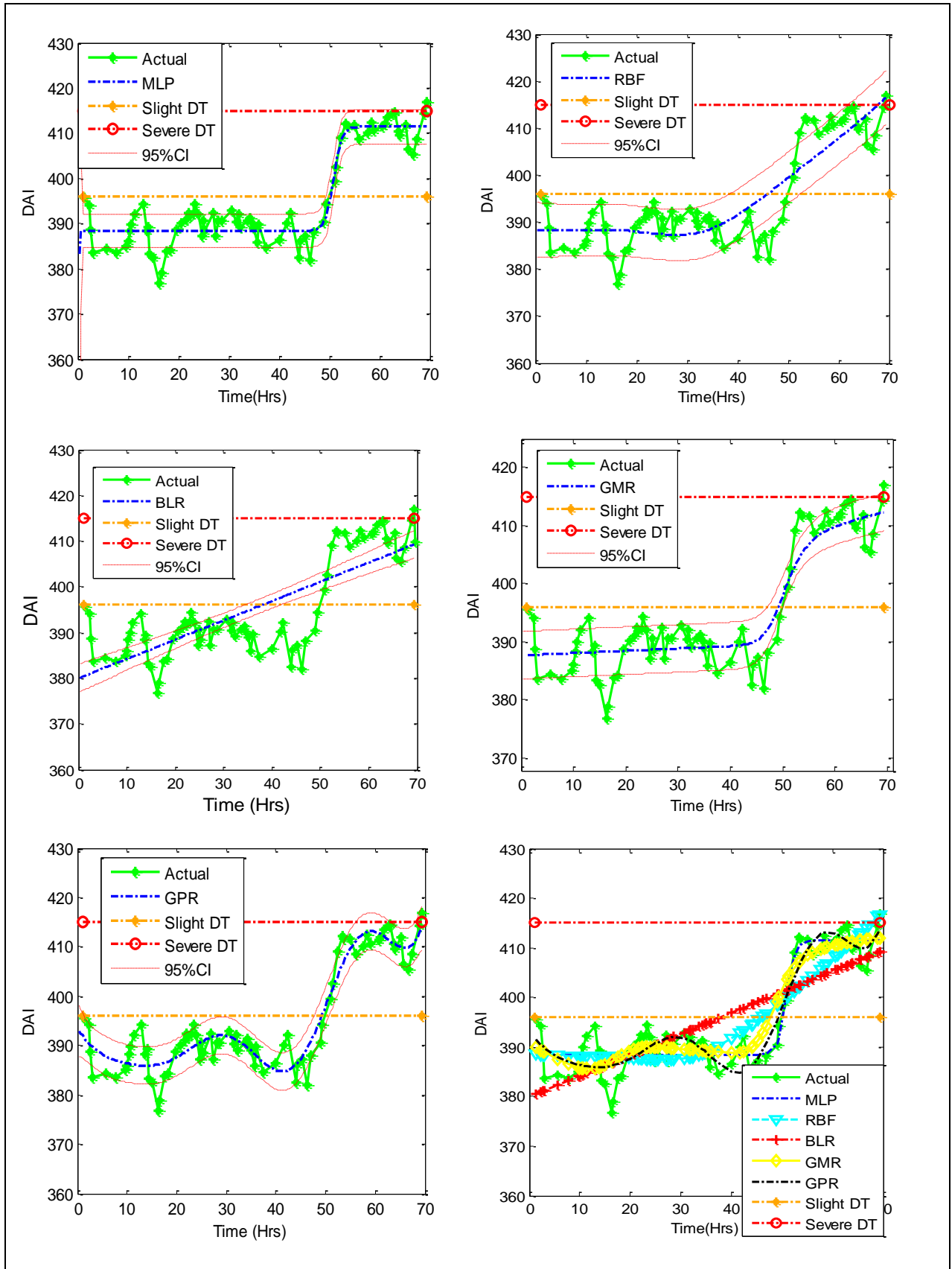


Figure 6: Prediction for the whole bearing life for bearing 1 using different methodologies based on dependent samples

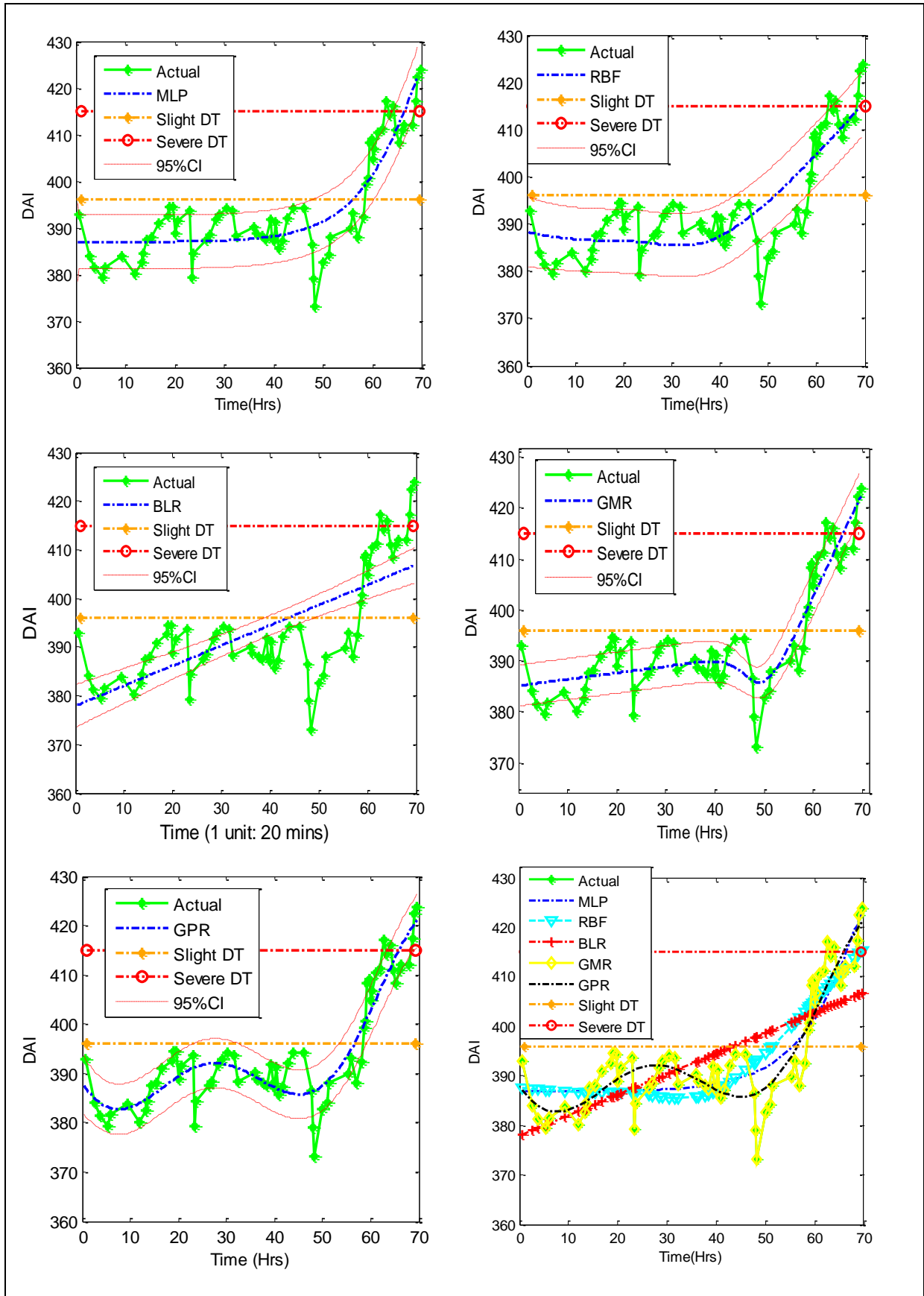


Figure 7: Prediction for the whole bearing life for bearing 2 using different methodologies based on dependent samples

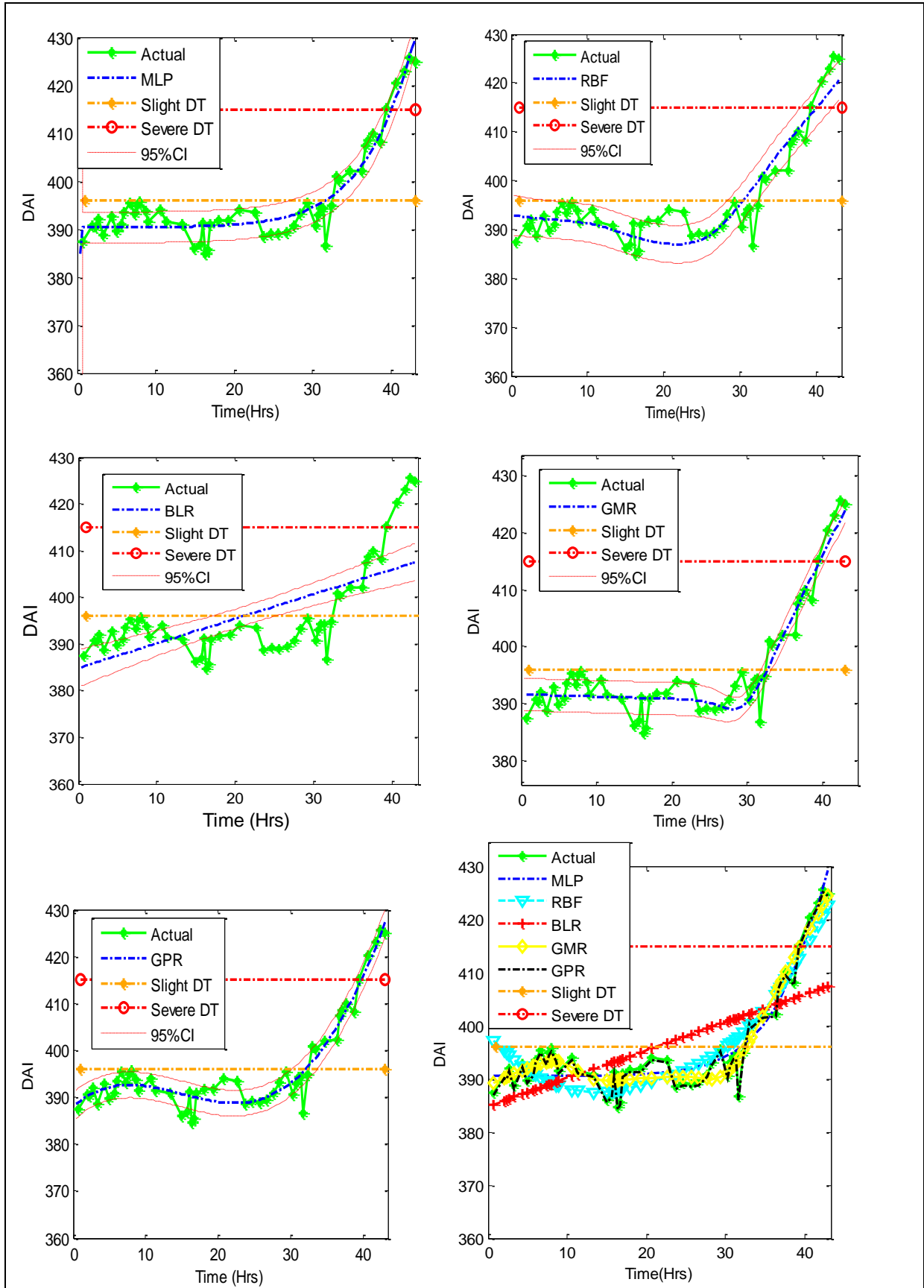


Figure 8: Prediction for the whole bearing life for bearing 3 using different methodologies based on dependent samples

### ***RUL using Gaussian process regression (GPR)***

The degradation assessment index (DAI) was used as input into the GPR. The GPR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI was calculated and shown in Figures 3, 4 and 5 for Bearings 1, 2 and 3 respectively. The GPR predictions of the incipient bearing damage and RUL of Bearings 1, 2 and 3 were subsequently plotted as shown in the bottom left hand of Figures 6, 7 and 8 respectively.

### **Model evaluation of the dependent samples**

After the training process, the prediction was done with data points equal to the training data points. The predicted and actual remaining useful life (RUL) plots of all the models were plotted in the bottom right hand of figures 6, 7 and 8 respectively.

The MAPE and RMSE between the output and real values are observed as plotted in figures 3, 4 and 5 for bearings 1, 2 and 3 for all the five models namely: MLP, RBF, BLR, GMR and GPR. All the models attempted to predict damage and RUL to a great degree. For bearing 1, the MAPE prediction errors from the models were 0.7049, 0.7206, 0.7637, 1.0769 and 1.3637 for GPR, MLP, GMR, RBF, and BLR respectively from the least to the highest prediction errors. Similarly, the RMSE prediction errors from the models were 3.4042, 3.5873, 3.7066, 5.2222 and 6.6570 for GPR, MLP, GMR, RBF, and BLR respectively from the least to the highest prediction errors. The worst prediction was that of the BLR which was a linear line across the non linear model. The GMR and MLP also modelled damage and RUL with little error. However, the best predictive model for bearing 1 was the GPR.

Similarly, for bearing 2, the MAPE prediction errors from the models were 0.9206, 0.9233, 1.3171, 1.1554 and 1.6747 for GPR, GMR, MLP, RBF, and BLR respectively from the least to the highest prediction errors. Similarly, the RMSE prediction errors from the models were 4.616, 4.4732, 5.5746, 6.4306 and 8.2592 for GPR, GMR, MLP, RBF, and BLR respectively from the least to the highest prediction errors. The worst prediction was that of the BLR which was a linear line across the non linear model. The GMR and MLP also modelled damage and RUL with little error. However, the best predictive model for bearing 2 was the GPR.

Finally for bearing 3, the MAPE prediction errors from the models were 0.1333, 0.4872, 0.6077, 0.8306 and 1.5108 for GPR, GMR, MLP, RBF, and BLR respectively from the least to the highest prediction errors. Similarly, the RMSE prediction errors from the models were 0.5282, 2.438, 2.9565, 3.8994 and 7.3283 for GPR, GMR, MLP, RBF, and BLR respectively from the least to the highest prediction errors. The worst prediction was that of the BLR which was a linear line across the non linear model. The GMR and MLP also modelled damage and RUL with little error. However, the best predictive model for bearing 3 was the GPR.

It could be seen that the GPR model consistently had the least error for all the three bearings. It was therefore concluded that the GPR model predicts damage and RUL better than the other models.

## Predictions based on independent samples

The predictions in this section are based on independent observations whereby two different sets of bearings are trained together and hence used as the training set while a third bearing data is used as the test set.

### *RUL using multi-layer perceptron (MLP) regression*

The MLP neural network was trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE between the predicted and the actual DAI are shown in figures 9, 10 and 11 for bearings 1, 2 and 3 respectively. The MLP neural network approach was used to monitor the trend of the incipient bearing damage and RUL of bearings 1, 2 and 3 are shown at the top right hand corner of figures 12, 13 and 14 respectively.

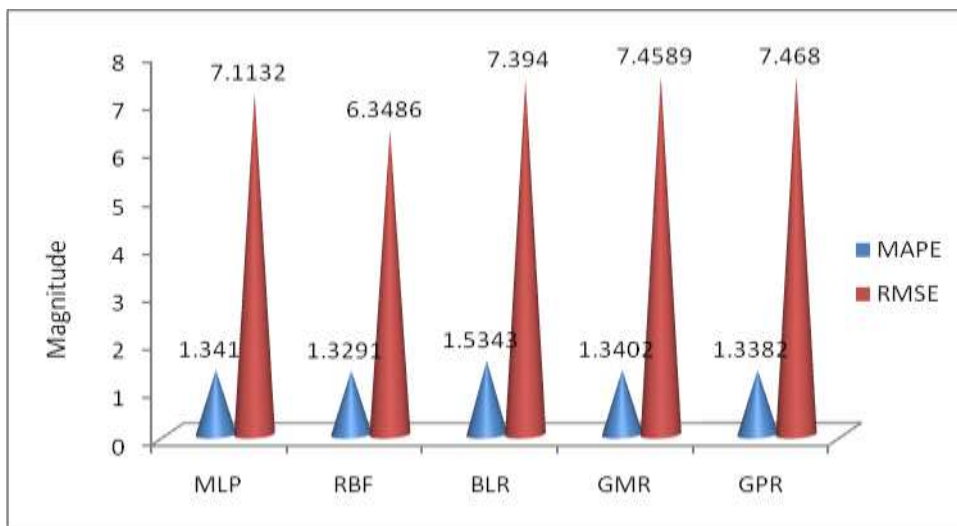


Figure 9: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for bearing 1 based on the independent samples

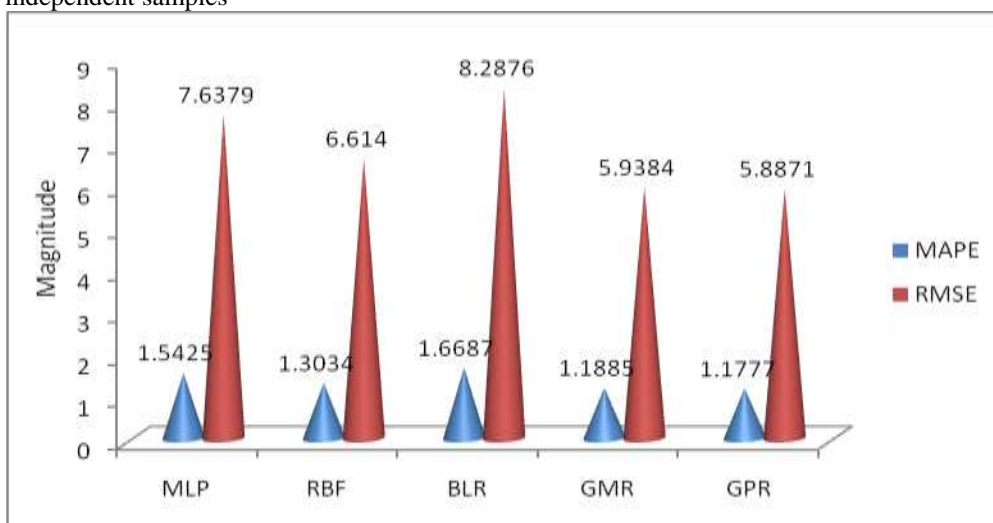


Figure 10: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for bearing 2 based on the independent samples

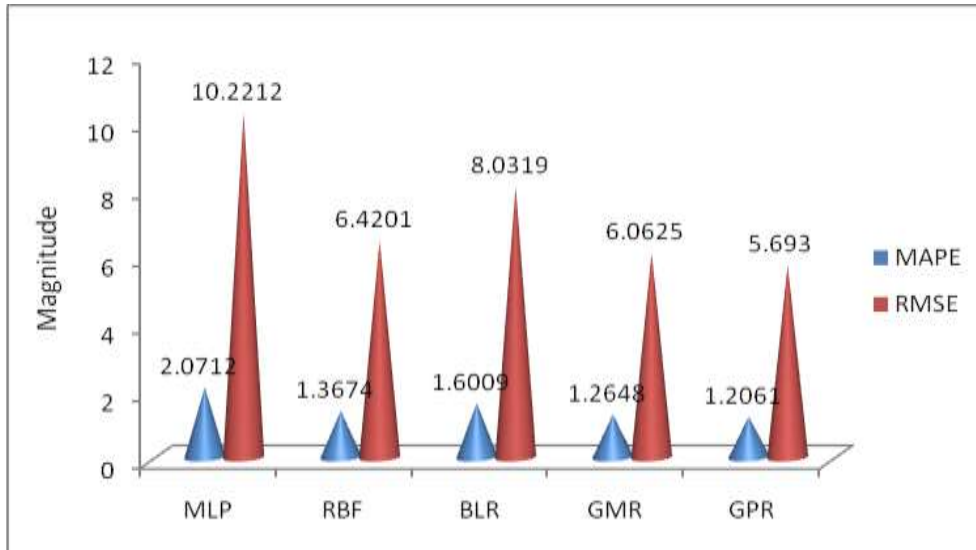


Figure 11: RMSE and MAPE for MLP, RBF, BLR, GMR and GPR models for bearing 3 based on the independent samples

### ***RUL using radial basis function (RBF) regression***

The degradation assessment index (DAI) was used as input into the RBF. The RBF was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the original DAI and shown in figures 9, 10 and 11 for bearings 1, 2 and 3 respectively. The RBF predictions of the incipient bearing damage and RUL of bearings 1, 2 and 3 were subsequently plotted as shown at the top right hand corner of figures 12, 13 and 14 respectively.

### ***RUL using Bayesian linear regression (BLR)***

Similarly, the degradation assessment index (DAI) was used as input into the BLR. The BLR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the original DAI and shown in figures 9, 10 and 11 for bearings 1, 2 and 3 respectively. The BLR predictions of the incipient bearing damage and RUL of bearings 1, 2 and 3 were subsequently plotted as shown in the middle left hand of figures 12, 13 and 14 respectively.

### ***RUL using Gaussian mixture regression (GMR)***

Furthermore, the degradation assessment index (DAI) was used as input into the GMR. The GMR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual DAI and shown in figures 9, 10 and 11 for bearings 1, 2 and 3 respectively. The GMR predictions of the incipient bearing damage and RUL of bearings 1, 2 and 3 were subsequently plotted as shown in the middle right hand of figures 12, 13 and 14 respectively.

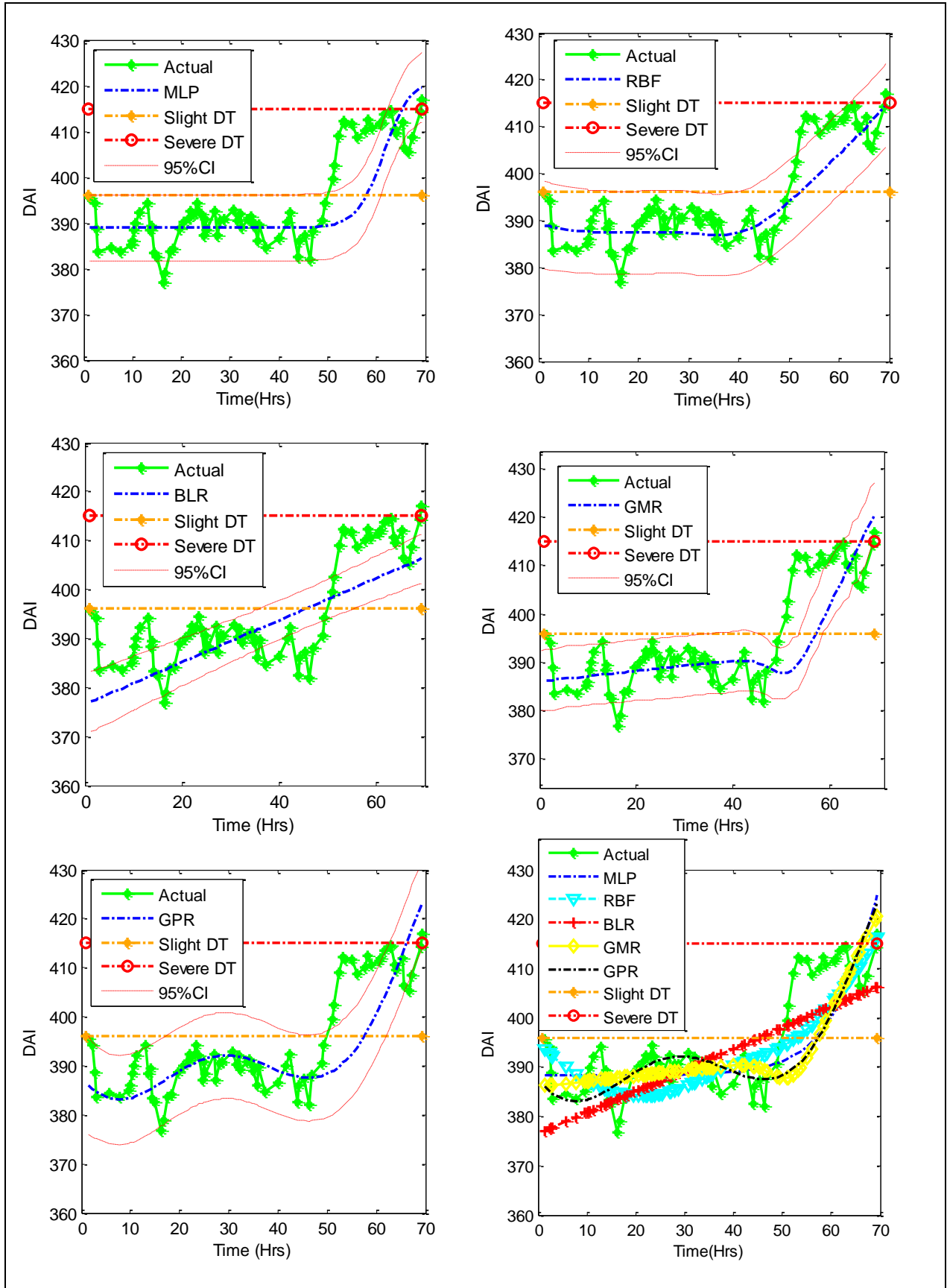


Figure 12: Prediction for the whole bearing life using bearing 2 and 3 as training set and bearing 1 as test set based on different methodologies and independent samples



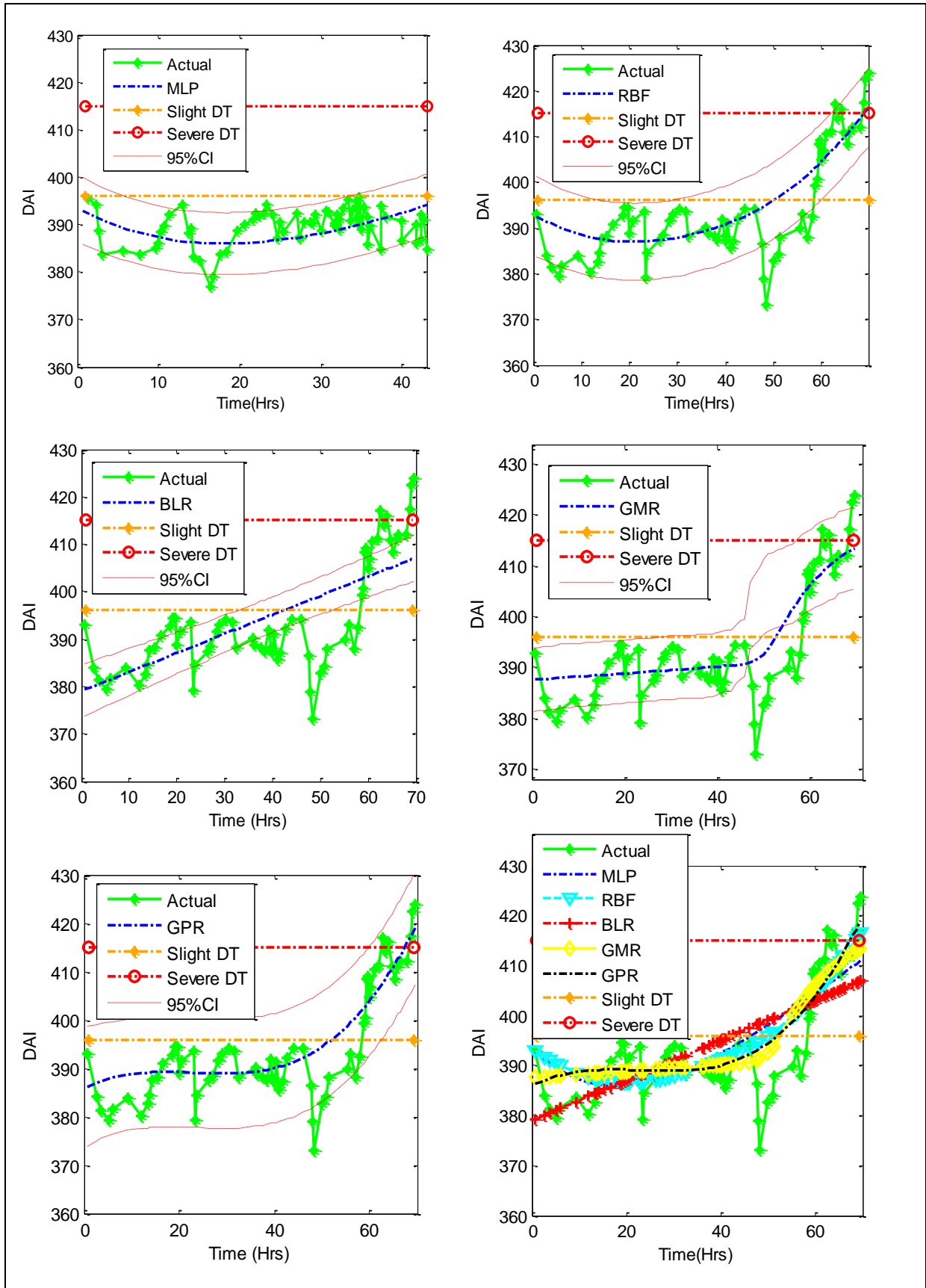


Figure 13: Prediction for the whole bearing life using bearing 1 and 3 as training set and bearing 2 as test set based on different methodologies and independent samples

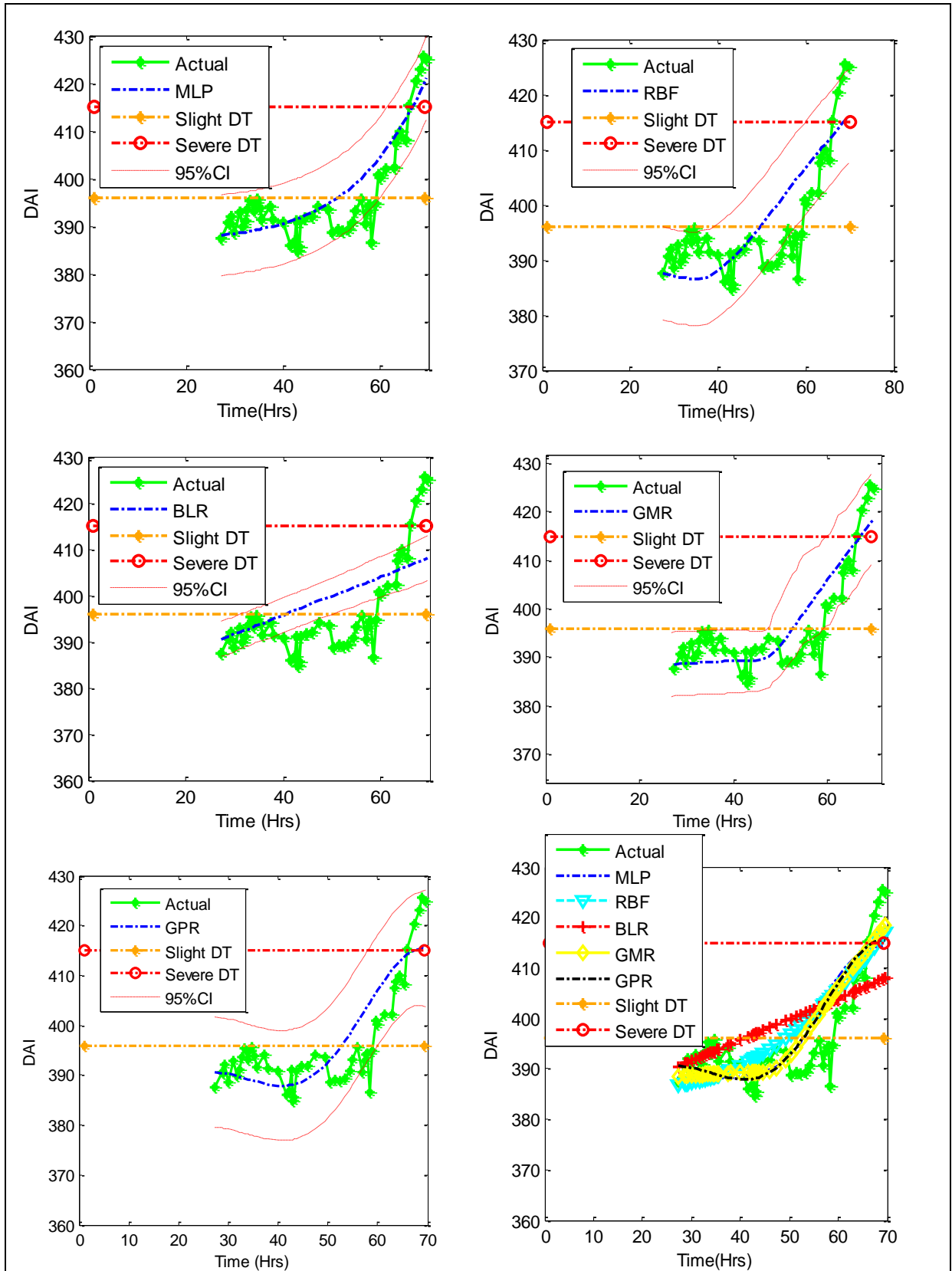


Figure 14: Prediction for the whole bearing life using bearing 1 and 2 as training set and bearing 3 as test set based on different methodologies and independent samples

### ***RUL using Gaussian process regression (GPR)***

The degradation assessment index (DAI) was used as input into the GPR. The GPR was then trained with the DAI which had been obtained from the bearing data at dynamic loadings conditions. The MAPE and RMSE were again computed between the predicted and the actual bearing DAI and shown in figures 9, 10 and 11 for bearings 1, 2 and 3 respectively. The GPR predictions of the incipient bearing damage and RUL of bearings 1, 2 and 3 were subsequently plotted as shown in the bottom left hand of figures 12, 13 and 14 respectively.

### **Model evaluation based on independent samples**

After the training process, the prediction was done with data points equal to the training data points. The predicted and actual remaining useful life (RUL) plots of all the models were plotted in the bottom right hand of figures 12, 13 and 14 respectively.

The MAPE and RMSE between the output and real values are observed as plotted in figures 9, 10 and 12 for bearings 1, 2 and 3 for all the five models namely: MLP, RBF, BLR, GMR and GPR. All the models attempted to predict damage and RUL to a great degree.

For bearing 1, using the independent approach the MAPE prediction errors from the models were 1.3291, 1.3382, 1.3402, 1.341 and 1.5343 for RBF, GPR, GMR, MLP and BLR respectively for bearing 1 from the least to the highest prediction errors. However, the RMSE prediction errors from the models were 6.3486, 7.1132, 7.394, 7.4589 and 7.468 for RBF, MLP, BLR, GMR and GPR respectively from the least to the highest prediction errors. The worst prediction was that of the GPR model. The best predictive model for bearing 1 was the RBF.

Similarly, for bearing 2, the MAPE prediction errors from the models was 1.1777, 1.1885, 1.3034, 1.5425 and 1.5425 for GPR, GMR, RBF, MLP, and BLR respectively from the least to the highest prediction errors. Similarly, the RMSE prediction errors from the models were 5.8871, 5.9384, 6.614, 7.6379 and 8.2876 for GPR, GMR, RBF, MLP, and BLR respectively from the least to the highest prediction errors. The worst prediction was that of the BLR which was a linear line across the non linear model. The GMR and RBF also modelled damage and RUL with little error. However, the best predictive model for bearing 2 was the GPR.

Finally for bearing 3, the MAPE prediction errors from the models were 1.2061, 1.2648, 1.3674, 1.6009 and 2.0712 for GPR, GMR, RBF, MLP and BLR respectively from the least to the highest prediction errors. Similarly, the RMSE prediction errors from the models were 5.693, 6.0625, 6.4201, 8.0319 and 10.2212 for GPR, GMR, RBF, MLP and BLR respectively from the least to the highest prediction errors. The worst prediction was that of the BLR which was a linear line across the non linear model. The GMR and RBF also modelled damage and RUL with little error. However, the best predictive model for bearing 3 was the GPR.

Overall, the GPR model had the least error for all the three bearings. It was therefore concluded that the GPR model predicts damage and RUL better than the other models.

### **Comparison of model performance based on dependent and independent samples**

The ranks of each prediction model according to whether the training and tests samples are dependent or independent are presented in tables 1, 2 and 3 for bearings 1, 2 and 3 respectively. While some of the models are sensitive to the type of sample, others are not. For example the GPR and BLR models ranked mainly 1<sup>st</sup> and last respectively in both the

dependent and independent samples whereas RBF and MLP ranked differently. However, the errors obtained from the cross validation based on independent samples (see figures 9 to 11) were relatively larger than those from the dependent samples (see figures 3 to 5) which could be an indication that the latter slightly overfitted the models.

Table 1: Ranking of models based on dependent and independent samples for bearing 1

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	1	2	5
GMR	3	3	3	4
RBF	4	4	1	1
MLP	3	2	4	2
BLR	5	5	5	3

Table 2: Ranking of models based on dependent and independent samples for bearing 2

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	2	1	1
GMR	2	1	2	2
RBF	4	4	3	3
MLP	3	3	4	4
BLR	5	5	5	5

Table 3: Ranking of models based on dependent and independent samples for bearing 3

Models	Dependent samples		Independent samples	
	MAPE	RMSE	MAPE	RMSE
GPR	1	1	1	1
GMR	2	2	2	2
RBF	4	4	3	3
MLP	3	3	5	5
BLR	5	5	4	4

## Conclusion

This paper proposes a novel approach to damage detection and prediction of remaining useful life of slow rotating bearings. During this investigation, three healthy slow rotating bearings were run to the point of failure. A degradation assessment index, DAI, which was obtained by the integration of polynomial kernel principal component analysis (PKPCA), Gaussian mixture model (GMM) and exponentially weighted moving average (EWMA) was used in slow rotating bearing prognostics. The slight and severe degradation thresholds are obtained through the use of the kernel density estimation (KDE) technique on the healthy and slightly degraded bearing data respectively. The DAI is used in the prediction of bearing damage and RUL using the MLP, RBF, BLR, GMR and GPR models respectively. Predictions were obtained using test and training sets from both dependent and independent samples. The models were able to predict damage and RUL of the slow rotating bearing. Overall, the GPR had the least mean absolute percentage and root mean square errors in this investigation for the slow rotating bearings and is robust to dependent and independent samples under varying operating conditions. Hence, the GPR is chosen as the most efficient model for prediction of remaining useful life of slow rotating bearings. This proposed

approach is useful and its application can be extended to the condition monitoring of other mechanical and allied systems.

## References

Al-Raheem K.F., Abdul-Karem W. ( 2010). Rolling bearing fault diagnostics using artificial neural networks based on Laplace wavelet analysis, *International Journal of Engineering, Science and Technology*, Vol. 2, No. 6, pp. 278-290,.

An, D., Choi, J-H.and Kim, N. H. (2012). A comparison study of methods for parameter estimation in the physics-based prognostics. Annual Conference of Prognostics and Health Management Society, Hyatt Regency Minneapolis, Minnesota, USA, 23 -27 Sep., 2012.

An, D., Kim, N. H. and Choi, J-H. (2013). Options for prognostics methods: A review of data-driven and physics-based prognostics. Annual Conference of the Prognostics and Health Management Society, New Orleans, 14-17 October, 2013.12.06

Arlot, S. (2010) A survey of cross-validation procedures for model selection. *Statistics Surveys*, 4, 40–79.

Bishop, C. M. (2006) *Pattern Recognition and Machine Learning Springer*, Cambridge, U.K.

Bolander, N., Qiu, H., Eklund, N., Hindle, E. and Rosenfeld, T. (2009). Physics-based remaining useful life prediction for aircraft engine bearing prognosis. Annual Conference of the Prognostics and Health Management Society, San Diego, CA September 27 – October 1, 2009.

Camcia, F., Medjaher, K., Zerhounib, N. and Nectoux, P. (2012) Feature evaluation for effective bearing prognostics. *Quality and Reliability Engineering International*, 1-15. DOI : 10.1002/qre.1396.

Calinon S. (2009). Robot Programming by Demonstration: A Probabilistic Approach, EPFL/CRC Press, 2009.

Chatzis *et al.* 2012

Chen, T. and Ren, J. (2009). Bagging for Gaussian process regression. *Neurocomputing*, 72 (7-9), 1605-1610.

Gebraeel N., Lawley M., Liu R., Parmeshwaran V. (2004). Residual life predictions from vibration-based degradation signals: a neural network approach. *IEEE Transactions on Industrial Electronics*, 51, 694–700.

Goebel, K., Saha, B., & Saxena, A. (2008). A comparison of three data - driven techniques for prognostics. Proceedings of the 62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT), May 6-8, Virginia Beach, VA.

Heyns, T., de Villiers, J.P. and Heyns, P.S. (2012). Consistent haul road condition monitoring by means of vehicle response normalisation with Gaussian processes. *Engineering Applications of Artificial Intelligence*, Vol. 25, Issue 8, December 2012, pp. 1752–1760.

Hippert, H.S. and Taylor, J.W. (2010). An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-term load forecasting. *Neural Networks*, 23, 386-395.

Hong S, and Zhou Z., (2012a). Remaining useful life prognosis of bearing based on a Gaussian process regression. 2012 5th International Conference on BioMedical Engineering and Informatics (BMEI 2012), Chongqing, China.

Hong, S., and Zhou, Z. (2012b). Application of Gaussian process regression for bearing degradation assessment. [Information Science and Service Science and Data Mining \(ISSDM\), 2012 6th International Conference on New Trends](#), Taipei, 23-25 Oct. 2012, 644 - 648.

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

Lee, J.-M., Yoo, C., Choi, S.W., Vanrolleghem, P.A., and Lee, I.-B. (2004). Nonlinear process monitoring using kernel principal component analysis. *Chemical Engineering Science*, 59, 223–234.

Liu, D., Pang J., Zhou, J, and Pang Y., (2012). Data driven prognostics for Lithium-ion battery based on Gaussian process regression. Accessed online at: <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06228848>

Malhi, A. and Gao, R.X. (2004). “PCA-based feature selection scheme for machine defectClassification” *IEEE Transactions on Instrumentation and Measurement*, 53 (6), pp. 1517–1525.

Marble, S. and Morton, B. (2005). Predicting the Remaining Life of Propulsion System Bearings, in *Proceedings of IEEE Aerospace Conference*, Big Sky, MO.

Nabney, I.T. (2002). *NETLAB Algorithms for Pattern Recognition*. Springer Publication, Great Britain.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 16(2), 134–153.

Rasmussen C. E. and Williams C. K. I. (2006). *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge MA, 2006.

Saha, B., Goebel, K., and Christophersen, J. (2009). Comparison of prognostic algorithms for estimating remaining useful life of batteries. *Transactions of the Institute of Measurement & Control*, 31(3-4), 293-308.

Saxena, A., Celaya, J., Saha, B., Saha, S. and Goebel, K. (2009). On applying the prognostic performance metrics. Annual Conference of the Prognostics and Health Management Society, 2009, 1-13.

Saxena, A. (2010). Prognostics, the science of prediction. Annual Conference of the Annual Conference of the Prognostics and Health Management Society, Portland, OR, 10 -14 October 2010.

Schölkopf, B., Smola, A., and Muller, K.R. (1998) Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1319.

Schölkopf, B., Smola, A., and Muller, K.R (1999). *Kernel principal component analysis*. In Advances in Kernel Methods - Support Vector Learning, MIT Press, 327-352.

Si, X.S., Wang, W., Hu, C-H., Zhou, D-H., (2011). Remaining useful life estimation- A review of the statistical data driven approaches. *European journal of operational research*, 213, 1-14.

Skabar A. (2007). Mineral potential mapping using Bayesian learning for multilayer perceptrons, *Mathematical Geology*, Vol. 39, pp. 439-451.

Wang G., Qian L. and Guo Z. (2013). Continuous tool wear prediction based on Gaussian mixture regression model, *International Journal Advance Manufacturing and Technology*, Vol. 66, pp. 1921–1929, 2013.

Wang, M. and Wang, J. (2012). CHMM for tool condition monitoring and remaining useful life prediction. *Int J Adv Manuf Technol*, 59, 463–471.