Combined Probability Approach and Indirect Data-Driven Method for Bearing Degradation Prognostics

Wahyu Caesarendra, Achmad Widodo, Pham Hong Thom, Bo-Suk Yang, and Joga Dharma Setiawan

***Abstract*—This study proposes an application of relevance vector machine (RVM), logistic regression (LR), and autoregressive moving average/generalized autoregressive conditional heteroscedasticity (ARMA/GARCH) models to assess failure degradation based on run-to-failure bearing simulating data. Failure degradation is calculated by using an LR model, and then regarded as the target vectors of the failure probability for training the RVM model. A multi-step-ahead method-based ARMA/GARCH is used to predict censored data, and its prediction performance is compared with one of Dempster-Shafer regression (DSR) method. Furthermore, RVM is selected as an intelligent system, and trained by run-to-failure bearing data and the target vectors of failure probability obtained from the LR model. After training, RVM is employed to predict the failure probability of individual units of bearing samples. In addition, statistical process control is used to analyze the variance of the failure probability. The result shows the novelty of the proposed method, which can be considered as a valid machine degradation prognostic model.**

***Index Terms*—Autoregressive moving average, censored data, Dempster-Shafer regression, generalized autoregressive conditional heteroscedasticity, prognostics, relevance vector machine, run-to-failure.**

ACRONYMS

ARMA	Autoregressive moving average
BPFI	Ball pass frequency inner race
BPFO	Ball pass frequency outer race
BSF	Ball spin frequency DSR Dempster-Shafer regression
GARCH	Generalized autoregressive conditional heterocedascity
LCL	Lower control limit
LR	Logistic regression
MLE	Maximum likelihood estimation
RVM	Relevance vector machine
SVM	Support vector machine
UCL	Upper control limit

NOTATION

c	Coefficient of ARMA model
$g(\vec{\mathbf{x}})$	Logit model
G_1 and A_1	Parameter of GARCH model
$K(\mathbf{x}, \mathbf{x}_i)$	Kernel function of RVM
k	Coefficient of GARCH model
$L(\phi_1, \theta_1, \sigma_\varepsilon^2)$	Likelihood function for the ARMA (1, 1) model
$L(k, G_1, A_1)$	Likelihood function for the GARCH (1, 1) model
N	Number of data for each data input of RVM
$P(\vec{x})$	Probability of some output event of logistic function
$p(\mathbf{w}\|\boldsymbol{\alpha})$	Prior probability distribution over the weights
$p(\mathbf{t}\|\mathbf{w}, \sigma^2)$	Likelihood of RVM data set
$p(\mathbf{w}\|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)$	Posterior distribution of weight
$\{\mathbf{t}\}_{n=1}^N$	Target vector for RVM training
$\mathbf{w}_i$	Weight vector obtained from RVM training
$\vec{x}(x_1, x_2, \ldots x_k)$	Input vector of logistic regression
$\{\mathbf{x}\}_{n=1}^N$	Set of data input for RVM training
$\boldsymbol{\alpha}$	Vector of $(N \times 1)$ hyperparameters
$\beta_1, \beta_2, \ldots, \beta_k$	Regression coefficients
ε_n	Noise process of RVM with mean 0, and variance σ^2
ε_t	White noise of ARMA model

Manuscript received March 25, 2010; revised July 08, 2010 and August 01, 2010; accepted August 08, 2010. Date of publication: February 07, 2011; date of current version: March 02, 2011 . This work was supported by the Brain Korea 21 project. Associate Editor: W. Wang.

W. Caesarendra and P. H. Thom are with the School of Mechanical Engineering, Pukyong National University, South Korea (e-mail: caesarwahyu@yahoo.co.id).

A. Widodo and J. D. Setiawan are with Mechanical Engineering Department of Diponegoro University, Semarang, Indonesia (e-mail: awid@ undip.ac.id; joga.setiawan@alum.mit.edu).

B.-S. Yang is with the School of Mechanical Engineering and the Intelligent Mechanics Lab (IML) and Research Centre of Intelligent Machine Condition Monitoring and Diagnosis (IMCMD), Pukyong National University, South Korea (e-mail: bsyang@pknu.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TR.2011.2104716

ϕ_1 and θ_1	Parameter of ARMA model
σ_t^2	Conditional variance

I. Introduction

In maintenance activities, a failure degradation assessment can be used as important information to indicate the performance of a part or machine at a particular or a range of time through condition monitoring, and fault diagnostics systems. To create the appropriate decision in system health management, and predict the impending failure, the condition monitoring and fault diagnostics of machine performance should be complemented with an appropriate prognostics system. Prognostics will reduce considerable maintenance cost (for instance by avoiding unscheduled maintenance, and by increasing equipment usage), and operational safety improvement. Therefore, finding solutions for the prognostic problem is a challenge. Several studies of machine prognostics are presented in [1]–[3].

This study proposes a combination of a probability approach, and an indirect data-driven method in RVM, LR, and ARMA/GARCH models are utilized.

Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution [4]. Kurtosis is also one of the statistical time domain features that indicates the performance degradation of bearings, and gives potential damage detections at an earlier stage. When the defect impacts the rolling element bearings, it produces a signal which has probability density that is more peaked than that of normal bearings. In this case, kurtosis is the best measure for indicating the peaked distribution related to a rolling element bearing fault. The literature [5]–[7] studied the usage of kurtosis for assessing the bearing condition, and showed the effectiveness of kurtosis in bearing defect detection. The LR model is utilized to address the probability of bearing samples. For example, we have a large number of bearing samples. We run the bearings from new condition, and stop it at particular future time. Some bearing samples will fail at different times, and others are censored. The failure time and censored data are regarded as the target vector input for training the RVM model. Furthermore, ARMA/GARCH is employed to predict the censored data. The prediction performance of the ARMA/GARCH model will be compared with that of the DSR method. Moreover, the RVM model is used for training the kurtosis bearing data, and target vectors of failure probability obtained from the LR model. After training, the RVM model is employed to predict the failure probability of individual bearing samples. In addition, to yse the variance of failure probability, statistical process control (SPC) is employed to analyze the variance of failure probability in this study. The result shows that the proposed method can be considered as a machine degradation prognostic model.

II. Theoretical Background

A. Logistic Regression (LR)

LR is a variation of the ordinary regression method which is used where the dependent variable is a dichotomous variable (which is usually represented as the occurrence or non-occurrence of some output events, usually coded as 0 or 1). The goal of LR is to find the most fitting model to describe the relationship between the dichotomous characteristic of the dependent variable, and a set of independent variables [8]. Here, LR is used to obtain the bearing failure probability between incipient failure time to final failure time. Then, the probability result is treated as a target vector input of RVM. In the LR approach, the dependent variable is the probability of an event occurrence. Hence, the output value has a discrete number of responses which are constrained from 0 (functional) to 1 (failure). The logistic function is

$$P(\vec{x}) = \frac{1}{1+e^{-g(\vec{x})}} = \frac{e^{g(\vec{x})}}{1+e^{g(\vec{x})}} \tag{1}$$

where $P(\vec{x})$ is the probability of some output event, $\vec{x}(x_1, x_2, \ldots x_k)$ is an input vector, corresponding to the s-independent variables (predictors), and $g(\vec{\mathbf{x}})$ is the logit model. The logit model of multiple logistic regressions can be written as

$$g(\vec{\mathbf{x}}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k \tag{2}$$

where $g(\vec{\mathbf{x}})$ is a linear combination of the s-independent variables $x_1, x_2, \ldots, x_k$, and α, $\beta_1, \beta_2, \ldots, \beta_k$ are known as the regression coefficients. These coefficients can be calculated by applying MLE after transforming the dependent variable into a logit variable.

B. Relevance Vector Machine (RVM)

RVM is a Bayesian form representing a generalized linear model of identical functional form to that of SVM. It differs from SVM in the case of calculating the solution which provides probabilistic interpretation of its outputs [9]. As a supervised learning, RVM starts with a set of data input $\{\mathbf{x}\}_{n=1}^{N}$, and their corresponding target vector $\{\mathbf{t}\}_{n=1}^{N}$. The aim is to determine a model of the dependency of the target vectors on the inputs in order to make accurate prediction of $\mathbf{t}$ for the unseen value of $\mathbf{x}$. Typically, the predictions are based on a function $y(\mathbf{x})$ defined over the input space, and the learning process of inferring the parameter of this function.

In the context of SVM, this function takes the form of

$$y(\mathbf{x}) = \sum_{1}^{N} \mathbf{w}_i K(\mathbf{x}, \mathbf{x}_i) + w_0 \tag{3}$$

RVM seeks to predict a target $\mathbf{t}$ for any query of $\mathbf{x}$ such that

$$\mathbf{t} = y(\mathbf{x}) + \varepsilon_n \tag{4}$$

The likelihood can be written as

$$p(\mathbf{t}|\mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{N-2} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{t} - \mathbf{\Phi w}\|^2\right\} \tag{5}$$

where Φ is the $N \times (N+1)$ design matrix with $\Phi_{nm} = \{1, K(\mathbf{x}_i, \mathbf{x}_1), K(\mathbf{x}_i, \mathbf{x}_2), \ldots, K(\mathbf{x}_i, \mathbf{x}_N)\}^T$.

The MLE of $\mathbf{w}$ and σ^2 in (5) often result in over fitting. Therefore, Tipping [10] recommended imposition of some prior constrains on the parameters $\mathbf{w}$ by adding a complexity to the likelihood or error function. This prior information controls the generalization ability of the learning process. Typically, new higherlevel parameters are used to constrain an explicit zero-mean Gaussian prior probability distribution over the weights

$$p(\mathbf{w}|\boldsymbol{\alpha}) = \prod_{i=0}^{N} N\left(w_i|0, \alpha_i^{-1}\right) \tag{6}$$

where $\boldsymbol{\alpha}$ is a vector of $(N \times 1)$ hyperparameters that controls how far from zero each weight is allowed to deviate [11].

Using Bayes' rule, the posterior overall unknowns could be computed, given the defined non-informative prior-distributions

$$p(\mathbf{w}, \alpha, \sigma^2|\mathbf{t}) = \frac{p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma)}{\int p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)d\mathbf{w}d\boldsymbol{\alpha}d\sigma^2} \tag{7}$$

However, we cannot compute the solution of the posterior in (7) directly because we cannot perform the normalizing integral $p(\mathbf{t}) = \int p(\mathbf{t}|\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2)d\mathbf{w}d\boldsymbol{\alpha}d\sigma^2$.

Instead, we decompose the posterior as

$$p(\mathbf{w}, \boldsymbol{\alpha}, \sigma^2|\mathbf{t}) = p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha}, \sigma^2|\mathbf{t}) \tag{8}$$

to facilitate the solution. The posterior distribution of weights is given by

$$p(\mathbf{w}|\mathbf{t}, \boldsymbol{\alpha}, \sigma^2) = \frac{p(\mathbf{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}, \boldsymbol{\alpha})}{p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2)} \tag{9}$$

Equation (9) has an analytical solution where the posterior covariance and mean are

$$\sum = (\boldsymbol{\Phi}^T\mathbf{B}\boldsymbol{\Phi} + \mathbf{A})^{-1}, \quad \boldsymbol{\mu} = \sum \boldsymbol{\Phi}^T\mathbf{B}\mathbf{t} \tag{10}$$

with $\mathbf{A} = diag(\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{N+1})$, and $\mathbf{B} = \sigma^{-2}\mathbf{I}$.

Note that σ^2 is also treated as a hyperparameter which may be estimated from the data. Therefore, machine learning becomes a search for the hyperparameter posterior in the most probable way. Predictions for a new data are then made according to integration of the weights to obtain the marginal likelihood for the hyperparameter

$$\begin{aligned} p(\mathbf{t}|\boldsymbol{\alpha}, \sigma^2) &= \int p(\mathbf{t}|\mathbf{w}, \sigma^2)p(\mathbf{w}|\boldsymbol{\alpha})d\mathbf{w} \\ &= (2\pi)^{N/2}|\mathbf{B}^{-1} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^T|^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}\mathbf{t}^T(\mathbf{B}^{-1} + \boldsymbol{\Phi}\mathbf{A}^{-1}\boldsymbol{\Phi}^T)^{-1}\mathbf{t}\right\} \end{aligned} \tag{11}$$

C. ARMA/GARCH Model

The ARMA (1, 1) model is expressed as

$$y_t = c + \phi_1 y_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t. \tag{12}$$

To estimate the coefficients c, ϕ_1, and θ_1, the error term ε_t is assumed to be white noises which are s-independent following normal distributions with zero-mean and constant variance σ_ε^2. The likelihood function for the ARMA (1, 1) model is

$$\begin{aligned} L\left(\phi_1, \theta_1, \sigma_\varepsilon^2\right) &= \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \\ &\quad \times \exp\left\{-\frac{\left(y_t - c - \phi_1 y_{t-1} - \theta_1 \varepsilon_{t-1}^*\right)^2}{2\sigma_\varepsilon^2}\right\} \end{aligned} \tag{13}$$

The log likelihood function, neglecting the constant term, is

$$\begin{aligned} l\left(\phi_1, \theta_1, \sigma_\varepsilon^2\right) &= -(T-1)\log\sigma_\varepsilon - \frac{1}{2\sigma_\varepsilon^2} \\ &\quad \times \sum_{t=2}^{T}\left(y_t - c - \phi_1 y_{t-1} - \theta_1 \varepsilon_{t-1}^*\right)^2 \end{aligned} \tag{14}$$

where $\varepsilon_t^* = y_{t-1} - c - \phi_1 y_{t-2} - \theta_1 \varepsilon_{t-2}^*$ for $t = 3, \ldots, T$ are obtained recursively.

The GARCH (1, 1) model is expressed as

$$\sigma_t^2 = k + G_1\sigma_{t-1}^2 + A_1\varepsilon_{t-1}^2 \tag{15}$$

where ε_t is inferred from ARMA (1, 1) model, and assumed to have the conditional variance σ_t^2.

In the case of a Gaussian ε_t, the likelihood function is

$$L(k, G_1, A_1) = \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} \exp\left\{-\frac{\varepsilon_t^2}{2\sigma_t^2}\right\} \tag{16}$$

The log likelihood function, neglecting the constant term, is

$$l(k, G_1, A_1) = -\frac{1}{2}\sum_{t=2}^{T}\left\{\log\sigma_t^2 + \frac{\varepsilon_t^2}{\sigma_t^2}\right\} \tag{17}$$

where $\sigma_t^2 = k + G_1\sigma_{t-1}^2 + A_1\varepsilon_{t-1}^2$, are obtained recursively.

GARCH prediction is implemented after the functional form of the model has been specified, and all parameters: r, m, ϕ_i, θ_j (for ARMA); p, q, G_i, A_j (for GARCH) have been estimated. The ARMA/GARCH model can be used to predict the multi-step-ahead value of the time series. In this paper, the ARMA/GARCH model is used to predict the kurtosis. This method has been utilized for estimating and forecasting machine health conditions from low methane compressor data [12].

D. Statistical Process Control (SPC)

In this study, we introduce SPC into the prognosis application to analyze the variance of failure time. By applying SPC into the prognosis area, feedback on the variance and reliability of parts can be sent to the manufacturer once parts are operated under same loading condition. A primary tool used for SPC is the control chart. If the chart indicates that the process is currently under control, then it can be used with confidence to predict the future performance of the process. On the contrary, if the chart indicates that the process being monitored, is not under control, the pattern which it reveals can help to determine the source of variation to be eliminated to bring the process back into control. To control the process, upper and lower warning limits are important features in control charts.

The formulas of LCL, and UCL with 3-standard errors is

$$LCL = \bar{\bar{x}} - 3s_{\bar{x}}, \quad UCL = \bar{\bar{x}} + 3s_{\bar{x}}, \quad s_{\bar{x}} = \frac{\bar{R}}{d_2\sqrt{n}} \tag{18}$$

where $\bar{\bar{x}}$ defines the average of averages, $s_{\bar{x}}$ is standard error, $\bar{R}$ is range (maximum value–minimum value), d_2 is a control chart constant based on the number of subgroup, and n is the subgroup size.

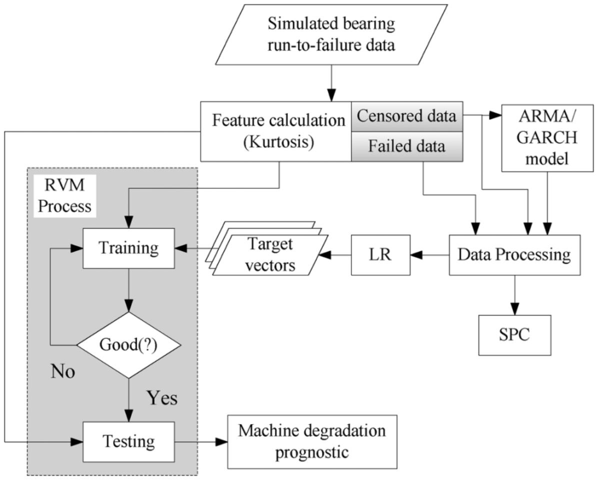

Fig. 1. Schematic diagram of the machine degradation prognostic model.

III. Methodology

The proposed method shown in Fig. 1 is employed for bearing failure simulating data. One-dimensional statistic, namely kurtosis, is calculated initially. This feature can be used to represents the bearing failure degradation from normal to the final failure condition. In addition, the ARMA/GARCH model is employed to predict the kurtosis of censored data up to reach a predetermined final failure threshold. The time series prediction method using DSR is employed to compare with the prediction result of the ARMA/GARCH model. The result shows that ARMA/GARCH prediction is better than DSR, and deserves to be used further. Then, failure degradation is calculated by using the LR method, which provides target vectors of failure probability. RVM is used for training the run-to-failure kurtosis data, and target vectors of failure probabilities estimated by LR. Furthermore, RVM is used to predict the failure time of bearing. SPC is used as an additional process to analyze the variance of incipient failure, and final failure of bearing samples. The results of SPC can give the information whether bearing sample is failed over the tolerance limit or not. If the variances are small and do not exceed the limit, the prediction result will be more accurate.

To evaluate the training performance, the root mean square error $(RMSE)$, and the correlation (R) are calculated. The $RMSE$, and R formula are respectively given by

$$RMSE=\left(\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)^2\right)^{1/2} \quad (19)$$

$$R=\frac{Cov(y_t,\hat{y}_t)}{\sigma_{y_t}\sigma_{\hat{y}_t}}=\frac{1}{\sigma_{y_t}\sigma_{\hat{y}_t}N}\sum_{t=1}^{N}(y_t-\bar{y}_t)(\hat{y}_t-\bar{\hat{y}}_t) \quad (20)$$

where R is the correlation coefficient, and $Cov(y_t,\hat{y}_t)$ is covariance between actual and validated values. Variable y_t indicates an actual value, and $\hat{y}_t$ refers to a validated value. In addition, $\bar{y}_t$, and $\bar{\hat{y}}_t$ denote the average result of actual, and validated values, respectively.

Moreover, σ_{y_t} is the standard deviation of the actual value, and $\sigma_{\hat{y}_t}$ is the standard deviation of the validated value. The symbol n denotes the number of predictable data. The smaller value of $RMSE$ indicates a higher accuracy of validation, and a high correlation value expresses a good validation.

TABLE I
Specifications of the Bearing Used in the Simulation Study

Properties	Description
Pitch diameter	23 mm
No. of rolling elements	9
Ball diameter	8 mm
Contact angel	0°

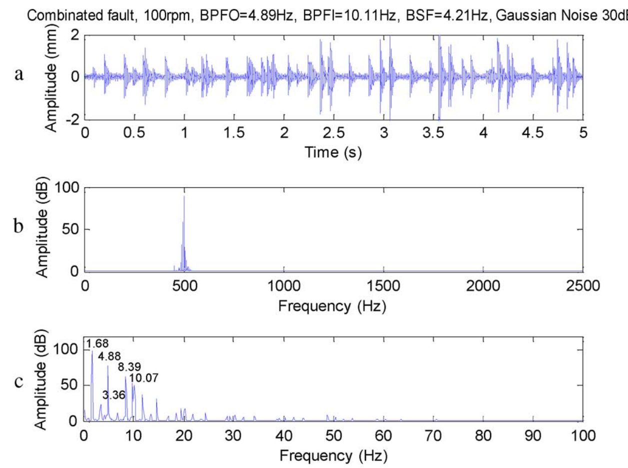

Fig. 2. Accumulated signal of three bearing defects: (a) time domain plot of the raw signal; (b) frequency spectrum of the raw signal; and (c) fault detection after demodulation.

IV. Simulating Data

The proposed method is validated by using simulating data of bearing defect degradation. We develop vibration condition monitoring data that represents defect propagation of rolling element bearing using a MATLAB program. The properties of the rolling element bearing in the simulation are in Table I.

Data consisted of bearing outer-race, bearing inner-race, and ball fault defect at a rotating speed of 100 rpm, and a sampling frequency 5 kHz. Fig. 2 shows the accumulated signal of the three defects on the bearing. The calculated frequencies of an outer-race fault (BPFO), inner-race fault (BPFI), and ball fault defect (BSF) are indicated on the top of the Fig. 2. Fig. 2(a) shows the signal over time. This signal is converted to the frequency domain using a fast Fourier transform (FFT), as shown in Fig. 2(b). This figure reveals that the spectrum is dominated by high frequency resonant signals. To separate the bearing fault frequency signal from these dominant signals, the vibration signals are band-pass filtered and rectified. Fig. 2(c) shows that the calculation results are closely matched to the fault simulation peaks that were detected at 4.88 Hz $(1 \times \mathrm{BPFO})$, 10.07 Hz $(1 \times \mathrm{BPFI})$, and 8.39 Hz $(2 \times \mathrm{BSF})$.

The three simulated signals are repeatedly generated from the computer program based on equations presented in [13], [14]. All simulated signals have defect impulses that increase over time. Bearing degradation signals possess an inherent exponential growth [15], [16]. The result of feature calculation using kurtosis is shown in Fig. 3. In this figure, kurtosis of bearing sample 15 reached the predetermined final failure threshold before the observation stopped at time step 100 (this is referred as failure data), while the kurtosis of bearing sample 2 did not hit

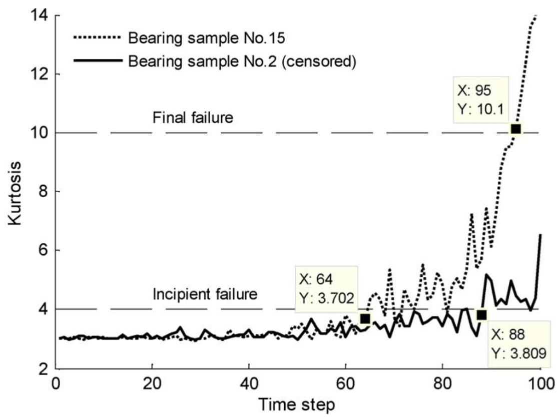

Fig. 3. Kurtosis of two bearing samples.

the predetermined final failure threshold, in other word bearing sample 2 is not failure up to time step of 100 (this is referred as censored data). Moreover, this censored data will be predicted to calculate the final failure time.

V. Results and Discussion

To acquire run-to-failure bearing data, 25 bearing samples were simulated in the time step range of 0 to 100. After calculating the kurtosis, the predefined thresholds of incipient failure, and final failure were set-up as shown in Fig. 3.

Fig. 3 also shows that the kurtosis value from time step 0 until time step approximately 45 is 3. This value is similar to the expected value of kurtosis for bearing normal is near 3.

As failures develop over time, the kurtosis signal will exceed incipient and final failure thresholds. Then, these data are used for calculating the failure probability from the incipient failure time until the final failure using LR. In Fig. 3, the kurtosis signal of bearing 15 has an incipient failure time, and final failure time of 64, and 95, respectively. We conducted simulations for 25 bearings. When the simulation stopped at 100 units of time, 20 bearings had failed, and 5 bearings were not failed, so were censored.

This study applies single logistic regression, which means it models only one s-independent variable. The independent variable is estimated from the incipient and final failure times of kurtosis data, denoted as x_1. The parameters α, and β_1 are calculated by using MLE. Therefore, the logit model corresponding to (2) is determined by

$$g(\vec{\mathbf{x}}) = 4.3318 + 0.2172x_1 \tag{21}$$

Failure probability results according to (1) can be estimated as shown in Fig. 4. This figure demonstrates the failure probability of 24 bearings from simulated data. The failure probability is calculated from the incipient failure (failure probability equal to 0) until the final failure occurred (failure probability equal to 1) [17]. Moreover, this result is regarded as the target vector which will be used for training the RVM model. Fig. 4 also shows the failure degradation starting from the incipient failure until the final failure is between time steps 48 and 100. Within the time step of 0 to 47, the condition can be considered to be a normal state.

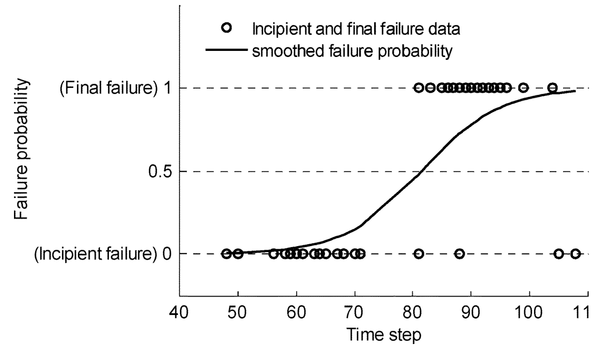

Fig. 4. Logistic regression result of simulated data [13].

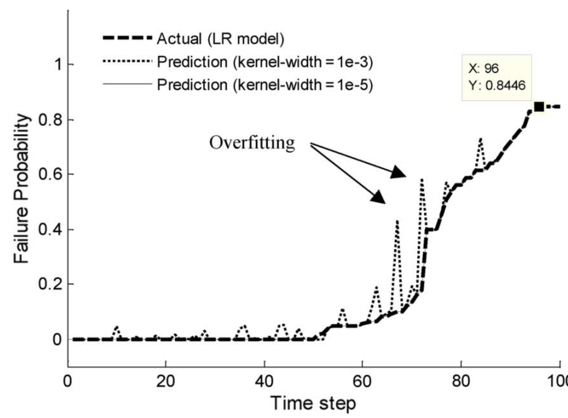

Fig. 5. Final failure prediction of bearing sample 15.

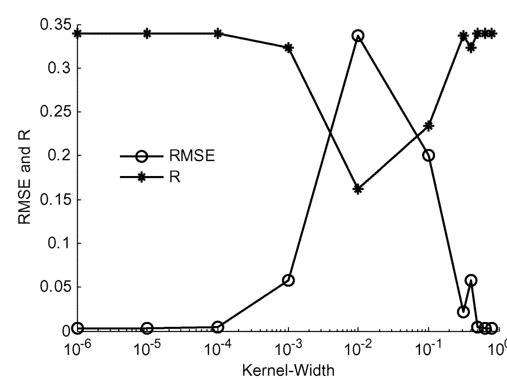

Fig. 6. Kernel-width selection of the RVM training stage (bearing sample 15).

RVM is trained by input data obtained from the kurtosis of simulated bearing failure data, and the target vector of failure probability estimated by LR. For RVM, 24 kurtosis data are utilized for training, and the remaining 1 data is used for testing. In this study, a Gaussian kernel is employed, and the kernel-width is searched in the range of $\{10^{-0.1}, 10^{-0.2}, \ldots, 10^{-6}\}$ to obtain an optimum RVM training process. The training process calculates the weight, and the bias. After training, RVM is employed to predict the failure time of bearing sample 15, which has a final failure time at 95, as shown in Fig. 3.

Fig. 5 shows the final failure predictions of bearing sample 15. Fig. 6 informs the kernel-width selection of RVM training. The optimum values of kernel-width are 10^{-4}, and $10^{-0.5}$, where kernel-widths less than 10^{-4}, and bigger than $10^{-0.5}$, give the same optimal value of $RMSE$, and R. The range between 10^{-3}

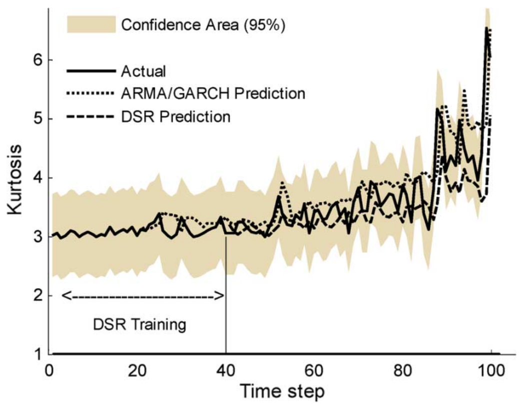

Fig. 7. ARMA/GARCH model prediction.

TABLE II
MULTI-STEP-AHEAD PREDICTION OF THE ARMA/GARCH MODEL

Number of bearing sample	Final failure (time step)	*RMSE*
2	104	0.378
10	102	0.466
12	108	0.483
14	105	0.469
20	104	0.375

and 10^{-1} of kernel-width results in over fitting shown by higher, and lower value of $RMSE$, and R, respectively. Using this optimal kernel-width value, the final failure prediction of bearing sample 15 reaches the failure state at 96.

The performance of prediction can be simply calculated as

$$\text{Accuracy} = \left(1 - \frac{|t_a - t_p|}{t_a}\right) \times 100\%$$
$$= \left(1 - \frac{|95 - 96|}{95}\right) \times 100\% = 98.95\%$$

where t_a refer to actual failure time, and t_p is the prediction of failure time. This prediction result seems overestimated, but the accuracy of 98.95% is acceptable for building the prognostic model.

To predict and estimate the final failure of censored data, the ARMA/GARCH model is employed. Fig. 7 shows the comparison between the prediction of DSR [18] and the ARMA/ GARCH model. DSR used one-step-ahead prediction of future states based on the training step. In this work, we set 40 data for training. Fig. 7 shows the prediction of the ARMA/GARCH model and DSR compared with the actual data. The prediction performance is signified by *RMSE*. The *RMSE* of the ARMA/GARCH model, and DSR is 0.3801, and 0.5588, respectively. Because the performance prediction of the ARMA/GARCH model is better than the DSR, the ARMA/GARCH model will be utilized for further processes. Notice that both prediction results are not exceeding the confidence area (95%). Table II shows the multi-step ahead prediction result of censored data using the ARMA/GARCH model. The observation is stopped at time step 104. Moreover, bearing samples 10, 12, 14, and 20 are used for training; and bearing sample 2 is used for testing. The failure probability of censored data prediction is also calculated using LR, then used for training the RVM. Table III shows the selected optimal kernel-width. The result of final failure predictions of bearing sample 2 is shown is Fig. 8. The accuracy of the prediction similar to (21) is 97.12%. This result seems to be an underestimation, and smaller than the prediction of bearing sample 15. However, in real applications, underestimation prediction is more reliable than overestimation prediction.

TABLE III
PERFORMANCE OF RVM TESTING FOR BEARING SIMULATION NO. 2

Kernel-width	*RMSE*	*R*
10^{-2}	1.537	0.69
10^{-3}	0.167	0.92
10^{-4}	0.041	0.98
10^{-5}	0.0006	0.99

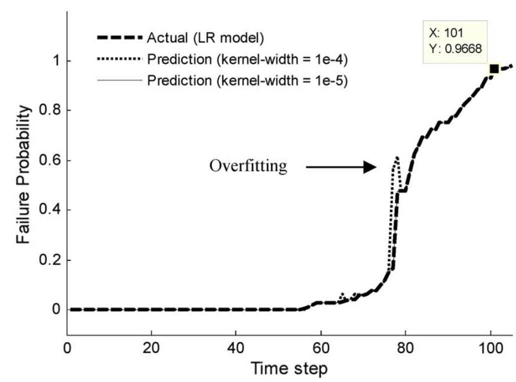

Fig. 8. Final failure prediction of bearing sample No. 2.

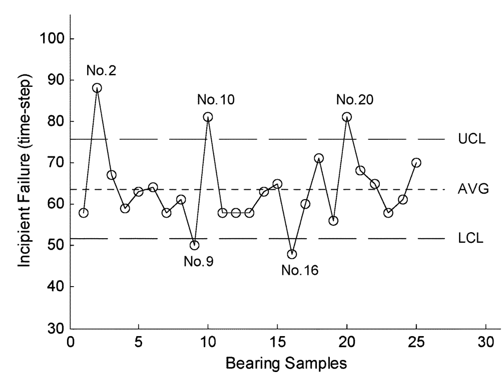

Fig. 9. SPC result of incipient failure data.

$$\text{Accuracy} = \left(1 - \frac{|t_a - t_p|}{t_a}\right) \times 100\%$$
$$= \left(1 - \frac{|104 - 101|}{104}\right) \times 100\% = 97.12\%$$

Figs. 9 and 10 show the SPC results of incipient, and final failure of bearing samples respectively. As previously mentioned, 25 bearing samples were simulated under the same bearing defect over time-steps. Five groups are set, and each group consists of 5 subgroups.

The average value of all sub averages and ranges are calculated to obtain the standard error. Based on (18) and (19), LCL and UCL can be estimated. In Fig. 8, the LCL, and UCL values are 51.59, and 75.53, respectively. In Fig. 9, the LCL, and UCL values are 83.29, and 102.78. Fig. 9 shows that bearing samples

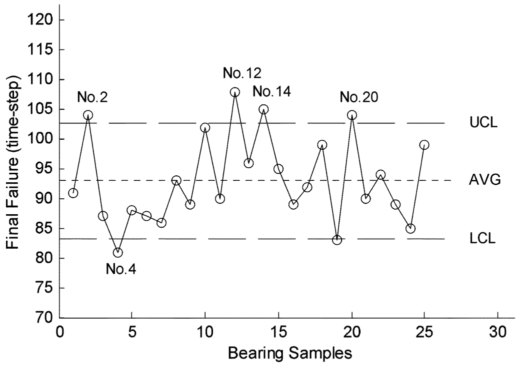

Fig. 10. SPC result of final failure data.

2, 9, 10, 16, and 20 exceed the LCL or UCL; while bearing samples 2, 4, 12, 14, and 20 are over the limit, as shown in Fig. 10. In the case of the experimental data, the results of SPC will give the benefit or feedback recommendation for bearing manufacturers.

In one type of bearing under similar loading conditions, we can recognize whether the failure time of individual bearing is different or not. If the variance of failure time is small, we can assume that the quality of bearings is good. Unfortunately, gathering the experimental run-to-failure bearing data is not an easy task. Thus, this paper employs simulated bearing data to give the illustration. The combination of RVM and LR has been validated using experimental data as discussed in [17].

VI. Conclusion

This study presents an assessment of failure degradation of simulated bearing fault data. Failure degradation in the form of target vectors is calculated by using LR. The best model of RVM is obtained from training process. Over fit prediction appears when the selected kernel-width value is improper. The ARMA/GARCH model is employed to predict the censored data. In addition, SPC is used to analyze the variance of failure data. Even though this proposed method is only applied to simulated data, the result illustrates the capability of the proposed method for failure degradation assessment.

References

[1] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, pp. 1483–1510, 2006.

[2] A. Heng, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, pp. 724–739, 2009.

[3] R. Khotamasu, S. H. Huang, and W. H. Ver Dui, "System health monitoring and prognostics—A review of current paradigms and practices," *International Journal of Advanced Manufacturing and Technology*, vol. 28, pp. 1012–1024, 1996.

[4] H. H. Lee, N. T. Nguyen, and J. M. Kwon, "Bearing diagnosis using time-domain features and decision tree," in *Lecture Notes in Computer Science*. Berlin/Heidelberg, Germany: Springer, 2007.

[5] X. Niu, L. Zhu, and H. Ding, "New statistical moments for the detection of defects in rolling elements bearing," *International Journal of Advanced Manufacturing Technology*, vol. 26, pp. 1268–1274, 2005.

[6] H. R. Martin and F. Honarvar, "Application of statistical moments for the bearing failure detection," *Applied Acoustic*, vol. 44, pp. 67–77, 1995.

[7] T. Williams, X. Ribadeneira, S. Billington, and T. Kurfess, "Rolling element bearing diagnostics in run-to-failure lifetime testing," *Mechanical Systems and Signal Processing*, vol. 15, pp. 979–993, 2001.

[8] J. Yan, M. Koc, and J. Lee, "A prognostic algorithm for machine performance assessment and its application," *Production Planning and Control*, vol. 15, no. 8, pp. 796–801, 2004.

[9] M. E. Tipping *et al.*, "The relevance vector machine," in *Advances in Neural Information Processing System*, S. Solla, T. Leen, and K. R. Muller, Eds. Cambridge, MA: MIT Press, 2000, pp. 287–289.

[10] M. E. Tipping, "Spare bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

[11] B. Schölkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA: MIT Press, 2002.

[12] H. T. Pham and B. S. Yang, "Estimation and forecasting of machine health condition using ARMA/GARCH model," *Mechanical Systems and Signal Processing*, vol. 24, pp. 546–558, 2010.

[13] P. D. McFadden and J. D. Smith, "Model for the vibration produced by a single point defect in rolling element bearing," *Journal of Sound and Vibration*, vol. 96, pp. 69–82, 1984.

[14] Y. F. Wang and P. J. Kootsookos, "Modeling of low shaft speed bearing faults for condition monitoring," *Mechanical Systems and Signal Processing*, vol. 12, no. 3, pp. 415–426, 1998.

[15] N. Z. Gebraeel, M. A. Lawley, R. Li, and J. K. Ryan, "Residual-life distribution from component degradation signals: A Bayesian approach," *IIE Transaction*, vol. 37, pp. 543–557, 2005.

[16] N. Gebraeel, "Sensory-updated residual life distribution for components with exponential degradation pattern," *IEEE Trans. Automation Science and Engineering*, vol. 3, no. 4, pp. 382–393, 2006.

[17] W. Caesarendra, A. Widodo, and B. S. Yang, "Application of relevance vector machine and logistic regression for machine degradation assessment," *Mechanical Systems and Signal Processing*, vol. 24, pp. 1161–1171, 2010.

[18] G. Niu and B. S. Yang, "Dempster-Shafer regression for multi-step ahead time series prediction towards data-driven machinery prognosis," *Mechanical Systems and Signal Processing*, vol. 23, pp. 740–751, 2009.

Wahyu Caesarendra received the B.Eng. degree from Diponegoro University, Indonesia, in 2005; and the M.Eng. degree from Pukyong National University, South Korea in 2010. Currently He is a Ph.D. student at School of Mechanical, Materials & Mechatronic Engineering, University of Wollongong, Australia. His research interest is machine condition monitoring and prognostics.

Achmad Widodo received the B.Eng., and M. Eng. degrees in Mechanical Engineering from Diponegoro University, and Bandung Institute Technology, Indonesia, in 1998 and 2002, respectively. He received the Ph.D. degree in Mechanical Engineering from Pukyong National University, Busan, South Korea in 2007. He is a lecturer in the Mechanical Engineering Department of Diponegoro University, Indonesia. His research interests include machine condition monitoring, fault diagnosis, and prognosis.

Pham Hong Thom received the B.Eng degree from Ho Chi Minh City University of Technology, Vietnam, in 2007; and the M.Eng. degree from Pukyong National University, South Korea in 2010. His research interest is in machine condition prognostics.

Bo-Suk Yang is a professor at the Department of Mechanical & Automotive Engineering, Pukyong National University, Korea. He received the Ph.D. degree in Mechanical Engineering from Kobe University, Japan in 1985. Now he is the leader of the Intelligent Mechanics Lab, and the Research Centre of Intelligent Machine Condition Monitoring and Diagnosis at PKNU.

Joga Dharma Setiawan received the B.Sc. degree in Mechanical Engineering from Northeastern University, Boston, USA in 1992; and the M.Sc. degree in Aeronautics/Astronautics from Massachusetts Institute of Technology, Cambridge, USA in 1996. He received the Ph.D. degree in Mechanical Engineering from Michigan State University, USA in 2001. He is currently a lecturer in the Mechanical Engineering Department of Diponegoro University, Indonesia. His research interests include dynamics and control of mechanical systems.