Health Condition Prediction of Gears Using a Recurrent Neural Network Approach

Zhigang Tian and Ming J. Zuo, *Senior Member, IEEE*

***Abstract*—The development of accurate health condition prediction approaches has been a key research topic in condition based maintenance (CBM) in recent years. However, current health condition prediction approaches are not accurate enough, which has become the bottleneck for achieving the full power of CBM. Neural network based methods have been considered to be a very promising category of methods for equipment health condition prediction. In this paper, we propose a neural network prediction model called extended recurrent neural network (ERNN). An ERNN based approach is developed for health condition prediction of gearboxes based on the vibration data collected from a gearbox experimental system. The results demonstrate the capability of the ERNN based approach for producing satisfactory health condition prediction results. A comparative study based on the gearbox experiment data further establishes ERNN as an effective recurrent neural network model for equipment health condition prediction.**

***Index Terms*—Gearbox, health condition, prediction, recurrent neural network.**

ACRONYMS

CBM	Condition based maintenance
FCRNN	Fully connected recurrent neural network
ERNN	Extended recurrent neural network
NMSE	Normalized mean squared error

NOTATION

$h(t)$	The function to fit the ERNN training data set
α	The scale parameter in $h(t)$
β	The shape parameter in $h(t)$
K	A constant to scale the function value in $h(t)$
n	The number of data points considered
$y(k)$	The actual value at time k
$\hat{y}(k)$	The estimated value with neural networks at time k
n_{FC}	The number of neurons in FCRNN
n_h	The number of hidden neurons in ERNN

Manuscript received February 23, 2009; revised December 22, 2009; accepted February 01, 2010. Date of publication October 21, 2010; date of current version November 30, 2010. This work was supported by the Natural Sciences and Engineering Research Council of Canada. Associate Editor: L. Cui.

Z. Tian is with Concordia Institute for Information Systems Engineering, Concordia University, Montreal, QC H3G 2W1 Canada (e-mail: tian@ciise.concordia.ca).

M. J. Zuo is with the Department of Mechanical Engineering, University of Alberta, Edmonton, AB T6G 2G8 Canada (e-mail: ming.zuo@ualberta.ca).

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.2010.2083231

I. INTRODUCTION

CONDITION BASED MAINTENANCE (CBM) is an advanced maintenance strategy to achieve the reliable, cost-effective operation of engineering systems. CBM is based on the understanding that a piece of equipment goes through multiple degraded states before it fails. These degraded states, or health conditions, can be monitored and predicted, and optimal maintenance actions can be scheduled for improving reliability while minimizing total operation costs [1], [2]. In this work, we focus on gears, which are critical components in aircraft systems, manufacturing systems, etc. The development of accurate health condition prediction approaches has been a key research topic in CBM in recent years [3]–[5]. However, current health condition prediction approaches are not accurate enough, which has become the bottleneck for achieving the full power of CBM.

One category of equipment health condition prediction methods are based on damage propagation physics [6]. However, damage propagation processes are typically very complex, and accurate physical models are difficult to build for many components and systems, which limits the application of this category of methods. Neural network based methods, which belong to data-driven methods, have been considered to be very promising for health condition prediction due to the adaptability, nonlinearity, and arbitrary function approximation ability of neural networks. Neural network methods do not assume the mathematical model of the damage propagation, but aim at modeling the degradation processes based on the collected condition monitoring data using neural networks, and perform health condition prediction. Huang et al. predicted the health condition of ball bearings using self-organizing maps, and back propagation neural networks methods based on vibration signals [7]. Lee et al. proposed to extract an overall health indicator based on the collected condition data, and predict future health indicator values using the autoregressive moving average (ARMA) method, and Elman neural networks [8]. Gebraeel et al. developed feedforward neural networks based methods for predicting the remaining useful life of ball bearings [9]. Tse and Atherton compared the prediction performance of various prediction approaches based on vibration signals from several industrial machines, and concluded that the Jordan network was superior to both statistical approaches

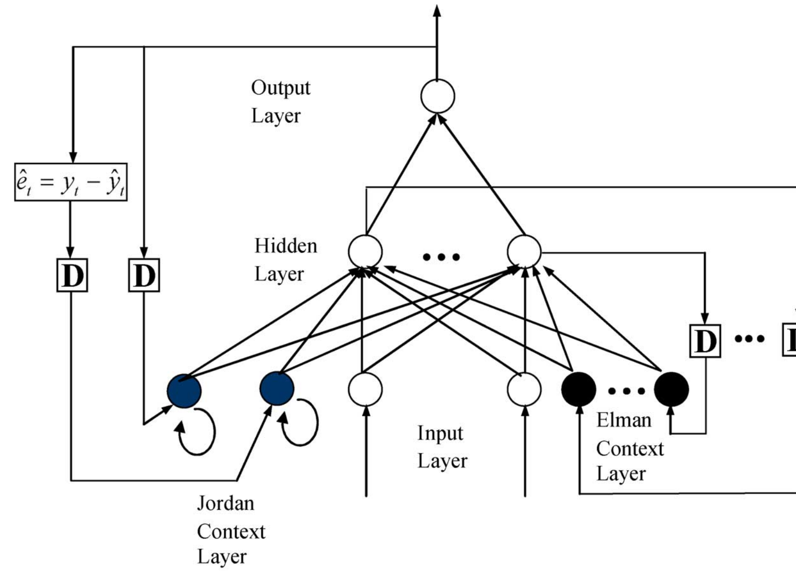

Fig. 1. The structure of ERNN.

and feedforward neural networks in one-step-ahead time series prediction [10].

For the purpose of time series prediction, a neural network can be considered to be a general nonlinear mapping between a subset of the past time series and the future time series values. The neural network models that have been used for prediction include feedforward neural networks, the Jordan network, the Elman network, and the fully connected recurrent neural network (FCRNN) [11]. The last three types of neural networks fall into the category of recurrent neural networks. Studies show that neural network approaches perform better than statistical autoregressive moving average (ARMA) methods, and recurrent neural networks are better than feedforward neural networks in time series prediction [12]. Thus, in this study, we focus on recurrent neural network prediction models. Part of this paper was published in the proceedings of the Sixth International Conference on Reliability, Maintainability & Safety [13].

II. The Extended Recurrent Neural Network Model

In this section, we propose the ERNN model, a new recurrent neural network prediction mode. The structure of the proposed ERNN is shown in Fig. 1. ERNN has two context layers, called the Elman context layer, and the Jordan context layer, respectively. The Elman context layer is the same as the context layer in the Elman network. In the Jordan context layer, there are two neurons with self-feedbacks: one neuron obtains inputs from the actual output $\hat{y}_t$ of the network after a delay of one time unit, and from itself; while the other neuron obtains inputs from the output error $\hat{e}_t$ of the network after a delay of one time unit, and from itself. For the purpose of predicting time series, ERNN has only one neuron in the output layer. We use 2 neurons in the input layer because it has been reported that every data point in a time series is only strongly dependent on the immediate past two values. The linear activation function, i.e. transform function, is used in the output layer, the Jordan context layer, and the Elman context layer. The sigmoid activation function is used in the hidden layer.

The activations of the hidden layer, V, is given by

$$\boldsymbol{V} = f^h\left((\boldsymbol{W}^h, \boldsymbol{W}^J, \boldsymbol{W}^E) \cdot (\boldsymbol{X}, \boldsymbol{Z}^J, \boldsymbol{Z}^E)^T\right), \tag{1}$$

where f^h is the sigmoid activation function. X is the input vector plus an additional column with all the elements equal to 1, Z^J is the activation vector of the Jordan context layer, and Z^E is the activation vector of the Elman context layer. W^h is the trainable weights from the input layer to the hidden layer, W^J is the weights from the Jordan context layer to the hidden layer, and W^E is the weights from the Elman context layer to the hidden layer. The activation of the output layer y is

$$y = f^o\left(\boldsymbol{W}^o \cdot \begin{pmatrix}\boldsymbol{V}\\ 1\end{pmatrix}\right). \tag{2}$$

where f^o is the linear activation function, and the bias in the output layer is incorporated into W^o. The activations of the Jordan and Elman context layers are

$$\boldsymbol{Z}^{J,(k)} = \begin{pmatrix}\boldsymbol{y}^{(k-1)}\\ \boldsymbol{e}^{(k-1)}\end{pmatrix} + \beta \cdot \boldsymbol{Z}^{J,(k-1)}, \quad \text{and} \tag{3}$$

$$\boldsymbol{Z}^{E,(k)} = \boldsymbol{V}^{(k-1)}, \tag{4}$$

where $\boldsymbol{Z}^{J,(k)}$ denotes the activation vector of the Jordan context layer at time point k, $y^{(k-1)}$ and $e^{(k-1)}$ are the output value and the output error respectively of the network at time point $(k-1)$, and β denotes the self-feedback connection weight in the Jordan context layer which is a fixed value between 0 and 1.

Theoretically, the original Elman network with feedback connections from the hidden layer to the context layer is capable of representing an arbitrary dynamic system, while the original Jordan network does not have this capability. Moreover, the context layer in the Elman network can store information for future references. Therefore, the incorporation of the Elman context layer in the proposed ERNN is expected to enhance its ability to model nonlinear, dynamic systems such as nonlinear time series.

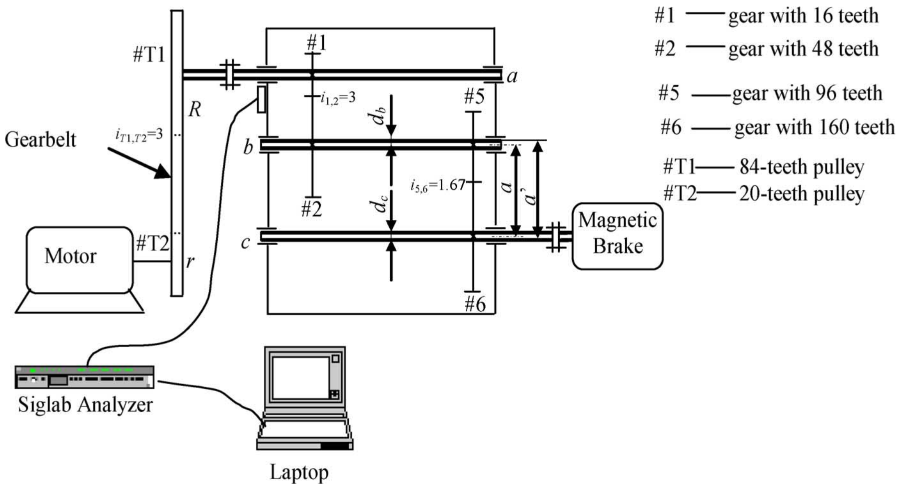

Fig. 2. The gearbox experimental system.

When the ERNN is used for predicting a certain time series, say series $\{x(t)\}$, it should be trained with $x(t-1), x(t-2), \ldots, x(t-p)$ as inputs, and $x(t)$ as the desired output. In other words, the current desired output is one of the inputs at the next time point. Therefore, feeding the current actual output of the network back to the Jordan context layer, and then presenting it as an additional input to the hidden layer, is helpful for the network to learn the time series. This is why the Jordan context neuron that receives the actual output of the network is meaningful. The self-feedbacks of the Jordan context neurons will improve the dynamic property of the proposed ERNN, and make it more stable. In addition, these self-feedbacks enable the Jordan context layer to store information not only of the current output, and the current output error at the current time point, but also of the output, and the output error values at many preceding time points. This feature will further enhance the proposed ERNN's ability to learn temporal sequences.

The ARMA model has been widely used for time series prediction. The ARMA model uses both the time series values, and the prediction errors at the previous time points as inputs to predict the value at the next time point. In such a model, the prediction errors at previous time points influence the predicted value at the next time point. We have incorporated this idea used in ARMA into the proposed ERNN network. A Jordan context neuron is used to feed the output error back to the hidden layer. The proposed ERNN is expected to perform better than models without such a context neuron.

Based on the discussions above, it can be seen that the architecture of the proposed ERNN is designed to be suitable for time series prediction. We expect that the proposed ERNN will perform better than other neural network models for prediction of time series.

According to the structure of the proposed ERNN model, presented in Fig. 1, there are two neurons in the Jordan context layer, and the number of the neurons in the Elman context layer is equal to that in the hidden layer. Thus, we can control the number of neurons in the hidden layer, and thus the number of neurons in the Elman context layer, to fit different problems properly. The general rule is, when the problem is large with a large number of data points, we need to use more neurons in the hidden layer. Meanwhile, we should try to minimize the number of hidden neurons to achieve good generalization capability. ERNN can be trained using the exact gradient based training algorithm [11]. The details of the training algorithm are omitted here.

III. The Gearbox Experimental System, and Data Processing

An approach based on the ERNN model is used to predict the health condition of the gearbox in the gearbox experimental system at the Reliability Research Lab of the University of Alberta, Canada [14].

A. The Gearbox Experimental System, and the Collected Vibration Data

The setup of the gearbox experimental system is shown in Fig. 2, along with some system parameters [14]. The speed of the motor is 2400 rpm; that is, the motor frequency is 40 HZ. The load applied to the gearbox is 40 Nm. An accelerated run-to-failure test of the gearbox was conducted. The vibration data was collected, and then re-sampled so that the interval between two neighboring points is 8 minutes. The total duration of the data set used in this study is 18.8 hours with 141 vibration data points. At the end of the 18.8 hours, a failure occurred. The gearbox was opened after the failure, and it was found that gears 5 and 6 were damaged with broken teeth, and out of meshing.

The root mean square (RMS) value is relatively flat before data point 110, and then it starts to increase, suggesting a degradation process of the gearbox. We are interested in predicting the future health condition of the gearbox, particularly when the gearbox starts to deteriorate, using an ERNN based approach. So we only use part of the vibration data, data points 90 to 141, for training ERNN for health condition prediction. The RMS plot with only these data points is shown in Fig. 3. We use data points 90 to 135, represented by "o" in Fig. 3, to train ERNN.

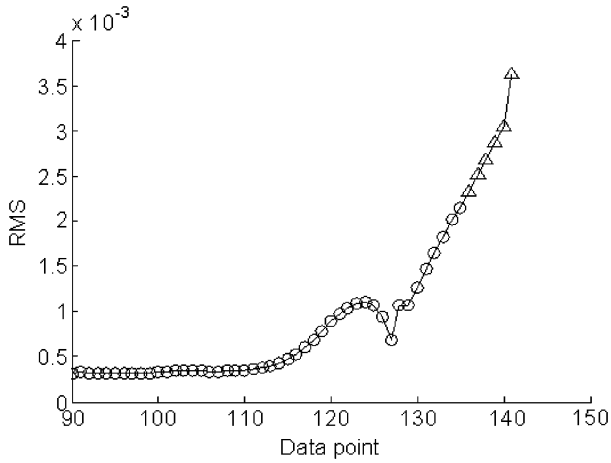

Fig. 3. The RMS data points 90 to 141.

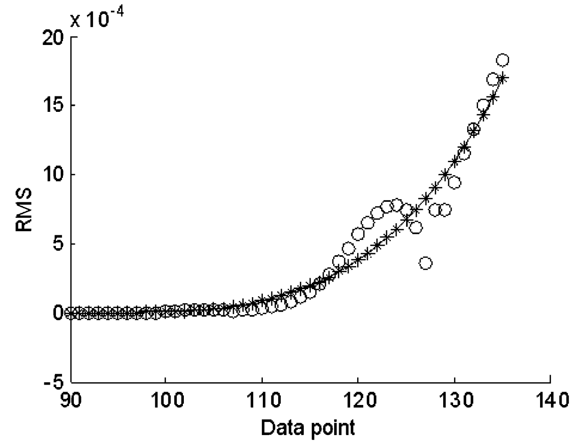

Fig. 4. The original and fitted RMS data sets: (o) the original data set; * the fitted data set.

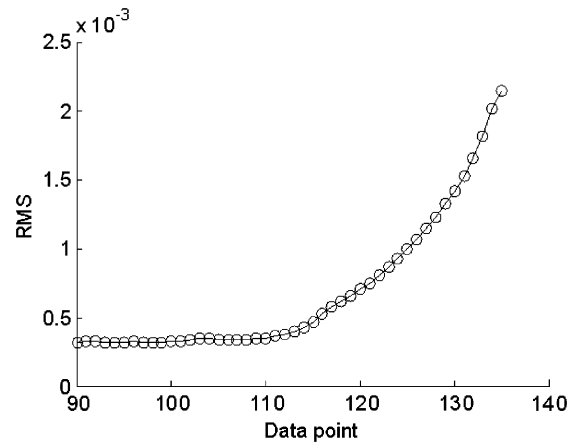

Fig. 5. The training data set for ERNN.

Data points 136 to 141, represented by "△", are used to test the prediction performance of the ERNN based approach.

B. Processing the Vibration Data With a Weibull Failure Rate Function Based Approach

From the RMS data set shown in Fig. 3, the vibration level starts to increase at approximately data point 110, which corresponds to 14.6 hours from the start of the continuous experiment. The understanding is that the vibration level increases as the gearbox deteriorates. However, the RMS curve fluctuates roughly between data points 117 and 131, which is likely due to environment noise or intervention. This will affect the prediction performance if we directly feed this data set to ERNN. Thus, we decided to fit the training set, data points 90 to 135, and use the fitted data points to replace the fluctuating segment of the data set, so as to remove the negative effect of the fluctuation on the prediction performance.

The vibration level, as shown in Fig. 3, roughly indicates the health condition of the gearbox. Meanwhile, in reliability engineering, it is the failure rate that indicates the health condition of a piece of equipment. The Weibull distribution is the most powerful, flexible lifetime distribution, and it is flexible enough to model the "wear out" portion of a piece of equipment's life. Thus, we propose to use a function based on the Weibull failure rate function to fit the training data set. The function used takes the form [19]

$$h(t) = K \frac{\beta}{\alpha^{\beta}} t^{\beta-1}. \quad (5)$$

We use this function to fit the training data set; and the fitted data, along with the original data, are shown in Fig. 4. The original data set is represented by "o", and the fitted data set is represented by "*". We replace data points 117 to 131 in the original data set with the corresponding portion in the fitted data set. The resulting training set is shown in Fig. 5, which shows a generally steady increase of the vibration level with the deterioration of the gearbox.

IV. GEARBOX HEALTH CONDITION PREDICTION USING THE ERNN BASED APPROACH

In this section, we use the gearbox dataset obtained from the previous section to investigate the performance of the ERNN based approach in predicting the gearbox health condition, and perform a comparative study between the ERNN and FCRNN based methods.

A. Gearbox Health Condition Prediction Results Using the ERNN Based Approach

Using the training data set with data points from 90 to 135, as shown in Fig. 5, we train the ERNN model. Based on the size of the problem, we use 2 hidden neurons, and 1 output neuron in the ERNN model. The one-step-ahead prediction performance is investigated. That is, first we train the ERNN model using data points 90 to 135, and use the trained ERNN model to predict the vibration RMS value for data point 136, then compare it with the actual RMS value for this point. Then we use actual data points 90 to 136 to train the ERNN model, use the trained ERNN model to predict the vibration RMS value for data point 137, and compare it with the actual RMS value for this point. Similarly, we perform one-step-ahead prediction for data points 138, 139, 140, and 141. The training and prediction processes

TABLE I
PREDICTION RESULTS FOR THE GEARBOX EXPERIMENT DATA

Prediction Measure	ERNN ($n_h = 2$)	FCRNN ($n_{FC} = 3$)	ERNN ($n_h = 1$)	FCRNN ($n_{FC} = 2$)
NMSE	0.1183	0.3238	0.1648	0.4337
Average normalized error	4.88%	7.01%	3.53%	9.56%

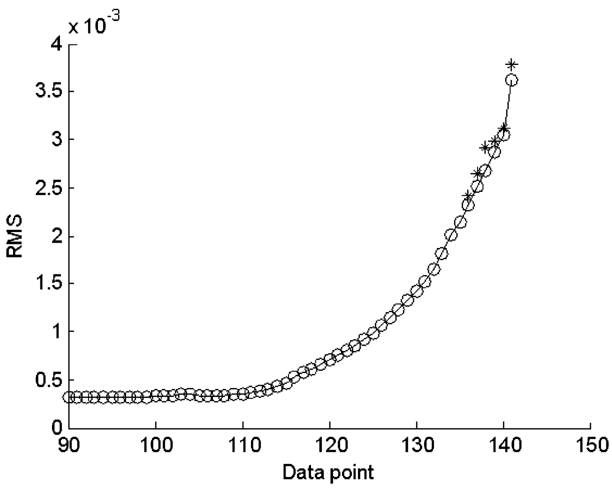

Fig. 6. The prediction results: (○) the actual data points; (*) the predicted data points.

are conducted 10 times to obtain the average prediction performance using the ERNN model.

We use the normalized mean squared error (NMSE) to compare the performances of different methods. The NMSE is given as

$$NMSE = \frac{\sum_{k=1}^{n} (y_k - \hat{y}_k)^2}{\sum_{k=1}^{n} (y_k - \bar{y})^2}, \tag{6}$$

where $\bar{y}$ is the mean of y_k $(k = 1, 2, \ldots, n)$. The NMSE is not affected by the data set size. The average training NMSE is 0.0012, and the average prediction NMSE is 0.1183. The prediction results are shown in Fig. 6, where the "○" points represent the real data, and the "*" points represent the average prediction values.

We might as well use another index to measure the prediction performance, given as

$$\text{Average normalized error} = \frac{1}{n} \sum_{k=1}^{n} \frac{|y_k - \hat{y}_k|}{y_k}, \tag{7}$$

where n is the number of data points investigated for the one-step-ahead prediction, which equals 6 in this problem. The average normalized error in the gearbox condition prediction is 0.0488, or 4.88%. In another words, in the gearbox run-to-failure experiment, using the ERNN based approach to predict the vibration level at the next time point, i.e. 8 minutes from the current time point, the average normalized prediction error is 4.88%, which is satisfactory. Here, we used the data from an accelerated failure experiment with a total duration of 18.8 hours. If we project it out to a practical gearbox under heavy load condition with a life expectancy of, say, 1 year, the duration between two data points will be approximately 2.6 days, and the maintenance staff will have time to act on health condition prediction results.

B. Comparative Study Between the ERNN and FCRNN Based Methods

To ensure a fair comparison, we use two input neurons, and one output neuron in ERNN; and use two external inputs, and one output in FCRNN. We want to keep the numbers of the trainable parameters, or trainable weights, of ERNN and FCRNN as close as possible, so that the differences in their prediction performances are only a result of the differences in their model architectures. For the ERNN model used in Section IV-A with 2 hidden neurons, the number of the trainable parameters is 19. Thus, to compare with this ERNN model, we use 3 neurons in the FCRNN model, which has 18 trainable parameters.

Based on the gearbox experiment data, we follow the same procedure as that in Section IV-A to perform one-step-ahead prediction using FCRNN. The average prediction is 0.3238, and the average normalized error is 7.01%. The one-step-ahead prediction results with ERNN and FCRNN are summarized in Table I, where n_h represents the number of hidden neurons in ERNN, and n_{FC} represents the number of neurons in FCRNN. From the results, we can see that both ERNN and FCRNN can produce reasonably good gearbox health condition prediction results. Secondly, according to both the NMSE measure values, and the average normalized error measure values, ERNN with 2 hidden neurons performs better than FCRNN with 3 neurons, while they have a close number of trainable parameters (19 versus 18).

We also compare another pair of ERNN and FCRNN models: an ERNN model with 1 hidden neuron, and a FCRNN model with 2 neurons, which both have 10 trainable parameters, and the results are also listed in Table I. The ERNN model again demonstrates better prediction performance than the FCRNN model. From Table I, comparing to the ERNN model with 2 hidden neurons, the ERNN model with 1 hidden neuron produces worse NMSE results, but better average normalized error result. We have also investigated ERNN models with other numbers of hidden neurons, and found that these two ERNN models actually give the best results. We have also investigated other FCRNN models with different numbers of neurons, and found that the FCRNN model with 3 neurons produces the best prediction results. To summarize, the results in this section suggest that both ERNN and FCRNN models can produce reasonably good gearbox health condition prediction results. For two ERNN and FCRNN models with similar numbers of trainable parameters, the ERNN model produces better prediction results.

V. Conclusions

Neural networks based methods have been considered to be very promising for equipment health condition prediction. In this paper, we propose a recurrent neural network prediction model called ERNN. The ERNN based approach is developed for the health condition prediction of gearboxes based on the vibration data collected from a gearbox experimental system. The results demonstrate the capability of the ERNN based approach for producing satisfactory health condition prediction results. The comparative study based on the gearbox experiment data further demonstrates ERNN as an effective neural network model for health condition prediction.

References

[1] A. K. S. Jardine, D. M. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] S. K. Yang, "A condition-based failure-prediction and processing-scheme for preventive maintenance," *IEEE Trans. Reliability*, vol. 52, no. 3, pp. 373–383, Sep. 2003.

[3] N. Gebraeel, A. Elwany, and J. Pan, "Residual life predictions in the absence of prior degradation knowledge," *IEEE Trans. Reliability*, vol. 58, no. 1, pp. 106–117, Mar. 2009.

[4] N. Gebraeel and J. Pan, "Prognostic degradation models for computing and updating residual life distributions in a time-varying environment," *IEEE Trans. Reliability*, vol. 57, no. 4, pp. 539–550, Dec. 2009.

[5] F. Peysson, M. Ouladsine, and R. Outbib, "A generic prognostic methodology using damage trajectory models," *IEEE Trans. Reliability*, vol. 58, no. 2, pp. 277–285, Jun. 2009.

[6] D. J. Inman, C. R. Farrar, and V. Lopes, Jr., *Damage Prognosis: For Aerospace, Civil and Mechanical Systems*. New York: John Wiley and Sons, 2005.

[7] R. Q. Huang, L. F. Xi, and X. L. Li, "Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods," *Mechanical Systems and Signal Processing*, vol. 21, no. 1, pp. 193–207, 2007.

[8] J. Lee, J. Ni, D. Djurdjanovic, H. Qiu, and H. Liao, "Intelligent prognostics tools and e-maintenance," *Computers in Industry*, vol. 57, pp. 476–489, 2006.

[9] N. Gebraeel and M. A. Lawley, "A neural network degradation model for computing and updating residual life distributions," *IEEE Trans. Automation Science and Engi.*, vol. 5, no. 1, pp. 154–163, 2008.

[10] P. W. Tse and D. P. Atherton, "Prediction of machine deterioration using vibration based fault trends and recurrent neural networks," *Journal of Vibration and Acoustics, Transactions of ASME*, vol. 121, pp. 355–362, 1999.

[11] R. Rojas, *Neural Networks: A System Introduction*. Berlin, Germany: Springer, 1996.

[12] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 240–254, 1994.

[13] Z. Tian, M. J. Zuo, and H. Z. Huang, "Extended recurrent neural network for time series prediction," in *Proceedings of the Sixth International Conference on Reliability, Maintainability & Safety*, Xi'an, China, 2004, pp. 669–674.

[14] X. F. Fan, H. Lin, and M. J. Zuo, Gearbox Lifetime Assessment Experiment—Experiment Design & Operation Guide University of Alberta, Oct. 2004, Technical report.

[15] W. Kuo and M. J. Zuo, *Optimal Reliability Modeling: Principles and Applications*. Hoboken: John Wiley & Sons, Inc., 2003.

Zhigang Tian is currently an Assistant Professor at Concordia Institute for Information Systems Engineering at Concordia University, Montreal, Canada. He received his Ph.D. degree in 2007 in Mechanical Engineering at the University of Alberta, Canada; and his M.S. degree in 2003, and B.S. degree in 2000 both in Mechanical Engineering at Dalian University of Technology, China. His research interests focus on reliability analysis and optimization, prognostics, condition monitoring, and maintenance optimization. He is a member of IIE and INFORMS.

Ming J. Zuo (S'89–M'89–SM'00) received the Bachelor of Science degree in Agricultural Engineering in 1982 from Shandong Institute of Technology, China; and the Master of Science degree in 1986, and the Ph.D. degree in 1989, both in Industrial Engineering from Iowa State University, Ames, Iowa, U.S.A. He is currently Professor in the Department of Mechanical Engineering at the University of Alberta, Canada. His research interests include system reliability analysis, maintenance modeling and optimization, signal processing, and fault diagnosis. He is Associate Editor of IEEE Transactions on Reliability, Regional Editor for North and South American region for International Journal of Strategic Engineering Asset Management, and Editorial Board Member of IIE Transactions, Journal of Traffic and Transportation Engineering, International Journal of Quality, Reliability and Safety Engineering, and International Journal of Performability Engineering. He also served as Department Editor of IIE Transactions (2005–2008). He is Senior Member of IEEE and IIE, and Founding Fellow of the International Society of Engineering Asset Management.