

Statistical aspects in neural network for the purpose of prognostics<sup>†</sup>Dawn An<sup>1,2</sup>, Nam H. Kim<sup>1</sup> and Joo-Ho Choi<sup>2,\*</sup><sup>1</sup>Department of Mechanical and Aerospace Engineering, University of Florida, 32611, Gainesville, FL, USA<sup>2</sup>Department of Aerospace and Mechanical Engineering, Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea

(Manuscript Received September 14, 2014; Revised November 30, 2014; Accepted December 10, 2014)

## Abstract

Neural network (NN) is a representative data-driven method, which is one of prognostics approaches that is to predict future damage/degradation and the remaining useful life of in-service systems based on the damage data measured at previous usage conditions. Even though NN has a wide range of applications, there are a relatively small number of literature on prognostics compared to the usage in other fields such as diagnostics and pattern recognition. Especially, it is difficult to find studies on statistical aspects of NN for the purpose of prognostics. Therefore, this paper presents the aspects of statistical characteristics of NN that are presumable in practical usages, which arise from measurement data, weight parameters related to the neural network model, and loading conditions. The Bayesian framework and Johnson distribution are employed to handle uncertainties, and crack growth problem is addressed as an example.

**Keywords:** Distribution-type data; Johnson distribution; Neural network; Prediction uncertainty; Prognostics

## 1. Introduction

Prognostics illustrated in Fig. 1 is to predict future damage/degradation and the remaining useful life (RUL) of in-service systems based on the damage data obtained at previous usage conditions, which facilitate condition-based maintenance known as cost effective maintenance strategy in company with diagnostics. Once a damage model (Black solid curve) is determined based on damage data at previous times under a given usage condition (Black dots) or under the various usage conditions (Grey dots), RUL which is remaining time/cycles before required maintenance can be predicted by progressing the damage state until it reaches the threshold. In general, prognostics methods can be categorized into data-driven [1], physics-based [2], and hybrid [3] approaches, based on the usage of information. Data-driven approaches use information from collected data to identify the characteristics of damage state without using any specific physical model; physics-based ones combine the physical model describing the behavior of damage with measured data; and hybrid ones integrate the other two methods to improve the prediction performance.

Since the physical model describing the behavior of damage rarely exists, data-driven approaches have a wide range of applications. It includes neural network (NN) [4, 5], Gaussian

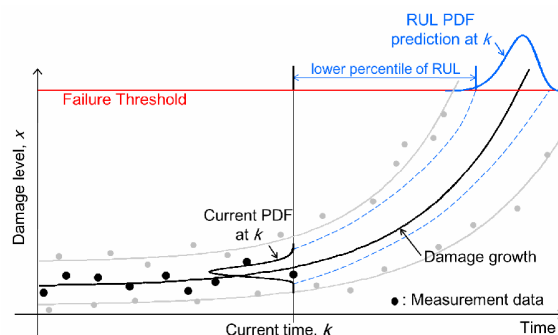


Fig. 1. Illustration of prognostics.

process regression [6, 7], relevance vector machine [8], least square regression [9], etc. Among these algorithms, NN is a representative data-driven method, in which a network model learns a way to produce a desired output such as future damage level reacting to given inputs such as previous damage level and usage conditions instead of physical model. The learning process is the same as finding weight parameters associated with the network model by minimizing the mean square error between measurement data and network outputs, which is called training process, and the data used for training expressed as dotted markers in Fig. 1 is called training data.

In general, weight parameters are obtained as deterministic values by using an optimization process, and prediction uncertainties are added with confidence bounds based on nonlinear regression and/or the error between NN outputs and training

\*Corresponding author. Tel.: +82 2 300 0117, Fax.: +82 2 3158 2191

E-mail address: jhchoi@kau.ac.kr

<sup>†</sup>This paper was presented at the 8th CJK-OSM, Gyeongju, Korea, May 2014.

Recommended by Guest Editor Joo-Ho Choi

© KSME & Springer 2015

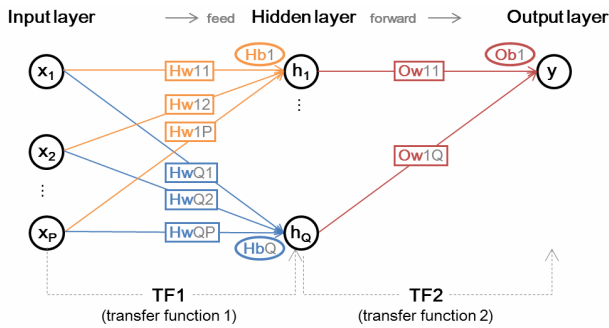


Fig. 2. Illustration of typical network model: FFNN.

data [10–13]. It, however, is difficult to find global optimum of parameters due to measurement noise, a small number of data compared to the number of parameters, and the complexity of damage growth, which can yield a significant error in prediction results. On the other hand, Bayesian NN (BNN) [14, 15] has been proposed to resolve local optimum problem, which provides distribution of prediction results caused by measurement error and uncertainty in parameters that are identified as distributions based on Bayes' theorem instead of deterministic values given by an optimization process. There are no literatures that employ BNN for the purpose of prognostics, though. Liu et al. [16] repeated NN process 50 times to predict battery's RUL, which is similar to BNN in a sense of employing randomness of weight parameters.

In addition to general statistical aspects mentioned in the previous paragraph, additional issues that are presumable in practical usages are also addressed. Data used for input variables have mostly been considered as deterministic values, but they can be distributed. In such a case, there are no clear damage indicators, many number of damage data are given at the same usage conditions from the same system, and usage conditions such as loading conditions can also have uncertainties and need to be considered as distributions. This case as well as general statistical aspects will be considered with a crack growth example.

The paper is organized as follows: in Sec. 2, the process of NN is explained for the purpose of prognostics with a crack growth example; and in Sec. 3, statistical aspects are considered based on the understanding of NN, followed by discussions and conclusions in Sec. 4.

## 2. Neural network

### 2.1 Network model

A typical architecture of NN is feed-forward neural network (FFNN) [17], which is illustrated in Fig. 2. In the figure, circles represent nodes (Also called neuron or unit), and a set of nodes in the same column is called a layer. The nodes in the input and output layer, respectively, represent input variables and response variable. Since the given information for data-driven approaches are only measurement data, previous damage data and the current damage data are, respectively, usually

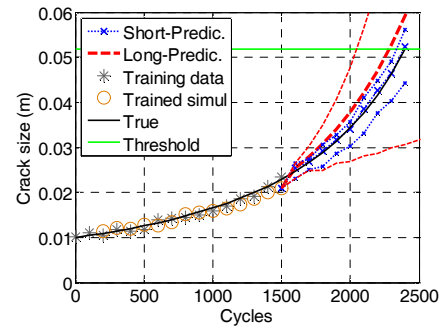


Fig. 3. Example of NN-based prognostics with crack growth example.

employed for input and output variables. And then, the number of nodes in the hidden layer can be adjusted to properly express the mechanism between input and output by receiving signals from input layers and forwarding them to the output layer. Even though the network model that includes selecting the number of hidden nodes, hidden layers and input nodes has an effect on the prediction results, it is not considered here because the network problem is a different issue from statistical ones as well as trial-and-error methods are often used to determine a suitable network model.

Once the network model is determined, the model is functionalized using transfer functions and weight parameters. Transfer functions characterize the relationship between layers, and several types of transfer function are available such as sigmoid, inverse, and linear function [18]. Usually, the tangent sigmoid and pure linear functions are employed as a common practice. Weight parameters include weights for the interconnected nodes and biases that are added to inputs of transfer functions [16, 19], which are shown as rectangles and ellipses in Fig. 2, respectively. The process of finding the weight parameters is called training or learning, and to accomplish that, usually many sets of training data are required.

In general, FFNN is often called a back-propagation neural network (BPNN) because weight parameters are obtained through the learning/optimization algorithm [20] that adjusts weight parameters through backward propagation of errors between actual output (Training data) and the one from the network model based on gradient descent optimization methods. In other words, FFNN and BPNN are, respectively, to calculate the response forward and to update weight parameters based on the response backward. Once the network model learns enough the relationship between inputs and output, it can be used for the purpose of prognosis. In the following, the process of NN-based prognostics becomes specified with crack growth example.

### 2.2 The process of NN with a crack growth example

Fig. 3 shows an example of NN-based prognostics for a crack growth problem. The star markers are assumed as crack growth data measured at every 100 cycles in a fuselage panel under repeated pressurization loadings, which are generated

based on Paris model [21] with true damage growth parameters  $m_{\text{true}} = 3.8$ ,  $C_{\text{true}} = 1.5 \times 10^{-10}$ , the initial half crack size  $a_0 = 10$  mm, load magnitude  $\Delta\sigma = 80$  MPa, and random noise that is uniformly distributed between  $-1.0$  mm and  $+1.0$  mm. Note that the true values of parameters are used only for the purpose of generating measurement data in this paper.

The network model is constructed based on aforementioned FFNN with two input nodes, one hidden layer with one node; and thus, the number of total weight parameters become 5 including three weights ( $2 \times 1 + 1 \times 1$ ) and two biases ( $1 + 1$ ). For input variables, damage data ( $x_{k-2}, x_{k-1}$ ) at the previous two 100 cycles are used, and the current damage data ( $x_k$ ) becomes the output,  $k$  is the current time index. If  $k = 16$  (The current cycle is 1500 cycles), 14 sets of input and output data are available, which are the training data used to obtain weight parameters via optimization process. Then future damages ( $x_{k+1}^p, x_{k+2}^p, x_{k+3}^p, \dots$ ) are predicted based on the obtained weight parameters and the previous damage data, i.e., input variables. According to the previous damage data used as inputs, prediction methods can be divided into short term prediction and long term prediction. Short term prediction is one-step ahead prediction since it uses only measured data for input, e.g.,  $x_{k+1}, x_{k+2}$  are inputs to predict  $x_{k+3}^p$ . On the other hand, long term prediction is multi-step ahead prediction since it utilizes predicted results as inputs, e.g.,  $x_{k+1}^p, x_{k+2}^p$  are inputs to predict  $x_{k+3}^p$ .

Future damage prediction results are shown in Fig. 3. In the figure, thick dotted curve and thick dashed curve are, respectively, the median of short term prediction and long term prediction obtained by repeating NN 30 times, and their thin curves mean 90% confidence intervals. The wide range of long term prediction interval means that the results become significantly different whenever the NN process is performed due to the local optimum problem, even though the training simulation results shown as circles are close to the training data shown as gray star makers. Nevertheless, NN can be used for the purpose of prognostics by employing proper statistical methods. Although repeating the process to obtain statistical distribution can be a way, a more logical method is introduced in the next section.

### 3. Statistical aspects in NN

In the following subsections, different statistical aspects that are presumable in practical usages are considered according to given information.

#### 3.1 Uncertainty in prognostics results

The first case is a common condition caused by noise in measurement data and parameter identification, and it is to identify the weight parameters as distribution based on Bayesian framework. Bayesian inference is a statistical method in which observations are used to estimate and update unknown

parameters such as weight parameters in the form of a probability density function (PDF). Bayesian inference is based on the following Bayes' theorem [22]:

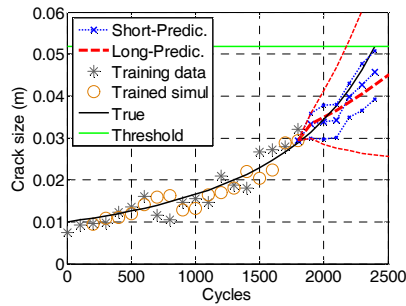
$$p(\boldsymbol{\theta}|\mathbf{z}) \propto L(\mathbf{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (1)$$

where  $\boldsymbol{\theta}$  is a vector of unknown parameters,  $\mathbf{z}$  a vector of observed data,  $L(\mathbf{z}|\boldsymbol{\theta})$  the likelihood,  $p(\boldsymbol{\theta})$  the prior PDF of  $\boldsymbol{\theta}$ , and  $p(\boldsymbol{\theta}|\mathbf{z})$  the posterior PDF of  $\boldsymbol{\theta}$  conditional on  $\mathbf{z}$ . The likelihood is the PDF value of  $\mathbf{z}$  conditional on given  $\boldsymbol{\theta}$ , and the prior information can be given, assumed, or not considered. The reliability of posterior PDF increases as more data are used, which gives more accurate and precise prediction results of damage and RUL.

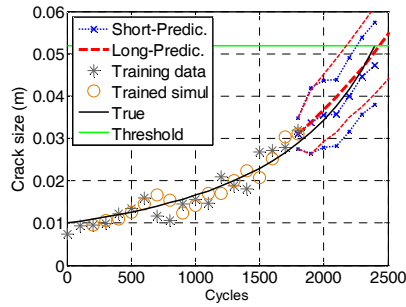
Fig. 4 shows the comparison between repeating NN (RNN) and Bayesian NN (BNN). Figs. 4(a) and (b) are the same condition as the previous example in Sec. 2.2 but with a larger level of noise,  $\pm 5$  mm. Figs. 4(c) and (d) are also crack growth problem, but they are based on Huang's model [23] that express crack growth under variable amplitude loading condition, which is employed to show the case of complex damage model. Although the actual crack size cannot be reduced, the measured crack size can be reduced due to the random nature of noise. In both cases of large noise and complex model, BNN outperforms RNN in terms of accuracy and precision of future damage prediction. The two cases means severe prediction conditions, but such conditions are more likely to be in real damage data. If the damage data have small level of noise and the damage growth increase monotonically, it will be more efficient to use RNN as in Ref. [16]. This can be explained by the following two reasons: (1) the results obtained by repeating NN more than 30 times do not change much with other attempts, which gives more reliable results compared to use NN just one time with confidence bounds, and (2) since it is hard to identify the distributions of weight parameters as the number of parameters depending on network model increases, the performance of BNN is deteriorated in constructing network model.

#### 3.2 Distribution-type data

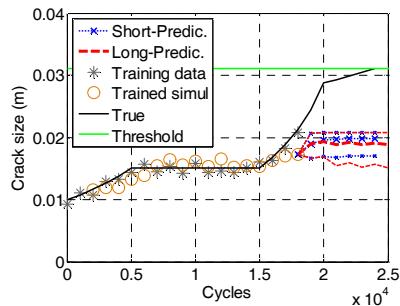
Input variables of NN are composed of damage data and usage conditions that are considered as deterministic values, and never considered as distributions. However, input variables can be distributed in such cases: many number of damage data are given at the same usage conditions from the same system, usage conditions such as loading conditions are uncertain, and there are no clear damage indicators. Johnson distribution [24] having four parameters, four quantiles corresponding to probabilities of 0.0668, 0.3085, 0.6915 and 0.9332, is employed to predict future damage distribution. Fig. 5 shows examples of Johnson distribution in cases of normal and beta distribution. The black solid curves are exact probability density function (PDF) from each distribution, and the bars are the results from Johnson distribution using four quantiles rep-



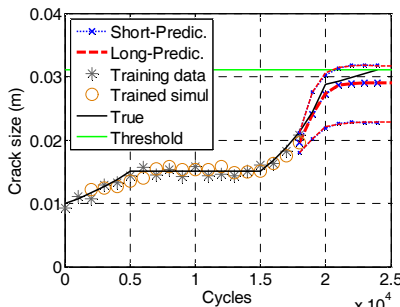
(a) RNN with large noise



(b) BNN with large noise



(c) RNN with complex model

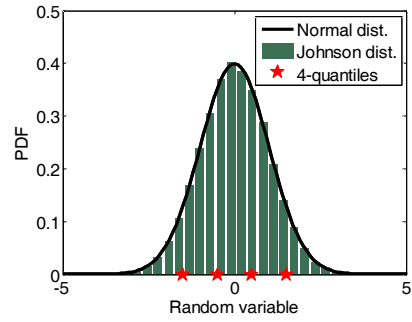


(d) BNN with complex model

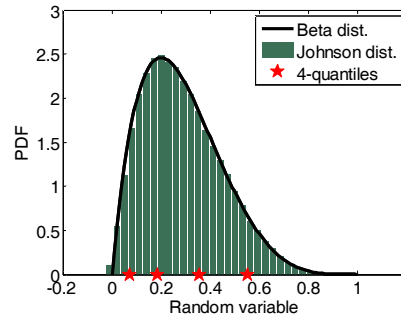
Fig. 4. Comparison example of NN-based prognostics with crack growth example.

resented as red star markers. Johnson distribution can express any other distribution types when the four quantiles are correctly given.

The same crack growth example as in the previous one is again employed to demonstrate the case of random input variable. Distributed synthetic data are generated from the pertur-

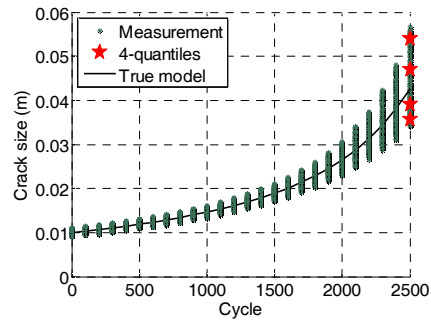


(a) Standard normal distribution

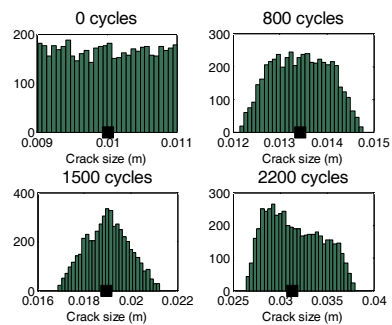


(b) Beta distribution with  $\alpha = 2, \beta = 5$

Fig. 5. Example of Johnson distribution.



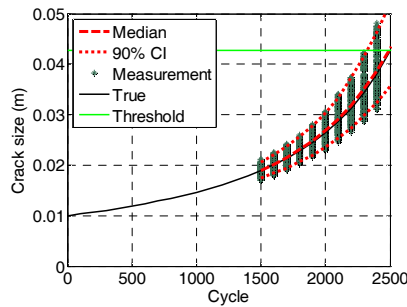
(a) Measured data at every 100 cycles



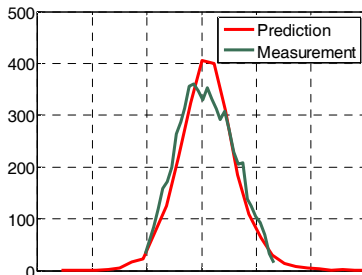
(b) Distribution of measured data

Fig. 6. Distributed synthetic data.

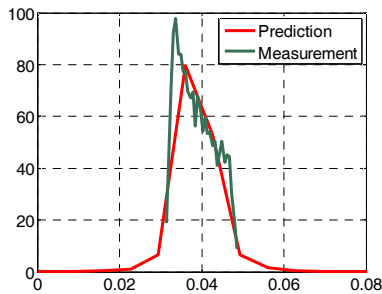
bation of Paris model parameter  $m$  with a deterministic load magnitude, and then a small level of noise is added:  $m \sim U(3.8 - 0.027, 3.8 + 0.027)$ ,  $\Delta\sigma = 78$  MPa,  $noise \sim U(-1, +1)$  mm whose result is shown in Fig. 6. Each cycle has 5000



(a) Damage growth prediction



(b) Damage distribution at 1600 cycles



(c) Damage distribution at 2400 cycles

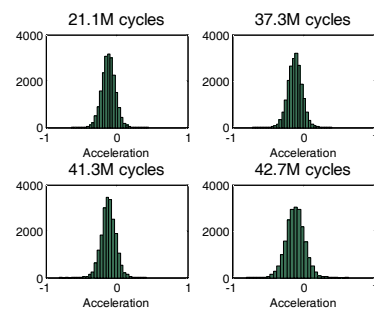
Fig. 7. Damage prediction results.

samples as the measurement data, whose distribution at 0, 800, 1500, and 2200 cycles are shown in Fig. 6(b) with their true damage size shown as black squares. It is shown that the shape of distributions is changed as cycle increases. Four quantiles whose example at 2500 cycles is shown as red star markers in Fig. 6(a) are used for input variables. Since there were two input variables and one output variable in the previous study, total numbers of input and output variables become eight (2 × 4) and four (1 × 4), respectively.

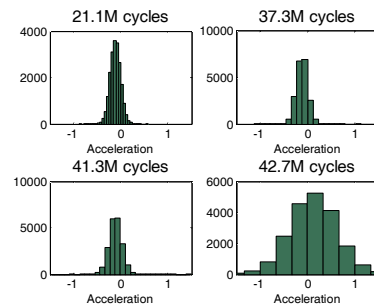
Fig. 7 shows damage prediction results from RNN with Johnson distribution at 1500 cycles. In Fig. 7(a), the median of future damage growth is very close to the true one, and 90% confidence interval also covers damage distribution at every cycle. Figs. 7(b) and (c) show comparison of damage distribution between predicted one and measured one at 1600 and 2400 cycles, and their errors are listed in Table 1. The maximum magnitude of error is 5.75% at 2400 cycles that is 900 cycles ahead prediction from 1500 cycles. These results show that NN using Johnson distribution is applicable for prediction

Table 1. Errors between prediction results and measurement at 1500 cycles.

Cycles		1600	1800	2000	2200	2400
6.7Q	Measurement	0.0186	0.0210	0.0239	0.0276	0.0327
	Prediction	0.0186	0.0210	0.0240	0.0279	0.0332
	Error (%)	0.03	0.02	0.58	0.95	1.82
30.9Q	Measurement	0.0196	0.0221	0.0253	0.0296	0.0354
	Prediction	0.0197	0.0225	0.0261	0.0308	0.0374
	Error (%)	0.63	1.54	2.89	4.18	5.75
69.1Q	Measurement	0.0207	0.0237	0.0278	0.0333	0.0413
	Prediction	0.0207	0.0237	0.0276	0.0329	0.0402
	Error (%)	0.10	0.01	0.59	1.35	2.66
93.3Q	Measurement	0.0218	0.0251	0.0296	0.0363	0.0463
	Prediction	0.0218	0.0253	0.0299	0.0363	0.0453
	Error (%)	0.32	0.82	0.78	0.09	2.20



(a) Acceleration: failure not occurred



(b) Acceleration: failure occurred

Fig. 8. Real damage data.

of damage distribution.

Lastly, Fig. 8 shows real distribution-type measurement data from the bearing provided by the Center for Intelligent Maintenance Systems [25]. Vibration signal is monitored using accelerometer during one second with 20 kHz sampling rate at specific intervals. The distributions in Figs. 8(a) and (b) are, respectively, observed from a bearing without failure and a bearing with failure. While the distribution of the case without failure does not changes much, the distribution with failure becomes wider and its mode shifts to the value greater than zero as cycles increase. Even though it has not been fully explored to consider the change of distribution as the damage indicator (There are no clear criteria of damage threshold yet),

the results in this section show that this method has a possibility to be employed for prognosis when there are no clear/deterministic damage indicators like this bearing problem.

#### 4. Conclusions

The goal of this paper is to address the aspects of statistical aspects in NN that are presumable in practical usages. As the first case, RNN and BNN are compared in terms of prediction uncertainty that is general statistical aspect related with noise in measurement data and weight parameters in NN model. BNN outperforms RNN under severe prediction conditions such as large level of noise in data and complex damage growth. In another case, random input variables are handled by employing Johnson distribution to NN. Future damage distribution are well predicted, and the results show that the method considering the change of distribution has a possibility to be employed for prognosis when there are no clear damage indicators.

#### Acknowledgment

This work was supported by a grant from the International Collaborative R&D Program (0420-2011-0161) of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), funded by the Korean government's Ministry of Knowledge Economy.

#### References

- [1] M. A. Schwabacher, A survey of data-driven prognostics, *AIAA Infotech@Aerospace Conference*, Reston, VA (2005).
- [2] J. Luo, K. R. Pattipati, L. Qiao and S. Chigusa, Model-based prognostic techniques applied to a suspension system, *IEEE Transactions on System, Man and Cybernetics*, 38 (5) (2008) 1156-1168.
- [3] J. Yan and J. Lee, A hybrid method for on-line performance assessment and life prediction in drilling operations, *IEEE International Conference on Automation and Logistics*, Jinan, Shandong, China (2007).
- [4] K. Chakraborty, K. Mehrotra, C. K. Mohan and S. Ranka, Forecasting the behavior of multivariate time series using neural networks, *Neural Networks*, 5 (1992) 961-970.
- [5] X. Yao, Evolving artificial neural networks, *Proceedings of the IEEE*, 87 (9) (1999) 1423-1447.
- [6] M. Seeger, Gaussian processes for machine learning, *International Journal of Neural Systems*, 14 (2) (2004) 69-106.
- [7] S. Mohanty, R. Teale, A. Chattopadhyay, P. Peralta and C. Willhauck, Mixed Gaussian process and state-space approach for fatigue crack growth prediction, *International Workshop on Structural Health Monitoring*, 2 (2007) 1108-1115.
- [8] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, *Journal of Machine Learning Research*, 1 (2001) 211-244.
- [9] V. T. Tran and B. S. Yang, Data-driven approach to machine condition prognosis using least square regression tree, *Journal of Mechanical Science and Technology*, 23 (2009) 1468-1475.
- [10] G. Chryssolouris, M. Lee and A. Ramsey, Confidence interval prediction for neural network models, *IEEE Transactions on Neural Networks*, 7 (1) (1996) 229-232.
- [11] R. D. Veaux, J. Schumi, J. Schweinsberg and L. H. Ungar, Prediction intervals for neural networks via nonlinear regression, *Technometrics*, 40 (4) (1998) 273-282.
- [12] L. Yang, T. Kavli, M. Carlin, S. Clausen and P. F. M. Groot, An evaluation of confidence bound estimation methods for neural networks, *European Symposium on Intelligent Techniques*, Aachen, Germany (2000).
- [13] J. A. Leonard, M. A. Kramer and L. H. Ungar, A neural network architecture that computes its own reliability, *Computers in Chemical Engineering*, 16 (9) (1992) 819-835.
- [14] J. F. G. de Freitas, Bayesian methods for neural networks, *PhD Thesis*, University of Cambridge, UK (2003).
- [15] R. M. Neal, Bayesian learning for neural networks, *Ph.D. Thesis*, University of Toronto, Ontario, Canada (1995).
- [16] J. Liu, A. Saxena, K. Goebel, B. Saha and W. Wang, An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries, *Annual Conference of the Prognostics and Health Management Society*, Portland, Oregon, USA (2010).
- [17] D. Svozil, V. Kvasnička and J. Pospíchal, Introduction to multi-layer feed-forward neural networks, *Chemometrics and Intelligent Laboratory Systems*, 39 (1997) 43-62.
- [18] W. Duch and N. Jankowski, Survey of neural transfer functions, *Neural Computing Surveys*, 2 (1999) 163-212.
- [19] A. E. Firth, O. Lahav and R. S. Somerville, Estimating photometric redshifts with artificial neural networks, *Monthly Notices of the Royal Astronomical Society*, 339 (2003) 1195, DOI: 10.1046/j.1365-8711.2003.06271.
- [20] D. E. Rumelhart, G. E. Hinton and R. J. Williams, Learning internal representations by error propagation, *In Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: foundations, MIT Press (1986) 318-362.
- [21] P. C. Paris and F. Erdogan, A critical analysis of crack propagation laws, *Transactions of the ASME, Journal of Basic Engineering, Series D*, 85 (3) (1963) 528-534.
- [22] T. Bayes, An essay towards solving a problem in the doctrine of chances, *Philosophical Transactions of the Royal Society of London*, 53 (1763) 370-418.
- [23] X. Huang, M. Torgeir and W. Cui, An engineering model of fatigue crack growth under variable amplitude loading, *International Journal of Fatigue*, 30 (1) (2008) 2-10.
- [24] N. L. Johnson, Systems of frequency curves generated by methods of translation, *Biometrika*, 36 (1949) 149-176.
- [25] J. Lee, H. Qiu, G. Yu, J. Lin and Rexnord Technical Services, *Bearing data set*, IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, <http://ti.arc.nasa.gov/project/prognostic-data-repository>, NASA Ames, Moffett Field, CA (2007).



**Dawn An** received the B.S. degree and M.S. degree of mechanical engineering from Korea Aerospace University in 2008 and 2010, respectively. She is now a joint Ph.D. student at Korea Aerospace University and the University of Florida. Her current research is

focused on prognostics methods for practical considerations, such as a large noise in data, insufficient number of data, and indirectly measured data.



**Nam-Ho Kim** received the B.S. degree of mechanical engineering from Seoul National University in 1989, the M.S. and Ph.D. degrees of mechanical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1991 and the University of Iowa in 1999, respectively. He worked as a Postdoctoral

Associate at the University of Iowa from 1999 to 2001. He joined the Dept. of Mechanical & Aerospace Engineering at the University of Florida, in 2002 and is now Professor. His current research is focused on design under uncertainty, design optimization of automotive NVH problem, shape DSA of transient dynamics (Implicit/explicit) and structural health monitoring.



**Joo-Ho Choi** received the B.S. degree of mechanical engineering from Hanyang University in 1981, the M.S. and Ph.D. degrees of mechanical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1983 and 1987, respectively. During the year 1988, he worked as a Postdoctoral

Fellow at the University of Iowa. He joined the School of Aerospace and Mechanical Engineering at Korea Aerospace University, Korea, in 1997 and is now Professor. His current research is focused on the reliability analysis, design for lifetime reliability, and prognostics and health management.