Statistical Aspects in Neural Network for the Purpose of Prognostics

Dawn An[1]
University of Florida, Gainesville, FL, 32611
Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea

Nam Ho Kim[2]
University of Florida, Gainesville, FL, 32611

and

Joo Ho Choi[3]
Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea

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.

I. Introduction

PROGNOSTICS illustrated in Figure 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 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

[1] Joint Ph.D. student, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611, Student member.
[2] Associate professor, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611, Member.
[3] Professor, School of Aerospace and Mechanical Engineering, Korea Aerospace University, 100 Hwajeon-dong, Deokyang-gu, Goyang-City, Gyeonggi-do, Korea, 412-791.

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 Figure 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 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 Section Ⅱ, the process of NN is explained for the purpose of prognostics with a crack growth example; and in Section Ⅲ, statistical aspects are considered based on the understanding of NN, followed by discussions and conclusions in Section Ⅳ.

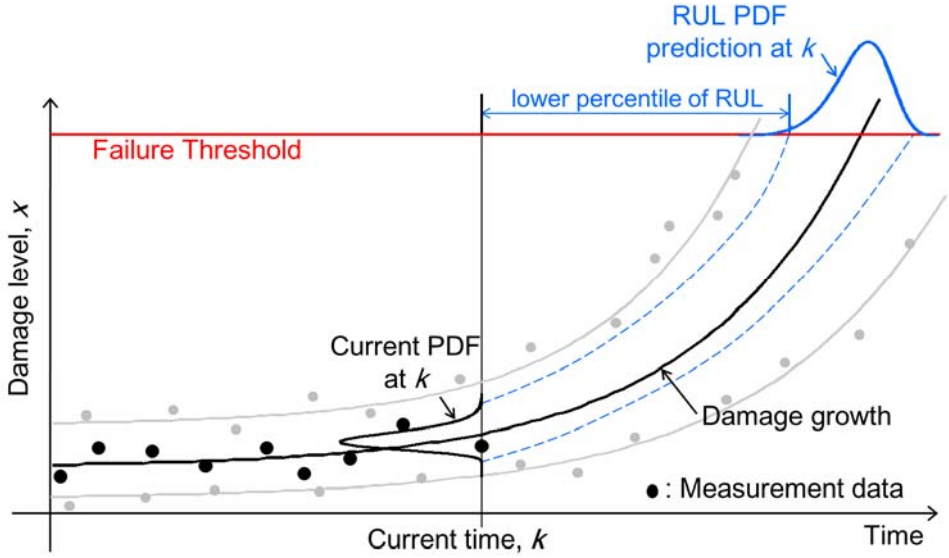


Figure 1. Illustration of prognostics.

II. Neural Network

A. Typical Network Model

A typical architecture of NN is feed-forward neural network (FFNN),[17] which is illustrated in Figure 2. In the figure, circles represent nodes (also called neuron or unit), and each 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 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 each layer, 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 way. 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 Figure 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.

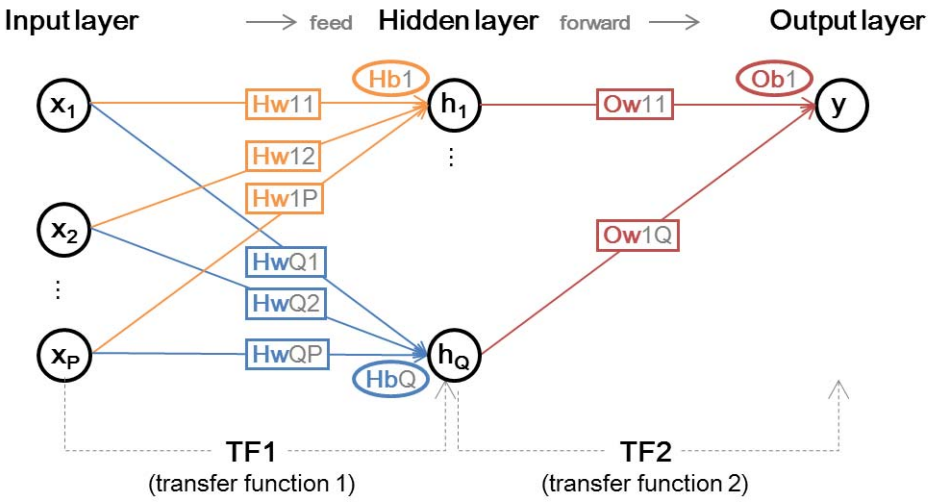


Figure 2. Illustration of typical network model: FFNN.

B. The Process of NN with a Crack Growth Example

Figure 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 $a_0 = 10\text{mm}$, load magnitude $\Delta\sigma$ =80MPa, 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 ($2\times1+1\times1$) 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 ($x_{k+1}^p, x_{k+2}^p, x_{k+3}^p, \ldots$) 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^p_{k+3}. On the other hand, long term prediction is multi-step ahead prediction since it utilizes predicted results as inputs, e.g., x^p_{k+1}, x^p_{k+2} are inputs to predict x^p_{k+3}.

Future damage prediction results are shown in Figure 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.

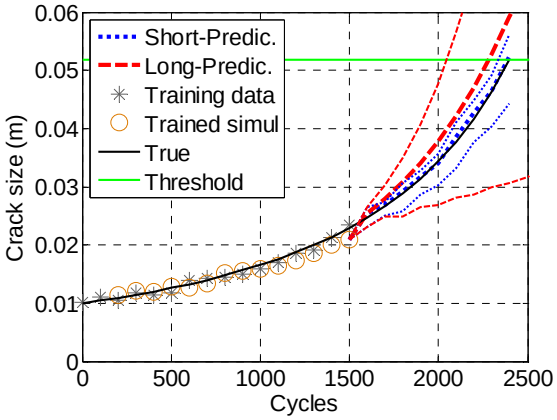


Figure 3. Example of NN-based prognostics with crack growth example.

III. Statistical Aspects in NN

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

A. Prediction Uncertainty

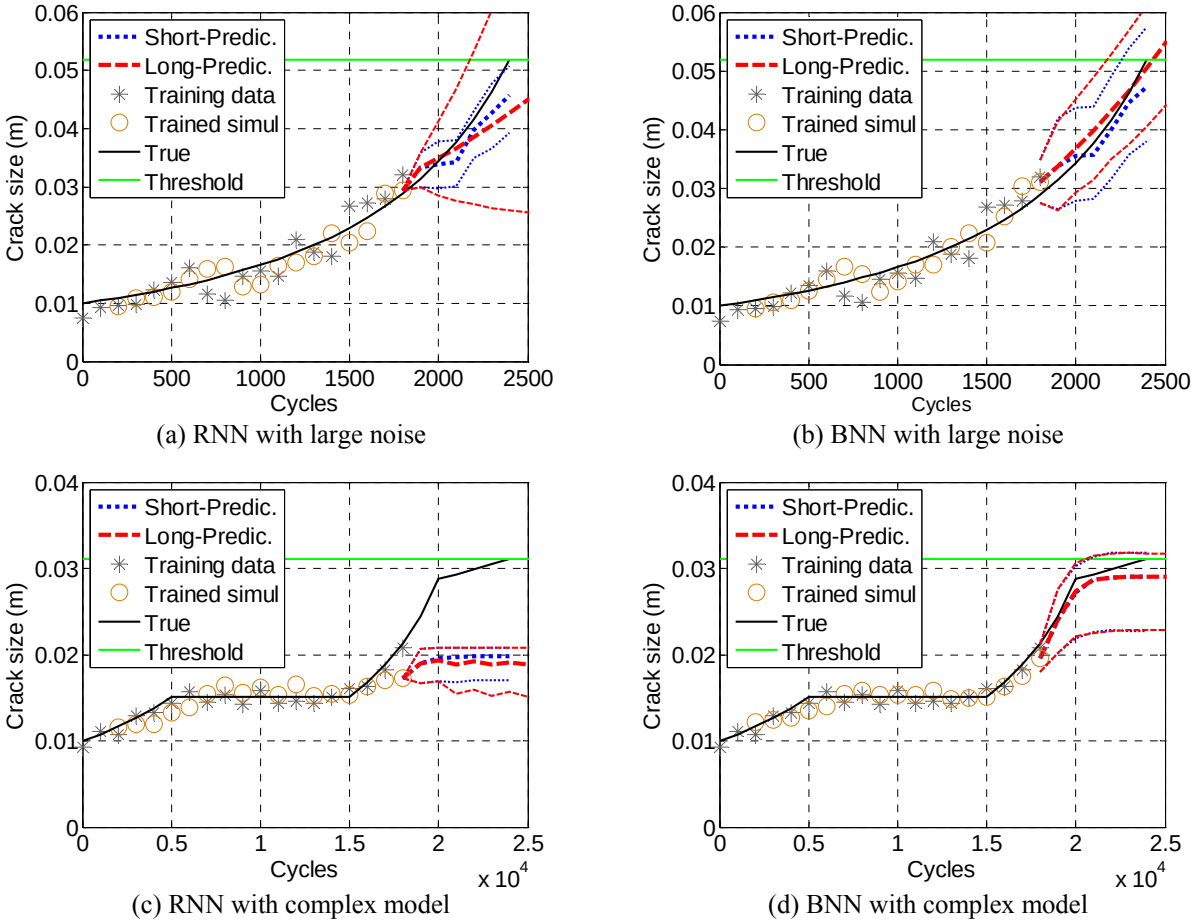
The first case is a common condition caused by noise in measurement data and parameter identification, and it is to identify the weight parameters as distribution based on Bayesian framework. Bayesian inference is a statistical method in which observations are used to estimate and update unknown parameters such as weight parameters in the form of a probability density function (PDF). Bayesian inference is based on the following Bayes' theorem:[22]

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

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.

Figure 4 shows the comparison between repeating NN (RNN) and Bayesian NN (BNN). Figure 4 (a) and (b) are the same condition as the previous example in Section Ⅱ.B but with a larger level of noise, ± 5mm. Figure 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. In both cases, 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 Liu et al.[16] did. There are two reasons why: (1) the results obtained by repeating NN more than 30 times do not much change with other attempts, which gives more reliable results compared to use NN just one time

with confidence bounds, and (2) since it grows hard to identify the distribution of weigh parameters as the number of parameters depending on network model increases, BNN is interrupted to construct network model flexibly.



(a) RNN with large noise

(b) BNN with large noise

(c) RNN with complex model

(d) BNN with complex model

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

B. Input Variable Uncertainty

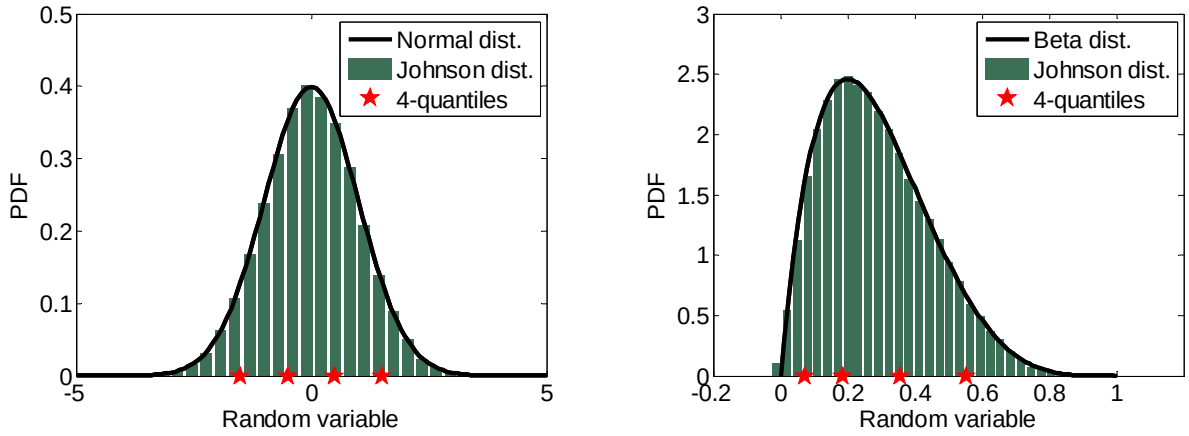
Input variables of NN are composed of damage data and usage conditions that are considered as deterministic values, and never considered as distributions. However, input variables can be distributed in such cases: many number of damage data are given at the same usage conditions from the same system, 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 0.0668, 0.3085, 0.6915 and 0.9332, is employed to prediction future damage distribution. Figure 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 represented 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 the previous one is again employed to demonstrate the case of random input variable. Distributed synthetic data are generated from the load magnitude $\Delta\sigma$ =78MPa, the perturbation of Paris model parameter m and small noise level: $m \sim U[3.8-0.027,\ 3.8+0.027]$, $noise \sim U[-1,+1]mm$, whose result is shown in Figure 6. Each cycle has 5000 samples as the measurement data, whose distribution at 0, 800, 1500, and 2200 cycles are shown in Figure 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 Figure 6(a) are used for input variables. Since there were two input variables and one output variable in the previous study, total number of input and output variables becomes eight and four, respectively.

Figure 7 shows damage prediction results at 1500 cycles. In Figure 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. Figure 7(b)

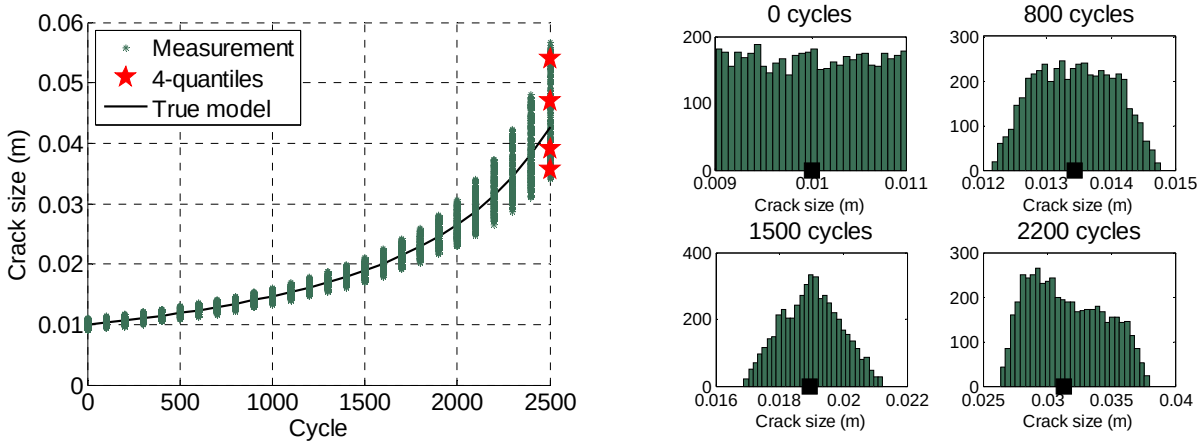
and (c) show comparison of damage distribution between predicted one and measured one at 1600 and 2400 cycles, and their errors are listed in Table 1. The maximum magnitude of error is 5.75% at 2400 cycles that is 900 cycles ahead prediction from 1500 cycles. These results show that NN using Johnson distribution is applicable for prediction of damage distribution.

Lastly, Figure 8 shows real measurement data from the bearing provided by the Center for Intelligent Maintenance Systems.[25] Vibration signal is monitored using accelerometer during one second with 20kHz sampling rate at specific intervals. The distributions in Figure 8(a) and Figure 8(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 gets wider and its mode shift 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 damage indicators like this bearing problem.



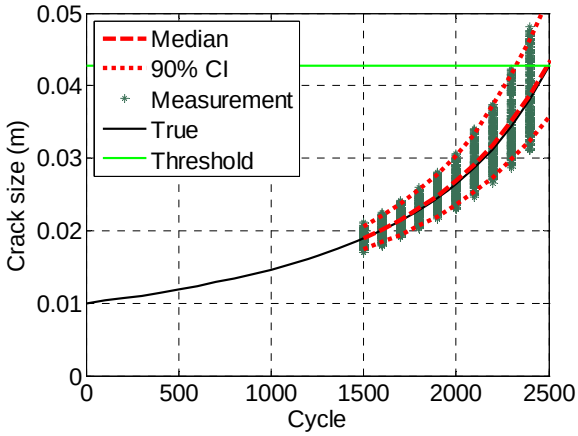
(a) standard normal distribution (b) beta distribution with α=2, β=5

Figure 5. Example of Johnson distribution.

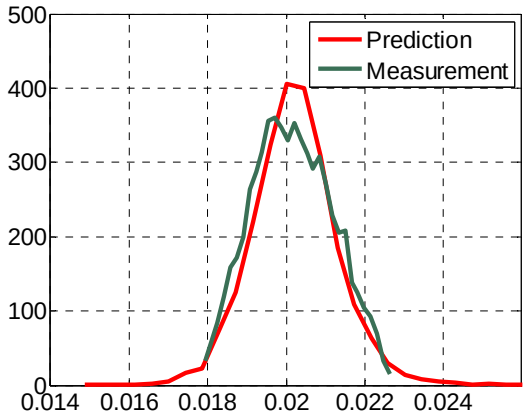


(a) measured data at every 100 cycles (b) distribution of measured data

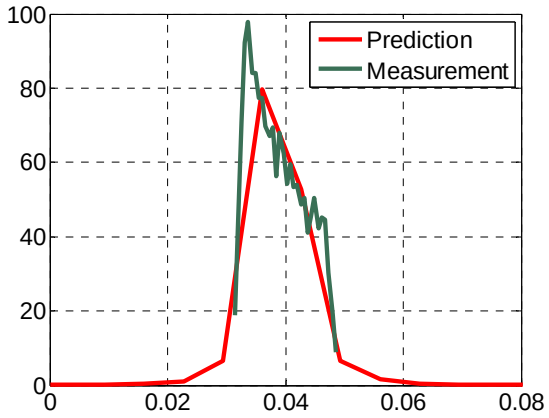
Figure 6. Distributed synthetic data.



(a) damage growth prediction



(b) damage distribution at 1600 cycles

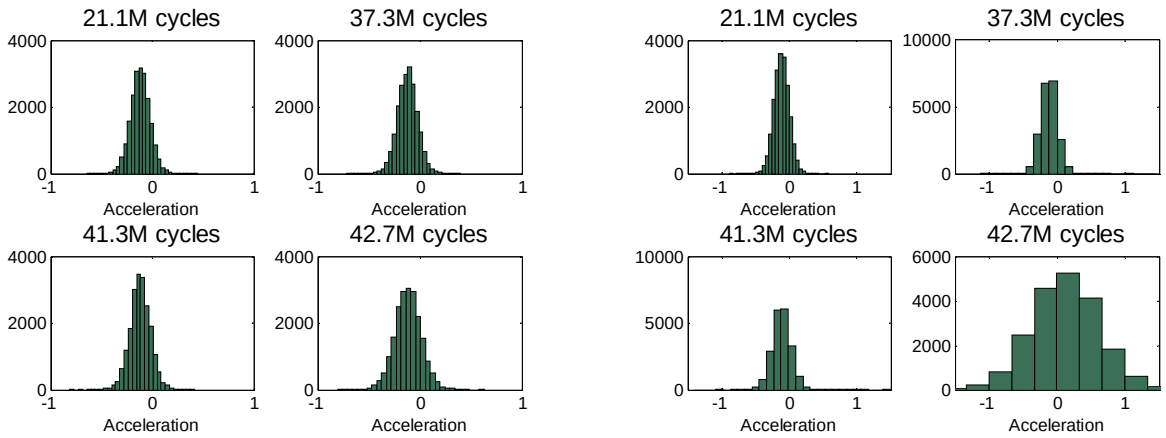


(c) damage distribution at 2400 cycles

Figure 7. Damage prediction results.

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

Cycles		1600	1800	2000	2200	2400
	measurement	0.0186	0.0210	0.0239	0.0276	0.0327
	prediction	0.0186	0.0210	0.0240	0.0279	0.0332
	error (%)	**0.0346**	**0.0158**	**0.5782**	**0.9487**	**1.8163**
	measurement	0.0196	0.0221	0.0253	0.0296	0.0354
	prediction	0.0197	0.0225	0.0261	0.0308	0.0374
	error (%)	**0.6279**	**1.5374**	**2.8931**	**4.1835**	**5.7526**
	measurement	0.0207	0.0237	0.0278	0.0333	0.0413
	prediction	0.0207	0.0237	0.0276	0.0329	0.0402
	error (%)	**0.1004**	**0.0141**	**0.5928**	**1.3502**	**2.6584**
	measurement	0.0218	0.0251	0.0296	0.0363	0.0463
	prediction	0.0218	0.0253	0.0299	0.0363	0.0453
	error (%)	**0.3191**	**0.8199**	**0.7796**	**0.0934**	**2.1976**



(a) acceleration: failure not occurred (b) acceleration: failure occurred

Figure 8. Real damage data.

IV. 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.

Acknowledgments

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] Schwabacher, M. A., "A survey of data-driven prognostics," *AIAA Infotech@Aerospace Conference*, Reston,VA, September 26-29, 2005.

[2] Luo, J., Pattipati, K. R., Qiao, L., and Chigusa, S., "Model-based Prognostic Techniques Applied to a Suspension System," *IEEE Transactions on System, Man and Cybernetics*, Vol. 38, No. 5, 2008, pp. 1156-1168.

[3] Yan, J. and Lee, J., "A Hybrid Method for On-line Performance Assessment and Life Prediction in Drilling Operations," *IEEE International Conference on Automation and Logistics*, Jinan, Shandong , China , August 18-21, 2007.

[4] Chakraborty, K., Mehrotra, K., Mohan, C. K., and Ranka, S., "Forecasting the Behavior of Multivariate Time Series Using Neural Networks," *Neural Networks*, Vol. 5. 1992, pp. 961-970.

[5] Yao, X., "Evolving Artificial Neural Networks," *Proceedings of the IEEE*, Vol. 87, No. 9, September 1999, pp. 1423-1447.

[6] Seeger, M., "Gaussian Processes for Machine Learning," *International Journal of Neural Systems*, Vol. 14, No. 2, 2004, pp. 69-106

[7] Mohanty, S., Teale, R., Chattopadhyay, A., Peralta, P., and Willhauck, C., "Mixed Gaussian Process and State-Space Approach for Fatigue Crack Growth Prediction," *International Workshop on Structural Heath Monitoring*, Vol. 2, 2007, pp. 1108-1115.

[8] Tipping, M. E., "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211-244.

[9] Tran, V. T. and Yang, B. S., "Data-Driven Approach to Machine Condition Prognosis Using Least Square Regression Tree," *Journal of Mechanical Science and Technology*, Vol. 23, 2009, pp. 1468-1475.

[10] Chryssoloiuris, G., Lee, M., and Ramsey, A., "Confidence Interval Prediction for Neural Network Models," *IEEE Transactions on Neural Networks*, Vol. 7, No. 1, January 1996, pp. 229-232.

[11] Veaux, R. D., Schumi, J., Schweinsberg, J., and Ungar, L. H., "Prediction Intervals for Neural Networks via Nonlinear Regression," *Technometrics*, Vol. 40, No. 4, November 1998, pp. 273-282.

[12] Yang, L., Kavli, T., Carlin, M., Clausen, S., and Groot, P. F. M., "An Evaluation of Confidence Bound Estimation Methods for Neural Networks," *European Symposium on Intelligent Techniques 2000*, Aachen, Germany, September 14-15, 2000.

[13] Leonard, J. A., Kramer, M. A., and Ungar, L. H., "A Neural Network Architecture that Computes its Own Reliability," *Computers in Chemical Engineering*, Vol. 16, No. 9, 1992, pp. 819-835.

[14] Freitas, de J. F. G., "Bayesian Methods for Neural Networks." *PhD Thesis*. University of Cambridge, UK, 2003.

[15] Neal, R. M., "Bayesian Learning for Neural Networks." *PhD Thesis*, University of Toronto, Ontario, Canada, 1995.

[16] Liu, J., Saxena, A., Goebel, K., Saha, B., and Wang, W., "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, October 10-16, 2010.

[17] Svozil, D., Kvasnička, V., Pospíchal, J., "Introduction to Multi-Layer Feed-Forward Neural Networks," *Chemometrics and Intelligent Laboratory Systems*, Vol. 39, 1997, pp. 43-62.

[18] Duch, W. and Jankowski, N., "Survey of Neural Transfer Functions," *Neural Computing Surveys*, Vol. 2, 1999, pp. 163-212.

[19] Firth, A. E., Lahav, O., Somerville, R. S., "Estimating Photometric Redshifts with Artificial Neural Networks," *Monthly Notices of the Royal Astronomical Society*, Vol. 339, 2003, 1195, DOI: 10.1046/j.1365-8711.2003.06271.

[20] Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation," *In Parallel Distributed Processing: Ex-plorations in the Microstructure of Cognition*, Vol. 1: foundations, MIT Press, 1986, pp. 318-362.

[21] Paris, P. C. and Erdogan, F., "A Critical Analysis of Crack Propagation Laws," *Transactions of the ASME, Journal of Basic Engineering, Series D*, Vol. 85, No. 3, 1963, pp. 528-534.

[22] Bayes, T., "An Essay Towards Solving a Problem in the Doctrine of Chances," *Philosophical Transactions of the Royal Society of London*, Vol. 53, 1763, pp. 370-418.

[23] Huang, X., Torgeir, M., and Cui, W., "An Engineering Model of Fatigue Crack Growth under Variable Amplitude Loading." *International Journal of Fatigue*, Vol. 30, No. 1, 2008, pp. 2-10.

[24] Johnson, N. L., "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, Vol. 36, 1949, pp. 149-176.

[25] Lee, J., Qiu, H., Yu, G., Lin, J., and Rexnord Technical Services., "Bearing Data Set," IMS, University of Cincinnati. NASA Ames Prognostics Data Repository, 2007, http://ti.arc.nasa.gov/project/prognostic-data-repository, NASA Ames, Moffett Field, CA.