Options for Prognostics Methods: A review of data-driven and physics-based prognostics

Dawn An[1], Nam Ho Kim[2], and Joo-Ho Choi[3]

[1,2]*University of Florida, Gainesville, FL, 32611, USA*
dawnan@ufl.edu
nkim@ufl.edu

[1,3]*Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea*
dawnan@ufl.edu
jhchoi@kau.ac.kr

ABSTRACT

Condition-based maintenance is a cost effective maintenance strategy, in which maintenance schedules are predicted based on the results provided from diagnostics and prognostics. Although there are several reviews on diagnostics methods and CBM, a relatively small number of reviews on prognostics are available. Moreover, most of them either provide a simple comparison of different prognostics methods or focus on algorithms rather than interpreting the algorithms in the context of prognostics. The goal of this paper is to provide a practical review of prognostics methods so that beginners in prognostics can select appropriate methods for their field of applications in terms of implementation and prognostics performance. To achieve this goal, this paper introduces not only various prognostics algorithms, but also their attributes and pros and cons using simple examples.

1. INTRODUCTION

Prognostics is to predict future behavior of damage/degradation and the remaining useful life (RUL) of in-service system, which facilitates condition-based maintenance whose schedule is planned according to predicted results based on diagnosis. There are a large number of publications on condition-based maintenance (Jardine, Lin, & Banjevic, 2006; Grall, Bérenguer, & Dieulle, 2002; Yam, Tse, Li, & Tu, 2001) and diagnostics methods (Martin, 1994; Samuel & Pines, 2005; Singh & Al Kazzaz, 2003; Juricic, Znidarsic, & Fussel, 1997; Sugumaran, Sabareesh, & Ramachandran, 2008; Yan & Gao, 2007; Samanta & Al-Balushi, 2003). On the other hand, a relatively small number of reviews on prognostics are available (Si, Wang, Hu, & Zhou, 2011; Srivastava & Das, 2009; Goebel, Saha, & Saxena, 2008; Saha, Goebel, & Christophersen, 2009; Xing, Williard, Tsui, & Pecht, 2011; Zhang & Lee, 2011). Most of them provide a simple comparison of different prognostics methods using a specific application, a summary of pros and cons of algorithms, or a collection of papers. In addition, most review papers focused on algorithms rather than interpreting the algorithms in the context of prognostics. Therefore, the goal of this paper is to provide a practical review of prognostic methods so that beginners in prognostics can select appropriate methods for their field of applications. To achieve this goal, this paper introduces not only various prognostics algorithms, but also their attributes and pros and cons, so that engineers can choose the best algorithm for their field of applications in terms of implementation and prognostics performance.

In general, prognostics methods can be categorized into data-driven, physics-based, and hybrid approaches. Data-driven approaches use information from previously collected data (training data) to identify the characteristic of the currently measured damage state and to predict the future trend. Physics-based approaches assume that a physical model describing the behavior of damage is available, and combine the physical model with measured data to identify model parameters and to predict the future behavior. Hybrid approaches combine the above-mentioned two methods to improve the prediction performance (Mohanty, Teale, Chattopadhyay, Peralta, & Willhauck, 2007; Sankavaram, Pattipati, Kodali, Pattipati, Azam, Kumar, & Pecht, 2009; Cheng & Pecht, 2009; Xu & Xu, 2011; Xing, Miao, Tsui, & Pecht, 2011). The last approach, however, is not mature yet, and will not be considered in this paper.

Data-driven and physics-based approaches have different properties that will contribute to the preference of each algorithm. Providing a standard to select the best algorithm is important for a practical review of prognostic methods, and considering practically available conditions is also important. In this paper, therefore, the following conditions will be considered: the number of data sets, the level of noise and bias in obtained data, availability of loading conditions and physical models, and complexity of damage growth behavior. Typical prognostics algorithms are tested under such conditions, and the algorithms' attributes and pros and cons are provided based on the results.

The paper is organized as follows: in Sections 2 and 3, reviews on the data-driven and physics-based approaches are presented, respectively. In Section 4, several case studies are presented to analyz different prognostics methods, followed by conclusions in Section 5.

2. REVIEWS ON DATA-DRIVEN APPROACHES

Data-driven approaches use information from collected data to identify the characteristics of damage state and predict the future state without using any particular physical model. Instead, mathematical models with weight parameters are employed. The weight parameters are determined based on the training data that are obtained under the various usage conditions. Since the data-driven approaches depend on the trend of data, which often has a distinct characteristic near the end of life, it is powerful in predicting near-future behaviors, especially toward the end of life.

The data-driven approaches are divided into two categories: (1) the artificial intelligence approaches that include neural network (NN) (Chakraborty, Mehrotra, Mohan, & Ranka, 1992; Krogh, 2008; Yao, 1999) and fuzzy logic (Zio & Maio, 2010; Gouriveau, Dragomir, & Zerhouni, 2008), and (2) the statistical approaches that include gamma process (Pandey & Noortwijk, 2004; Dickson & Waters, 1993), hidden Markov model (Rabiner, 1989), and regression-based model such as Gaussian process (GP) regression (Mackay, 1997; Seeger, 2004), relevance vector machine (Tipping, 2001), and least square regression (Tran & Yang, 2009; Bretscher, 1995), etc. Among these algorithms, NN and GP are commonly used for prognostics and will be discussed in the following sections.

2.1. Neural Network (NN)

NN is a representative data-driven method, in which a network model learns a way to produce a desired output, such as the level of degradation or lifespan, reacting to given inputs, such as time and usage conditions. This method mimics the human nervous system, which responds and adapts to a stimulus. Once the network model learns enough the relationship between inputs and output, it can be used for the purpose of diagnosis and prognosis. A typical architecture, feed-forward neural network (FFNN) (Svozil, Kvasnička, & Pospíchal, 1997), is illustrated in Figure 1. In the figure, circles represent nodes (also called a 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. The number of nodes in the hidden layer can be adjusted to properly express the mechanism between input and output. Once the network model learned enough, 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 (Duch & Jankowski, 1999). Weight parameters include weights for the interconnected nodes and biases that are added to inputs of transfer functions (Liu, Saxena, Goebel, Saha, & Wang, 2010; Firth, Lahav, & Somerville, 2003). The process of finding the weight parameters is called training or learning, and to accomplish that, 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 (Rumelhart, Hinton, & Williams, 1986) 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 addition to FFNN, there exists recurrent (Bodén, 2002), fuzzy (Liu & Li, 2004), wavelet (He, Tan, & Sun, 2004), associative-memory (Bicciato, Pandin, Didonè, & Bello, 2001), modular (Happel & Murre, 1994), and hybrid (Psichogios & Ungar, 1992; Rovithakis, Maniadakis, & Zervakis, 2004) neural network.

In the following, three important issues are discussed for NN-based prognostics.

2.1.1. Issue 1: Network Model Definition (the Number of Node and Layer)

The first issue is the definition of the network model that includes selecting the number of hidden nodes, hidden layers and input nodes. Trial-and-error methods are often used to determine a suitable network model. Lawrence, Giles, and Tsoi (1998) and Doukim, Dargham, and Chekima

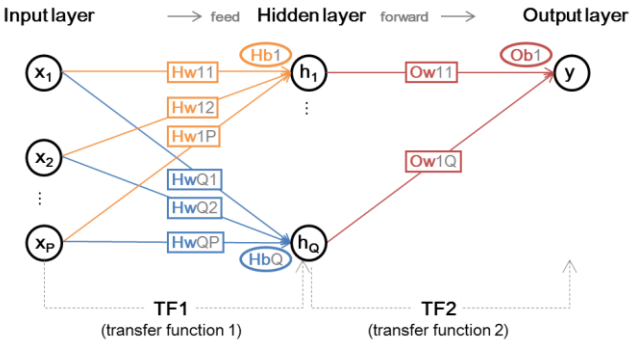


Figure 1. Illustration of typical network model: FFNN

(2010) investigated the usage of mean square error in order to find the optimal number of hidden nodes. Gómez, Franco, and Jérez (2009) used generalization complexity to determine the number of nodes and showed that the results were quite close to the optimum. Although one or two hidden layers are generally used, there is no fixed rule. Ostafe (2005) presented a method using pattern recognition to determine the number of hidden layers. The problem of determining the number of input nodes occurs when input variables affecting the output are not clear; various inputs possibly affecting the response can be considered or only data trace is applicable. Chakraborty et al. (1992) compared the prediction results of flour prices using variable network models. In such a case, actual past values are used for inputs, but how many past values should be used is unclear. In fact, the accuracy of prediction results is not proportional to the number of input nodes. Chang and Hsieh (2011) also researched to select the optimal input layer neurons using particle swarm optimization. Therefore, defining a proper neural network model can be difficult for new users without having much experience.

2.1.2. Issue 2: Optimal Parameters (Finding Weights and Biases)

Once a network model is defined, the next issue is to find weight parameters related with the model. In NN, no matter how complex the relationship between input and output layer is, it is possible to express the relationship by augmenting the number of hidden layers and hidden nodes. However, when the BP algorithm is used, the following problems exist: (1) the global optimum of many weight parameters is extremely difficult to find, and (2) the convergence rate is very low and depends on the initial estimates. For these reasons, there have been many efforts to improve the drawbacks of the BP algorithm. Salomon and Hemmen (1996) presented a dynamic self-adaptation algorithm to accelerate the steepest descent method, and Chen, Lin, Tseng, and Lin (2006) applied the simulated annealing algorithm to search for the best BP parameters such as learning rate, momentum and the number of hidden nodes. Also, Subudhi, Jena, and Gupta (2008) proposed a technique combining the genetic algorithm and differential evolution with BP, and Nawi, Ransing, and Ransing (2007) presented a technique combining the conjugate gradient optimization algorithm with the BP algorithm. There are many ensemble techniques to improve the performance of a single algorithm (Navone, Granitto, Verdes, & Ceccatto, 2001; Jacobs, 1995; Drucker, Cortes, Jackel, LeCun, & Vapnik, 1994; Krogh & Vedelsby, 1995; Perrone & Cooper, 1993; Naftaly, Intrator, & Horn, 1997), and the other efforts are found in the Refs.(Jardine et al., 2006; Salomon and Hemmen, 1996; Nawi et al., 2007) However, finding good weight parameters is still challenging, and the performance of NN algorithm deteriorates with non-optimal weight parameters.

2.1.3. Issue 3: Uncertainty From Data and Optimization Process

Last but not least, uncertainty in noise and bias in training data is an important issue in NN, as most measured data include them. The bias here is different from the bias as weight parameters; here the bias is the error caused by sensors, such as calibration error. In terms of noise, it is common to provide confidence bounds based on nonlinear regression and/or the error between NN outputs and training data (Chryssoloiuris, Lee, & Ramsey, 1996; Veaux, Schumi, Schweinsberg, & Ungar, 1998; Yang, Kavli, Carlin, Clausen, & Groot, 2000; Leonard, Kramer, & Ungar, 1992). Bootstrapping (Efron & Tibshirani, 1994) can also be applied, which can be easily implemented by running Matlab NN toolbox several times because Matlab uses different subsets of given training data for obtaining weight parameters. Furthermore, running NN several times can relieve the concerns about initial weight parameters for optimization by setting different initial parameters automatically. For example, Liu et al. (2010) used the repeating method with 50 attempts to predict battery's RUL with uncertainty. Actually, a basic method to handle uncertainty in NN is the probabilistic neural network (PNN) (Specht, 1990) using Parzen estimator (Parzen, 1962). However, most papers employ PNN for classification or risk diagnosis (Petalas, Spyridonos, Glotsos, Cavouras, Ravazoula, & Nikiforidis, 2003; Giurgiutiu, 2002; Mao, Tan, & Ser, 2000), and prognostics ones are rarely found except for the study by Khawaja, Vachtsevanos, and Wu (2005). They introduced a way to obtain not only confidence bounds but also confidence distribution based on PNN to predict a crack on a planetary gear plate. Unfortunately, bias in measured data cannot be handled with data-driven approaches because the approaches are based on measured data, and there are no parameters related with bias.

2.2. Gaussian Process (GP) Regression

GP is a commonly used method among regression-based data-driven approaches for prognostics, whose conceptual property is illustrated in Figure 2. An outstanding property of GP is that simulated outputs are smoothly constructed making exactly the same value as every measured point (data) as blue-dashed curve in Figure 2. The reason for this can be explained with following GP model that is composed of a global model ($\mathbf{f}\boldsymbol{\beta}$) and departures ($\mathbf{r}\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta}\right)$):

$$y^{*}=\mathbf{f}\boldsymbol{\beta}+\mathbf{r}\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta}\right) \quad (1)$$

where y^{*} is a simulated GP output at an arbitrary input vector, $\mathbf{x}^{*}$ whose size is $1\times p$ (p is the number of input variables), $\mathbf{f}$ is the known function of $\mathbf{x}^{*}$ and determines polynomial order of global model, $\boldsymbol{\beta}$ is the regression coefficient and obtained by $\left(\mathbf{F}^{T}\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^{T}\mathbf{R}^{-1}\mathbf{Y}$, and $\mathbf{r}$ is a

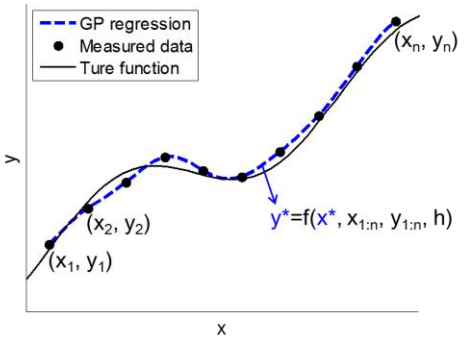


Figure 2. Illustration of GP regression

$1\times n$ (n is the number of measured data) vector, which represents a correlation between $\mathbf{x}^*$ and a $n\times p$ input matrix of all measured points, $\mathbf{X}$. The rest capital letters $\mathbf{R}$, $\mathbf{Y}$, and $\mathbf{F}$ that are the same property as their small letters' have the same size in terms of row vector as their small letters, but the size of column vector is the same as number of measured data, n.

If correlation terms ($\mathbf{r}$ and $\mathbf{R}$) and departures term are ignored in Eq. (1), it becomes that $y^*=\mathbf{f}\boldsymbol{\beta}$ with $\boldsymbol{\beta}=\left(\mathbf{F}^T\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{Y}$, which is the ordinary least square regression (Tran & Yang, 2009). Consequently, GP is distinguishable from least squarie in terms of that simulated outputs penetrate every measured point with two assumptions that (1) GP model is a combination of global model and local departures, and (2) the error between every points is correlated. Assumption (1) is already reflected in the GP model, and assumption (2) is determined by a type of covariance function and scale parameters (or hyperparameters) related with them. Once scale parameters are obtained based on the measured points (training data) using optimization algorithm, GP model can be used to predict future behavior of damage. Lastly, the name, Gaussian process comes from the assumption that each point is normally distributed. Based on this assumption with multivariate normal distribution, more mathematical/probabilistic derivation can be done in terms of probabilistic parameters of Gaussian distribution; the mean expressed in Eq. (1) and the variance found in the Refs. (Gelman, Carlin, Stern, & Rubin, 2004; Santner, Williams, & Notz, 2003).

2.2.1. Issue 1: Model Problem (Covariance Function)

In common with NN, the performance of GP largely depends on models. In this case, covariance function and polynomial order of global model are related with GP model, but order of global model is less important as it is often handled with constant value. There are various types of covariance functions such as radial basis (or squared exponential), rational quadratic, neural network, Matern, periodic, constant and linear (Rasmussen & Williams, 2006; Williams, 1997). Mohanty, Das, Chattopadhyay, and Peralta (2009) compared the prediction results of crack length under variable loading from radial basis function (RBF) covariance function and neural network based (NN-based) covariance function, and showed that RBF-based GP model outperformed NN-based one in their application. As part of an effort to resolve the selection of covariance function, research on nonstationary covariance functions that is a model to adapt to variable smoothness and can be constructed by adding or multiplying simple covariance functions has been conducted. Paciorek and Schervish (2004) introduced a class of nonstationary covariance functions so that the model adapts to variable smoothness, and compared the results from stationary GP. From their research, it was concluded that the results from nonstationary GP are better than stationary GP, but pointed out that simplicity loss of the algorithm occurs as the nonstationary GP requires more parameters than a stationary GP. Brahim-Belhouari and Bermak (2004) used nonstationary GP to predict respiration signal, and compared with a GP model with an exponential covariance function, and Liu, Pang, Zhou, and Peng (2012) used the combination of three covariance functions to predict lithium-ion battery degradation (state of health, SOH).

2.2.2. Issue 2: Optimization Problem (Finding Scale Parameters)

Determining the scale parameters related with covariance function is also important, since they determine the smoothness of regression model. In general, the parameters are obtained based on equivalent likelihood function (Sacks, Welch, Mitchell, & Wynn, 1989) via optimization algorithm. It, however, is a difficult task to search their optimum values, and even if they are found they are not always the best selection (An & Choi, 2012). Since the scale parameters are seriously affected by input and output values, input and output values are applied as normalized values in most cases. Mohanty et al. (2009), however, studied the performance to prediction crack growth according to three different types of scaling. Neal (1998) considered the scale parameters as distributions rather than deterministic values, and An and Choi (2012) showed that the GP models with scale parameters identified as distributions outperform the one using optimal deterministic parameter.

2.2.3. Issue 3: Data Problems (Num. of Data and Uncertainty)

Even though large number of training data is usually profitable for increasing accuracy of prediction results, it's not always acceptable for GP because it also increases computational costs to calculate the inversion of the covariance matrix (Eq. (1)) as well as generates singularity. It is suggested that inversion by direct method may become

prohibitive when the number of the data points is greater than 1000 (MacKay, 1997). As a solution to relieve such problem, the methods to select a subset of data points are usually employed (Lawrence, Seeger, & Herbrich, 2003; Smola & Bartlett, 2001; Foster, Waagen, Aijaz, Hurley, Luis, Rinsky, Satyavolu, Way, Gazis, & Srivastava, 2009). While Melkumyan and Ramos (2009) suggested new covariance function based on cosine function that inherently provides a sparse covariance matrix. In terms of uncertainty, it's determined with Gaussian noise as mentioned before. Mohanty et al. (2007) and Liu et al. (2012), respectively, showed the predictive confidence interval of crack length and SOH of lithium-ion battery using GP.

3. REVIEWS ON PHYSICS-BASED APPROACHES

Physics-based approaches combine the physical damage model with measured data to predict future behavior of damage and the RUL, which is illustrated in Figure 3. The behavior of the physical model depends on the model parameters that are estimated and updated in company with damage state based on the measured data. Finally, the RUL is predicted by progressing the damage state until it reaches a threshold as the dashed curves in the Figure 3. Similar issues are addressed in this approach as the previous ones.

3.1. Issue 1: Model Problem (Physical Model Accuracy)

Since the physics-based approaches employ a physics model describing the behavior of damage, it has advantages in predicting long term behaviors of damage. However, model validation should be carried out since such models contain many assumptions and approximations. There have been much literature on model validation using statistical methods such as hypothesis test and Bayesian method to calibrate and improve the model by comparing with observation (Rebba, Huang, Liu, & Mahadevan, 2006; Rebba, Mahadevan, & Huang, 2006; Kleijnen, 1995; Sargent, 2009). In general, the number of model parameters increases as model complexity increases, which makes it difficult to identify the model. Recently, Coppe, Pais, Haftka, and Kim (2012) showed that the issue of model

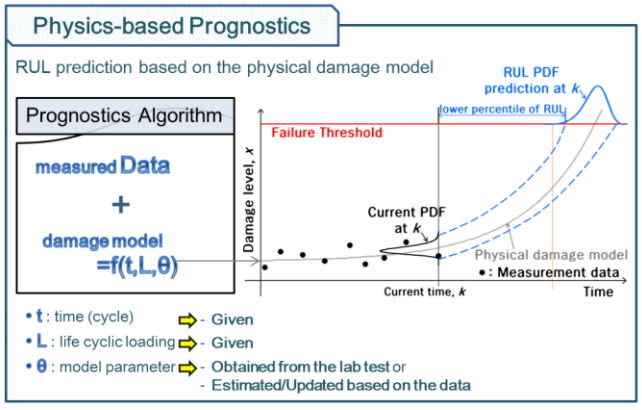


Figure 3. Illustration of physics-based prognostics

accuracy can be relieved by identifying equivalent model parameters of the simpler model. They showed that a simple Paris model with an assumed stress-intensity factor can be used for predicting crack growth of complex geometries by adjusting the model parameters to compensate for the error in the simple model. Although this is limited to the case of a similar model form, cumbersome efforts to validate the model accuracy can be eased off.

3.2. Issue 2: Model Parameter (Physical Model Parameters, Noise and Bias)

3.2.1. Introduction to Physics-Based Algorithms

Once a physical model is available, model parameter identification becomes the most important issue, which is performed with an estimation algorithm based on measured data with a usage condition. In fact, estimation algorithms become criteria to classify physics-based approaches. There are several algorithms such as Kalman filter (Kalman, 1960), extended Kalman filter (Julier & Uhlmann, 2004), particle filter (PF) (Doucet, De Freitas, & Gordon, 2001), and Bayesian method (BM) (Kramer & Sorenson, 1998). These algorithms are based on the Bayesian inference (Bayes, 1763), in which observations are used to estimate and update unknown parameters in the form of a probability density function (PDF). The updated PDF is called the posterior distribution, which is obtained by multiplying the prior distribution that is prior knowledge or information of the unknown parameters and the likelihood function that is the PDF value of measured data conditional on the given parameters.

There are several researches dealing with parameter estimation in terms of prognostics. DeCastro, Tang, Loparo, Goebel, and Vachtsevanos (2009) used extended Kalman filter to estimate model parameters and predict RUL for crack growth on a planetary carrier plate. Orchard and Vachtsevanos (2007) estimated the crack closure effect using PF for RUL prediction of a planetary carrier plate based on vibration-based feature. Daigle and Goebel (2011) used PF to estimate wear coefficients by considering multiple damage mechanisms in centrifugal pumps. An, Choi, Schmitz, and Kim (2011) estimated wear coefficients to predict the joint wear volume of slider-crank mechanism based on BM. Among the aforementioned algorithms, PF is the most commonly used for prognostics. In the following, PF and BM are discussed.

- ***Particle filter (PF)***

PF is the most commonly used algorithm in the prognostics field, in which the posterior distribution of model parameters is expressed as a number of particles (or samples) and their weights as shown in Figure 4. There are three steps in PF process: (1) prediction step - posterior distributions of the model parameters ($\boldsymbol{\theta}$) at the previous ($k-1$th) step are

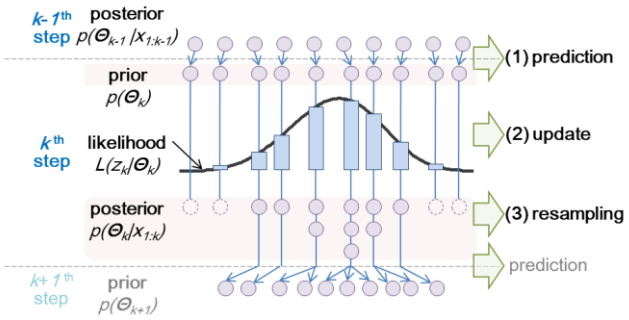


Figure 4. Illustration of PF process

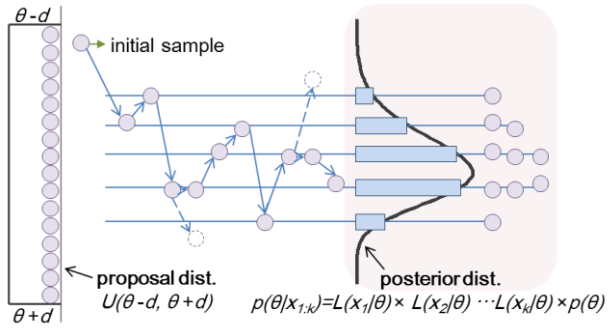


Figure 5. Illustration of BM process with MCMC.

used for the prior at the current (k th) step, and the damage state at the current time is transmitted from the previous one based on a damage model (physical model), (2) update step - model parameters and damage state are updated based on the likelihood function combined with measurement data (x), and (3) resampling step - particles of the prior distribution are resampled based on their weights expressed as vertical-rectangular in Figure 4 by duplicating or eliminating the samples with high or low weight, respectively. The resampled result corresponds to the posterior distribution at the current step and is also used as the prior distribution at the next ($k+1$th) step. That means the Bayesian update is processed in a sequential way with particles in PF.

Since prediction results from PF depends on initial distributions of parameters (posterior distribution at $k=0$), correct information or proper assumption for initial distributions is one of the most important issues to find model parameters. Another important point is accumulated sampling error that occurs during the update process. In other words, the particle depletion problem can occur since those particles with a very small weight are eliminated, while those particles with a high weight are duplicated. There have been researches to recover the particle diversity during update process. A common practice is to add random sample ξ_{k-1} from arbitrary distribution during the prediction step; that is, $\theta_k = \theta_{k-1} + \xi_{k-1}$ so that identical particles are not generated (Higuchi, 1997; Kitagawa, 1987; Wang, Liao, & Xing, 2009). This method, however, can change probabilistic characteristic of parameters as well as increase the variance of parameters. Gilks and Berzuini (2001) proposed a resample-move algorithm based on PF and Markov Chain Monte Carlo (MCMC) method (Andrieu, Freitas, Doucet, & Jordan, 2003), Kim and Park (2011) introduced the maximum entropy particle filter and demonstrated the effectiveness of the proposed technique by applying it to highly nonlinear dynamical systems.

- ***Bayesian method (BM)***

The Bayesian update is processed with an overall way in BM; the posterior distribution is obtained as an equation by multiplying all the likelihood function given by k number of data. Once the posterior distribution is available, a sampling method can be used to draw samples from the posterior distribution. Therefore, the estimation performance in BM depends on sampling methods, MCMC method which has been recognized as a computationally effective means is usually employed. MCMC is based on a Markov chain model of random walk as shown in Figure 5. It starts from generating an arbitrary initial sample (old sample) and a new sample from an arbitrary proposal distribution with the centered at the old sample. The two samples are compared with each other based on an acceptance criterion, from which either one is selected. In Figure 5, two circles with dashed line means new samples not selected, and in this case, the old one is selected. This process is repeated as many as the number of particles in PF.

Even though there is no accumulated sampling error in BM, there still exists some error caused by sampling method, random walk. Initial sample, proposal distribution for new sample, and acceptance ratio to the old sample have an effect on the sampling results; with improper setting, it could be not converged or show stationary chain that old sample is selected continually. There are some researches to reduce those effects by utilizing marginal density function for proposal distribution (Rubin, 1998; An & Choi, 2013). Gelfand and Sahu (1994) presented two distinct adaptive strategies to accelerate the convergence of a MCMC algorithm. More literatures are found in the Ref. (Andrieu et al. 2003).

3.2.2. Correlation Issue between Model Parameters

One of the most challenging parts in model parameter identification is correlation between model parameters. Without properly identifying correlation, the predicted RUL can be significantly different from reality. An, Choi, and Kim (2012) studied the correlation in a crack growth problem, in which correlation between the parameters was well identified, but eache parameter was not accurately identified under a large level of noise because of the correlations: correlation between the two Paris parameters and correlation between bias and the initial crack size. The

prediction results of damage growth and RUL, however, were reliable since many combinations of the correlated parameters can yield the same prediction results.

3.3. Issue 3: Uncertainty From Data (Noise and Bias)

Since damage cannot be directly measured in many cases, a damage quantification process is required from sensor measurement data, which is called structural health monitoring (SHM). SHM data could include a large level of noise and bias due to sensor equipment, and there are several researches dealt with the analysis of noise and bias in SHM data. Gu, Azarian, and Pecht (2008) presented a prognostics approach which detects the performance degradation of multilayer ceramic capacitors under temperature-humidity-bias conditions. Coppe, Haftka, Kim, and Yuan (2009) showed that the uncertainty in structure-specific damage growth parameters can be progressively reduced in spite of noise and bias in sensor measurements. Guan, Liu, Saxena, Celaya, and Goebel (2009) considered various uncertainties from measurements, modeling, and parameter estimations to describe the stochastic process of fatigue damage accumulation based on a maximum entropy-based general framework. It is concluded that convergence with large noise becomes slow, and positive and negative bias, respectively, effect on early and late prediction.

4. Case Study to Select Pertinent Method

Prognostics algorithms including NN, GP, PF and BM are analyzed and compared, so that engineers can choose the best algorithm for their field of applications. Since there are many variations each algorithm, the most common and basic ones are employed.

4.1. Problem Definition

4.1.1. Given Information for Case Study

Paris model (Paris & Erdogan, 1963) and Huang's model (Huang, Torgeir, & Cui, 2008) are, respectively, employed for a simple behavior of damage growth and complex behavior of damage growth, which are shown in Figure 6. In each model, there are ten sets of data under different loading conditions. Different level of noise and bias are artificially added to the data in Figure 6. Bias is considered as -2 mm, and noise is uniformly distributed between $-u$ mm and $+u$ mm. Three different levels of u are considered: 0, 1, and 5 mm. Ten data sets are numbering, one data set (usually #8) will be used for the set to be predicted and the other sets will be used for training data.

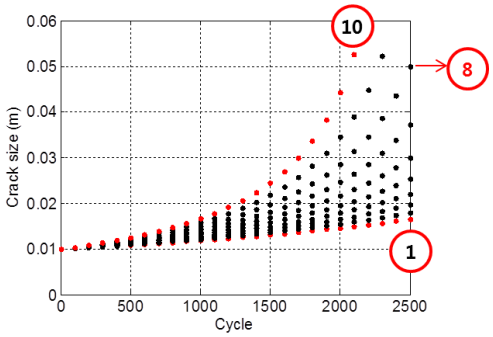
4.1.2. Definition of Algorithm Conditions

For the case of NN, the network model is constructed based on FFNN with three input nodes, one hidden layer with two nodes. Then, the number of total weight parameters become 11 including eight weights ($3\times2+2\times1$) and three biases (2+1). Since there is one hidden layer, two transfer functions are required, and as a common way, the tangent sigmoid and pure linear functions are employed. For GP model, linear or second order polynomial function is employed for the global model, and one parameter radial basis covariance function is employed as follow:

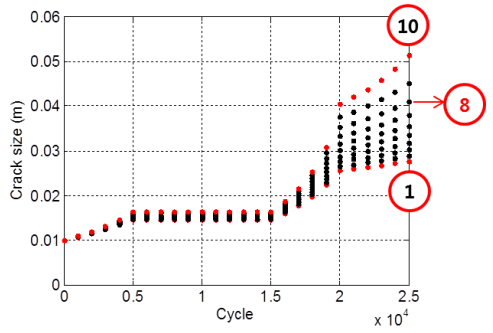
$$R\left(\mathbf{x}^i,\mathbf{x}^j\right)=\exp\left[-\left(d/h\right)^2\right],\ d=\left|\mathbf{x}^i-\mathbf{x}^j\right|,\quad i,j=1,\ldots,n \quad (2)$$

where h is a scale parameter to be identified, $\mathbf{x}^i,\mathbf{x}^j$ are vector of input variables, n is total number of training data; in this case, $n=k-3$, and k is the current time index.

For the input variable in NN and GP, the previous three damage data (x_{k-3}, x_{k-2}, x_{k-1}) are used, and the current damage data (x_k) becomes the output. Also, loading conditions can be utilized by adding to the input nodes. In this case, loading condition at the current cycle is added to the forth input node. If $k=16$, 13 sets of input and output data are available, which are used to obtain weight parameters, and then future damages ($x^p_{k+1}, x^p_{k+2}, x^p_{k+3},\ldots$) are predicted based on the obtained parameters and the previous damage data. 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, x_{k+1}, x_{k+2} are inputs to predict x^p_{k+3}. On



a) simple model



b) complex model

Figure 6. Problem definition for case study

the other hand, long term prediction is multi-step ahead prediction since it utilizes predicted results as inputs, e.g., x_k, x_{k+1}^p, x_{k+2}^p are inputs to predict x_{k+3}^p. In terms of prediction uncertainty, while noise in training data can be handled during GP, bootstrapping is employed by repeating NN 30 times to handle uncertainty from data noise and optimization process in NN.

Models for physics-based approaches are the same as Paris model and Huang's model employed to generate data sets. There are, respectively, two and six model parameters in Paris model and Huang's model, and the data sets in Figure 6 were generated with following true model parameters:

Paris model ture parameters: m=3.8, C = 1.5e-10

Huang's model true parameters: $m = 3.1$, C = 5.5e-11, $m = 3.1$, C = 5.5e-11, $\Delta K_{th} = 5.2$, $\beta = 0.2$, $n = 2.8$, $\sigma_y = 580$

4.2. Case Study Results

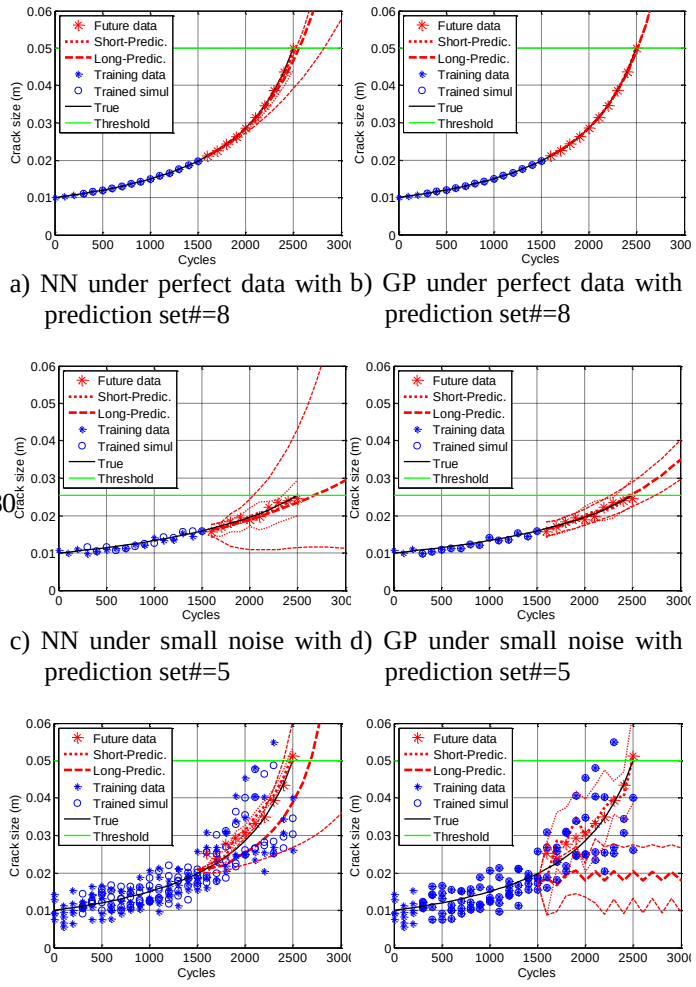
4.2.1. Data-Driven Results

At first, the performance of training data is compared. Figure 7 shows the comparison between NN and GP with different levels of noise. In the figure, the red and blue star markers are, respectively, future damage data and training data up to the current cycle (1500 cycles), and circle markers are simulation results from each algorithm. Thick dotted curve and thick dashed curve are, respectively, medians of short term prediction and long term prediction, and their thin curves represent 90% confidence intervals.

GP show exact result under perfect data (no noise) and outperform NN under small noise in terms of both short term and long term prediction. Long term prediction using GP is not available under large noise and many sets of training data. In this case, NN outperform GP in both short term and long term prediction, and long term prediction results get better as the number of cycles increases. The reason why GP is better than NN for small noise and small number of data is because of correlation property. Large data affect adversely the covariance matrix manipulation. Also, GP model penetrates every training data points assuming each data are correlated, which means that the GP behaves poorly when the data include large noise because the noisy relation is reflected to the predictions. While the reason for better results from NN under large noise and many data is that increasing data have no effect on network model but gives more information. Also, combination of transfer function is much less restricted to the level of noise.

4.2.2. Physics-Based Results

As mentioned before, bias in obtained data cannot be handled with data-driven approaches. In contrast, physics-based one can do it by adding bias to unknown parameter,



a) NN under perfect data with prediction set#=8

b) GP under perfect data with prediction set#=8

c) NN under small noise with prediction set#=5

d) GP under small noise with prediction set#=5

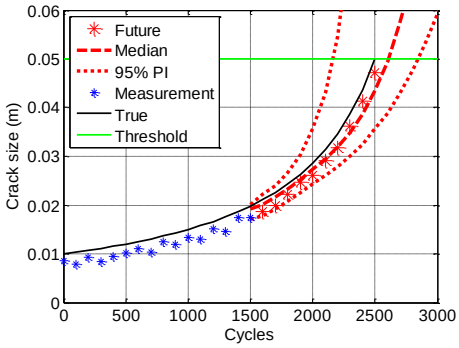
e) NN under large noise with prediction set#=8 and training set#=6,7,9,10

f) GP under large noise with prediction set#=8 and training set#=6,7,9,10

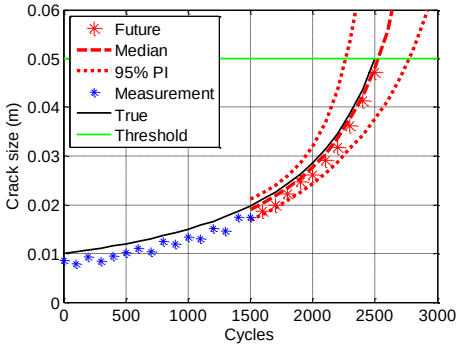
Figure 7. Results from data-driven approaches with the case of simple damage growth

whose results are shown in Figure 8. In the figure, measurement data up to the current cycle (blue star makers) are biased; measured crack size is consistently less than the true one. The medians of prediction results (dashed curves) at the current cycle are close to the true one, which means bias is well identified and compensated. Since physical model describes behavior of damage data accurately, the difference between model and data can be obtained as a constant at any cycle, which satisfies bias in measurement data is a constant. Further study for noise and bias in physic-based approaches is found in the Ref. (An et al., 2012).

The difference between PF and BM is negligible in view of the prediction results because the two methods have the same foundation with the same physical model. The only differences are the way of updating distributions and generating samples. BM is faster than PF because the



a) PF



b) BM

Figure 8. Results from physics-based approaches with the case of simple damage growth under small noise and negative bias

posterior distribution is given as a single equation and there is no accumulated sampling error. Even if PF has an accumulated sampling error during the updating process, it predicts well because the updating process occurs along with damage propagation. However, BM is not practical for them because of tremendous computational costs (Gilks & Berzuini, 2001; An, Choi, & Kim, 2012; Storvik, 2002). This is a key difference between the two methods, and more detailed comparison between PF and OBM can be found in the reference by An et al. (2012).

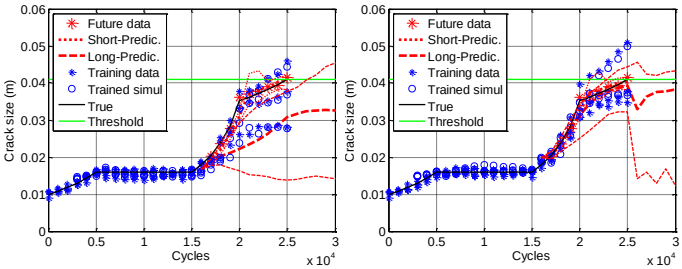
4.2.3. Case Study for Comparison between NN and PF

For data-driven approaches, NN is considered to predict damage growth in a complex model, because it is difficult to use a proper correlation function for GP to predict future due to retardation portion in obtained data. To find out how many data sets are required to obtain proper prediction results, different numbers of training sets are randomly selected. The results from this case study are presented in Figure 9 (a) and (b). Based on the training data sets, NN well predicts future damage as shown in Figure 9 (a). It, however, is limited to short term prediction, and proper results for long term prediction could not be found with different attempts. If there are available loading conditions, medians of short term and long term prediction become similar to each other with at least three sets of training data as shown in Figure 9 (b).

If there are available physical model as well as loading conditions, it might be clear that using physics-based approaches for the case of complex model outperforms data-driven ones. Since BM has a difficulty for the complex model due to extremely expensive computational costs, PF is only considered, and the results are shown in Figure 9 (c). The median of prediction is still not accurate, but uncertainty covers that. Also, since this physical model largely depends on the initial damage, if the initial distribution of damage is also updated, median will close to the true one. For example, Figure 9 (d) shows the results with true value of initial damage.

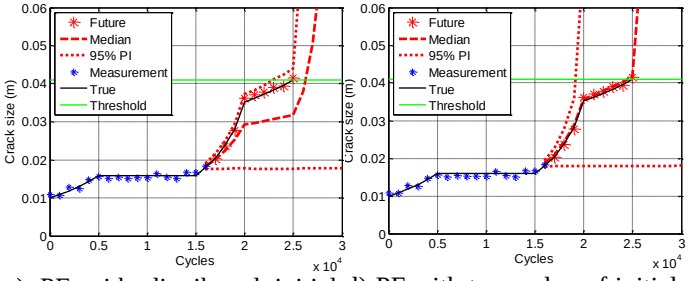
4.3. Results summary

In terms of algorithms, results from case studies can be summarized as follow: GP works well when the correlation function can be well defined such as the case of small noise data and simple models. It is easy to implement and fast to calculate. NN is proper to apply for the case of large noise and complex models with many training data sets. Even so, NN can be applied for small noise and simple models, which has a wide range of applications. It is, however, challenging to obtain many sets of data in realistic applications. PF and BM are less affected by the level of noise and model complexity, but they can be employed only if a physical model and loading conditions are given. The



a) NN with training set#=2, 7, 9

b) NN with training set#=6, 7, 10 and loading conditions



c) PF with distributed initial damage size

d) PF with true value of initial damage size

Figure 9. Comparison of NN and PF with the case of complex damage growth under small noise

results from the two methods are not much different, but PF and BM, respectively, have advantage in terms of wide range of applications and a fast calculation. Further, the case of no loading conditions and no physical models, short term prediction can be done by using data-driven approaches with at least three data sets. For long term predictions, loading conditions are additionally required.

5. CONCLUSIONS

This paper provides a practical review of both data-driven and physics-based approaches for the purpose of prognostics. As common prognostics algorithms, NN, GP, PF and BM are introduced and employed for case studies under practical conditions to discuss about attributes, pros and cons, and applicable conditions. Even if advanced algorithms are available, the basic algorithms are employed in this study, and the results are analyzed focusing on their intrinsic properties. This will be helpful for the beginners in prognostics to choose the best algorithm for their field of applications.

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

An, D., Choi, J. H., Schmitz, T. L., & Kim, N. H. (2011). In-situ monitoring and prediction of progressive joint wear using Bayesian statistics. *Wear*, vol. 270(11-12), pp. 828-838.

An, D., & Choi, J. H. (2012). Efficient reliability analysis based on Bayesian framework under input variable and metamodel uncertainties. *Structural and Multidisciplinary Optimization*, vol. 46, pp. 533-547.

An, D., Choi, J. H., & Kim, N. H. (2012). A comparison study of methods for parameter estimation in the physics-based prognostics. *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, April 23-26, Honolulu, Hawaii.

An, D., Choi, J. H., & Kim, N. H. (2012). Identification of correlated damage parameters under noise and bias using Bayesian inference. *Structural Health Monitoring*, vol. 11(3), pp. 293-303.

An, D., & Choi, J. H. (2013). Improved MCMC method for parameter estimation based on marginal probability density function. *Journal of Mechanical Science and Technology*, vol. 27(6).

Andrieu, C., Freitas, de N., Doucet, A., & Jordan, M. (2003). An introduction to MCMC for machine learning. *Machine Learning*, vol. 50(1), pp. 5-43.

Bayes, T. (1763.) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, vol. 53, pp. 370-418.

Belhouari, S. B., & Bermak, A. (2004). Gaussian process for nonstationary time series prediction. *Computational Statistics & Data Analysis*, vol. 47, pp. 705-712.

Bicciato, S., Pandin, M., Didonè, G., & Bello, C. D. (2001). Analysis of an associative memory neural network for pattern identification in gene expression data. *Biokdd01, Workshop on Data Mining in Bioinformatics*, August 26, San Francisco, CA.

Bodén, M. (2002). A guide to recurrent neural networks and backpropagation. in *The DALLASproject. Report from the NUTEK-supported project AIS-8: Application of Data Analysis with Learning Systems, 1999-2001*. Holst, A. (ed.), SICS Technical Report T2002:03, SICS, Kista, Sweden.

Bretscher, O. (1995). *Linear Algebra with Applications*, 3rd ed. Upper Saddle River NJ: Prentice Hall.

Chakraborty, K., Mehrotra, K., Mohan, C. K., & Ranka, S. (1992). Forecasting the behavior of multivariate time series using neural networks. *Neural Networks*, vol. 5, pp. 961-970.

Chang, J. F., & Hsieh, P. Y. (2011). Particle swarm optimization based on back propagation network forecasting exchange rates. *International Journal of Innovative Computing, Information and Control*, vol. 7(12), pp. 6837-6847.

Chen, S. C., Lin, S. W., Tseng, T. Y., & Lin, H. C. (2006). Optimization of back-propagation network using simulated annealing approach. *IEEE International Conference on Systems, Man, and Cybernetics*, October 8-11, Taipei, Taiwan.

Cheng, S., & Pecht, M. (2009). A fusion prognostics method for remaining useful life prediction of electronic products. *5th Annual IEEE Conference on Automation Science and Engineering*, August 22-25, Bangalore, India.

Chryssoloiuris, G., Lee, M., & Ramsey, A. (1996). Confidence interval prediction for neural network models. *IEEE Transactions ON Neural Networks*, vol. 7(1), pp. 229-232.

Coppe, A., Haftka, R. T., Kim, N. H., & Yuan, F. G. (2009). Reducing uncertainty in damage growth properties by structural health monitoring. *Annual Conference of the Prognostics and Health Management Society*, September 27-October 1, San Diego, CA.

Coppe, A., Pais, M. J., Haftka, R. T., & Kim, N. H. (2012). Remarks on using a simple crack growth model in predicting remaining useful life. *Journal of Aircraft*, vol. 49, pp. 1965-1973.

Daigle, M., & Goebel, K. (2011). Multiple damage progression paths in model-based prognostics. *IEEE Aerospace Conference*, March 05-12, Big Sky, Montana.

DeCastro, J. A., Tang, L., Loparo, K. A., Goebel, K., & Vachtsevanos, G. (2009). Exact nonlinear filtering and prediction in process model-based prognostics. *Annual Conference of the Prognostics and Health Management Society*, September 27-October 1, San Diego, CA.

Dickson, D. C. M., & Waters, H. R. (1993). Gamma processes and finite time survival probabilities. *Astin Bulletin*, vol. 23(2), pp. 259-272.

Doucet, A., De Freitas, N., & Gordon, N. J. (2001). *Sequential Monte Carlo methods in practice*. Springer-Verlag.

Doukim, C. A., Dargham, J. A., & Chekima, A. (2010). Finding the number of hidden neurons for an MLP neural network using coarse to fine search technique. *10th International Conference on Information Science, Signal Processing and their Applications*, May 10-13, Kuala Lumpur, Malaysia.

Drucker, H., Cortes, C., Jackel, L. D., LeCun, Y., & Vapnik, V. (1994). Boosting and other ensemble methods. *Neural Computation*, vol. 6(6), pp. 1289-1301.

Duch, W., & Jankowski, N. (1999). Survey of neural transfer functions. *Neural Computing Surveys*, vol. 2, pp. 163-212.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. FL: Chapman & Hall/CRC.

Firth, A. E., Lahav, O., & Somerville, R. S., (2003). Estimating photometric redshifts with artificial neural networks. *Monthly Notices of the Royal Astronomical Society*, vol. 339, pp. 1195-1202.

Foster, L., Waagen, A., Aijaz, N., Hurley, M., Luis, A., Rinsky, J., Satyavolu, C., Way, M. J., Gazis, P., & Srivastava, A. (2009). Stable and efficient gaussian process calculations. *Journal of Machine Learning Research*, vol. 10, pp. 857-882.

Gelfand, A. E., & Sahu, S. K. (1994). On Markov chain Monte Carlo acceleration. *Journal of Computational and Graphical Statistics*, vol. 3, pp. 261-276.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2004). *Bayesian data analysis*. NY: Chapman and Hall.

Gilks, W. R., & Berzuini, C. (2001). Following a moving target-Monte Carlo inference for dynamic Bayesian models. *Royal Statistical Society B*, vol. 63, Part 1, pp. 127-146.

Giurgiutiu, V. (2002). Current issues in vibration-based fault diagnostics and prognostics. *SPIE's 9th Annual International Symposium on Smart Structures and Materials and 7th Annual International Symposium on NDE for Health Monitoring and Diagnostics*, March 17-21, San Diego, CA.

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

Gómez, I., Franco, L., & Jérez, J. M. (2009). Neural network architecture selection: can function complexity help?. *Neural Processing Letters*, vol. 30, pp. 71-87.

Gouriveau, R., Dragomir, O., & Zerhouni, N. (2008). A fuzzy approach of online reliability modeling and estimation. *Proceedings of the 34th European Safety Reliability & Data Association, ESReDA Seminar and 2nd Joint ESReDA/ESRA Seminar on Supporting Technologies for Advanced Maintenance Information Management*, May 13-14, San Sebastian, Spain.

Grall, A., Bérenguer, C., & Dieulle, L. (2002). A condition-based maintenance policy for stochastically deteriorating systems. *Reliability Engineering and System Safety*, vol. 76(2), pp. 167-180.

Gu, J., Azarian, M. H., & Pecht, M. G. (2008). Failure prognostics of multilayer ceramic capacitors in temperature-humidity-bias conditions. *2008 International Conference on Prognostics and Health Management*, October 6-9, Denver, Colorado.

Guan, X., Liu, Y., Saxena, A., Celaya, J., & Goebel, K. (2009). Entropy-Based probabilistic fatigue damage prognosis and algorithmic performance comparison. *Annual Conference of the Prognostics and Health Management Society*, September 27-October 1, San Diego, CA.

Happel, B. L. M., & Murre, J. M. J. (1994). The design and evolution of modular neural network architectures. *Neural Networks*, vol. 7, pp. 985-1004.

He, Y., Tan, Y., & Sun, Y. (2004) Wavelet neural network approach for fault diagnosis of analogue circuits. *IEE Proceedings-Circuits, Devices and Systems*, vol. 151(4), pp. 379-384.

Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation*, vol. 59(1), pp. 1-23.

Huang, X., Torgeir, M., & Cui, W. (2008). An engineering model of fatigue crack growth under variable amplitude loading. *International Journal of Fatigue*, vol. 30(1), pp. 2-10.

Jacobs, R. A. (1995). Methods for combining experts' probability assessments. *Neural Computation*, vol. 7(5), pp. 867-888.

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

Julier, S. J., & Uhlmann, J. K. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, vol. 92(3), pp. 401-422.

Juricic, D., Znidarsic, A., & Fussel, D. (1997). Generation of diagnostic trees by means of simplified process models and machine learning. *Engineering Applications of Artificial Intelligence*, vol. 10(1), pp. 15-29.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transaction of the ASME-Journal of Basic Engineering*, No. 82, pp. 35-45.

Khawaja, T., Vachtsevanos, G., & Wu, B. (2005). Reasoning about uncertainty in prognosis: a confidence prediction neural network approach. *2005 Annual Meeting of the North American Fuzzy Information Processing Society*, June 22-25, Ann Arbor, Michigan.

Kim, S., & Park, J. S. (2011). Sequential Monte Carlo filters for abruptly changing state estimation. *Probabilist Engineering Mechcnics*, vol. 26, pp. 194-201.

Kitagawa, G. (1987). Non–Gaussian state space modeling of nonstationary time series (with discussion). *Journal of the American Statistical Association*, vol. 82(400), pp. 1032-1063.

Kleijnen, J. P. C. (1995). Statistical validation of simulation models. *European Journal of Operational Research*, vol. 87, pp. 21-34.

Kramer, S. C., & Sorenson, H. W. (1998). Bayesian parameter estimation. *IEEE Transactions on Automatic Control*, vol. 33(2), pp. 217-222.

Krogh, A. (2008). What are artificial neural networks?. *Nature Biotechnology*, vol. 26(2), pp. 195-197.

Krogh, A., & Vedelsby, J. (1995). Neural network ensembles, cross validation, and active learning. In Tesauro G., Touretzky D., & Leen T. (Eds.), *Advances in Neural Information Processing Systems* (231-238). The MIT Press.

Lawrence, N., Seeger, M., & Herbrich, R. (2003). Fast sparse Gaussian Process methods: the information vector machine. In Becker S., Thrun S., & Obermayer K. (Eds.), *Advances in Neural Information Processing Systems*(625-632), Vol. 15, MIT Press.

Lawrence, S., Giles, C. L., & Tsoi, A. C. (1998). *What size neural network gives optimal generalization? convergence properties of backpropagation*. Technical Reports. UM Computer Science Department, UMIACS.

Leonard, J. A., Kramer, M. A., & Ungar, L. H. (1992). A Neural network architecture that computes its own reliability, *Computers in Chemical Engineering*, vol. 16(9), pp. 819-835.

Liu, D., Pang, J., Zhou, J., & Peng, Y. (2012). Data-driven prognostics for lithium-ion battery based on Gaussian process regression. *2012 Prognostics and System Health Management Conference*, May 23-25, Beijing, China.

Liu, J., Saxena, A., Goebel, K., Saha, B., & Wang, W. (2010). An adaptive recurrent neural network for remaining useful life prediction of lithium-ion batteries. *Annual Conference of the Prognostics and Health Management Society*, October 10-16, Portland, Oregon.

Liu, P., & Li, H. (2004). *Fuzzy Neural Network Theory and Application*. Singapore: World Scientific.

MacKay, D. J. C. (1997). Gaussian processes-a replacement for supervised neural networks?, *Tutorial lecture notes for NIPS*, UK, http://www.inference.phy.cam.ac.uk/mackay/BayesGP.html.

Mackay, D. J. C. (1997). *Introduction to Gaussian processes*. Technical report. Cambridge University, UK, http://www.cs.utoronto.ca/~mackay/gpB.pdf

Mao, K. Z., Tan, K.-C., & Ser, W. (2000). Probabilistic neural-network structure determination for pattern classification. *IEEE Transactions on Neural Networks*, vol. 11(4), pp. 1009-1016.

Martin, K. F. (1994). A review by discussion of condition monitoring and fault diagnosis in machine tools. *International Journal of Machine Tools and Manufacture*, vol. 34(4), pp. 527-551.

Melkumyan, A., & Ramos, F. (2009). A sparse covariance function for exact gaussian process inference in large datasets. *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence*, July 11-17, Pasadena, CA.

Mohanty, S., Teale, R., Chattopadhyay, A., Peralta, P., & Willhauck, C. (2007). Mixed Gaussian process and state-space approach for fatigue crack growth prediction. *International Workshop on Structural Heath Monitoring*, vol. 2, pp. 1108-1115.

Mohanty, S., Das, D., Chattopadhyay, A., & Peralta, P. (2009). Gaussian process time series model for life prognosis of metallic structures. *Journal of Intelligent Material Systems and Structures*, vol. 20, pp. 887-896.

Naftaly, U., Intrator, N., & Horn, D. (1997). Optimal ensemble averaging of neural networks. *Network: Computation in Neural Systems*, vol. 8(3), pp. 283-296.

Navone, H. D., Granitto, P. M., Verdes, P. F., & Ceccatto, H. A. (2001). A Learning algorithm for neural network ensembles. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*. No. 12, pp 70-74.

Nawi, N. M., Ransing, R. S., & Ransing, M. R. (2007). An improved conjugate gradient based learning algorithm for back propagation neural networks. *International Journal of Computational Intelligence*, vol. 4(1), pp. 46-55.

Neal, R. M. (1998). Regression and classification using Gaussian process priors. In Bernardo J. M., Berger J. O., Dawid A. P., & Smith A. F. M. (Eds.), *Bayesian statistics* (475-501), Vol. 6, NY: Oxford University Press.

Orchard, M. E., & Vachtsevanos, G. J. (2007). A particle filtering approach for on-line failure prognosis in a planetary carrier plate. *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 7(4), pp. 221-227.

Ostafe, D. (2005). Neural network hidden layer number determination using pattern recognition techniques. *2nd Romanian-Hungarian Joint Symposium on Applied computational Intelligence*, May 12-14, Timisoara, Romania.

Paciorek, C., & Schervish, M. J. (2004). Nonstationary covariance functions for gaussian process regression. In

Thrun S., Saul L., & Schölkopf B. (Eds.), *Advances in neural information processing systems*, Vol. 16, MIT Press.

Pandey, M. D., & Noortwijk, J. M. V. (2004). Gamma process model for time-dependent structural reliability analysis. In Watanabe E., Frangopol D. M., & Utsonomiya T. (Eds.), *Bridge maintenance, safety, management and cost* (101-102). London: Taylor and Francis Group.

Paris, P. C., & Erdogan, F. (1963). A critical analysis of crack propagation laws. *Transactions of the ASME, Journal of Basic Engineering, Series D*, vol. 85(3), pp. 528-534.

Parzen, E. (1962). On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, vol. 33(3), pp. 1065-1076.

Perrone, M. P., & Cooper, L. N. (1993). When networks disagree: ensemble method for neural networks. In Mammone R. J. (Ed.), *Neural Networks for Speech and Image processing*, Chapman-Hall.

Petalas, P., Spyridonos, P., Glotsos, D., Cavouras, D., Ravazoula, P., & Nikiforidis, G. (2003). Probabilistic neural network analysis of quantitative nuclear features in predicting the risk of cancer recurrence at different follow-up times. *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, September 18-20, Rome, Italy.

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network-first principles approach to process modeling. *AIChE Journal*, vol. 38(10), pp. 1499-1511.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, vol. 77(2), pp. 257-286.

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

Rebba, R., Huang, S., Liu, Y., & Mahadevan, S. (2006). Statistical validation of simulation models. *International Journal of Materials and Product Technology*, vol. 25(1/2/3), pp. 164-181.

Rebba, R., Mahadevan, S., & Huang, S. (2006). Validation and error estimation of computational models. *Reliability Engineering and System Safety*, vol. 91, pp. 1390-1397.

Rovithakis, G. A., Maniadakis, M., & Zervakis, M., (2004). A hybrid neural network/genetic algorithm approach to optimizing feature extraction for signal classification. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 34(1), pp. 695-702.

Rubin, D. B. (1998). Using the SIR algorithm to simulate posterior distributions. In Bernardo J. M., DeGroot M. H., Lindley D. V., & Smith A. F. M. (Eds.), *Bayesian statistics*(395-402), vol. 3, Cambridge, MA: Oxford University Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. *Parallel Distributed Processing: Ex-plorations in the Microstructure of Cognition*, vol. 1: foundations, MIT Press, pp. 318-362.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, vol. 4(4), pp. 409-423.

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

Salomon, R., & Hemmen, J. L. V. (1996). Accelerating backpropagation through dynamic self-adaptation. *Neural Networks*, vol. 9(4), pp. 589-601.

Samanta, B., & Al-Balushi, K. R. (2003). Artificial neural network based fault diagnostics of rolling element bearings using time-domain features. *Mechanical Systems and Signal Processing*, vol. 17(2), pp. 317–328.

Samuel, P. D., & Pines, D. J. (2005). A review of vibration-based techniques for helicopter transmission diagnostics. *Journal of Sound and Vibration*, vol. 282(1-2), pp. 475-508.

Sankavaram, C., Pattipati, B., Kodali, A., Pattipati, K., Azam, M., Kumar, S., & Pecht, M. (2009). Model-based and data-driven prognosis of automotive and electronic systems. *5th Annual IEEE Conference on Automation Science and Engineering*, August 22-25, Bangalore, India.

Santner, T. J., Williams, B. J., & Notz, W. I. (2003). *The design and analysis of computer experiments*. NY: Springer Verlag.

Sargent, R. G. (2009). Verification and validation of simulation models. *Proceedings of the 2009 Winter Simulation Conference*, December 13-16, Austin, TX.

Seeger, M. (2004). Gaussian processes for machine learning. *International Journal of Neural Systems*, vol. 14(2), pp. 69-106.

Si, X. S., Wang, W., Hu, C. H., & Zhou, D. H (2011). Remaining useful life estimation – A review on the statistical data driven approaches. *European Journal of Operational Research*, vol. 213, pp. 1-14.

Singh, G. K., & Al Kazzaz, S. A. S. (2003). Induction machine drive condition monitoring and diagnostic research-a survey. *Electric Power Systems Research*, vol. 64, pp. 145-158.

Smola, A. J., & Bartlett, P. L. (2001). Sparse greedy Gaussian process regression. In Leen T. K., Dietterich T. G., & Tresp V. (Eds.), *Advances in Neural Information Processing Systems* (619-625), Vol. 13. MIT Press.

Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*, vol. 3, pp. 109-118.

Srivastava, A. N., & Das, S. (2009). Detection and prognostics on low-dimensional systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 39(1), pp. 44-54.

Storvik, G. (2002). Particle filters in state space models with the presence of unknown static parameters. *IEEE Transactions on Signal Processing*, vol. 50(2), pp. 281-289.

Subudhi, B, Jena, D., & Gupta, M. M. (2008). Memetic differential evolution trained neural networks for nonlinear system identification. *IEEE Region 10 Colloquium and the Third International Conference on Industrial and Information Systems*, December 8 -10, Kharagpur, India.

Sugumaran , V., Sabareesh, G. R., & Ramachandran, K. I. (2008). Fault diagnostics of roller bearing using kernel based neighborhood score multi-class support vector machine. *Expert Systems with Applications*, vol. 34, pp. 3090-3098.

Svozil, D., Kvasnička, V., & Pospíchal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, vol. 39, pp. 43-62.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, vol. 1, pp. 211-244.

Tran, V. T., & Yang, B. S. (2009). Data-driven approach to machine condition prognosis using least square regression tree. *Journal of Mechanical Science and Technology*, vol. 23, pp. 1468-1475.

Veaux, R. D., Schumi, J., Schweinsberg, J., & Ungar, L. H. (1998). Prediction intervals for neural networks via nonlinear regression. *Technometrics*, vol. 40(4), pp. 273-282.

Wang, W. P., Liao, S., & Xing, T. W. (2009). Particle filter for state and parameter estimation in passive ranging. *IEEE International Conference on Intelligent Computing and Intelligent Systems*. November 20-22, Shanghai, China.

Williams, C. K. I. (1997). Computing with infinite networks. In Mozer M. C., Jordan M. I., & Petsche T. (Eds.), *Advances in neural information processing systems*, Cambridge, MA: MIT Press.

Xing, Y., Miao, Q., Tsui, K.-L., & Pecht, M. (2011). Prognostics and health monitoring for lithium-ion battery," *2011 IEEE International Conference*, July 10-12, Beijing, China.

Xing, Y., Williard, N., Tsui, K.-L., & Pecht, M. (2011). A comparative review of prognostics-based reliability methods for lithium batteries. *Prognostics and System Health Management Conference*, May 24-25, Shenzhen, China.

Xu, J., & Xu, L. (2011). Health management based on fusion prognostics for avionics systems," *Journal of Systems Engineering and Electronics*, vol. 22(3), pp. 428–436.

Yam, R. C. M., Tse, P. W., Li, L., & Tu, P. (2001). Intelligent predictive decision support system for condition-based maintenance. *International Journal of Advanced Manufacturing Technology*, vol. 17(5), pp. 383-391.

Yan, R., & Gao, R. X. (2007). Approximate entropy as a diagnostic tool for machine health monitoring. *Mechanical Systems and Signal Processing*, vol. 21, pp. 824–839.

Yang, L., Kavli, T., Carlin, M., Clausen, S., & Groot, P. F. M. (2000). An evaluation of confidence bound estimation methods for neural networks. *European Symposium on Intelligent Techniques 2000*, September 14-15, Aachen, Germany.

Yao, X. (1999). Evolving artificial neural networks. *Proceedings of the IEEE*, vol. 87(9), pp. 1423-1447.

Zhang, J., & Lee, J. (2011). A review on prognostics and health monitoring of li-ion battery. *Journal of Power Sources*, vol. 196, pp. 6007-6014.

Zio, E., & Maio, F. D. (2010). A data-driven fuzzy approach for predicting the remaining useful life in dynamic failure scenarios of a nuclear system. *Reliability Engineering and System Safety*, vol. 95, pp. 49-57.

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 the Bayesian inference, correlated parameter identification and the methodology for prognostics and health management and structural health monitoring.

Nam Ho Kim received the B.S. degree of mechanical engineering from Seoul National University in 1989, the M.S. degree and Ph.D. degree 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 Associate 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. degree and Ph.D. degree 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 life-time reliability, and prognostics and health management.