

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

Dawn An<sup>1</sup>

*University of Florida, Gainesville, FL, 32611  
Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea*

Nam Ho Kim<sup>2</sup>

*University of Florida, Gainesville, FL, 32611*

and

Joo Ho Choi<sup>3</sup>

*Korea Aerospace University, Goyang-si, Gyeonggi-do, 412-791, Korea*

**Condition-based maintenance (CBM) 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, pros and cons using simple examples.**

## I. Introduction

CONDITION-based maintenance (CBM) illustrated in Figure 1 is a cost-effective maintenance strategy, in which maintenance schedules are predicted based on the results provided from diagnostics and prognostics. Diagnostics characterizes the status of damage through detection, isolation and identification using collected data. Structural health monitoring provides an accurate quantification of degradation and damage at an early stage without intrusive and time-consuming inspections by utilizing on-board sensors/actuators. Based on diagnostics information, prognostics predicts the future behavior of damage and the remaining useful life (RUL), which is remaining time/cycles before required maintenance. The collected data that is important for both diagnostics and prognostics can be classified into two types; (1) event data or run-to-failure data and (2) condition monitoring (CM) data, which provides the information of damage, including wear volume, crack size, vibration signal, oil debris, thermography, and acoustics. Some CM data are direct, such as crack size, while others are indirect, such as oil debris. In the case of indirect CM data, signal processing and damage quantification process are required to be used to diagnostics and prognostics.

There are several reviews on diagnostics methods<sup>1-4</sup> as well as a large number of publications on diagnostics,<sup>5-10</sup> signal processing and damage quantification.<sup>11-13</sup> For example, Jardine et al.<sup>1</sup> provided an overall review on diagnostics and prognostics in terms of implementing CBM and decision making. There are also other references for CBM.<sup>14,15</sup> On the other hand, however, a relatively small number of reviews on prognostics are available.<sup>16-21</sup> Most

---

<sup>1</sup> Joint Ph.D. student, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611, Student member.

<sup>2</sup> Associate professor, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL, 32611, Member.

<sup>3</sup> Professor, School of Aerospace and Mechanical Engineering, Korea Aerospace University, 100 Hwajeon-dong, Deokyang-gu, Goyang-City, Gyeonggi-do, Korea, 412-791.

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, 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, which is illustrated in Figure 2. 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. There are two main differences between data-driven and physics-based approaches; (1) availability of a physical model and (2) usage of training data to identify the characteristic of the damage state. Hybrid approaches combine the above-mentioned two methods to improve the prediction performance.<sup>22-26</sup> 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 shape. Typical prognostics algorithms are tested under such conditions, and the algorithms' attributes, 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, in which the attribute of different prognostics methods are analyzed in order to help selecting an appropriate method, followed by conclusions in Section 5.

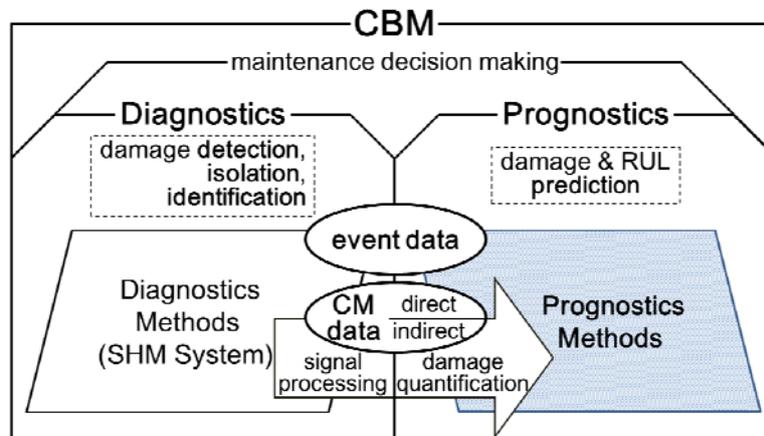


Figure 1. CBM building blocks.

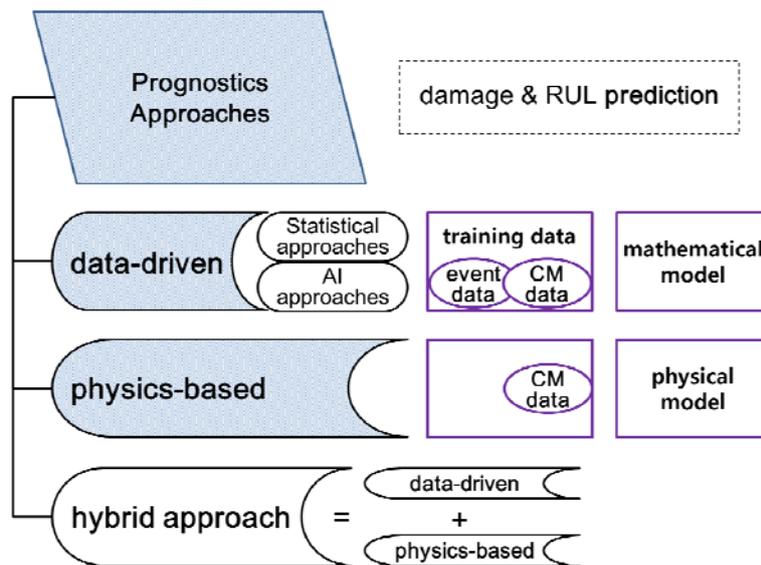


Figure 2. Categorization and definition of prognostics methods.

## II. 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 or weight parameters are employed. The weight parameters are determined based on the training data that are obtained under the various usage conditions or at previous times under a given usage condition. 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)<sup>27-29</sup> and fuzzy logic,<sup>30,31</sup> and (2) the statistical approaches that include gamma process,<sup>32,33</sup> hidden Markov model (HMM),<sup>34</sup> and regression-based model such as Gaussian process (GP) regression,<sup>35,36</sup> relevance vector machine (RVM),<sup>37</sup> and least square (LS) regression,<sup>38,39</sup> etc. Among these algorithms, NN and GP are commonly used for prognostics and will be discussed in the following sections.

### A. 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),<sup>40</sup> is illustrated in Figure 3. 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. 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. 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.<sup>41</sup> Weight parameters include weights for the interconnected nodes and biases that are added to inputs of transfer functions.<sup>42,43</sup> 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<sup>44</sup> 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. Recurrent NN (RNN)<sup>45</sup> is the next most common

architecture, whose concept is not much different from FFNN except that it has local feedback connections between input and hidden/output nodes. Since Levenberg-Marquardt (LM) back-propagation algorithm<sup>46</sup> is usually employed for RNN's learning algorithm, RNN and LMNN are sometimes used for the same meaning. In addition to FFNN and RNN, there exists fuzzy-neural,<sup>47</sup> wavelet,<sup>48</sup> associative-memory,<sup>49</sup> modular,<sup>50</sup> and hybrid<sup>51,52</sup> neural network.

In the following, three important issues are discussed for NN-based prognostics. It is necessary that users must be aware of these issues in order to use the algorithm properly.

- **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 et al.<sup>53</sup> and Doukim et al.<sup>54</sup> investigated the usage of mean square error in order to find the optimal number of hidden nodes. Gómez et al.<sup>55</sup> 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. Otafe<sup>56</sup> 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.<sup>27</sup> 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. It is observed by Chakraborty et al.<sup>27</sup> that the accuracy of prediction results is not proportional to the number of input nodes. Chang and Hsieh<sup>57</sup> 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.

- **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<sup>58</sup> presented a dynamic self-adaptation algorithm to accelerate the steepest descent method, and Chen et al.<sup>59</sup> 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 et al.<sup>60</sup> proposed a technique combining the genetic algorithm and differential evolution with BP, and Nawi et al.<sup>61</sup> 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,<sup>62-67</sup> and the other efforts are found in the Refs.<sup>46,58,61</sup> However, finding good weight parameters is still challenging, and the performance of NN algorithm deteriorates with non-optimal weight parameters.

- **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.<sup>68-71</sup> Bootstrapping<sup>72</sup> can also be applied, which can be easily implemented by running Matlab NN toolbox several times because Matlab uses training data for obtaining weight parameters, validation and test. 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.<sup>42</sup> 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)<sup>73</sup> using Parzen estimator.<sup>74</sup> However, most papers employ PNN for classification or risk diagnosis,<sup>75-77</sup> and prognostics ones are rarely found except for the study by Khawaja et al.<sup>78</sup> 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. Although NN algorithms can handle uncertainty in the regress process, they are not accustomed to deal with the noise and bias in data.

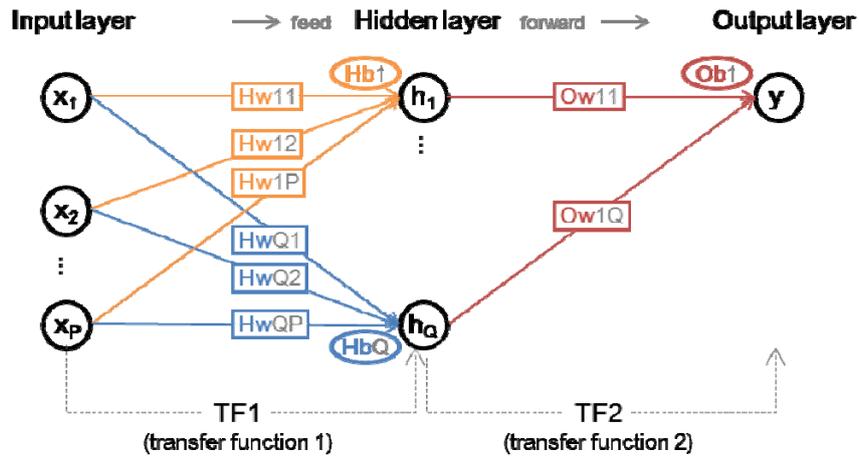


Figure 3. Illustration of typical network model: FFNN.

## B. GP Regression

GP is a commonly used method among regression-based data-driven approaches for prognostics, whose conceptual property is illustrated in Figure 4. 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 4. The reason for this can be explained with following GP model that is composed of a global model ( $\mathbf{f}\boldsymbol{\beta}$ ) and departures ( $\mathbf{r}\mathbf{R}^{-1}(\mathbf{Y} - \mathbf{F}\boldsymbol{\beta})$ ):

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

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  $(\mathbf{F}^T \mathbf{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{R}^{-1} \mathbf{Y}$ , and  $\mathbf{r}$  is a  $1 \times n$  ( $n$  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  $\boldsymbol{\beta} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{Y}$ , which is the LS regression<sup>38</sup> that might be easy to understand for all. Consequently, GP is distinguishable from LS 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.<sup>79,80</sup>

### ▪ 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.<sup>81,82</sup> Mohanty et al.<sup>83</sup> 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<sup>84</sup> 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<sup>85</sup> used nonstationary GP to predict respiration signal, and compared with a GP model with an exponential covariance function, and Liu et al.<sup>86</sup> used the combination of three covariance functions to predict lithium-ion battery degradation (state of health, SOH).

▪ **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<sup>87</sup> 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.<sup>88</sup> 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.,<sup>83</sup> however, considered about proper scaling procedures to take account of the scale of input and output variables, and obtained better results than ones from the procedure using just normalized values. Neal<sup>89</sup> considered the scale parameters as distributions rather than deterministic values, and An and Choi<sup>88</sup> showed that the GP models with scale parameters identified as distributions outperform the one using optimal deterministic parameter.

▪ **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 said that direct methods of inversion may become prohibitive when the number of the data points is greater than 1000.<sup>90</sup> As a solution to relieve such problem, the methods to select a subset of data points are usually employed.<sup>91-93</sup> While Melkumyan and Ramos<sup>94</sup> 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.<sup>22</sup> and Liu et al.<sup>86</sup>, respectively, showed the prediction confidence interval of crack length and SOH using GP.

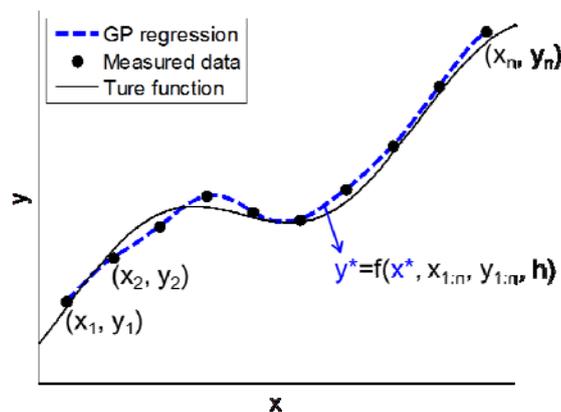


Figure 4. Illustration of GP regression.

### III. Reviews on Physics-Based Approaches

Physics-based approaches combine the physical damage model with measured data to predict future behavior of degradation or damage, and predict the RUL, which is illustrated in Figure 5. The behavior of the physical model depends on the model parameters that are obtained from laboratory test, or 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 5. There are similar issues to data-driven approaches, but

algorithms of physics-based approaches are not as much different each other as data-driven approaches. Therefore, the following subsections are built around issues in physics-based approaches rather than their algorithms.

### A. Issue 1: Model Problem (Physical Model Accuracy)

Since the physics-based approaches assume that a physics model describing the behavior of damage is available, 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.<sup>95-98</sup> In general, the number of model parameters increases as model complexity increases, which makes it difficult to identify the model. Recently, Coppe et al.<sup>99</sup> showed that the issue of model accuracy can be relieved by identifying equivalent model parameters that can compensate for the model error because prediction results are obtained from the combination of the model and its parameters. They showed that a simple Paris model with an assumed stress-intensity factor can be used for predicting crack growth of complex geometries by compensating for the error in the simple model by adjusting model parameters. Although this is limited to the case of a similar model form, cumbersome efforts to validate the model accuracy can be eased off.

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

#### a) 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 (KF),<sup>100</sup> extended Kalman filter (EKF),<sup>101</sup> particle filter (PF),<sup>102</sup> and Bayesian method (BM).<sup>103</sup> These algorithms are based on the Bayesian inference,<sup>104</sup> 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 et al.<sup>105</sup> used EKF to estimate model parameters and predict RUL for crack growth on a planetary carrier plate. Orchard and Vachtsevanos<sup>106</sup> estimated the crack closure effect using PF for RUL prediction of a planetary carrier plate based on vibration-based feature. Daigle and Goebel<sup>107</sup> used PF to estimate wear coefficients by considering multiple damage mechanisms in centrifugal pumps. An et al.<sup>108</sup> 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 (a.k.a. sequential Monte Carlo method<sup>105,106</sup>) 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 6. There are three steps in PF process: (1) prediction step - posterior distributions of the model parameters ( $\theta$ ) at the previous ( $k-1$ ) step are used for the prior at the current ( $k$ ) 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 6 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 ( $k+1$ ) 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  $\xi_{k-1}$  from arbitrary distribution during the prediction step; that is,  $\theta_k = \theta_{k-1} + \xi_{k-1}$  so that identical particles are not generated.<sup>107,109-111</sup> This method, however, can change probabilistic characteristic of parameters as well as increase the variance of parameters. Gilks and Berzuini<sup>112</sup>

proposed a resample-move algorithm based on PF and Markov Chain Monte Carlo (MCMC) method,<sup>113</sup> Kim and Park<sup>114</sup> 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, Markov chain Monte Carlo (MCMC) method<sup>113</sup> 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 7. It starts from to generate an arbitrary initial sample (old sample) and a new sample from arbitrary proposal distribution. The new sample is estimated with an old sample based on a criterion to be selected or not. In Figure 7, 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 and sampling way for new sample concerned with proposal distribution and 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.<sup>115,116</sup> Gelfand and Sahu<sup>117</sup> presented two distinct adaptive strategies to accelerate the convergence of a MCMC algorithm. More literatures are found in the Ref.<sup>113</sup>

**b) 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 et al.<sup>118</sup> predicted crack growth and RUL under various levels of noise and bias, in which bias and initial crack size as well as two Paris model parameters were included as unknown parameters. The parameters were not accurately identified under a large level of noise because they are correlated each other: 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. It is concluded that even though the correlation between model parameters makes it difficult to identify the correct parameters, it does not effect on RUL prediction. If the correlation structure is not identified, uncertainty in the prediction results will be amplified because the domain of joint PDF becomes wider.

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

Even though the model parameters can be offered from off-line laboratory tests, they are not intrinsic parameters and largely depend on loading conditions, boundary conditions, and geometry. Thus, it is better to identify the parameters based on measured damage data from a specific component. 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 et al.<sup>119</sup> presented a prognostics approach which detects the performance degradation of multilayer ceramic capacitors under temperature-humidity-bias conditions. Coppe et al.<sup>120</sup> showed that the uncertainty in structure-specific damage growth parameters can be progressively reduced in spite of noise and bias in sensor measurements. Guan et al.<sup>121</sup> 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.

Noise in raw data, i.e., sensor signals, interrupts to extract degradation features, which naturally has an effect on prognostics results not only physics-based approaches but data-driven ones. To relieve the noise effect, de-noising is usually conducted in signal processing processes. Zhang et al.<sup>122</sup> proposed a de-noising scheme for improving the signal-to-noise ratio (SNR) and applied it to vibration signals obtained from a helicopter gearbox testbed. Qiu et al.<sup>123</sup> introduced an enhanced robust method for bearing prognostics, which includes a wavelet filter based method for extraction of the weak fault feature from the vibration signals and self-organizing map (SOM) based method for bearing degradation assessment. Abouel-Seoud et al.<sup>124</sup> introduced robust prognostics concepts based on an optimal wavelet filter for fault identification from acoustic emission (AE) monitoring and a statistical approaches for degradation assessment of a gear tooth crack.

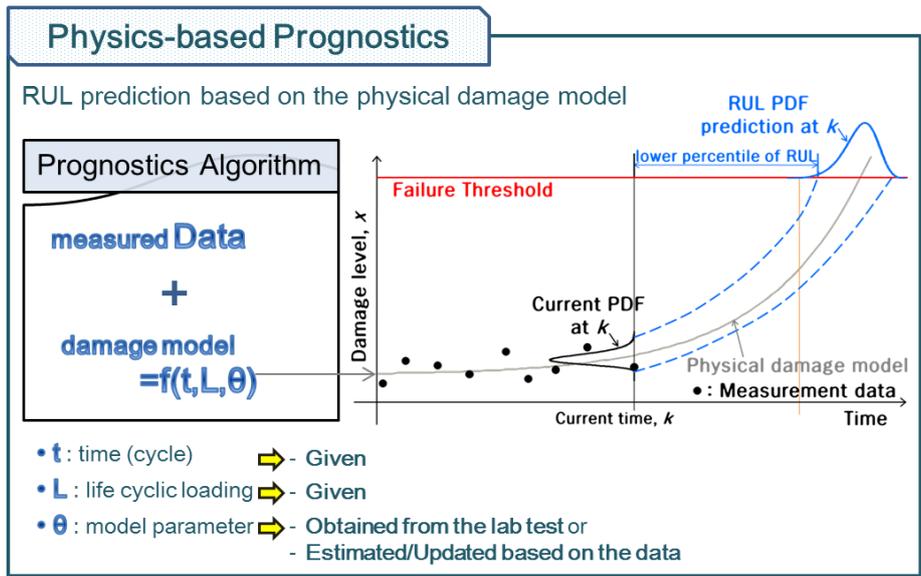


Figure 5. Illustration of physics-based prognostics.

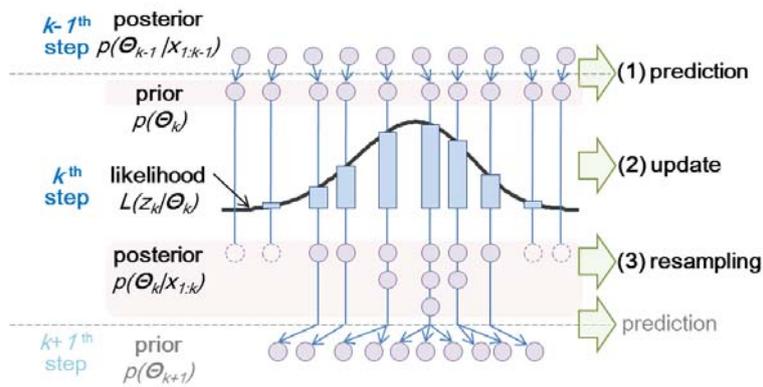


Figure 6. Illustration of PF process.

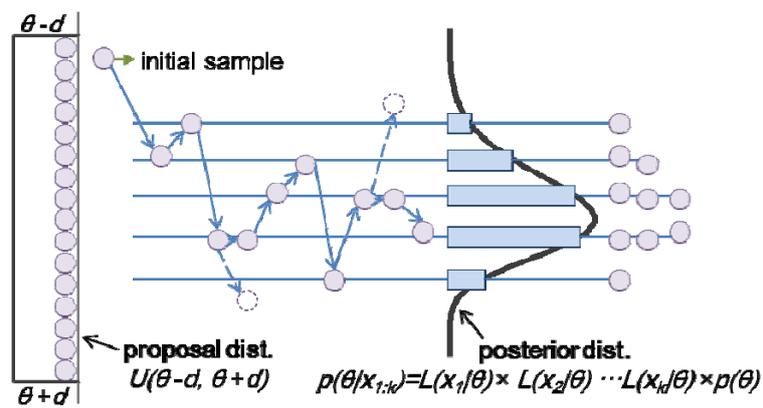


Figure 7. Illustration of BM process with MCMC.

#### IV. 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 improved versions of each algorithm, the most common and basic ones are employed. Each algorithm is tested under the cases that can be considered in real problems, and their attributes, pros and cons are discussed based on logic comes from their algorithm.

##### A. Problem Definition

###### a) Given information for case study

Paris model<sup>125</sup> and Huang's model<sup>126</sup> are, respectively, employed for a simple shape of damage growth and complex shape of damage growth, which are shown in Figure 8. On each model, there are ten sets of data under different loading conditions; ten constant loadings from 65 MPa to 83 MPa for Paris model (simple model) and ten sets of variable amplitude loadings from maximum loading 125 MPa to maximum loading 134 MPa for Huang's model (complex model). In terms of data, different level of noise and bias are added to the data in Figure 8. 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.

###### b) 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. Since input variables and hidden nodes increase the number of unknown parameters, their numbers should be adjusted according to the number of training data. In this case, the number of total weight parameters become 11 including eight weights ( $3 \times 2 + 2 \times 1$ ) 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:

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

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 ( $x_{k+1}^p, x_{k+2}^p, x_{k+3}^p, \dots$ ) 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  $x_{k+3}^p$ . On the other hand, long term prediction is multi-step ahead prediction since it utilizes predicted results as inputs, e.g.,  $x_k, x_{k+1}^p, x_{k+2}^p$  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 model parameters,  $m$  and  $C$  and six model parameters,  $m, C, \Delta K_{th}, \beta, n, \sigma_y$  in Paris model and Huang's model, and the data sets in Figure 8 were generated with following true model parameters:

Paris model parameters:  $m_{true} = 3.8, C_{true} = 1.5e-10$

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

Also, noise (measurement error) and bias can be added as unknown parameters; all unknown parameters are identified based on combination of physical model and measured data to predict future damage behavior.

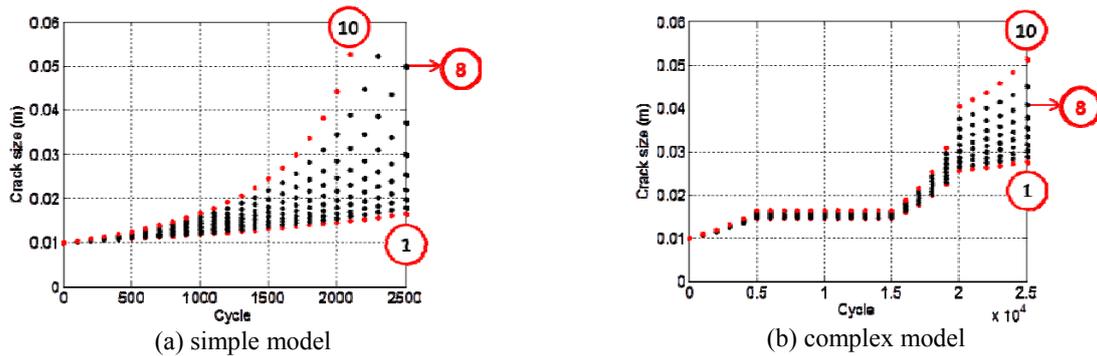


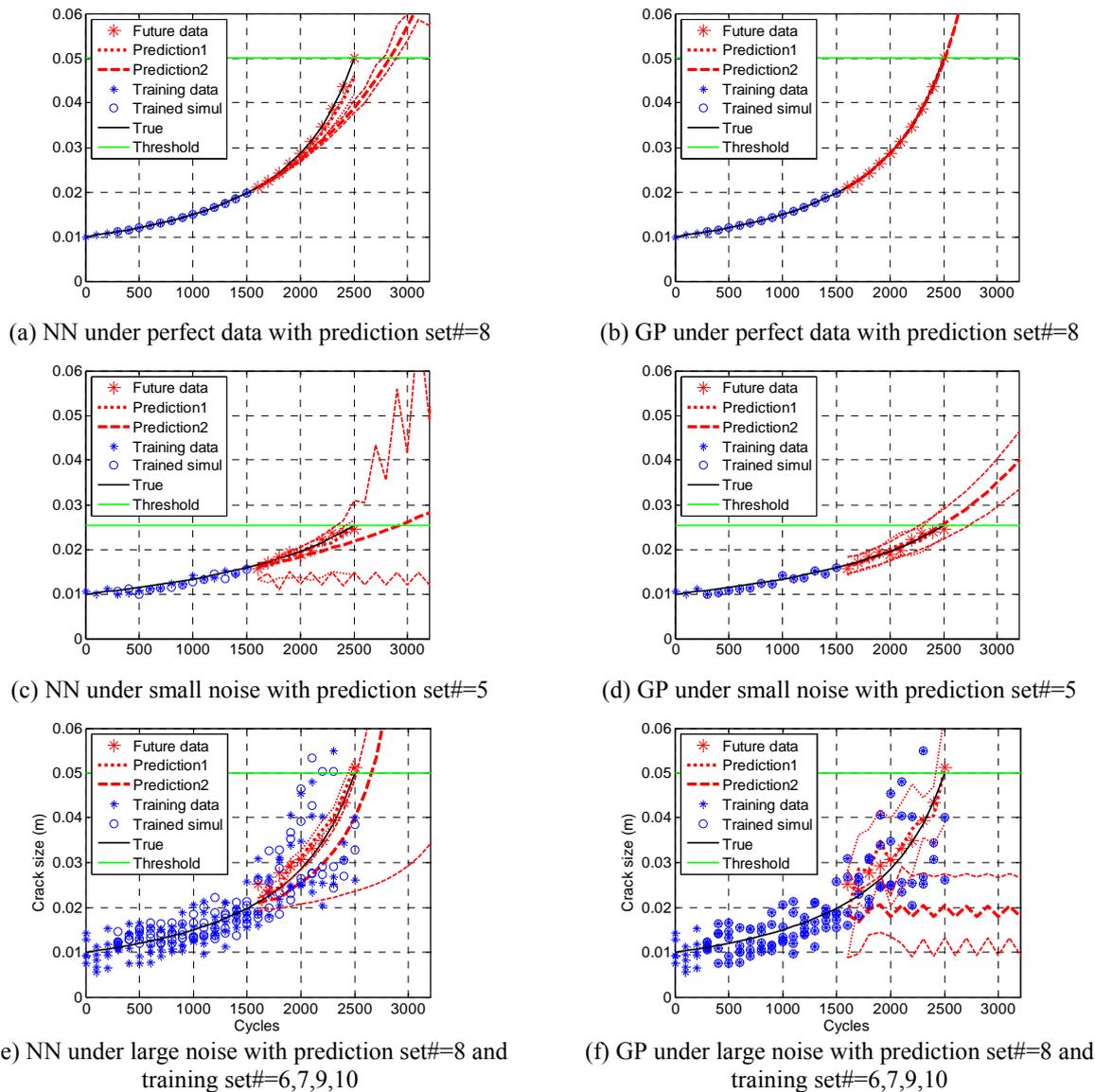
Figure 8. Problem definition for case study.

## B. Case Study Results

### a) Data-driven results

At first, the performance of training data is compared. Figure 9 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. Many sets of data do not have a good effect on condition of the correlation matrix. Also, GP is trained so that simulation results penetrate every training data points, and each data are correlated. This means that the results from GP get better as the noise level reduces because the noisy relation is applied for prediction points. 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.



**Figure 9. Results from data-driven approaches with the case of simple damage growth.**

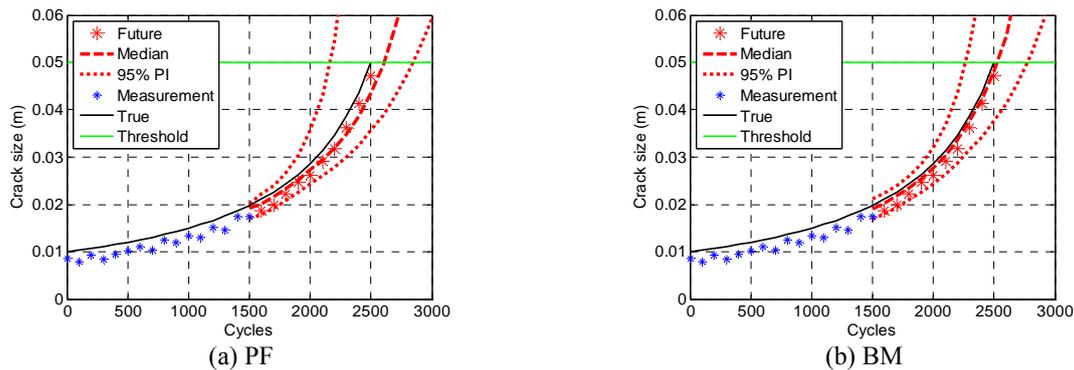
### b) Physics-based results

As mentioned before, bias in obtained data cannot be handled with data-driven approaches. In contrast, physics-based one can do it as shown in Figure 10. 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. Even though true values of measurement data and bias are unknown, it is known that the summation of them is a constant at each measurement cycle. Therefore the bias is considered as an unknown parameter and updated based on measured data containing the bias property. Further study for noise and bias in physic-based approaches is found in the Ref.<sup>118</sup>

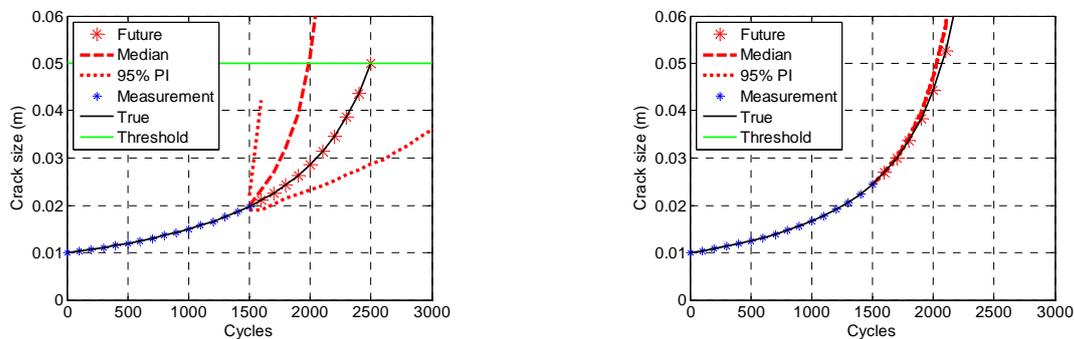
The difference between PF and BM is negligible. In fact, there are no outstanding differences between the two methods in terms of prediction results because the two methods have the same foundation with the same physical mode. The only differences are the way of updating distributions and generating samples. BM is faster than PF because the 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. From this attributes, PF can handle the case of fixed model parameters

obtained from laboratory test and express damage degradation in the form of increment. However, BM is not practical for them because of tremendous computational costs.<sup>112,127,128</sup> This is a key difference between the two methods, and more detailed comparison between PF and OBM can be found in An et al.<sup>127</sup>

Figure 11 shows different results from the two cases using parameters obtained from laboratory tests and updating the parameters. It is clear that error in parameters obtained from laboratory test produces improper results. For example, when the Paris model parameter,  $m$ , varies between 3.7 and 4.3 ( $m_{true} = 3.8$ ), the uncertainty in prediction is too wide, and the median of prediction strays from the true one as shown in Figure 11 (a). In contrast to this, the case of updating parameters shows good results even under wrong loading condition as shown in Figure 11 (b). Since the parameters from laboratory tests can be different from those in service due to environmental conditions, model parameters should be updated along with measurement data.



**Figure 10. Results from physics-based approaches with the case of simple damage growth under small noise and negative bias.**



**Figure 11. Comparison of the cases using parameters obtained from laboratory tests and parameters updated.**

### c) 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 12 (a) and (b). Three data sets among all ten are randomly selected from the usage conditions for prediction set. Based on the training data sets, NN well predicts future damage as shown in Figure 12 (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 that can be applied for NN, 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 12 (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 12 (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 12 (d) shows the results with true value of initial damage.

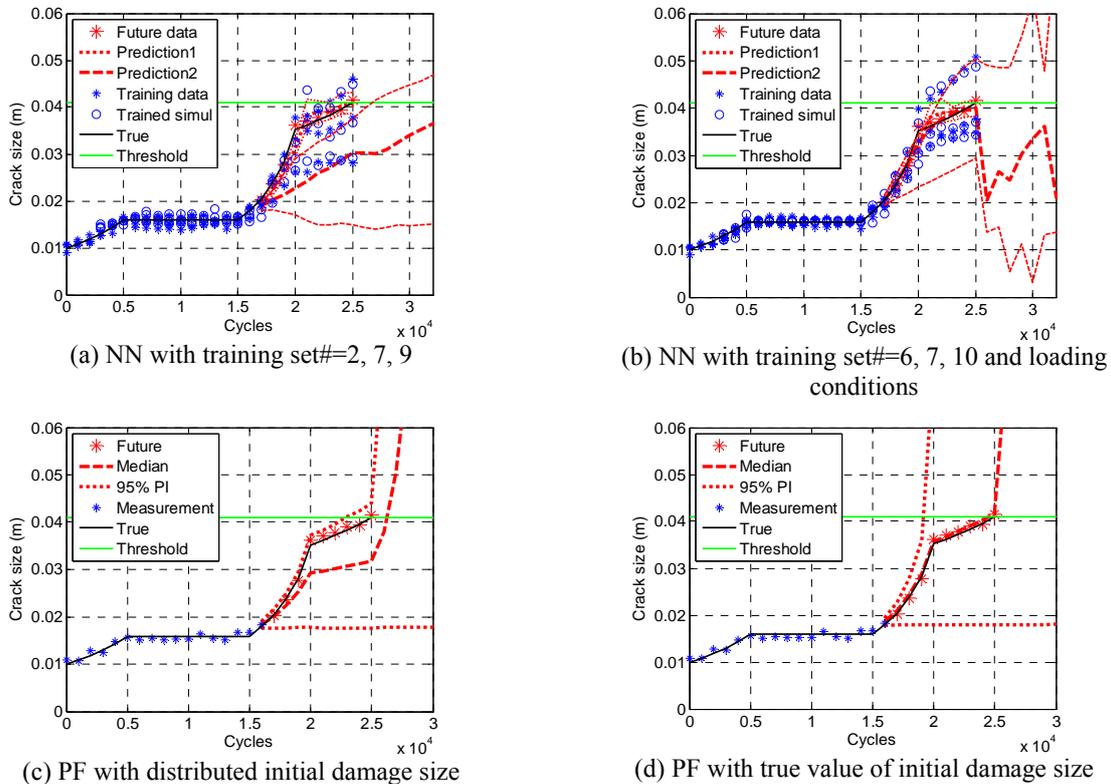


Figure 12. Comparison of NN and PF with the case of complex damage growth under small noise.

### C. 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 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.

## V. 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

- <sup>1</sup> Jardine, A. K. S., Lin, D., and Banjevic, D., "A Review on Machinery Diagnostics and Prognostics Implementing Condition-Based Maintenance," *Mechanical Systems and Signal Processing*, Vol. 20, No. 7, October 2006, pp. 1483-1510.
- <sup>2</sup> Martin, K. F., "A Review by Discussion of Condition Monitoring and Fault Diagnosis in Machine Tools," *International Journal of Machine Tools and Manufacture*, Vol. 34, No. 4, May 1994, pp. 527-551.
- <sup>3</sup> Samuel, P. D. and Pines, D. J., "A Review of Vibration-Based Techniques for Helicopter Transmission Diagnostics," *Journal of Sound and Vibration*, Vol. 282, No. 1-2, April 2005, pp. 475-508.
- <sup>4</sup> Singh, G. K. and Kazzaz, S. A. S. A., "Induction Machine Drive Condition Monitoring and Diagnostic Research-a Survey," *Electric Power Systems Research*, Vol. 64, 2003, pp. 145-158.
- <sup>5</sup> Bechhoefer, E., He, D., and Dempsey, P., "Gear Health Threshold Setting Based On a Probability of False Alarm," *Annual Conference of the Prognostics and Health Management Society*, Montreal, Quebec, Canada, September 25-29, 2011.
- <sup>6</sup> Giurgiutiu, V., *Structural Health Monitoring with Piezoelectric Wafer Active Sensors*, Academic Press, Elsevier Inc., MA, 2008.
- <sup>7</sup> Juricic, D., Znidarsic, A., and Fussel, D., "Generation of Diagnostic Trees by Means of Simplified Process Models and Machine Learning," *Engineering Applications of Artificial Intelligence*, Vol. 10, No. 1, February 1997, pp. 15-29.
- <sup>8</sup> Sugumaran, V., Sabareesh, G. R., and Ramachandran, K. I., "Fault Diagnostics of Roller Bearing Using Kernel Based Neighborhood Score Multi-Class Support Vector Machine," *Expert Systems with Applications*, Vol. 34, 2008, pp. 3090-3098.
- <sup>9</sup> Yan, R. and Gao, R. X., "Approximate Entropy as a Diagnostic Tool for Machine Health Monitoring," *Mechanical Systems and Signal Processing*, Vol. 21, 2007, pp. 824-839.
- <sup>10</sup> Samanta, B. and Al-Balushi, K. R., "Artificial Neural Network Based Fault Diagnostics of Rolling Element Bearings Using Time-Domain Features," *Mechanical Systems and Signal Processing*, Vol. 17, No. 2, 2003, pp. 317-328.
- <sup>11</sup> Radziński, M., Krawczuk, M., and Palacz, M., "Improvement of Damage Detection Methods Based on Experimental Modal Parameters," *Mechanical Systems and Signal Processing*, Vol. 25, 2011, pp. 2169-2190.
- <sup>12</sup> Ostachowicz, W., Kudela, P., Malinowski, P., and Wandowski, T., "Damage Localisation in Plate-Like Structures Based on PZT Sensors," *Mechanical Systems and Signal Processing*, Vol. 23, 2009, pp. 1805-1829.
- <sup>13</sup> Liew, C. K. and Veidt, M., "Pattern Recognition of Guided Waves for Damage Evaluation in Bars," *Pattern Recognition Letters*, Vol. 30, 2009, pp. 321-330.
- <sup>14</sup> Grall, A., Bérenguer, C., and Dieulle, L., "A Condition-Based Maintenance Policy for Stochastically Deteriorating Systems," *Reliability Engineering and System Safety*, Vol. 76, No. 2, May 2002, pp. 167-180.
- <sup>15</sup> Yam, R. C. M., Tse, P. W., Li, L., and Tu, P., "Intelligent Predictive Decision Support System for Condition-Based Maintenance," *International Journal of Advanced Manufacturing Technology*, Vol. 17, No. 5, 2001, pp. 383-391.
- <sup>16</sup> Si, X. S., Wang, W., Hu, C. H., Zhou, D. H., "Remaining Useful Life Estimation – A Review on the Statistical Data Driven Approaches," *European Journal of Operational Research*, Vol. 213, 2011, pp. 1-14.
- <sup>17</sup> Srivastava, A. N. and Das, S., "Detection and Prognostics on Low-Dimensional Systems," *IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 39, No. 1, January 2009, pp. 44-54.
- <sup>18</sup> Goebel, K., Saha, B., and Saxena, A., "A Comparison of Three Data-Driven Techniques for Prognostics," *Proceedings of the 62nd Meeting of the Society For Machinery Failure Prevention Technology (MFPT)*, pp. 119-131, Virginia Beach, VA, May, 2008.
- <sup>19</sup> Saha, B., Goebel, K., and Christophersen, J., "Comparison of Prognostic Algorithms for Estimating Remaining Useful Life of Batteries," *Transactions of the Institute of Measurement & Control*, Vol. 31, No. 3-4, June 2009, pp. 293-308.
- <sup>20</sup> Xing, Y., Williard, N., Tsui, K.-L., and Pecht, M., "A Comparative Review of Prognostics-Based Reliability Methods for Lithium Batteries," *Prognostics and System Health Management Conference*, Shenzhen, China, May 24-25, 2011.
- <sup>21</sup> Zhang, J. and Lee, J., "A Review on Prognostics and Health Monitoring of Li-Ion Battery," *Journal of Power Sources*, Vol. 196, 2011, pp. 6007-6014.
- <sup>22</sup> 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 Health Monitoring*, Vol. 2, 2007, pp. 1108-1115.
- <sup>23</sup> Sankavaram, C., Pattipati, B., Kodali, A., Pattipati, K., Azam, M., Kumar, S., and Pecht, M., "Model-based and Data-driven Prognosis of Automotive and Electronic Systems," *5th Annual IEEE Conference on Automation Science and Engineering*, Bangalore, India, August 22-25, 2009.
- <sup>24</sup> Cheng, S. and Pecht, M., "A Fusion Prognostics Method for Remaining Useful Life Prediction of Electronic Products," *5th Annual IEEE Conference on Automation Science and Engineering*, Bangalore, India, August 22-25, 2009.
- <sup>25</sup> Xu, J. and Xu, L., "Health Management Based on Fusion Prognostics for Avionics Systems," *Journal of Systems Engineering and Electronics*, Vol. 22, No. 3, June 2011, pp. 428-436.
- <sup>26</sup> Xing, Y., Miao, Q., Tsui, K.-L., and Pecht, M., "Prognostics and Health Monitoring for Lithium-ion Battery," *2011 IEEE International Conference*, pp. 242-247.

- <sup>27</sup> 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.
- <sup>28</sup> Krogh, A., "What are Artificial Neural Networks?," *Nature Biotechnology*, Vol. 26, No. 2, February 2008, pp. 195-197.
- <sup>29</sup> Yao, X., "Evolving Artificial Neural Networks," *Proceedings of the IEEE*, Vol. 87, No. 9, September 1999, pp. 1423-1447.
- <sup>30</sup> Zio, E. and Maio, F. D., "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, 2010, pp. 49-57.
- <sup>31</sup> Gouriveau, R., Dragomir, O., and Zerhouni, N., "A Fuzzy Approach of Online Reliability Modeling and Estimation," *Proceedings of Supporting Technologies for Advanced Maintenance Information Management*, San Sebastian, Spain: European Safety Reliability & Data Association; 2008. p. 1-12.
- <sup>32</sup> Pandey, M. D. and Noortwijk, J. M. V., "Gamma Process Model for Time-Dependent Structural Reliability Analysis, In: Watanabe E, Frangopol DM, Utsonomiya T, editors. Bridge Maintenance, Safety, Management and Cost." *Proceedings of the Second International Conference on Bridge Maintenance, Safety and Management (IABMAS)*, Kyoto, Japan, 18-22 October 2004. London: Taylor and Francis Group, 2004.
- <sup>33</sup> Dickson, D. C. M. and Waters, H. R., "Gamma Processes and Finite Time Survival Probabilities," *Astin Bulletin*. Vol. 23, No. 2, 1993, pp. 259-272.
- <sup>34</sup> Rabiner, L. R., "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceedings of the IEEE*, Vol. 77, No. 2, February, 1989, pp. 257-286.
- <sup>35</sup> Mackay, D. J. C., "Introduction to Gaussian Processes," *Technical report*, Cambridge University, UK, 1997. URL: <http://www.cs.utoronto.ca/~mackay/gpB.pdf>
- <sup>36</sup> Seeger, M., "Gaussian Processes for Machine Learning," *International Journal of Neural Systems*, Vol. 14, No. 2, 2004, pp. 69-106.
- <sup>37</sup> Tipping, M. E., "Sparse Bayesian Learning and the Relevance Vector Machine," *Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211-244.
- <sup>38</sup> 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.
- <sup>39</sup> Bretscher, O., *Linear Algebra with Applications*, 3<sup>rd</sup> ed., Prentice Hall, Upper Saddle River NJ, 1995.
- <sup>40</sup> 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.
- <sup>41</sup> Duch, W. and Jankowski, N., "Survey of Neural Transfer Functions," *Neural Computing Surveys*, Vol. 2, 1999, pp. 163-212.
- <sup>42</sup> 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.
- <sup>43</sup> 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.
- <sup>44</sup> Rumelhart, D. E., Hinton, G. E., and Williams, R. J., "Learning Internal Representations by Error Propagation," *In Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1: foundations, MIT Press, 1986, pp. 318-362.
- <sup>45</sup> Bodén, M., "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, 2002.
- <sup>46</sup> Wilamowski, B. M., Iplikci, S., Kaynak, O., and Efe, M. Ö., "An Algorithm for Fast Convergence in Training Neural Networks," *Proceedings of the International Joint Conference on Neural Networks*, Vol. 3, 2001, pp. 1778-1782.
- <sup>47</sup> Liu, P. and Li, H., *Fuzzy Neural Network Theory and Application*, World Scientific, Singapore, 2004.
- <sup>48</sup> He, Y., Tan, Y., and Sun, Y., "Wavelet Neural Network Approach for Fault Diagnosis of Analogue Circuits," *IEE Proceedings-Circuits, Devices and Systems*, Vol. 151, No. 4, 2004, pp. 379-384.
- <sup>49</sup> Bicciano, S., Pandin, M., Didonè, G., and Bello, C. D., "Analysis of an Associative Memory Neural Network for Pattern Identification in Gene Expression Data," *Biokdd01, Workshop on Data Mining in Bioinformatics*, pp. 22-30.
- <sup>50</sup> Happel, B. L. M. and Murre, J. M. J., "The Design and Evolution of Modular Neural Network Architectures," *Neural Networks*, Vol. 7, 1994, pp. 985-1004.
- <sup>51</sup> Psychogios, D. C. and Ungar, L. H., "A Hybrid Neural Network-First Principles Approach to Process Modeling," *AICHE Journal*, Vol. 38, No. 10, October, 1992, pp. 1499-1511.
- <sup>52</sup> Rovithakis, G. A., Maniadakis, M., and Zervakis, M., "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, No. 1, February, 2004, pp. 695-702.
- <sup>53</sup> Lawrence, S., Giles, C. L., Tsoi, A. C., "What Size Neural Network Gives Optimal Generalization? Convergence Properties of Backpropagation," *Technical Reports*, UM Computer Science Department, UMIACS, October 15, 1998.
- <sup>54</sup> Doukim, C. A., Dargham, J. A., and Chekima, A., "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*, Kuala Lumpur, Malaysia, 10-13 May, 2010.
- <sup>55</sup> Gómez, I., Franco, L., and Jérez, J. M., "Neural Network Architecture Selection: Can Function Complexity Help?," *Neural Processing Letters*, Vol. 30, 2009, pp. 71-87.

- <sup>56</sup> Ostafe, D., "Neural Network Hidden Layer Number Determination Using Pattern Recognition Techniques," *2nd Romanian-Hungarian Joint Symposium on Applied computational Intelligence*, Timisoara, Romania, 2005.
- <sup>57</sup> Chang, J. F. and Hsieh, P. Y., "Particle Swarm Optimization Based on Back Propagation Network Forecasting Exchange Rates," *International Journal of Innovative Computing, Information and Control*, Vol. 7, No. 12, December 2011, pp. 6837-6847.
- <sup>58</sup> Salomon, R. and Hemmen, J. L. V., "Accelerating Backpropagation through Dynamic Self-Adaptation," *Neural Networks*, Vol. 9, No. 4, 1996, pp. 589-601.
- <sup>59</sup> Chen, S. C., Lin, S. W., Tseng, T. Y., and Lin, H. C., "Optimization of Back-Propagation Network Using Simulated Annealing Approach," *IEEE International Conference on Systems, Man, and Cybernetics*, Taipei, Taiwan, October 8-11, 2006.
- <sup>60</sup> Subudhi, B. Jena, D., and Gupta, M. M., "Memetic Differential Evolution Trained Neural Networks for Nonlinear System Identification," *IEEE Region 10 Colloquium and the Third International Conference on Industrial and Information Systems*, Kharagpur, India, December 8 -10, 2008.
- <sup>61</sup> Nawi, N. M., Ransing, R. S., and Ransing, M. R., "An Improved Conjugate Gradient Based Learning Algorithm for Back Propagation Neural Networks," *International Journal of Computational Intelligence*, Vol. 4, No. 1, pp. 46-55.
- <sup>62</sup> Navone, H. D., Granitto, P. M., Verdes, P. F., and Ceccatto, H. A., "A Learning Algorithm for Neural Network Ensembles," *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial*. No. 12, 2001, pp 70-74.
- <sup>63</sup> Jacobs, R. A., "Methods for Combining Experts' Probability Assessments," *Neural Computation*, Vol. 7, No. 5, 1995, pp. 867-888.
- <sup>64</sup> Drucker, H., Cortes, C., Jackel, L. D., LeCun, Y., and Vapnik, V., "Boosting and Other Ensemble Methods," *Neural Computation*, Vol. 6, No. 6, 1994, pp. 1289-1301.
- <sup>65</sup> Krogh, A. and Vedelsby, J., "Neural Network Ensembles, Cross Validation, and Active Learning," in G. Tesauro, D. Touretzky and T. Leen, eds, *Advances in Neural Information Processing Systems*, Vol. 7, The MIT Press, 1995, pp. 231-238.
- <sup>66</sup> Perrone, M. P. and Cooper, L. N., "When Networks Disagree: Ensemble Method for Neural Networks," in R. J. Mammone, ed., *Neural Networks for Speech and Image processing*, Chapman-Hall, 1993.
- <sup>67</sup> Naftaly, U., Intrator, N., and Horn, D., "Optimal Ensemble Averaging of Neural Networks," *Network: Computation in Neural Systems*, Vol. 8, No. 3, 1997, pp. 283-296.
- <sup>68</sup> Chryssolouris, 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.
- <sup>69</sup> 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.
- <sup>70</sup> 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.
- <sup>71</sup> 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.
- <sup>72</sup> Efron, B. and Tibshirani, R. J., *An Introduction to the Bootstrap*, Chapman & Hall/CRC., FL, 1994.
- <sup>73</sup> Specht, D. F., "Probabilistic Neural Networks," *Neural Networks*, Vol. 3, 1990, pp. 109-118.
- <sup>74</sup> Parzen, E., "On Estimation of a Probability Density Function and Mode," *Annals of Mathematical Statistics*, Vol. 33, No. 3, 1962, pp. 1065-1076.
- <sup>75</sup> Petalas, P., Spyridonos, P., Glotsos, D., Cavouras, D., Ravazoula, P., and Nikiforidis, G., "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*, 2003, pp. 1024-1027.
- <sup>76</sup> Giurgiutiu V., "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*, San Diego, CA, March 17-21, 2002.
- <sup>77</sup> Mao, K. Z., Tan, K.-C., and Ser, W., "Probabilistic Neural-Network Structure Determination for Pattern Classification," *IEEE Transactions on Neural Networks*, Vol. 11, No. 4, July, 2000, pp. 1009-1016.
- <sup>78</sup> Khawaja, T., Vachtsevanos, G., and Wu, B., "Reasoning about Uncertainty in Prognosis: A Confidence Prediction Neural Network Approach," *2005 Annual Meeting of the North American Fuzzy Information Processing Society*, Ann Arbor, Michigan, June 22-25, 2005.
- <sup>79</sup> Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., *Bayesian Data Analysis*, Chapman and Hall, NY, 2004.
- <sup>80</sup> Santner, T. J., Williams, B. J., and Notz, W. I., *The Design and Analysis of Computer Experiments*, Springer Verlag, NY, 2003.
- <sup>81</sup> Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA, 2006.
- <sup>82</sup> Williams, C. K. I., "Computing with infinite networks," in M. C. Mozer, M. I. Jordan, and T. Petsche, eds., *Advances in Neural Information Processing Systems*, Vol. 9, MIT Press, Cambridge, MA, 1997.
- <sup>83</sup> Mohanty, S., Das, D., Chattopadhyay, A., and Peralta, P., "Gaussian Process Time Series Model for Life Prognosis of Metallic Structures," *Journal of Intelligent Material Systems and Structures*, Vol. 20, 2009, pp. 887-896.
- <sup>84</sup> Paciorek, C. and Schervish, M. J., "Nonstationary Covariance Functions for Gaussian Process Regression," in S. Thrun, L. Saul, and B. Schölkopf, eds., *Advances in Neural Information Processing Systems*, Vol. 16, MIT Press, 2004.
- <sup>85</sup> Belhouari, S. B., Bermak, A., "Gaussian Process for Nonstationary Time Series Prediction," *Computational Statistics & Data Analysis*, Vol. 47, 2004, pp. 705-712.

- <sup>86</sup> Liu, D., Pang, J., Zhou, J., and Peng, Y., "Data-driven Prognostics for Lithium-ion Battery Based on Gaussian Process Regression," *2012 Prognostics and System Health Management Conference*, Beijing, China, May 23-25, 2012.
- <sup>87</sup> Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409-423.
- <sup>88</sup> An, D. and Choi, J. H., "Efficient Reliability Analysis Based on Bayesian Framework under Input Variable and Metamodel Uncertainties," *Structural and Multidisciplinary Optimization*, Vol. 46, 2012, pp. 533-547.
- <sup>89</sup> Neal, R. M., "Regression and Classification Using Gaussian Process Priors," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, eds., *Bayesian Statistics*, Vol. 6, Oxford University Press, NY, 1998, pp. 475-501.
- <sup>90</sup> MacKay, D. J. C., "Gaussian Processes-A Replacement for Supervised Neural Networks?," *Tutorial lecture notes for NIPS*, UK, <http://www.inference.phy.cam.ac.uk/mackay/BayesGP.html>, 1997.
- <sup>91</sup> Lawrence, N., Seeger, M., and Herbrich, R., "Fast Sparse Gaussian Process Methods: The Information Vector Machine," in S. Becker, S. Thrun, and K. Obermayer, eds., *Advances in Neural Information Processing Systems*, Vol. 15, MIT Press, 2003, pp. 625-632.
- <sup>92</sup> Smola, A. J. and Bartlett, P. L., "Sparse Greedy Gaussian Process Regression," in T. K. Leen, T. G. Dietterich, and V. Tresp, eds., *Advances in Neural Information Processing Systems*, Vol. 13, MIT Press, 2001, pp. 619-625.
- <sup>93</sup> Foster, L., Waagen, A., Aijaz, N., Hurley, M., Luis, A., Rinsky, J., Satyavolu, C., Way, M. J., Gazis, P., and Srivastava, A., "Stable and Efficient Gaussian Process Calculations," *Journal of Machine Learning Research*, Vol. 10, 2009, pp. 857-882.
- <sup>94</sup> Melkumyan, A. and Ramos, F., "A Sparse Covariance Function for Exact Gaussian Process Inference in Large Datasets," *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence*, Pasadena, CA., July 11-17, 2009, pp. 1936-1942.
- <sup>95</sup> Rebba, R., Huang, S., Liu, Y., and Mahadevan, S., "Statistical Validation of Simulation Models," *International Journal of Materials and Product Technology*, Vol. 25, No. 1/2/3, 2006, pp. 164-181.
- <sup>96</sup> Rebba, R., Mahadevan, S., Huang, S., "Validation and Error Estimation of Computational Models," *Reliability Engineering and System Safety*, Vol. 91, 2006, pp. 1390-1397.
- <sup>97</sup> Kleijnen, J. P. C., "Statistical Validation of Simulation Models," *European Journal of Operational Research*, Vol. 87, 1995, pp. 21-34.
- <sup>98</sup> Sargent, R. G., "Verification and Validation of Simulation Models," *Proceedings of the 2009 Winter Simulation Conference*, Austin, TX, December 13-16, 2009.
- <sup>99</sup> Coppe, A., Pais, M. J., Haftka, R. T., and Kim, N. H., "Remarks on Using a Simple Crack Growth Model in Predicting Remaining Useful Life," *Journal of Aircraft*, accepted, 2012.
- <sup>100</sup> Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transaction of the ASME-Journal of Basic Engineering*, No. 82, March 1960, pp. 35-45.
- <sup>101</sup> Julier, S. J., Uhlmann, J. K., "Unscented Filtering and Nonlinear Estimation," *Proceedings of the IEEE*, Vol. 92, No. 3, 2004, pp. 401-422.
- <sup>102</sup> Doucet, A., De Freitas, N., and Gordon, N. J., *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.
- <sup>103</sup> Kramer, S. C., and Sorenson, H. W., "Bayesian Parameter Estimation," *IEEE Transactions on Automatic Control*, Vol. 33, No. 2, pp. 217-222.
- <sup>104</sup> 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.
- <sup>105</sup> DeCastro, J. A., Tang, L., Loparo, K. A., Goebel, K., Vachtsevanos, G., "Exact Nonlinear Filtering and Prediction in Process Model-Based Prognostics," *Annual Conference of the Prognostics and Health Management Society*, San Diego, CA, September 27-October 1, 2009.
- <sup>106</sup> Orchard, M. E., Vachtsevanos, G. J., "A Particle Filtering Approach for On-Line Failure Prognosis in a Planetary Carrier Plate," *International Journal of Fuzzy Logic and Intelligent Systems*, Vol. 7, No. 4, 2007, pp. 221-227.
- <sup>107</sup> Daigle, M., Goebel, K., "Multiple Damage Progression Paths in Model-Based Prognostics," *IEEE Aerospace Conference*, Big Sky, Montana, 2011.
- <sup>108</sup> An, D., Choi, J. H., Schmitz, T. L., and Kim, N. H., "In-Situ Monitoring and Prediction of Progressive Joint Wear Using Bayesian Statistics," *Wear*, Vol. 270, No. 11-12, May, 2011, pp. 828-838.
- <sup>109</sup> Higuchi, T., "Monte Carlo Filter Using the Genetic Algorithm Operators," *Journal of Statistical Computation and Simulation*, Vol. 59, No. 1, 1997, pp. 1-23.
- <sup>110</sup> Kitagawa, G., "Non-Gaussian State Space Modeling of Nonstationary Time Series (with Discussion)," *Journal of the American Statistical Association*, Vol. 82, No. 400, 1987, pp. 1032-1063.
- <sup>111</sup> Wang, W. P., Liao, S., Xing, T. W., "Particle Filter for State and Parameter Estimation in Passive Ranging," *IEEE International Conference on Intelligent Computing and Intelligent Systems*. Shanghai, China, 2009.
- <sup>112</sup> Gilks, W. R., Berzuini, C., "Following a Moving Target-Monte Carlo Inference for Dynamic Bayesian Models," *Royal Statistical Society B*, Vol. 63, Part 1, 2001, pp. 127-146.
- <sup>113</sup> Andrieu, C., Freitas, de N., Doucet, A., and Jordan, M., "An Introduction to MCMC for Machine Learning," *Machine Learning*, Vol. 50, No. 1, 2003, pp. 5-43.
- <sup>114</sup> Kim, S., Park, J. S., "Sequential Monte Carlo Filters for Abruptly Changing State Estimation," *Probabilist Engineering Mechanics*, Vol. 26, 2011, pp. 194-201.
- <sup>115</sup> Rubin, D. B., "Using the SIR Algorithm to Simulate Posterior Distributions," in J. M. Bernardo, M. H. DeGroot, D. V. Lindley, and A. F. M. Smith, eds., *Bayesian statistics*, Vol. 3, Oxford University Press, Cambridge, MA, 1998, pp. 395-402.

- <sup>116</sup> An, D. and Choi, J. H., "Improved MCMC Method for Parameter Estimation Based on Marginal Probability Density Function," *Journal of Mechanical Science and Technology*, in press, 2013.
- <sup>117</sup> Gelfand, A. E. and Sahu, S. K., "On Markov Chain Monte Carlo Acceleration," *Journal of Computational and Graphical Statistics*, Vol. 3, 1994, pp. 261-276.
- <sup>118</sup> An, D., Choi, J. H., and Kim, N. H., "Identification of Correlated Damage Parameters under Noise and Bias Using Bayesian Inference," *Structural Health Monitoring*, Vol. 11, No. 3, May, 2012, pp. 293-303.
- <sup>119</sup> Gu, J., Azarian, M. H., and Pecht, M. G., "Failure Prognostics of Multilayer Ceramic Capacitors in Temperature-Humidity-Bias Conditions," *2008 International Conference on Prognostics and Health Management*, Denver, Colorado, October 6-9, 2008.
- <sup>120</sup> Coppe, A., Haftka, R. T., Kim, N. H., and Yuan, F. G., "Reducing Uncertainty in Damage Growth Properties by Structural Health Monitoring," *Annual Conference of the Prognostics and Health Management Society*, San Diego, CA, September 27-October 1, 2009.
- <sup>121</sup> Guan, X., Liu, Y., Saxena, A., Celaya, J., and Goebel, K., "Entropy-Based Probabilistic Fatigue Damage Prognosis and Algorithmic Performance Comparison," *Annual Conference of the Prognostics and Health Management Society*, San Diego, CA, September 27-October 1, 2009.
- <sup>122</sup> Zhang, B., Khawaja, T., Patrick, R., Vachtsevanos, G., Orchard, M. E., and Saxena, A., "Application of Blind Deconvolution Denoising in Failure Prognosis," *IEEE Transactions on Instrumentation and Measurement*, Vol. 58, No. 2, February 2009, pp. 303-310.
- <sup>123</sup> Qiu, H., Lee, J., Linc, J., Yu, G., "Robust Performance Degradation Assessment Methods for Enhanced Rolling Element Bearing Prognostics," *Advanced Engineering Informatics*, Vol. 17, 2003, pp. 127-140.
- <sup>124</sup> Abouel-seoud, S. A., Elmorsy, M. S., and Dyab, E. S., "Robust Prognostics Concept for Gearbox with Artificially Induced Gear Crack Utilizing Acoustic Emission," *Energy and Environment Research*, Vol. 1, No. 1, December, 2011, pp. 81-93.
- <sup>125</sup> 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.
- <sup>126</sup> 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.
- <sup>127</sup> An, D., Choi, J. H., and Kim, N. H., "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*, Honolulu, Hawaii, April 23-26, 2012.
- <sup>128</sup> Storvik, G., "Particle Filters in State Space Models with the Presence of Unknown Static Parameters," *IEEE Transactions on Signal Processing*, Vol. 50, No. 2, May, 2002, pp. 281-289.