A Comparison Study of Methods for Parameter Estimation in the Physics-based Prognostics

Dawn An[1] and Joo-Ho Choi[2]

Korea Aerospace University, Goyang-City, Gyeonggi-do, 412-791, Korea

and

Nam Ho Kim[3]

University of Florida, Gainesville, FL, 32611

Prediction of remaining useful life of a system is important for safety and maintenance scheduling. In the physics-based prognostics, the accuracy of predicted remaining useful life is directly related to that of estimated model parameters. It, however, is not a simple task to estimate the model parameters because most real systems have multivariate model parameters, which are often correlated each other. This paper mainly discusses the difference in estimating model parameters among different prognostics methods: the particle filter method, the overall Bayesian method, and the incremental Bayesian method. These methods are based on the same theoretical foundation, Bayesian inference, but they are different from each other in the sampling scheme and/or uncertainty analysis process. A simple analytical example and the Paris model for crack growth are used to demonstrate the difference among the three methods in terms of prognostics metrics. The numerical results show that the particle filter and overall Bayesian methods outperform the incremental Bayesian method. Even though the particle filter shows slightly better results in terms of prognostics metrics, the overall Bayesian method is efficient when batch data exist.

I. Introduction

STRUCTURAL health monitoring (SHM) facilitates condition-based maintenance that provides a safe and cost-effective strategy by predicting the level of degradation or damage without intrusive and time-consuming inspections.[1] Since SHM systems can assess damage frequently, they can also be used to predict the future behavior of damage, which is important for safety and maintenance schedule management. SHM systems can have a significant impact on increasing safety by allowing evaluation of the system's health status and prediction of the remaining useful life (RUL), which is called prognostics.

In general, prognostics methods can be categorized into data-driven,[2] physics-based,[3] and hybrid[4] approaches. The data-driven method that does not use any particular physical model is powerful in predicting near-future behaviors using previously trained data, while the physics-based method has advantages in predicting long-term behaviors of the system by identifying model parameters. Since fatigue failures slowly progresses, the physics-based method fits better than the data-driven method. In the physics-based method, model parameter estimation has a great effect on evaluating the system's health status and predicting the RUL.

The objective of this paper is to compare different methods in estimating model parameters and predicting RUL: the particle filter (PF), the overall Bayesian method (OBM), and the incremental Bayesian method (IBM). These

[1] Graduate student, Department of Aerospace & Mechanical Engineering, Korea Aerospace University, 100 Hwajeon-dong, Deokyang-gu, Goyang-City, Gyeonggi-do, Korea, 412-791.

[2] Professor, School of Aerospace & Mechanical Engineering, Korea Aerospace University, 100 Hwajeon-dong, Deokyang-gu, Goyang-City, Gyeonggi-do, Korea, 412-791

[3] Associate professor, Mechanical and Aerospace Engineering department, University of Florida, Gainesville, FL, 32611, AIAA Member.

methods are based on the Bayesian inference in which the uncertainty in model parameters can be reduced using SHM data, which makes the estimated RUL more reliable. Each method is introduced in the following paragraphs.
PF is also known as a sequential Monte Carlo method[5,6] for sequentially updating a time-dependent system model based on Bayesian inference, where the posterior distribution is expressed as a number of particles and their weights. PF has been widely employed to estimate system states and model parameters.[5-9] Although many efforts have been directed to improve PF,[10-13] only sequential important resampling (SIR) is considered in this paper as it is the most commonly used algorithm.

OBM is to estimate the unknown parameters using a batch of measurement data. Samples of the updated posterior distribution of the model parameters based on Bayesian inference are drawn from a viable sampling method, such as the Markov Chain Monte Carlo (MCMC) simulation.[14-19] IBM is similar to OBM, but instead of using the final updated distribution, the update is performed incrementally with each measurement data.

It, however, should be noted that the three methods are based on the same theoretical background, the Bayesian estimation technique. Despite the same theoretical background, the results from these methods are different from each other. The reason is that the PF and IBM predict the time evolution of the growing damage state, while the OBM is based on the given entire set of data, not the time evolution. Another important reason is from the difference in the sampling scheme: resampling for PF and MCMC for OBM and IBM.

The purpose of this paper is to discuss the difference between the three methods, which has not been founded yet. The methods are compared by using a simple analytical example as well as the Paris model[20] that describes crack growth in a plate under mode I loading, and are evaluated by using established prognostics metrics.[21] The paper is organized as follows. In Section 2, PF, OBM, and IBM are introduced with a simple mathematical example. In Section 3, the three methods are compared by applying crack growth problem. In Section 4, the differences are discussed, and conclusions are presented in Section 5.

II. Three Methods for Bayesian-based Prognostics

Bayesian inference is a statistical method in which observations are used to estimate unknown model parameters or system states. The unknown parameters are represented as a probability density function (PDF), which is updated with observed data. Bayesian inference is based on the following Bayes' theorem:[22]

$$p(\boldsymbol{\Theta} \mid \mathbf{z}) \propto L(\mathbf{z} \mid \boldsymbol{\Theta})\, p(\boldsymbol{\Theta}) \tag{1}$$

where $\boldsymbol{\Theta}$ is a vector of model parameters or system states, $\mathbf{z}$ is a vector of observed data, $L(\mathbf{z} \mid \boldsymbol{\Theta})$ is the likelihood or the PDF value of $\mathbf{z}$ conditional on the given $\boldsymbol{\Theta}$, $p(\boldsymbol{\Theta})$ is the prior PDF of $\boldsymbol{\Theta}$, and $p(\boldsymbol{\Theta} \mid \mathbf{z})$ is the posterior PDF of $\boldsymbol{\Theta}$ conditional on $\mathbf{z}$. In the following subsections, three Bayesian-based methods, the particle filter (PF), the overall Bayesian method (OBM), and the incremental Bayesian method (IBM), are explained in a general way.

A. Particle Filter (PF)

The particle filter (PF, a.k.a. sequential Monte Carlo method[5,6]) is a method of estimating and sequentially updating a time-dependent system model based on Bayesian inference, where the posterior distribution is expressed as a number of particles and their weights. In the case of a linear system, the exact posterior distribution with a Gaussian noise can be obtained analytically using the Kalman filter[23] though the convergence of KF largely depends on the initial condition of the parameter and variance of the parameter. On the other hand, the posterior distribution of a non-linear system with non-Gaussian noise cannot be obtained exactly. In such a case, there are several suboptimal filters such as the extended Kalman filter that requires linearization,[24] the multiple model filter that performs the state estimation algorithm for multiple degradation models,[25] the unscented Kalman filter,[26] and PF[5-9] that are based on the sampling theory. This paper focuses on the original PF because it contains fundamental characteristics of other filtering methods.

It would be better to explain the importance sampling method[27] first to understand PF. In sampling-based methods, a number of samples can be used to approximate the posterior distribution. In order to approximate the distribution better with a limited number of samples, the importance sampling method assigns a weight to each sample (or particle) in proportion to an arbitrarily chosen importance distribution; therefore, the quality of estimation depends on the selected importance distribution. The weight is expressed as

$$w(\boldsymbol{\Theta}^i) = \frac{\pi(\boldsymbol{\Theta}^i)}{q(\boldsymbol{\Theta}^i)} = \frac{L(z \mid \boldsymbol{\Theta}^i)\, p(\boldsymbol{\Theta}^i)}{q(\boldsymbol{\Theta}^i)} \tag{2}$$

where $\pi\left(\mathbf{\Theta}^i\right)$ and $q\left(\mathbf{\Theta}^i\right)$ are i th particle's PDF value of the posterior distribution and an arbitrarily chosen importance distribution, respectively. From the viewpoint of Bayes' theorem, it is possible to use the prior distribution as an importance distribution because it is already available and close to the posterior distribution. Then, the Eq. (2) is reduced to the likelihood function by substituting the prior $p\left(\mathbf{\Theta}^i\right)$ for $q\left(\mathbf{\Theta}^i\right)$; this is called the Condensation (CONditional DENSity propagATION) algorithm, which is employed in this paper.

PF can be considered as a sequential importance sampling, which continuously performs the importance sampling whenever a new observed data is available. In addition, a resampling process, called sequential importance resampling (SIR), is required to solve the degeneracy phenomenon that decreases the accuracy in the posterior distribution. In SIR, those particles with a very small weight are eliminated, while those particles with a high weight are duplicated.

Conventionally, PF has been used to estimate the system state using measurement data given model parameters, which are obtained from laboratory tests.[5,8] However, the parameters from laboratory tests can be different from those in service due to different conditions. In such a case, PF can be used to estimate both the system state and model parameters; detailed procedure is as follows:[7,8]

* Given information:
  - State transition function f and measurement function h:

$$x_k = f\left(x_{k-1}, \boldsymbol{\theta}_k, \nu_k\right) \tag{3}$$

$$z_k = h\left(x_k, \omega_k\right) \tag{4}$$

where k is the time step index, x_k is the system state, $\boldsymbol{\theta}_k$ is a vector of model parameters, z_k is measurement data. The measurement function is needed when the system state cannot be measured directly. ν_k and ω_k are, respectively, process and measurement noise; in the case of Gaussian distribution, the noises are represented as $\nu_k \sim N\left(0, \sigma_p\right)$ and $\omega_k \sim N\left(0, \sigma_m\right)$.

1) Assume initial distribution of x_0 and $\boldsymbol{\theta}_0$. Set $k = 1$.
2) **Prediction**- First, generate N particles of $\boldsymbol{\theta}_k$ from $p\left(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}\right)$, which means a predicted distribution of $\boldsymbol{\theta}_k$ conditional on the previous distribution, $p\left(\boldsymbol{\theta}_{k-1}\right)$. It, however, should be noted that model parameters inherently do not depend on time evolution. Therefore, N particles of $\boldsymbol{\theta}_k$ become the same as the previous particles. In other words, there is no transition between $\boldsymbol{\theta}_{k-1}$ and $\boldsymbol{\theta}_k$, only update process is required. Next, sample N particles of x_k based on $p\left(x_k \mid x_{k-1}, \boldsymbol{\theta}_k\right)$, that is, N particles of each variables; process noise generated from $N\left(0, \sigma_p\right)$ as well as x_{k-1} and $\boldsymbol{\theta}_k$ are used in Eq. (3) to obtain N particles of x_k.
3) **Update**- Calculate weight for each particle in Step 2. In the case that measurement noise is normally distributed with zero mean and standard deviation, σ_m, the likelihood function and weight become as follows:

$$L\left(z_k \mid x_k^i, \boldsymbol{\theta}_k^i\right) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{1}{2}\frac{\left(z_k - h\left(x_k^i\right)\right)^2}{{\sigma_m}^2}\right], \tag{5}$$

$$w_k^i = \frac{L\left(z_k \mid x_k^i, \boldsymbol{\theta}_k^i\right)}{\sum_{j=1}^{N} L\left(z_k \mid x_k^j, \boldsymbol{\theta}_k^j\right)} \tag{6}$$

4) **Resample**- Resample the particles proportional to their weights. Several methods can be used for this purpose, and one thing among them is to repeat N times with the following way; (1) construct CDF from the likelihood function in Eq. (5). In other words, CDF is based on the weights in Eq. (6) which corresponds to PDF of z_k conditional on $x_k^i, \boldsymbol{\theta}_k^i$. (2) find particles of $x_k, \boldsymbol{\theta}_k$ which make CDF of z_k be the same value as (or the closest to) randomly chosen value from $U\left(0,1\right)$ which is the uniform distribution in the interval of $\left[0,1\right]$. By repeating N times this process, N samples of $x_k, \boldsymbol{\theta}_k$ are obtained, which represents an

approximation of the posterior distribution $p(x_k, \boldsymbol{\theta}_k \mid z_{0:k})$ obtained based on the weighted samples $\left(x_k^i, \boldsymbol{\theta}_k^i \mid w_k^i\right)$.

5) **Prognosis**- The system state x_{k+l} where t_{k+l} is the future time can be predicted based on the transition function and estimated model parameters until t_k. There are no new data and update process after t_k. Therefore, the weigh after t_k are kept unchanged with $w_k^i = 1/N$ since every particle is assigned to the same weight after resampling.

The process of PF with one parameter is illustrated in Figure 1. The length of vertical bars represents magnitude of the weight, and the particles in the prediction step are assigned to the weight at the update step and duplicated as much as the magnitude of the weight at the resampling step.

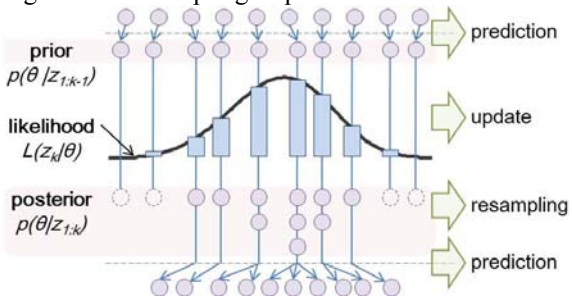


Figure 1. Illustration of the PF process with resampling

As a simple example, the following linear system is considered.

$$x_k = x_{k-1} + \theta_k \Delta t + v_k \tag{7}$$

where θ_k is a model parameter to be estimated, process noise v_k is normally distributed with zero mean and standard deviation σ_p=1. In this example, it is assumed that both the initial distributions of the system state and model parameter are uniform in the interval of $[-25,50]$. Let the measurement function $z_k = x_k + \omega_k$, where measurement noise ω_k is normally distributed with zero mean and standard deviation σ_m=10. Then, the likelihood function to calculate the weight becomes as follows:

$$L\left(z_k \mid x_k^i, \theta_k^i\right) = \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{1}{2}\frac{\left(z_k - x_k^i\right)^2}{\sigma_m^{\,2}}\right] \tag{8}$$

The results of estimated model parameter and system state with Δt=1 are shown in the Figure 2. N=5,000 particles (which is applied for all problems in this paper) are used with the data observed until $t = 20 (\text{or } k = 20)$. The estimated parameter converges to the true value ($\theta_{\text{true}} = 2$) with small confidence intervals. In addition, the prediction result of the system state is also fairly accurate.

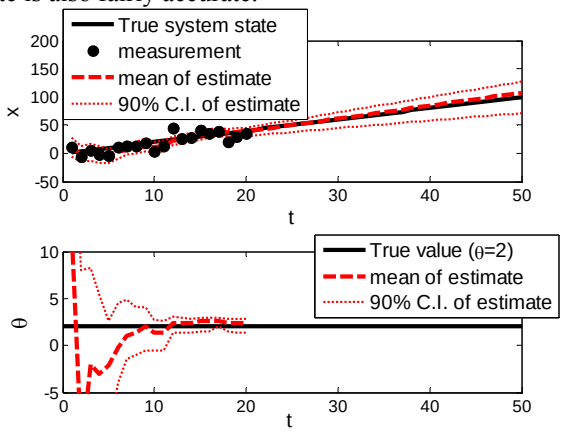


Figure 2. The results from PF in the simple example

B. Overall Bayesian Method (OBM)

The overall Bayesian method (OBM) is to estimate the unknown parameters using the measurement data in which the final posterior distribution is expressed as an equation. The posterior distribution is obtained by multiplying all the likelihood functions given by n number of data and the prior as

$$p(\boldsymbol{\theta} \mid \mathbf{z}) \propto L(z_1 \mid \boldsymbol{\theta}) \times L(z_2 \mid \boldsymbol{\theta}) \times \cdots \times L(z_n \mid \boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \tag{9}$$

Using Eq. (9), a sampling method, such as the Markov Chain Monte Carlo (MCMC) method,[19] can be used to construct the posterior distribution. The Metropolis-Hastings (M-H) algorithm is a typical method of MCMC, and the procedure is summarized in the following pseudo-code:

$$\begin{aligned}
&1.\ \text{Initialize } \boldsymbol{\theta}^0 \\
&2.\ \text{For } i = 1 \text{ to } N \\
&\quad -\ \text{Sample } u \sim U(0,1) \\
&\quad -\ \text{Sample } \boldsymbol{\theta}^* \sim q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{i-1}\right) \\
&\quad -\ \text{if } u < A\left(\boldsymbol{\theta}^{i-1}, \boldsymbol{\theta}^*\right) = \min\left\{1, \frac{p\left(\boldsymbol{\theta}^* \mid \mathbf{z}\right) q\left(\boldsymbol{\theta}^{i-1} \mid \boldsymbol{\theta}^*\right)}{p\left(\boldsymbol{\theta}^{i-1} \mid \mathbf{z}\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{i-1}\right)}\right\} \\
&\qquad\qquad \boldsymbol{\theta}^i = \boldsymbol{\theta}^* \\
&\quad \text{else} \\
&\qquad\qquad \boldsymbol{\theta}^i = \boldsymbol{\theta}^{i-1}
\end{aligned} \tag{10}$$

where $\boldsymbol{\theta}^0$ is a vector of initial values of the unknown model parameters to estimate, N is the number of samples, $p(\boldsymbol{\theta} \mid \mathbf{z})$ is the posterior distribution in Eq. (9), and $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{i-1}\right)$ is an arbitrary chosen proposal distribution which is used when a new sample $\boldsymbol{\theta}^*$ is to be drawn conditional on the current point $\boldsymbol{\theta}^{i-1}$. $q\left(\boldsymbol{\theta}^{i-1} \mid \boldsymbol{\theta}^*\right)$ is chosen to be a uniform distribution centered at $\left[\boldsymbol{\theta}^*\right]$ with the interval of $\pm \mathrm{d}$, where d is a vector for setting the sampling interval and is selected arbitrary based on the experience. The same interval is used for $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{i-1}\right)$. If the sample $\left[\boldsymbol{\theta}^*\right]$ is not accepted as an i th sample, the $i-1$ th sample becomes the i th sample; that is, the particular sample is doubly counted. This process is illustrated in Figure 3. Compared to PF, there is no need for resampling, but subsequent samples are affected by previous samples.

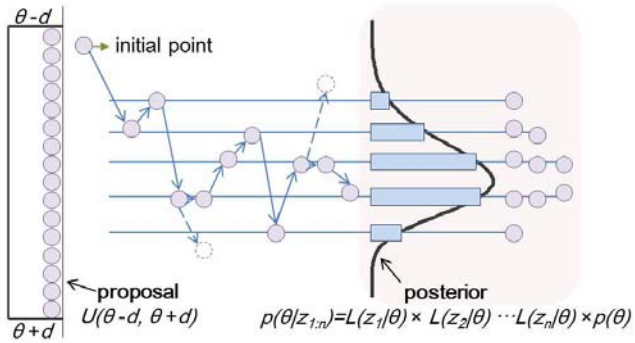


Figure 3. Illustration of the OBM process

In common with PF, the linear system in Eq. (7) is used for OBM, which can be rewritten in the following form after ignoring process noise, ν_k:

$$x_k = \theta \times t_k \tag{11}$$

In this case, the posterior distribution is as follows:

$$p(\theta \mid \mathbf{z}) \propto \frac{1}{\sqrt{2\pi}\sigma_m} \exp\left[-\frac{1}{2}\sum_{k=1}^{n} \frac{\left(z_k - x_k\right)^2}{\sigma_m^{\,2}}\right] \times U(-25, 50) \tag{12}$$

where n=20 data are used. The system state is calculated from Eq. (11) after model parameter is estimated. The results are shown in the Figure 4. They are similar to those from PF, but the intervals are narrower than PF because there is no process noise in OBM.

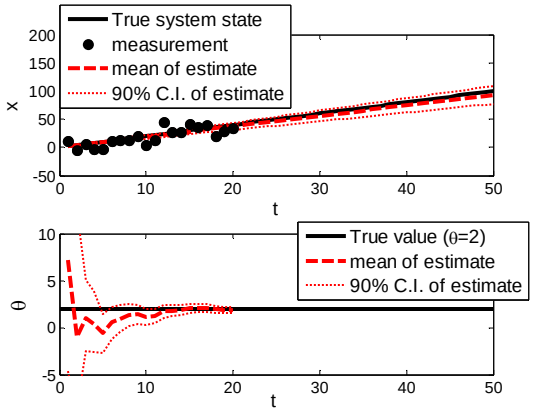


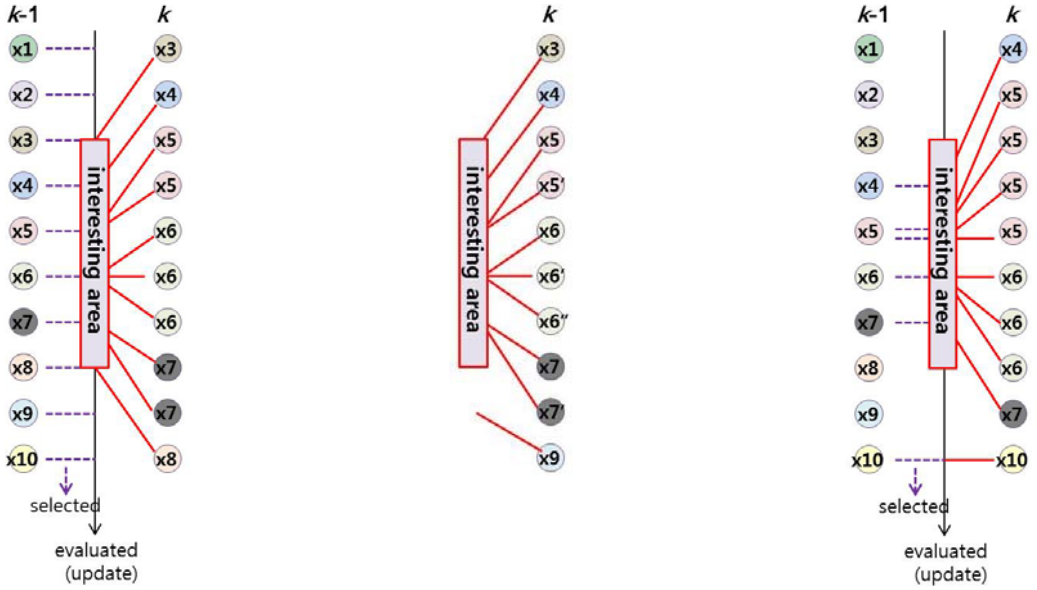
Figure 4. The results from OBM in the simple example

C. Incremental Bayesian Method (IBM)

IBM sequentially updates the distribution instead of using the final posterior distribution based on a batch of data. In this case, Eq. (10) can be changed to the following procedure:

$$
\begin{aligned}
&1.\text{For } k = 1 \text{ to } n \\
&2.\ \text{Initialize } \boldsymbol{\theta}_k^0 \sim p\left(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}\right) \\
&3.\ \text{For } i = 1 \text{ to } N \\
&\quad - \text{ Sample } u \sim U(0,1) \\
&\quad - \text{ Sample } \boldsymbol{\theta}_k^* \sim p\left(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}\right) \\
&\quad - \text{ if } u < A\left(\boldsymbol{\theta}_k^{i-1}, \boldsymbol{\theta}_k^*\right) = \min\left\{1, \frac{L\left(z_k \mid \boldsymbol{\theta}_k^*\right)}{L\left(z_k \mid \boldsymbol{\theta}_k^{i-1}\right)}\right\} \\
&\qquad\qquad \boldsymbol{\theta}_k^i = \boldsymbol{\theta}_k^* \\
&\quad \text{else} \\
&\qquad\qquad \boldsymbol{\theta}_k^i = \boldsymbol{\theta}_k^{i-1}
\end{aligned}
\tag{13}
$$

where $p\left(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_{k-1}\right)$ and $L\left(z_k \mid \boldsymbol{\theta}_k\right)$, respectively, are the sample set of prior and likelihood as shown in Eq. (8). The likelihood is the same as OBM, but the prior is given as a set of samples in IBM. Since it is not easy to express the sets of samples as equations, the samples are directly used as prior in the same was as PF. As a simple method to use samples itself as prior, samples obtained at the previous step replace the proposal distribution in MCMC as shown in Eq. (13). In this case, all prior samples are not selected for new sample, $\boldsymbol{\theta}_k^*$, because $\boldsymbol{\theta}_k^*$ is randomly selected from prior. This is different from PF where all prior samples once are selected then updated by applying the weight. In this context, the relationship of samples between previous step and current step in OBM is similar to that of PF rather than that of IBM since the posterior in OBM can be regarded the result from accumulated weight in PF. These characteristics of sampling in the three methods are illustrated in the Figure 5.

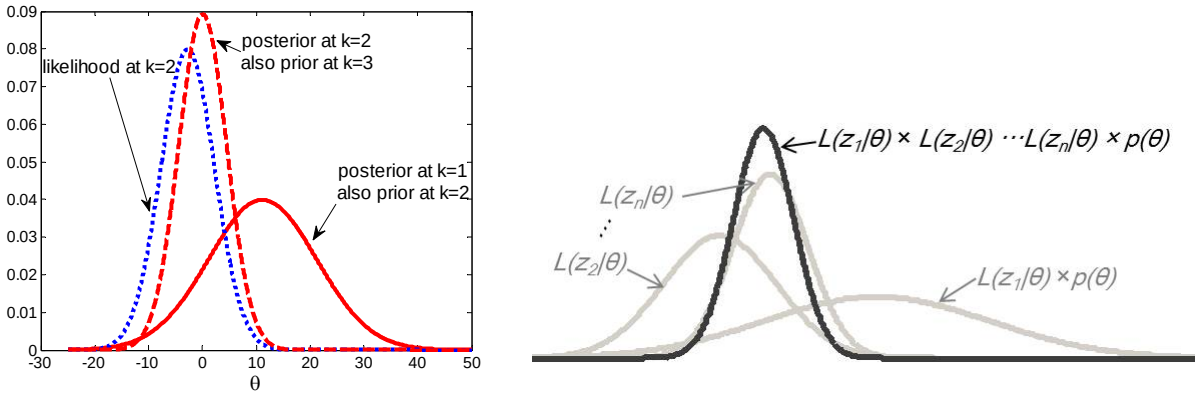


(a) PF (b) OBM (c) IBM

Figure 5. Illustration of sampling characteristic

D. Comparison of the Three Methods with the Same Configuration

In this subsection, the differences between the three methods are discussed using a same configuration. The model parameter estimation is focused rather than the state estimation, so the Eq. (11) is used for the state transition function or degradation model; that is, the system state is calculated from the estimated parameter without state transition. The procedure of three methods are illustrated as shown in Figure 6 and summarized in Table 1. The Figure 6(a) represents sequential updating process, in which distribution information to update is disseminated by samples. The posterior at k=1 depicted as solid curve becomes the prior at k=2 which is multiplied by likelihood at k=2 depicted as dotted curve to obtain the posterior distribution at k=2 depicted as dashed curve, which also becomes the prior at the next step. On the other hand, the posterior distribution in the overall method depicted as black solid curve in the Figure 6(b) is obtained resulting from multiplying all likelihood functions and prior distribution depicted as blur solid curve. Once the posterior is obtained, samples are drawn using MCMC. The more detailed procedure of three methods is listed in the Table 1. First, initial distribution is prerequisite for PF and is assumed as $\theta_0 \sim U(-25,50)$ which is wide enough not to affect the posterior distribution. For OBM and IBM, initial distribution is not required, but is given for the same configuration because initial distribution of PF corresponds to the prior information. Next, prediction and update step is performed sequentially in PF and IBM, while the posterior distribution expressed in the Eq. (12) is used for OBM. To implement the procedure, resampling and MCMC, respectively, are used for PF and IBM as well as OBM.



(a) sequential method: PF and IBM (b) overall method: OBM

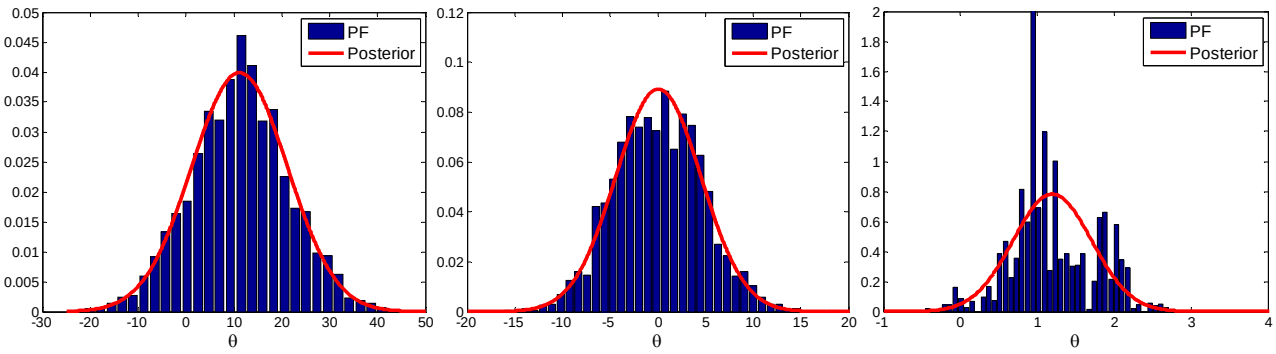
Figure 6. Illustration of the Bayesian inference

Table 1 The procedure of the three methods

Step	PF	OBM	IBM
initial distribution	$\theta_0 \sim U(-25,50)$	$U(-25,50)$	$\theta_0 \sim U(-25,50)$
prediction (prior)	$p(\theta_k \mid \theta_{k-1})$		$p(\theta_k \mid \theta_{k-1})$
update (likelihood)	$p(z_k \mid \theta_k^i)$ (Eq. (8))		$p(z_k \mid \theta_k)$ (Eq.(8))
sampling	resampling	MCMC	MCMC

The three methods are performed at every discrete time step, and the results of estimated parameter using PF are shown in Figure 7 and Table 2. In the figure, it can be found that the PDF shape becomes sparse as the update is progressed. Similar results can be observed in the case of IBM. Table 2 lists the Kolmogorov-Smirnov (K-S) test results that are the maximum difference of CDF value between the posterior distribution obtained by grid method and the distribution from the three methods. The CDF error between posterior obtained by grid method and sampling results becomes larger as the update is progressed in both PF and IBM. The reason is that sampling error caused by expressing the posterior distribution is accumulated as time evolves. On the other hand, the table shows that the samples fairly well represent the posterior distribution at all times in OBM since there is no accumulated sampling error.

In the usage aspect, PF and IBM are related with the time evolution, while OBM uses batch data for estimating the posterior distribution; PF and IBM are useful for estimating in real time as the particles are directly used for the prior information, and it is not necessary to preserve the data. On the other hand, OBM has no cumulated error and no need to suffer from utilizing prior information. PF and IBM are fundamentally the same method except sampling schemes, and if there is no cumulated error in the two methods, they are the same as OBM. In the next section, performances of the three methods are evaluated as well as the differences between the methods are discussed.



(a) k=1 (b) k=2 (c) k=10

Figure 7. The parameter estimation results from PF

Table 2 The K-S test to compare three methods

	k=1	k=2	k=10
PF	0.0262	0.0326	0.1186
OBM	0.0234	0.0286	0.0260
IBM	0.0278	0.0512	0.1130

III. Crack Growth Problem for Comparison of the Three Methods

A. Crack Growth Model

In this paper, a physical model for crack growth in an aircraft panel is used to compare the three methods: PF, OBM and IBM. This example is to estimate damage growth parameters in Paris model[20] based on the measured crack size over a number of cycles, which was addressed by An et al.[18] When the stress range due to the pressure differential is $\Delta\sigma$, the rate of damage growth can be written using the Paris model[20] as

$$\frac{da}{dN} = C(\Delta K)^m, \quad \Delta K = \Delta\sigma\sqrt{\pi a} \tag{14}$$

where a is the half crack size, N is the number of cycles, ΔK is the range of stress intensity factor, and m and C are damage growth parameters.

The synthetic measurement data are used for prognostics in this example, which are generated by (a) assuming that the true parameters, m_{true} and C_{true}, are known; (b) calculating the true crack sizes based on Eq. (14); and (c) adding random noise to the true crack size data. Once the synthetic data are obtained, the true values of crack sizes as well as the true values of model parameters are not used in the prognostics process. In this paper, the following values of parameters are used: $m_{\text{true}} = 3.8$, $C_{\text{true}} = 1.5\times10^{-10}$, $\Delta\sigma = 78.6\text{MPa}$, and initial half crack size, $a_0 = 10\text{mm}$. Also, it is assumed that the measurements are performed at every 100 cycles, and noise is uniformly distributed between -1*mm* and +1*mm*. In this study, the measurement noise is also considered as an unknown parameter.

B. Comparison of the Three Methods from Two Parameter Estimation

First, it is assumed that model parameter m and measurement noise σ_m are unknown and have prior $U(3.3, 4.0)$ and $U(0, 5)\,mm$, respectively. In this case, true value of model parameter C is used. Figure 8 shows the results from PF, where the parameters are converged to the true values as the number of cycles increase; i.e., more data are obtained. Other two methods also show a similar convergence. The K-S test results listed in the Table 3 shows that the error in PF is the largest among the three methods. However, the result of crack growth prediction from PF with estimated parameters at 1000 cycles is fairly good as shown in the Figure 9. Also, the error between the true crack growth and median of predicted crack growth and its standard deviation (std.) are listed in the Table 4. In the table, the errors of three methods are close, but the error of PF is the smallest, even though the error at the parameter estimation is the other way around. This means that while the parameter estimation cannot represent well the distribution shape due to sample duplication, the prediction results from PF can be accurate because information about the distribution are well contained in the duplicated samples. Therefore, it can be concluded that the future behavior of system can be predicted by estimating the model parameters based on the three methods.

Although the three methods represent similar results, Table 5 shows a difference in computational time, which is normalized by that of PF until 2500 cycles. In the case that data are added in stages (new data comes in at every 100 cycles), PF has the best efficiency. On the other hand, if batch data are available (all 25 data sets are available), computational time of OBM is considerably reduced in contrast to PF and IBM. This is because OBM can evaluate the posterior distribution by multiplying all likelihoods at the same time as in Eq. (9).

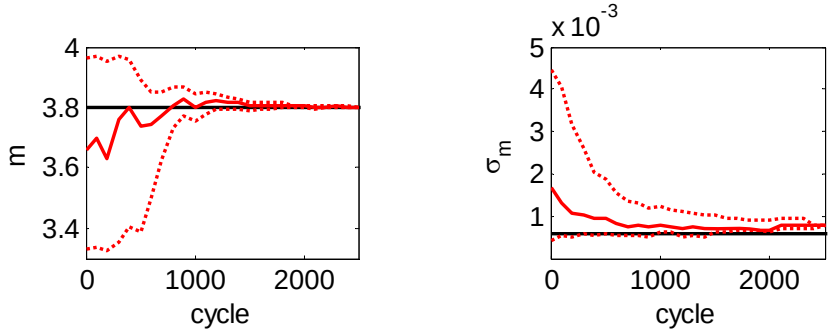


Figure 8. Convergence process of estimated parameters

Table 3 The K-S test to compare three methods at 1000 cycles

PF	OBM	IBM
0.0987	0.0632	0.0871

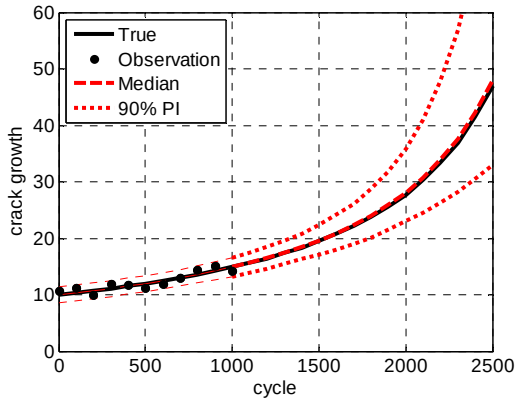


Figure 9. Crack growth prediction with estimated parameters at 1000 cycles

Table 4 The median and std. of predicted crack growth and the error with the true value

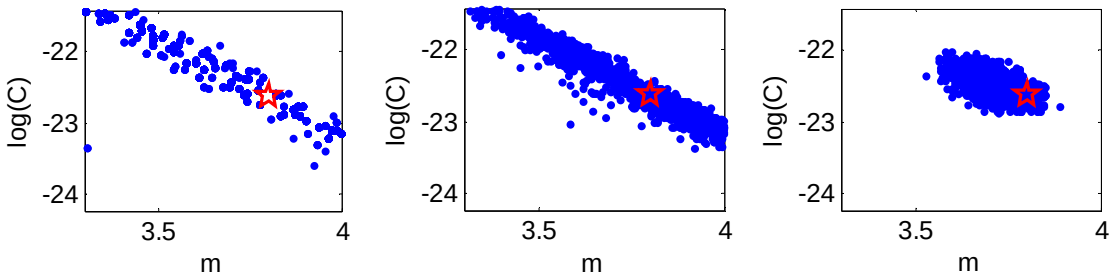
	1000 cycle		1500 cycle		2000 cycle	
	median	std.	median	std.	median	std.
true crack size	14.88		19.47		27.79	
PF	14.95	1.02	19.62	1.61	28.09	3.99
error (%)	0.45		0.78		1.08	
OBM	14.95	1.08	19.71	1.70	28.42	4.42
error (%)	0.50		1.20		2.28	
IBM	14.92	1.03	19.68	1.64	28.58	3.97
error (%)	0.24		1.09		2.86	

Table 5 Relative computational costs of three methods

	PF	OBM	IBM
data added in stages	1.0	2.13	1.70
batch data existence	1.0	0.08	1.70

C. Comparison of the Three Methods from Three Parameter Estimation

Next, two model parameters m and C, and measurement noise σ_m are considered as unknown parameters. m and σ_m have the same prior as in Section 3.3, and $\log(C)$ has prior $U\left(\log(3\mathrm{E}-11), \log(5\mathrm{E}-10)\right)$. The estimated model parameters from the three methods at 1000 cycles are shown in the Figure 10 with the true values in a star symbol. In this case, m and C are strongly correlated, and the difference in sampling results caused by the different procedures is better exposed than the case of two parameters. Although the samples from PF look scatted due to duplication of high weighted particles, the shape is similar to that of OBM, which is close to the exact posterior distribution. On the other hand, the samples from IBM are less scattered because some samples at *t*-1 are selected repeatedly.



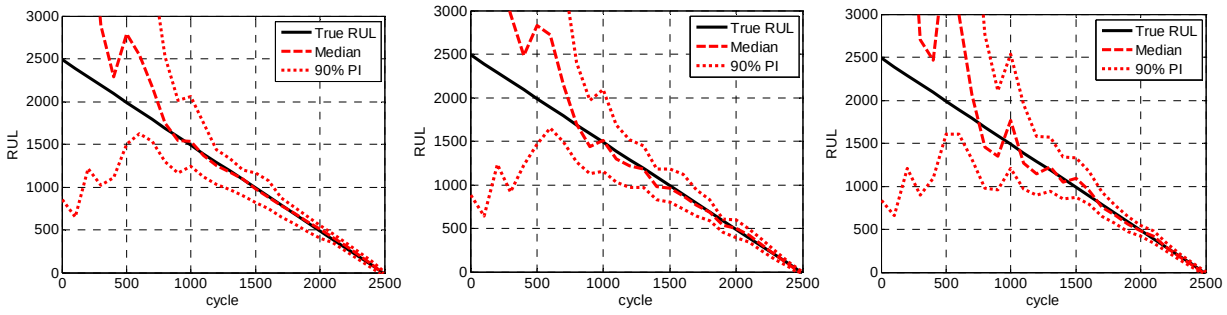
(a) PF (b) OBM (c) IBM

Figure 10. Samples of estimated parameters at 1000 cycles

RUL prediction is an ultimate goal in prognostics, and the results from three methods are shown in the Figure 11. After 1000 cycles, the median of predicted RUL is fairly accurate in PF and OBM. IBM produces somewhat worse results than PF and OBM, which is caused by the estimated parameters that do not represent overall posterior distribution as shown in the Figure 10(c). The results of RUL prediction are evaluated by using established prognostics metrics[21] to compare the three methods more objectively. In this paper, five performance metrics are used to evaluate the three methods; prognostic horizon (PH), α-λ accuracy, relative accuracy (RA), cumulative relative accuracy (CRA), and convergence.

PH is defined as the difference between the time index t_i when more than β % of the predictions distribution first meet the α accuracy zone and the time index t_{EOL} for end of life (EOL). The best score for PH is obtained when an algorithm always predicts within the desired accuracy zone. In α-λ accuracy, α and λ, respectively, are an accuracy modifier and time window modifier. $\lambda = 0$ when prediction is started, t_p, and $\lambda = 1$ at t_{EOL}. The result of α-λ accuracy becomes true or false. True is when more than β % of the predictions distribution at $t_\lambda = t_p + \lambda\left(t_{EOL} - t_p\right)$ is included in α-λ accuracy zone. RA is determined based on the error between true and prediction RUL at t_λ. CRA is the same to the accumulated RA from t_p to t_{EOL}. RA and CRA have the range from zero to one, and perfect value is one. The median of prediction RUL is used for RA and CRA. Lastly, convergence can be represented by the Euclidean distance between the center of mass of the area under the curve $M(i)$ and $\left(t_p, 0\right)$. The lower the distance is, the faster the convergence is. In this study, the following conditions are used; β =50, t_p =600, and t_{EOL} =2400, and the results are listed in the Table 6.

In the table, the best performances are represented with a bold-typed text. The reason why IBM has the worst performance is that the prior given in samples cannot be used effectively, so there is much error as the time evolves as shown in the Figure 10(c) which does not represent the posterior distribution well. But, the convergence performance of IBM is the best because the median of RUL at t_p =600 is much higher than other two methods. The performances of PF and OBM are similar, but PF has a little better performance than OBM. The reason is that sampling in OBM is not an easy task, especially under correlated parameters. It is difficult to confirm the convergence of sampling results at all time steps, because there might be improper samples. On the other hand, there are almost no samples which are not satisfied posterior distribution, and RUL can be exact since it is obtained in consequence of combination of the model parameters; scanty samples phenomenon is not too much trouble in the problem.



(a) PF (b) OBM (c) IBM

Figure 11. The RUL prediction

Table 6 Performance evaluation for three methods

	PH (α =10%)	α-λ accuracy (α =10%, λ =0.5,)	RA (λ =0.5)	CRA	Convergence ($M(i)$ is a non-negative prediction error)
PF	**1600**	**true**	**0.9940**	**0.9391**	735
OBM	**1600**	**true**	0.9731	0.9262	642
IBM	1300	false	0.8953	0.8958	**524**

IV. Discussion about the Difference between the Three Methods

In this section, the differences between the three methods are summarized. In conclusion, PF is useful for estimating in real time, and OBM is efficient when batch data exist. IBM is also for estimating in real time even though it has not been considered as a time concept, and the basic theory is the same to the PF. However, sampling method in PF (resampling) is more effective than one of IBM. The resampling method is a simple and effective way to represent the prior distribution. The sampling result from OBM is more accurate than that of PF because the latter can be changed from the initial distribution and shows scanty samples phenomenon. It is, however, difficult to obtain the accurate result because result from OBM can be significantly different from the proposal distribution. Since both PF and OBM have pros and cons, they can be used simultaneously in order to compensate for each other. Lastly, PF and IBM are fundamentally the same method except sampling schemes, and if there is no cumulated sampling error in the two methods, they will be the same to OBM. Summary of the difference between the three methods are listed in the Table 7.

Table 7 Summary of the difference between the three methods

	PF	OBM	IBM
initial distribution	should be assumed	not necessary	not necessary
levels of difficulty for performance	simple	complex	normal
accuracy of model parameter estimation	good	very good (but it is not easy to obtain it)	normal
evaluation of RUL prediction	very good	good	fair
computational costs (data added in stages)	low	high	lower than OBM
computational costs (batch data existence)	normal	very low	high
proper usage	progressed problem by time step with proper initial distribution	when batch data exist	-

V. Conclusion

In this paper, the three methods to estimate model parameters in the physics-based prognostics are compared and discussed; PF, OBM, and IBM are introduced with a simple mathematical example and evaluated by using established prognostics metrics with crack growth problem. The basic theory of IBM is the same to the PF in terms of estimating in real time. However, sampling methods of the two methods are different from each other, and sampling method in PF is more effective than one of IBM. On the other hand, OBM utilizes all the data available up to current time and evaluate the posterior distribution with one set of samples. The sampling result from OBM could be more accurate than that of PF. Consequently, PF and OBM have pros and cons, the former is useful for estimating in real time, but there is scanty samples phenomenon, while the latter is efficient under a batch of measurement data and has adequate samples, but the prediction results somewhat lack the accuracy comparing to PF. Therefore, they can be used complementally; e.g., PF is used from the latter stage after OBM is used until the early stage of growing damage state.

Acknowledgments

This work was supported by the International Cooperation of the Korea Insitute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korea government Ministry of Knowledge Economy (No. 20118520020010).

References

[1] Giurgiutiu, V., *Structural Health Monitoring: with Piezoelectric Wafer Active Sensors*, Academic Press (an Imprint of Elsevier), Burlington, MA, 2008.

[2] Schwabacher, M. A., "A survey of data-driven prognostics," *in AIAA Infotech@Aerospace Conference*, Reston,VA, 2005.

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

[4] Yan, J. and Lee, J., "A Hybrid Method for On-line Performance Assessment and Life Prediction in Drilling Operations," *in IEEE International Conference on Automation and Logistics*, 2007.

[5] Orchard, M. E., and 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.

[6] DeCastro, J. A., Tang, L., Loparo, K. A., Goebel, K., and Vachtsevanos, G., "Exact Nonlinear Filtering and Prediction in Process Model-based Prognostics," *Annual Conference of the Prognostics and Health Management Society*, 2009.

[7] Li, P., Goodall, R., and Kadirkamanathan, V., "Parameter Estimation of Railway Vehicle Dynamic Model using Rao-Blackwellised Particle Filter," *In Proceedings of the European Control Conference*, Cambridge, UK, 2003.

[8] Zio, E., and Peloni, G., "Particle Filtering Prognostic Estimation of the Remaining Useful Life of Nonlinear Components," *Reliability Engineering and System Safety*, Vol. 96, No. 3, 2011, pp. 403-409.

[9] Gašperin, M., Juričić, Đ., Boškoski, P., and Vižintin, J., "Model-based Prognostics of Gear Health using Stochastic Dynamical Models," *Mechanical Systems and Signal Processing*, Vol. 25, 2011, pp. 537-548.

[10] Wang, W., Liao, S., and Xing, T., "Particle Filter for State and Parameter Estimation in Passive Ranging," *Intelligent Computing and Intelligent Systems (ICIS), 2009 IEEE International Conference on*.

[11] Daigle, M., and Goebel, K., "Multiple Damage Progression Paths in Model-based Prognostics," *Aerospace Conference, 2011 IEEE*.

[12] Storvik, G., "Particle Filters in State Space Models with the Presence of Unknown Static Parameters." *IEEE Transactions on, Signal Processing,* Vol. 50, No. 2, 2002, pp. 281-289.

[13] Campillo, F., and Rossi, V., "Convolution Particle Filter for Parameter Estimation in General State-Space Models," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 45, 2009, pp. 1063-1072.

[14] An, D., Choi, J. H., Kim, N. H., and Pattabhiraman, S., "Fatigue Life Prediction based on Bayesian Approach to Incorporate Field Data into Probability Model," *Structural Engineering and Mechanics*, Vol. 37, No. 4, 2011, pp. 427-442.

[15] 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, 2011, pp. 828-838.

[16] Payne, S. J., "A Bayesian Approach for the Estimation of Model Parameters from Noisy Data Sets," *IEEE Signal Processing Letters*, Vol. 12, No. 8, 2005, pp. 553-556.

[17] Park, B. J., Zhang, Y., Lord, D., "Bayesian Mixture Modeling Approach to Account for Heterogeneity in Speed Data," *Transportation Research Part B-methodological*, Vol. 44, No. 5, 2010, pp. 662-673.

[18] An, D., Choi, J. H., and Kim, N. H., "Identification of Correlated Damage Parameters under Noise and Bias Using Bayesian Inference," *Structural Health Monitoring*, online 2011.10, doi: 10.1177/1475921711424520.

[19] 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.

[20] Paris, P. C., and Erdogan, F., "A Critical Analysis of Crack Propagation Laws," *ASME Journal of Basic Engineering*, Vol. 85, 1963, pp. 528–534.

[21] Saxena, A., Celaya, J., Saha, B., Saha, S., and Goebel, K., "On Applying the Prognostic Performance Metrics," *Annual Conference of the Prognostics and Health Management Society*, 2009.

[22] Bayes, T., "An Essay towards solving a problem in the doctrine of chances," *Philosophical Transactions of the Royal Society of London*, Vol. 53, 1763, pp. 370-418.

[23] Kalman, R. E., "A New Approach to Linear Filtering and Prediction Problems," *Transaction of the ASME—Journal of Basic Engineering*, Vol. 82, No. 1, 1960, pp. 35-45.

[24] Ristic, B., Arulampalam, S., and Gordon, N., *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House, 2004.

[25] Rago, C., Prasanth,R., Mehra, R. K., and Fortenbaugh, R., "Failure Detection and Identification and Fault Tolerant Control using the IMM-KF with applications to the Eagle-Eye UAV," *Proceedings of the 37th IEEE, Conference on Decision & Control*, Tampa, Florida, USA, December, 1998.

[26] Julier, S. J., and Uhlmann, J. K., "A New Extension of the Kalman Filter to Nonlinear Systems," *In Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls.*, 1997.

[27] Glynn, P. W., and Iglehart, D. L., "Importance Sampling for Stochastic Simulations," Management Science, Vol. 35, No. 11, 1989, pp. 1367-1392.