Bayesian Approach for Fatigue Life Prediction from Field Inspection

Dawn An[1] and Jooho Choi[2]
School of Aerospace & Mechanical Engineering, Korea Aerospace University, Goyang, Seoul, Korea

Sriram Pattabhiraman[3] and Nam Ho Kim[4]
University of Florida, Gainesville, FL 32611, USA

In fatigue life design of mechanical components, uncertainties arising from materials and manufacturing processes should be taken into account for ensuring reliability. A common practice is to apply a safety factor in conjunction with a physics model for evaluating the lifecycle, which most likely relies on the designer's experience. Due to conservative design, predictions are often in disagreement with field observations, which makes it difficult to schedule maintenance. In this paper, the Bayesian technique, which incorporates the field failure data into prior knowledge, is used to obtain a more dependable prediction of fatigue life. The effects of prior knowledge, noise in data, and bias in measurements on the distribution of fatigue life are discussed in detail. By assuming a distribution type of fatigue life, its parameters are identified first, followed by estimating the distribution of fatigue life, which represents the degree of belief of the fatigue life conditional to the observed data. As more data are provided, the values will be updated to reduce the confidence interval. The results can be used in various needs such as a risk analysis, reliability based design optimization, maintenance scheduling, or validation of reliability analysis codes. In order to obtain the posterior distribution, the Markov Chain Monte Carlo technique is employed, which is a modern statistical computational method which effectively draws the samples of the given distribution. Field data of turbine components are exploited to illustrate our approach, which counts as a regular inspection of the number of failed blades in a turbine disk.

Keywords: Fatigue life, Prior distribution, Posterior distribution, Bayesian approach, Markov Chain Monte Carlo Technique, Field Inspection, Turbine blade

Nomenclature

$P(x)$	=	Probability of random variable $X = x$
$P(x\|y)$	=	Conditional probability of $X = x$ with a given $Y = y$
$f_X(x)$	=	Probability density function of random variable X
$f_{XY}(x,y)$	=	Joint probability density function of random variables X and Y
$f_X(x\|y)$	=	Marginal probability density function
P_f	=	Probability of failure
μ, σ	=	Parameters for normal distribution
m, η	=	Parameters for Weibull distribution
$f_{\text{life}}(t)$	=	Distribution of fatigue life

[1] Graduate Student
[2] Professor, AIAA Member
[3] Graduate Research Assistant, Student Member
[4] Associate Professor, Corresponding author, Department of Mechanical and Aerospace Engineering, PO Box 116250, University of Florida, Gainesville, FL 32611-6250, AIAA Member

I. Introduction

PERFORMANCE of mechanical components undergoes a change by uncertainties such as environmental effects, dimensional tolerances, loading conditions, material properties and maintenance processes. Especially when the design criterion is fatigue life, it is significantly affected by system uncertainties. Even with today's modern computing systems, it is infeasible to include all the relevant uncertain variables into the analytical prediction, since many of the potential inputs are not characterized in the design phase. Approximation methods, such as the response surface method with Monte Carlo simulation (MCS)[1,2], were often employed to overcome excessive computational cost in reliability assessment.

To account for the unknown variables, common practices use so called "safety factors" or statistical minimum properties in conjunction with the analytical prediction when evaluating lifetimes. Due to these conservative estimations, analytical predictions are often in disagreement with field experience, and a gap exists in correlating the field data with the analytical predictions. Thus, there is an increasing need to improve the analytical predictions using field data, which collectively represents the real status of a particular machine.

Field failure data can be helpful in predicting fatigue life that has uncertainties due to the unknown potential inputs. Recently, for more reliable life prediction, many studies using field data have been undertaken[3]. In non-fatigue life prediction, Orchard *et al*.[4] used particle filtering and learning strategies to predict the life of a defective component. Marahleh *et al*.[5] predicted the creep life from test data, using the Larson-Miller parameter. Park *et al*.[6] used an energy-based approach to predict constant amplitude multiaxial fatigue life. Guo *et al*.[7] performed reliability analysis for wind turbines using maximum likelihood function, incorporating test data.

In this paper, the Bayesian technique is utilized to incorporate field failure data with prior knowledge to obtain the posterior distribution of the unknown parameters of the fatigue life[8]. The analytical predictions are obtained either from numerical models or laboratory tests. The field data, although noisy, invariably portray environmental factors, measurement errors, and loading conditions, or in short, reality. Since the predictions incorporate field experience, as time progresses and more data are available, the probabilistic prediction is continuously updated. This results in a continuous increase of confidence and accuracy of the prediction. In this paper, Markov Chain Monte Carlo (MCMC) technique is employed as an efficient means to draw samples of given distribution[9]. Consequently, the posterior distribution of the unknown parameters of the fatigue life is obtained in light of the field data collected from the inspection. Subsequently, fatigue life is predicted a posteriori based on the drawn samples. The resulting distributions can then be used directly in risk analysis, maintenance scheduling, and financial forecasting by both manufacturers and operators of heavy-duty gas turbines. This presents a quantification of the real time risk for direct comparison with the volatility of the power market.

The paper is organized as follows. In Section 2, the Bayesian technique is summarized, particularly with estimating the distribution of fatigue life through identifying the distribution of parameters. Section 3 discusses the effect of noise and bias on the accuracy of posterior distribution. In Section 4, five different cases are considered with varying priors and likelihoods, followed by conclusions in Section 5.

II. BAYESIAN TECHNIQUE FOR LIFE PREDICTION

In this section, Bayesian inference is explained in the view of updating distribution of fatigue life using test data. The Bayesian theorem is first presented in a general form, followed by a specific expression for estimating the distribution of fatigue life.

A. Bayes' theorem

Bayesian inference estimates the degree of belief in a hypothesis based on collected evidence. Bayes[10] formulated the degree of belief using the identity in conditional probability:

$$P(x \cap y) = P(x \mid y)P(y) = P(y \mid x)P(x) \tag{1}$$

where $P(x|y)$ is the conditional probability of $X = x$ given $Y = y$. In the case of estimating the probability of fatigue life using test data, the conditional probability of event X (i.e., fatigue life) when the probability of test Y is available can be written as

$$P(x \mid y) = \frac{P(y \mid x)P(x)}{P(y)} \tag{2}$$

where $P(x|y)$ is the posterior probability of fatigue life X for given test y, and $P(y|x)$ is called the likelihood function or the probability of obtaining test Y for a given fatigue life x. In Bayesian inference, $P(x)$ is called the prior probability, and $P(y)$ is the marginal probability of Y and acts as a normalizing constant. The above equation can be used to improve the knowledge of $P(x)$ when additional information $P(y)$ is available.

Bayes' theorem in Eq. (2) can be extended to the continuous probability distribution with probability density function (PDF), which is more appropriate for the purpose of the present paper. Let f_X be a PDF of fatigue life X. If the test measures a fatigue life Y, it is also a random variable, whose PDF is denoted by f_Y. Then, the joint PDF of X and Y can be written in terms of f_X and f_Y, as

$$f_{XY}(x,y) = f_X(x \mid Y = y) f_Y(y) = f_Y(y \mid X = x) f_X(x) \tag{3}$$

When X and Y are independent, the joint PDF can be written as $f_{XY}(x,y) = f_X(x) \cdot f_Y(y)$ and Bayesian inference cannot be used to improve the probabilistic distribution of $f_X(x)$. Using the above identity, the original Bayes' theorem can be extended to the PDF as[11]

$$f_X(x \mid Y = y) = \frac{f_Y(y \mid X = x) f_X(x)}{f_Y(y)} \tag{4}$$

Note that it is trivial to show that the integral of $f_X(x|Y = y)$ is one by using the following property of marginal PDF:

$$f_Y(y) = \int_{-\infty}^{\infty} f_Y(y \mid X = \xi) f_X(\xi) \mathrm{d}\xi \tag{5}$$

Thus, the denominator in Eq. (4) can be considered as a normalizing constant. By comparing Eq. (4) with Eq. (2), $f_X(x \mid Y = y)$ is the posterior PDF of fatigue life X given test $Y = y$, and $f_Y(y \mid X = x)$ is the likelihood function or the probability density value of test Y given fatigue life $X = x$.

When the analytical expressions of the likelihood function, $f_Y(y \mid X = x)$, and the prior PDF, $f_X(x)$, are available, the posterior PDF in Eq. (4) can be obtained through simple calculation. In practical applications, however, they may not be in the standard analytical form. In such a case, the Markov Chain Monte Carlo (MCMC) simulation method can be effectively used, which will be addressed in Section 2.3 in detail.

When multiple, independent tests are available, Bayesian inference can be applied either iteratively or all at once. When N number of tests are available; i.e., $\mathbf{y} = \{y_1, y_2, \ldots, y_N\}$, the Bayes' theorem in Eq. (4) can be modified to

$$f_X(x \mid Y = \mathbf{y}) = \frac{1}{K} \prod_{i=1}^{N} \left[f_Y(y_i \mid X = x) \right] f_X(x) \tag{6}$$

where K is a normalizing constant. In the above expression, it is possible that the likelihood functions of individual tests are multiplied together to build the total likelihood function, which is then multiplied by the prior PDF followed by normalization to yield the posterior PDF. On the other hand, the one-by-one update formula for Bayes' theorem can be written in the recursive form as

$$f_X^{(i)}(x \mid Y = y_i) = \frac{1}{K_i} f_Y(y_i \mid X = x) f_X^{(i-1)}(x), \quad i = 1, \ldots, N \tag{7}$$

where K_i is a normalizing constant at i-th update and $f_X^{(i-1)}(x)$ is the PDF of X, updated using up to ($i-1$)th tests. In the above update formula, $f_X^{(0)}(x)$ is the initial prior PDF, and the posterior PDF becomes a prior PDF for the next update.

In the view of Eqs. (6) and (7), it is possible to have two interesting observations. Firstly, the Bayes' theorem becomes identical to the maximum likelihood estimate when there is no prior information; i.e., $f_X(x)$ = constant. Secondly, the prior PDF can be applied either first or last. For example, it is possible to update the posterior distribution without prior information and then to apply the prior PDF after the last update.

An important advantage of Bayes' theorem over other parameter identification methods, such as the least square method and maximum likelihood estimate, is its capability to estimate the uncertainty structure of the identified parameters. These uncertainty structures depend on that of the prior distribution and likelihood function. Accordingly, the accuracy of posterior distribution is directly related to that of likelihood and prior distribution. Thus, the uncertainty in posterior distribution must be interpreted in that context.

B. Application to fatigue life estimation

In deriving Bayes' theorem in the previous section, two sets of information are required: a prior PDF and a likelihood function. In estimating fatigue life, the prior distribution can be obtained from numerical models and laboratory tests. Since they can be performed multiple times with different input parameters that represent various uncertainties, it is possible to evaluate the distribution of fatigue life, which can be served as a prior PDF of fatigue life.

On the other hand, the field data cannot be obtained in a laboratory environment. In this section, using field data in calculating the likelihood function is presented. When a gas turbine engine is built and installed in the field, the maintenance/repair reports include the history of the number of parts that were defective and replaced at specific operating cycles. Although these data are not obtained under a controlled laboratory environment, they represent reality with various effects of uncertainties in environmental factors, measurement errors, and loading conditions. Thus, it is desirable to use these data to update the fatigue life of the specific machine using Bayes' theorem.

The standard approach to applying Bayes' theorem is to use the field data to build the likelihood function, which is basically the same as the PDF form with fatigue life. However, different from specimen-level tests, the field data cannot be repeated multiple times to construct a distribution. Only one data point exists for the specific operation cycles. Thus, the original formulation of Bayes' theorem needs to be modified. First, instead of updating the PDF of fatigue life, it is assumed that the distribution type of fatigue life is known in advance. This can be a big assumption, but it is possible that different types of distributions can be assumed and the most conservative type can be chosen. Once the distribution type is selected, then it is necessary to identify distribution parameters. For example, in the case of normal distribution, the mean (μ) and standard deviation (σ) need to be identified. In this paper, these distribution parameters are assumed to be uncertain and Bayes' theorem is used to update their distribution; i.e., the joint PDF of mean and standard deviation will be updated. In this case, the vector of random variables is defined as $\mathbf{X} = \{\mu, \sigma\}$, and the joint PDF $f_{\mathbf{X}}$ is updated using Bayes' theorem. Initially, it is assumed that the mean and standard deviation are uncorrelated.

A field data set consists of the number of hours of operation until inspection (N_f), and the number of defective blades (r) out of the total number of blades (n). Thus, the field data are represented by $\mathbf{y} = \{N_f, n, r\}$, which are given in Table 1. Then, the likelihood function is the PDF f_Y for given $\mathbf{X} = \{\mu, \sigma\}$. Since the field data is given at fixed N_f and n, f_Y can be represented in terms of *r*. Unfortunately, the number of defective blades cannot be a continuous number because it is an integer. Thus, the likelihood function f_Y can be represented using the following probability mass function:

$$f_Y\left(\mathrm{y} \mid \mathbf{X} = \{\mu, \sigma\}\right) = \frac{n!}{r!(n-r)!}(P_f)^r(1-P_f)^{n-r} \tag{8}$$

where P_f is the probability of defects at given N_f for given $\mathbf{X} = \{\mu, \sigma\}$. Since the distribution of fatigue life is given as a function of $\mathbf{X}$, the probability of defects can be calculated by

$$P_f\left(\mu, \sigma\right) = \int_0^{N_f} f_{\text{life}}\left(t; \mu, \sigma\right) dt \tag{9}$$

where, f_{life} is the PDF of fatigue life distribution. The probability mass function in Eq. (8) is a binomial distribution, which models the probability distribution of having '*r*' defects out of '*n*' samples with defect probability of P_f. Figure 1 illustrates the relation in Eq. (9). The predictive distribution of life can be estimated using the mean and standard deviation obtained from the updated joint PDF as

$$f_{\text{life}}(t) = N(\mu, \sigma) \tag{10}$$

In this paper, since the distribution type of the fatigue life is unknown, two different types are assumed: normal and Weibull. The strategy is to select the one that can provide a more conservative estimate. In the case of the Weibull distribution, the scale and shape parameters (m, η) are used in the Bayesian technique.

As mentioned above, the predictive distribution of fatigue life depends on that of the prior distribution and likelihood function. The PDF of life $f_{\text{life}}(t|\mathbf{X})$ and the prior PDF $f(\mathbf{X})$ are therefore assumed as the following cases:

Case 1: $f_{\text{life}}(t \mid \mu,\sigma)$: normal dist. and $f(\mu,\sigma)$: non-informative prior (11)

Case 2: $f_{\text{life}}(t \mid \mu,\sigma)$: normal dist. and $f(\mu,\sigma) = N(1,0.5) \cdot N(0.154,0.077)$ (12)

Case 3: $f_{\text{life}}(t \mid \mu,\sigma)$: normal dist. and $f(\mu,\sigma) = N(1,0.5) \cdot \text{Inv-}\chi^2(4-1,0.154^2)$ (13)

Case 4: $f_{\text{life}}(t \mid \eta,m)$: Weibull dist. and $f(\eta,m)$: non-informative prior (14)

Case 5: $f_{\text{life}}(t \mid \eta,m)$: Weibull dist. and $f(\eta,m) = LogN(\text{-}0.049,0.472) \cdot LogN(1.912,0.472)$ (15)

For the likelihood function, normal and Weibull distributions are first considered as fatigue life distributions, and then, are used to calculate the fatigue failure probability P_f in likelihood calculation as in Eq. (8). In the model, the associated model parameters are μ, σ in the case of normal and m, η in the case of Weibull distribution, respectively. These are taken to be unknown and are estimated using the inspected data. In terms of prior distribution, the probability distribution of fatigue life that was obtained numerically by conducting reliability analysis is exploited, in which the mean and standard deviation of fatigue life are given by normalized fatigue life 1 and 0.154 respectively. Based on these values, different kinds of prior distributions for the two parameters are undertaken for the Cases 2, 3 and 5, in which the coefficient of variation (COV) for the mean and standard deviation are assumed as 0.5 in common.

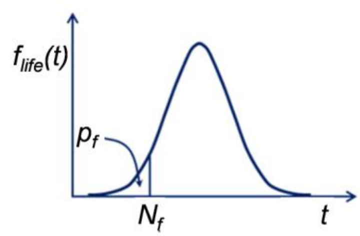


Figure 1: Probability of defects calculation from life distribution.

Table 1: Field data for inspected turbine blades

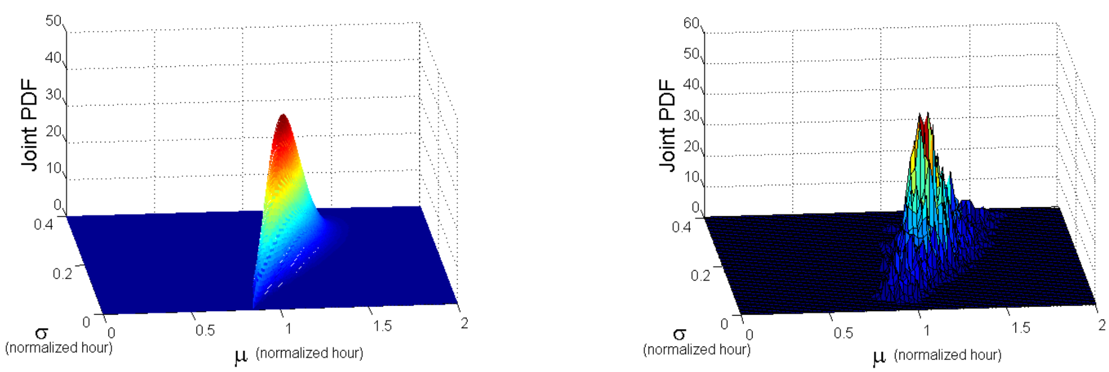
Engine	Hours(N_f)	Failed(r)/Total(n)	Engine	Hours(N_f)	Failed(r)/Total(n)
1	0.836	2/40	8	0.281	0/40
2	0.604	1/40	9	5.053	13/40
3	0.290	1/40	10	0.707	0/40
4	1.770	0/40	11	1.652	0/40
5	2.321	12/40	12	1.265	10/40
6	2.254	3/40	13	3.615	18/40
7	1.162	6/40			

C. MCMC simulation

Once the expression for the posterior PDF is available as in Eq. (6), one can proceed to sample from the PDF. A primitive way is to compute the PDF values at a grid of points after identifying the effective range, and sample by the inverse CDF method. This method, however, has several drawbacks such as the difficulty finding correct location and scale of the grid points, the spacing of the grid, and so on. Especially when a multi-variable joint PDF is required, the computational cost is proportional to N^m, where N is the number of grids in one-dimension and m is the number of variables. On the other hand, the MCMC simulation can be an effective solution as it is less sensitive to the number of variables[9]. The Metropolis-Hastings (M-H) algorithm is a typical method of MCMC, which is given in the case of two parameters (μ, σ) by the following procedure:

$$
\begin{aligned}
&1.\ \text{Initialize } \left[\mu^{(0)}, \sigma^{(0)}\right]. \\
&2.\ \text{For } i = 0 \text{ to } N-1 \\
&\quad -\ \text{Sample } u \sim U_{[0,1]}. \\
&\quad -\ \text{Sample } \left[\mu^*, \sigma^*\right] \sim q\left(\mu^*, \sigma^* \mid \mu^{(i)}, \sigma^{(i)}\right). \\
&\quad -\ \text{if } \ u < A\left(\left[\mu^{(i)}, \sigma^{(i)}\right], \left[\mu^*, \sigma^*\right]\right) \\
&\qquad \left(A = \min\left\{1, \frac{p\left(\mu^*, \sigma^*\right) q\left(\mu^{(i)}, \sigma^{(i)} \mid \mu^*, \sigma^*\right)}{p\left(\mu^{(i)}, \sigma^{(i)}\right) q\left(\mu^*, \sigma^* \mid \mu^{(i)}, \sigma^{(i)}\right)}\right\}\right) \\
&\qquad \left[\mu^{(i+1)}, \sigma^{(i+1)}\right] = \left[\mu^*, \sigma^*\right] \\
&\quad \text{else} \\
&\qquad \left[\mu^{(i+1)}, \sigma^{(i+1)}\right] = \left[\mu^{(i)}, \sigma^{(i)}\right]
\end{aligned}
\tag{16}
$$

where $\left[\mu^{(0)}, \sigma^{(0)}\right]$ is the initial value of unknown parameters to estimate, N is the number of iterations or samples, U is the uniform distribution, $p(\mu, \sigma)$ is the posterior PDF (target PDF), and $q(\mu, \sigma)$ is an arbitrarily chosen distribution. A uniform distribution is used in this study for the sake of simplicity. Thus, $q\left(\mu^{(i)}, \sigma^{(i)} \mid \mu^*, \sigma^*\right)$ and $q\left(\mu^*, \sigma^* \mid \mu^{(i)}, \sigma^{(i)}\right)$ become constants, and $q(\mu, \sigma)$ can be ignored. As an example of MCMC, the joint posterior PDF of the unknown parameters μ, σ of the fatigue life using only the first data is shown in Figure 2, which represents the degree of belief on the concerned parameters in the form of PDF. The joint posterior PDF using the grid method as well as MCMC sampling are shown in Figure 2(a) and (b), respectively. In Table 2, statistical moments by the two methods are compared. As shown in the table, the two methods agree quite closely but MCMC used $10(10)^3$ samples, whereas gird used $250(10)^3$ samples. The difference in the number of samples will be significantly increased as more variables are identified.



(a) using grid method (500 $\times$ 500 grid) (b) using MCMC (10^4 iterations)

Figure 2: Joint posterior PDF of Case 2 in Eq. (12) with one test data

Table 2: Statistical moments by the two methods

	μ_μ	μ_σ	σ_μ	σ_σ	$Cov(\mu, \sigma)$
Grid	1.1476	0.1854	0.1265	0.0656	0.0000
MCMC	1.1353	0.1797	0.1241	0.0656	0.0000

3. ANALYTICAL EXAMPLE

Although the Bayesian approach has been used extensively in literature[12], it is important to investigate its performance. In particular, it is important to characterize how this method identifies unknown parameters when the experimental data have noise and bias. In this section, the properties of Bayes' theorem are studied using analytical examples. In particular, the effects of noise and bias of data on the final distribution are discussed. The data here are simulated to demonstrate the new analytical technique. These data are then perturbed in order to study the effects of noise and bias errors on the algorithm.

A. Field data governed by a distribution

The first study is to test the accuracy of the updated distribution using Bayes' theorem. Table 3 shows an example of field data that are used in this section. The total number of blades is n = 50. The field data are generated from a distribution, $B \sim N(12000, 2000)$ (B is actually unknown distribution which should estimated based on the field data but is used to generate the field data) and the objective is to test if the updated distribution recovers the distribution *B*. For the prior distribution, the mean is uniformly distributed in the interval of [0, 30000], while the standard deviation is also uniformly distributed in the interval of [0, 4500]. Using the four sets of field data in Table 3, the posterior joint PDF are calculated using Bayes' theorem. After obtaining the posterior joint PDF, samples obtained from the MCMC simulation are used to predict the distribution of life. Figure 3 compares the updated distribution of fatigue life with the distribution *B*, along with the predicted probability of defects from field data. In this figure, the red curves represent the 5% lower bound, mean, and 95% upper bound from the left hand, respectively. It is clear that both distributions match each other quite well. When the field data are governed by a particular distribution, Bayes' theorem can reproduce the distribution well when more than 3 sets of data are available.

Table 3: Sample field data for inspected turbine blades

Engine	1	2	3	4
Operating hours	9303	14255	12700	11402
Number of defective blades	4	44	32	19

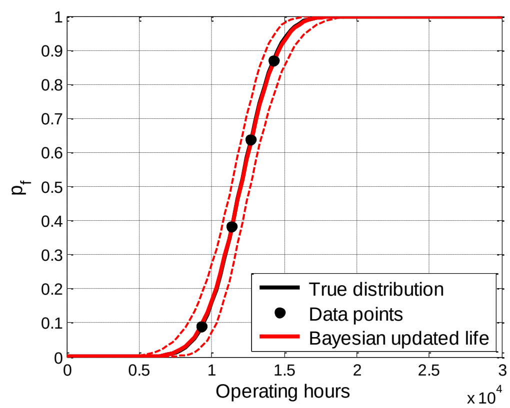


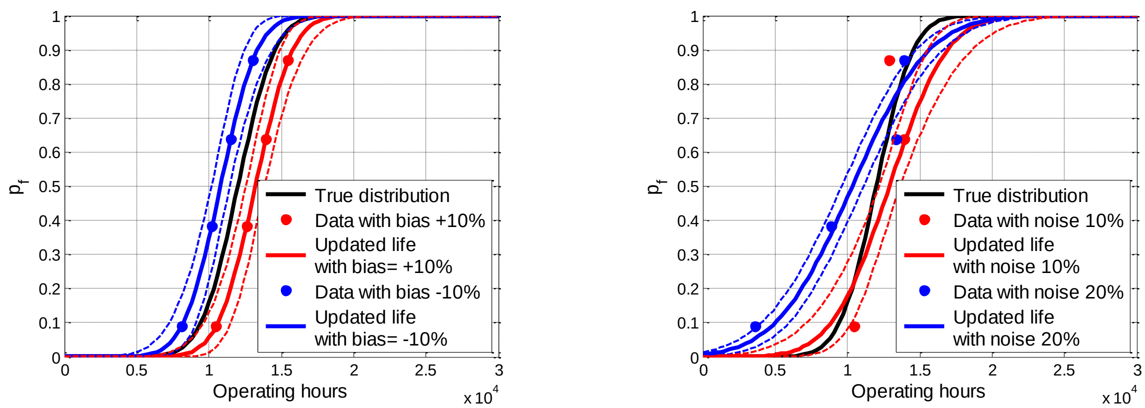
Figure 3: Comparison between the updated life distribution using Bayes' theorem and the original distribution

B. Effect of bias and noise

In practice, the field data are often accompanied by noise and bias. The former is caused by variability in measurement environment, while the latter represents systematic departure, such as device error. The difference is that the former is random, while the latter is deterministic, although its value is unknown. In some cases, a positive bias is consciously applied to remain conservative. This section analyzes the effect of bias and noise on the updated life distribution using the same sample data shown in Table 3.

First, the bias is given in terms of the operating hours of blades. For example, bias = 10% means that the operating hours of blades is 10% more than the nominal operating hours. Figure 4(a) shows the effect of bias on the distribution of fatigue life. Both positive and negative biases are considered. As expected a negative bias leads to a conservative estimate of the updated life distribution. From Table 4, it can be observed that the standard deviation of the life distribution remains about the same as the true value 2000, while the mean values are shifted by ±1200 due to the biases.

Next, the effect of noise in the data is investigated by randomly perturbing the original data by 10% and 20%. Figure 4 (b) shows the updated distribution of fatigue life with the two different levels of noise, and more specific results are shown in Table 4, which also includes the effect of the number of data. Different from the case of bias, as the number of data increases, the means converge to a nominal mean. However, the standard deviation is relatively insensitive to the number of data; it shows a slight increase with 50 points of data. Although the estimated distribution is inaccurate with high levels of noise, it tends to be conservative when the life with a low probability of failure is estimated. In general, bias can be identified with many data by using the least-square method or the Bayesian approach with the bias as an unknown system parameter[13].



(a) life distribution results due to ±10% bias (b) life distribution results due to 10% and 20% noise

Figure 4: Effect of bias and noise

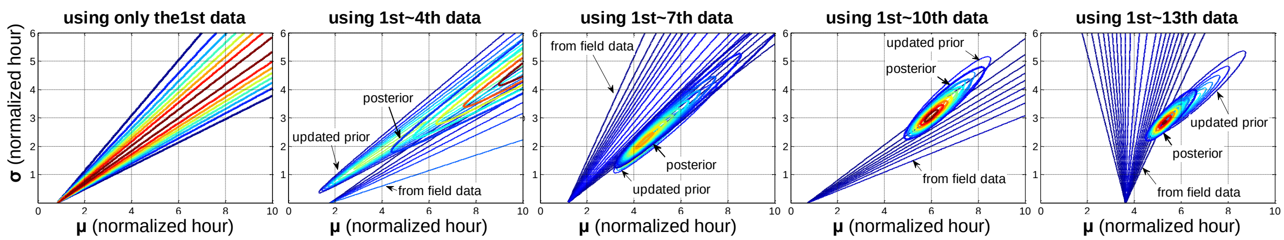
Table 4: Mean and standard deviation of unknown parameter μ, σ

	determ-inistic	bias		noise 10%			noise 20%		
		+10%	-10%	data=4	data=10	data=50	data=4	data=10	data=50
μ_μ	12008	13209	10795	12907	11854	12148	10350	11514	12024
Error (%)	0.07	10.08	(-)10.04	7.55	1.21	1.24	13.75	4.05	0.20
μ_σ	2054	2039	2017	3187	3024	3080	4166	4225	4473
Error (%)	2.69	1.94	0.86	59.34	51.18	53.98	108.32	111.23	123.65
σ_μ	404	404	397	453	408	196	453	347	208
σ_σ	322	319	326	640	381	157	262	211	89

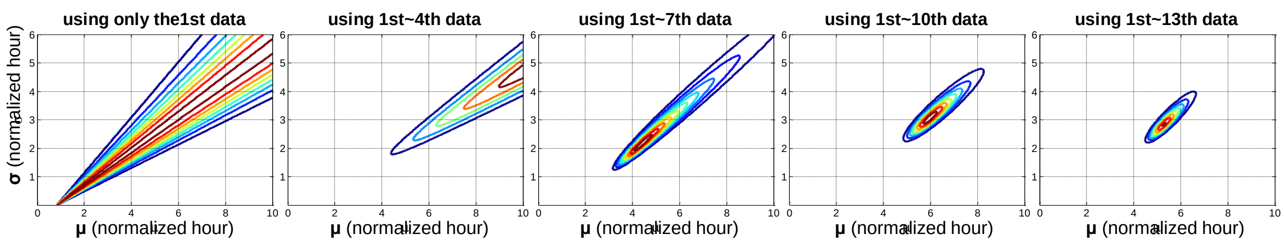
4. IDENTIFICATION OF MODEL PARAMETER AND PREDICTION OF LIFE DISTRIBUTION

A. Posterior distribution of model parameter

The results of Case 1 where the likelihood is a normal distribution and non-informative prior are shown in Figure 5. In Figure 5(a), the contours of prior distribution, likelihood function and joint posterior PDF of the unknown parameters are plotted. In these figures, the updated prior and the likelihood are obtained from the posterior distribution previously obtained and the inspection data, respectively. The posterior distribution is obtained by multiplying the prior and likelihood, and is used in the next updating step as the prior distribution. Since a non-informative prior is used, the first update is identical to the first likelihood function. Since data from a single test can be represented by infinite combinations of means and standard deviations, the contours of likelihood become straight lines. In Figure 5(b), the posterior PDFs are plotted in the form of contours. Even if the contours of individual likelihood functions are straight, the uncertainty in the posterior distribution is reduced and show correlation between the mean and standard deviation. It is shown that as more data are added, the location and range of μ, σ moves and narrows down to converge to a certain point. The results indicate our knowledge on the unknown parameters based on the field inspection.



(a) likelihood function and updated prior and posterior distribution



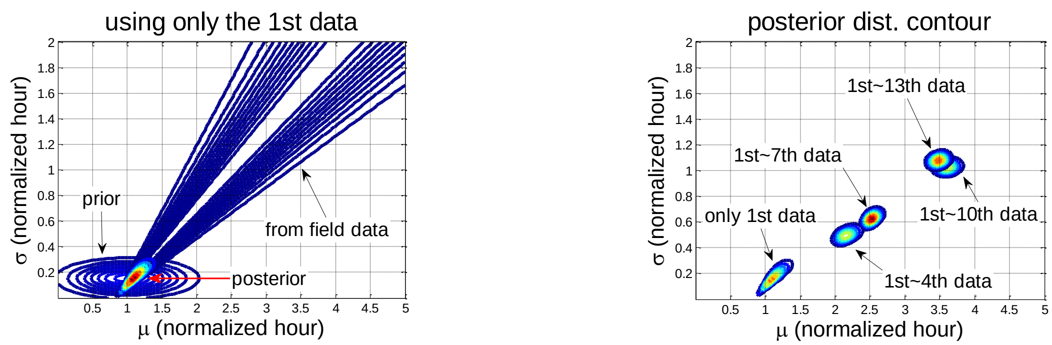
(b) contour of joint posterior PDF

Figure 5: Updated posterior PDF of Case 1.

The results of Cases 2 – 5 are shown in Figure 6~Figure 9. The results of Case 2 where the likelihood is still a normal distribution but with normally distributed priors are shown in Figure 6(a) and (b). Different from the non-informative prior case in Figure 5, the level of uncertainties remains almost constant, but their locations continuously change for different field data. As illustrated in Figure 6(a), the posterior distribution is already narrowed due to the prior distribution. However, since the prior distribution is quite different from the field data, the centers of posterior distributions gradually move as more field data are used. In addition, the final posterior distribution moves toward to the prior distribution compared to that of Figure 5. Figure 5 and Figure 6 clearly provide the effect of prior distribution. When there is a strong inclination to the prior distribution, then its use can yield the posterior distribution closer to it than the non-informative prior. Then due to field data that predict a longer fatigue life, the expected life will gradually increase.

The results of Case 3 where the likelihood is normal and the prior for the sigma is changed to a chi-square distribution with degree of freedom $n(=4)-1$ and scale parameter $\sigma_0 = 0.154$, which are more reasonable assumptions due to non-negativity, are shown in Figure 7(a) and (b). In this case, the standard deviation of the posterior distribution increases with more field data because the prior has a small standard deviation than that of field data. If the likelihoods are considered first with a non-informative prior, as is found in the 1st~4th data and 1st~10th data of Figure 7(c), and the posterior is obtained by applying the prior at the last stage, which is 1st~13th data in the figure, one gets the standard deviation decreased as more field data are added, with the final posterior being updated by the prior.

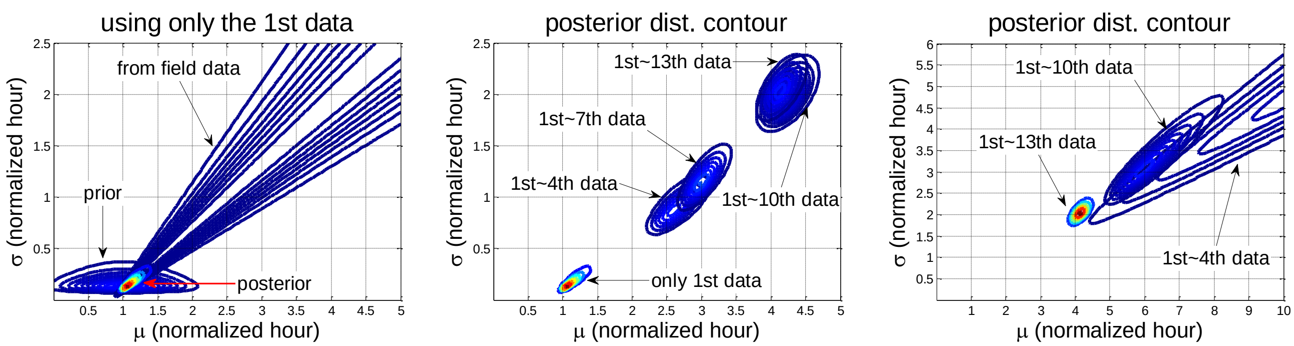
The results of Case 4 where the likelihood is the Weibull distribution with scale (η) and shape (m) parameter and a non-informative prior are shown in Figure 8, and the results of Case 5 where the likelihood is still the Weibull distribution but with the prior being lognormally distributed are shown in Figure 9. Both cases show the convergent behavior, but the range of m, η of Case 5 is narrower than that of Case 4 due to the prior. The case of a non-informative prior shows a strong correlation between the mean and standard deviation, which is significantly reduced with the lognormal prior. These cases are the most reasonable because the Weibull distribution is the best model for the lifetime[14].



(a) results using the first data2

(b) updated posterior PDF contours

Figure 6: Updated posterior PDF contours of Case 2



(a) results using only the first data

(b) updated posterior PDF

(c) updated posterior PDF with prior being applied at the last stage

Figure 7: Updated posterior PDF contours of Case 3

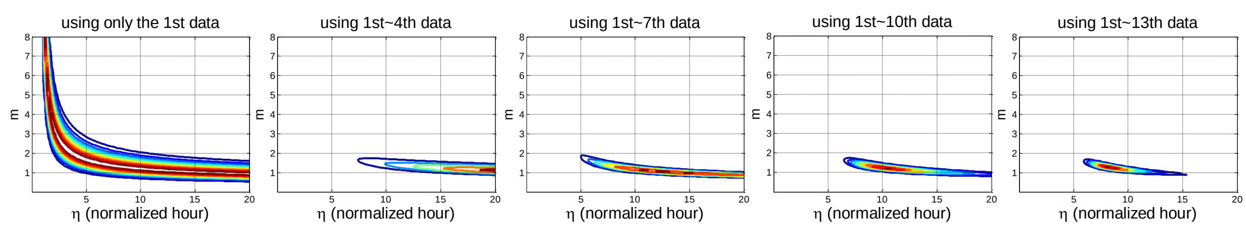


Figure 8: Updated posterior PDF contours of Case 4

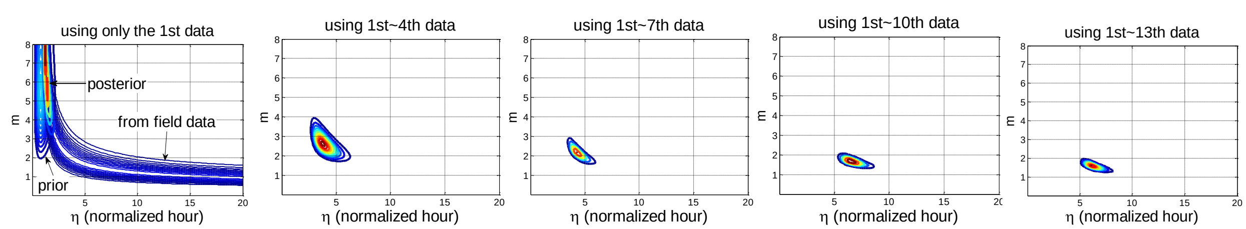
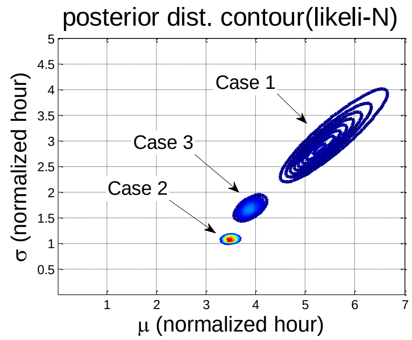
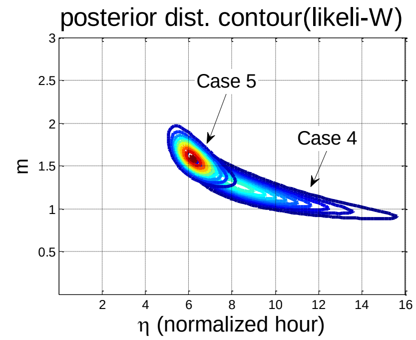


Figure 9: Updated posterior PDF contours of Case 5

The final posterior PDFs of all cases are shown in Figure 10. Figure 10(a) and (b) are the results of μ and σ of the normal distribution and η and m of the Weibull distribution, respectively. Case 1 in Figure 10(a) and Case 4 in Figure 10(b) are the results of non-informative priors. As was expected, the results are much wider than the others due to the non-informative prior. If specific prior information is available, the precision of the posterior distribution is increased.



(a) likelihood is normal distribution



(b) likelihood is Weibull distribution

Figure 10: Final posterior PDFs of all cases.

B. Posterior predictive distribution of fatigue life

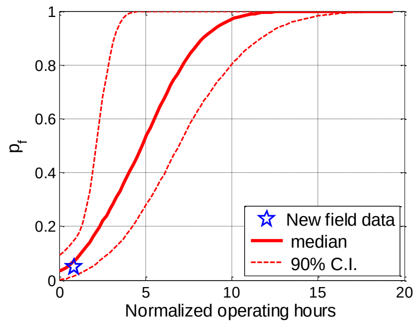
The posterior PDF obtained in Section 4.2 can be used to predict the probability of fatigue life. Recall that the Bayesian inference updates the distribution of mean and standard deviation of the life distribution. Each sample of mean and standard deviation yields a distribution of fatigue life, which can be represented by a cumulative distribution function (CDF). Thus, as a result of Bayesian inference, a distribution of CDF can be obtained. For explanatory purposes, this distribution of CDF can be represented by a confidence interval. It is expected that this interval will be wide when the uncertainty in the mean and standard deviation is large. In general, the uncertainty in mean and standard deviation reduces with greater numbers of field data, and the confidence interval will also be reduced with more field data.

Figure 11 shows the updating process of the predicted fatigue life CDF along with field data for Case 1. The red stars are the field data at the current update, while the blue stars are the field data up to the previous update. In the same figure, the dashed red curve is the mean CDF of fatigue life, while the two solid red curves are 5% and 95% confidence bounds. Compared to the significant noise in the field data, the confidence interval of the predicted CDF is progressively reduced. In order to accommodate safety, it is advised to take a 5% lower bound of the CDF.

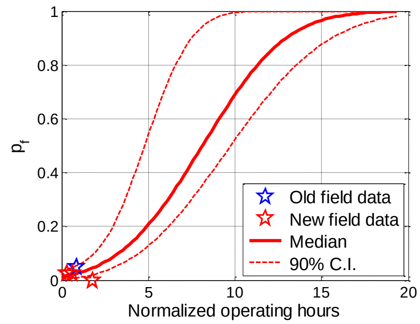
In order to show the difference between initial and final distributions of fatigue life, Figure 12 plots the confidence intervals using prior and the final posterior distribution of mean and standard deviation for Cases 2, 3, and 5. It is clear that all initial distributions are overly conservative. This often happens because analytical models often assume all input random variables are independent, when they may be correlated. In addition, material properties and design loads are often chosen to be conservative.

For the purpose of planning scheduled maintenance, it is often required to choose either 1% or 10% life, which is also called B1 or B10 life. Table 5 shows the confidence intervals of B1 and B10 life, along with lower- and upper-bounds of the 90% confidence intervals. In Cases 1 and 3, which employed a normal model, negative values for the life are calculated due to wrong assumptions on the model; the normal distribution can have negative lives. On the other hand, Cases 4 and 5 are reasonable because they only allow positive values.

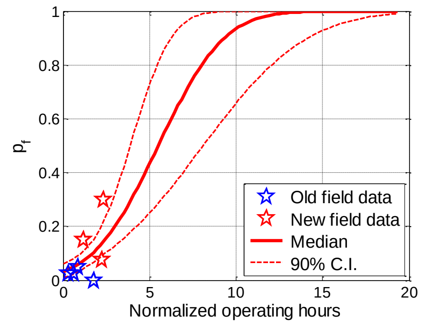
Since the type of distribution is assumed initially, it is advised to choose the most conservative one, which is Case 4 in this study. Although Case 4 does not have prior information compared to Case 5, its confidence interval has significantly been reduced using enough field data; the confidence intervals at 1% P_f are 0.245 and 0.267, respectively. However, this may not always be true. For example, Cases 4 and 5 are compared in Figure 13 and Table 6 when only the first data are used in Bayesian inference. Without having the prior distribution, Case 4 shows a significant uncertainty, and the lower-bound of 1% P_f is 0.019, which is overly conservative and costly due to frequent maintenance. Thus, when the number of field data is small or the field data have excessive noise, the prior information plays an important role to determine acceptable maintenance intervals. In summary, the ability to use the prior properly and the ability to choose appropriate statistical model are the main advantages of Bayesian inference over the regression method.



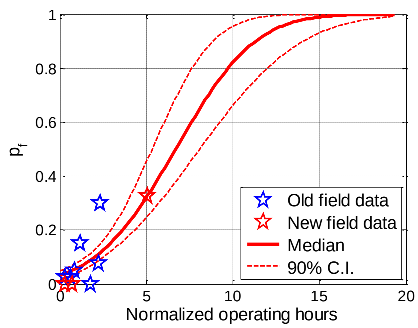
(a) using only the 1st data



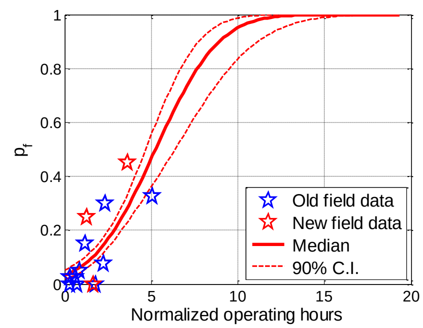
(b) using 1st~4th data



(c) using 1st~7th data

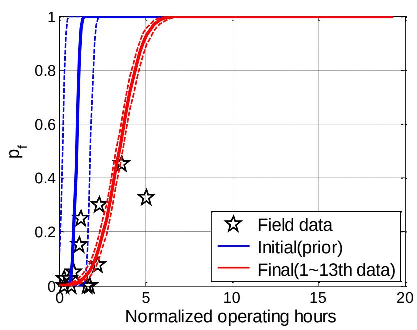


(d) using 1st~10th data

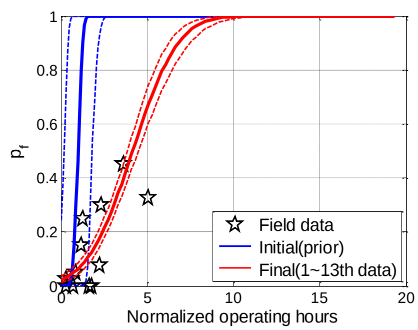


(e) using 1st~13th data

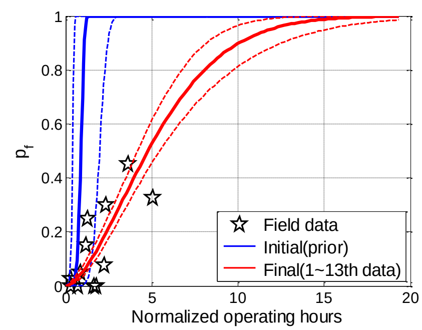
Figure 11: Updated Process of Case 1



(a) Case 2



(b) Case 3



(c) Case 5

Figure 12: Final updated distribution of fatigue life.

Table 5: Confidence interval of normalized fatigue life at the last stage

	1% P_f			10% P_f		
	5% lower	95% upper	interval	5% lower	95% upper	interval
Case 1	-2.261	-0.562	2.065	1.160	2.013	0.853
Case 2	0.658	1.272	0.615	1.872	2.320	0.448
Case 3	-0.459	0.429	0.888	1.461	2.019	0.558
Case 4	0.093	0.338	0.245	1.035	1.750	0.715
Case 5	0.229	0.496	0.267	1.223	1.769	0.546

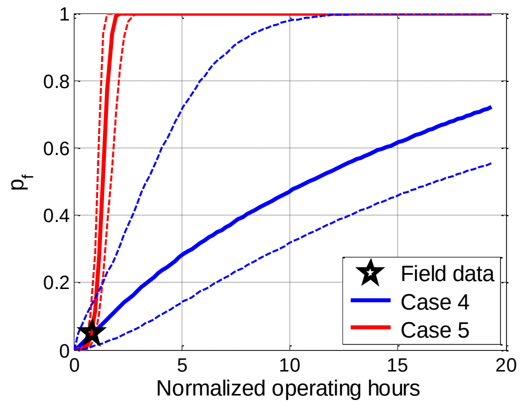


Figure 13: Comparison between Case 4 and Case 5

Table 6: Confidence interval of normalized fatigue life at the first stage

	1% P_f			10% P_f		
	5% lower	95% upper	interval	5% lower	95% upper	interval
Case 4	0.019	0.862	0.843	0.562	3.930	3.368
Case 5	0.473	0.776	0.303	0.789	1.155	0.366

5. CONCLUSIONS

In this paper, a Bayesian updating technique is presented, which incorporates statistical prediction with field data. By using MCMC simulation, samples of model parameters $\boldsymbol{\theta}$ (μ, σ or m, η) are drawn effectively, which are parameters of the fatigue life distribution. After obtaining samples for a joint posterior PDF of $\boldsymbol{\theta}$, the fatigue life prediction results are obtained, which have a CDF in confidence intervals due to the uncertainties of the model parameters. If there is specific prior information of model parameters, the precision of the posterior distribution is increased. In case the type of the distribution of the likelihood is not known a priori, it is advised to choose the most conservative type after examining several candidates as was found in this study.

6. ACKNOWLEDGEMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2008-02-010 and 2009-0081438)

7. REFERENCES

1. Voigt, M., Mücke, R., Vogeler, K., Oevermann, M., "Probabilistic Lifetime Analysis of Turbine Blades based on a combined direct Monte Carlo and Response Surface Approach", ASME Publication, Paper No. GT2004-53439, ASME Turbo Expo, June 2004, Vienna, Austria.
2. Weiss, T., Voigt, M., Schlums, H., Mücke, R., Becker, K.-H., Vogeler, K., "Probabilistic Finite-Element-Analysis on turbine blades", Paper No. GT2009-59877, ASME Turbo Expo, June 2009, Orlando, FL.
3. Marahleh, G., A. R. I. Kheder, and H. F. Hamad. Creep-Life Prediction of Service-Exposed Turbine Blades. Materials Science, 42(4), 2006.
4. Orchard, M., Wu, B., Vachtsevanos, G., "A Particle Filtering Framework for Failure Prognosis", World Tribology Congress III Sep. 12-16, 2005, Washington, D.C., USA
5. Marahleh, G., Kheder, A.R.I., Hamad, H.F., "Creep-Life Prediction of Service-Exposed Turbine Blades", Materials Science, Vol. 42, No. 4, 2006
6. Park, J., Nelson, D., "Evaluation of an energy-based approach and a critical plane approach for predicting constant amplitude multiaxial fatigue life", International Journal of Fatigue Vol. 22, 2000, pp. 23–39

7. Guo, H., Watson, S., Tavner, P., Xiang, J., Reliability analysis for wind turbines with incomplete failure data collected from after the date of initial installation", Reliability Engineering & System Safety Vol. 94, Issue 6, June 2009, pp. 1057-1063
8. Kim, N. H., S. Pattabhiraman, and L.A. Houck III. Bayesian Approach for Fatigue Life Prediction from Field Data. 2010 ASME Turbo Expo Gas Turbine Technical Congress and Exposition, Glasgow, Scotland, UK, June 14-18, 2010.
9. Andrieu, C., N. de Freitas, A. Doucet, and M. Jordan. An introduction to MCMC for Machine Learning. Machine Learning, 50:5-43, 2003.
10. Bayes, Thomas (1763). "An Essay towards solving a Problem in the Doctrine of Chances.". Philosophical Transactions of the Royal Society of London 53: 370–418.
11. Athanasios Papoulis (1984), Probability, Random Variables, and Stochastic Processes, second edition. New York: McGraw-Hill.
12. Gelman, A., J. B. Carlim, H. S. Stern, and D. B. Rubin. Bayesian Data Analysis (second ed.). New York, Chapman & Hall/CRC, 2004.
13. 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, 2009
14. Haldar, A., and Mahadevan, S., 2000, Probability, Reliability, and Statistical Methods in Engineering Design, John Wiley & Sons Inc., New York.