Comparison of Two Probabilistic Fatigue Damage Assessment Approaches Using Prognostic Performance Metrics

Xuefei Guan[1], Ratneshwar Jha[1], Yongming Liu[2], Abhinav Saxena[3], Jose Celaya[3] and Kai Goebel[4]**

[1] *Department of Mechanical & Aeronautical Engineering, Clarkson University, Potsdam, NY 13699-5725, USA*
guanx@clarkson.edu
rjha@clarkson.edu

[2] *Departments of Civil & Environmental Engineering, Clarkson University, Potsdam, NY 13699-5710, USA*
yliu@clarkson.edu

[3] *SGT, NASA Ames Research Center, Moffett Field, CA 94035, USA*
abhinav.saxena@nasa.gov
jose.r.celaya@nasa.gov

[4] *NASA Ames Research Center, Moffett Field, CA 94035, USA*
kai.goebel@nasa.gov

ABSTRACT

[*]In this paper, two probabilistic prognosis updating schemes are compared. One is based on the classical Bayesian approach and the other is based on newly developed maximum relative entropy (MRE) approach. The algorithm performance of the two models is evaluated using a set of recently developed prognostics-based metrics. Various uncertainties from measurements, modeling, and parameter estimations are integrated into the prognosis framework as random input variables for fatigue damage of materials. Measures of response variables are then used to update the statistical distributions of random variables and the prognosis results are updated using posterior distributions. Markov Chain Monte Carlo (MCMC) technique is employed to provide the posterior samples for model updating in the framework. Experimental data are used to demonstrate the operation of the proposed probabilistic prognosis methodology. A set of prognostics-based metrics are employed to quantitatively evaluate the prognosis performance and compare the proposed entropy method with the classical Bayesian updating algorithm. In particular, model accuracy, precision, robustness and convergence are rigorously evaluated in addition to the qualitative visual comparison. Following this, potential development and improvement for the prognostics-based metrics are discussed in detail.

1 INTRODUCTION

Fatigue damage is a critical issue in many structural and non-structural systems, such as aircraft, critical civil structures, and electronic components. The estimation of the reliability and remaining useful life (RUL) is important in condition-based maintenance of a system so that unit replacements can be done in time prior to catastrophic failures. Several physics-based models have been proposed in order to describe the fatigue process and predict the damage propagation; among these, Paris-type crack growth laws (Paris and Erdogan, 1963; Forman *et al.,* 1967; Walker, 1970) are most commonly used (Bourdin *et al.*, 2008). However, experimental data indicate that fatigue crack propagation is not a smooth, stable and well ordered process (Virkler *et al.,* 1979), thus a deterministic model is not capable of quantifying the crack growth subject to various uncertainties associated with the

[*] This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
[**] Corresponding author

fatigue damage. Uncertainties arising from a number of sources, such as measurement errors, model prediction residuals, and non-optimal parameter estimation, affect the quality of life predictions. These uncertainties need to be carefully included and managed in the prognosis process for risk management and decision-making.

In order to model the stochastic process of fatigue propagation and gain knowledge about a target system via monitoring system responses, probabilistic updating methods based on Bayes theorem have been used to evaluate the probability density functions (PDF) of input parameters using response measurements. For example, see (Madsen, 1997; Zhang and Mahadevan, 2000; Perrin *et al.*, 2007). Entropy methods, such as Maximum Entropy (MaxEnt) theorem (Jaynes, 1957; Jaynes, 1979; Skilling, 1988) and relative entropy methods (van Campenhout and Cover, 1981; Haussler, 1997), are alternative approaches for probability assignment and updating and have been used in many applications such as statistical mechanics (Caticha and Preuss, 2004; Tseng and Caticha, 2008), quantum physics (Hiai and Petz, 1991; Vedral, 2002), and fatigue prognosis (Guan *et al.,* 2009a,b). This paper has two objectives; the first is to develop a general prognosis approach based on maximum relative entropy (MRE) principles for probabilistic fatigue damage prognosis and compare it to the classical Bayesian approach, and the other is to explore prognosis metrics to evaluate prognosis performance quantitatively. One of the advantages of the proposed MRE approach is that the resulting confidence bounds are narrower compared to the classical Bayesian method, which is beneficial for decision making in a health management setting. The rest of the paper is organized as follows. In section 2, we review the classical Bayesian approach and formulate a general MRE updating and prognosis framework. Section 3 presents two application examples and methodology validation. Section 4 discusses algorithmic performance metrics and extends the two examples of section 3 in this context. Following that is the discussion and conclusion.

2 PROBABILISTIC MODEL UPDATING

In this section, both the classical Bayesian probability updating and a general MRE prognosis framework for fatigue damage problems are introduced. To evaluate the posterior probability distribution, Markov Chain Monte Carlo (MCMC) simulation is then introduced and employed in this framework to approximate the target distribution. For a generic inference problem with an uncertain parameter vector $\theta \in \Theta$, the posterior PDF of θ is inferred on the basis of three pieces of information: the prior knowledge about θ (the prior PDF of θ), the observation of a response event/variable $x \in \mathrm{X}$, and the known relationship between x and θ (the likelihood function based on physical/mathematical models). The search space for desired posterior PDF of θ is $\mathrm{X} \times \Theta$. Both Bayesian and MRE are capable of performing the search for an optimized posterior. However, these two approaches are based on different mechanisms. This is discussed in details in the following paragraphs.

2.1 Classical Bayesian model updating

Bayes' theorem provides a model for inductive inference or the learning process. A Bayesian posterior PDF is a measure of known information about parameters with uncertainty. Bayes' theorem is a means for combining the observation regarding the related parameters through the likelihood function (Gregory, 2005). Let $\mu(\theta)$ be a prior distribution, $p(\theta)$ be a posterior distribution and $L(x' \mid \theta)$ be the likelihood equation with parameter vector θ and response variable x'. According to Bayes' theorem, the optimized posterior that reflects the fact that we observed x' is

$$p(\theta) \propto \mu(\theta) \cdot L(x' \mid \theta) \tag{1}$$

The Bayesian formulation of a posterior is straightforward and has an enormous variety of applications, such as the model updating in fatigue analysis (Dey *et. al.,* 1998). Detailed derivation and demonstration can be found in the referred article and is not repeated here. One issue with the classical Bayesian approach is that only response observations can be used for updating. Other types of information, such as the expected value of a parameter and statistical moments, cannot be directly incorporated into the classical Bayesian framework. For example, coupon level experiment testing and failure analysis can reflect statistical features of batch productions. The statistical information can further help to improve the individual prognosis performance. In order to include this type of information in the probabilistic prognosis and model updating, an entropy-based probabilistic inference framework has been developed. Details are discussed below.

2.2 MRE approach for model updating

The relative information entropy, also referred to as Kullback-Leibler divergence (Kullback and Leibler, 1951), of two PDFs $f_1(\theta)$ and $f_2(\theta)$ is defined as,

$$I(f_1 : f_2) = -\int_\Theta d\theta \cdot f_1(\theta) \log\left(f_1(\theta) / f_2(\theta)\right) \tag{2}$$

where θ is the parameter vector and Θ is the associated vector space. The axioms of maximum entropy indicate that the form of Eq. (2) is the unique entropy representation for inductive inference (Skilling, 1988).

The three axioms are:

1. Locality – Local information has local effects.
2. Coordinate invariance – The ranking of the two probability densities should not depend on the system coordinates. This indicates that the coordinates carry no information.
3. Consistency for independent subsystem – For a system composed of subsystems that are believed to be independent; it should not make a difference whether the inference treats them separately or jointly.

Using the similar notation above, let $\mu(x,\theta)$ be a prior joint distribution and $p(x,\theta)$ be a posterior joint distribution. According to the entropy axioms, the selected joint posterior is the one that maximizes the relative entropy $I(p:\mu)$ in Eq. (3), subject to all available constraints, such as statistical moments and measures of a response event/variable.

$$I(p:\mu) = -\int dxd\theta \cdot p(x,\theta)\log\left(p(x,\theta)/\mu(x,\theta)\right) \quad (3)$$

In Eq. (3), $\mu(x,\theta) = \mu(\theta)\mu(x \mid \theta)$ contains all prior information. $\mu(x \mid \theta)$ is the likelihood function and $\mu(\theta)$ is the prior PDF. The same relationship applies to the joint posterior $p(x,\theta)$. When new information is available in the form of a constraint, the updating procedure will search in the space of $X \times \Theta$ for a posterior which maximizes $I(p:\mu)$. Measurements of the response variable x can be used to perform the updating, which is performed in a similar way as the classical Bayesian updating. The benefit of MRE updating is that it can incorporate other information for inference, which cannot be included in the classical Bayesian updating. For example, the expected value of a function of θ or the empirical judgment on the mean value of θ can be used in MRE updating (Giffin and Caticha, 2007). This flexibility of applicable information can pose more constraints on a posterior thus yield a more accurate result given that those constraints are justified. Following the derivation of MRE posterior (Caticha and Giffin, 2006), if a new observation x' is obtained, the posteriors that reflect the fact x is now known to be x' is a constraint such that

$$c_1 : p(x) = \int d\theta \cdot p(x,\theta) = \delta(x - x') \quad (4)$$

Other information in the form of moment constraints, such as the expected value of some function $g(\theta)$, can be formulated as

$$c_2 : \int dxd\theta \cdot p(x,\theta)g(\theta) = \langle g(\theta) \rangle = G \quad (5)$$

The normalization constraint is

$$c_3 = \int dxd\theta \cdot p(x,\theta) = 1 \quad (6)$$

Maximizing Eq. (3), subject to constraints Eqs. (4-6), the desired posterior can be obtained as

$$p(\theta) \propto \mu(\theta)\mu(x' \mid \theta)e^{\beta \cdot g(\theta)}. \quad (7)$$

The detailed derivation of Eq. (7) and the computation of β can be found in (Guan *et. al.*, 2009a). The right side of Eq. (7) consists of three terms. $\mu(\theta)$ is the parameter prior, $\mu(x' \mid \theta)$ is the likelihood, and $e^{\beta \cdot g(\theta)}$ is the exponential term introduced by moment constraints. Eq. (7) is similar to Bayesian posterior except for the additional exponential term. This equation further indicates that, if no moment constraint is available, i.e., β is zero, MRE updating will be identical to Bayesian updating. In other words, Bayesian updating is a special case of MRE updating. Similar to that of a Bayesian updating problem, the likelihood function is usually constructed using the physics-based model depending on different realistic applications.

2.3 Fatigue mechanism model and likelihood function construction

In this section, a general procedure of constructing the likelihood equation is presented. Let d be a response variable measure of our target system and y be the prediction value of a prediction model M. If the model is sufficiently accurate to describe the system output, the observed value is equal to model prediction value, i.e. $y = d$. However, noise and errors usually exist for both modeling and measurements. Incorporating a modeling error term e and a measurement error term ε into consideration and assuming both errors are additive, we have,

$$d = M(x \mid \theta) + e + \varepsilon; \quad (9)$$

where $M(x \mid \theta)$ is the model prediction and θ is the model parameter vector. The two uncorrelated error terms e and ε usually can be described using two independent zero-mean normal variables. Replacing the two error terms with a total error term $\tau = (e+\varepsilon) \sim Normal(0,\sigma_\tau)$, the likelihood function of multiple observations can be constructed as

$$L(d_{1\ldots n} \mid \theta) = \frac{1}{\left(\sqrt{2\pi}\sigma_\tau\right)^n} \exp\left\{-\sum_{i=1}^{n} \frac{\left[d_i - M(x \mid \theta)\right]^2}{2\sigma_\tau^2}\right\}. \quad (10)$$

, where $L(d_{1\ldots n} \mid \theta)$ is the joint PDF of observations given parameter θ. Substituting Eq. (10) in Eq. (7), an MRE posterior of θ is derived to be

$$p(\theta) \propto \mu(\theta) \frac{1}{\sigma_\tau^{\;n}} \exp\left\{-\sum_{i=1}^{n} \frac{\left[d_i - M(x \mid \theta)\right]^2}{2\sigma_\tau^2}\right\} e^{\beta g(\theta)}. \quad (11)$$

For fatigue damage model $M(x \mid \theta)$, various deterministic models have been proposed to describe the fatigue crack accumulation, among which Paris type of laws are commonly used in cycle based fatigue crack growth calculation. In this study, Paris model (Paris and Erdogan, 1963) is employed for illustration purposes. In a realistic situation, other model might be adopted accordingly. Let a be the crack length, N be the number of cycles, the Paris' law reads,

$$\frac{da}{dN} = c(\Delta K)^m = c\left(\Delta\sigma \cdot \sqrt{\pi a} \cdot F(a)\right)^m \quad (12)$$

where c and m are material constants, ΔK is the variation of stress intensity factor in one cycle of stress amplitude $\Delta\sigma$, and $F(a)$ is the geometric correction factor. The crack size can be calculated by solving Eq. (12) numerically given c, m, and N. Early studies show that $\log(c)$ follows a normal distribution and m follows a truncated normal distribution (Kotulski, 1998). Assuming $\log(c)$ and m are independent variables and combining Eq. (12) with Eq. (11), the joint posterior can be expressed as Eq. (13) using MRE formulation.

$$\begin{aligned} p_i(\log(c), m) \propto \\ \frac{1}{\sigma_c} \exp\left\{-\frac{1}{2}\left(\frac{\log(c)-\zeta_c}{\sigma_c}\right)^2 + \beta_c g_c(\log(c))\right\} \cdot \\ \frac{1}{\sigma_m} \exp\left\{-\frac{1}{2}\left(\frac{m-\zeta_m}{\sigma_m}\right)^2 + \beta_m g_m(m)\right\} \cdot \\ \frac{1}{\sigma_\tau^n} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{d_i - M_i(a_i \mid c, m, N_i)}{\sigma_\tau}\right)^2\right\} \end{aligned} \quad (13)$$

Setting β to zero in Eq. (13) gives Bayesian formulation of the same problem. The PDF of one parameter can be obtained by integrating over the rest of the parameters. But for a large dimension parameter space, more general and computationally efficient methods, such as sampling techniques, might be applied.

2.4 MCMC simulation method

Direct evaluation of the PDF in Eq. (13) is difficult because of the multi-dimensional integration needed for normalization. In order to circumvent the direct evaluation of Eq. (13), Markov Chain Monte Carlo sampling technique is used in this study. MCMC was first introduced by (Metropolis *et al.,* 1953) as a method to simulate a discrete-time homogeneous Markov chain. The merit of MCMC is that it overcomes the normalization of Eq. (13) and ensures that the state of the chain converges to the target distribution after a large number of steps from an arbitrary initial start. The widely used random walk algorithm, Metropolis-Hastings algorithm (Hastings, 1970), is summarized here.

The transition between two successive samples x_t and x_{t+1} is defined by Eq. (14).

$$x_{t+1} = \begin{cases} \tilde{x} \sim q(X \mid x_t) & \textit{with probability } \alpha(x_t, \tilde{x}) \\ x_t & \textit{else} \end{cases}, \quad (14)$$

where $q(X \mid x_t)$ is the transition distribution, and $\alpha(x_t, \tilde{x}) = \min(1, r)$ is the acceptance probability. The Metropolis ratio r is defined as,

$$r = \frac{p(\tilde{x})}{p(x_t)} \frac{q(x_t \mid \tilde{x})}{q(\tilde{x} \mid x_t)} \quad (15)$$

where $p(\cdot)$ is the posterior probability representation. In our case, $p(\cdot)$ is computed using Eq. (13). For a symmetric transition distribution of $q(\cdot)$, such as a normal distribution, the property of $q(x_t \mid \tilde{x}) = q(\tilde{x} \mid x_t)$ simplifies Metropolis ratio in Eq. (15) to $r = p(\tilde{x})/p(x_t)$. In this paper, 100,000 posterior samples of $(\log(c), m)$ are generated with a 5% burn-in period using a normal transition distribution. Additionally, the moment information of these samples is then integrated into the proposed MRE updating procedure.

3 APPLICATION EXAMPLES

Two fatigue crack growth experimental datasets are used to demonstrate the proposed MRE updating procedure and show the benefits of this approach.

3.1 Virkler's 2024-T3 aluminum alloy experimental data

An extensive fatigue crack growth data under constant loading for Al 2024-T3 plate specimens with center through cracks was collected in (Virkler *et al.,* 1979). The dataset consists of 68 fatigue crack growth trajectories and each trajectory contains 164 measurement points. All specimens have the same geometry, i.e., an initial crack size $a_i = 9mm$, length $L = 558.8mm$, width $w = 152.4mm$ and thickness $d = 2.54mm$. The loading information is $\Delta\sigma = 48.28MPa$ and stress ratio $R = 0.2$. The geometry correction factor for Virkler's experiments is $F(a) = 1/\sqrt{\cos(\pi a / w)}$. Kotulski (1998) reported the statistical information of the parameters in Paris' law, namely, mean values $\zeta_c = \langle \log(c) \rangle = -26.155$ and $\zeta_m = \langle m \rangle = 2.874$ with standard deviations $\sigma_c = 0.968$ and $\sigma_m = 0.164$, respectively. Assuming the total error term is $\sigma_\tau = 0.1mm$ (see Eq. (8)) and substituting the statistics information into Eq. (12) with $g_c(c) = \log(c)$ and $g_m = m$, the updating procedure can be performed when observation data become available.

One crack growth trajectory in Virkler's dataset was selected arbitrarily for fatigue crack length prediction updating from (Ostergaard and Hillberry, 1983). Five data points in the early stage of the crack propagation are randomly chosen to represent the measured ground truth values of crack length a obtained from health monitoring system or nondestructive inspection. These data points are listed in Table 1.

Predictions from MRE updating and Bayesian updating procedures are shown in Figure 1. As can be seen, MRE updating gives a narrower prognosis confidence interval as compared to classical Bayesian updating. It further justifies that the additional moment constraints imposed on the posterior yield a more compact results.

Table 1: Data used for updating (Virkler's dataset)

Number	Crack size (mm)	Cycle
1	9.7330	21269
2	10.5272	42734
3	11.2557	56392
4	12.1708	73161
5	15.0549	110487

3.2 McMaster's 2024-T351 aluminum alloy experimental data

A large set of 2024-T351 aluminum alloy experimental data under constant and variable loading conditions were obtained in (McMaster and Smith, 1999). The experimental data of center-cracked specimens with length $L = 250mm$, width $w = 100mm$ and thickness $t = 6mm$ under constant loading $\Delta\sigma = 65.7MPa$ and stress ratio $R = 0.1$ are used in this paper. Priors of the parameters are obtained by $\log(da/dN) \sim \log(\Delta K)$ regression using the experimental data. Five data points as shown in Table 2 are chosen arbitrarily to be used as sensor measurements from health monitoring system in order to perform the updating. The prior PDFs is artificially set as $\zeta_c = -26.5$ and $\zeta_m = 2.9$, which is not sufficiently accurate enough to match the experimental records as seen in Figure 2.

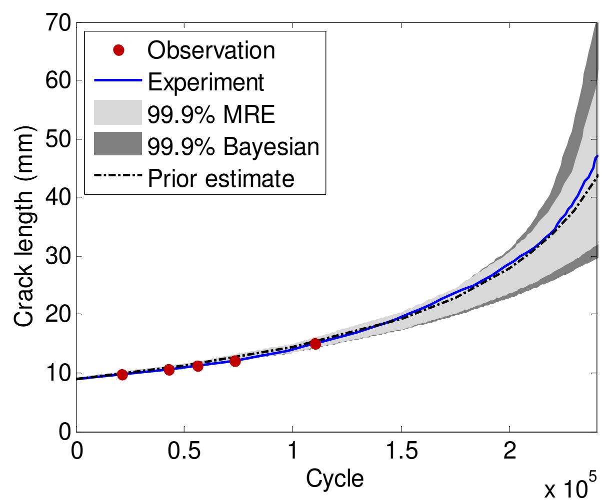

Figure 1: MRE and Bayesian prognosis (Virkler's dataset)

The predictions of MRE and Bayesian updatings are further shown in Figure 2, where interval predictions obtained by MRE updating are much narrower than that by Bayesian updating.

Table 2: Data used for updating (McMaster's dataset)

Number	Crack size (mm)	Cycle
1	11.3611	4875
2	11.9282	8475
3	12.3254	11550
4	13.8563	17775
5	14.8771	21375

MRE updating shows the advantages over Bayesian updating in two application examples (visual observation). This is more likely due to the additional statistical moment constraints of MCMC samples added to posteriors. To quantify the performance, prognosis metrics need to be considered to provide a rigorous comparison between MRE updating and Bayesian updating as given below.

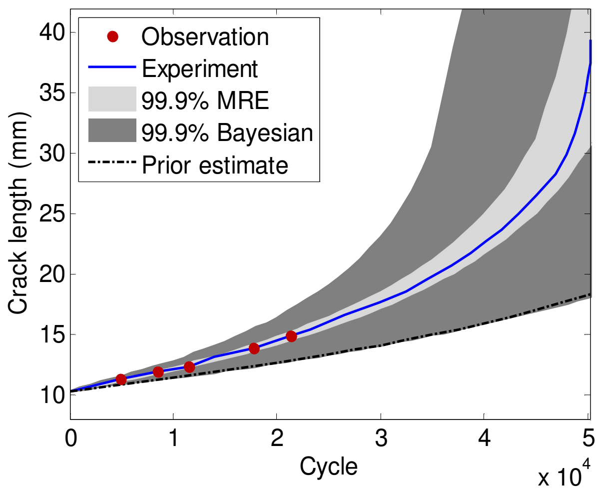

Figure 2: MRE and Bayesian prognosis (McMaster's dataset)

4 METRIC-BASED PERFORMANCE EVALUATION OF THE MODEL

Various metrics are available to quantify the performance of prognosis algorithms (Saxena *et al.,* 2008). In this section, classical error based statistical measures and several prognosis metrics are applied to quantify the prediction performance of application examples in the previous section.

4.1 Statistical metrics

Metrics, such as mean squared error (MSE), mean absolute percentage error (MAPE), average bias, sample standard deviation (STD), and their variations are widely used in medicine and finance fields where large datasets are available for statistical data analysis (Saxena *et al.,* 2008). The results for those classical metrics shown in Table 3 and Table 4, (rows 1-4) are computed using the prediction residuals (the difference between actual RUL and predicted RUL) obtained after the fifth updating. The proposed MRE approach shows its advantages over Bayesian method in all cases.

4.2 Prognosis metrics

The metrics mentioned in Section 4.1 are general purpose metrics and were not specifically designed for prognosis. In (Saxena *et al.,* 2009a) authors proposed several metrics, such as Prognostic Horizon (PH), Alpha-Lambda (α-λ) Performance, Relative Accuracy (RA), Cumulative Relative Accuracy (CRA), and Convergence; that were designed specifically for prognosis to incorporate the prediction distributions and the structure of the prognostics process. These metrics help assess how well prediction estimates improve over time as more measurement data become available. For readers' reference, we present a brief definition of these metrics here.

1. Prognostic Horizon is defined as the length of time before end-of-life (EoL) when an algorithm starts predicting within specified accuracy limits. These limits are specified as $\pm\alpha\%$ of the true EoL.

2. α-λ Accuracy determines whether predictions from an algorithm are within $\pm\alpha\%$ accuracy of the true RUL at a given time instant, specified by the parameter λ. For instance a $\lambda = 0.5$ would specify midway between the first time a prediction is made and EoL.

3. Relative Accuracy quantifies the percent accuracy with respect to actual RUL at a given time (specified by λ). It's an accuracy measure normalized by RUL, signifying that predictions closer to EoL should be more accurate and precise.

4. Cumulative Relative Accuracy is a weighted average of RAs computed at different time instances. Weights can be assigned to the predictions based on how critical they become as EoL approaches, and hence the accuracy of the predictions.

5. Convergence quantifies the rate at which any performance metric of interest improves to reach its desired value as time passes by.

For more description, implementation details and application examples on these metrics; the reader may referred to (Saxena *et al.*, 2009a). In general, these metrics were designed to capture the time varying aspects of prognostics. As more data become available prognostic estimates get revised. It is, therefore, important to track how well an algorithm performs as time passes by as opposed to evaluating performance at one specific time instant only. Further, these metrics also incorporate the notion of increased criticality as EoL approaches, which imply that a successful prognosis algorithm should improve as the system approaches its EoL.

In this paper we compare the two approaches based on Bayesian and MRE updating. In addition to evaluating performance based on prognosis metrics, we also include some classical statistical metrics. For this purpose, in our approach we include an additional updating point from the end of time series to establish EoL and compute the RUL curves. Results obtained from this evaluation exercise are presented next.

Performance results for Virkler's dataset

The visual results for PH and α-λ accuracy are shown in Figure 3. Numerical values of those metrics are listed in Table 3. For computing CRA (see Table 3), the starting point is cycle zero because the specimens have initial cracks. We evaluated RA at 20, 40, 60, and 80% of EoL and did not use weighting factors. This assumes that relative accuracy is equally weighted at all time instants. Though, this may not always be preferable, a simplistic evaluation was carried out to observe the natural behavior of the algorithm itself.

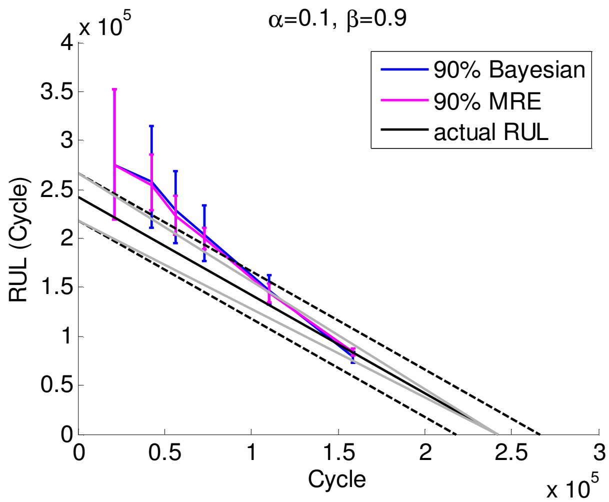

Figure 3: Performance comparison for PH and α-λ accuracy at α=0.1 (10% error bound) on Virkler's dataset

Figure 3 compares the prediction horizon for the two algorithms with 10% error bound around EoL value. Using the strict definition for PH as laid out in (Saxena *et al.,* 2009b), we observed that MRE yields a larger PH. The plot of PH performance in Figure 3 shows that 90% MRE interval prediction enters the 90% accuracy zone at the fifth updating, while Bayesian prediction enters the zone at the sixth updating showing that MRE is slightly better than Bayesian. It is worth mentioning that there is no specific reason to choose $\beta = 0.9$, which is very conservative and strict. Typically 50% corresponds to evaluating mean value being inside the alpha bounds. It depends on specific reliability requirement and actual application constraints to pick up a proper value. In general, it indicates that, for engineering practice, the proposed MRE can give an informative prediction at an earlier stage of the whole lifecycle.

Table 3 Comparison of metrics between MRE and Bayesian approaches (Virkler's dataset)

Metric	**MRE**	**Bayesian**
MAPE	8.66	10.93
Average Bias (cycles)	10956.27	14051.92
STD (cycles)	7628.77	9115.78
MSE($cycle^2$)	178.23×10^6	280.5×10^6
PH(cycle)	132016	83583
RA $_{\lambda=0.4}$	0.92	0.89
CRA	0.89	0.87
Convergence	74365.72	77349.24

Looking at Table 3 one can see that on Virkler's dataset MRE performs better than Bayesian approach under all performance measures. One must note that although classical metrics conclude the same as the new prognostics metrics, they do not take into account the time varying nature of the prognostics and hence may not always be useful in practice.

Performance results for McMaster's dataset

Next, we perform a similar analysis for the McMaster's dataset. The visual results for PH and α-λ accuracy metrics comparing Bayesian and MRE updating are shown in Figure 4. The rest of the metrics are included in Table 4. Looking at these results, the general conclusion about the superior performance of the MRE procedure from Virkler's dataset is further strengthened. The MRE's superior performance over Bayesian approach is attributed to the ability to incorporate additional knowledge about the system using additional constraints.

Table 4 Comparisons of metrics between MRE and Bayesian approaches (McMaster's dataset)

Metric	**MRE**	**Bayesian**
MAPE	4.06	22.53
Average Bias (cycles)	418.76	4561.93
STD (cycles)	1413.53	6888.38
MSE ($cycle^2$)	2.17×10^6	68.26×10^6
PH(cycle)	32475	N/A
RA $_{\lambda=0.4}$	0.99	0.86
CRA	0.95	0.87
Convergence	13757.94	22175.16

For this dataset, these metrics clearly distinguish the two approaches and show better outcomes from the MRE method. For example, the PH and α-λ performance metrics shown in Figure 4 present clear visual comparisons, e.g., the prognosis bounds obtained by MRE enters the cone area at the fourth updating which is earlier than that of Bayesian.

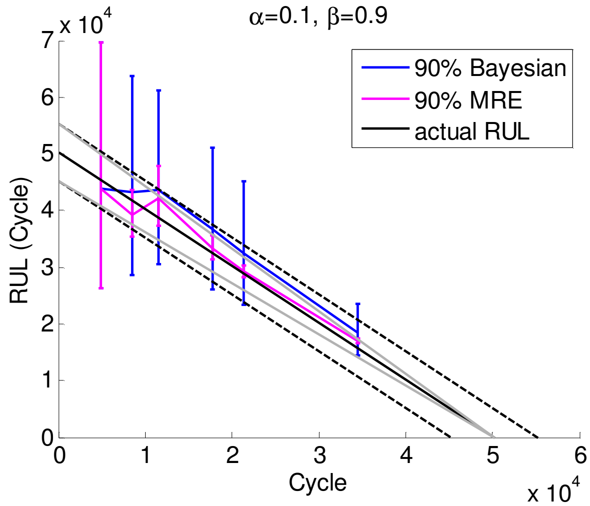

Figure 4: Performance comparison for PH and α-λ accuracy at α=0.1 (10% error bound) on McMaster's dataset

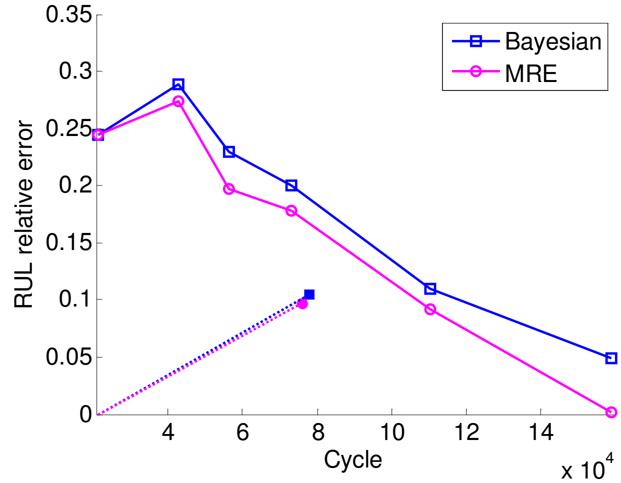

Figure 5: Comparison of convergence performance on Virkler's dataset

5 DISCUSSION

As observed in the previous section, there are a few aspects where these metrics can be further enhanced to improve performance evaluation.

The significant difference between the PHs for the two algorithms may also be an artifact of the frequency at which these algorithms make a prediction.

We also observed that in a probabilistic prognosis updating scheme, the selection of priors may produce different prognosis results and affect the performance. Consequently, different updating methods may exhibit different robustness with inappropriate priors. Next, we discuss some of these issues as they relate to prognosis metrics.

5.1 Convergence metric

The convergence metric computes a value to quantify how fast prognostic estimates improve and converge towards the ground truth. A metric like convergence is meaningful only if an algorithm improves with time and passes various criteria defined by other prognostic metrics. For example, the convergence in terms of RUL relative error (RE) defined in Eq. (16), which is the difference between an actual response measure (R) and the inferred value (R_0) divided by the actual response measure, for Virkler's dataset, shows a monotonic decreasing trend after the second update (Figure 5). Both MRE and Bayesian methods show diverging trends for McMaster's dataset (Figure 6).

$$RE := (R_0 - R)/R \tag{16}$$

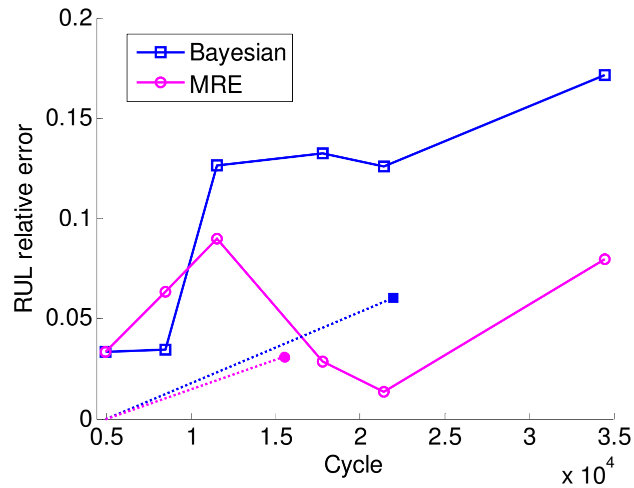

Figure 6: Comparison of convergence performance on McMaster's dataset

The results (converging and diverging trends) suggests that a metric like convergence will not make complete sense if the algorithms do not show improvements with time and hence additional fine tuning of the algorithms is required. The length of the dash line (Figure 5 and Figure 6) between the coordinate origin and the centric point of the area covered by the RE curves serves as a quantitative value of convergence metric. The details of that can be found in (Saxena *et al.*, 2009a). It is worth mentioning that different applications may require different measures instead of RE and the choice of measures depends on which aspect of the algorithmic convergence we would like to investigate.

5.2 Robustness metric

From the above examples, it is shown that the selection of a prior PDF is critical for a meaningful prognosis using probabilistic updating schemes such as Bayesian and MRE. An inaccurate prior may render a poor prediction of RUL. For example, when the prior prediction shown in Figure 7 is very different from the actual distribution, the Bayesian predictions lead to inaccurate estimates with very wide confidence bounds. The MRE updating approach performs well while using the same inaccurate prior distributions.

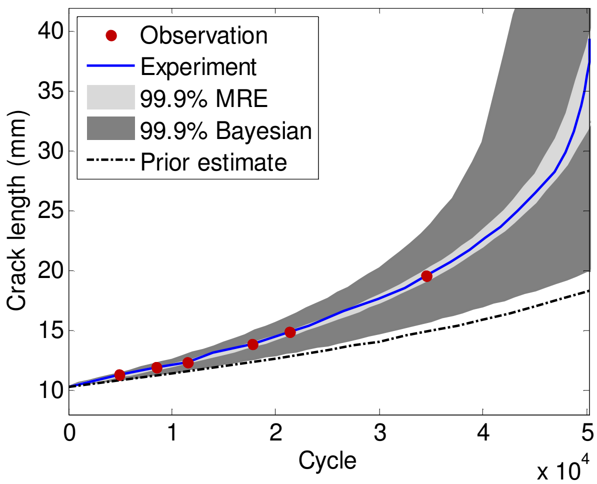

Figure 7: MRE and Bayesian prognosis with an inaccurate prior (McMaster's dataset)

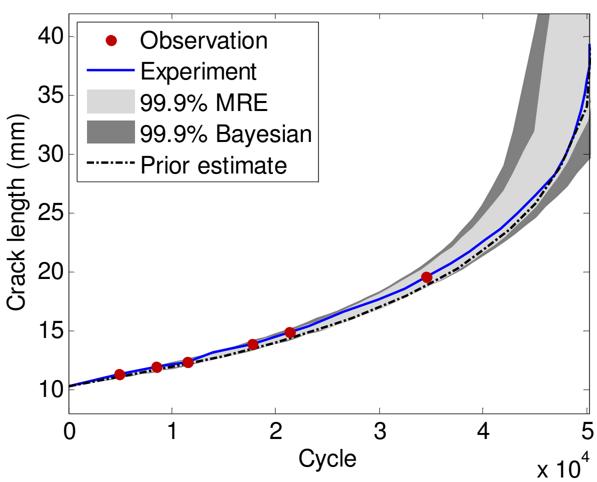

Figure 8: MRE and Bayesian prognosis with an accurate prior (McMaster's dataset)

On the other hand, starting with a relatively accurate prior prediction, both MRE and Bayesian give similar predictions as shown in Figure 8. It is valuable to define a robustness metric that can quantify the sensitivity of different algorithms with respect to the algorithm parameters, such as prior distribution, initial conditions, and training data size.

A preliminary study on the robustness metric is shown below. The basic idea is to quantify the change of prognosis confidence bounds due to the changing of algorithm parameter values. The range of investigated parameter is first defined based on specific application requirements (e.g., 10% variation around the mean value) or based on the underlining physics requirement (e.g., parameter should be non-negative). In this paper, we used a parameter η to specify the range of interested parameter (i.e., the parameter is in the range of mean$\pm\eta$). For a robust algorithm, the change of algorithm parameters will not affect the prognosis confidence bounds much. In view of this, the area in a confidence bound vs. parameter variation plot is a good indication of algorithm robustness (shaded area in Figures 9 and 10). In order to perform the metric comparison across different parameter spaces, a normalization process is proposed. A reference area is defined by specifying an allowable prediction error level (e.g., ±20% in the current investigation). This allowable level is expressed using parameter δ. The reference area can be calculated as $4\eta\delta$ and is shown as the area by the dashed lines in Figures 9 and 10. Mathematically, the robustness metric R_b can be defined as

$$R_b = \frac{\int_{Xmean-\eta}^{Xmean+\eta} f(x)dx}{4\eta\delta} \tag{17}$$

where x is the investigated algorithm parameter and $f(x)$ is the confidence bound variation function with respect to x. The physical meaning of Eq. (17) is the shaded area normalized by the dashed line area in Figures 9 and 10.

The performance of the two updating algorithms is investigated using the above mentioned robustness metric for Virkler's dataset first. In this case, $\eta = 0.02$ and $\delta = 0.2$ is used to investigate the parameter m in the crack growth law (Eq. (12)). The mean value of m is 2.874. All predictions are made after six updatings and the 99% confidence bounds is shown in Figure 9. The robustness metric (Eq. (17)) of the Bayesian approach is 2.6 while that for the MRE approach is 0.7. The similar investigation if performed for McMaster's dataset with the mean value of m equaling to 2.9. The robustness metric of the Bayesian and MRE approach are 3.0 and 0.4, respectively. The metric configuration and the visual comparison for McMaster's dataset are shown in Figure 10.

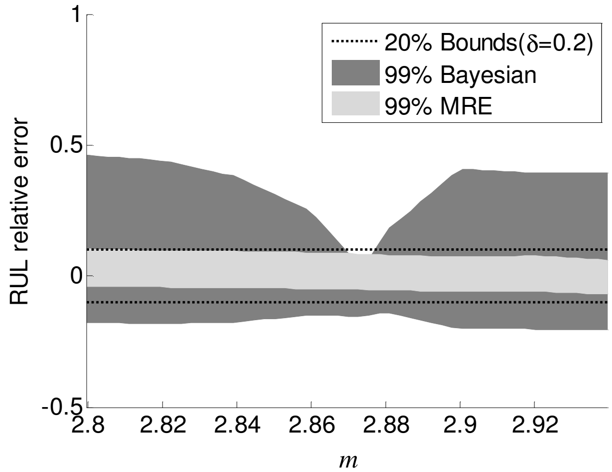

Figure 9: Comparison of robustness metric after six updatings with varying values of *m* in prior PDF (Eq. 13) for parameter m (Virkler's dataset)

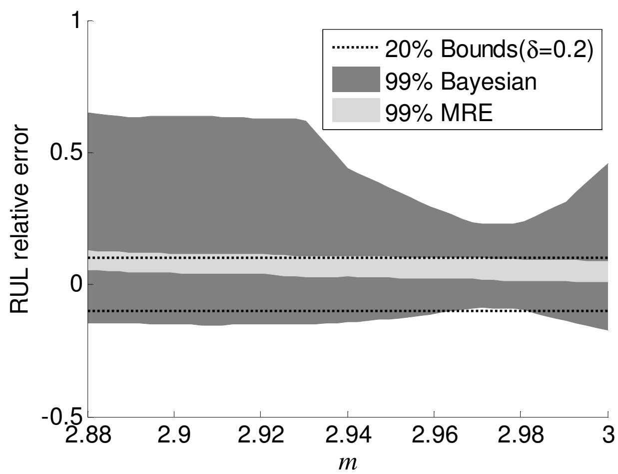

Figure 10: Comparison of robustness metric after six updatings with varying values of *m* in prior PDF (Eq. 13) for parameter m (McMaster's dataset)

From the above results we can see that, under this specific parameter configuration, MRE exhibits more robust against the variation of *m* in prior PDFs. It may be valuable in a practical perspective since most of the time an accurate prior is difficult to obtain with a limited data source. One issue with this robustness metric is that it does not reflect how the performance changes with time. More complicated metrics based on this idea maybe developed by adding another dimension to record the performance variation with time. Since Bayesian updating algorithms are associated with many factors, such as the total number of updating points, the training data size, noise levels, etc., further studies are needed to establish such concepts regarding the algorithmic robustness.

To make further comparison between different Bayesian updating and prognosis approaches, more data points and even the whole dataset can be used as observation data to see with enough measures of response whether MRE and Bayesian give similar prognosis results and show convergence. Though in practice it is more desirable to get an early stage accurate prognosis, it is necessary to explore the characteristics of different updating algorithms using experimental data as we showed in previous sections.

6 CONCLUSION

A general framework for probabilistic prognosis using maximum entropy approach, MRE, is proposed in this paper to include all available information and uncertainties for RUL prediction. Prognosis metrics are used for model comparison and performance evaluation. Several conclusions can be drawn based on the results in the current investigation:

- The proposed MRE updating approach results in more accurate and precise prediction compared with the classical Bayesian method.
- The classical Bayesian method is a special case of the proposed MRE approach and MRE approach is more flexible to include additional information for inference, which cannot be handled by the classical Bayesian method.
- The prognosis metrics can be successfully used for algorithm comparison and can give quantitative values in model (algorithm) performance evaluation.
- A robustness metric measuring the updating algorithmic sensitivity to prior uncertainty is proposed and applied to both Bayesian and MRE updating approaches. The application examples show that MRE exhibits more robustness against the uncertainty introduced by parameter distribution priors in the sense of prognosis performance.
- It is important to realize when to apply these metrics to arrive at meaningful interpretations. For instance, use of the convergence metric makes sense only when the algorithm predictions converge (get better) with time.

ACKNOWLEDGMENT

The research reported in this paper was supported in part by the NASA ARMD/AvSP IVHM project under NRA NNX09AY54A. The support is gratefully acknowledged.

NOMENCLATURE

$I(\cdot)$ Relative information entropy

$\mu(\cdot)$ Prior PDF
$p(\cdot)$ Posterior PDF
$L(\cdot)$ Likelihood function
$M(\cdot)$ Model prediction of crack length
$F(\cdot)$ Geometric correction factor
N_i Number of cycles
d_i Actual crack length

REFERENCES

(Bourdin *et. al.,* 2008) B. Bourdin, G. A. Francfort and J. J. Marigo (2008). *The Variational Approach to Fracture*, Netherlands: Springer.

(Caticha and Preuss, 2004) A. Caticha and R. Preuss (2004). Maximum entropy and Bayesian data analysis: Entropic prior distributions. *Physical Review E*. vol. 70(4), 046127.

(Caticha and Giffin, 2006) A. Caticha and A. Giffin (2006). Updating probabilities. In A.M. Djafari (Ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering, AIP Conference Proceedings*, 31: 872.

(Dey, 1998) D. Dey, P. Mueller, and D. Sinha (1998). *Practical Nonparametric and Semiparametric Bayesian statistics*. Springer, pp. 339-350.

(Forman *et. al.,* 1967) R. G. Forman, V. E. Kearney, and R. M. Engle (1967). *Numerical analysis of crack propagation in a cyclic-loaded structure*. Journal of Basic Engineering, Transactions of American Society of Mechanical Engineers, vol 89(D), pp. 459-464.

(Giffin and Caticha, 2007) A. Giffin and A. Caticha (2007). Updating probabilities with data and moments. In K. Knuth (Ed.), *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, AIP Conference Proceedings, 954:74

(Gregory, 2005) *Bayesian Logical Data Analysis for the physical science*. 1st ed. Cambridge University Press.

(Guan *et. al.,* 2009a) X. Guan, R. Jha, and Y. Liu (2009). Probabilistic fatigue damage prognosis using maximum relative entropy approach. *Journal of Intelligent Manufacturing*. DOI: 10.1007/s10845-009-0341-3.

(Guan *et. al.,* 2009b) X. Guan, Y. Liu, A. Saxena, J. Celaya, and K. Geobel (2009). Entropy-based probabilistic fatigue damage prognosis and algorithm performance comparison. *Annual Conference of the Prognostics and Health Management. PHM 2009.* 28 September – 2 October 2009.

(Hastings, 1970) W. K. Hastings (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, vol. 57(1), pp. 97-109.

(Haussler, 1997) D. Haussler (1997). A general minimax result for relative entropy. *IEEE transactions on information theory*, vol 43(4), pp. 1276-1280.

(Hiai and Petz, 1991) F. Hiai and Denes Petz (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in Mathematical Physics*, vol. 143(1), pp. 99-114.

(Jaynes, 1957) E. T. Jaynes (1957). Information theory and statistical mechanics. *Physical Review*, vol. 106, pp. 620-630 and vol. 108: 171-190.

(Jaynes, 1979) E.T. Jaynes (1979). Where do we stand on Maximum Entropy. In R.D. Levine and M. Tribus (Eds.) *The Maximum Entropy Formalism*. M.I.T. Press, Cambridge, MA. pp. 15-119.

(Kotulski, 1998) Z. A. Kotulski (1998). On efficiency of identification of a stochastic crack propagation model based on Virkler experimental data. *Archives of Mechanics*, vol. 50(5), pp. 829-847.

(Kullback and Leibler, 1951) S. Kullback and R. A. Leibler (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, vol. 22, pp.79-86.

(Madsen, 1997) H. Madsen (1997). Stochastic modeling of fatigue crack growth and inspection. In C.G. Soares (Ed.), *Probabilistic methods for structure design, collection Solid Mechanics and its applications*, pp. 59-83. Netherlands: Kluwer Academic Publishers.

(McMaster and Smith, 1999) F. J. McMaster and D. J. Smith (1999). Effect of load excursions and specimen thickness on crack closure measurements, *Advances in Fatigue Crack Closure Measurements and Analysis: Second Volume, ASTM STP 1343*, pp 246-264. R. C. McClung and J. C. Newman, Jr., Eds., West Conshohocken, Pennsylvania: American Society for Testing and Materials.

(Metropolis *et al.*, 1953) N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, vol. 21(6), pp.1087-1092.

(Ostergaard and Hillberry, 1983) D. F. Ostergaard and B. M. Hillberry (1983). Characterization of the variablility in fatigue crack propagation data, In J. M. Bloom and J. C. Ekvall (Eds.), *Probabilistic Fracture Mechanics and Fatigue Methods: Application for Structural Design and Maintenance*, ASTM STP 798, pp. 97-115.

(Paris and Erdogan, 1963) P. Paris and F. Erdogan

(1963). A critical analysis of crack propagation laws. *Trans. The ASME, J. Basic Eng.* pp. 528-534.

(Perrin *et. al.*, 2007) F. Perrin, B. Sudret, and M. Pendola (2007). Bayesian updating of mechanical models – Application in fracture mechanics. *In 18ème Congrès Français de Mécanique*, CFM 2007, Grenoble.

(Saxena *et al.*, 2008) A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher (2008). *Metrics for evaluating performance of prognostic techniques.* International Conference on Prognostics and Health Management, PHM 2008. 6-9 Oct. 2008, pp: 1-17.

(Saxena *et al.,* 2009a) A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel (2009). *Evaluating algorithm performance metrics tailored for prognostics.* IEEE Aerospace conference, 7-14 March 2009, pp. 1-13.

(Saxena *et al.,* 2009b) A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Geobel (2009). *On applying the prognosis metrics*. Annual Conference of the Prognostics and Health Management, PHM 2009. 28 September – 2 October 2009.

(Skilling, 1988) J. Skilling (1988). The axioms of maximum entropy. In G.J. Erickson and C.R. Smith (Eds.), *Maximum Entropy and Bayesian Methods, in Science and Engineering, volume 1*. pp 173-187, Netherlands: Kluwer Academic Publishers.

(Tseng and Caticha, 2008) C. Tseng and A. Caticha (2008). Using relative entropy to find optimal approximations: An application to simple fluids. *Physica A*, vol. 387, pp. 6759-6770.

(van Campenhout and Cover, 1981) J. van Campenhout and T. Cover (1981). Maximum entropy and conditional probability. *IEEE Transactions on information theory*, vol. 27(4), pp. 483-489.

(Vedral, 2002) V. Vedral (2002). The role of relative entropy in quantum information theory. *Review of Modern Physics*, vol. 74(1), pp 197-234.

(Virkler *et al.*, 1979) D. A. Virkler, B. M. Hillberry, and P. K. Goel (1979). The statistical nature of fatigue crack propagation, *ASME Journal of Engineering Materials and Technology*, vol. 101, pp.148-153.

(Walker 1970) K. Walker. The effect of stress ratio during crack propagation and fatigue for 2024-T3 and 7075-T6 aluminum. In *Effects of environment and complex load history on fatigue life*. ASTM STP 462, pp. 1-14. American Society for Testing and Material.

(Zhang and Mahadevan, 2000) R. Zhang and S. Mahadevan (2000). Model uncertainty and Bayesian updating in reliability-based inspection. *Structure Safety*, vol. 22(2), pp.145-160.

Xuefei Guan is a PhD student in the department of mechanical and aeronautical engineering at Clarkson University. He received his B.S. degree in Reliability Engineering and M.S. degree in Aeronautical Engineering from Beihang University, China in 2005 and 2008, respectively. Prior to his study at Clarkson, he worked as a software engineer in Thomson Broadband (Beijing) R&D Co., LTD. His research interests are probabilistic prognosis, Bayes and entropy based methods and applications, Markov Chain Monte Carlo simulation, and system reliability.

Ratneshwar Jha is an Associate Professor in the Department of Mechanical and Aeronautical Engineering at Clarkson University. His research interests include structural health monitoring, modeling of composite and smart structures, adaptive control of structural vibrations, intelligent flight controls, and multidisciplinary design optimization. Dr. Jha is an Associate Fellow of AIAA and a member of ASME and ASEE. Dr. Jha earned PhD in Mechanical Engineering from Arizona State University in 1999, MS in Aerospace Engineering from Georgia Institute of Technology in 1983, and B. Tech in Aeronautical Engineering from Indian Institute of Technology in 1981. Dr. Jha worked in the aerospace industry from 1983 to 1995 where he led a team of engineers working on conceptual and preliminary designs of combat aircraft.

Yongming Liu is an assistant Professor in the department of civil and environmental engineering. His research interests include fatigue and fracture analysis of metals and composite materials, probabilistic methods, computational mechanics, and risk management. He completed his PhD at Vanderbilt University, and obtained his Bachelors' and Masters' degrees from Tongji University in China. Dr. Liu is a member of ASCE and AIAA and serves on several technical committees on probabilistic methods and advanced materials.

Abhinav Saxena is a Research Scientist with Stinger Ghaffarian Technologies at the Prognostics Center of Excellence NASA Ames Research Center, Moffet Field CA. His research focus lies in developing and evaluating prognostic algorithms for engineering systems using soft computing techniques. He is a PhD in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta. He earned his B.Tech in 2001 from Indian Institute of Technology (IIT) Delhi, and Masters Degree in 2003 from Georgia Tech. Abhinav has been a GM manufacturing scholar and is also a member of IEEE, AAAI and ASME.

Jose R. Celaya is a staff scientist with Stinger Ghaffarian Technologies at the Prognostics Center of

Excellence, NASA Ames Research Center. He received a Ph.D. degree in Decision Sciences and Engineering Systems in 2008, a M. E. degree in Operations Research and Statistics in 2008, a M. S. degree in Electrical Engineering in 2003, all from Rensselaer Polytechnic Institute, Troy New York; and a B.S. in Cybernetics Engineering in 2001 from CETYS University, Mexico.

Kai Goebel received the degree of Diplom-Ingenieur from the Technische Universität München, Germany in 1990. He received the M.S. and Ph.D. from the University of California at Berkeley in 1993 and 1996, respectively. Dr. Goebel is a senior scientist at NASA Ames Research Center where he leads the Diagnostics & Prognostics groups in the Intelligent Systems division. In addition, he directs the Prognostics Center of Excellence and he is the Associate Principal Investigator for Prognostics of NASA's Integrated Vehicle Health Management Program. He worked at General Electric's Corporate Research Center in Niskayuna, NY from 1997 to 2006 as a senior research scientist. He has carried out applied research in the areas of artificial intelligence, soft computing, and information fusion. His research interest lies in advancing these techniques for real time monitoring, diagnostics, and prognostics. He holds eleven patents and has published more than 100 papers in the area of systems health management.