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

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

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

[2] *Department of Civil & Environmental Engineering, Clarkson University, Potsdam, NY, 13699-5712, USA*
yliu@clarkson.edu

[3] *SGT, NASA Ames Research Center, Moffett Field, CA, 94035, USA*
abhinav.saxena@nasa.gove
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 nonstructural 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 models (Paris & Erdogan, 1963; Forman et al., 1967; Walker, 1970) are most commonly used (Bourdin, Francfort, & Marigo, 2008). However, experimental data indicate that fatigue crack propagation is not a smooth, stable and well ordered process (Virkler, Hillberry, & Goel, 1979), thus a deterministic model is not capable of quantifying the crack growth subject to various uncertainties associated with the 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 & Sorensen, 1990; Zhang & Mahadevan, 2000). Entropy methods, such as Maximum Entropy (MaxEnt) methods (Jaynes, 1957, 1979; Skilling, 1988) and relative entropy methods (Van Campenhout & Cover, 1981; Haussler, 1997), are alternative approaches for probability assignment and updating and have been used in many applications such as statistical mechanics (Caticha & Preuss, 2004; Tseng & Caticha, 2008), quantum physics (Hiai & Petz, 1991; Vedral, 2002), and fatigue prognosis (Guan, Jha, & Liu, 2009). 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. To the best knowledge of the authors, this is the first attempt to apply the MRE method as a general methodology in fatigue damage problems. 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 are discussions 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 $x \in X$, 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 $X \times \Theta$. 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 $p(\theta)$ denote the prior PDF of θ. According to the Bayes' theorem, the posterior PDF of a variable θ that reflects the fact that we observed x' is

$$p(\theta|x') \propto p(\theta)p(x'|\theta) \tag{1}$$

The Bayesian formulation of a posterior is straightforward and has an enormous variety of applications. 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 & Leibler, 1951), of two PDFs $p_1(\theta)$ and $p_2(\theta)$ is defined as,

$$I(p_1 : p_2) = -\int_{\Theta} p_1(\theta)\ln\frac{p_1(\theta)}{p_2(\theta)}\mathrm{d}\theta, \tag{2}$$

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 independent; it should not make a difference whether the inference treats them separately or jointly.

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

$$I(q : p) = -\int_{X\times\Theta} q(x, \theta)\ln\frac{q(x, \theta)}{p(x, \theta)}\mathrm{d}x\mathrm{d}\theta. \tag{3}$$

In Eq. (3), $p(x, \theta) = p(x)p(x|\theta)$ contains all prior information, $p(x|\theta)$ is the conditional PDF or likelihood function and $p(\theta)$ is the prior PDF of θ. The same relationship applies to $q(x, \theta)$. 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 $I(q : p)$. 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 from experiments or the empirical judgment on the mean value of θ. This flexibility of applicable information can pose more constraints on a posterior thus yield a more accurate result given that those constraints are justified. 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

$$c_1 : q(x) = \int_{\Theta} q(x, \theta)\mathrm{d}\theta = \delta(x - x'). \tag{4}$$

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

$$c_2 : \int_{X\times\Theta} q(x, \theta)g(\theta)\mathrm{d}x\mathrm{d}\theta = \langle g(\theta)\rangle . \tag{5}$$

The normalization constraint is

$$c_3 : \int_{X\times\Theta} q(x, \theta)\mathrm{d}x\mathrm{d}\theta = 1. \tag{6}$$

Maximizing Eq. (3) using the method of Lagrange multipliers, subject to constraints Eqs. (4-6) and the posterior PDF of θ is obtained as

$$q(\theta) \propto p(\theta)p(x'|\theta)\exp[\beta g(\theta)]. \tag{7}$$

The detailed derivation of Eq. (7) and the computation of the Lagrange multiplier β can be found in (Guan et al., 2009). The right side of Eq. (7) consists of three terms. $p(\theta)$ is the parameter prior, $p(x'|\theta)$ is the likelihood, and $\exp[\beta g(\theta)]$ 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 uncertainty term e and a measurement noise term ϵ into consideration and assuming both errors are additive to obtain

$$d = M(\theta) + e + \epsilon, \tag{8}$$

where $M(\theta)$ is the deterministic model prediction and θ is the associated model parameter variable. Without the evidence that e and ϵ are correlated to each other, the two terms are assumed to be two independent zero-mean normal variables and can be collected as a new normal variable $\tau = (e + \epsilon) \sim \mathrm{Norm}(0, \sigma_\tau)$, the likelihood function for multiple observations can be constructed as

$$p(d_1, \ldots, d_n|\theta) = \frac{1}{(\sqrt{2\pi}\sigma_\tau)^n}\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{d_i - M_i(\theta)}{\sigma_\tau}\right)^2\right]. \tag{9}$$

Substituting Eq. (9) in Eq. (7), the MRE posterior of θ is obtained as

$$p(\theta|d_1, \ldots, d_n) \propto p(\theta)\exp\left[-\frac{1}{2}\sum_{i=1}^{n}\left(\frac{d_i - M_i(\theta)}{\sigma_\tau}\right)^2 + \beta g(\theta)\right]. \tag{10}$$

For fatigue damage model $M(\theta)$, various deterministic models have been proposed to describe the fatigue crack accumulation, among which Paris type of models are commonly used in cycle based fatigue crack growth calculation. In this study, Paris model (Paris & 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' model reads,

$$\frac{\mathrm{d}a}{\mathrm{d}N} = c(\Delta K)^m = c[\Delta\sigma\sqrt{\pi a}F(a)]^m, \tag{11}$$

where c and m are model parameters, $\Delta\sigma$ is the stress variation during one cyclic load, ΔK is the variation of stress intensity in one cyclic load, and $F(a)$ is the geometry correction factor. The crack size can be calculated by solving Eq. (11) given the parameter c and m and the applied number of loading cycles N. Early studies have show that $\ln c$ follows a normal distribution and m follows truncated normal distribution (Kotulski, 1998). Given this information, the posterior of the joint distribution of $(\ln c, m)$ can be expressed as,

$$\begin{aligned} p(\ln c, m) \propto& \\ &\exp\left[-\frac{1}{2}\left(\frac{\ln c - \mu_{\ln c}}{\sigma_{\ln c}}\right)^2 + \beta_{\ln c}g_{\ln c}(\ln c)\right] \times \\ &\exp\left[-\frac{1}{2}\left(\frac{m - \mu_m}{\sigma_m}\right)^2 + \beta_m g_m(m)\right] \times \\ &\exp\left[-\frac{1}{2}\left(\frac{d_i - M_i(\ln c, m)}{\sigma_\tau}\right)^2\right]. \end{aligned} \tag{12}$$

Setting $\beta_{\ln c}$ and β_m to zero in Eq. (12) gives the 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. (12) is difficult because of the multi-dimensional integration needed for normalization. In order to circumvent the direct evaluation of Eq. (12), 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. (12) 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. (13).

$$x_{t+1} = \begin{cases} \tilde{x} \sim \pi(X|x_t) & \text{with probability } \alpha(x_t, \tilde{x}) \\ x_t & \text{else} \end{cases} \tag{13}$$

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

$$r = \frac{p(\tilde{x})}{p(x_t)} \frac{\pi(x_t|\tilde{x})}{\pi(\tilde{x}|x_t)}, \tag{14}$$

where $p(\cdot)$ is the target distribution. In our case, $p(\cdot)$ is Eq. (12). For a symmetric transition distribution $\pi(\cdot)$, such as a normal distribution, the property of $\pi(x_t|\tilde{x}) = \pi(\tilde{x}|x_t)$ reduces Eq. (14) to $r = p(\tilde{x})/p(x_t)$. In this study, 100,000 samples of $(\ln c, m)$ are generated with a 5% burn-in period using a normal transition distribution. In addition, 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 $a_i = 9\text{mm}$, length $L = 558.8\text{mm}$, width $w = 152.4\text{mm}$ and thickness $t = 2.54\text{mm}$. The loading information is $\Delta\sigma = 48.28\text{MPa}$ and stress ratio $R = 0.2$. The geometry correction factor for these specimens is $F(a) = 1/\sqrt{\cos(\pi a/w)}$. (Kotulski, 1998) reported the statistical information of the parameters in Paris' model, namely, mean values $\mu_{\ln c} = -26.155$ and $\mu_m = 2.874$ with standard deviations $\sigma_{\ln c} = 0.968$ and $\sigma_m = 0.164$, respectively. Assuming the total error term is $\tau = 0.1\text{mm}$ and substituting the statistics information into Eq. (12) with $g_{\ln c} = \ln c$ and $g_m(m) = m$, 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 & Hillberty, 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 obtained from health monitoring system or nondestructive inspection. These data points are listed in Table 1.

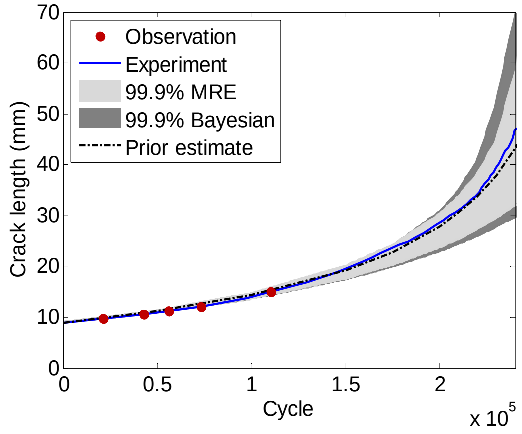

Figure 1. MRE and Bayesian prognosis (Virkler's dataset)

Number	Crack size (mm)	Cycles
1	9.733	21269
2	10.527	42734
3	11.256	56392
4	12.171	73161
5	15.055	110487

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

Predictions from MRE updating and Bayesian updating procedures are shown in Figure 1. To keep the figure clear, the median prediction (expected value) is omitted. 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.

3.2 McMaster's 2024-T351 aluminum alloy experimental data

In (McMaster & Smith, 1999), a large set of 2024-T351 aluminum alloy experimental data under constant and variable loading conditions were reported. The experimental data of center-cracked specimens with length $L = 250\text{mm}$, width $w = 100\text{mm}$ and thickness $t = 6\text{mm}$ under constant loading $\Delta\sigma = 65.7\text{MPa}$ and stress ratio $R = 0.1$ are used. Priors of the parameters are obtained by $\ln(\mathrm{d}a/\mathrm{d}N) \sim \ln(\Delta K)$ regression using the experimental data. Five data points as shown in Table 2 are chosen arbitrarily to represent sensor measurements from health monitoring system. The prior

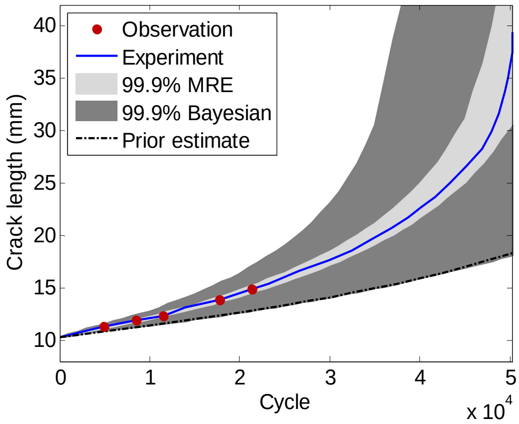

Figure 2. MRE and Bayesian prognosis (McMaster's dataset)

PDFs are artificially set as $\mu_{\text{lnc}} = -26.5$ and $\mu_m = 2.9$, which is not sufficiently accurate enough to match the experimental records as seen in Figure 2. Predictions of crack growth trajectories are also shown in Figure 2, where interval predictions obtained by MRE updating are narrower than that by Bayesian updating. One interesting observation is that the difference between MRE and Bayesian interval predictions in Figure 2 is larger than that in Figure 1. One possible explanation is that the prior PDF settings in the two datasets have different level of uncertainties. The prior PDFs in Figure 1 are sufficiently accurate. We can observe this because the prior point estimate in black dash line (computed using the mean value reported by Kotulski) is very close to the experiment records in solid blue line. For the McMaster's dataset, the prior PDFs for the Paris' equation parameters are artificial set. The prior estimate is far from the experiment records. The affect of prior PDFs settings is further discussed in Section 5. In the two examples, MRE updating shows the advantages over Bayesian updating by 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.

Number	Crack size (mm)	Cycles
1	11.361	4875
2	11.928	8475
3	12.325	11550
4	13.856	17775
5	14.877	21375

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

4. Metric-based Performance Evaluation

Various metrics are available to quantify the performance of prognosis algorithms (Saxena, Celaya, Balaban, 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, Celaya, Balaban, 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 statistical metrics mentioned above are general purpose metrics and were not specifically designed for prognosis. In (Saxena, Celaya, Saha, Saha, & Goebel, 2008) 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. $\alpha - \lambda$ 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, Celaya, Saha, et al., 2008). 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 $\alpha - \lambda$ 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 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, Celaya, Saha, Saha, & Goebel, 2009), 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 , 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. The statistical metrics, MAPE, Average Bias, STD, MSE, are computed after the fifth updating. The prognosis metrics of PH, RA, and CRA are computed using the 90% interval predictions of RUL at each updating points. For the convergence metric, the median

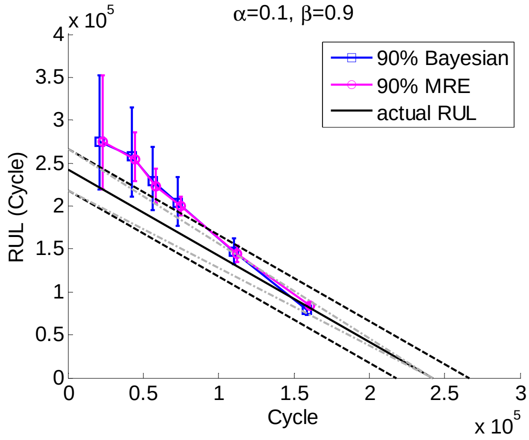

Figure 3. Performance comparison for PH and $\alpha - \lambda$ accuracy at $\alpha = 0.1$ (10% error bound) on Virkler's dataset

prediction is used here. 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.

Metrics	MRE	Bayesian
MAPE	8.66	10.93
Average Bias(cycle)	10956.27	14051.92
STD(cycle)	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

Table 3. Comparison of metrics between MRE and Bayesian approaches (Virkler's dataset, statistical metrics (rows 1-4) are computed after fifth updating)

Performance Results for McMaster's Dataset

A similar analysis for the McMaster's dataset is performed. The visual results for PH and $\alpha - \lambda$ accuracy metrics comparing Bayesian and MRE updating are shown in Figure 4. The rest of the metrics are included in Table 4. 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. For this dataset, these met-

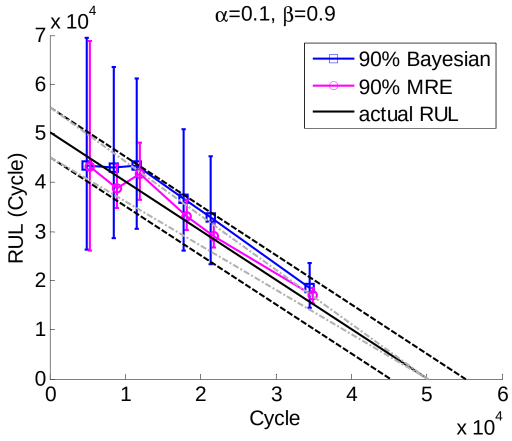

Figure 4. Performance comparison for PH and $\alpha-\lambda$ accuracy at $\alpha = 0.1$ (10% error bound) on McMaster's dataset

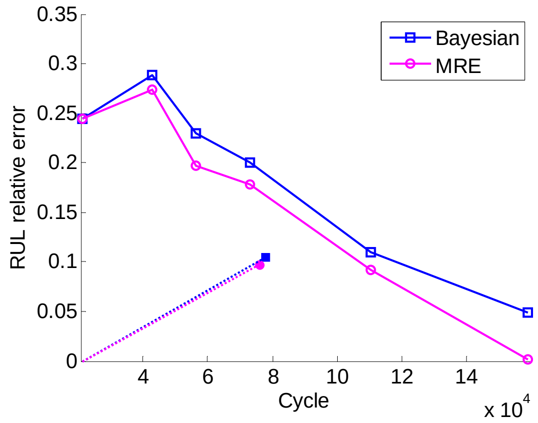

Figure 5. Comparison of convergence performance on Virkler's dataset

rics clearly distinguish the two approaches and show better outcomes from the MRE method. For example, the PH and $\alpha - \lambda$ 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.

Metrics	MRE	Bayesian
MAPE	4.06	22.53
Average Bias(cycle)	418.76	4561.93
STD(cycle)	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

Table 4. Comparisons of metrics between MRE and Bayesian approaches (McMaster's dataset, statistical metrics (rows 1-4) are computed after fifth updating)

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. (15), which is the difference between an actual response measure (R) and the inferred value (R_0) divided by the actual response measure. The result of 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). 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, Celaya, Saha, et al., 2008). 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.

$$\text{RE} := \left| \frac{R_0 - R}{R} \right|. \tag{15}$$

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

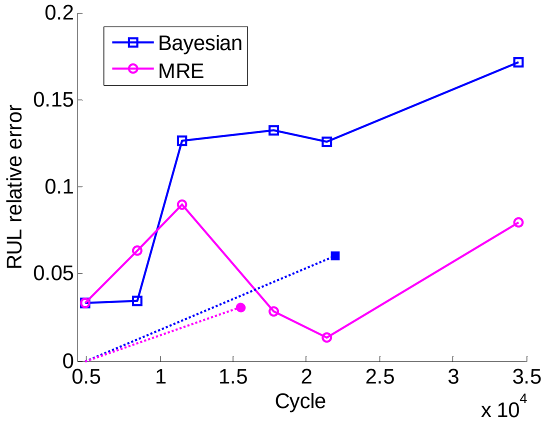

Figure 6. Comparison of convergence performance on McMaster's dataset

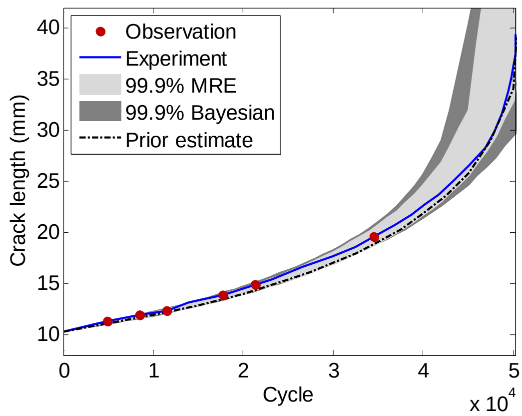

Figure 8. MRE and Bayesian prognosis with an accurate prior (McMaster's dataset)

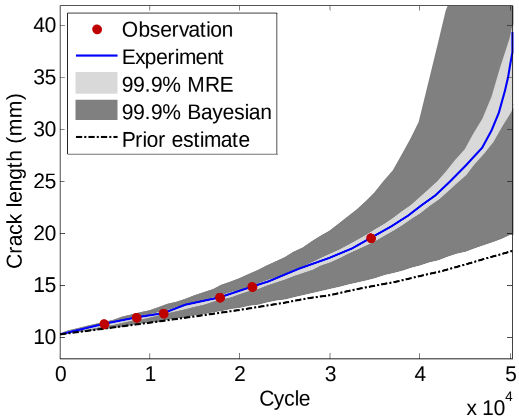

Figure 7. MRE and Bayesian prognosis with an inaccurate prior (McMaster's dataset)

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 crack growth trajectory, 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. 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., $\pm 20\%$ 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

$$R_b := \frac{\int_{x_{\text{mean}}-\eta}^{x_{\text{mean}}+\eta} f(x)\mathrm{d}x}{4\eta\delta}, \tag{16}$$

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. (16) 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, $\eta = 0.2$ and $\delta = 0.2$ are used to investigate the parameter m in the crack growth model (Eq. (12)). The mean value of m is 2.874. All predictions are made after six updatings and the 99% confidence bounds are shown in Figure 9. The robustness metric (Eq. (16)) 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.

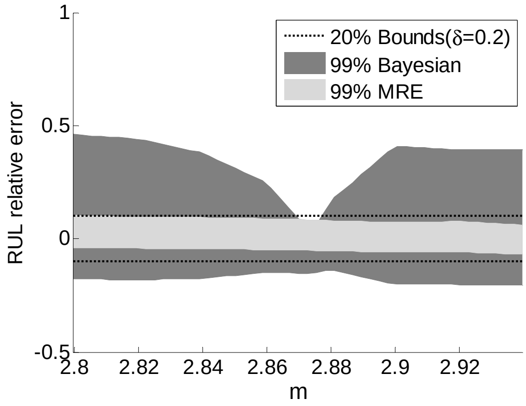

Figure 9. Comparison of robustness metric after six updatings with varying values of m in prior PDF (Eq. 12) for parameter m (Virkler's dataset)

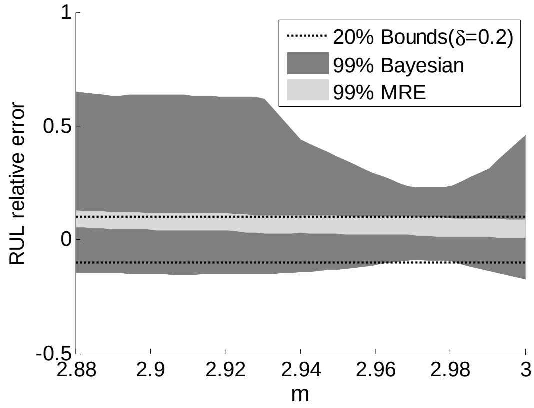

Figure 10. Comparison of robustness metric after six updatings with varying values of m in prior PDF (Eq. 12) for parameter m (McMaster's dataset)

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. In fatigue damage problems, the model parameters are usually tuned using extensive experiments on standard specimens. The realistic systems are usually different from specimens in geometric dimensions, loading profiles, and usage environment. Extensive experiments on the actual engineering systems are sometimes prohibitive due to the time and cost constraints. Therefore, 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
$p(\cdot)$	Probability distribution
$M(\cdot)$	Fatigue crack growth model
$F(\cdot)$	Geometry correction factor
a	Crack length
N	Number of loading cycles
d	Crack length measurements

REFERENCES

Bourdin, B., Francfort, G., & Marigo, J. (2008). The variational approach to fracture. *Journal of Elasticity*, *91*(1), 5–148.

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

Forman, R., Kearney, V., & Engle, R. (1967). Numerical analysis of crack propagation in cyclic-loaded structures. *Journal of Basic Engineering*, *89*(3), 459–464.

Gregory, P. (2005). *Bayesian logical data analysis for the physical sciences: a comparative approach with Mathematica support*. Cambridge Univ Pr.

Guan, X., Jha, R., & Liu, Y. (2009). Probabilistic fatigue damage prognosis using maximum entropy approach. *Journal of Intelligent Manufacturing*, 1-9. (DOI: 10.1007/s10845-009-0341-3)

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

Haussler, D. (1997). A general minimax result for relative entropy. *Information Theory, IEEE Transactions on*, *43*(4), 1276–1280.

Hiai, F., & Petz, D. (1991). The proper formula for relative entropy and its asymptotics in quantum probability. *Communications in mathematical physics*, *143*(1), 99–114.

Jaynes, E. (1957). Information theory and statistical mechanics. II. *Physical review*, *108*(2), 171.

Jaynes, E. (1979). Where do we stand on maximum entropy. *The maximum entropy formalism*, 15–118.

Kotulski, Z. (1998). On efficiency of identification of a stochastic crack propagation model based on Virkler experimental data. *Archives of Mechanics*, *50*, 829–848.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 79–86.

Madsen, H., & Sorensen, J. (1990). Probability-based optimization of fatigue design, inspection and maintenance. *Integrity of offshore structures–4*, 421.

McMaster, F., & Smith, D. (1999). Effect of load excursions and specimen thickness on crack closure measurements. *Advances in Fatigue Crack-closure Measurements and Analysis*, 246–264.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., et al. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, *21*(6), 1087.

Ostergaard, D., & Hillberty, B. (1983). Characterization of the variability in fatigue crack propagation data. *Probabilistic Frature Mechanics and Fatigue Methods: Applications for Structural Design and Maintenance*, 97.

Paris, P., & Erdogan, F. (1963). A critical analysis of crack propagation laws. *Journal of Basic Engineering*, *85*(4), 528–534.

Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., et al. (2008). Metrics for evaluating performance of prognostic techniques. In *International Conference on Prognostics and Health Management (PHM08).*

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2008). Evaluating algorithm performance metrics tailored for prognostics. In *Aerospace conference, 2009 IEEE* (pp. 1–13).

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2009). On applying the prognostic performance metrics. In *Proceedings of the Annual Conference of the Prognostics and Health Management Society (PHM), San Diego, CA.*

Skilling, J. (1988). The axioms of maximum entropy. *Maximum Entropy and Bayesian Methods in Science and Engineering, volume 1*, 173–187.

Tseng, C., & Caticha, A. (2008). Using relative entropy to find optimal approximations: An application to simple fluids. *Physica A: Statistical Mechanics and its Applications*, *387*(27), 6759–6770.

Van Campenhout, J., & Cover, T. (1981). Maximum entropy and conditional probability. *Information Theory, IEEE Transactions on*, *27*(4), 483–489.

Vedral, V. (2002). The role of relative entropy in quantum information theory. *Reviews of Modern Physics*, *74*(1), 197.

Virkler, D., Hillberry, B., & Goel, P. (1979). The Statistical Nature of Fatigue Crack Propagation. *Journal of Engineering Materials and Technology*, *101*, 148.

Walker, K. (1970). The Effect of Stress Ratio During Crack Propagation and Fatigue for 2024-T3 and 7075-16 Aluminum. In *Effects of environment and complex load history on fatigue life: a symposium* (p. 1).

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

Xuefei Guan is a PhD student in the department of mechanical and aeronautical engineering at Clarkson University. He received his Bachelor's degree in Reliability Engineering and Master's degree in Aeronautical Engineering from Beihang University, China in 2005 and 2008, respectively. His re-

search 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 Bachelor's and Master's 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.