Gaussian Process Time Series Model for Life Prognosis of Metallic Structures

SUBHASISH MOHANTY,* SANTANU DAS, ADITI CHATTOPADHYAY AND PEDRO PERALTA

Mechanical and Aerospace Engineering, Arizona State University, Tempe, AZ-85287, USA

ABSTRACT: Al 2024-T351 fatigue specimens have been modeled using a kernel-based multivariate Gaussian Process approach. The Gaussian Process model projects fatigue affecting input variables to output crack growth by probabilistically inferring the underlying nonlinear relationship between input and output. The Gaussian Process approach not only explicitly models the uncertainty due to scatter in material microstructure parameter but it also implicitly models the loading sequence effect due to variable loading. The loading sequence effect is modeled through the Gaussian Process optimal hyperparameters by using the crack length data observed over the entire domain of spectrum loading. The performance in the crack growth prediction is evaluated for two covariance functions, a radial basis-based, anisotropic, covariance function and a neural network-based isotropic covariance function. Furthermore, the performance of different types of scaling, used to scale the input–output data space, is tested. It is found that for the radial basis-based anisotropic covariance function with normalized scaling, the prediction error is consistently lower compared to other combinations. In addition, the Gaussian Process model allows determination of the collapse load condition, which is a desirable feature for the online health monitoring and prognosis.

Key Words: prognosis, fatigue crack growth, 2024-T351 aluminum alloy, variable loading, Gaussian Process, covariance function, maximum likelihood optimization, hyperparameters.

INTRODUCTION

SIGNIFICANT research and development efforts are currently being invested in the development of on-board systems for integrated health management and prognosis of aerospace structures. A simple form of prognostics, based on constant life diagram (Miner, 1945), is being widely used for fatigue life prediction of metallic structural components. However, these predictions are not based on measured characteristics of the individual components. Another method that is widely practiced and has constantly been improved is the physics-based approach (Suresh, 1991). These reported physics-based prognosis models use linear elastic fracture mechanics (LEFM) to account for fatigue failure. A third approach is the data mining, data driven or the machine learning method, which uses historical and current usage data to 'learn' a model of system behavior.

Fatigue crack growth in aircraft structural components is influenced by service load. Conventionally, fatigue life of aircraft structural components under service loading is often analyzed and predicted based on crack growth rates obtained from constant-amplitude fatigue testing data. It is well known that fatigue crack growth is a random phenomenon (Ditlevsen and Olsen, 1986; Spencer and Tang, 1989; Liu and Mahadevan, 2007) because of scatter in microstructural parameters. One approach to probabilistic modeling of fatigue crack growth is to generate the necessary stochastic information by multiplying the deterministic dynamics (e.g., Paris-Erdogan equation) of fatigue crack growth with a non-negative random process. The parameters in such a randomized model are inferred from experimental data usually found from constant-amplitude loading fatigue tests. In contrast to the fatigue crack growth due to constant amplitude loading, crack growth caused by variable amplitude loadings is characterized by retardation and acceleration effects (Elber, 1970; Newman, 1982, 1984), which extend or reduce the lifetime of structures. Experimental investigations show that single mode I overloads lead to extensive retardation effects, which are influenced by the overload ratio and the baseline level loading, in which the single overloads are interspersed (Sander and Richard, 2005). Currently there are many physics-based models (Newman, 1982; 1992; Harter, 1999; Ray and Patnakar, 2001a, 2001b) with empirical parameters available to model crack growth with retardation and acceleration effects. These models reasonably capture the dynamics of the fatigue crack growth under variable loading in a deterministic framework. However, these models do not explicitly

*Author to whom correspondence should be addressed.
E-mail: subhasish.mohanty@asu.edu
Figures 1–6 and 8–10 appear in color online: http://jim.sagepub.com

1045-389X/09/08 0887–10 $10.00/0 DOI: 10.1177/1045389X08099602

model the uncertainty in crack growth that arises due to scatter in micro structural properties and subsequent uncertainty propagation due to loading sequence effect.

The present article presents a data-driven Gaussian Process-based probabilistic approach (GP) (Mackay, 1998; Rasmussen and Williams, 2006), which not only takes care of the uncertainty that arises due to scatter in material properties, but also accounts for the loading sequence effect due to variable loading. The loading sequence effect has been implicitly modeled over GP hyperparameters while probabilistically mapping the GP input space to the GP output space. The input space includes variables such as number of fatigue cycles, minimum load, maximum load, and load ratio, whereas the output space describes the crack length at that point in the input space. It is noted that though the first three independent variables ideally suffice to model the input space, an additional dependant variable such as load ratio is included in the input space to give extra weight to loading parameters and hence the loading sequence effect. In the Bayesian approach to the regression problem, such as in Bayesian Neural Network (NN) (Neal, 1996), a prior probability distribution over the model parameters ascertains a prior probability distribution over underlying function space and hence in the output prediction. This prior has been combined with a noise model to yield a posterior distribution over functions that can then be used for output predictions. In general, the prior over individual model parameters coupled with noise model has a complex form in the function space. The idea of GP modeling is, without parameterizing (e.g., as in NN (Hidetoshi et al., 1996)) the model function, to place a prior directly on the functions space. It has been shown that many Bayesian regression models based on NN converge to GP only in the limit of an infinite network (Neal, 1996) or in the realistic sense for a large finite network. As the complexity of the regression model increases due to variable loading, NN with a large network may not be advisable for computational efficiency, which motivates the use of a GP model for the current prognostic approach.

THE FATIGUE CRACK GROWTH PREDICTION PROBLEM

The prediction of fatigue crack growth can be treated as a time series evolution problem. Most common time series are the result of unknown or incompletely understood systems. Here the fatigue crack growth at any time t can be defined as a function $a(\vec{x})$ of an input variable $\vec{x}$, where both a and $\vec{x}$ vary with time t. The main characteristics of the function $a(\vec{x})$ is that its evolution cannot be described exactly, but rather can be expressed probabilistically. In addition, to infer the function value, past observations should be taken into account. So with given information of crack lengths $a_1, a_2, a_3, \ldots, a_N$, respectively, at time $t_1, t_2, t_3, \ldots, t_N$, it is possible to infer the crack length a_{N+1} at time t_{N+1}. It is noted that the known inputs at time $t_1, t_2, t_3, \ldots, t_N, t_{N+1}$ are, respectively, $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots, \vec{x}_N, \vec{x}_{N+1}$, where the individual $\vec{x}_{i,i=1,2,\ldots,N,N+1}$ are vectors with individual components corresponding to variables that affect the fatigue crack growth.

Prediction in a Bayesian Framework

The goal of a probabilistic Bayesian forecasting approach is to compute the posterior distribution $f(a_{N+1}|D = \{\vec{x}_i, a_i\}_{i=1,\ldots,N}, \vec{x}_{N+1})$, i.e., to capture the probability distribution of the random crack length a_{N+1} given a random test input $\vec{x}_{N+1}$ and a set of N training data points described as $D = \{\vec{x}_i, a_i\}_{i=1,\ldots,N}$. In the Bayesian framework the predictive distribution with mean and variance can be found by conditioning the crack lengths $a_1, a_2, a_3, \ldots, a_{N+1}$ that are affected by the corresponding random inputs $\vec{x}_1, \vec{x}_2, \vec{x}_3, \ldots, \vec{x}_N, \vec{x}_{N+1}$. Now we can define a prior over the space of possible functions to model the random crack length as $f(a|\alpha)$, where α are some hyperparameters which may take care of random load sequence effect in the form of curve fitting. We can also define a prior over the noise $f(\vartheta|\beta)$, where ϑ is some appropriate noise vector that arises due to scatter in material microstructure and β is another set of hyperparameters used to model the uncertainty due to scatter. Now if the hyperparameters α and β are given, the conditional probability can be expressed as:

$$f\left(\vec{a}_{N+1}|\{\vec{x}_{i=1,\ldots,N}, \alpha, \beta\}\right) = \int \left(\vec{a}_{N+1}|\{\vec{x}_{i=1,\ldots,N}, a, \vartheta\}\right) f(a|\alpha) f(\vartheta|\beta) \mathrm{d}a \mathrm{d}\vartheta \quad (1)$$

where a and ϑ denote the underlying function which, respectively, corresponds to crack length and noise due to scatter. Since $a_1, a_2, a_3, \ldots, a_N$, and a_{N+1} are conditioned random variables in the observed set of crack lengths, the conditional distribution of a_{N+1} can be written as follows:

$$f\left(a_{N+1}|D = \{\vec{x}_i, a_i\}_{i=1,\ldots,N}, \vec{x}_{N+1}, \alpha, \beta\right) = \frac{f\left(\vec{a}_{N+1}|\{\vec{x}_i\}_{i=1,\ldots,N+1}, \alpha, \beta\right)}{f\left(\vec{a}_N|\{\vec{x}_i\}_{i=1,\ldots,N}, \alpha, \beta\right)}. \quad (2)$$

Predicting with GP

To evaluate Equation (2) it is necessary to evaluate the integral given in Equation (1). However, in general, Equation (1) is complicated to evaluate. The standard approach to evaluate the integral in Equation (1) is by a

method called evidence maximization (Mackay, 1992) or by numerically integrating the integral by Monte Carlo simulation (Neal, 1993). However, assuming the overall (including noise) underlying function $a_{i=1,\ldots,N,N+1,\ldots,}$ follows a Gaussian distribution, the exact analytical form of Equation (1) is as follows:

$$f(\vec{a}_N|\{x_i\}_{i=1,\ldots,N}, C_N) = \frac{1}{(2\pi)^{N/2} det(C_N)^{1/2}} \exp\left(-\frac{1}{2}(\vec{a}_N - \mu)^T C_N^{-1}(\vec{a}_N - \mu)\right) \quad (3)$$

where μ is the function mean and C_N is a $N \times N$ covariance matrix. The individual elements c_{mn} of the covariance matrix C_N can be found from a parameterized covariance function that will be described in the next section. Assuming zero mean function distribution, Equation (2) can be written as:

$$f\left(a_{N+1}|D = \left\{\vec{x}_i, a_i\right\}_{i=1,\ldots,N}, \vec{x}_{N+1}, c_{mn}(x_m, x_n, \Theta)_{m,n=1,2,\ldots,N+1}\right) = \sqrt{\frac{det(C_N)}{(2\pi)det(C_{N+1})}} \exp\left(-\frac{(a_{N+1} - \hat{a}_{N+1})^2}{2\sigma^2_{\hat{a}_{N+1}}}\right) \quad (4)$$

where $\hat{a}_{N+1}$ is the predicted mean at time t_{N+1} and is given by:

$$\hat{a}_{N+1} = \vec{k}^T C_N^{-1} \vec{a}_N; \quad k_i = c(\vec{x}_{N+1}, \vec{x}_i)_{i=1,2,\ldots N} \quad (5)$$

whereas $\sigma^2_{\hat{a}_{N+1}}$ is the predicted variance at time t_{N+1} and is given by:

$$\sigma^2_{\hat{a}_{N+1}} = \kappa - \vec{k}^T C_N^{-1} \vec{k}; \quad k_i = c(\vec{x}_{N+1}, \vec{x}_i)_{i=1,2,\ldots N}; \quad \kappa = c(\vec{x}_{N+1}, \vec{x}_{N+1}). \quad (6)$$

Parameterizing the Covariance Function

For any two sets of input vectors x_m and x_n, the covariance or kernel function used in Equation (4) has the following form:

$$c_{mn}(x_m, x_n, \Theta) = c_f(x_m, x_n, \Theta) + c_{\text{noise}}(x_m, x_n, \Theta) \quad (7)$$

where c_f is associated with an interpolation function and c_{noise} is associated with the noise model. There are many possible choices of prior interpolation covariance functions. From a modeling point of view, the objective is to specify a prior covariance that contains our assumptions about the structure of the process being modeled. Formally, it is required to specify a function that will generate a positive definite covariance matrix for any set of inputs. In this article, the performance of an anisotropic covariance function based on a radial basis function (RBF) and a NN-based isotropic covariance function is tested. The RBF covariance function (Rasmussen and Williams, 2006) has the following form:

$$c_f(x_m, x_n, \Theta) = \theta_1 \exp\left(-\frac{1}{2}\sum_{l=1}^{N_d}\frac{(\vec{x}_m - \vec{x}_n)^2}{r_l^2}\right) + \theta_2. \quad (8)$$

Whereas NN-based (Rasmussen 2006; Williams 1997) covariance function has the following form:

$$c_f(x_m, x_n, \Theta) = \theta_1 Sin^{-1}\frac{\vec{x}_m^T \theta_3 \vec{x}_n}{\sqrt{(1 + \vec{x}_m^T \theta_3 \vec{x}_m)(1 + \vec{x}_n^T \theta_3 \vec{x}_n)}} + \theta_2. \quad (9)$$

In Equation (8) $r^2_{\ell=1,\ldots,N_d} \in \Theta$ are length scale dependent hyperparameters corresponding to individual input space variables such as fatigue cycles, minimum load, maximum load, load ratio, etc., with N_d as the number of input space variables. Also in Equation (7) the second term associated with the noise model only contributes to the diagonal term of the covariance matrix, and for an input independent noise model, c_{noise}, can be written in terms of a hyperparameter θ_4 such that:

$$c_{\text{noise}}(x_m, x_n, \Theta) = \delta_{mn}\theta_4 \quad (10)$$

where δ_{mn} is the Kronecker delta with value one when $m = n$ and zero when $m \neq n$.

Hyperparameter Determination

So far we have only considered the properties of the prediction model for fixed values of the hyperparameters. In this section we will discuss how to obtain the hyperparameters Θ for a fixed training data set $D = \left\{\vec{x}_i, a_i\right\}_{i=1,\ldots,N}$. Ideally one should integrate over all possible hyperparameters in order to obtain best possible predictions of the function value a_{N+1} at time t_{N+1}. So we can write:

$$f\left(a_{N+1}|D, \vec{x}_{N+1}, C(\cdot)\right) = \int f\left(a_{N+1}|D, \vec{x}_{N+1}, C(\cdot), \Theta\right) f(\Theta|D, C(\cdot))\mathrm{d}\Theta. \quad (11)$$

The above integral is as complex as the integral given in Equation (1) and also difficult to evaluate for a complex problem with several hyperparameters and a multiple input space. Out of the two possible approaches e.g., the Maximum evidence (Mackay, 1992) and the Monte Carlo approach (Neal, 1993) we will only discuss the use of the maximum evidence approach to evaluate

the integral. Using maximum evidence approach, Equation (11) can be written in its approximate form as:

$$f\left(a_{N+1}|D,\vec{x}_{N+1},C(\cdot)\right)\cong f\left(a_{N+1}|D,\vec{x}_{N+1},C(\cdot),\Theta_{\mathrm{MAP}}\right). \tag{12}$$

The approximation in Equation (12) is based on the assumption that the posterior distribution over Θ i.e., $f(\Theta|D,C(\cdot))$ has a sharp peak around Θ_{MAP}. This approximation is generally good and predictions are often found to be very close to those obtained using the true predictive distribution (Mackay, 1993). Now to find the peak location of $f(\Theta|D,C(\cdot))$ the posterior of its need to be optimized and the posterior can be written as:

$$f(\Theta|D,C(\cdot))=\frac{f\left(\vec{a}_N|\{\vec{x}_i\}_{i=1,2,\ldots N},C(\cdot),\Theta\right)f(\Theta)}{f\left(\vec{a}_N|\{\vec{x}_i\}_{i=1,2,\ldots N},C(\cdot)\right)}. \tag{13}$$

In Equation (13) the denominator (i.e., evidence) is independent of Θ and can be ignored in the optimization process. On the other hand, the other two terms, the likelihood $f\left(\vec{a}_N|\{\vec{x}_i\}_{i=1,2,\ldots,N},C(\cdot),\Theta\right)$ and the prior $f(\Theta)$, need to be considered in the optimization of $f(\Theta|D,C(\cdot))$. Now with the assumption that all $a_{i=1,2,\ldots,N}$ follow a Gaussian distribution and using Equation (3), the logarithm of the objective function can be written as:

$$\begin{aligned}L\equiv\mathrm{Log}(f(\Theta|D,C(\cdot)))&=-\frac{1}{2}\mathrm{Log}(\det C_N)\\&-\frac{1}{2}\vec{a}_N^{\,T}C_N^{-1}\vec{a}_N\frac{N}{2}\mathrm{Log}(2\pi)+\mathrm{Log}f(\Theta).\end{aligned} \tag{14}$$

The log-likelihood function L in Equation (14) is generally multi-modal and can be optimized using any multi-variate optimization algorithm. In the present work the conjugate gradient method is used to optimize the log-likelihood function and to obtain the optimized hyper-parameters. Note that it is common practice (Gibbs, 1997) to ignore the log prior term in Equation (14) due to the absence of knowledge about Θ. The resulting solution may not be always a realistic solution. However, it can be assumed that $\mathrm{Log}f(\Theta)$ is implicitly modeled through the optimization of the log-likelihood L. The details about the optimization process are discussed later.

DATA SETS AND COVARIANCE MATRIX

As mentioned before, the GP model is a data-driven approach. Correct predictions using this model depend on proper selection of the input–output data set, scaling of data, and on proper selection of covariance function. The following subsections discuss the above-mentioned issues in more details.

Fatigue Test Experiment

GP data sets were generated from experimental observations. Compact Tension (CT) samples, 6.31 mm thick, made of 2024 T351 Aluminum were used. These specimens were fabricated according to ASTM E647-93 with an average width of 25.53 mm (from the center of the pin hole to the edge of the specimen) and an average height of 30.6 mm. The experiments were performed in an Instron 1331 servo-hydraulic load frame operating at 20 Hz. Four CT test samples, labeled CT417a, CT419a, CT421a, and CT423a were tested to generate the training and test data for the GP model. To simulate typical flight maneuvering conditions, a variable load spectrum was programmed into the digital controller of the load frame. The spectrum, as coded to the load cell, is shown in Figure 1. The loading spectrum shown in Figure 1 had six low-load regimes, each of which vary from a minimum load of 219 lb to maximum load of 328 lb. The first low-load regime operated for 5 Kcycles and the other five load regimes operate for 40 Kcycles each. The loading spectrum also had a medium-load regime of 1kcycles that operated just after the first low-load regime with a minimum load of 122 lb and maximum load of 553 lb. In addition, the load spectrum had five high-load regimes of 1 Kcycle each. The minimum and maximum loads in the high-load regimes were, respectively, 87 lb and 1000 lb. However, due to noise and compliance effects, the load that was actually applied to structure was somewhat different than that programmed to the controller. The minimum and maximum loads applied to the structure were measured through the load cell and are shown in Figure 2. Before the specimens were loaded with the actual loading spectrum shown in Figure 1, cracks were grown to a particular length to decrease scatter due to crack nucleation. It should be noted that for each sample the frame was stopped and the corresponding load cell outputs were recorded at chosen instances during the experiment. Each time the test was stopped, a high-resolution picture of the cracked sample was taken

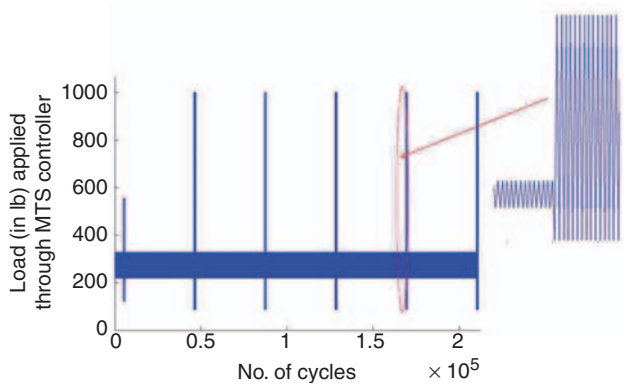

Figure 1. *Actual load spectrum programmed to MTS controller.*

using a camera (Sony XCD-SX910CR) to generate the output space crack length data. The corresponding measured crack lengths for different samples are depicted in Figure 3.

Input–Output Data Space and its Scaling

The total number of data points (comprising load cell observations, number of cycles elapsed, and corresponding crack length) for samples CT417a, CT419a, CT421a, and CT423a are, respectively, 29, 32, 28, and 28, comprising a total number of data points 117. Out of these observed data, GP input space $\{\vec{x}_i\}_{i=1,\ldots,N}$ has been constructed using variables such as number of loading cycles elapsed, minimum and maximum loads measured by the load cell and the load ratio. As mentioned earlier, in addition to the first three independent input variables, the fourth dependent variable, the load ratio, is included to the input space to give an extra weight to the loading variable and hence the associated loading sequence effect. So $\vec{x}_i$ is a 4D quantity (1×4 vector) with different input values at different observation instances $i = 1, \ldots, N$. Similarly the output space parameter a_i comprises the observed crack length at various instances of time, but unlike the input space parameters $\vec{x}_i$ the output space parameter is a scalar at a given instant of time. A chart of the input–output relation has been depicted in Equation (15):

$$D = \begin{bmatrix} \overbrace{\begin{matrix} x_{1,1} & x_{1,2} & x_{1,3} & x_{1,4} \\ x_{2,1} & x_{2,2} & x_{2,3} & x_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{i,1} & x_{i,2} & x_{i,3} & x_{i,4} \\ \vdots & \vdots & \vdots & \vdots \\ x_{N,1} & x_{N,2} & x_{N,3} & x_{N,4} \end{matrix}}^{\text{Input data matrix}} & \overbrace{\begin{matrix} a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_N \end{matrix}}^{\text{Target vector}} \end{bmatrix} \tag{15}$$

Once the training input–output space $D = \{\vec{x}_i, a_i\}_{i=1,\ldots,N}$ is fixed, scaling of the data is performed to avoid ill conditioning and subsequent erratic predictions. For instance, in the present prediction problem the number of cycles elapsed as one input space variable generally has large numerical value compared to other variables such as minimum load, maximum load, and load ratio. It is noted that the GP work wells when the individual input–output space variables have similar distributions i.e., similar mean and variance. To counter this problem of ill conditioning, all the input and output space variables have to be properly scaled. In this article, performance of three different types of scaling, logarithmic, normalized, and log-normalized scaling, are studied. In logarithmic scaling each input–output space variable is logarithmically scaled, whereas in normalized scaling each variable is scaled between zero and one, and in the case of log-normalized scaling each variable is first logarithmically scaled and then normalized between zero and one.

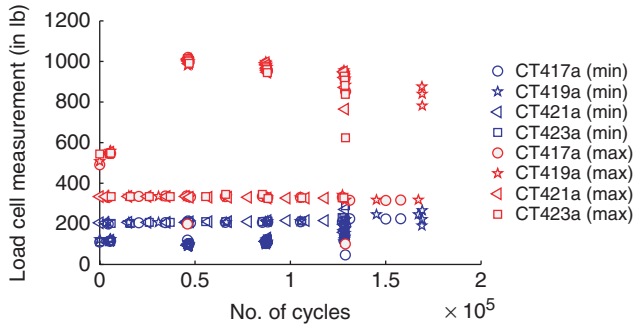

Figure 2. *Minimum and maximum load (in lb) measured through MTS load cell.*

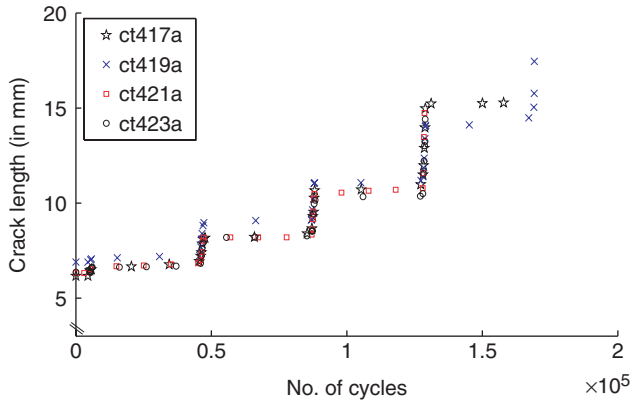

Figure 3. *Measured crack length (using high-resolution digital camera) from different samples.*

Optimal Hyperparameters and Covariance Matrix

After appropriate scaling, the minimization of the negative log-likelihood function (or the maximization of the likelihood function as given in Equation (14) are performed by using a conjugate gradient optimization algorithm. This minimization process gives the hyperparameters which have been optimally adopted to the training data. Though the hyperparameters do not directly correlate to any of the conventional physics-based fatigue growth parameters (such as Paris exponent and coefficients, constraint factor that relates the plastic zone size, and loading sequence effect) in the GP approach it can be envisioned that these physics-based parameters are implicitly modeled through the optimized hyperparameters. Figure 4 shows a typical example of the convergence of the negative log-likelihood function for crack growth law prediction in the CT419a sample. It is noted that while performing the optimization, the hyperparameters have been initialized to perform the conjugate gradient search process.

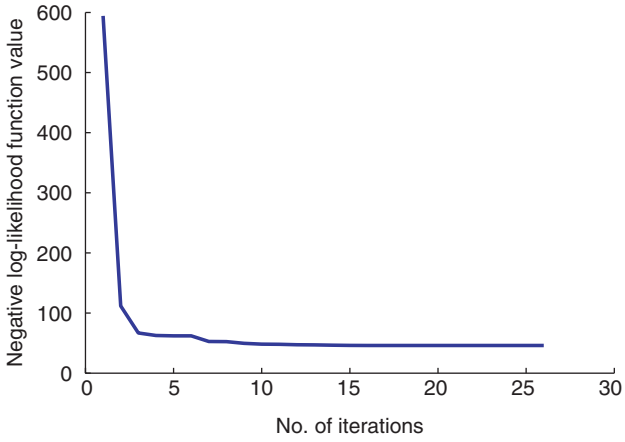

Figure 4. *Optimization of the log-likelihood function for finding the optimal hyperparameters. This plot shows the convergence of the optimization algorithm.*

At different iterations in different search directions the covariance matrix gets updated along with the new set of hyperparameters. The hyperparameter values obtained after 25 iterations were used for prediction of future or unseen crack growth. To generate the individual element of covariance matrix, either the radial basis covariance function given by Equation (8) or the NN covariance function given by Equation (9) has been used. The outcome of the GP model with different covariance functions and a comparative study on their performance with different scaling is discussed in the following sections.

NUMERICAL RESULTS

Once the optimized hyperparameters are obtained from a given set of training inputs, predictions can be made either for continuous or discrete input spaces. However, in the present work a discontinuous input space has been used because the crack length data are measured from experiment by stopping the Instron machine at discrete instances. These discrete predictions have been compared with the corresponding true crack length and the prediction accuracy has been evaluated using two different performance metrics, under various test cases. The two metrics used are normalized mean square error (NMSE) and cumulative square error (CSE), and their corresponding definitions are given below:

$$\text{NMSE} = \frac{1}{N\sigma_T^2(a^{\text{true}})}\sum_{i=1}^{N}(a^{\text{predicted}} - a^{\text{true}})^2 \qquad (16)$$

$$E_{\text{CSE}}^{j} = \sum_{i=1}^{j}(a_i^{\text{predicted}} - a_i^{\text{true}})^2. \qquad (17)$$

In Equation (16) $'N'$ is the total number of training data points, a^{true} is the true crack length value as found from experiment, $a^{\text{predicted}}$ is the predicted crack length from GP model, and $\sigma_T^2(a^{\text{true}})$ is the variance of true crack length data.

Performance Evaluation of Different Covariance Functions and Scaling

Table 1 shows a comparison of the resulting NMSE's under different test cases. For example, the NMSE for prediction of crack length for sample ct423a (test case A) with RBF kernel and normalized scaling is 0.210. For this particular case the data from the other three samples (ct417a, ct419a, and ct421a) have been used as training data. Similarly, the table shows the NMSE for other test cases with different kernel and scaling combinations. From the table it is found that the RBF kernel with normalized and log-normalized scaling has slightly better performance in terms of NMSE when compared to other kernel and scaling combinations. Figure 5 shows the CSE for crack growth prediction of CT417a sample using different scaling and covariance function types. From Figure 5 it can be noted that in most of the fatigue loading regime, the CSE of prediction with RBF kernel and normalized scaling has been consistently lower compared to predictions with the same kernel but with the log-normalized scaling. Also, it can be found from both Table 1 and Figure 5 that the prediction with NN-based kernel and log-normalized scaling performs worse compared to predictions with other combinations of kernel and scaling. Although the NN kernel with log scaling shows the least amount of CSE during the final failure regime, it develops a larger error much earlier than the RBF kernel with normalized scaling. This results in a reduced prediction horizon compared to the RBF kernel with normalized scaling. Here it is important to note that the threshold on the CSE is application-specific and this threshold determines that horizon of prediction. Since there is a finite increment (≥ 0) of crack growth over the fatigue loading regime, the CSE is always monotonically increasing with cycles. The inferior performance (in terms of CSE) of the NN kernel with logarithmic and normalized scaling compared to the RBF kernel has been attributed to the fact that the NN kernel has a single length scale hyperparameter (θ_3 in Equation (9)) as compared to four different length scale hyperparameters for the RBF kernel (r_ℓ^2 in Equation (8)). Figure 6 shows the crack length prediction for a typical test case of CT421a with RBF kernel and normalized scaling. From this figure it can be observed that there is a good correlation between prediction and true value, though the difference between prediction and true value increases with the number of fatigue cycles elapsed. The higher variation towards the final failure is because of increased nonlinearity in fatigue crack growth due to a larger plastic zone size in front of the crack tip. The effect of higher nonlinearity can be better modeled using a higher number of training data points.

Table 1. Values of the NMSE of the predicted crack length for several test cases using GP algorithm. Two different covariance functions are used with normalized, logarithmic, and log-normalized data sets.

Kernel	Scaling	Test-A (ct423a)	Test-B (ct421a)	Test-C (ct419a)	Test-D (ct417a)
RBF	Normalized	0.210	0.170	0.097	0.219
RBF	Log	0.296	0.127	0.106	0.230
RBF	Log-normalized	0.291	0.088	0.108	0.112
NN	Normalized	0.245	0.091	0.098	0.134
NN	Log	0.168	0.140	0.127	0.220
NN	Log-normalized	0.458	0.117	0.130	0.145

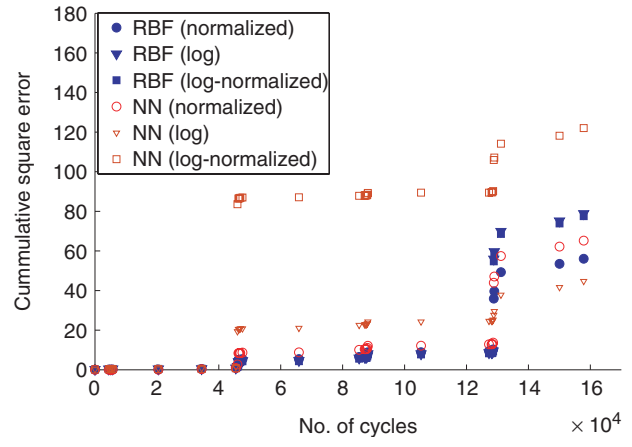

Figure 5. *CSE for crack growth prediction of CT417a sample using different scaling and covariance function types.*

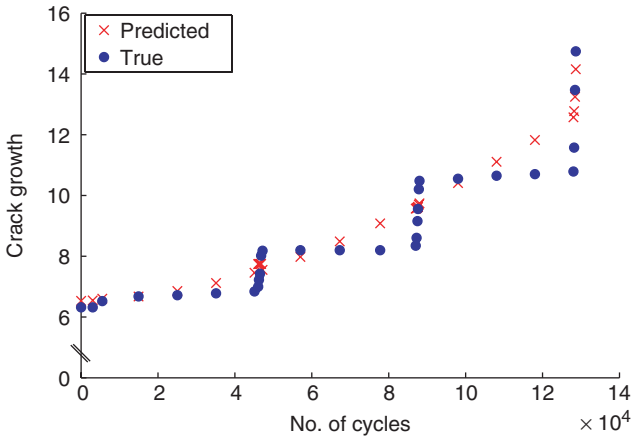

Figure 6. *Crack growth prediction for CT421a with RBF covariance function and normalized scaling.*

Influence of Number of Input Variables on Prediction Performance

The influence of the number of input space variables on prediction performance is also evaluated by considering different input variables in training and test input data. A comparative result of NMSE's for sample CT417a, CT419a, CT421a, and CT423a are shown in Figure 7 for the RBF kernel with different scaling. The number '4' in the horizontal indicates all four input variables (number of cycles, minimum load, maximum load, and load ratio) are considered for generating the input space, whereas the number '3' symbolizes minimum load, maximum load, and load ratio as training space variables, the number '2' symbolizes maximum load and load ratio as training space variables and the number '1' symbolizes only the load ratio as one input space variable. From Figure 7 it can be seen that in general the NMSE has lower value in the case of four input variables compared to three, two, or one variable input spaces.

Influence of Number of Data Points on Prediction Performance

In addition to the above parametric study, the effect of the number of data points in the training data set for finding the optimal hyperparameters and on subsequent prediction performance has been studied. Figure 8 shows the box-plot of NMSE with different percentage of data points randomly selected from the original training space. A total of 100 different combinations of data set were drawn from the original training data set to perform 100 different numerical experiments. The randomized input data points were selected from the training input space (data from CT417a, CT421a, and CT423a) for testing sample CT419a. With 100% training data points the NMSE is shown in the horizontal dashed line and is unique. On the other hand, for the case with 75%, 50% and 25% of the total data points there is variation in the NMSE which is evident from the three box-plot shown in Figure 8. As expected, with decreasing training data points the median NMSE shifted away from the NMSE with 100% data points. Also, it can be seen that the inter-quartile range, which is the difference between 1st quartile and third quartile, increases with reductions in the number of training data points. This shows higher uncertainty in the prediction of hyperparameters, which in turn results in a more inconsistent prediction of crack length.

Collapse Point Forecasting

In prognostics, apart from forecasting statements on the upcoming future events, it is desirable to generate warning signals, ahead in time, regarding any possible catastrophic failures anticipated with confidence. In this section, we review a methodology (Srivastava and Das, 2008) that uses GP to generate a precursor signal which can be used to forecast and map the complete failure of the test samples. As explained earlier, in the GP the predicted output is associated with an 'error bar' (Equation 6) that describes the confidence limit on that prediction. The idea to flag a warning signal in the pre-collapse stage is based on the fact that, given a training set, the prediction variance (uncertainty)

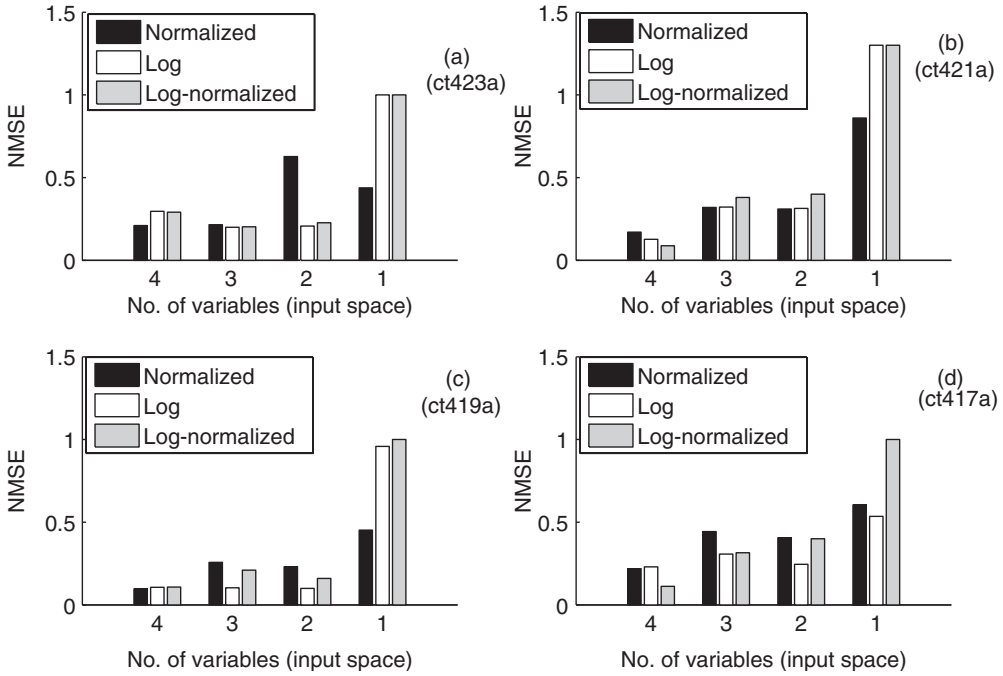

Figure 7. *Influence of the number of input variables on the NMSE prediction using the GP model.*

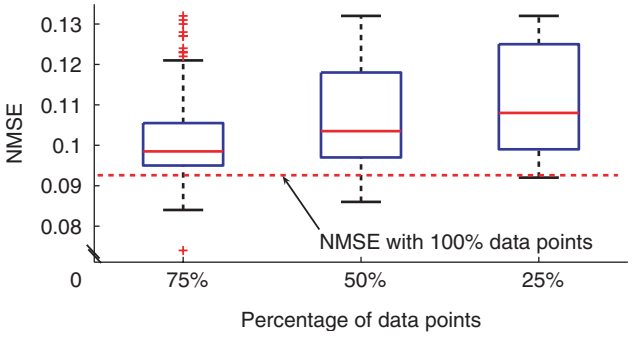

Figure 8. *Percentage of training data used to evaluate the hyperparameters and their influence on the NMSE prediction for sample CT421a.*

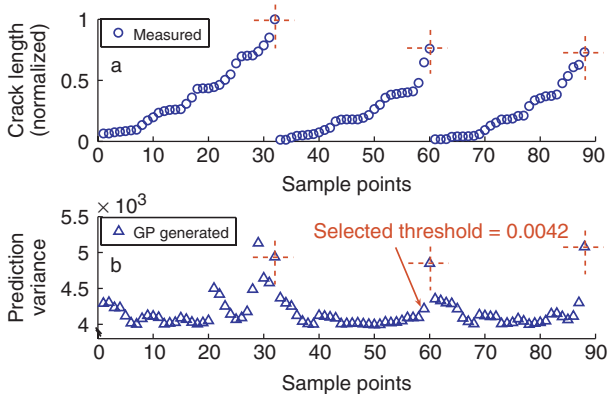

Figure 9. *Determination of the threshold of uncertainty for collapse point prediction from the experimental crack growth data of CT417a, CT419a, and CT423a specimens.*

associated with the forecast of a future (system) collapse of any test case (unknown) will intuitively be greater than predicting a similar collapse example (seen) from the subset of the historical data. The prognostic signal is generated by imposing a threshold on the estimated uncertainties associated with the predictions of the test case. Here the basic assumption is that the training set consists of several examples of collapse patterns and the developed GP model has adequate knowledge of these observed patterns. Once the model is ready, predictions for the same training samples have been made and the associated variance has been obtained. Figure 9 shows the measured crack lengths for different samples and their respective predicted variance at different instances of time. Among all the collapse points (indicated with dotted lines in Figure 9), the ones with minimum variance have been noted and the threshold is set to a value that precedes the minimum variance at that collapse point. For example, at sample point 60 the predicted variance is of minimum and is approximately equal to 0.00475. So a threshold value of 0.0042, which is well below the minimum predicted variance, is chosen to flag a warning signal. For any test case (considered unknown), once the prediction variance crosses this threshold the prognosis signal is set to 'high' status, indicating a probable collapse or a abrupt change in system dynamics. Figure 10 shows the prognostics signal for CT421a sample. The top half of Figure 10 shows the true (experimental) catastrophic failure of CT421a sample, whereas the bottom half shows the GP prediction of warning signal for a possible catastrophic failure. The figure shows that the GP warns of a possible catastrophic failure approximately 200 fatigue

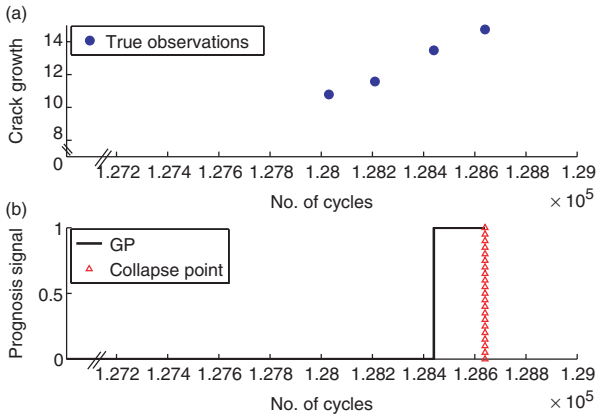

Figure 10. *GP-generated prognostic signal indicating the possible collapse point for CT421a sample.*

Table 2. Number of cycles by which the prognostic signal leads the true collapse point. The prognosis signal is set to high (unity) once the uncertainty associated with each prediction crosses a predefined threshold.

Kernel	Scaling	Test-A (ct423a)	Test-B (ct421a)	Test-C (ct419a)	Test-D (ct417a)
RBF	**Normalized**	**7830**	**24020**	**200**	**90**
RBF	Log	26740	24020	0	90
RBF	Log-normalized	26740	40070	0	90

cycles before the real failure. It is noted that the threshold variance is to be selected in a case-by-case basis depending on the level of confidence required. In addition, Table 2 shows the prediction of number of cycles to real catastrophic failure for various CT samples using the RBF kernel. This table also shows the comparison of predicted number of cycles to failure obtained using different scaling. It is seen that GP with normalized scaling and RBF kernel is able to generate an advance warning signal for all test cases, whereas for log and log-normalized scaling it fails to detect the collapse event in advance for specimen CT419a. This observation further implies that GP is not only sensitive to proper selection of covariance function but also sensitive to selection of proper scaling procedures. The correctness of the warning signal can be improved by taking frequent measurements (using potential drop measurement method or other suitable monitoring method) around the zone where the growth rate of the fatigue crack is much higher. Hence there is an immense potential of using such warning signals for prognosis purpose on real-life systems, especially where the historical data have some correlation to the associated collapse patterns.

CONCLUSION

A covariance function (kernel)-based GP approach has been developed for the time series prediction of fatigue crack growth function under variable loading. From the GP prediction model it is found that there is a good correlation between crack growth prediction and experiment. The prediction result indicates that the GP is able to capture the highly nonlinear dynamics of the crack growth function associated with variable amplitude loading conditions. This implies that the parameters affecting fatigue under variable loading can be implicitly modeled through the GP model using the input–output space correlation. Also, since GP approach models the joint distribution of individual distribution at different instances over the entire fatigue loading regime, it automatically accounts for crack growth acceleration and retardation effect due to random loading. Parametric studies have been conducted to evaluate two different kernels and three different types of scaling. It is found that the anisotropic RBF-based kernel with normalized scaling gives consistently better prediction performance over other kernel functions and scaling combinations. In addition, it is found that the GP algorithm can automatically predict the onset of a collapse mode, and hence can provide early warning of a potential catastrophic failure. It is expected that the developed concepts could be implemented in mission-critical aerospace systems, for online diagnosis and prognosis.

ACKNOWLEDGMENTS

This research was supported by the Air Force Office of Scientific Research, grant number: FA9550-06-1-0309, with Technical Monitor Dr Victor Giurgiutiu.

REFERENCES

Ditlevsen, O. and Olsen, R. 1986. "Statistical Analysis of the Virkler Data on Fatigue Crack Growth," *Engineering Fracture Mechanics*, 25(2):177–195.

Elber, W. 1970. "Fatigue Crack Closure under Cyclic Tension," *Eng Fract Mech.*, 2:37–45.

Gibbs, M.N. 1997. "Bayesian Gaussian Processes for Regression and Classification," PhD thesis, Department of Physics, University of Cambridge.

Harter, J.A. 1999. "AFGROW Users Guide and Technical Manual," Report No. AFRL-VA-WP-1999-3016, Air Force Research Laboratory.

Hidetoshi, F., Mackay, D.J.C. and Bhadeshia, H. 1996. "Bayesian Neural Network Analysis of Fatigue Crack Growth Rate in Nickel Base Supper Alloys," ISIJ International., 36(11):1373–1382.

Liu, Y. and Mahadevan, S. 2007. "Stochastic Fatigue Damage Modeling under Variable Amplitude Loading," *International Journal of Fatigue*, 29(6):1149–1161.

MacKay, D.J.C. 1992. "Bayesian Interpolation," *Neural Computation*, 4(3):415–447.

MacKay, D.J.C. 1993. "Hyperparameters: Optimize or Integrate Out?," In: Heidbreder, G. and Barbara, S. (eds), *Maximum Entropy and Bayesian Methods*, pp.43–60, Kluwer, Dordrecht.

MacKay, D.J.C. 1998. "Introduction to Gaussian Processes," In: Bishop, C.M. (ed.), *Neural Networks and Machine Learning*, Vol. 168, pp. 133–165. Springer, Berlin.

Miner, M.A. 1945. "Cumulative Damage in Fatigue," *Journal of Applied Mechanics 12, Trans. ASME*, A-159–164.

Neal, R.M. 1993. "Probabilistic Inference using Markov Chain Monte Carlo Method," Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.

Neal, R.M. 1996. *Bayesian Learning for Neural Networks*, Springer-Verlag, New York.

Newman Jr, J.C. 1982. "Prediction of Fatigue Crack Growth under Variable-amplitude and Spectrum Loading using a Closure Model," ASTM STP 761:255–277.

Newman Jr, J.C. 1984. "A Crack-opening Stress Equation for Fatigue Crack Growth," *International Journal of Fracture*, 24:R131–R135.

Newman Jr, J.C. 1992. "FASTRAN-II – A Fatigue Crack Growth Structural Analysis Program," NASA Technical Memorandum 104159, Langley Research Center.

Rasmussen, C. and Williams, C. 2006. *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA.

Ray, A. and Patankar, R.P. 2001a. "Fatigue Crack Growth under Variable Amplitude Loading – Part I: Model Formulation in State-Space Setting," *Journal of Applied Mathematical Modeling*, 25(11):979–994.

Ray, A. and Patankar, R.P. 2001b. "Fatigue Crack Growth under Variable Amplitude Loading – Part II: Code Development and Model Validation," *Journal of Applied Mathematical Modeling*, 25(11):995–1013.

Sander, M. and Richard, H.A. 2005. "Finite Element Analysis of Fatigue Crack Growth with Interspersed Mode I and Mixed Mode Overloads," *International Journal of Fatigue*, 27:905–913.

Spencer Jr, B.F. and Tang, J. 1989. "Stochastic Approach to Modeling Fatigue Crack Growth," *AIAA Journal*, 27:1628–1635.

Srivastava, A.N. and Das, S. 2008. "Detection and Prognostics on Low Dimensional Systems," Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews (in press, paper available at http://dashlink.arc.nasa.gov/static/dashlink/media/topic/LaserPrognosis-Ashok.pdf).

Suresh, S. 1991. *Fatigue of Materials*, Cambridge University Press, Cambridge, UK.

Williams C.K.I. 1997. "Computing with infinite networks," In: Mozer, M.C., Jordan, M.I., and Petsche, T., (eds.), *Advances in Neural Information Processing Systems*, Vol. 9. MIT Press, Cambridge, MA.