ABSTRACT
This paper presents a hybrid model of fatigue crack growth in metallic alloys that combines a physics-based state-space model and a data-driven Gaussian Process (GP) and can be applied for variable-amplitude aircraft service loads. The physics driven part has a state variable formulation with underlying physics based on the crack closure concept used in the FASTRAN-II model. The data driven part of the model uses a kernel based GP regression model, which is being considered as the equivalent of an infinite neuron neural network model. The state space part is linked to the data driven part via a constraint factor $\alpha$, which accounts for the effects of the stress state on the plastic-zone size. Through simulations it is shown that the hybrid model gives better predictions of fatigue crack length and growth rate as compared to both the pure physics model and the pure data driven GP model.

INTRODUCTION
Intensive research and development efforts are currently being dedicated to on-board systems for integrated health management of aerospace structures, which need to be complemented by a prognosis approach. In this regard, a simple form of prognostics, known as a life usage model, is widely used and is based on gathering statistical information about how long a component lasts before failure, and uses these statistics, collected from a large population sample, to make remaining life predictions for individual components. However, these predictions are not based on measured characteristics of the individual components. Another method that is widely practiced and constantly been improved is a physics based approach. This is primarily based on fracture mechanics, as far as fatigue failure is concerned. 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. A detailed review on various data driven prognosis models is given in reference [1]. Prediction of fatigue crack growth, even under constant amplitude loading, is not easy. This is mainly because the manner in which various parameters, such as loads, material properties and crack geometries, interact with each other to affect the crack propagation is not clearly understood [2]. In this sense, prediction of fatigue crack growth under typical service loads experienced by aircraft structures is an even more involved task because of loading transient effects in the fatigue crack growth rates, which may affect the fatigue life significantly. A vast majority of fatigue crack growth models [3], has been restricted to constant-amplitude stress cycles, where the crack opening stress is assumed to be constant. These models cannot be used directly for
variable-amplitude service loading, because they cannot capture the intrinsic dynamic behavior. The FASTRAN-II [4-6] and the AFGROW [7] models capture the dynamics under variable amplitude by incorporating the crack opening stress as a physical variable in the fatigue crack growth model. Under variable-amplitude loading, these models typically rely on a memory-dependent variable (e.g., crack opening stress) that requires knowing the load history. In general the model is more accurate if the load history is considered, but that is not always the case. Accounting for the load history makes a model more complicated and time-consuming for calculations, and load histories are not always available. It would be beneficial to establish a procedure whereby load history is not explicitly required but is accounted for in other ways, such as an additional equation for an extra (state) variable with memory, which would likely make the model more computationally efficient. The state space model presented in [8-11], though based on the original physical concept of the FASTRAN model [4-6], significantly reduces the computation time and memory requirements. A detailed comparison of computational time requirements under different variable load cases for both the FASTRAN and state space models is given in [9]. The lower computational requirements for the state space model would also help in possible future implementations of on-board computations for a prognosis model.

A parallel effort is being undertaken by the information science community to develop and improve data driven approaches. In the context of structural life prediction, many of the existing approaches to data driven prognosis have used Multi-layer perceptron (MLP) artificial neural networks to model the prognosis system [12, 13]. However, these networks have drawbacks: Firstly, the learning algorithm is generally believed to be implausible from a biological point of view, for example in requiring synapses to act bi-directionally, and being very slow to train. Secondly, there is no uncertainty in a neural network implemented in this way: given an input, the network produces a unique output. This is undesirable – the real world is dominated by uncertainty, and predictions without it are of limited value.

The above discussion motivates exploring other methods to predict structural damage if its growth is highly uncertain and non-linear. Unlike neural networks, GP regression [14-16] is not based on a biological model, provides an uncertainty measure and does not require the lengthy ‘training’ that a neural network does.

The physics-based state-space and the data-driven GP models have their own merits and demerits. Though the state-space model has lower computational cost than the FASTRAN model, it does not consider the uncertainty that is prevalent in fatigue of metallic structures. The GP model, on the other hand, has explicit uncertainty measures and it requires less training data and computational time compared to neural networks, but it lacks the ability to retain the shape/trend of a prediction in the absence of large amounts of physical information. The present paper discusses a hybrid approach, which integrates the state-space and GP models to address these problems.

STATE-SPACE AS A PHYSICS-BASED MODEL

In the two state-space [8-10] model used here the inputs are maximum stress $S_{k}^{\text{max}}$ and minimum stress $S_{k}^{\text{min}}$, whereas the outputs are crack-length increments $\Delta a_{k}$ and crack opening stress $S_{k}^{o}$. The crack growth equation in the state-space model is similar to the Paris equation [3] modified for crack closure, which has been used in FASTRAN [6]. Equation1 describes the Paris model and the first output state:
\[ \Delta a_k = h(\Delta K_k^{eff}) \quad \text{with} \quad h(0) = 0 \]
\[ \Delta K_k^{eff} = \sqrt{a_{k-1}} F(a_{k-1}, w) \left\{ \left( S_k^{\min} - \max(S_k^{\min}, S_{k-1}^{\min}) \right) U(S_k^{\min} - S_{k-1}^{\min}) \right\} \]

Where, \( h(\cdot) \) is a non-negative monotonically increasing function, \( a_{k-1} \) and \( S_{k-1}^{\min} \) are the crack length and the crack opening stress in the \((k-1)\)th cycle; \( w \) is the width of the sample and \( F(\cdot, \cdot) \) is a geometric factor that for the CT sample is given by [17]:
\[ F(a_{k-1}, w) = \frac{1}{\sqrt{\alpha}} \frac{2 + \alpha}{(1 - \alpha)^{3/2}} (0.886 + 4.69\alpha - 13.32\alpha^2 + 14.72\alpha^3 - 5.6\alpha^4) ; \quad \alpha = \frac{a_{k-1}}{w} \]

The second output state, the crack opening stress \( S_k^{\infty} \) is given by:
\[ S_k^{\infty} = \left( \frac{1}{1 + \eta} \right) S_{k-1}^{\infty} + \left( \frac{\eta}{1 + \eta} \right) S_{k-1}^{\max} + \left( \frac{1}{1 + \eta} \right)(S_k^{\max} - S_{k-1}^{\max}) U(S_k^{\max} - S_{k-1}^{\max}) \]
\[ + \left( \frac{1}{1 + \eta} \right)(S_k^{\max} - S_k^{\max-old}) U(S_k^{\min} - S_{k-1}^{\min})[1 - U(S_k^{\max} - S_{k-1}^{\max})] \]

Where, in Eq. 1 and 3, \( U(x) \) is the unit Heaviside (step) function and for thickness \( t' \), yield stress \( S' \), and Young’s modulus \( E \) the parameter \( \eta = (t'S')/(2wE) \). In Eq. (3) the crack opening stress under constant-amplitude load is given as:
\[ S_k^{\max} = S_k^{\max} \left( S_k^{\max}, S_k^{\min}, \alpha_k, F(a_{k-1}, w) \right) \]
\[ \left\{ \max \{ (A_k^0 + A_k^1 R_k + A_k^2 (R_k)^2 + A_k^3 (R_k)^3), R_k \} S_k^{\max} \quad \text{for} \quad R_k \geq 0 \right\} \]
\[ \left\{ (A_k^0 + A_k^1 R_k) S_k^{\max} \quad \text{otherwise} \right\} \]

Where \( S_k^{\max-old} = S_k^{\max} \left( S_k^{\max}, S_k^{\min}, \alpha_k, F(a_{k-1}, w) \right) \), \( R_k = S_k^{\min} / S_k^{\max} \) is the \( k \)th cycle load ratio and \( A_k^i (i=0, 1, 2, 3) \) are empirical parameters, the details of which can be found in the FASTRAN manual [6] or from [8-11]. The constraint factors \( \alpha_k \) used to find the empirical parameters \( A_k^i \) are calculated from:
\[ \alpha_k = \alpha_k^{\min} + (\alpha_k^{\max} - \alpha_k^{\min}) \left\{ \frac{\ln(\Delta a_k)}{\ln(\Delta a_k^{\max})} - \frac{\ln(\Delta a_k^{\min})}{\ln(\Delta a_k^{\max})} \right\} \]

**GAUSSIAN PROCESS AS A DATA-DRIVEN MODEL**

The goal of Gaussian process forecasting is to compute the distribution \( p(a^{(N+1)} | D = \{ \tilde{x}^{(n)}, a^{(n)} \}_{n=1}^{N}, \tilde{x}^{(n+1)} \} \), i.e., to compute the probability distribution of the damage index \( a^{(N+1)} \) given a test input \( \tilde{x}^{(n+1)} \) and a set of \( N \) training points \( D = \{ \tilde{x}^{(n)}, a^{(n)} \}_{n=1}^{N} \). Let us define our damage index data vector \( \tilde{a} \) to be a Gaussian process with a covariance matrix \( C_N \) and a mean vector \( \mu = 0 \). By doing this we have bypassed the step of expressing individual priors on the noise and the modeling function by combining both priors into the covariance matrix \( C_N \). Now the Gaussian distribution over \( a_{N+1} \) can be written as
\[ p(a_{N+1} | D, C(x_i, x_j), x_{N+1}, \Theta) = \frac{p'(a_{N+1} | C(x_i, x_j), x_{N+1}, \{x_N\} \Theta)}{p(a_N | C(x_i, x_j), \{x_N\} \Theta)} \]

\[ = \frac{1}{Z} \exp \left( -\frac{(a_{N+1} - \hat{a}_{N+1})^2}{2\sigma^2_{a_{N+1}}} \right) \]  

(6)

Where \( Z \) is an appropriate normalizing constant and the mean and variance of the new distribution are, respectively, defined as:

\[ \hat{a}_{N+1} = k^{T} C^{-1}_{N} a_{N} \quad \sigma^2_{a_{N+1}} = k - k^{T} C^{-1}_{N} k \]  

(7)

There are many possible choices of prior 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, we are required to specify a function that will generate a positive definite covariance matrix for any set of inputs. A simple non-stationary neural network based [14] covariance function is used for the current Gaussian process model and is given by:

\[ C_N = \Theta_i \text{Sin}^{-1}(\frac{(\hat{a}_i^p)^T \Theta_2 (\hat{a}_j^q)}{\sqrt{(1 + (\hat{a}_i^p)^T \Theta_2 (\hat{a}_j^p))(1 + (\hat{a}_j^q)^T \Theta_2 (\hat{a}_j^q))}}) + \Theta_3 \]  

(8)

The parameters \( \Theta_i \), \((i=1,2,3)\) are adjusted to maximize the log likelihood \( L \), given by

\[ L = -\frac{1}{2} \log \det C_N - \frac{1}{2} a^T C^{-1}_N a - \frac{N}{2} \log 2\Pi \]  

(9)

These hyper parameters are initialized to reasonable values and then, the conjugate gradient method is used to search for their optimal values.

**HYBRID MODEL**

In the FASTRAN model the constraint factor \( \alpha_k \) is used to account for the effects of the stress state on plastic-zone size, which elevates the tensile flow stress for the intact elements [6] in the plastic zone in front of the crack tip. The constraint factor varies with the crack growth rate [5], which in turn depends on the crack-length and stress amplitude. Note that the stress state effect on the plastic-zone size is highly dependent on the microstructural parameters of the material. Due to this fact, calculation of \( \alpha_k \) using Eq. (5) may sometimes result in an erroneous prediction. In addition, the error accumulates due to the power law prediction of the crack growth rate. The current hybrid formulation allows calculating the constraint factor using Eq. (5), but in the current approach it is suggested to use input crack length in Eq. (5) from a GP data-driven model rather than from an error accumulating state-space model. This not only incorporates uncertainty, which exists due to material and load variability, into the state-space model, but also avoids error accumulation. However, note that there should be enough data available for the training of the GP model, otherwise error accumulation similar to that of the “pure” state-space model would occur.

**EXPERIMENTAL RESULTS**

**Crack Growth Measurement:** Compact Tension (CT) samples 6.31 mm thick made of Al 2024 T3 were used. The CT specimens were fabricated according to ASTM
E647-93 with a width of 25.4 mm (from the center of the pin hole to the edge of the specimen) and an initial notch length of 5 mm.

The experiments were performed in an Instron 1331 servo-hydraulic load frame operating at 20 Hz. Two samples, CT407a and CT408a, were tested. To simulate real flight conditions a typical center wing load spectrum was programmed into the digital controller of the load frame. The spectrum, as measured by the load-cell, is shown in Fig. (1).

**Fractography:** The fracture surface of sample CT-408a was examined using Scanning Electron Microscopy (SEM) in an FEI XL-30 operating at 15 kV. The fracture surface showed features typical of fatigue crack propagation that correlated closely with the load spectrum used to propagate the crack. Examples of the features observed and their correlation with the crack length are shown in Fig. 2.

![Fig. 1 Typical center wing flight spectrum](image)

![Crack Growth Direction](image)

![Fig. 2 Crack growth curves for CT407a and 408a and striation patterns for CT408a](image)

There were two different regions forming bands that alternated along the crack growth direction. One type of region was between 130 and 160 micrometers long and had small striation spacing, indicating slow crack growth. These regions correlated with the low load periods of the load spectrum. An example of the fracture surface appearance for a region of slow crack growth is shown in Fig. (2a). This corresponds to a location within a few millimeters from the notch tip. The slow crack growth in these regions, which correspond to a periods right after a stage of tensile overload, can be attributed to a combination of the lower driving force and the well-known retardation effect associated to tensile overloads during fatigue crack growth [3].
The second type of region had lengths between 400 and 1000 micrometers and had larger striation spacing, indicating faster crack growth. The positions of these regions correlate with the overload periods of the load spectrum used. An example of the fracture surface for one of these regions is shown in Fig. (2b). This location is close to the middle of the unbroken ligament. The figure also suggests that ductile rupture is starting to contribute, given the presence of a dimple (enclosed in the circle).

Finally, an example of the fracture surface for the fast crack growth region closer to the final fracture is shown in Fig. 2(c). The magnification of this micrograph is lower than in the two previous pictures, so that the overall aspect of the fracture surface could be shown. The fracture mechanism was a combination of striations and beach marks, as well as ductile rupture, as expected for unstable fatigue crack growth.

SIMULATION RESULTS

“Pure” State Space Model: The material, Paris’ law \((C \text{ and } m \text{ in } C(\Delta K_{eq})^m)\) and constraint factor constants (used in Eq. 5) are taken from [9] and are as follows: \(E=71.75 \text{ GPa}, \ S_y=328 \text{ MPa}, \text{ ultimate strength } S_u=473 \text{ MPa}, \ C=5 \times 10^{-11} \text{ m/cycle}(\text{MPa.m}^{1/2})^m, \ m=4.07, \ \alpha_{\text{max}}=1.8, \ \alpha_{\text{min}}=1.1, \ \Delta a_{\text{max}}=7.1 \times 10^{-7} \text{ m}, \) and \(\Delta a_{\text{max}}=9.8 \times 10^{-8} \text{ m}.\) The load spectrum used for the simulation was taken directly from the load cell output and it was shown in Fig. (1). Two cases with \(\alpha_{\text{max}}=(f/100) \times 1.8, \ f=100 \text{ and } f=85, \) are simulated and the resulting crack lengths are shown in Fig. (3).

The purpose of running the state space model for two different values of \(\alpha_{\text{max}}\) is to check the sensitivity of the model predictions to variations in this parameter, which is supposed to be assumed constant. It is found from Fig. (3) that the state space model gives a fairly good prediction up to the first large overload at 47 kilocycles. After 47 kilocycles the simulated crack growth rate increases rapidly, leading to a large mismatch between experiment and prediction. Note that the predicted crack growth rate is very sensitive to the chosen value of \(\alpha_{\text{max}},\) as expected. This implies that proper selection of \(\alpha_{\text{max}}\) is necessary. However, it is not always possible to select a proper \(\alpha_{\text{max}}\) (hence \(\alpha\)) due to loading and material uncertainties. This uncertainty in the value of \(\alpha\) leads to the idea of using a data-driven approach, as discussed next.

“Pure” Gaussian Process Model: The input in the GP model is number of cycles whereas the output is crack length. This output is predicted as a time series using Eq. (7), which provides the expected value, i.e., the mean, and the variance of the target data points given the current system state. The training and test data for the simulation are taken from the experimental fatigue crack growth data obtained for sample CT-408a. The first step is using the training data to find the optimum values of the hyper parameters of the covariance function using the conjugate gradient optimization technique by minimizing the negative logarithm likelihood function. The numerical value of this function versus the number of search iterations is shown in Fig. (4).
It is evident that the function converges to its lowest value, which in turn leads to a prediction of the optimal set of hyperparameters. These hyperparameters were used to predict the crack lengths for two different cases: in the first case the training data are taken from the first 14 experimental points, i.e., up to 80 kilocycles (refer to Fig. 5) and continuous crack growth data are predicted from 80 to 170 kilocycles.

In the second case the full set of 32 experimental points are used as the training data to predict a smooth crack growth curve between 40 to 170 kilocycles. The corresponding expected values and their 3σ confidence intervals are plotted in Fig. 5.

From this figure it can be seen that as the number of training points increases the procedure provides a better 3σ confidence interval. However, note from Fig. 5b that even though the test data are available for the whole domain of interest, the GP model is unable to retain the shape of the measured crack growth curve. This is because of the lack of physical information in the Gaussian process model, which results in a smooth “curve fit” of the data. This suggests that a “hybrid” physics and data-driven approach should be used. Results for this case are discussed in the next subsection.

**Hybrid SS-GP (State Space and Gaussian Process) Model:** for the hybrid approach suggested above, the input crack length \( \alpha_k \) in Eq. (5) is being constantly supplied from a GP data driven model rather than from an error accumulating state space model. Once \( \alpha_k \) is calculated, the future crack length \( \alpha_k \) is recalculated using the state-space model and the process continues in a loop for the fatigue cycles of interest. The training of the GP model is performed with the full set of experimental points from sample CT-408a, whereas the validation of the hybrid model prediction is done against the experimental data from sample CT-407a. Note that the case of \( \alpha_{\text{max}}^{\text{max}} =1.8 \) is considered. The crack growth rate is shown in Fig. (6) and it is evident that the hybrid model results in better prediction than those of the state-space model while reproducing the shape of the experimental crack growth curve fairly well.
CONCLUSION

The hybrid state space and Gaussian Process (SS-GP) model gives a better prediction of crack length evolution and crack growth rate compared to both a physics based “pure” state space model and a data driven “pure” GP model. The hybrid model not only implicitly incorporates uncertainty into physics based state space model but also avoids exponential error accumulation due to the Paris type formulation.

ACKNOWLEDGEMENT

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