Probabilistic Prognosis of Non-Planar Fatigue Crack Growth

Patrick E. Leser[1], John A. Newman[1], James E. Warner[1], William P. Leser[1], Jacob D. Hochhalter[1], and Fuh-Gwo Yuan[2]

[1] *NASA Langley Research Center, Hampton, VA, 23681, USA*
patrick.e.leser@nasa.gov

[2] *North Carolina State University, Raleigh, NC, 27695, USA*

ABSTRACT

Quantifying the uncertainty in model parameters for the purpose of damage prognosis can be accomplished utilizing Bayesian inference and damage diagnosis data from sources such as non-destructive evaluation or structural health monitoring. The number of samples required to solve the Bayesian inverse problem through common sampling techniques (e.g., Markov chain Monte Carlo) renders high-fidelity finite element-based damage growth models unusable due to prohibitive computation times. However, these types of models are often the only option when attempting to model complex damage growth in real-world structures. Here, a recently developed high-fidelity crack growth model is used which, when compared to finite element-based modeling, has demonstrated reductions in computation times of three orders of magnitude through the use of surrogate models and machine learning. The model is flexible in that only the expensive computation of the crack driving forces is replaced by the surrogate models, leaving the remaining parameters accessible for uncertainty quantification. A probabilistic prognosis framework incorporating this model is developed and demonstrated for non-planar crack growth in a modified, edge-notched, aluminum tensile specimen. Predictions of remaining useful life are made over time for five updates of the damage diagnosis data, and prognostic metrics are utilized to evaluate the performance of the prognostic framework. Challenges specific to the probabilistic prognosis of non-planar fatigue crack growth are highlighted and discussed in the context of the experimental results.

1. INTRODUCTION

Probabilistic damage prognosis is an essential aspect of any aerospace structural health management system. The ability to predict, with confidence and in the face of uncertainties, how damage will propagate in a structure can be an invaluable tool to operators making mission- or safety-critical decisions. The applications of probabilistic damage prognosis are widespread, including but not limited to concepts such as condition-based maintenance (Farrar & Worden, 2012) and long endurance missions (e.g., unmanned aircraft systems or spacecraft) for which maintenance is not an option. The degree of fidelity in the models used to make these forecasts can vary from simple approximations to high-fidelity finite element models with millions of degrees of freedom. Recently, NASA has adopted an approach referred to as digital twin (DT) as part of the Convergent Aeronautics Solutions (CAS) project. Although a commonly used term that is sometimes considered synonymous with structural health management, this particular form of DT focuses on the high-fidelity modeling and coupling of fluid-structure interactions, material response, and multi-scale damage growth. In regards to fatigue damage prognosis in particular, this DT framework rests upon a foundation comprising the concept of damage and dura-

bility simulators (DDSim) (Emery, Hochhalter, Wawrzynek, Heber, & Ingraffea, 2009).

DDSim is a hierarchical prognosis methodology consisting of three primary levels. Level I deals with a rapid search and assessment of structural sub-regions possessing the potential to initiate life-limiting damage. Using linear elastic fracture mechanics (LEFM), this search is conducted at a subset of nodes in a finite element (FE) model where stresses from the FE solution inform analytical models to approximate damage driving forces. Level I aims to provide an estimate of the number of cycles to failure, or remaining useful life (RUL), with damage initiating at various locations in the FE model. At Level II, these approximations are improved upon using high-fidelity fracture simulations. The potential for life-limiting damage obtained from Level I is used to inform the initial damage conditions (i.e., most likely sizes, locations, and orientations) chosen for these simulations so as to reduce the computational effort involved. Level II is used to more accurately predict the structure's RUL at the macroscale. Both of these levels are typical instances of damage prognosis methodologies. However, a third level exists that uses high-fidelity models to predict the number of cycles consumed by damage initiation and microscale damage growth, two mechanisms that often account for the majority of a structure's fatigue life (Vasudevan, Sadananda, & Glinka, 2001). The work presented herein focuses on Level II, but with the intention of utilizing information from Levels I and III.

While microscale growth can consume a significant portion of a structure's fatigue life, the risk associated with life predictions made in the macroscale growth regime of Level II is higher; this is the regime in which detectable cracks typically lie, and thus is the regime right before failure or retirement (Farrar & Lieven, 2007; Chang, Markmiller, Ihn, & Cheng, 2007; Banerjee, S, & Chijioke, 2014). With higher risk comes higher impact, as accurate and reliable predictions at this stage of fatigue damage growth have the potential to safely and significantly extend the useful life of a structure. Furthermore, since structural health monitoring (SHM) or non-destructive evaluation (NDE) can be used for damage diagnosis at these length scales, a unique opportunity exists in which these data can be used to reduce uncertainty in RUL forecasts on a structure-by-structure basis. Although not critical for the understanding of this paper, it should be noted that SHM is defined by its use of mounted sensors for online damage diagnostics (i.e., monitoring during operation), while NDE refers to more traditional methods for damage detection that are carried out offline (Farrar & Worden, 2007).

There has been extensive research in the area of coupling SHM and NDE with damage prognosis, a few notable examples of which are as follows. Liu and Mahadevan proposed a new way to quantify the uncertainty in equivalent initial flaw sizes, which they used in conjunction with an analytical fatigue crack growth model to produce probabilistic predictions of fatigue life in metallic specimens (Liu & Mahadevan, 2009). Gobbato et al. combined NDE with probabilistic models of both damage evolution and future aerodynamic loading through Bayesian inference (Gobbato, Conte, Kosmatka, & Farrar, 2012). Peng et al. demonstrated a direct link between Lamb wave-based damage detection and probabilistic prognosis for a metallic lap joint using analytical crack growth laws and Bayesian inference (Peng, He, et al., 2015). In recent years, researchers have also investigated fatigue damage prognosis for composite materials, joining Lamb wave-based diagnosis with analytical stiffness degradation models through Bayesian approaches (Chiachıo, Chiachıo, Saxena, Rus, & Goebel, 2013; Peng, Liu, Saxena, & Goebel, 2015). Note that all of these works are probabilistic. Without probability, it is often impossible to make predictions with confidence. Without confidence, the risk in making mission- or safety-critical decisions becomes unacceptable.

Much of the research conducted and discussed thus far would be considered Level I approaches. The true key to Level II lies in the fidelity of the modeling. Application to damage in real-world structures is key to unlocking the potential of damage prognosis. High-fidelity FE models are now capable of modeling the complexities of real-world damage, but these models are often prohibitively time-intensive. This limitation is the primary reason these models are seldom used in probabilistic prognosis and do not appear in the afore-

mentioned research efforts. Many of these works rely on Bayesian inference and sampling techniques such as Markov chain Monte Carlo (MCMC) to develop probabilistic life predictions. MCMC can require anywhere from thousands to millions of simulations to provide meaningful results (Smith, 2013). FE-based simulations which conservatively take on the order of a few hours to complete could result in total prediction times on the order of years. While efforts are being made to parallelize these statistical techniques (Ter Braak, 2006; Vrugt et al., 2009; Laloy & Vrugt, 2012; Neiswanger, Wang, & Xing, 2014), parallelization is, in general, not a feasible option at this time due to the serial nature of MCMC techniques.

Some researchers have taken strides to bridge the gap between probabilistic prognosis and high-fidelity modeling, primarily through the use of machine learning. In general, the time consuming aspects of high-fidelity damage growth simulations can be replaced by a surrogate model, trained via supervised machine learning, which can quantitatively represent the primary features of the high-fidelity model being replaced but at a decreased computational cost. Sankararaman et al. trained surrogate models to replace expensive finite element solutions of crack driving forces for a cylindrical specimen subjected to multi-axial loading using a characteristic plane approach (Sankararaman, Ling, Shantz, & Mahadevan, 2011; Sankararaman, Ling, & Mahadevan, 2011). Expanding upon this work, Ling and Mahadevan coupled the surrogate model-based, characteristic plane approach with damage diagnosis data to forecast fatigue damage growth in aluminum specimens with quantified uncertainties (Ling & Mahadevan, 2012). As a natural extension, Hombal et al. developed a two-stage planar approximation for non-planar crack growth (Hombal, Ling, Wolfe, & Mahadevan, 2012). A more advanced surrogate modeling methodology was proposed by Hombal and Mahadevan to predict three-dimensional damage growth under multi-axial, time varying fatigue loading (Hombal & Mahadevan, 2013). The complex crack growth was simulated in a reduced-order space, allowing for supervised learning without the need for planar approximations. The resultant crack growth simulations benefited from a significant reduction in computation times.

The aforementioned research into rapid, high-fidelity damage prognosis is promising; however, there remains a dearth of flexibility in these proposed models. For example, the three-dimensional surrogate modeling approach presented in (Hombal & Mahadevan, 2013) is arguably the most sophisticated technique discussed above since it does not require any planar assumptions. However, since the crack growth steps are internal to the surrogate model, the crack growth parameters are fixed unless accounted for in the initial training matrix. At best, this prevents the use of model selection algorithms to determine an ideal crack growth law. At worst, this means that uncertainty in crack growth rate parameters cannot be accounted for in a prognostic framework. The latter issue is detrimental, as a large portion of the uncertainty in RUL predictions for fatigue-driven damage results from scatter in the crack growth rate parameters (Johnston, 1983; Gope, 1999). While surrogate modeling is an excellent approach for reducing computation times of high-fidelity models, care must be taken not to restrict the dimensionality of the parameter space considered in the Bayesian inverse problem. This is especially important since supervised learning with high-dimensional input spaces can cause both a decrease in predictive accuracy and an increase in storage requirements.

Recently, (Leser et al., 2016) proposed an alternative approach to reducing computation times associated with high-fidelity damage growth modeling. To address the issue of flexibility, the surrogate model is confined to only a subcomponent of the overall damage growth model. Particularly, the surrogate model replaces the high-fidelity, FE-based computation of the damage driving forces; e.g., stress-intensity factors (SIF) or energy release rates. It was determined that the solution of the finite element system of equations was the primary driver of the exorbitant computation times. By restricting the surrogate model to this portion of the modeling process, model parameters such as those associated with the crack growth rate were not dependent on the machine learning process. As a result, both high degrees of fidelity and flexibility were achieved while simultaneously reducing

computation times by over three orders of magnitude. The present paper expands on this work by providing a more rigorous demonstration of the predictive capabilities offered by the surrogate modeling approach in the context of non-planar crack growth. In particular, the effects of noise in two dimensions is discussed along with an example of model discrepancy and how it poses dangers in the context of probabilistic prognosis.

2. PROGNOSIS FRAMEWORK

The prognosis framework is composed of the following four components: (i) parameter estimation and uncertainty quantification through Bayesian inference, (ii) global sensitivity analysis to determine the parameters that contribute the most variance to the final prediction, (iii) the crack growth model, and (iv) the prognostic metrics used to quantitatively assess the performance of the framework. While presented in the order above for clarity, the first three components interact with each other at various stages when forming the prognosis and are order-independent.

2.1. Uncertainty Quantification and Propagation

The methods used here are based on the incorporation of SHM or NDE with a given damage model in order to make a prediction of remaining useful life for a unique component or structure. This can be accomplished by making discrete observations of the damage state throughout the life of the monitored component, combining this information with prior knowledge (e.g., knowledge of the component geometry, applied loads, and the material), and inversely quantifying the uncertainties in the damage model through Bayesian inference. This uncertainty can then be propagated back through the model, allowing for extrapolation to future time instances. Bayes Theorem also provides flexibility in how these predictions are updated as more observations are made.

The relationship between a parameter-dependent model response, $f_k(Q)$, measurement errors, ε_k, and experimental measurements,Υ_k, is given by the statistical model

$$\Upsilon_k = f_k(Q) + \varepsilon_k, \tag{1}$$

where $f_k(Q)$, ε_k, and Υ_k are random variables and the measurement errors are assumed to be unbiased, independent and identically distributed. Here, Q, also a random variable, denotes the model parameters, and has realizations q. The index of the available observations, $k = 1, ..., n_{obs}$ where n_{obs} is the total number of observations. The solution to the Bayesian inverse problem is the posterior density, $\pi(q|\upsilon_{obs})$, which is the best estimate of the parameter densities based on experimental observations and prior knowledge of the parameter distributions, $\pi_0(q)$. Formally, the relationship between the posterior density, the prior density, and the observations is given by Bayes' Theorem, which takes the form

$$\pi(q|\upsilon_{obs}) = \frac{\pi(\upsilon_{obs}|q)\pi_0(q)}{\pi(\upsilon_{obs})} = \frac{\pi(\upsilon_{obs}|q)\pi_0(q)}{\int_{\mathbb{R}^p} \pi(\upsilon_{obs}|q)\pi_0(q)dq} \tag{2}$$

where p is the number of parameters and defines the dimension of the integral in Equation 2. Assuming normally distributed errors, $\varepsilon_k \sim N(0, \sigma^2)$, damage diagnosis data are incorporated through the likelihood,

$$\pi(\upsilon_{obs}|q) = \frac{1}{(2\pi\sigma^2)^{n_{obs}/2}} e^{-SS_q/2\sigma^2}, \tag{3}$$

where SS_q is the sum of squares error between the model response and the observed data, defined as

$$SS_q = \sum_{k=1}^{n_{obs}} [\upsilon_{obs,k} - f_k(q)]^2. \tag{4}$$

Equation 2 can be solved directly for simple problems of low dimensionality, but, for most practical models, the direct solution to the inverse problem becomes intractable. Quadrature and sparse grid techniques can be used in certain cases with low dimension; i.e., $p \leq 6$ (Smith, 2013). An alternative approach is to use Markov chain Monte Carlo (MCMC) techniques. Utilizing assumed attributes of the posterior density, such as those used to define Equation 3, Markov chains can be constructed for the model parameters based on the observed measurements. The stationary distributions of the chains con-

structed in this manner is the sought-after posterior density, thus approximating a solution to the inverse problem.

In general, when using MCMC techniques, increasing the number of parameters requires an increasing number of samples to determine the stationary distribution of the Markov chains. Before implementing a MCMC technique for parameter estimation, a global sensitivity analysis can be conducted that allows for a quantified ranking of parameter contribution to the uncertainty in the predicted quantity of interest. For large problems, reducing the dimension of the parameter space is often necessary. Regardless, global sensitivity analysis also provides insight into the problem and can prove valuable whether parameters are eliminated from consideration during the parameter estimation process or not.

2.2. Global Sensitivity Analysis

Variance-based global sensitivity analysis can be conducted to determine the contribution of parameter uncertainty to the output uncertainty (Saltelli, A., Ratto, M., Andrtes, T., Campolongo, F., Cariboni, J., Gatelli, D., Michaela, S. & Tarantola, S., 2008; Smith, 2013). By analyzing parameter sensitivity, non-influential parameters can be identified; i.e., the parameter, Q_i for $i = 1, \ldots, p$ where the influence, $\mathcal{I}(Q_i) \approx 0$. These are parameters that do not have a significant effect on the output uncertainty. As such, these parameters can be removed from the Bayesian inference procedure. This is critical because the number of samples required to reliably represent the posterior distribution is directly dependent on the number of parameters, p, i.e., the dimension of the multivariate distribution. Fewer samples means a faster solution to the inverse problem.

Russian mathematician Ilya M. Sobol´ is credited with developing the variance-based measures of sensitivity referred to as the Sobol´ indices. Consider the nonlinear model

$$Y = f(Q_1, Q_2, \ldots, Q_p). \tag{5}$$

Assuming the model inputs, Q_i, are independent, the output variance of Y can be represented by the ANOVA-HDMR decomposition (Rabitz & Aliş, 1999; Ma & Zabaras, 2010). Here, the truncated, second-order decomposition is used such that

$$\mathrm{Var}(Y) = \sum_{i=1}^{p} V_i + \sum_{1 \le i < j \le p} V_{ij} \tag{6}$$

where

$$V_i = \mathrm{Var}[\mathbb{E}(Y|q_i)], \tag{7}$$

$$V_{ij} = \mathrm{Var}[\mathbb{E}(Y|q_i, q_j)] - V_i - V_j. \tag{8}$$

Dividing both sides of Equation 6 by $\mathrm{Var}(Y)$ results in the sensitivity decomposition,

$$1 = \sum_{i=1}^{p} S_i + \sum_{1 \le i < j \le p} S_{ij}. \tag{9}$$

Here, S_i are the first-order sensitivity indices and, from Equations 6, 7, and 9,

$$S_i = \frac{V_i}{\mathrm{Var}(Y)} = \frac{\mathrm{Var}[\mathbb{E}(Y|q_i)]}{\mathrm{Var}(Y)} \tag{10}$$

These indices measure the main effect of the parameter Q_i on the output variance, $\mathrm{Var}(Y)$. For a purely additive model, $\sum_{i=1}^{p} S_i = 1$. However, it is also important to consider the nonadditive features of the model. Equation 9 is a finite series consisting of $p + \frac{p(p-1)}{2}$ terms. For high-dimensional problems, the calculation of all first- and second- order indices becomes impractical, thus motivating the total-effect indices. The total-effect indices for inputs Q_i are given by the sum of all the sensitivity terms in Equation 9,

$$S_{T_i} = S_i + \sum_{\substack{j=1 \\ j \neq i}}^{p} S_{ij}. \tag{11}$$

By decomposing the variance conditional on $q_{\sim i}$ instead of q_i, where the subscript $\sim i$ implies all variables except those of index i, the total-effect indices can be formally expressed as

$$S_{T_i} = 1 - \frac{\mathrm{Var}[\mathbb{E}(Y|q_{\sim i})]}{\mathrm{Var}(Y)} = \frac{\mathbb{E}[\mathrm{Var}(Y|q_{\sim i})]}{\mathrm{Var}(Y)}. \quad (12)$$

This term is a measure of the output variance attributed to Q_i including all of the variance caused by any interactions of any order with $Q_{\sim i}$. To deem Q_i non-influential, it is necessary and sufficient that the total-effect indices be equal to zero, or

$$\mathbb{I}(Q_i) = 0 \iff S_{T_i} = 0. \quad (13)$$

Note that this inherently implies that $S_i = 0$, since $S_{T_i} \geq S_i$. It follows that a parameter can be fixed for $S_{T_i} \simeq 0$, albeit with model error due to the approximation (Saltelli, A., Ratto, M., Andrtes, T., Campolongo, F., Cariboni, J., Gatelli, D., Michaela, S. & Tarantola, S., 2008; Smith, 2013).

The evaluation of each sensitivity index through a double-loop, brute-force Monte Carlo approach with M samples for each loop requires M^2 model evaluations. Since the choice of M should be a function of the dimension, p, for a model with more than a few parameters, this method becomes computationally prohibitive. However, Saltelli (Saltelli, A., Ratto, M., Andrtes, T., Campolongo, F., Cariboni, J., Gatelli, D., Michaela, S. & Tarantola, S., 2008) derived a formulation to obtain these indices with only $M(p + 2)$ model evaluations. The Saltelli algorithm was implemented in the current work to compute both the first order and total-effect indices. It should be noted that this method relies on the assumption that all of the parameters are independent (i.e., uncorrelated).

2.3. Fatigue Crack Growth Model

The most expensive aspect of high-fidelity, FE-based fracture modeling is the solution of the finite element equations. The remeshing at each growth step is relatively fast. The mesh-independent tracking of the crack geometry in three-dimensional space is even faster, which is the basis for the model developed in (Leser et al., 2016) and used in the present work. The fatigue crack growth model is essentially a three dimensional fracture mechanics algorithm wrapped around a set of externally trained surrogate models which replace the finite element analysis (FEA) component of high-fidelity fracture modeling. The code tracks a given crack front in space and evolves the crack based on driving forces obtained from the surrogate models. The code was developed using the Python computing language (Van Rossum & Drake, 2011) as a general tool to reduce computation times associated with high-fidelity fracture simulation and is not limited to the work presented herein.

A diagram illustrating the geometrical and symbolic basis of the model is shown in Figure 1. The user defines the geometry of the part or component in which the crack resides, the initial crack front, and any required model parameters (e.g., Paris' Law parameters). A component geometry is established as a set of points, $G \subset \mathbb{R}^3$, a crack is initialized within this component, and the crack is grown step-by-step by either a finite number of cycles per step or a median crack front extension. For any time, t, the crack front, $\Gamma(t)$, exists in Cartesian space as a $3 \times \Omega$ array, where Ω is the total number of explicitly defined crack front points, γ_ω. Each crack front point for $\omega = 1, \ldots, \Omega$ is a point in the Cartesian frame. In this sense, $\gamma_\omega \in \Gamma \subset G$ at any time, t. For each growth step, crack driving forces at each point are returned from the surrogate models based on the boundary conditions applied to the component and the current geometrical state of the crack front. These driving forces dictate how the crack evolves. At each growth step, any crack front points that have grown outside of the geometry (i.e., $\gamma_\omega \notin G$) are deleted, and a new crack front Γ is derived using a polynomial or spline fit. The curve is extrapolated to the points at which it intersects with the component or part geometry boundaries, and then Ω evenly spaced points are interpolated between (and including) these end points. The time varying crack path can be tracked by storing the $\Gamma(t)$ instance at each growth step.This process is iterated until a stopping condition is reached.

FEA-based methods traditionally compute the crack driving forces at each growth step by either an explicit representation where the crack surface and component assembly is remeshed at each step, or through the use of enriched elements (e.g., XFEM). In either case, the system is solved and the displacements are used to compute the crack driving forces. As discussed, the finite element solution is typically the most

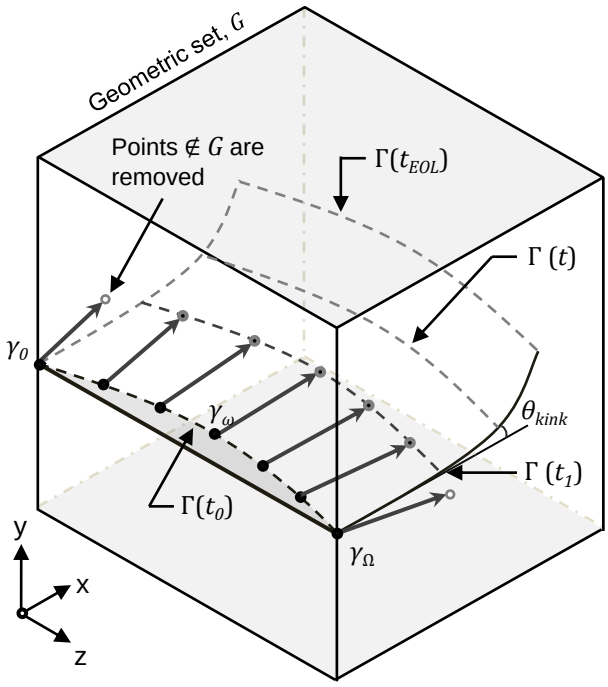

Figure 1. General illustration of the fatigue crack growth model geometrical framework.

computationally intensive part of this process. Therefore, the proposed crack growth model replaces this process with a surrogate modeling approach. Here, a large number, Φ, of crack growth simulations are run using a FE-based, high-fidelity fracture mechanics code *a priori*, or without knowledge of the true crack path. Each of these simulations are then broken down by growth step, meaning that each explicitly modeled crack front and its corresponding driving force profile are treated as a single data point to be used to train the surrogate models via supervised machine learning. If each simulation provides η unique couples of crack front and corresponding driving force profiles, the collection of all crack fronts from all training simulations make up a training dataset of size $\sum_{s=1}^{\Phi} \eta_s$. For example, 30 crack growth simulations where $\eta_s = 100$ for $s = 1, \ldots, 30$ yields $3,000$ data points that can be used for training.

Training the surrogate models in this sense has three primary advantages. First, machine learning requires a sampling or grid scheme that provides adequate coverage of the multidimensional parameter space. For example, the starting coordinates of the crack growth training simulations should be distributed over the space of potential or expected crack initiation sites. While it is impossible to simulate growth along all of the infinite paths that exist for a given geometry, the goal is to achieve an adequate distribution of cracks with which to train the surrogate model, which is problem-dependent. Levels I and III of the DDSim concept (Emery et al., 2009) are useful here, as they can be used to focus the chosen initial conditions (e.g., starting location and starting orientation) used for the training simulations. Secondly, by basing the training data on expected crack paths, all of the crack geometries included in the training set are reliably admissible and cracks that are known to be infeasible based on the mechanics of the problem are excluded. This not only avoids wasting computation time on unnecessary simulations but also can lead to better surrogate performance as it limits extraneous training data. Finally, all of the time-consuming simulations can be conducted in advance and in parallel utilizing high-performance computing, removing them from the serial sampling procedures used for Bayesian inference (e.g., MCMC).

Apart from reducing the computation times of high-fidelity fracture simulations by orders of magnitude (Leser et al., 2016), the hybrid fracture mechanics and machine learning approach also enables a separation of model parameters that dictate crack path (e.g., initiation location and boundary conditions) and those that dictate crack growth rate (e.g., Paris' Law parameters). As a result, only the former need to be considered in the machine learning procedure, which reduces upfront computational costs and adds a significant degree of flexibility. For example, crack growth can be executed using a variety of crack growth rate models without the need to retrain. Additionally, the crack path history is easily accessible at user-defined crack growth increments, which is ideal when solving the Bayesian inverse problem.

2.4. Prognostic Metrics

It is critical when dealing with prognostics to be able to evaluate the performance of the proposed approach. To this end, prognostic metrics developed by (Saxena, Celaya, Saha, Saha, & Goebel, 2010) were used herein to evaluate the performance of the proposed approach. The hierarchical metrics

allow for evaluation of the performance over time, which is desirable because as more information is obtained, better predictions can be made. It should be noted that, since the current work involves fatigue crack growth, all times are measured in cycles. In the case of damage prognosis, RUL is typically of interest and is defined as

$$\mathrm{RUL} = t_{\mathrm{EOL}} - t_D, \tag{14}$$

where t_{EOL} is the end of life (EOL) or time of failure, and t_D is the time at which the last diagnostic measurement was taken. Saxena et al. refer to this as a moving horizon, where RUL is not only a function of the EOL, but also of the current time. Since they are independent, the balance between the rate at which predictions improve and the time rate of change is critical. If the predictions do not improve quickly enough, they may never prove useful. The problem is exacerbated by the fact that the prognosis procedure cannot be conducted in real time (i.e., there is a finite amount of time after the last measurement before a prediction is made). In this paper, this finite time will be assumed negligible, since the speed of the prognosis was not a primary focus. Mathematically,

$$t_{\mathrm{pred}_\rho} = t_{D_\rho} \text{ for } \rho = 1, \ldots, P \tag{15}$$

where P is the total number of predictions made during the component or structure's lifespan, t_{pred_ρ} is the time at which the ρ^{th} prediction was made, and, as before, t_{D_ρ} is the time that the diagnosis data for t_{pred_ρ} were gathered. Henceforth, only t_D will be used for simplicity.

The prognostic metrics applied in this work are as follows:

1. RUL vs. time plot: the basis for all of the prognostic metrics, it is a plot of the RUL predictions and uncertainty over the life of the monitored component or structure.
2. Prognostic horizon, PH: a measure of time at which the prediction reaches a desired level of accuracy with respect to the EOL.
3. α-λ performance: a measure of how accurate the prediction is with respect to the RUL at a given time. Note that α and λ are parameters which are used to define the metric and are defined in the subsequent paragraphs.

The RUL vs. time plot is based around the true RUL which is plotted as a straight line, about which RUL predictions and error bounds are drawn for qualitative assessment of the prognostic algorithm's performance as $t \to t_{\mathrm{EOL}}$. The RUL predictions are represented by box plots. Here, the predicted mean and median are represented by a small square symbol and a line dividing the box, respectively. The upper and lower quartiles of the data are represented by the extent of the box, and the whiskers, represented as capped dotted lines, denote the range of the data.

The prognostic horizon and the α-λ performance metrics are more quantitative than the plot alone and depend on three terms: α, β and λ. The first is a percent error where $\alpha \in [0, 1]$ and is used to define a set of upper and lower error bounds, α^+ and α^-, respectively. Since the two metrics considered here deal with accuracy about two separate quantities, the EOL and RUL, the definition of these bounds differ for each. The same value of α can and should be used for both metrics. The β parameter is used to define a portion of probability and is used as a threshold. This value should be set equal to the percentage of the probability in the predicted RUL PDF that the user desires to lie within the α-bounds, where $\beta \in [0, 1]$. In this way, β is dependent on the purpose or mission of the component or structure being monitored. Finally, $\lambda \in [0, 1]$ and is a time window modifier that simply normalizes the region of the time axis between the time the first prediction is made, t_{D_1}, and t_{EOL}, such that

$$t_{\lambda_\rho} = t_{D_1} + \lambda(t_{\mathrm{EOL}} - t_{D_1}). \tag{16}$$

The α-bounds for the prognostic horizon metric are defined as

$$\alpha^{\pm} = RUL \pm \alpha \cdot t_{\mathrm{EOL}}, \tag{17}$$

where RUL is the true value of RUL calculated using the true EOL, t_{EOL}, and Equation 14. The actual metric is defined as

$$PH = t_{\text{EOL}} - t_D^*, \tag{18}$$

where t_D^* is the first time, t_{D_ρ}, for which the percentage of probability of the predicted PDF (obtained via integration or a probability mass approximation) within the α-bounds is greater than or equal to β. Hence, the PH metric is a measure of time indicating how long it took to reach an acceptable accuracy in the EOL forecast.

Contrary to the prognostic horizon, the α-λ metric quantifies the accuracy of the prediction of the time-varying RUL, and, therefore, the α-bounded region for this metric decreases as $t \to t_{\text{EOL}}$. These bounds are defined by

$$\alpha^{\pm} = RUL \cdot (1 \pm \alpha). \tag{19}$$

The metric takes the form of the Boolean expression

$$\alpha{-}\lambda \text{ accuracy} = \begin{cases} 1 & \text{if } \pi(RUL_{\text{pred}})_{\alpha^-}^{\alpha^+} \geq \beta \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $\pi(RUL_{\text{pred}})_{\alpha^-}^{\alpha^+}$ is the predicted RUL probability that lies within the α-bounds. This metric is computed at each t_{λ_ρ}. In the present work, these time values always correspond to t_{D_ρ} for $\rho = 1, \ldots, P$. The only difference between these time axes is simply that one is normalized over the region in which predictions are made and the other is based on the true time.

3. METHODS

3.1. Experimental Setup

To achieve controllable non-planar crack growth, an experiment was developed based on the work in (Ingraffea, Grigoriu, & Swenson, 1991). Two holes were drilled in an edge-notched specimen as shown in Figure 2. When loaded in tension, mixed-mode driving forces were induced by the presence of the holes, causing any cracks growing through their region of influence to kink and grow toward the closest hole.

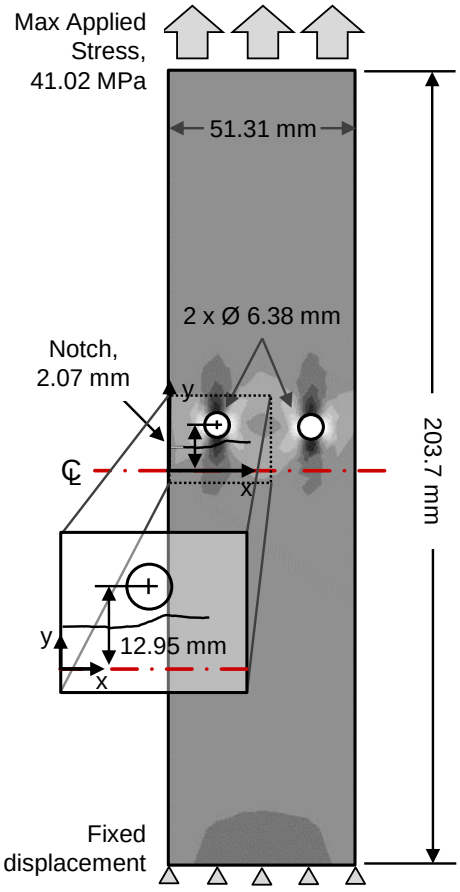

Figure 2. Description of the two-hole specimen used in the experiment. For qualitative purposes, the Von-Mises stress field is overlain on the schematic, highlighting the primary regions of influence of the holes. Darker shades correspond to higher stresses.

Depending on the y-coordinate of the notch, herein referred to as y_0, various degrees of kinking could be achieved. At higher y_0 values, the crack would grow into the hole. At low y_0 values, the crack would grow with minimal kinking. In between these two extremes the crack would slope up toward the hole, peak, and then partially slope back down to a relatively horizontal growth condition before crossing an instability threshold and ultimately failing.

A random notch location between the aforementioned extremes was chosen for the current experiment at $y_0 = 6.73$ mm, and the edge notch was cut at that location to a length of 2.16 mm using electrical discharge machining (EDM). The

dimensions of the specimen were measured as-manufactured using a ruler and calipers. The crack was grown from the notch under a constant amplitude fatigue stress of 41.0 MPa with a load ratio, $R = 0.1$, and a frequency of 10 Hz. Crack tip location was measured at fixed cycle intervals using a traveling optical microscope. The observed crack path from the experiment is plotted in Figure 2. The crack grew into the influence of the hole, exhibited non-planar growth, and then grew to failure. The crack shown is the final configuration before failure was deemed to have occurred. Note that, while through-the-thickness measurements were not available, the crack growth was modeled in all three dimensions to demonstrate the capabilities of the model.

The goal of the present work was to utilize damage diagnosis data to accurately predict, with quantified uncertainty, the RUL of a specimen containing an evolving non-planar crack. Predictions were to be made at multiple times during the specimen lifespan to demonstrate the effect of data on the prognosis. For the purpose of this experiment, the damage diagnosis data took the form of crack tip locations in the x-y plane over time. Since the intention of the presented prognosis framework was to utilize noisy data from in-situ SHM or automated NDE diagnosis systems, the high-precision visual measurements were augmented with Gaussian white noise, $\epsilon \sim N(0, \sigma^2)$, in an attempt to replicate results typical of these systems. The variance of the noise distribution was loosely based on the Rose criterion for image processing in which a signal-to-noise-ratio (SNR) ≥ 5 is required to reliably distinguish image features (Rose, 2013). The SNR was defined here as the reciprocal of the coefficient of variation (Parzen, 1961) for both the x and y one dimensional data arrays, such that

$$\text{SNR}_x = \frac{\mu_x}{\sigma_x}, \ \text{SNR}_y = \frac{\mu_y}{\sigma_y}. \tag{21}$$

Here, μ and σ are the mean and standard deviation of the measurements, respectively.

Assuming that an in-situ SHM system would likely be less accurate than an NDE scan, a mounted piezoelectric sensor array was chosen as inspiration for the noise model. A simple linear array would likely have to be mounted somewhere toward the bottom of the specimen and oriented in the x-direction. Therefore, it was also assumed that $\text{SNR}_x > \text{SNR}_y$ since the incident waves would be approximately perpendicular to the crack faces growing in the x-direction. Based on these assumptions, $\text{SNR}_x = 5$ and $\text{SNR}_y = 2.5$ for the experiment. The resulting dataset was thinned by 20% and divided into 5 intervals; i.e., five predictions of RUL would be made, each after a new interval of data was appended to the total set. The vector of times at which these data were gathered are $t_D = [350, 500, 650, 800, 950]^{\mathrm{T}} \times 10^3$ cycles[1]. Figure 3 shows the data divided into intervals overlain on the visually observed crack, near the left hole. Future work should implement an actual NDE or SHM system to gather the data to avoid the above assumptions.

3.2. Surrogate Training

The surrogate models responsible for returning crack driving forces were trained using the high-fidelity FE-based fracture software FRANC3D[2] in conjunction with the FE software Abaqus.[3] Training simulations were generated by varying the following two parameters: (i) the initial starting position, y_0, which is the only parameter directly affecting crack path with respect to the hole, and (ii) the initial crack length a_0, which is required for initiation from a straight notch of an unknown length.

The training data were developed in three steps. Originally, the number of complete simulations $\Phi = 30$ as shown in Figure 4. All of the simulations were fixed at a single a_0 and varied y_0 only. Next, 330 additional, single-step simulations with 11 different values for a_0 per each of the original y_0 values were added to the dataset under the assumption that these small cracks would result in paths nearly identical to the original 30 simulations. Finally, the dataset was augmented by manually varying the crack front shape at each growth step, in all simulations. Five different crack front shapes were used ranging from perfectly straight (i.e., a midpoint extension of

[1] Note that this time vector and all subsequent times herein will be reported as a count of complete fatigue cycles. True time can be obtained by dividing the number of cycles by the frequency used for the experiment, 10 Hz

[2] See reference ("FRANC3D Reference Manual, Version 6", 2011)

[3] See reference ("Abaqus/CAE User's Manual, Version 6.12", 2012)

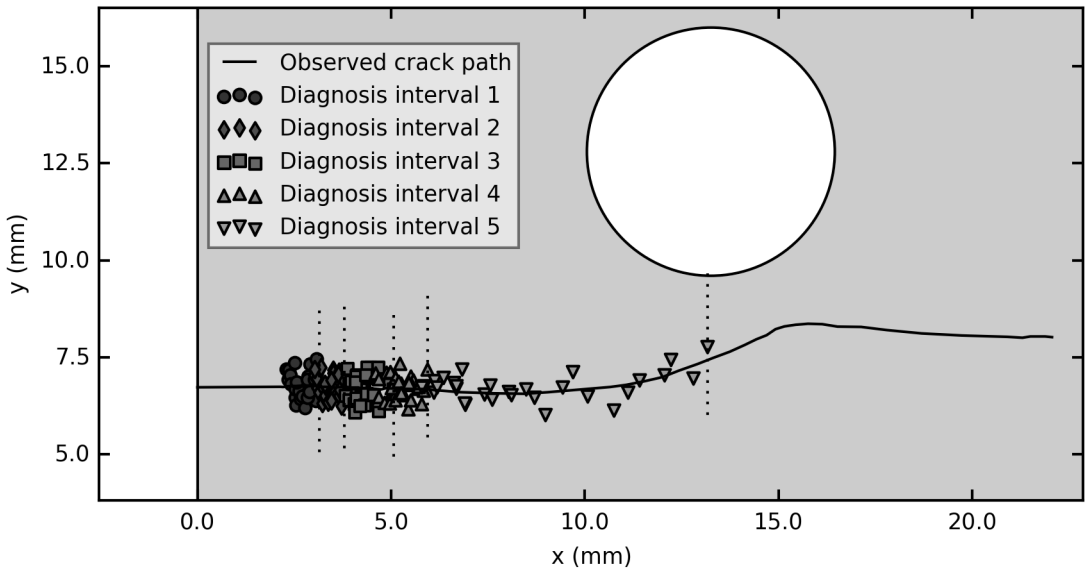

Figure 3. Damage diagnosis obtained via visual measurements of the crack tip location with added noise and ($\text{SNR}_x = 5$ and $\text{SNR}_y = 2.5$). The dotted lines represent the boundaries of the diagnosis intervals and represent the five points in time at which predictions were made; $t_D = [350, 500, 650, 800, 950]^{\text{T}} \times 10^3$ cycles, respectively.

zero for a parabolic curve) to an exaggerated curve (i.e., a midpoint extension of 0.127 mm). These augmented shapes allowed the surrogates to reach equilibrium during growth in spite of small numerical errors in the crack propagation. In total, the training simulations resulted in $16,229$ training points, each consisting of one unique crack front and its corresponding SIF profiles (i.e., $K_I, K_{II},$ and K_{III} at each front point, γ_ω). The crack growth algorithm and surrogate model were verified using FRANC3D simulations that were not part of the original training set, and validation was carried out using experimental crack growth data. The details of the training process, the verification, and the validation were reported in (Leser et al., 2016) and will be omitted here for brevity.

3.3. Detailing the Crack Growth Model

The conceptual framework of the crack growth model was presented earlier in Section 2.3. Due to its flexibility, a variety of fracture mechanics models can be plugged into this framework. The specific equations used to dictate the fracture growth in the present work are outlined in this section. For more details, the reader is referred to the work in (Leser et al., 2016), which used the same crack growth model.

To incorporate the effects of the load ratio, R, on the crack

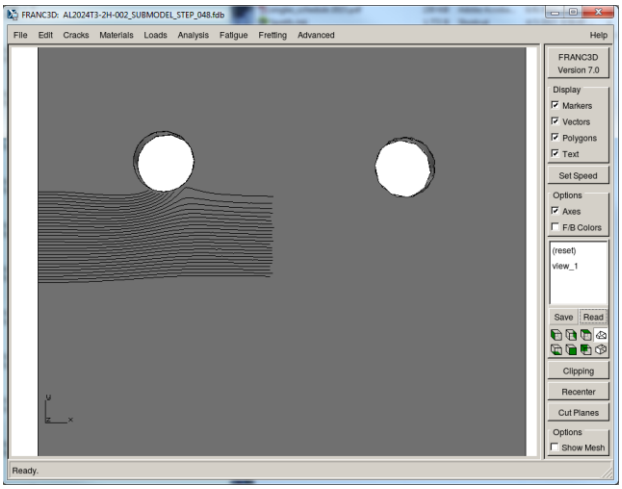

Figure 4. Depiction of the original 30 crack growth simulations comprising the base of the training dataset. Screenshot taken from the fracture analysis software FRANC3D.[1]

growth rate, $\frac{da}{dN}$, Walker's modification of the Paris' Law (Walker, 1970),

$$\frac{da}{dN} = C\left[\frac{\Delta K}{(1-R)^{1-m}}\right]^n, \tag{22}$$

was chosen as the crack growth rate model. Here, C, m and n are empirical constants and

$$\Delta K = K_{I,max} - K_{I,min}. \tag{23}$$

where K_I is the mode I SIF. Note that $R = \frac{K_{I,min}}{K_{I,max}}$. The mode II SIF, K_{II} plays a significant role in the kinking behavior of the crack. In the present work, the maximum tangential stress criterion (Erdogan & Sih, 1963) is used to dictate how the crack turns under mixed-mode SIFs,

$$\theta_{kink} = \arg\max_{\theta}\Big(K_I^r(\theta)\Big), \tag{24}$$

where, ignoring high-order stress terms,

$$\begin{aligned} K_I^r(\theta) &= \sigma_{\theta\theta}\sqrt{2\pi r} \\ &= \cos\frac{\theta}{2}\Big[K_I\cos^2\frac{\theta}{2} - \frac{3}{2}K_{II}\sin(\theta)\Big]. \end{aligned} \tag{25}$$

is the resolved mode I SIF and $\sigma_{\theta\theta}$ is the tangential stress. Combining Equations 23 and 25, an effective equivalent ΔK can be defined as

$$\Delta K_{ee} = K_{I,max}^r - K_{I,min}^r. \tag{26}$$

which can then be applied to Equation 22.

In the three-dimensional implementation of the crack growth model, these equations are applied to each crack front point, γ_ω, to determine the crack growth rate, $\frac{da}{dN}_\omega$. The crack front is then advanced using a median extension approach, where the median extension is defined as

$$\Delta a_\omega = \Delta a_{median}\left[\frac{\frac{da}{dN}_\omega}{\frac{da}{dN}_{median}}\right], \tag{27}$$

and Δa_{median} is a user-defined value, and $\frac{da}{dN}_{median}$ is the median of $\frac{da}{dN}_\omega$ for $\omega = 1, \ldots, \Omega$. Using the median extension approach requires that the number of cycles to grow the front at each point be computed through integration of the crack growth rate equation over the Δa_ω,

$$\Delta N_\omega = \int_0^{\Delta a_\omega} C^{-1}\left[\frac{\Delta K_{ee}}{(1-R)^{1-m}}\right]^{-n} d(\Delta a_\omega). \tag{28}$$

This means that the number of cycles returned at each front point will be different, albeit often with a standard deviation, $\sigma_{N_\omega} < 1$ cycle. The final cycle count used for the crack growth step, $\Delta N^* = \sum_{\omega=1}^{\Omega} \Delta N_\omega$. It should also be noted that the SIF term is a continuous function of the crack length, so an assumption must be made as to its functional form. This relationship was assumed to be linear in the current work.

Finally, failure was dictated by a critical mode I SIF criterion. If, at any crack front point,

$$K_{I,max} > K_{IC} \tag{29}$$

where K_{IC} is the critical mode I SIF and is a material property, then the failure is deemed to have occurred. Uncertainty in K_{IC} were not considered. Instead, the experimentally observed crack geometry was inserted into a FE model of the specimen using FRANC3D, and the value of $K_{I,max}$ at failure was calculated and used to set $K_{IC} = 833.96\ \mathrm{MPa}\sqrt{\mathrm{mm}}$.

4. RESULTS

4.1. Sensitivity Analysis

Global sensitivity analysis was conducted using Saltelli's algorithm (Saltelli, A., Ratto, M., Andrtes, T., Campolongo, F., Cariboni, J., Gatelli, D., Michaela, S. & Tarantola, S., 2008). The parameters considered in the sensitivity analysis were the initial y-coordinate of the crack, y_0, the initial crack length, a_0, and the Walker model parameters, C, n, R, and m. The parameter space was defined as shown in Table 1. The initial distributions were chosen to encompass the extremes of

the knowledge of the specimen and previous research of the crack growth rates in aluminum alloy 2024-T3. While the RUL is the primary output of interest for prognosis, the cycle values at EOL over the parameter space defined in Table 1 varied over multiple orders of magnitude, causing difficulties when calculating the sensitivity indices. Instead, an approach was adopted in which the variance of crack tip coordinates in the x and y directions due to the varying parameters were used to calculate the first-order and total-effect Sobol´ indices over time.

The sensitivity analysis results for the x-coordinate are shown in Figure 5. The figure shows Sobol´ indices plotted over a normalized time for simplicity (i.e., cycles divided by true end of life). Examining the first-order indices, S_i, the Paris coefficient, C, and the Paris exponent, n, are the most influential parameters, which is to be expected as these are the two primary drivers of the crack growth rate. The effect of the load ratio, R, was also evident, and its Sobol´ indices overtook n as the number of cycles approaches the end of life. The remaining parameters, a_0 and y_0, are approximately zero over time. While not clear in the figures, both the first-order and total-effect indices for a_0 exhibit peaks at time zero, which is intuitive since it was the only parameter affecting the value of x for that instant. This peak was ignored since these values quickly decline for times greater than zero. From Equation 13, the total-effect indices must equal zero for a parameter to be considered non-influential, which is the case for both a_0 and y_0. All of the crack growth rate parameters showed an appreciable total-effect index and, therefore, were considered influential with respect to the crack growth in the x-direction.

Sensitivity indices for the crack tip y-coordinate over time are shown in Figure 6. As expected, only y_0 has an appreciable first-order effect, with the remainder of model parameters at or near zero. Note that some of the values were negative, which is an artifact of the approximations associated with small or moderate sample sizes used with Saltelli's algorithm (Smith, 2013). The total-effect Sobol´ indices show a marked increase over the first-order effects for the crack growth model parameters, implying interactions among these parameters. This result may be somewhat artificial or un-

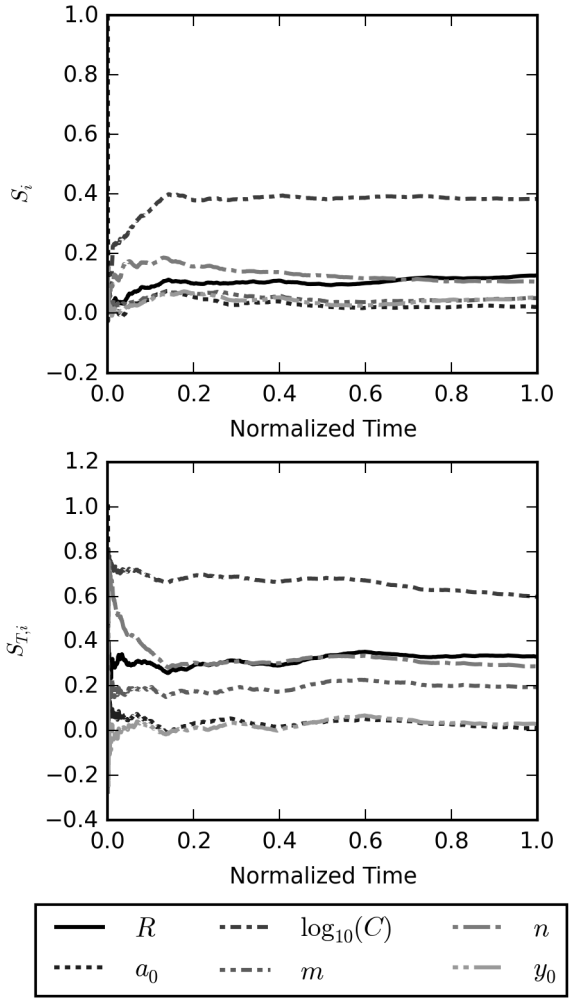

Figure 5. First-order and total-effect Sobol´ indices for crack tip x location over 1×10^6 cycles (normalized)

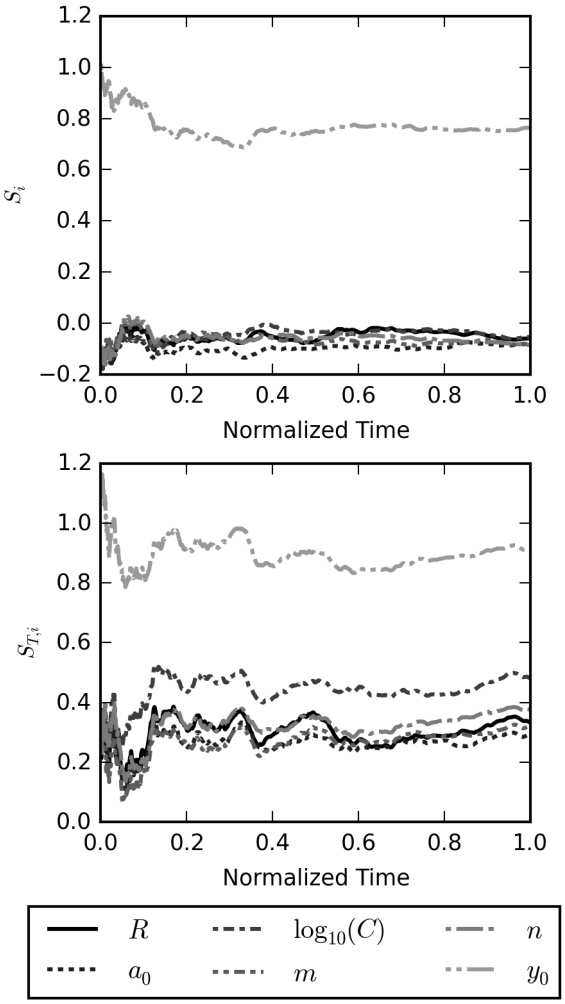

Figure 6. First-order and total-effect Sobol´ indices for crack tip y location over 1×10^6 cycles (normalized)

trustworthy, however. For example, it was not expected that a_0 or any of the growth rate parameters would not have an effect on the y-coordinate of the crack tip since it was assumed that the crack path was independent of these factors. If the lower grouping of $S_{T,i}$ values were considered zero as intuition might suggest (i.e., an artificial bias was introduced), the only two non-zero parameters would be y_0 and C. The influence of C could possibly be explained by small kink angle errors or large crack growth rates near the point at which the crack approaches the hole and demonstrates significant nonlinear behavior.

Based on all of the information obtained from the global sensitivity analysis, only a_0 could be fixed with reasonable confidence that it was non-influential. The remainder of the parameters were considered in the Bayesian inference problem. Regardless, valuable information about the effect and interaction of parameters in the context of the crack growth model was gained. Furthermore, a global sensitivity analysis, besides the obvious utility, also tests the robustness of the model as it requires evaluations over the entire parameter space.

Table 1. Parameter distributions for global sensitivity analysis

Parameter	Distribution	Description
a_0	$U(0.05, 0.1)$	Initial crack length
$\log_{10}(C)$	$U(-10, -7)$	Log of the Paris' Law coefficient
m	$U(0.0, 1.0)$	Load ratio exponent
n	$U(2.0, 5.0)$	Paris' Law exponent
R	$U(0.0, 0.99)$	Load ratio
y_0	$U(0.18, 0.32)$	Crack initiation site

4.2. Prognosis

4.2.1. Parameter estimation & uncertainty quantification

A solution to the Bayesian inverse problem of Equation 2 was obtained using the experimentally observed crack tip measurements in conjunction with Markov chain Monte Carlo (MCMC) sampling. Parameter estimation using MCMC was conducted five times in total using the data in Figure 3, once for each interval as marked. Specifically, the uncertainty in parameters was estimated at 350, 500, 650, 800, and 950 thousand fatigue cycles into the experiment, respectively. The data at each interval consisted of all of the measurements up to the respective cycle; i.e., the data used at a particular cycle included all of the measurements from the beginning of the test up to that point in time. Non-informative, or uniform, prior distributions were used for all of the parameters at each interval as shown in Table 2. This implies that each MCMC run was considered independent from the others, and no prior information was passed between the subsequent runs as might be the case for a true Bayesian updating scheme. The adaptive Metropolis algorithm included in the PyMC python package (Patil, Huard, & Fonnesbeck, 2010) was used to generate 2×10^5 samples, with a conservative burn-in of 1×10^5 samples to ensure Markov chain stabilization. Geweke's time-series approach was utilized to diagnose chain convergence (Geweke, 1992).

Table 2. Parameter distributions for Markov chain Monte Carlo sampling

Parameter	Distribution	Description
$\log_{10}(C)$	$U(-20,-1)$	$\log_{10}$ of the Paris' Law coefficient
m	$U(0.5,0.7)$	Load ratio exponent
n	$U(1.0,20)$	Paris' Law exponent
R	$U(0.09,0.11)$	Load ratio
y_0	$U(0.01,0.365)$	Crack initiation y-coordinate

Parameter estimation results in the form of marginal probability distribution functions (PDF) for all parameters and for all five data intervals are shown in Figure 7. The bounds used during parameter estimation for R and m were significantly tighter than those used during the sensitivity analysis. The bounds on these parameters were based on a conservative interpretation of the expected accuracy of the load frame used in the experiments. These parameters were unidentifiable given the current set of data, and thus the MCMC procedure returned uniform distributions (i.e., returned the prior distribution). It is possible that this was due to the tight bounds enforced through prior knowledge. Examining the remaining parameters in the figure, the uncertainty decreases as more data are added. However, it is also important to notice that the distributions exhibit a significant shift at time t_{D_5} after the last interval, a topic that will be discussed more in the subsequent sections.

Figure 8 shows the pairwise plots of the sampled parameters after the fourth data interval was obtained. The samples indicate a strong correlation between the Paris' Law parameters, C and n. This correlation is commonly reported in fracture mechanics literature (Cortie, 1991; Carpinteri & Paggi, 2007). This correlation indicates that these two parameters are not mutually identifiable. The correlation also violates the assumptions of Saltelli's global sensitivity analysis algorithm, which was used herein and assumes independent parameters. Future work should investigate other methods such as those presented in (Xu & Gertner, 2008) or (Li & Mahadevan, 2016) which do not require this assumption. The shift in parameter distributions can be seen again by comparing Figure 8 with the pairwise plots from the fifth update shown in Figure 9.

4.2.2. Uncertainty propagation and RUL prediction

Once the uncertainty in the model parameters was quantified, the prognosis could be formed by propagating these uncertainties through the crack growth model and extrapolating forward in time to the failure condition given by Equation 29. This process was conducted using a Monte Carlo sampling approach in which $25,000$ samples were drawn from the parameter distributions and passed to the model, which then simulated crack propagation up to the EOL condition. The final cycle counts for all $25,000$ simulations were aggregated to form an EOL distribution. Using Equation 14, the EOL distribution can easily be converted to an RUL distribution. This procedure was repeated five times, once for each interval, and all simulations were conducted in parallel utilizing high performance computing.

The resulting RUL distributions are presented in the form of an RUL vs. time plot in Figure 10. The shaded strip and cone regions represent the $\alpha = 0.1$ accuracy intervals associated with the PH and α-λ prognostic metrics, respectively. As discussed previously, the constant-width strip region can be interpreted as an accuracy measure of how well the EOL was predicted, while the cone indicates how well the RUL was predicted. The cone shape is an artifact of the true RUL being a monotonically decreasing function of time. Similar to the parameter estimation results, the first four RUL predictions exhibit convergence with a steady decrease in associated uncertainty. However, it can now be observed that these predictions actually diverge from the true solution. Again, after the last diagnosis interval at t_{D_5}, the PDF shifts.

While it was difficult to determine whether the shift shown in the parameter PDFs resulted in better accuracy, the shift in the RUL distribution can be quantified since the ground truth value is known. As shown, the final prediction enters both α bounded regions. Figure 11 shows the predicted PDF of the RUL at this point (i.e., obtained after the fifth data interval) for clarity. The double dotted lines represent the α-bounds associated with the α-λ metric. The difference between the mean prediction and the true RUL is $4,625$ cycles, or a percent error of 5.9%. While this result may seem satisfactory,

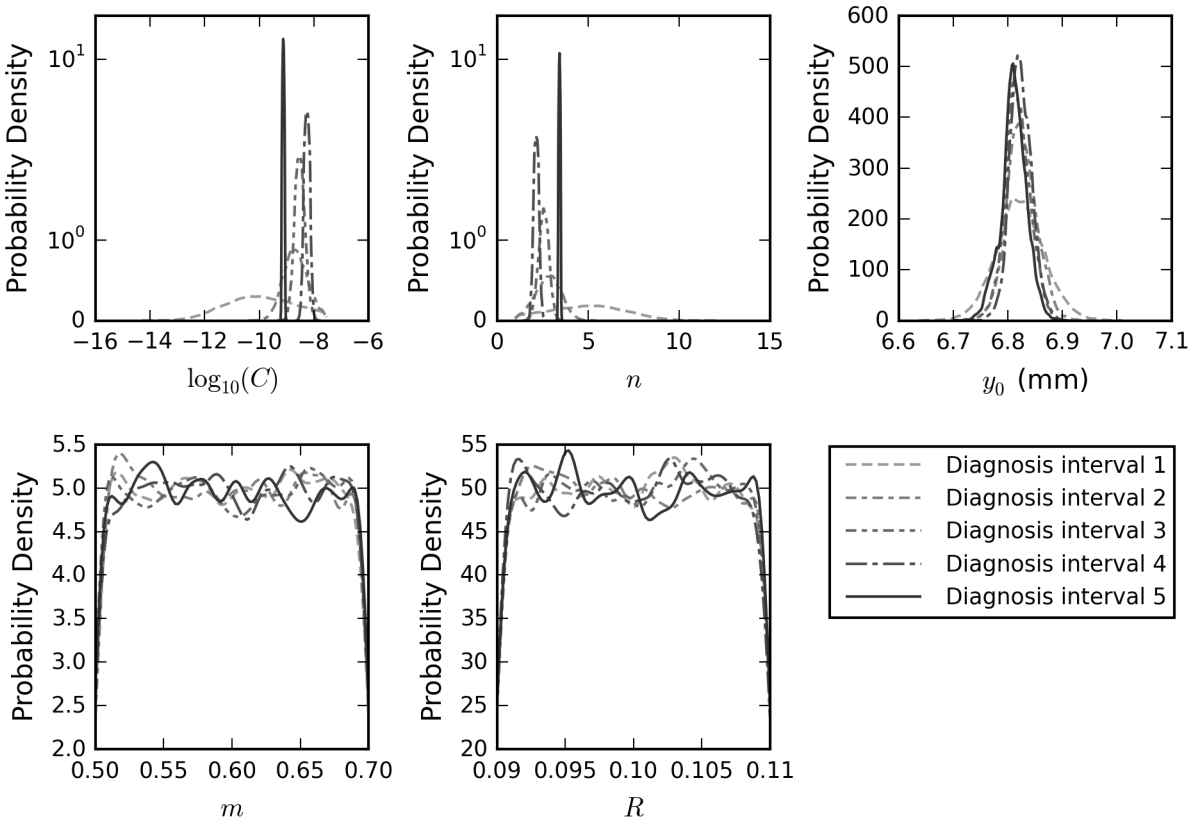

Figure 7. Resulting PDFs from the Bayesian parameter estimation process using MCMC, updated over the 5 diagnosis intervals. These intervals correspond to the values in $t_D = [350, 500, 650, 800, 950]^{\mathrm{T}} \times 10^3$ cycles, respectively, which represent the times at which the last data point in the interval was collected.

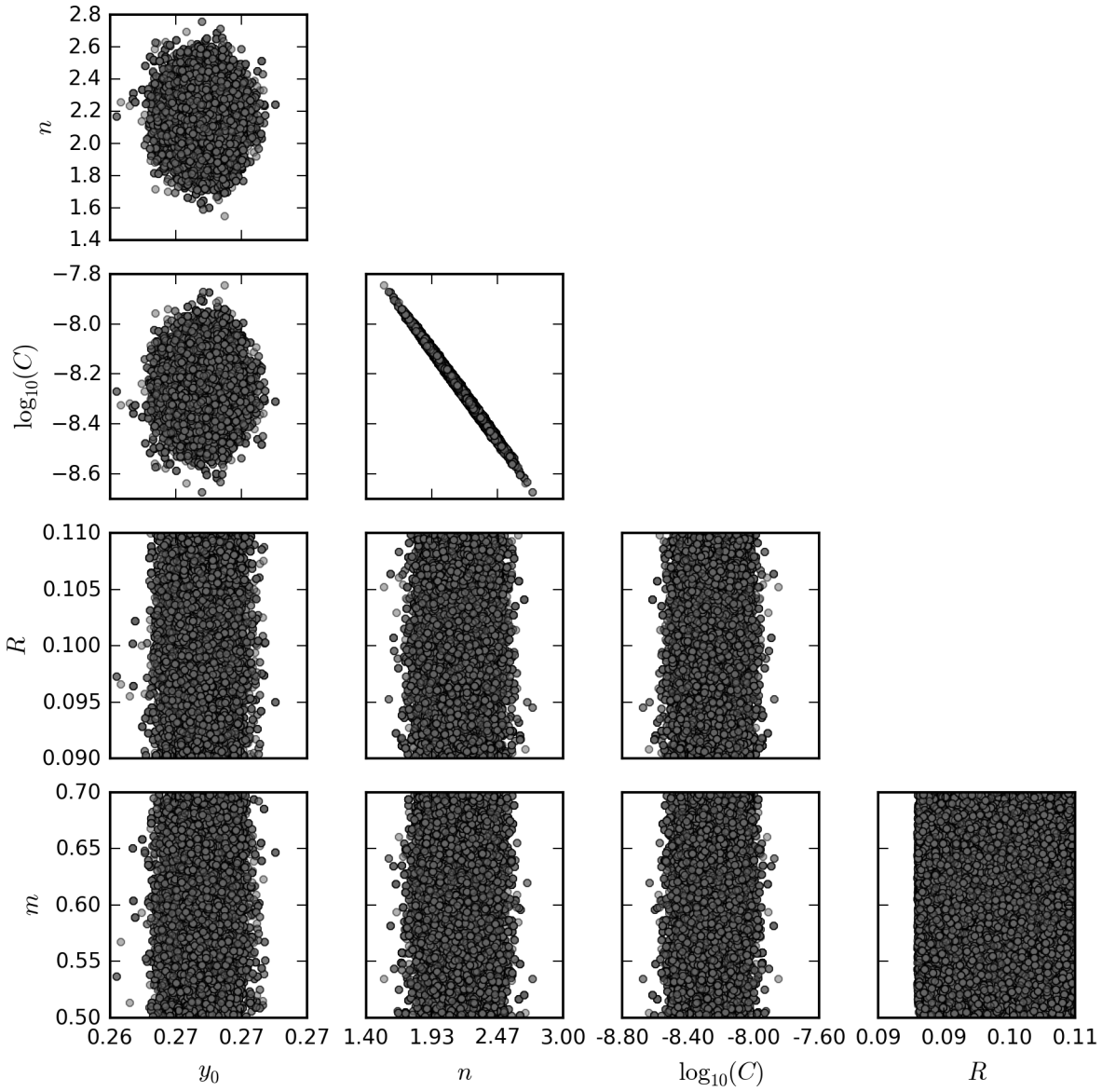

Figure 8. Joint samples obtained from the parameter estimation for the 4^{th} diagnosis interval, or $t_{D_4} = 800 \times 10^3$ cycles.

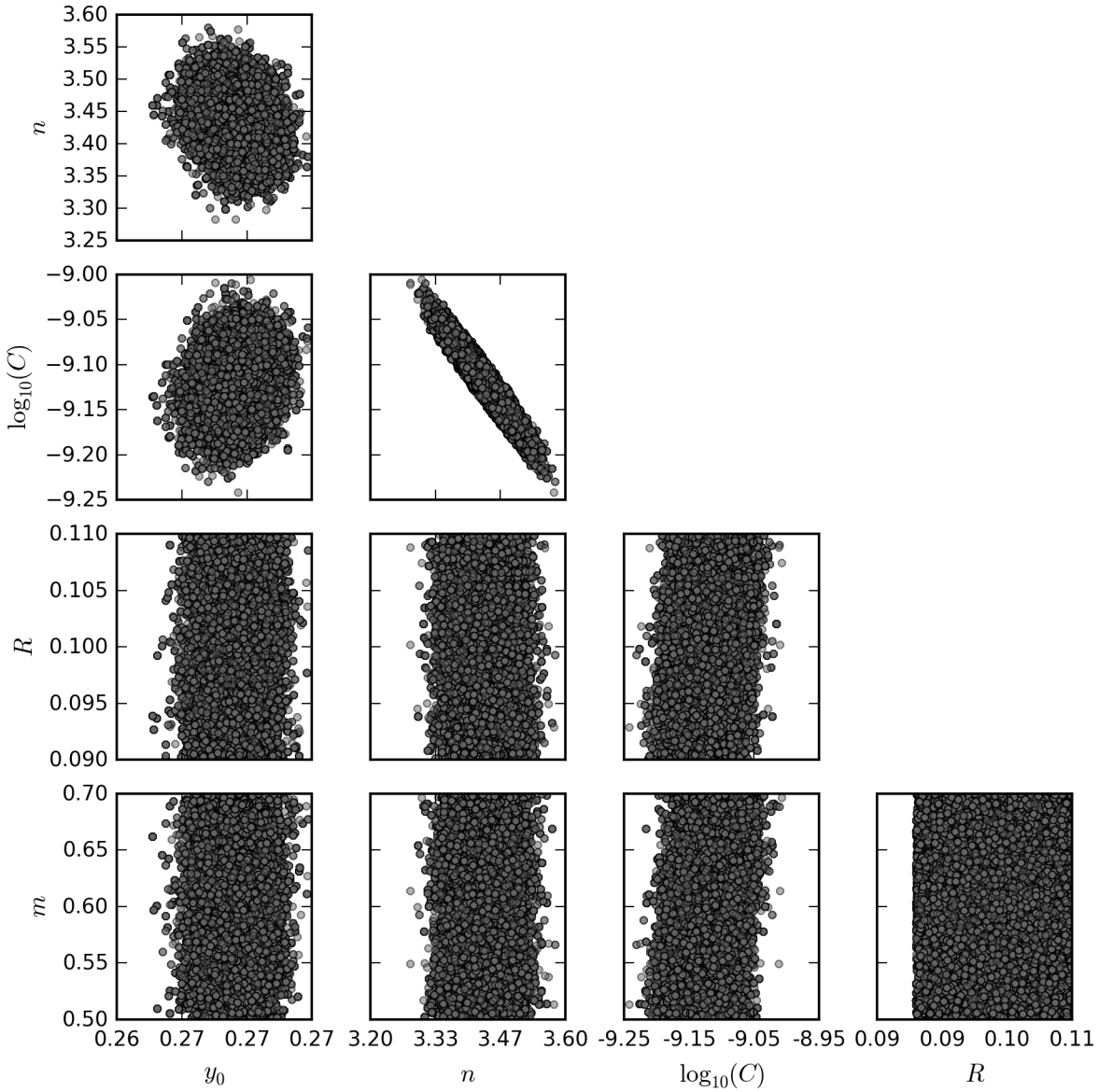

Figure 9. Joint samples obtained from the parameter estimation for the 5^{th} diagnosis interval, or $t_{D_5} = 950 \times 10^3$ cycles.

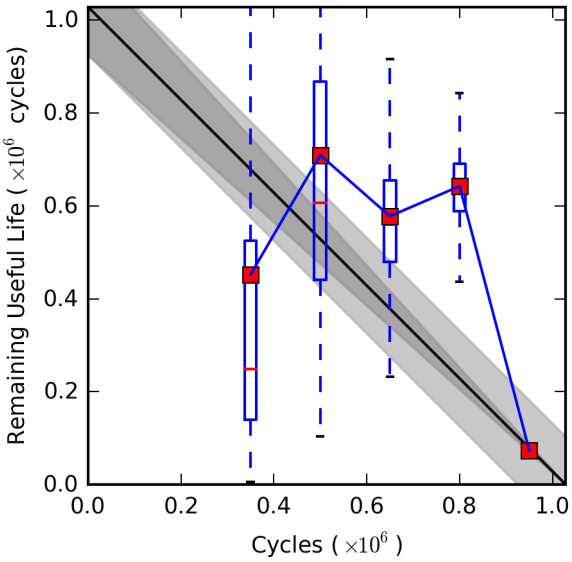

Figure 10. RUL vs. time plot for the five updates of the diagnosis dataset. The solid black line represents the experimentally observed RUL over time. The shaded strip (constant width) and the cone-shaped regions represent the $\alpha = 0.1$ accuracy bounds corresponding to the PH and α-λ metrics, respectively. The probabilistic prediction of RUL at each time, t_{D_i}, is represented by a box plot as described in Section 2.4.

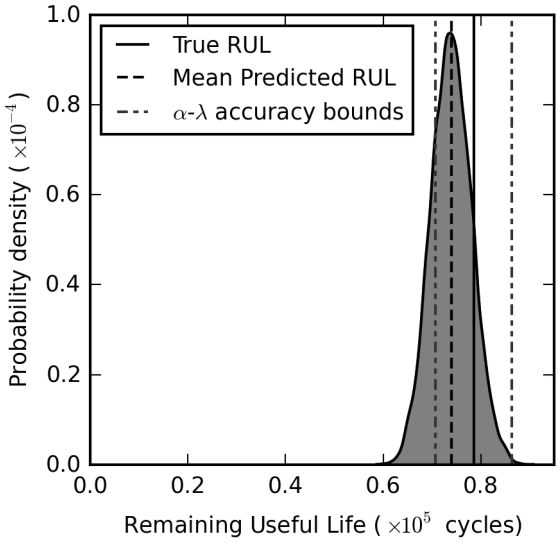

Figure 11. PDF of the final RUL prediction (i.e., obtained using all five intervals of damage diagnosis data).

Table 3. α-λ test results with $\alpha = 0.1$ and $\beta = 0.25$.

Cycle	λ	**Passes α-λ test**
350, 000	0.00	False
500, 000	0.22	False
650, 000	0.44	False
800, 000	0.66	False
950, 000	0.88	True

the ability to accurately predict the RUL is only part of a successful damage prognosis scheme.

An accurate prediction must be made early enough for the proper action to be taken, which is the motivation of the prognosis metrics presented in (Saxena et al., 2010) discussed earlier. Setting $\alpha = 0.1$ and $\beta = 0.25$, the prognostic horizon, $PH = 78,537$ cycles. This represents the earliest time that at least 25% of the probability in the forecasted RUL distribution lies within the α-bounds. Here, PH corresponds to the prediction made after the fifth data interval was gathered (see Figure 10). Even if we set $\beta = 0.1$, the prognostic horizon, $PH = 528,537$. This corresponds to the second data interval. However, as shown in Figure 10, the following data update results in a RUL distribution that has near zero probability within the α-bounds.

Proceeding further through the prognostic metric hierarchy, the α-λ metrics for the five data intervals are shown in Table 3. This metric clearly captures the poor performance of the metric for the first four data intervals, and quantifies the percentage of time between the first prediction and the EOL. This metric utilizes the cone-shaped bounds, which, in general, provide a better prognostic evaluation since the RUL is the true quantity of interest. Only 22% of the time available for predictions remains when the predicted RUL distribution satisfies the α-λ metric. The α-λ test can also be visualized as demonstrated in Figure 12. As opposed to presenting a Boolean pass or fail, this figure shows the probability of the RUL PDF that lies within the cone-shaped α-bounds, regardless of the value of β. The double x-axes show the time in cycles and as a percentage of time in the available prediction window. The divergence from the true RUL is clearly shown here between the second and fourth intervals along with the sudden shift in accuracy after the fifth interval. The improvement in accuracy results in 78% of the predicted probability lying within the α-bounds, but with only 12% of the λ-window remaining.

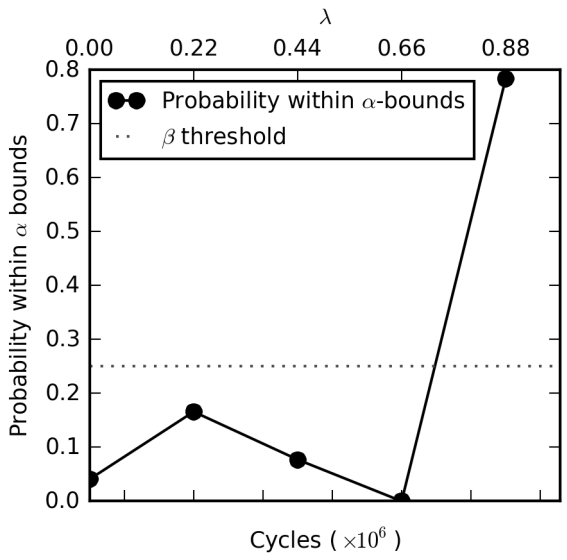

Figure 12. Probability within the α-λ metric α-bounds (i.e., the cone-shaped bounds). The λ axis represents the percentage of time between the first prediction and the EOL.

5. DISCUSSION

The divergence of the predictions would be concerning if the presented specimen was instead a real-world structure or safety-critical component operating in the field. The idea that the forecasts could shift from the true value while decreasing in uncertainty is dangerous as it can result in either false-positives or false-negatives, both of which could have drastic consequences. The first suspect in the search for a reason behind this divergence is the model. However, the model clearly shows that it is capable of capturing the true RUL given adequate data. The model cannot be entirely ruled out without further testing, as there may be a form of model discrepancy present in the region of growth occurring around the fifth data update. Model discrepancy refers to cases where the model is incapable of capturing certain physically observed behavior. In a simple form, it is analogous to a polynomial curve fit without enough degrees of freedom. Future work should investigate the existence and effect of potential model discrepancies not only of the proposed surrogate modeling approach, but also of the training simulator FRANC3D. As shown in (Leser et al., 2016), the proposed model matches almost exactly with the FRANC3D simulations for the validation cases.

The parameters chosen for the analysis represent another way in which the model could be causing the poor predictions. The choice to remove the parameter a_0 based on the sensitivity analysis could have unforeseen consequences. The results shown in Figure 6 show signs of numerical errors in the global sensitivity analysis implementation. Perhaps this parameter was more influential than shown. Furthermore, upon examination of the differences between the joint samples in Figures 8 and 9, there is a change observed in the parameter relationships. First, the sharp, almost linear correlation between C and n that is so commonly reported in fracture mechanics literature is much less pronounced. Additionally, slight correlations now can be observed between the Paris' Law parameters and y_0, as well as between C and both Walker parameters. This indicates that the interactions within the parameter space enforced by the prior distributions may not completely understood. It is possible that the Walker parameters require a less restrictive prior, and that the uncertainty in the loading was greater than previously thought. Future work should include a more expansive study of these correlations that appear after the fifth update.

The model may be a part of the issues highlighted herein, but, based on the sudden shift in parameter distributions and the final PDF of the RUL, a stronger argument could be made against the data instead. The crack growth data in the fifth interval are the first of which that capture the primary non-planar features of the crack behavior. It is not a coincidence that this coincides with the shift in the prediction, and it is likely that the shift would be more gradual if given one data point at a time from this interval. This result is intuitive. However, a more important concern is whether or not the data in the second through fourth intervals are somehow negatively impacting the results. This is the region where a divergence from the true RUL is observed, and thus this is a reasonable assumption. Even if it is accepted that forecasts will only be usable after data have been gathered in the non-planar region, the question remains whether or not the data before that point are valuable or detrimental. It is possible that removing the middle three intervals and only using the first and fifth might reduce the uncertainty in the final predic-

tion. After all, if accurate predictions can only be made after $800,000$ cycles into the fatigue life, it would still be desirable to improve the accuracy within the time that remains. High levels of accuracy in this region are especially important considering how close it is to the EOL.

All of the above issues are driven by the fact that the crack growth is non-planar, which supports the idea that real-world prognosis must be of concern to researchers in the field of damage prognostics. The number of issues in all phases of a prognostic framework appear to increase with complexity in the crack growth behavior. Data become harder to obtain and, therefore, more uncertain, and the modeling is more complicated and increases the likelihood of model discrepancies or user error. Both of these issues then complicate the Bayesian inference and uncertainty propagation, culminating in difficulties reaching the final goal of an accurate forecast of future damage growth.

6. CONCLUSION

A flexible, high-fidelity yet rapid probabilistic framework for fatigue damage prognosis was demonstrated for a metallic specimen exhibiting non-planar crack growth. Damage diagnosis data in the form of visual crack tip measurements with added noise were used to quantify the uncertainty in crack growth model parameters through Bayesian inference and Markov chain Monte Carlo sampling techniques. The computational burden associated with high-fidelity crack growth models was neutralized through the use of a previously developed modeling approach in which expensive finite element-based computation of crack driving forces was replaced with an efficient surrogate model. Driving forces returned by the surrogate model were fed to a fully three-dimensional fracture mechanics algorithm for crack propagation. The flexibility of the crack growth model was demonstrated through a global sensitivity analysis followed by the parameter estimation procedure. Following their quantification, parameter uncertainties were propagated forward in time to generate a probabilistic prediction of the specimen's remaining useful life.

Results showed an initial divergence from the true solution followed by a sharp increase in accuracy near the specimen end of life. The shift in accuracy corresponded to the acquisition of observations of the primary non-planar features of the crack, suggesting model discrepancy and faults in the data (e.g., lack of data or misleading data) as possible reasons for the early divergence in forecasts. Future work should investigate the role of both factors, as a system whose uncertainty decreases but diverges from the truth is potentially dangerous. Upon inclusion of the data from the primary non-planar region of crack growth, the new predicted mean converged to the true remaining useful life value with a percent error of only 5.9% and with 78% of the predicted probability lying within 10% of the true value.

The discrepancy between the final prediction and those before it is an important result from this work, as it demonstrates not only the importance of rigorous prognostic verification and validation, but it also offers a glimpse into some of the real-world challenges that need to be addressed in the damage prognosis field. The issues highlighted herein can be directly traced to the non-planar nature of the crack, suggesting that models must be developed to accurately capture this behavior. Continued research in the area of rapid, high-fidelity prognostic methods is paramount if health management tools such as NASA's Digital Twin are to reach their full potential.

REFERENCES

Abaqus/CAE User's Manual, Version 6.12 [Computer software manual]. (2012).

Banerjee, S., S, P., & Chijioke, A. (2014). Quantification of memory effect in composites under fatigue for precursor damage analysis. In H. Kim, D. Whisler, Z. Chen, C. Bisagni, M. Kawai, & R. Krueger (Eds.), *Proceedings of the American Society for Composites 29th Technical Conference* (p. 382-391). Lancaster, PA: DEStech Publications, Inc.

Carpinteri, A., & Paggi, M. (2007). Self-similarity and crack growth instability in the correlation between the paris constants. *Engineering Fracture Mechanics*, *74*(7), 1041–1053.

Chang, F.-K., Markmiller, J. F., Ihn, J.-B., & Cheng, K. Y. (2007). A potential link from damage diagnostics to health prognostics of composites through built-in sensors. *Journal of Vibration and Acoustics*, *129*(6), 718–729.

Chiachıo, J., Chiachıo, M., Saxena, A., Rus, G., & Goebel, K. (2013). An energy-based prognostics framework to predict fatigue damage evolution in composites. In S. Sankararaman & I. Roychoudhury (Eds.), *Proceedings of the Annual Conference of the Prognostics and Health Management Society* (pp. 363–371). New York: Prognostics and Health Management Society.

Cortie, M. (1991). The irrepressible relationship between the paris law parameters. *Engineering Fracture Mechanics*, *40*(3), 681–682.

Emery, J., Hochhalter, J., Wawrzynek, P., Heber, G., & Ingraffea, A. (2009). DDSim: A hierarchical, probabilistic, multiscale damage and durability simulation system–part I: methodology and level I. *Engineering Fracture Mechanics*, *76*(10), 1500–1530.

Erdogan, F., & Sih, G. (1963). On the crack extension in plates under plane loading and transverse shear. *Journal of Fluids Engineering*, *85*(4), 519–525.

Farrar, C. R., & Lieven, N. A. (2007). Damage prognosis: the future of structural health monitoring. *Philosophical Transactions of the Royal Society of London, Series A*, *365*(1851), 623–632.

Farrar, C. R., & Worden, K. (2007). An introduction to structural health monitoring. *Philosophical Transactions of the Royal Society of London, Series A*, *365*(1851), 303–315.

Farrar, C. R., & Worden, K. (2012). *Structural health monitoring: a machine learning perspective*. John Wiley & Sons.

FRANC3D Reference Manual, Version 6 [Computer software manual]. (2011).

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo & J. Berger (Eds.), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting* (pp. 169–193). Oxford, UK: Oxford University Press.

Gobbato, M., Conte, J. P., Kosmatka, J. B., & Farrar, C. R. (2012). A reliability-based framework for fatigue damage prognosis of composite aircraft structures. *Probabilistic Engineering Mechanics*, *29*, 176 - 188.

Gope, P. (1999). Determination of sample size for estimation of fatigue life by using Weibull or log-normal distribution. *International Journal of Fatigue*, *21*(8), 745–752.

Hombal, V., Ling, Y., Wolfe, K., & Mahadevan, S. (2012). Two-stage planar approximation of non-planar crack growth. *Engineering Fracture Mechanics*, *96*, 147–164.

Hombal, V., & Mahadevan, S. (2013). Surrogate modeling of 3D crack growth. *International Journal of Fatigue*, *47*, 90–99.

Ingraffea, A., Grigoriu, M., & Swenson, D. (1991). Representation and probability issues in the simulation of multi-site damage. In S. Atluri, S. Sampath, & P. Tong (Eds.), *Structural integrity of aging airplanes* (pp. 183–197). Berlin: Springer-Verlag.

Johnston, G. (1983). Statistical scatter in fracture toughness and fatigue crack growth rate data. In J. Bloom & J. Ekvall (Eds.), *Probabilistic fracture mechanics and fatigue methods: Applications for structural design and maintenance* (pp. 42–66). Philadelphia: American Society for Testing and Materials.

Laloy, E., & Vrugt, J. A. (2012). High-dimensional posterior exploration of hydrologic models using multiple-try DREAM (ZS) and high-performance computing. *Water Resources Research*, *48*(1), 239–249.

Leser, P. E., Hochhalter, J. D., Warner, J. E., Newman, J. A., Leser, W. P., Wawrzynek, P. A., & Yuan, F.-G. (2016). IWSHM 2015: Probabilistic fatigue damage prognosis using surrogate models trained via three-dimensional finite element analysis. *Structural Health Monitoring (currently online only)*. doi: 10.1177/1475921716643298.

Li, C., & Mahadevan, S. (2016). An efficient modularized sample-based method to estimate the first-order Sobol´ index. *Reliability Engineering & System Safety*, *153*, 110–121.

Ling, Y., & Mahadevan, S. (2012). Integration of structural health monitoring and fatigue damage prognosis. *Mechanical Systems and Signal Processing*, *28*, 89–104.

Liu, Y., & Mahadevan, S. (2009). Probabilistic fatigue life prediction using an equivalent initial flaw size distribution. *International Journal of Fatigue*, *31*(3), 476–487.

Ma, X., & Zabaras, N. (2010). An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *Journal of Computational Physics*, *229*(10), 3884–3915.

Neiswanger, W., Wang, C., & Xing, E. (2014). Asymptotically exact, embarrassingly parallel MCMC. In N. Zhang & J. Tian (Eds.), *Proceedings of the 30th Conference on Uncertainty in Artificial Intelligence* (pp. 623–632). Oregon: AUAI Press Corvallis.

Parzen, E. (1961). Mathematical considerations in the estimation of spectra. *Technometrics*, *3*(2), 167-190.

Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. *Journal of Statistical Software*, *35*(4), 1–81.

Peng, T., He, J., Xiang, Y., Liu, Y., Saxena, A., Celaya, J., & Goebel, K. (2015). Probabilistic fatigue damage prognosis of lap joint using Bayesian updating. *Journal of Intelligent Material Systems and Structures*, *26*(8), 965–979.

Peng, T., Liu, Y., Saxena, A., & Goebel, K. (2015). In-situ fatigue life prognosis for composite laminates based on stiffness degradation. *Composite Structures*, *132*, 155–165.

Rabitz, H., & Aliş, Ö. F. (1999). General foundations of high-dimensional model representations. *Journal of Mathe-*

matical Chemistry, *25*(2-3), 197–233.

Rose, A. (2013). *Vision: human and electronic*. Springer Science & Business Media.

Saltelli, A., Ratto, M., Andrtes, T., Campolongo, F., Cariboni, J., Gatelli, D., Michaela, S. & Tarantola, S. (2008). *Global sensitivity analysis: the primer*. Chichester, West Sussex, UK: John Wiley & Sons.

Sankararaman, S., Ling, Y., & Mahadevan, S. (2011). Uncertainty quantification and model validation of fatigue crack growth prediction. *Engineering Fracture Mechanics*, *78*(7), 1487–1504.

Sankararaman, S., Ling, Y., Shantz, C., & Mahadevan, S. (2011). Uncertainty quantification in fatigue crack growth prognosis. *International Journal of Prognostics and Health Management*, *2*(1), 1–15.

Saxena, A., Celaya, J., Saha, B., Saha, S., & Goebel, K. (2010). Metrics for offline evaluation of prognostic performance. *International Journal of Prognostics and Health Management*, *1*(1), 4–23.

Smith, R. C. (2013). *Uncertainty quantification: Theory, implementation, and applications*. Philadelphia: SIAM.

Ter Braak, C. J. (2006). A Markov chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing*, *16*(3), 239–249.

Van Rossum, G., & Drake, F. J. (2011). *An introduction to Python*. Godalming, Surrey, UK: Network Theory Ltd.

Vasudevan, A., Sadananda, K., & Glinka, G. (2001). Critical parameters for fatigue damage. *International Journal of Fatigue*, *23*, 39–53.

Vrugt, J. A., Ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, *10*(3), 273–290.

Walker, K. (1970). The effect of stress ratio during crack propagation and fatigue for 2024-T3 and 7075-T6 aluminum. In M. Rosenfield (Ed.), *Effects of environment and complex load history on fatigue life* (pp. 1–14). Philadelphia: American Society for Testing and Materials.

Xu, C., & Gertner, G. Z. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, *93*(10), 1563–1573.