A Novel Computational Methodology for Uncertainty Quantification in Prognostics Using The Most Probable Point Concept

Shankar Sankararaman¹ and Kai Goebel²

¹ SGT Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA shankar.sankararaman@nasa.gov

² NASA Ames Research Center, Moffett Field, CA 94035, USA kai.goebel@nasa.gov

ABSTRACT

This paper develops a novel computational approach to quantify the uncertainty in prognostics in the context of conditionbased monitoring. Prognostics consists of two major steps; first, it is necessary to estimate the state of health at any time instant, and then, it is required to predict the remaining useful life of the engineering component/system of interest. While the topic of estimation has been addressed through different types of Bayesian tracking techniques, this paper primarily focuses on the second aspect of future prediction and remaining useful life computation, which is influenced by several sources of uncertainty. Therefore, it is important to identify these sources of uncertainty, and quantify their combined effect on the remaining useful life prediction. The computation of uncertainty in remaining useful life can be treated as an uncertainty propagation problem which can be solved using probabilistic techniques. This paper investigates the use of the Most Probable Point approach (which was originally developed to estimate the failure probability of structural systems) for calculating the probability distribution of the remaining useful life prediction. The proposed methodology is illustrated using a battery which is used to power an unmanned aerial vehicle.

1. INTRODUCTION

Research in the past few years has been advocating the use of an onboard health management system in engineering systems used for time-critical, safety-critical, and cost-critical missions. An accurate health management system constantly monitors the performance of the engineering system, performs diagnosis (fault detection, isolation, and estimation), prognosis (predict possible failures in the future and estimate remaining useful life) and aid online decision-making (fault mitigation, fault recovery, mission replanning, etc.). The prediction of remaining useful life is an important aspect of prognostics, and is directly useful in different types of decisionmaking. This paper focuses on the calculation of remaining useful life in the context of model-based prognostics and condition-based monitoring.

In practical applications, there are several sources of uncertainty which affect the performance of both the engineering system and the health management system. For example, the loading conditions and operating conditions of the engineering system may be random in nature. The sensors, which are part of the health management system, may not be accurate due to measurement errors, and this may prevent accurate estimation of the system state. The system models which are used by the health management system for estimation and prediction may have certain errors. As a result of the presence of such uncertianties, it is important to rigorously account for the sources of uncertainty during diagnosis, prognosis, and decision-making. While the topic of uncertainty quantification in diagnosis has gained attention in literature (Sankararaman & Mahadevan, 2011b, 2013), the importance of uncertainty significantly increases in the context of prognosis, since the focus is on predicting future behavior, which is far more challenging and uncertain than fault diagnosis. The primary objective of this paper is to develop a computational methodology which can quantify the combined effect of the various sources of uncertainty on prognostics, and estimate the overall uncertainty in the remaining useful life (RUL) prediction.

In the past, several researchers have used different types of methods to quantify uncertainty in prognostics. Tang et al. (Tang, Kacprzynski, Goebel, & Vachtsevanos, 2009) discuss the use of Bayesian tracking algorithms for uncertainty quantification and management in prognostics for In-

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

tegrated Vehicle Health Management (IVHM) systems. The "Damage Prognosis Project" at Los Alamos National Laboratory (C. R. Farrar, Lieven, & Bement, 2005; Inman, Farrar, Junior, & Junior, 2005) exclusively dealt with prognosis and uncertainty quantification applied to structual composites. Several researchers worked on this project and published articles that deal with model development, verification, validation, prediction, and uncertainty quantification; the conclusions of this project have been documented by Farrar et al (C. Farrar & Lieven, 2007). Sankararaman et al. (Sankararaman, Ling, Shantz, & Mahadevan, 2009, 2011) quantified the uncertainty in fatigue crack growth prognosis in metals, by using finite element models (for structural analysis), crack growth models (to predict future crack growth), and Monte Carlo simulation (for uncertainty quantification). Gu et al. (Gu, Barker, & Pecht, 2007) also used Monte Carlo simulation in order to compute the uncertainty in damage in electronics subjected to random vibration. In some practical applications, Monte Carlo simulation using exhaustive sampling may be computationally expensive, and this challenge has inspired the development of intelligent sampling-based algorithms (DeCastro, 2009; Orchard, Kacprzynski, Goebel, Saha, & Vachtsevanos, 2008; Daigle, Saxena, & Goebel, 2012) and mathematical techniques such as relevance vector machines (Saha & Goebel, 2008) and principle component analysis (Usynin & Hines, 2007), that can reasonably approximate the uncertainty in the prognostic predictions. Further, Bayesian and maximum relative entropy methods (Guan et al., 2011) have also been used for estimating uncertainty in prognostics.

The above described methods for uncertainty quantification are mainly based on sampling techniques, and may require several thousands of samples to accurately calculate the uncertainty in RUL prediction. This may be computationally expensive for online monitoring, and therefore, Sankararaman et al. (Sankararaman, Daigle, Saxena, & Goebel, 2013) discussed analytical approaches for predicting the uncertainty quantification. These analytical methods are based on firstorder Taylor's series expansion, and in particular, one method known as the Inverse First-order Reliability Method was implemented for calculating the uncertianty in RUL prediction. This method was originally developed by structural engineers for calculating the failure probability of structural engineering applications. This method is based on the concept of Most Probable Point (MPP) Estimation, and this paper further investigates the applicability of this approach to different types of loading conditions and RUL calculation. Note that the term "reliability method" is avoided in the rest of the present paper in order to avoid confusion with "reliability testing" methods for prognostics, since the proposed methodology is primarily applicable to condition-based online monitoring.

The rest of the paper is organized as follows. Section 2 discusses a framework for uncertainty quantification in prognostics, and explains that the problem of estimating the uncertainty in the RUL prediction can be viewed as an uncertainty propagation problem. Section 3 discusses the importance of future loading conditions, and describes three different types of future loading scenarios for prognostics. Section 4 explains the proposed computational methodology for quantifying prognostics uncertainty and estimating the uncertainty in the remaining useful life prediction; this methodology is illustrated using a numerical example in Section 5. Finally, conclusions are drawn in Section 6.

2. UNCERTAINTY QUANTIFICATION IN CONDITION-BASED MONITORING AND PROGNOSTICS

This section discusses the need for uncertainty quantification in prognostics and health monitoring. First, a model-based framework for prognostics is presented, and then, the various sources of uncertainty are discussed with reference to this framework. Finally, it is illustrated that uncertainty quantification in prognostics can be viewed as an uncertainty propagation problem.

2.1. Model-based Framework for Prognostics

The goal of prognostics is to predict the future behavior of a component/system at any generic instant of prediction, denoted by t_P . This is accomplished by estimating the states of the systems at all time instants $t > t_P$. The inputs $(\mathbf{u}(t))$ and outputs $(\mathbf{y}(t))$ to the system are known until the prediction time $t = t_P$, and in order to perform prediction, the future inputs i.e., $\mathbf{u}(t) \forall t > t_P$ also need to be available. A generalized prognostics architecture is showed in Fig. 1.

The first step in prognostics is to estimate the state at time t_P . Using the estimated state, the second step is to predict future states until failure; thereby, the end-of-life (EOL) and the remaining useful life (RUL) can be predicted.

State-space models are used for both estimation and prediction. Consider a generic state space model which is used to continuously predict the state of the system, as:

$$\dot{\mathbf{x}}(t) = \mathbf{f}(t, \mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t), \mathbf{v}(t))$$
(1)

where $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ is the state vector, $\boldsymbol{\theta}(t) \in \mathbb{R}^{n_{\theta}}$ is the parameter vector, $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ is the input vector, $\mathbf{v}(t) \in \mathbb{R}^{n_v}$ is the process noise vector, and \mathbf{f} is the state equation. This state equation can be constructed using physics-based principles, or using data-driven techniques.

While Eq. 1 is used for state prediction, actual sensor measurements (which are available until time $t = t_P$) are used for state estimation. The sensor measurements are modeled using a generic output equation, such as:

$$\mathbf{y}(t) = \mathbf{h}(t, \mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t), \mathbf{n}(t))$$
(2)

where $\mathbf{y}(t) \in \mathbb{R}^{n_y}$, $\mathbf{n}(t) \in \mathbb{R}^{n_n}$, and \mathbf{h} denote the output vec-



Figure 1. Model-Based Prognostics Architecture

tor, measurement noise vector, and output equation respectively. Note that output measurements are available only until time $t = t_P$. Therefore, the output equation is used only in the estimation stage, and not in the prediction stage.

(Note: While the above framework for state-space representation is general, some practical systems are time-invariant. Therefore, **f** and **h** may not depend on *t*. Sometimes, the output equation **h** may depend only on the states $\mathbf{x}(t)$ and not the inputs $\mathbf{u}(t)$. While the proposed methodology is presented using the general framework, simplifications and assumptions may be included depending on the physical system under consideration.)

2.2. State Estimation

Bayesian tracking approaches as Kalman filtering, particle filtering, etc. can be used for state estimation. These methods use Bayes theorem to update the uncertainty in the states continuously as a function of time, as and when new measurements are available. While particle filtering is the most general method that can account for different distribution types and account for non-linearity, Kalman filtering can be used only when Eq. 1 is linear and all the uncertain quantities are Gaussian. When the uncertain quantities are Gaussian, the extended Kalman filter can be applied by linearizing Eq. 1. Similarly, the unscented Kalman filter (Daigle et al., 2012) can be used to approximate the mean and variance of response quantities (future states and remanining useful life prediction), even for non-linear models.

2.3. Prognostics: Future Prediction

Prognostics consists of predicting future behavior of engineering systems, identifying possible failure modes, and estimating the remaining useful life. Estimating system state is an essential precursor to prognostics, because damage and/or system faults can be represented using state variables or a collection of state variables. Therefore, predicting how the damage will progress or how the fault will grow is equivalent to estimating future states of the systems, based on the information available at the prediction time instant t_P . In order predict future behavior and thereby perform prognosis until end-of-life (denoted by $E(t_P) = t + R(t_P)$, where $R(t_P)$ is the remaining useful life), the following pieces of information are necessary:

- 1. State prediction model, as in Eq. 1.
- 2. Present state estimate $(\mathbf{x}(t_P))$; using the present state estimate and the state space equations in Eq. 1, the future states $(\mathbf{x}(t_P), \mathbf{x}(t_P+1), \mathbf{x}(t_P+2), ..., \mathbf{x}(t_P+R(t_P)))$ can be calculated.
- 3. Future loading $(\mathbf{u}(t_P), \mathbf{u}(t_P+1), \mathbf{u}(t_P+2), ..., \mathbf{u}(t_P+R(t_P)))$; these values are needed to calculate the future state values using the state space equations.
- 4. Parameter values from time t_P until time $t_P + R(t_P)$ (denoted by $\theta(t_P), \theta(t_P + 1), ..., \theta(t_P + R(t_P))$).
- 5. Process noise $(\mathbf{v}(t_P), \mathbf{v}(t_P + 1), \mathbf{v}(t_P + 2), ..., \mathbf{v}(t_P + R(t_P)))$.

While writing " $\theta(t_P)$, $\theta(t_P + 1)$, ..., $\theta(t_P + R(t_P))$ ", note that unit time discretization has been used for the sake of illustration. During implementation, appropriate time-discretization values need to be selected.

During prognosis, all the future states can be predicted as a function of the above quantities, and in a practical engineering system, all of them may be potentially uncertain. First, the state-prediction model is prone to have modeling errors. Second, the state estimate at time t_P is uncertain as a result of the Bayesian tracking method used for estimation. Third, future loading cannot be precisely known in many applications, and therefore, uncertainty regarding future loading needs to be considered. Fourth, model parameters are usually estimated using filtering; but it is challenging to know future parameter values. In this paper, model parameter values are assumed to be constant over time and precisely known (without uncertainty). Fifth, process noise is conventionally assumed to follow a probability distribution (usually, Gaussian with zero mean and a specified variance), and needs to be accounted for in prognostics.

Since prognostics needs to be performed until failure, a boolean function (Sankararaman et al., 2013) of the states is necessary to define end-of-life of the engineering system. Such a boolean function can be defined as:

$$T_{EOL}(\mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t)) = \begin{cases} 1, & c(\mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t)) \leq 0\\ 0, & \text{otherwise,} \end{cases}$$
(3)

where $c(\mathbf{x}(t), \boldsymbol{\theta}(t), \mathbf{u}(t)) \leq 0$ represents failure criterion. There may be multiple failure criteria too (Daigle et al., 2012; Sankararaman et al., 2013) in some cases, and therefore, the definition of T_{EOL} needs to account for all such failure criteria.

In fact, it can be easily shown that T_{EOL} is a function of the above list of quantities, and this functional relationship can be expressed through a graphical flowchart, as shown in Fig. 2.

As seen in Fig. 2, both the End-of-life (EOL, denoted by $E(t_P)$) and the remaining useful life (RUL, denoted by $R(t_P)$) can be calculated as a function of the above list of quantities. Let the function, which predicts the RUL, be denoted as:

$$R = G(\boldsymbol{X}),\tag{4}$$

where X represents the concatenated vector of quantities contained in (1) present state estimates $(\mathbf{x}(t_P))$; (2) future loading values $(\mathbf{u}(t_P), \mathbf{u}(t_P + 1), \mathbf{u}(t_P + 2), ..., \mathbf{u}(t_P + R(t_P)))$; and (3) future process noise values. Again, note that unit time discretization has been used for illustration. Since these quantities are uncertain, the problem of estimating the uncertainty in prognostics, and thereby computing the uncertainty in EOL and RUL can be viewed as propagating the uncertainty in X through G (Sankararaman & Goebel, 2013a).

3. FUTURE LOADING CONDITIONS

In order to perform efficient prognosis, it is necessary to know what loading conditions the system will experience in the future. However, in many practical applications, it is challenging to be able to precisely predict future loading. Therefore, it is essential to estimate the uncertainty in future loading conditions and incorporate this information in prognostics. Time-series analysis techniques and signal processing methods can be used to represent future loading conditions, continuously as a function of time. Further, different types of engineering application may require different types of techniques for loading characterization and uncertainty representation. Therefore, a good prediction methodology should be able to account for different types of representation.

Three different types of future loading conditions - constant amplitude loading, white noise loading, variable amplitude loading - are discussed in this paper. Uncertainty representation for each of the above types of loading conditions are explained in the remainder of this section. Sample loading trajectories are graphically explained through appropriate illustrations.

3.1. Type I: Constant Amplitude Loading

This is the simplest form of loading, where the value of **u** is constant at all time instants. However, the constant value is assumed to be random, and one random variable is sufficient to represent uncertainty in this type of loading con-

dition. Multiple realizations of constant amplitude loading conditions are shown in Fig. 3.



Figure 3. Type-I Loading: Multiple Realizations

3.2. Type II: White Noise Loading

Now, the probability distribution of $\mathbf{u}(t)$ is assumed to be known. For the sake of simplicity, this probability distribution is assumed to be constant at all time instants. Therefore, at any time instant, the loading value is selected by sampling from this probability distribution. Let $f_{\mathbf{U}(t)}(\mathbf{u}(t))$ denote the probability density function (PDF) of this distribution. Loading values at multiple time instants are independently sampled from this distribution. A typical realization of this type of loading condition is shown in Fig. 4.



The number of random variables necessary to represent such a type of loading condition depends on the number of time steps from prediction time t_P until end-of-life $E(t_P)$, which in turn depends on the chosen time-discretization level. Therefore, the number of variables may range from a few tens to several thousands, and this poses a computational challenge for uncertainty propagation. Therefore, a new methodology is proposed in this paper to overcome this challenge.

Note that the value of loading varies from time to time, since it is sampled individually at every time instant. Suppose that the time-variant process is replaced with a time-invariant constant value denoted by \mathbf{u}^{E} . In other words,

$$\mathbf{u}(t) = \mathbf{u}^E \quad \forall \quad t \in [t_P, t_P + R(t_P)] \tag{5}$$



Figure 2. Uncertainty in Prognostics: An Uncertainty Propagation Problem

The above equation means that the same realization of loading will be used for prediction at every future time instant. In order for this to be valid, it is important to choose a suitable probability distribution for \mathbf{u}^E , so that the effect of propagating this distribution through G is equivalent to propagating the original distribution of $\mathbf{u}(t)$ through G. This can be accomplished by computing the likelihood of \mathbf{u}^E such that Eq. 5 is satisfied. In other words, any value of \mathbf{u}^E has an associated probability with which Eq. 5 is satisfied; the likelihood of \mathbf{u}^E is proportional to this probability. The probability distribution of the true loading distribution can be used to calculate this likelihood, as:

$$L(\mathbf{u}^E|R(t_P)) \propto \prod_{t=t_P}^{t=t_P+R(t_P)} f_{\mathbf{U}(t)}(\mathbf{u}(t) = \mathbf{u}^E)$$
(6)

where $f_{\mathbf{U}(t)}(\mathbf{v}(t))$ is the probability density function of the true loading $\mathbf{u}(t)$. Also note that the likelihood function is conditioned on the RUL and written as $R(t_P)$. Further, the above equation assumes that the loading values at two different times are independent of each other. If any statistical dependence is unknown, then it can be easily included in the above equation by conditioning appropriately. Having calculated the likelihood, the PDF of \mathbf{u}^E can be calculated as (Sankararaman & Mahadevan, 2011a):

$$f_{\mathbf{U}^E}(\mathbf{u}^E|R(t_P)) = \frac{L(\mathbf{u}^E|R(t_P))}{\int_D L(\mathbf{u}^E|R(t_P))d\mathbf{u}^E}$$
(7)

In Eq. 7, the domain of integration D is chosen such that $\mathbf{u}^E \in D$ if and only if $L(\mathbf{u}^E) \neq 0$. Now, propagating the uncertainty in \mathbf{u}^E through G is equivalent to propagating the

uncertainty in $\mathbf{u}(t)$ through G. Therefore, \mathbf{u}^E is referred to as equivalent time-invariant loading and its distribution is referred to as the equivalent time-invariant loading distribution.

The use of the equivalent time-invariant loading variable reduces the number of variables to the same number as the constant loading case, and therefore facilitates computation for uncertainty propagation. Now, the time-invariant equivalent variable \mathbf{u}^E is used in \mathbf{X} in Eq. 4, instead of the true loading values. The corresponding probability distribution (which is actually a function of $R(t_P)$) will be used in uncertainty propagation to calculate uncertainty in RUL, as explained later in Section 4.

Note that the above equivalent time-invariant concept can be used to address process noise (Sankararaman & Goebel, 2013b), since the process noise is also treated as white noise (Gaussian, in several models) in many practical applications.

3.3. Type III: Variable Amplitude Loading

In this type loading, multiple time-windows of varying lengths are considered, and within each time-window, the loading is assumed to be a constant. The number of timewindows is assumed to be known; the length of each time window is assumed to be a random variable, and the constant amplitude for each time window is also considered to be a random variable. Therefore, if there are five time-windows, then ten random variables are needed to represent the entire loading trajectory. Each realization of the set of random variables will correspond to one particular loading trajectory. Multiple realizations of a variable amplitude loading scenario are depicted in Fig. 5.



Figure 5. Type-III Loading: Multiple Realizations

Since the number of random variables is small (linear function of number of windows), the proposed uncertainty quantification methodology can be directly applied for this type of loading condition.

4. MOST PROBABLE POINT APPROACH FOR UNCER-TAINTY QUANTIFICATION

This section develops the proposed methodology for quantifying prognostics uncertainty and estimating the uncertainty in the remaining useful life prediction. A generic methodology is presented using the functional relationship $R = G(\mathbf{X})$, explained earlier in Eq. 4 in Section 2. It is also explained as to how the methodology can be adapted for the different types of loading conditions discussed in the previous section.

4.1. Most Probable Point Concept

The Most Probable Point (MPP) concept was originally developed by structural engineers to predict the failure probability of structural engineering applications. In this paper, this concept is used for uncertainty quantification in prognostics and remaining useful life prediction.

Consider the functional relationship R = G(X). The goal of uncertainty propagation is to compute the uncertainty in R, given the uncertainty in X. In order words, the goal is to compute the probability density function (PDF) or cumulative distribution function (CDF) of R, based on the probability distribution of X. Let $f_X(x)$ and $F_X(x)$ denote the PDF and CDF of X respectively. Note that an upper case letter is used to represent the random variable and a corresponding lower case letter to represent a generic realization of that random variable.

If all the variables X are Gaussian (i.e., normal) and if G is linear, then it can be easily proved that R is also Gaussian, and the statistics of R can be calculated analytically. In the context of prognostics, even if the state-space models and the EOL threshold function are linear, their combination renders G non-linear. Therefore, it is necessary to estimate the distribution of R by considering non-linear func-

tions and non-normal variables since the uncertain quantities may not necessarily follow Gaussian distributions. This is accomplished through a two-step procedure; first, all the uncertain quantities are transformed into Gaussian variables using well-known standard normal transformations, and then, the non-linear function is linearized using first-order Taylor's series expansion, as explained below.

1. Standard normal transformation: First, it is necessary to transform all the variables in X to equivalent normal distributions. For the sake of uniformity, all the variables are transformed to the standard normal distribution. There are several two-parameter and three-parameter transformations discussed in the literature (Haldar & Mahadevan, 2000). This paper uses a simple one-parameter transformation. Consider a single random variable X_i (instead of the vector denoted by X) with PDF $f_{X_i}(x_i)$ and CDF $F_{X_i}(x_i)$. Then, based on the probability integral transform concept, every x_i can be transformed into a corresponding u_i using the equation:

$$u_i = \Phi^{-1}(F_{X_i}(X_i = x_i)) \tag{8}$$

where $\Phi^{-1}(.)$ refers to the inverse of the standard normal distribution function (Haldar & Mahadevan, 2000). Now, u_i is function of x_i , and for any chosen distribution for X_i , the corresponding U_i follows the standard normal distribution N(0, 1). Eq. 8 first calculates the CDF which is equivalent to transforming the original variable to the standard uniform distribution (upper and lower bounds of 0 and 1 respectively), and then calculates the inverse CDF of the standard normal distribution, thereby transforming to the standard normal distribution. Note that the above transformation is performed for each variable X_i individually, and hence it is not directly applicable if the variables X are statistically dependent or correlated. In such cases, it is necessary to transform X into an uncorrelated standard normal space. Such a transformation can be generically represented as U = T(X), and the corresponding inverse transformation can be written as $X = T^{-1}(U)$. Several mathematical transforms have been discussed in the literature for this purpose; for example, refer to Liu and Der Kiureghian (Liu & Der Kiureghian, 1986) for details regarding the Morgenstern transform (Morgenstern, 1956) and the Nataf transform (Nataf, 1962).

2. Linearize "G" using Taylor's Series: Now that all the variables X have been transformed into equivalent standard normal variables U, the next task is to linearize G(X) using Taylor's series methodology. Obviously, the point or location of linearization will affect the estimate of the statistics of R, and hence needs to be chosen carefully. The Most Probable Point (MPP) concept guides in choosing the point of linearization, as explained below.



Figure 6. Most Probable Point Estimation

Instead of trying to estimate the complete statistics of Rby choosing one "global" linearization point, the concept of MPP advocates to solve a "local" problem by focusing on the CDF value at a *particular* realization of R, i.e., $P(R \leq$ r) = $F_R(r) = \eta$. Then, the method linearizes the curve represented by the equation G(x) - r = 0, by choosing an appropriate point of linearization. It is evident that an arbitrary location (say, mean of X) cannot be chosen as the location of linearization, since it may not even satisfy the above equation for a generic value of r. Therefore, the first condition the point of localization must satisfy is the equation $G(\mathbf{x}) - r = 0$. Of the several points that lie on the curve represented by the equation G(x) - r = 0, the point which has the highest likelihood of occurrence is chosen and therefore, the point of linearization is called as the Most Probable Point (MPP). The likelihood of occurrence is proportional to the PDF value and the joint PDF needs to be maximized; in practice, this maximization can be easily performed in the standard normal space. In the standard normal space U, the closer any point is to the origin, the higher is its likelihood of occurrence. Therefore, the MPP is estimated through an optimization problem by estimating the point on the curve $G(\mathbf{x}) - r = 0$ which is closest the origin in the standard normal space, as shown in Fig. 6.

4.2. Calculating the CDF of RUL

The MPP can also be described as the minimum distance (measured from the origin in the standard normal space) point on the curve represented by the equation G(x) - r = 0. Let β denote this minimum distance, i.e., the distance of MPP from the origin in the standard normal space. Then, it can be easily proved (Haldar & Mahadevan, 2000) that:

$$F_R(r) = P(R \le r) = \eta = \Phi(-\beta) \tag{9}$$

where $\Phi(.)$ represents the standard normal distribution function. Thus, estimation of MPP directly leads to the calculation of the CDF $F_R(r)$, only locally at R = r. In a practical problem, the goal is to compute the uncertainty in R, and therefore, it may not be possible to choose a suitable value for r. Inversely, given the value of η , it is also possible to calculate the value of r which satisfies Eq. 9, using optimization. An iterative, numerical procedure for such an optimization is outlined below:

- 1. Given a value of η , compute β such that $\eta = \Phi(-\beta)$.
- 2. Initialize counter j = 0 and start with an initial guess for the MPP, i.e., $x^j = \{x_1^j, x_2^j, \dots, x_i^j, \dots, x_n^j\}$, and a corresponding initial guess for r is obtained.
- 3. Transform into uncorrelated standard normal space u = T(x) and calculate $u^j = \{u_1^j, u_2^j, \dots, u_i^j, \dots, u_n^j\}$. During this transformation, the original distributions of the variables are used. In the case of process noise, the equivalent time-invariant process noise distribution is used. If Gaussian white process loading is considered, then, the equivalent time-invariant loading distribution is used. Recall that the time-invariant distribution is a function of r, which was calculated in the previous step.
- 4. Compute the gradient vector in the standard normal space, i.e., $\alpha = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, a column vector where

$$\alpha_i = \frac{\partial G}{\partial u_i} = \frac{\partial G}{\partial x_i} \times \frac{\partial x_i}{\partial u_i} \tag{10}$$

The above derivatives depends both on G, and the chosen transformation T.

5. In the iterative procedure, the next point u^{j+1} is calculated as:

$$u^{j+1} = -\frac{\alpha}{|\alpha|}\beta \tag{11}$$

6. Transform back into the original space using X = T⁻¹(U). In other words, compute x^{j+1}, and continue starting from Step 3 until the iterative procedure converges. Using tolerance limits δ₁ and δ₂, convergence can be verified if the following two criteria are satisfied: (i) the point lies on the curve of demarcation, i.e., |G(x^j) − r| ≤ δ₁; and (2) the solution does not change between two iterations, i.e., |x^{j+1} − x^j| ≤ δ₂.

Note that the above iterative algorithm relies on the existence of a unique minimum distance point. If G is non-convex or if there are multiple local minima, then the above algorithm may not be able to identify the optimal MPP. This may happen when the probability distribution of RUL is multimodal. The applicability of the inverse-FORM approach to such cases needs to be investigated in future research. In this paper, it is assumed that the MPP can be identified using the above algorithm, usually within four or five iterations.

Hence, given a value of η , the value of $F_R(r)$ can be calculated using the above algorithm. Note that η denotes a probability level, and by choosing multiple values of η such as



Figure 7. Battery equivalent circuit

0.01, 0.1, 0.2, ... 0.9, 0.99, it is possible to estimate the entire CDF of R. In the context of condition-based monitoring, this procedure is repeated at every time instant prognosis needs to be performed. Note that the proposed methodology is an analytical procedure and can produce repeatable (deterministic) calculations as against Monte Carlo sampling-based approaches. This is an important criteria for existing verification, validation, and certification protocols in the aerospace domain. Further, this methodology requires a few tens of prognostic evaluations (in contrast with several hundreds of evaluations required by sampling-based methods), and therefore, directly aids in real-time, online prognosis.

5. CASE STUDY: BATTERY PROGNOSTICS

The proposed methods are illustrated using a lithium-ion battery that powers an unmanned aerial vehicle (Saha, Quach, & Goebel, 2012) at NASA Langley Research Center. This unmanned aerial vehicle is being used as a test-bed for prognostics and decision-making at NASA Ames Research Center and NASA Langley Research Center.

5.1. Description of the Model

The battery model, extended from that used by Daigle et al. (Daigle et al., 2012) for prognosis, is similar to the models presented by Chen and Rincon-Mora (Chen & Rincon-Mora, 2006). The model is based on an electrical circuit equivalent as shown in Fig. 7, where the large capacitance C_b holds the charge q_b of the battery. The nonlinear C_b captures the opencircuit potential and concentration overpotential. The R_{sp} - C_{sp} pair captures the major nonlinear voltage drop due to surface overpotential, R_s captures the so-called Ohmic drop, and R_p models the parasitic resistance that accounts for selfdischarge. This empirical battery model is sufficient to capture the major dynamics of the battery, but ignores temperature effects and other minor battery processes. The governing equations for the battery model are presented in continuous time below. The implementation of the proposed methodology considers a discrete-time version with a discrete timestep of 1 second.

The state-of-charge, SOC, is computed as

$$SOC = 1 - \frac{q_{max} - q_b}{C_{max}},\tag{12}$$

where q_b is the current charge in the battery (related to C_b), q_{max} is the maximum possible charge, and C_{max} is the maximum possible capacity. The resistance related to surface overpotential is a nonlinear function of SOC:

$$R_{sp} = R_{sp_0} + R_{sp_1} \exp\left(R_{sp_2}(1 - SOC)\right), \tag{13}$$

where R_{sp_0} , R_{sp_1} , and R_{sp_2} are empirical parameters. The resistance, and, hence, the voltage drop, increases exponentially as SOC decreases.

Voltage drops across the individual circuit elements are given by

$$V_b = \frac{q_b}{C_b},\tag{14}$$

$$V_{sp} = \frac{q_{sp}}{C_{sp}},\tag{15}$$

$$V_s = \frac{q_s}{C_s},\tag{16}$$

$$V_p = V_b - V_{sp} - V_s, (17)$$

where q_{sp} is the charge associated with the capacitance C_{sp} , and q_s is the charge associated with C_s . The voltage V_b is also the open-circuit voltage of the battery, which is a nonlinear function of SOC (Chen & Rincon-Mora, 2006). This is captured by expressing C_b as a third-order polynomial function of SOC:

$$C_b = C_{b_0} + C_{b_1}SOC + C_{b_2}SOC^2 + C_{b_3}SOC^3$$
(18)

The terminal voltage of the battery is

$$V = V_b - V_{sp} - V_s.$$
 (19)

Currents associated with the individual circuit elements are given by

$$\dot{i}_p = \frac{V_p}{R_p},\tag{20}$$

$$i_b = i_p + i, \tag{21}$$

$$i_{sp} = i_b - \frac{V_{sp}}{R_{sp}},\tag{22}$$

$$i_s = i_b - \frac{V_s}{R_s}.$$
(23)

The charges are then governed by

$$\dot{q}_b = -i_b,\tag{24}$$

$$\dot{q}_{sp} = i_{sp},\tag{25}$$

$$\dot{q}_s = i_s. \tag{26}$$

It is of interest to predict the end-of-discharge as defined by a voltage threshold V_{EOD} (16 V, in this example). So, C_{EOL} consists of only one constraint:

$$c_1: V > V_{EOD}.\tag{27}$$

The parameters of the battery model are assumed to be deterministic and are shown in Table 1. If the parameters are uncertain, and described through probability distributions, then parameter uncertainty can also be easily included, as indicated in Fig. 2.

In Table 1, all voltages are measured in Volts, resistances are measured in Ohms, charges are measured in Coulombs, and capacitances are measured in Coulombs per Volt (or Farads). Note that C_{b_0} , C_{b_1} , C_{b_2} , and C_{b_3} are simply fitting parameters in Eq. 18 and do not have physical meaning.

Table 1. Battery Model Parameters

Parameter	Value	Unit		
C_{b_0}	19.80	Farad (F)		
$C_{b_1}^{0}$	1745.00	Farad (F)		
C_{b_2}	-1.50	Farad (F)		
C_{b_3}	-200.20	Farad (F)		
R_s°	0.0067	Ohm (Ω)		
C_s	115.28	Farad (F)		
R_p	1×10^4	Ohm (Ω)		
C_{sp}	316.69	Farad (F)		
R_{sp_0}	0.0272	Ohm (Ω)		
R_{sp_1}	1.087×10^{-16}	Ohm (Ω)		
R_{sp_2}	34.64	(No unit)		
q_{max}	3.11×10^4	Coulomb (C)		
C_{max}	30807	Coulomb (C)		

The following sections deal with the different sources of uncertainty that affect the RUL prediction, and implement the proposed MPP-based methodology.

5.2. Future Loading Uncertainty

As explained earlier in Section 3, three types of future loading uncertainty are discussed in this paper, and uncertainty quantification results are presented for each type. In this numerical example, loading refers to the current drawn by the battery.

- 1. **Type-I**: The first type of future loading condition is constant amplitude loading condition. The constant amplitude (in Amps) is considered to be normally distributed (N(35, 5)), and this distribution is truncated at a specified lower bound (5.0) and upper bound (80).
- 2. **Type-II**: The second type of future loading condition is white noise, i.e., at every future time instant the loading value is drawn from a particular distribution. In this paper, the distribution is chosen to be truncated normal (N(35,5)) with a specified lower bound (5.0) and upper bound (80), and the loading values at multiple time instants are considered to be independent of each other. Note that the statistics are identical to that of Type-I loading scenario, but the actual loading trajectory is completely different.
- 3. **Type-III**: The third type of future loading condition is chosen to be variable amplitude loading with 6 different segments. The time-length (T) of each segment is chosen

at random and within each time-segment, the amplitude is considered to be constant; further, the constant amplitude (I) is also chosen randomly. Therefore, there are 12 random variables each of which is assumed to follow a truncated normal distribution. This truncated normal distribution is represented using the mean (μ), standard deviation (σ), lower bound (l) and upper bound (u) of random variable, and the statistics of the 12 random variables are tabulated in Table 2. The six segments in Table 2 correspond to multiple flight segments such as takeoff, climb, cruise, landing, etc. Note that the statistics for the amplitude are identical to that of Type-I and Type-II loading, but the actual loading trajectory is completely different.

Such statistics are chosen in order to enable comparison between the three types of loading condition; since the actual amplitudes are similar, it is expected that that overall uncertainty in RUL prediction for each loading scenario should not be wholly different from another.

Table 2. Variable Amplitude Loading: Statistics

Segment	I_{μ}	I_{σ}	I_l	I_u	T_{μ}	T_{σ}	T_l	T_u
Ι	35	5	5	80	60	10	50	75
II	35	5	5	80	120	10	90	140
III	35	5	5	80	90	10	70	115
IV	35	5	5	80	120	10	100	145
V	35	5	5	80	90	10	75	120
VI	35	5	5	80	60	10	40	80

5.3. Process Noise Uncertainty

At any time instant, there are three states, and hence three process noise terms. Typically, the statistics of process noise are calculated as a result of the parameter estimationg procedure. In this study, all the three process noise terms were determined to have zero mean and variances equal to $1, 1 \times 10^{-4}$, and 1×10^{-6} respectively. For the sake of illustration, it is assumed that the three process noise terms are statistically independent, and further, these process noise values at two different time instants are also statistically independent of each other. In this case, it can be shown that, if the true distribution of the process noise follows a normal distribution with mean 0 and standard deviation σ , then the equivalent time-invariant process noise follows a normal distribution with mean 0 and standard deviation $\frac{\sigma}{\sqrt{R}}$, where R is the remaining useful life prediction calculated using G.

5.4. State Uncertainty

Typically, the state estimation, which is an inverse problem, is addressed using a filtering technique that can continuously estimate the uncertainty in the state when measurements are continuously available as a function of time. In this paper, the focus is on prognostic uncertainty, and the state estimation is not explicitly carried out. The state values are assumed to be available, and the uncertainty in the states are predetermined based on the authors' past experiences with the use of filtering techniques for the above described problem. There are three state variables ((1) charge in C_b ; (2) charge in C_{sp} ; and (3) charge in C_s) in this example and at any time instant, they are assumed to be normally distributed with a specified mean. First, the mean of the initial states are chosen to be 3.1×10^4 , 0, and 0 respectively, and the mean values of the states at other time instants are provided in Fig. 8—10. The standard deviation of the states is chosen to 0.1 times the mean of the states; if the mean is zero, then the standard deviation is chosen to be equal to 0.1.







5.5. Prognostics Uncertainty Quantification

The results of prognostics uncertainty quantification are discussed in this section. As explained through Fig. 2, the aforementioned sources of uncertainty affect future state prediction and therefore, the remaining useful life prediction. The proposed MPP-based methodology is used to calculate the uncertainty in RUL, and the 90% probability bounds (estimated by repeating the iterative algorithm for $\eta = 0.05$ and $\eta = 0.95$) of RUL corresponding to the three different loading types are plotted in Fig. 11—13.



Figure 11. Type-I Loading: 90% Bounds of RUL



Figure 12. Type-II Loading: 90% Bounds of RUL



Figure 13. Type-III Loading: 90% Bounds of RUL

Note that the uncertainty is high at initial time instants, and then gradually decreases near the end-of-life of the battery. Initially, the uncertainty in RUL is high because it is necessary to predict at a farther time instant; future loading and the associated uncertainty need to be considered for a longer period of time. However, at a latter time instant, future loading needs to be assumed for a reduced period of time and hence, the uncertainty in the RUL decreases. In fact, any good prognostic algorithm should depict this behavior, i.e., the prediction of RUL at a later time instant must have lower uncertainty than the prediction at an earlier time instant. From Fig. 11 -13, it can be seen that the uncertainty in the RUL prediction is similar for three loading cases; particularly, the uncertainty in the case of Type-III loading was observed to be significantly smaller than the other two loading scenarios. For example, the 90% RUL bounds for Type-I, Type-II, and Type-III loading scenarios at the initial time instant ($t_P = 0$ seconds) are [586 1137], [593 1154], and [667 967] respectively. Further, in this example, the RUL prediction was found to be the most sensitive to the first state variable, i.e., charge in C_b . That is why the RUL prediction in Fig. 11 - 13 is similar to the state evolution in Fig. 8.

For the sake of verification, the computation of RUL was also performed using Monte Carlo sampling and the solutions in Fig. 11-13 were compared. While Fig. 11-13 show the RUL values corresponding to $\eta = 0.05, \eta = 0.50$, and $\eta = 0.95$, all computations were actually performed for 13 different η values (0.01, 0.05, 0.1, 0.2 ... 0.9, 0.95, 0.99) in order to construct the entire CDF of RUL. This CDF was compared with the Monte Carlo estimate and the solution from the proposed method was in good agreement with the Monte Carlo estimate. For example, the comparison between Monte Carlo sampling and MPP-based method in the case of the three loading scenarios (at selected time instants) is shown in Fig. 14—16. As seen from these figures, the error in the proposed methodology is very small, with respect to the Monte Carlo solution. Further, while the Monte Carlo solution required thousand evaluations of G, the proposed MPP-based method required much fewer evaluations. The precise number of evaluations for the MPP-based method depends on the selected number of η values and dimension of X; typically, favourable results are obtained by using about one-tenth of the number of samples required for Monte Carlo sampling. Considering that the proposed method requires much less computational power than Monte Carlo, and that it may be computationally infeasible to perform Monte Carlo sampling at every time instant, it is evident that the MPP-based procedure provides a good alternative for uncertainty quantification in prognostics.

6. CONCLUSION

This paper presented a new computational methodology for quantifying uncertainty in prognostics, in the context of condition-based monitoring. First, a model-based computational framework for prognostics was discussed, and the dif-



Figure 14. Comparison at Time = 400 s: Type-I Loading



Figure 15. Comparison at Initial Time: Type-II Loading

ferent sources of uncertainty were analyzed in the context of this framework. It was demonstrated that the problem of quantifying uncertainty in prognostics and predicting the end-of-life can be posed as an uncertainty propagation problem. An analytical methodology, based on the Most Probable Point (MPP) concept, was proposed to estimate the uncertainty in end-of-life prediction and thereby the remaining useful life prediction. The Most Probable Point concept was originally developed by structural engineers to compute the failure probability of structural engineering applications, and in this paper, this approach has been extended to quantify prognostics uncertainty.

Further, different types of future loading conditions were discussed for prognostics, and it was explained that the overall prediction methodology may need to be adapted to suit each type of loading condition. In this paper, uncertainty quantification methodology was developed for three types of loading conditions: constant amplitude loading, Gaussian white noise loading, and variable amplitude loading, and demonstrated using a numerical example of a battery used to power an unmanned aerial vehicle. In this paper, it was assumed that the information regarding future loading uncertainty was available. Future work may address the characterization of future loading uncertainty based on possible maneuvers and trajectories, and characterization of model uncertainty. Fur-



Figure 16. Comparison at Initial Time: Type-III Loading

ther, the applicability of the proposed methodology may also be investigated to practical situations where the probability distribution of the remaining useful life prediction may be multi-modal in nature.

ACKNOWLEDGMENT

The work reported herein was in part funded by the NASA System-wide Satefy Assurance Technologies (SSAT) project under the Aviation Safety (AvSafe) Program of the Aeronautics Research Mission Directorate (ARMD), and by the NASA Automated Cryogenic Loading Operations (ACLO) project under the Office of the Chief Technologist (OCT) of Advanced Exploration Systems (AES). The authors would also like to thank Dr. Matthew Daigle at NASA Ames Research Center for providing the model for the lithium-ion battery.

REFERENCES

- Chen, M., & Rincon-Mora, G. A. (2006). Accurate electrical battery model capable of predicting runtime and I-V performance. *IEEE Transactions on Energy Conversion*, 21(2), 504 - 511.
- Daigle, M., Saxena, A., & Goebel, K. (2012). An efficient deterministic approach to model-based prediction uncertainty estimation. In Annual Conference of the Prognostics and Health Management Society.
- DeCastro, J. A. (2009). Exact nonlinear filtering and prediction in process model-based prognostics. In Annual Conference of the Prognostics and Health Management Society. San Diego, CA..
- Farrar, C., & Lieven, N. (2007). Damage prognosis: the future of structural health monitoring. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1851), 623– 632.
- Farrar, C. R., Lieven, N. A., & Bement, M. T. (2005). An introduction to damage prognosis. *Damage Prognosis: For Aerospace, Civil and Mechanical Systems*, 1–12.

- Gu, J., Barker, D., & Pecht, M. (2007). Uncertainty assessment of prognostics of electronics subject to random vibration. In *Aaai fall symposium on artificial intelli*gence for prognostics (pp. 50–57).
- Guan, X., Liu, Y., Jha, R., Saxena, A., Celaya, J., & Geobel, K. (2011). Comparison of two probabilistic fatigue damage assessment approaches using prognostic performance metrics. *International Journal of Prognostics* and Health Management, 1, 005.
- Haldar, A., & Mahadevan, S. (2000). *Probability, reliability, and statistical methods in engineering design*. John Wiley & Sons, Inc.
- Inman, D. J., Farrar, C. R., Junior, V. L., & Junior, V. S. (2005). Damage prognosis for aerospace, civil and mechanical systems. Wiley.
- Liu, P.-L., & Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2), 105–112.
- Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitt. Math. Statist*, 8(1), 234–235.
- Nataf, A. (1962). Détermination des distributions de probabilités dont les marges sont données. *Comptes rendus de lacademie des sciences*, 225, 42–43.
- Orchard, M., Kacprzynski, G., Goebel, K., Saha, B., & Vachtsevanos, G. (2008). Advances in uncertainty representation and management for particle filtering applied to prognostics. In *Prognostics and health management*, *international conference on* (p. 1-6).
- Saha, B., & Goebel, K. (2008). Uncertainty management for diagnostics and prognostics of batteries using bayesian techniques. In *Aerospace Conference*, 2008 IEEE (p. 1 -8).
- Saha, B., Quach, C. C., & Goebel, K. (2012). Optimizing battery life for electric uavs using a bayesian framework. In Aerospace Conference, 2012 IEEE.
- Sankararaman, S., Daigle, M., Saxena, A., & Goebel, K. (2013). Analytical algorithms to quantify the uncertainty in remaining useful life prediction. In *Aerospace Conference*, 2013 IEEE (pp. 1–11).
- Sankararaman, S., & Goebel, K. (2013a). Remaining useful life estimation in prognosis: An uncertainty propagation problem. In 2013 AIAA Infotech@Aerospace Conference.
- Sankararaman, S., & Goebel, K. (2013b). Uncertainty quantification in remaining useful life of aerospace components using state space models and inverse form. In *Proceedings of the 15th Non-Deterministic Approaches Conference.*
- Sankararaman, S., Ling, Y., Shantz, C., & Mahadevan, S. (2009). Uncertainty quantification in fatigue damage prognosis. In Annual Conference of the Prognostics and Health Management Society.
- 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).

- Sankararaman, S., & Mahadevan, S. (2011a). Likelihoodbased representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliability En*gineering & System Safety, 96(7), 814 - 824.
- Sankararaman, S., & Mahadevan, S. (2011b). Uncertainty quantification in structural damage diagnosis. *Structural Control and Health Monitoring*, 18(8), 807–824.
- Sankararaman, S., & Mahadevan, S. (2013). Bayesian methodology for diagnosis uncertainty quantification and health monitoring. *Structural Control and Health Monitoring*, 20(1), 88–106.
- Tang, L., Kacprzynski, G. J., Goebel, K., & Vachtsevanos, G. (2009). Methodologies for uncertainty management in prognostics. In *Aerospace Conference*, 2009 IEEE (pp. 1–12).
- Usynin, A., & Hines, J. W. (2007). Uncertainty management in shock models applied to prognostic problems. In *Artificial Intelligence For Prognostics: Papers From The AAAI Fall Symposium.*

BIOGRAPHIES



Shankar Sankararaman received his B.S. degree in Civil Engineering from the Indian Institute of Technology, Madras in India in 2007 and later, obtained his Ph.D. in Civil Engineering from Vanderbilt University, Nashville, Tennessee, U.S.A. in 2012. His research focuses on the various aspects of uncertainty quantification, integration, and management in different types of aerospace, mechanical, and civil engineering systems. His research interests include probabilistic methods, risk and reliability analysis, Bayesian networks, system health monitoring, diagnosis and prognosis, decisionmaking under uncertainty, treatment of epistemic uncertainty, and multidisciplinary analysis. He is a member of the Non-Deterministic Approaches (NDA) technical committee at the American Institute of Aeronautics, the Probabilistic Methods Technical Committee (PMC) at the American Society of Civil Engineers (ASCE), and the Prognostics and Health Management (PHM) Society. Currently, Shankar is a researcher at NASA Ames Research Center, Moffett Field, CA, where he develops algorithms for uncertainty assessment and management in the context of system health monitoring, prognostics, and decision-making.



Kai Goebel is the Deputy Area Lead for Discovery and Systems Health at NASA Ames where he also directs the Prognostics Center of Excellence. After receiving the Ph.D. from the University of California at Berkeley in 1996, Dr. Goebel worked at General Electrics Corporate Research Cen-

ter in Niskayuna, NY from 1997 to 2006 as a senior research scientist before joining NASA. He has carried out applied research in the areas of artificial intelligence, soft computing, and information fusion and his interest lies in advancing these techniques for real time monitoring, diagnostics, and prognostics. He holds 17 patents and has published more than 250 papers in the area of systems health management.