An Uncertainty Quantification Framework for Prognostics and Condition-Based Monitoring

Shankar Sankararaman
SGT Inc., NASA Ames Research Center, Moffett Field, CA 94035, USA
Kai Goebel†
NASA Ames Research Center, Moffett Field, CA 94035, USA

This paper presents a computational framework for uncertainty quantification in prognostics in the context of condition-based monitoring of aerospace systems. The different sources of uncertainty and the various uncertainty quantification activities in condition-based prognostics are outlined in detail, and it is demonstrated that the Bayesian subjective approach is suitable for interpreting uncertainty in online monitoring. A state-space model-based framework for prognostics, that can rigorously account for the various sources of uncertainty, is presented. Prognostics consists of two important steps. First, the state of the system is estimated using Bayesian tracking, and then, the future states of the system are predicted until failure, thereby computing the remaining useful life of the system. The proposed framework is illustrated using the power system of a planetary rover test-bed, which is being developed and studied at NASA Ames Research Center.

I. Introduction

Condition-based monitoring and maintenance¹ have been emerging as new technologies in the field of prognostics and health management of aerospace systems. Prognostics and health management deals with continuous monitoring of the performance of an engineering system, checking for the presence of faults, predicting future failures, estimating the remaining useful life, and aiding decision-making activities such as fault mitigation, mission re-planning, etc. Two different types of approaches, in general, have been pursued for prognostics and health management. While the first approach is based on testing, the second approach is based on condition-based monitoring. While testing needs to be performed on multiple aircraft components and systems, condition-based maintenance emphasizes monitoring the performance of each individual component or system, and recommends maintenance decisions based on information collected through health monitoring.

When a health monitoring system is aboard an engineering system, there are several sources of uncertainty that affect the performances both the health monitoring system and the engineering system. It is important to identify the various sources of uncertainty, and understand how they affect prognostics. In the recent past, several researchers have developed different types of methods for uncertainty quantification in prognostics. For example, the Damage Prognosis project² at Los Alamos National Laboratory discussed uncertainty in fatigue crack growth by considering different sources of variability, in the context of structural health monitoring. Sankararaman et al.³ developed a computational approach to account for variability, data uncertainty, and model uncertainty in crack growth prognosis. However, many of these methods are suitable only for offline prognostics, i.e., testing before and after operation, and not for condition-based monitoring during operation.

Developing methods for uncertainty quantification in the context of condition-based monitoring is challenging because uncertainty methods are generally computationally expensive, whereas condition-based monitoring and prognostics require real time computational power and results for decision-making. An important

¹Post-doctoral Research Scholar, NASA Ames Research Center, Moffett Field, CA 94035, USA.
²Intelligent Systems Division, NASA Ames Research Center, Moffett Field, CA 94035, USA
aspect of condition-based monitoring is the prediction of remaining useful life, and several publications\textsuperscript{4,5} have discussed the importance of quantifying the uncertainty in remaining useful life prediction. Daigle et al.\textsuperscript{6} and Sankararaman et al.\textsuperscript{7} discussed sampling methods and analytical methods for estimating the uncertainty in the remaining useful life prediction for prognostics.

While the significance of uncertainty in condition-based monitoring has been emphasized, it is necessary to develop a framework which can account for the various sources of uncertainty in condition-based monitoring, and aid operational decision-making. This paper develops a general computational framework for uncertainty quantification in prognostics, in the context of condition-based monitoring, and explains the interpretation of uncertainty in condition-based monitoring. First, the various uncertainty quantification activities related to prognostics, and the different sources of uncertainty in condition-based monitoring are discussed in Section II. Finally, the proposed uncertainty quantification framework for prognostics and health monitoring is presented in Section III, and illustrated through an example in Section IV.

II. Uncertainty in Prognostics

It is important to discuss the effects of uncertainty on prognostics, particularly in the context of condition-based monitoring before the computational framework for uncertainty quantification can be developed. First, the various activities related to uncertainty quantification are identified, and then the different sources of uncertainty are discussed. Finally, the interpretation of uncertainty in prognostics is discussed; this interpretation is important since condition-based monitoring deals with the performance and monitoring of one particular system, and there is no “variability” in the context of frequentist analysis.

A. Activities Related to Uncertainty Quantification

In the context of prognostics and health management, uncertainty has been discussed from the perspectives of quantification, representation, and management, in various publications.\textsuperscript{8–11} While all the three are different processes, they are often confused with each other and interchangeably used.\textsuperscript{12} For the purpose of clarity, four different uncertainty-related activities are detailed below:

1. **Uncertainty Representation and Interpretation:** The first step is the representation of uncertainty, which may be guided by the choice of the modeling and simulation framework. There are several methods for uncertainty representation that vary in the level of granularity and detail. A probabilistic framework is used for uncertainty representation in this paper, and the interpretation of this uncertainty is discussed later in this section.

2. **Uncertainty Quantification:** The second step of uncertainty quantification deals with identifying and characterizing the various sources of uncertainty that can prospectively affect the prognostics calculations.

3. **Uncertainty Propagation:** The third step of uncertainty propagation is most relevant to prognostics, since it accounts for all the uncertainties previously quantified and uses this information in order to predict the future behavior, remaining useful life, and the associated uncertainty.

4. **Uncertainty Management:** Uncertainty management consists of two major aspects. The first aspect is to inspect whether it is possible to improve the uncertainty estimates, and the second aspect of uncertainty management deals with how the available uncertainty information needs to be used in the decision-making process.

B. Sources of Uncertainty

It has been conventional to classify the different sources of uncertainty into aleatory (physical variability) and epistemic (lack of knowledge), where epistemic uncertainty consists of data uncertainty and model uncertainty.\textsuperscript{13} However, in the case of condition-based monitoring, there is only one particular system being monitored, and not multiple realizations of a population, and therefore, it is not meaningful to discuss variability. For example, the system is at a particular state at any time instant and there is nothing variable about it. Variability would need to be accounted for only in the case of reliability-testing methods where multiple components/systems are tested. Therefore, the sources of uncertainty in prognostics are classified into:
1. **Modeling uncertainty**: Predicting the future is the most important aspect of prognostics, and typically, a physics-based model or a data-driven model is used for predicting future behavior. This model is usually represented using state-space equations. Modeling uncertainty represents the difference between the predicted response and the true response (which can neither be known nor measured accurately), and comprises of several parts such as model parameters, model form, process noise, etc.

2. **Present uncertainty**: The first step of prognostics is to estimate the condition/state of the component/system at any time instant. Output data (usually collected through sensors) is used to estimate the state and many filtering approaches are able to estimate the state, and calculate the uncertainty associated with the state estimate.

3. **Future uncertainty**: The most important source of uncertainty in the context of prognostics is due to the fact that the future is unknown, i.e. both the loading and operating conditions are not known precisely. The future behavior (i.e., the response of the system to the loading and operating conditions) needs to be estimated using a model; the usage of a model imparts additional uncertainty as explained earlier.

C. **Interpretation of Uncertainty**

Since condition-based monitoring focuses on the operation of one particular engineering component/system, there is no frequentist-related variability in this context, and the Bayesian interpretation of uncertainty is suitable for understanding the meaning of the aforementioned sources of uncertainty. Hence, all the different uncertainties need to interpreted subjectively, i.e., pertaining to the belief of the analyst, and as supported by the available evidence. For example, when a filtering approach calculates the uncertainty in a state estimate, the “true” state is actually deterministic, and the estimated uncertainty reflects the analyst’s belief or knowledge regarding the state uncertainty. Similarly, the future loading uncertainty reflects the analyst’s belief of what is going to happen in the future. Therefore, filtering approaches such as particle filtering, Kalman filtering, etc. are popularly called as Bayesian tracking approaches, not only because they use Bayes theorem recursively, but also interpret the resultant uncertainty from a Bayesian perspective.

### III. Uncertainty Quantification Framework for Health Management

This section develops the proposed framework for prognostics and health management in the context of condition-based monitoring. This framework accounts for the different types of uncertainty discussed earlier in Section II, and can accommodate both physics-based and data-driven approaches.

A. **Prognostics Architecture**

The goal of prognostics to predict the future behavior of a component/system at any genetic 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 \( (u(t)) \) and outputs \( (y(t)) \) to the system are known until the prediction time \( t = t_P \), and in order to perform prediction, the future inputs i.e., \( u(t) \forall t > t_P \) are also need to be available. A generalized prognostics architecture is showed in Fig. 1.

```
1. Estimation
2. Prediction
3. Estimate End-of-Life
```

**Figure 1: Model-Based Prognostics Architecture**

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 and the remaining useful life can be predicted.
B. Modeling

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

\[ x(t) = f(t, x(t), \theta(t), u(t), v(t)) \]  

(1)

where \( x(t) \in \mathbb{R}^{n_x} \) is the state vector, \( \theta(t) \in \mathbb{R}^{n_{\theta}} \) is the parameter vector, \( u(t) \in \mathbb{R}^{n_u} \) is the input vector, \( v(t) \in \mathbb{R}^{n_v} \) is the process noise vector, and \( f \) 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:

\[ y(t) = h(t, x(t), \theta(t), u(t), n(t)) \]  

(2)

where \( y(t) \in \mathbb{R}^{n_y} \), \( n(t) \in \mathbb{R}^{n_n} \), and \( h \) denote the output vector, 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.

C. State Estimation

Bayesian tracking approaches as Kalman filtering, particle filtering, etc. can be used for estimation purpose. 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.

Daigle et al.\(^6\) used unscented Kalman filters to approximate the mean and variance of states, and this approach is applicable even for non-linear models. This method is based on generating weighted samples and continuously updating the probability distributions of the state variables using Bayesian filtering.

D. Prediction

Having estimated the uncertainty in the states at time \( t_p \), this uncertainty, along with the other uncertainties associated with the model and the future prediction, need to be propagated through Eq. 1 until failure occurs. The function through which these sources of uncertainty need to be propagated is denoted by “\( Y = G(X) \)”, and this function is a combination of the the state space equations that are used for prediction and the threshold function that is used to check whether failure has occurred. Note that \( X \) denotes the list of uncertain quantities, and \( Y \) denotes the quantity of interest (remaining useful life or end-of-life, as desired). Thus, prognosis becomes an uncertainty propagation problem,\(^{14} \) i.e., it is necessary to propagate the uncertainty in \( X \) through \( G \) to compute the uncertainty in \( Y \), and can be solved using a variety of statistical methods. Such methods can be classified into two categories: sampling-based methods and analytical methods. Some of the commonly used sampling-based methods are:

1. **Importance Sampling**: This algorithm does not generate random realizations of \( X \) from the original distribution. Instead, random realizations are generated from a proposal density function, statistics of \( Y \) are estimated and then corrected based on the original density values and proposal density values.

2. **Stratified Sampling**: In this sampling approach, the overall domain of \( X \) is divided into multiple sub-domains and samples are drawn from each sub-domain independently. The process of dividing the overall domain into multiple sub-domains is referred to as stratification. This method is applicable when sub-populations within the overall population are significantly different.

3. **Latin Hypercube Sampling**: This is a sampling method commonly used in design of computer experiments. When sampling a function of \( N \) variables, the range of each variable is divided into \( M \) equally probable intervals, thereby forming a rectangular grid. Then, sample positions are chosen such that there is exactly one sample in each row and exactly one sample in each column of this grid. Each resultant sample is then used to compute a corresponding realization of \( Y \), and thereby the PDF \( f_Y(y) \) can be calculated.
4. **Unscented Transform Sampling**: Unscented transform sampling is a sampling approach which focuses on estimating the mean and variance of $Y$ accurately, instead of the entire probability distribution of $Y$. Certain pre-determined sigma points are selected in the $X$ space and these sigma points are used to generate corresponding realizations of $Y$. Using weighted averaging principles, the mean and variance of $Y$ are calculated. While this approach has been predominantly used in the context of state estimation, it can also be used for prediction.

Some of the commonly used analytical methods are:

1. **First Order Second Moment Method**: This method uses only the mean and variance of all the uncertain quantities and the first-order Taylor's series expansion of $G$, and estimates the mean and variance of the response quantity $Y$.

2. **First Order Reliability Method**: This method calculates the CDF function $F_Y(y)$ by linearizing $G$ around the so-called most probable point. While this approach is an approximation, it can estimate the CDF with reasonable accuracy in many practical applications.

3. **Inverse First Order Reliability Method**: This method is the inverse of the first-order reliability method, i.e., it calculates the value of $y$ that corresponds to a given value of $\beta$ such that $F_Y(y) = \beta$. By repeating this approach for several values of $\beta$, the entire cumulative distribution function can be easily calculated, thereby estimating the uncertainty in $Y$.

In addition to the aforementioned methods, there are several surrogate modeling approaches that have been commonly used by researchers for the purpose of uncertainty quantification. Such surrogate modeling approaches include regression techniques, polynomial chaos expansion, kriging, etc. Each of these methods use different types of basis functions and one may approximate $G$ better than the other; therefore, it may be useful to study such surrogate modeling approaches for uncertainty quantification in prognostics.

Even though there are several such uncertainty propagation methods, it is important to investigate their applicability to prognostics. While some methods like Monte Carlo sampling can be very accurate, they may not be readily suitable for online health monitoring due to the computational cost involved. On the hand, analytical methods, though based on approximations, can obtain quick results that can be processed during the operation of the engineering system. Even after selecting a suitable approach for uncertainty propagation, there are several practical challenges involved during implementation. Some of them are described below:

1. **Dimensionality**: Since state space models are solved using time-discretization, the number of time-steps until failure is a function of the chosen discretization level. At each time-instant, there are different uncertain variables such as loading, state estimates, model parameters, model process noise, etc. If it is necessary to consider a few thousands of time-steps, then it may be necessary to consider several thousands of random variables during uncertainty propagation.

2. **Future Loading**: The most challenging issue in prognostics is to estimate the future loading and the associated uncertainty. It is necessary to identify possible future maneuvers and quantify the loading demand for each maneuver in order to address this challenge.

3. **Model Uncertainty**: Quantifying model uncertainty is a challenge that is still being addressed for systems which don’t exhibit time-dependent behavior. While literature pertaining to time-dependent systems have addressed model uncertainty mainly through the inclusion of process noise, further research is necessary to develop methods to quantify model form uncertainty and model parameter uncertainty.

The first challenge of dimensionality is a computational issue, and is being studied by several researchers. Though the proposed framework for uncertainty quantification is capable of including future loading uncertainty, future model uncertainty, etc., the real challenge lies in characterizing these sources of uncertainty. In fact, these challenges are also still being studied, and future work needs to address the estimation of such sources of uncertainty. It must be noted that the proposed framework for uncertainty quantification in prognostics is general, and at present, it is assumed that all these sources of uncertainty can be characterized and included in prognostics.
IV. Application to a Planetary Rover’s Power System

This section discusses the application of the proposed computational framework for condition-based monitoring to the power system of a planetary rover. The power system consists of several lithium batteries, and is based on the empirical model used by Daigle et al.\textsuperscript{6} for prognosis. The complete details of the application, numerical results, and conclusions will be presented in the final paper.

A. Battery Model

The battery model is based on an electrical circuit equivalent as shown in Fig. 2.\textsuperscript{6} The large capacitance $C_b$ holds the charge $q_b$ of the battery. The nonlinear $C_b$ captures the open-circuit 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 self-discharge. This empirical battery model is sufficient to capture the major dynamics of the battery while ignoring temperature effects and additional minor 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 time-step of 1 s.

![Figure 2: Battery equivalent circuit](image)

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

$$SOC = 1 - \frac{q_{\text{max}} - q_b}{C_{\text{max}}},$$  \hspace{1cm} (3)

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

$$R_{sp} = R_{sp0} + R_{sp1} \exp (R_{sp2}(1 - SOC)),$$  \hspace{1cm} (4)

where $R_{sp0}$, $R_{sp1}$, and $R_{sp2}$ 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

$$V_b = \frac{q_b}{C_b},$$  \hspace{1cm} (5)

$$V_{sp} = \frac{q_{sp}}{C_{sp}},$$  \hspace{1cm} (6)

$$V_s = \frac{q_s}{C_s},$$  \hspace{1cm} (7)

$$V_p = V_b - V_{sp} - V_s,$$  \hspace{1cm} (8)

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 $SOC$\textsuperscript{6}. This is captured by expressing $C_b$ as a third-order polynomial function of $SOC$:

$$C_b = C_{b0} + C_{b1}SOC + C_{b2}SOC^2 + C_{b3}SOC^3$$  \hspace{1cm} (9)
The terminal voltage of the battery is
\[ V = V_b - V_{sp} - V_s. \] (10)

Currents associated with the individual circuit elements are given by
\[ i_p = \frac{V_p}{R_p}, \] (11)
\[ i_b = i_p + i, \] (12)
\[ i_{sp} = i_b - \frac{V_{sp}}{R_{sp}}, \] (13)
\[ i_s = i_b - \frac{V_s}{R_s}. \] (14)

The charges are then governed by
\[ \dot{q}_b = -i_b, \] (15)
\[ \dot{q}_{sp} = i_{sp}, \] (16)
\[ \dot{q}_s = i_s. \] (17)

In the case of the battery, the event \( E \) we are interested in predicting is EOD. \( T_E \) is specified as \( V < V_{EOD} \), where \( V_{EOD} = 2.5 \text{ V} \).

B. Numerical Values

The parameter values of the battery model are given 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 \( C_{b0}, C_{b1}, C_{b2}, \) and \( C_{b3} \) are simply fitting parameters in Eq. 9 and do not have physical meaning.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C_{b0} )</td>
<td>19.80 F</td>
</tr>
<tr>
<td>( C_{b1} )</td>
<td>1745.00 F</td>
</tr>
<tr>
<td>( C_{b2} )</td>
<td>-1.50 F</td>
</tr>
<tr>
<td>( C_{b3} )</td>
<td>-200.20 F</td>
</tr>
<tr>
<td>( R_s )</td>
<td>0.0067 Ω</td>
</tr>
<tr>
<td>( C_s )</td>
<td>115.28 F</td>
</tr>
<tr>
<td>( R_p )</td>
<td>( 1 \times 10^4 ) Ω</td>
</tr>
<tr>
<td>( C_{sp} )</td>
<td>316.69 F</td>
</tr>
<tr>
<td>( R_{sp0} )</td>
<td>0.0272 Ω</td>
</tr>
<tr>
<td>( R_{sp1} )</td>
<td>( 1.087 \times 10^{-16} ) Ω</td>
</tr>
<tr>
<td>( R_{sp2} )</td>
<td>34.64</td>
</tr>
<tr>
<td>( q_{max} )</td>
<td>( 3.11 \times 10^4 ) C</td>
</tr>
<tr>
<td>( C_{max} )</td>
<td>30807 C</td>
</tr>
</tbody>
</table>

For the battery model, the three state quantities are given by three charges: \( q_b, q_{sp}, \) and \( q_s \). All parameters are assumed constant and completely known. The current in the battery is chosen to be input and two cases of input loading are considered: constant amplitude loading and variable amplitude loading. In constant amplitude loading, the future current demand is assumed to be constant, but the constant value is a random variable that follows normal distribution \( \text{N}(1.375, 1/6) \), but is truncated at a specified lower bound (0.75) and an upper bound (2.00). In the variable amplitude loading scenario, constant amplitude loading is assumed for a particular time segment; the length of the time segment and the constant amplitude for that time segment are both chosen at random. This numerical example considers consecutive time segments; the
time-length of each segment is assumed to be uniformly distributed (with lower bound $T_L = 500$ seconds and standard deviation $T_U = 1000$ seconds) and the constant amplitude for that time segment is assumed to be normally distributed (with mean $I_\mu = 1.375$ amps and standard deviation $I_\sigma = 1/6$ amps). Multiple realizations of the loading trajectory (current versus time) are indicated in Fig. 3.

![Figure 3: Current versus Time: Multiple Realizations](image)

C. Results of Uncertainty Quantification

The results of quantifying the uncertainty in the end-of-life prediction (which corresponds to the end-of-discharge of the battery), continuously as a function of time, are shown in Fig. 4.

![Figure 4: Results of Uncertainty Quantification](image)

It can be seen that the uncertainty in the end-of-discharge prediction decreases with time. Initially, it is necessary to assume loading conditions for a longer period of time and therefore, there is a large amount of uncertainty. In later stages of monitoring, loading conditions need to be assumed for a shorter period of time, and therefore, results in decreased uncertainty in the end-of-discharge prediction.

The most important goal during uncertainty quantification is to be able to accurately capture the entire probability distribution. Sometimes, the probability distribution of end-of-life may be multi-modal in nature; sometimes, the shape of the probability distribution may significantly change shape over the course of time. Therefore, computational methods need to be able to address such issues. As stated earlier, the methods discussed in Section III need to be studied in detail and their applicability to prognostics and condition-based monitoring must be investigated.
V. Conclusion

This paper presented a computational framework for uncertainty quantification in prognostics, in the context of condition-based monitoring. An overview of the importance of uncertainty in prognostics was presented, and the proposed uncertainty framework was discussed in detail. This framework consists of two steps: the first step involves estimating the state of the system using Bayesian tracking methods and the second step involves predicting the future behavior by propagating the uncertainty in the state estimates, model uncertainty, and uncertainty in the future loading (and usage) conditions through the system model until the failure threshold is reached. This indicates the end of life (EOL), and the uncertainty in the EOL can be calculated using different types of uncertainty propagation techniques.

It is important to accurately estimate the entire probability distribution of EOL while being able to compute this probability distribution in real time so that the method can be suitable for online health monitoring. While some methods (like Monte Carlo sampling) are accurate, they may be computationally intensive and therefore, not suitable for online health monitoring. Some other analytical methods may be approximate but suitable for online health monitoring. Future work will investigate the applicability of different types of uncertainty quantification methods to prognostics and identify those methods which can be implemented in online health monitoring and condition-based maintenance.

Acknowledgments

The work reported herein was in part funded by the NASA System-wide Safety Assurance Technologies (SSAT) project under the Aviation Safety (AvSafe) Program of the Aeronautics Research Mission Directorate (ARMD), and by the NASA Advanced Cryogenic Loading Operations (ACLO) project under the Office of the Chief Technologist (OCT) of Advanced Exploration Systems (AES).

References