Uncertainty in Prognostics and Health Management: An Overview

Shankar Sankararaman[1] and Kai Goebel[2]

[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 presents an overview of various aspects of uncertainty quantification in prognostics and health management. Since prognostics deals with predicting the future behavior of engineering systems and it is almost practically impossible to precisely predict future events, it is necessary to account for the different sources of uncertainty that affect prognostics, and develop a systematic framework for uncertainty quantification and management in this context. Researchers have developed computational methods for prognostics, both in the context of testing-based health management and condition-based health management. However, one important issue is that, the interpretation of uncertainty for these two different types of situations is completely different. While both the frequentist (based on the presence of true variability) and Bayesian (based on subjective assessment) approaches are applicable in the context of testing-based health management, only the Bayesian approach is applicable in the context of condition-based health management. This paper explains that the computation of the remaining useful life is more meaningful in the context of condition-based monitoring and needs to be approached as an uncertainty propagation problem. Numerical examples are presented to illustrate the various concepts discussed in the paper.

1. INTRODUCTION

Prognostics is the art of predicting the future behavior of engineering systems, analyzing possible failure modes, and estimating the remaining useful life (RUL) of such systems. Since it is practically impossible to precisely predict future events and future behavior, it is imperative for an efficient and accurate Prognostics and Health Management (PHM) system to account for the different sources of uncertainty that are associated with system behavior and quantify the com-

bined effect of these sources of uncertainty on prognostics and remaining useful life prediction in order to facilitate risk-informed decision-making.

Existing methods for prognostics and health management can be broadly classified as being applicable to two different types of situations: testing-based prognostics and condition-based prognostics. Methods for testing-based prognostics are based on rigorous testing before and/or after operating an engineering system (offline), whereas methods for condition-based prognostics are based on monitoring the performance of the engineering system during operation (online). Researchers have developed computational methods for both testing-based and condition-based prognostics and health management. Both data-driven methods and model-based approaches been pursued for these purposes. While some of the initial research efforts did not explicitly account for the effects of uncertainty, some of the later efforts have exclusively focused on uncertainty quantification and management in prognostics.

Several researchers have developed methods for uncertainty quantification in crack growth analysis (Sankararaman, Ling, Shantz, & Mahadevan, 2011; Sankararaman, Ling, & Mahadevan, 2011), structural damage prognosis (Farrar & Lieven, 2007; Coppe, Haftka, Kim, & Yuan, 2010), electronics (Gu, Barker, & Pecht, 2007), and mechanical bearings (Liao, Zhao, & Guo, 2006), primarily in the context of offline testing. Such approaches may be applicable to smaller components since it is possible and affordable to perform laboratory tests until these components fail. However, it may not practically feasible to extend this approach to large scale expensive systems that cannot be tested. Further, the estimation of remaining useful life is more significant in an online health monitoring context where the performance of a system under operation needs to be monitored and its remaining useful life needs to be calculated. Engel et. al (Engel, Gilmartin, Bongort, & Hess, 2000) discuss several issues involved in the estimation of remaining useful life in online prognostics and health monitoring. Though some of the initial studies on remaining useful life prediction lacked uncertainty mea-

sures (Celaya, Saxena, Kulkarni, Saha, & Goebel, 2012), researchers have recently started investigating the impact of uncertainty on estimating the remaining useful life. For example, there have been several efforts to quantify the uncertainty in remaining useful life of batteries (Saha & Goebel, 2008) and pneumatic valves (Daigle & Goebel, 2010) in the context of online health monitoring. Different types of sampling techniques (Daigle, Saxena, & Goebel, 2012) and analytical methods (Sankararaman, Daigle, Saxena, & Goebel, 2013) have been proposed to predict the uncertainty in the remaining useful life.

A review of the aforementioned articles reveal that there exist several challenges in applying uncertainty quantification methods for prognostics. The primary challenge lies in the understanding the philosophical differences between testing-based health management and condition-based health management, since these differences significantly influence the interpretation of uncertainty (Sankararaman & Goebel, 2013c; Celaya, Saxena, & Goebel, 2012). Such interpretation is key to guide different types of decision-making activities during the operation of engineering systems.

The paper focuses on providing an overview of the state-of-the-art in the topic of uncertainty quantification and management in prognostics and health monitoring. To begin with, the significance of uncertainty in prognostics is explained in detail in Section 2. Then, the various aspects of uncertainty in testing-based health management and condition-based health management are discussed in detail in Section 3 and 4, and the differences between these two approaches are clearly explained. It is also explained that the prediction of remaining useful life is more meaningful only in the context of condition-based health management, and this topic is discussed in further detail. Numerical examples are presented in Sections 3 and Section 4, to illustrate the various concepts discussed in this paper. Finally, conclusions are presented in Section 5.

2. SIGNIFICANCE OF UNCERTAINTY IN PROGNOSTICS

In an ideal scenario, it would be possible to perfectly and precisely predict the behavior of engineering systems and facilitate decision-making with a significant amount of trust and confidence. However, this is not possible in practical engineering applications. First of all, it is almost impossible to be able to accurately predict the operating conditions and environmental conditions under which the system operates. Further, the future loading demands on the system cannot be precisely known in advance; for example, the future behavior of a simple electric vehicle depends upon several factors such as the driving terrain, climatic conditions, desired speed and acceleration, characteristics, properties, and parameters of the internal batteries, remaining charge, etc. While some factors are internal to the engineering system, other factors are external to the system. In order to be able to account for all of

these factors and perform prognostics, it is necessary to acknowledge the presence of uncertainty in all of these factors and develop a systematic framework in order to account for these uncertainties in prognostics.

When information regarding uncertainty is used for decision-making, it can lead quantifying the amount of risk involved in different types of decisions. Risk consists of two important components: the likelihood of occurrence of adverse events and the cost associated with the occurrence of adverse events. While the latter can be easily quantified by analyzing the different types of losses that occur due to such occurrence of adverse events, the former can only be quantified by rigorously accounting for the different sources of uncertainty in prognostic and decision-making activities.

It is a common misconception that the effect of uncertainty can be included at latter stages of the analysis when the fundamental deterministic problem has been solved without accounting for uncertainty. It is necessary to account for uncertainty right from the initial stages of system-level conception through analysis, design, testing, and operations. During these stages, there are several types of activities that need to be performed in order to accurately account for the effect of uncertainty in prognostics.

In the context of prognostics and health management, uncertainties have been discussed from representation, quantification, and management points of view (Hastings, D. and McManus, H., 2004; Orchard, Kacprzynski, Goebel, Saha, & Vachtsevanos, 2008; Tang, Kacprzynski, Goebel, & Vachtsevanos, 2009). While these three are different processes, they are often confused with each other and interchangeably used. In this paper, the various tasks related to uncertainty quantification and management are classified into four, as explained below. These four tasks need to performed in order to accurately estimate the uncertainty in the RUL prediction and inform the decision-maker regarding such uncertainty.

1. **Uncertainty Representation and Interpretation:** The first step is uncertainty representation and interpretation, which in many practical applications, is guided by the choice of modeling and simulation frameworks. There are several methods for uncertainty representation that vary in the level of granularity and detail. Some common theories include classical set theory, probability theory, fuzzy set theory, fuzzy measure (plausibility and belief) theory, rough set (upper and lower approximations) theory, etc. Amongst these theories, probability theory has been widely used in the PHM domain (Celaya, Saxena, & Goebel, 2012); even within the context of probabilistic methods, uncertainty can be interpreted and perceived in two different ways: frequentist (classical) versus subjective (Bayesian). While the former interpretation of uncertainty implies that uncertainty exists only when there is natural randomness across multiple nominally identi-

cal experiments, the latter facilitates associating uncertainty even with events that are not random and such uncertainty is simply reflective of the analyst's regarding the occurrence or non-occurrence of such events.

2. **Uncertainty Quantification:** The second step is uncertainty quantification, that deals with identifying and characterizing the various sources of uncertainty that may affect prognostics and RUL estimation. It is important that these sources of uncertainty are incorporated into models and simulations as accurately as possible. The common sources of uncertainty in a typical PHM application include modeling errors, model parameters, sensor noise and measurement errors, state estimates (at the time at which prediction needs to be performed), future loading, operating and environmental conditions, etc. The goal in this step is to address each of these uncertainties separately and quantify them using probabilistic/statistical methods. The Kalman filter is essentially a Bayesian tool for uncertainty quantification, where the uncertainty in the states is estimated continuously as a function of time, based on data which is also typically available continuously as a function of time.

3. **Uncertainty Propagation:** The third step is uncertainty propagation and is most relevant to prognostics, since it accounts for all the previously quantified uncertainties and uses this information to predict (1) future states and the associated uncertainty; and (2) remaining useful life and the associated uncertainty. The former is computed by propagating the various sources of uncertainty through the prediction model. The latter is computed using the estimated uncertainty in the future states along with a Boolean threshold function which is used to indicate end-of-life. In this step, it is important to understand that the future states and remaining useful life predictions are simply dependent upon the various uncertainties characterized in the previous step, and therefore, the distribution type and distribution parameters of future states and remaining useful life should not be arbitrarily chosen. Sometimes, a normal (Gaussian) distribution has been assigned to the remaining useful life prediction; such an assignment is erroneous and the true probability distribution of RUL needs to be estimated though rigorous uncertainty propagation of the various sources of uncertainty through the state space model and the EOL threshold function, both of which may be non-linear in practice.

4. **Uncertainty Management:** The fourth and final step is uncertainty management, and it is unfortunate that, in several articles, the term "Uncertainty Management" has been used instead of uncertainty quantification and/or propagation. As a result, there are few publications that directly address the issue of uncertainty management. In general, uncertainty management is a term used to refer to different activities which aid in managing uncertainty in condition-based maintenance during real-time operation. There are several aspects of uncertainty management. One aspect of uncertainty management attempts to answer the question: "Is it possible to improve the uncertainty estimates?" The answer to this question lies in identifying which sources of uncertainty are significant contributors to the uncertainty in the RUL prediction. For example, if the quality of the sensors can be improved, then it may be possible to obtain a better state estimate (with lesser uncertainty) during Kalman filtering, which may in turn lead to a less uncertain RUL prediction. Another aspect of uncertainty management deals with how uncertainty-related information can be used in the decision-making process. Future research needs to significantly focus on the different aspects of uncertainty management and develop computational methods for this purpose.

Most of the research in the PHM community pertains to the topics of uncertainty quantification and propagation; few articles have directly addressed the topic of uncertainty management. Even within the realm of uncertainty quantification and propagation, the estimates of uncertainty have sometimes been misinterpreted. For example, when statistical principles are used to estimate a parameter, there is an emphasis on calculating the estimate with the minimum variance. When this principle is applied to RUL estimation, it is important not to arbitrarily reduce the variance of RUL itself. Celaya et al. (Celaya, Saxena, & Goebel, 2012) explored this idea and explained that the variance of RUL needs to be carefully calculated by accounting for the different sources of uncertainty. The calculation of RUL is, arguably, the most important component of a prognostics and health management system, and this topic of discussed in detail, in the rest of this paper. Though the majority of this paper focuses on calculating RUL in the context of condition-based monitoring, some fundamental principles of testing-based health management are discussed, particularly from the perspective of uncertainty quantification, in order to explain the philosophical differences between these two approaches.

3. TESTING-BASED HEALTH MANAGEMENT

In testing-based prognostics (referred to as "reliability-based testing" in some publications), the remaining useful life is typically calculated by testing multiple nominally identical specimens of the engineering component/system. It may be noted that the term "remaining" in "remaining useful life" may not be applicable to all types of testing. This is because, testing is typically carried out before the engineering system is under operation. The term "time-to-failure" is more appropriate for testing-based health management. It is important not to confound "time-to-failure" and "remaining useful life". The appropriate interpretation of the latter will be clarified

in the next section, while discussing about condition-based health management.

Assume that a set of run to failure experiments have been performed with high level of control, ensuring same usage and operating conditions. The time to failure for all the n samples ($r_i; i = 1$ to n) are measured. It is important to understand that *different* time-to-failure values are obtained due to inherent variability across the n different specimens, thereby confirming the presence of physical probabilities or true randomness. The various factors that contribute are:

1. Inherent variability in properties and characteristics of the nominally identical specimens

2. Inherent variability across the loading conditions experienced by each of the individual specimens

3. Inherent variability in operating and environmental conditions for each of the individual specimens

Assume that these random samples belong to an underlying probability density function (PDF) $f_R(r)$, with expected value $E(R) = \mu$ and variance $Var(R) = \sigma^2$. The goal of uncertainty quantification is to characterize this probability density function based on the available n data. Theoretically, an infinite amount of data is necessary to accurately estimate this PDF; however, due to the presence limited data, the estimated PDF is not accurate. Hence, lack of infinite data adds some additional uncertainty to the aforementioned list of sources of uncertainty. Statistical approaches, both frequentist and subjective, express uncertainty regarding the estimate itself. However, frequentist and subjective analysts quantify and express this uncertainty in completely different ways. The following discussion is based on the assumption that the underlying PDF $f_R(r)$ is Gaussian, since closed form expressions for uncertainty are readily available for this case. Whenever appropriate and necessary, remarks are provided for non-Gaussian distributions.

3.1. Confidence Intervals: Frequentist Approach

Since R is Gaussian, estimating the parameters μ and σ is equivalent to estimating the PDF. In the context of physical probabilities (frequentist approach), the "true" underlying parameters μ and σ are referred to as "population mean" and "population standard deviation" respectively. Let $\bar{x}$ and s denote the mean and the standard deviation of the available n data. As stated earlier, due to the presence of limited data, the sample parameters ($\bar{x}$ and s) will not be equal to the corresponding population parameters (μ and σ). The fundamental assumption in this approach is that, since there are true but unknown population parameters, it is meaningless to talk about the probability distribution of any population parameter. Instead, the sample parameters are treated as random variables, i.e., if another set of n data were available, then another realization of $\bar{x}$ and s would have been obtained. Using the sample parameters (μ and σ) and the number of data

available (n), frequentists construct confidence intervals on the population parameters.

Confidence intervals can be constructed for both μ and σ (Haldar & Mahadevan, 2000). Consider multiple nominally identical specimens of an engineering component. The term "nominally identical" implies that there is inherent variability in the properties and behavior of these specimens. Suppose that these specimens have been subjective to failure analysis, and their time-to-failure times are available. If the true probability distribution of time-to-failure across multiple specimens is *assumed* to be Gaussian, the $(1 - \alpha)\%$ confidence interval of the mean run-to-failure time can be calculated as:

$$\left[\bar{x} - t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{\frac{\alpha}{2}} \frac{s}{\sqrt{n}} \right],$$

where $\bar{x}$, s, and n denote the sample mean, sample standard deviation, and number of samples respectively. If the run-to-failure times are given by $\{100, 105, 98, 110, 92, 97, 85, 120, 93, 101\}$, then $\mu = 100.10$, $s = 9.87$, $n = 10$, and the 95% confidence interval on the mean run-to-failure is given by [93.98, 106.22]. Using the properties of the chi-square distribution (χ^2), the confidence interval on the variance can be calculated as:

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \right].$$

For this numerical example, the corresponding confidence interval on the standard deviation is given by [6.79, 18.02]. While the above expressions for confidence intervals on mean and standard deviation are applicable only to Gaussian distributions, similar confidence intervals can also be constructed for other types of distributions; in general, it is easier to construct confidence intervals for mean than it is for standard deviation (or equivalently, variance).

Nevertheless, it is important that these confidence intervals be interpreted correctly. To begin with, the above confidence intervals will decrease as more data is available; therefore, the width of these confidence intervals is simply related to the number of data. The actual uncertainty in the run-to-failures times is given only by the estimate of the standard deviation, and this uncertainty is the result of variability (in material properties, operating conditions, etc.) across all the nominally identical specimens. Further, as stated earlier, the interpretation of confidence intervals may be confusing and misleading. A 95% confidence interval on μ *does not* imply that "the probability that μ lies in the interval is equal to 95%"; such a statement is wrong because μ is purely deterministic and physical probabilities cannot be associated with it. The random variable here is in fact $\bar{x}$, and the confidence interval is calculated using $\bar{x}$. Therefore, the correct implication is that "the probability that the estimated confidence interval contains the true population mean is equal to 95%". Thus, it

4

is easy to understand that, the width of the confidence intervals is indicative of lack of infinite data and the actual value of the standard deviation is indicative of the uncertainty in R.

A practical challenge is that, in many applications, it may not be possible to know what type of probability distribution (for example, Gaussian distribution had been "assumed" in the above discussion) needs to be assumed to in order to calculate the above confidence intervals; obviously, the procedure for calculation of confidence intervals depends on the choice of distribution type (Gaussian, Weibull, lognormal, etc.), and the presence of such distribution type uncertainty further adds to the confusion regarding the interpretation of confidence intervals. As the sample size increases, the confidence intervals for the mean and standard deviation may get narrower. This may be misleading since the confidence intervals should be interpreted only based on the underlying assumption of distribution type (which might have been wrong to begin with). Computational methods are being developed to deal with distribution type uncertainty (Sankararaman & Mahadevan, 2013a), however they have not been implemented in prognostics and health management applications.

3.2. Probability Distribution: Bayesian Approach

Alternatively, it is also possible to address the problem of computing $f_R(r)$ purely from a subjective (Bayesian) point of view. One important difference now is that the Bayesian approach does not clearly differentiate between "sample parameters" and "population parameters". The probability distribution of μ is directly computed using the available data (recall that this was impossible in the frequentist approach since μ is the underlying mean that is precise but unknown), and this uncertainty is referred to as the analyst's degree of belief for the underlying true parameter μ. Similarly, the probability distribution of σ can also be computed using Bayes' theorem.

Consider a set of time-to-failure times, given by r_i ($i = 1$ to n). In order to compute the probability distribution of μ and σ, the first step is construct their joint likelihood as (Sankararaman & Mahadevan, 2011):

$$L(\mu, \sigma) \propto \prod_{i=1}^{m} f_R(r_i|\mu, \sigma) \qquad (1)$$

The maximum likelihood estimate of the parameters $\boldsymbol{P}$ can be calculated by maximizing the above expression. Instead of maximizing the likelihood, the entire likelihood function can be used to construct the PDF of the distribution parameters. Further, sometimes time-to-failure data may also be available in terms of intervals. For example, intermittent inspections may be performed to check whether failure has occurred in a specimen; if failure is found to have occurred between 10 minutes and 11 minutes, the resultant time to failure is actually an interval. The above likelihood-based approach can also be extended to account for interval data, in order to com-

pute the uncertainty in the distribution parameters.

This approach is generally applicable for any type of parametric probability distribution, where the probability density function (PDF) can be expressed as $f_R(r|\boldsymbol{P})$. If R is Gaussian, then $\boldsymbol{P}$ represents the vector of mean and standard deviation. Let $f(\boldsymbol{P})$ denote the joint PDF of the distribution parameters $\boldsymbol{P}$. It is easy to apply Bayes theorem, choose uniform prior density ($f'(\boldsymbol{P}) = h$), and calculate the joint PDF as:

$$f(\boldsymbol{P}) = \frac{hL(\boldsymbol{P})}{\int hL(\boldsymbol{P})d\boldsymbol{P}} = \frac{L(\boldsymbol{P})}{\int L(\boldsymbol{P})d\boldsymbol{P}} \qquad (2)$$

Note that the uniform prior density function can be defined over the entire admissible range of the parameters $\boldsymbol{P}$. For example, the mean of a normal distribution can vary in $(-\infty, \infty)$ while the standard deviation can vary in $(0, \infty)$ because the standard deviation is always greater than zero. Both these prior distributions are improper prior distributions because they do not have finite bounds.

For the above numerical example, i.e., if the run-to-failure times are given by $\{100, 105, 98, 110, 92, 97, 85, 120, 93, 101\}$, the probability distribution of μ and σ can be calculated as shown in Figs. 1 and 2.
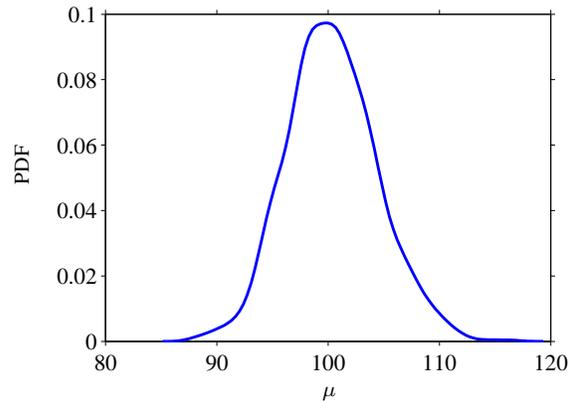

Figure 1. PDF of μ

Recall that one realization of the parameters (μ and σ) uniquely define the PDF $f_R(r)$. However, since the parameters are themselves uncertain, R is now represented by a family of distributions (Sankararaman & Mahadevan, 2011, 2013b). This family of distributions will shrink to the true underlying PDF as the number of available data increases, and asymptotic PDF (as the number data increases) is simply reflective of the variability (in material properties, operating conditions, etc.) across all the nominally identical specimens. Alternative to the family of PDFs approach, a single unconditional PDF of X, which includes both the variability in X and the uncertainty in the distribution parameters $\boldsymbol{P}$, as:

$$f'_R(r) = \int f_R(r|\boldsymbol{P})f(\boldsymbol{P})d\boldsymbol{P} \qquad (3)$$

5
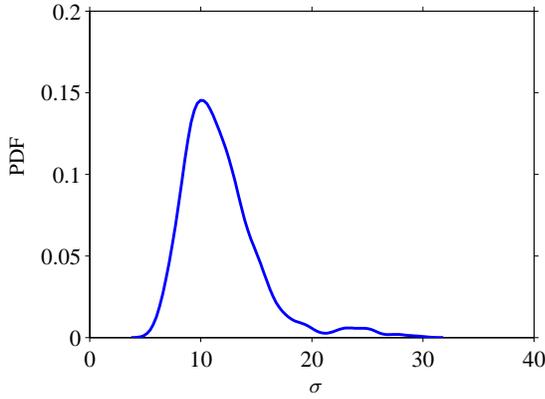
Figure 2. PDF of σ

Note that the RHS of Eq. 3 is not conditioned on $\boldsymbol{P}$ anymore. Some researchers refer to this PDF $f'_R(r)$ as the predictive PDF (Kiureghian, 1989) of R. The predictive PDF for the above numerical example is shown in Fig. 3.
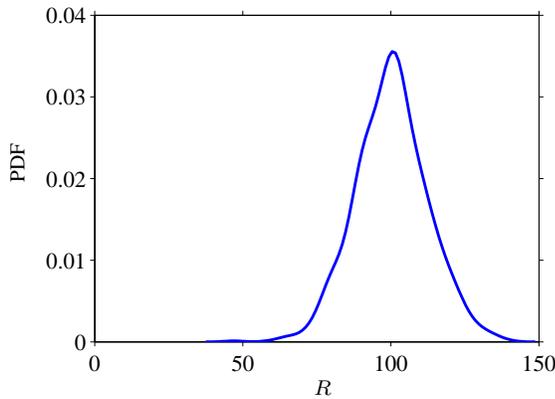
Figure 3. Predictive PDF of R

Note that the predictive PDF $f'_R(r)$ will indicate the presence of larger uncertainty in R than the original PDF $f_R(r)$, because the former accounts for the lack of infinite data. As the number of data increases, $f'_R(r)$ will tend towards $f_R(r)$. Of course, this is true only when the correct distribution type was assumed for R; in many cases, the choice of distribution type (referred to as "statistical model" by some researchers) is a challenge by itself, and contributes to additional uncertainty (Sankararaman & Mahadevan, 2013a).

4. CONDITION-BASED HEALTH MANAGEMENT

Most of the discussion pertaining to testing-based prognostics is not applicable to condition-based monitoring and prognostics. The distinctive feature of condition-based monitoring is that each component/subsystem/system is considered by itself, and therefore, "variability across specimens" is non-existent. Any such "variability" is spurious and must not be considered. At any generic time instant t_P at which prognos-

tics needs to be performed, the component/subsystem/system is at a specific state. The actual state of the system is purely deterministic, i.e., the true value of each state is completely precise, however unknown. Therefore, if a probability distribution is assigned for this state, then this distribution is simply reflective of the analyst's knowledge regarding this state and cannot be interpreted from a frequentist point of view. Thus, by virtue of definition of condition-based monitoring, physical probabilities are not present here, and a subjective (Bayesian) approach is only suitable for uncertainty quantification.

The goal in condition-based prognostics is, at any generic time instant t_P, to predict the remaining useful life of the component/subsystem/system as condition-based estimate of the usage time left until failure. Such computation needs to be, ideally, performed in real-time. In other words, the performance of the system during its operation needs to be analyzed, possible failure modes and future degradation needs to be prediction, and the remaining useful life needs to be computed while the system is under operation. These calculations help in operational decision-making activities such as path planning, mission routing, etc.

The following prognostics architecture can be used to achieve these goals. First, measurements until time t_P are used to estimate the state at time t_P. Then, using a degradation-prediction model (that may be model-based or data-driven), future state values (corresponding to time instants greater than t_P) are computed, and the first time time instant at which a failure threshold is true is calculated; this information is then used to calculate the remaining useful life. In order to forecast future state values, it is also necessary to assume future loading conditions (and operating conditions), and this is a major challenge in condition-based prognostics. Typically, the analyst subjectively assumes statistics for future loading conditions based on past experience and existing knowledge; thus, the subjective interpretation of uncertainty is clearly consistent across the entire condition-based monitoring procedure, and therefore, inferences made out of condition-based monitoring also need to be interpreted subjectively. The prediction of degradation (forecasting of future state values) is stopped when failure is reached, as indicated by a boolean threshold function that checks whether failure has occurred or not. This indicates the end-of-life (EOL) and the EOL can be directly used to compute the remaining useful life (RUL) prediction. Note that it is important to interpret the uncertainty in EOL and RUL subjectively.

4.1. Illustrative Example

Consider a generic engineering component whose health state at any time instant is given by $x(t)$. Consider a simple degradation model, where the rate of degradation of the health state (that decreases with time, due to the presence of damage) is

proportional to the current health state. This can be mathematically expressed as:

$$\dot{x}(t) \propto x(t), \qquad (4)$$

where the constant of proportionality is a negative number. Since differential equations are usually solved by considering discrete time instants, the above equation can be rewritten as:

$$x(k+1) = a.x(k) + b, \qquad (5)$$

where k represents the discretized time-index. The condition that "the constant of proportionality in Eq. 4 is negative" is equivalent to the condition that "$a < 1$ in Eq. 5". For the sake of illustration, let a denote the loading on the system, b denote the model of the degradation model above, and let a and b be constant and time-invariant. In practical examples, more than one variable may be necessary to represent the loading conditions and there may be multiple model parameters and state variables; further, the loading variables and model parameters may also be time-varying, just like the state x.

In order to compute the remaining useful life, it is necessary to chose a threshold function that defines the occurrence of failure. Since $x(k)$ is a decreasing function, the threshold function will indicate that failure occurs when the state value x becomes smaller than a critical lower bound (l), and the first time instant at which this event occurs indicates the end of life, and this time instant can be used to calculate the RUL. Therefore, the remaining useful life (r, an instance of the random variable R) is equal to the smallest n such that $x(n) < l$. Therefore RUL can be calculated as

$$r = \inf\{n : x(n) < l\}, \qquad (6)$$

For a given value of $x(0)$ (or $x(t_P)$, where t_P denotes the time at which prediction needs to be performed), a, b, it possible to calculate the end-of-life and remaining useful life using the above set of equations. However, in practical conditions, all of these are uncertain. However, note that the uncertainty in $x(0)$, a, b are related only to the knowledge regarding this particular unit and not an ensemble of units; recall that an ensemble of nominally identical units was considered earlier in Section 3. The presence of these uncertainties leads to uncertainty in the RUL prediction. This leads to the obvious question: How to compute the uncertainty in RUL? Prior to answering this question, the next subsection lists the different sources of uncertainty in generic condition-based prognostic applications.

4.2. Sources of Uncertainty

Typically, researchers have classified the different sources of uncertainty into different categories in order to facilitate uncertainty quantification and management. While it has been customary to classify the different sources of uncertainty into

aleatory (arising due to physical variability) and epistemic (arising due to lack of knowledge), such a classification may not be suitable for prognostics in the context of condition-based monitoring and RUL prediction because, as mentioned earlier, "true variability"' is not present in condition-based monitoring. A completely different approach for classification, particularly applicable to condition-based monitoring, is proposed in this paper.

1. **Present uncertainty:** Prior to prognosis, it is important to be able to precisely estimate the condition/state of the component/system at the time at which RUL needs to be predicted. Typically, damage (or faults) are expressed in terms of states, and therefore, estimating the state is equivalent to estimating the extent of damage (or fault). This is related to state estimation and is commonly addressed using filtering. Output data (usually collected through sensors) is used to estimate the state and many filtering approaches (Kalman filtering, particle filtering, etc.) are able to provide an estimate of the uncertainty in the state. In the illustrative example, the state uncertainty is equal to the uncertainty associated with $x(0)$. Practically, it is possible to improve the estimate of the states and thereby reduce this uncertainty, by using better sensors and improved filtering approaches. It is important to understand that the system is at particular state at any time instant, and the aforementioned uncertainty simply describes the lack of knowledge regarding the "true" state of the system.

2. **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. the loading, operating, environmental, and usage conditions are not known precisely, and it is important to assess this uncertainty before performing prognosis. In the illustrative example, the future uncertainty is equal to the uncertainty regarding the loading value, i.e., a, from the time of prediction until the time of failure. If there is no uncertainty regarding the future, then there would be no uncertainty regarding the *true* remaining useful life of the engineering component/system. However, this true RUL needs to be estimated using a model; the usage of a model imparts additional uncertainty as explained below.

3. **Modeling uncertainty:** It is necessary to use a functional degradation model in order to predict future state behavior, i.e. model the response of the system to anticipated loading, environmental, operational, and usage conditions. Further, the end-of-life is also defined using a Boolean threshold functional model, that is used to indicate whether failure has occurred or not. These two models are jointly used to predict the RUL, and they may either be physics-based or data-driven. It may be practically impossible develop models that accurately predict the underlying reality. Modeling uncertainty represents

the difference between the predicted response and the true response (that can neither be known nor measured accurately), and comprises of several parts: model parameters, model form, and process noise. While it may be possible to quantify these terms until the time of prediction, it is challenging to know their values at future time instants. In the illustrative example, Eq. 5 represents the degradation model, $x(n) < l$ represents the Boolean threshold function that indicates failure, b is a model parameter, and the uncertainty in b corresponds to one aspect of modeling uncertainty. Another aspect is the choice of the "linear" form of the model in Eq. 5; the underlying physical phenomena may differ from this assumption.

4. **Prediction method uncertainty:** Even if all the above sources of uncertainty can be quantified accurately, it is necessary to quantify their combined effect on the RUL prediction, and thereby, quantify the overall uncertainty in the RUL prediction. It may not be possible to do this accurately and this leads to additional uncertainty. For example, when sampling-based approaches are used for prediction, the use of limited number of samples causes uncertainty regarding the estimated probability distribution.

4.3. Computing Uncertainty in RUL

The goal in condition-based prognostics is to meaningfully integrate the degradation equation along with the failure threshold equation, and account for the different sources of uncertainty in $x(0)$, a, and b, and thereby, estimate the uncertainty in the remaining useful life. For any given realization of x_0, a, and b, it is possible to compute the first time instant (indicates the end-of-life) at which the failure threshold criteria will be valid, i.e., calculate the smallest value of n at which $x(n) < l$. The challenge is to compute the combined effect of uncertainty in $x(0)$, a, and b on RUL, and estimate the probability distribution of RUL.

It can be easily demonstrated that the state value at any future time instant can be expressed as a function of the initial state $x(0)$, as:

$$x(n) = a^n.x(0) + \sum_{j=0}^{j=n-1} a^j b \qquad (7)$$

Note that that $x(n)$ is decreasing and failure happens when $x < l$. Therefore, the remaining useful life (r, an instance of the random variable R) is equal to the smallest n such that $x(n) < l$. Therefore RUL can be calculated as

$$r = \inf\{n : a^n.x(0) + \sum_{j=0}^{j=n-1} a^j b < l\}, \qquad (8)$$

Assuming that the chosen time-discretization level is infinites-

imally small, it is possible to directly estimate the RUL by solving the equation:

$$a^r.x(0) + \sum_{j=0}^{j=r-1} a^j.b = l. \qquad (9)$$

The above equation calculates the RUL (r) as a function of the initial state $x(0)$, a and b. Even if the only considered source of uncertainty is the state estimate $x(0)$ (that is, a and b are constants), RUL R follows a Gaussian distribution if and only if it is linearly dependent on $x(0)$. In other words, R follows a Gaussian distribution if and only if Eq. 9 can be rewritten as:

$$\alpha.r + \beta.x(0) + \gamma = 0 \qquad (10)$$

for some arbitrary values of α, β, and γ. If it were possible to estimate such values for α, β, and γ, the distribution of RUL can be obtained analytically.

In order to examine if this is possible, rewrite Eq. 9 as:

$$x(0) = \frac{1}{a^r}(l - \sum_{j=0}^{j=r-1} a^j.b) \qquad (11)$$

While $x(0)$ is completely on the left hand side of this equation, r appears not only as an exponent in the denominator but is also indicative of the number of terms in the summation on the right hand side of the above equation. Therefore, it is clear that the relationship between r and $x(0)$ is not linear. Therefore, even if the initial state ($x(0)$, a realization of $X(0)$) follows a Gaussian distribution, the RUL (r, a realization of R) does not follow a Gaussian distribution. Furthermore, it is not even possible to analytically estimate the distribution of RUL. Thus, it is clear that even for a simple problem consisting of linear state models, an extremely simple threshold function, and only one uncertain variable that is Gaussian, the calculation of the probability distribution of R is neither trivial nor straightforward.

Practical problems in the prognostics and health management domain may consist of:

1. Several non-Gaussian random variables that affect the RUL prediction,

2. A non-linear multi-dimensional state space model,

3. Uncertain future loading conditions

4. A complicated threshold function that may be defined in multi-dimensional space.

The fact that the distribution of RUL simply depends on quantities such as degradation model and model parameters, threshold function, state estimate, future loading conditions, etc., implies that it is technically inaccurate to artificially assign the probability distribution type (or any statistic such as the mean or variance) to RUL. It is important to understand that

RUL is a dependent quantity and that the probability distribution of RUL needs to be accurately estimated using computational approaches. It has been illustrated the the problem of computing the uncertainty in the RUL prediction can be posed as an uncertainty propagation problem (Sankararaman & Goebel, 2013b), and therefore, it may be helpful to investigate statistical uncertainty propagation techniques in order to accomplish this goal.

4.4. Uncertainty Propagation Methods

The most commonly used uncertainty propagation technique is Monte Carlo sampling (Caflisch, 1998), which is based on drawing random samples of independent quantities, and computing corresponding realizations of the dependent quantity (in this case, the RUL). For instance, in the conceptual example, if $x(0)$ follows a Gaussian distribution (with mean and standard deviation equal to 975 and 50 respectively), a follows a uniform distribution (with lower and upper bounds of 0.990 and 0.995), and b follows a uniform distribution (with lower and upper bounds of -0.005 and 0 respectively), then the RUL (defined by Eq. 6, where $l = 50$) can calculated as a probability distribution, using Monte Carlo sampling. Using unit discretization (i.e., the time interval between the kth and $(k+1)^t h$ instants is equal to one second) for solution, the resultant probability density function (PDF) is shown in Fig. 4. It is clear that this distribution is not a typical parametric distribution (such as normal, lognormal, etc.) and that is why rigorous uncertainty propagation methods are necessary to accurate estimate this PDF.
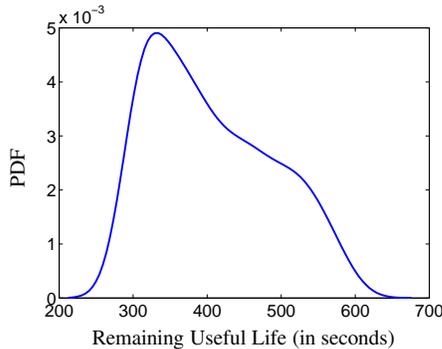

Figure 4. RUL: Conceptual Example

While Monte Carlo sampling can be accurate, it is computationally expensive and time-consuming, and therefore, researchers have focused on developing advanced methods that are computationally cheaper. These approaches include Latin hypercube sampling (Loh, 1996), adaptive sampling (Bucher, 1988), importance sampling (Glynn & Iglehart, 1989), unscented transform sampling (Van Zandt, 2001), etc. Alternatively, there are analytical methods such as the first-order second moment method (Dolinski, 1983), first-order reliability method (Hohenbichler & Rackwitz, 1983; Sankararaman & Goebel, 2013a), second-order reliability method (Der Ki-

ureghian, Lin, & Hwang, 1987), etc. In addition, there are also methods such as the efficient global reliability analysis (Bichon, Eldred, Swiler, Mahadevan, & McFarland, 2008) method which involve both sampling and the use of analytical techniques. All of these methods empirically calculate the probability distribution of RUL; while some of these methods calculate the PDF ($f_R(r)$) of RUL, some other methods calculate the CDF ($F_R(r)$), and some other methods directly generate samples from the desired probability density function ($f_R(r)$). Due to some limitations of each of these methods, it may not be possible to accurately calculate the actual probability distribution of R. Accurate calculation is possible only by using infinite samples for Monte Carlo sampling. Any other method (for example, the use of a limited, finite number of samples) will lead to uncertainty in the estimated probability distribution, and this additional uncertainty is referred to as prediction-method uncertainty. It is possible to decrease (and maybe eventually eliminate) this type of uncertainty either by using advanced probability techniques or powerful computing power.

It is necessary to further investigate the aforementioned uncertainty propagation methods, and identify whether they can be applied to prognostics health monitoring. Some earlier publications have investigated the use of certain methods such as Monte Carlo sampling, unscented transform sampling, first-order reliability methods, etc. in this regard.

5. CONCLUSION

This paper presented an overview of uncertainty quantification in prognostics and health management in engineering systems. First, the significance of the uncertainty in prognostics was explained, and the need for a systematic approach to account for uncertainty in prognostics was discussed. It was explained that four different activities — uncertainty representation and interpretation, uncertainty quantification, uncertainty propagation, and uncertainty management — need to be performed in order to rigorously include the effects of uncertainty in prognostics and provide useful information for decision-making under uncertainty. Researchers have pursued two different approaches for prognostics, and these two approaches are based on testing and condition-based assessment. The philosophical differences between these two approaches were explained and it was demonstrated that the concept of remaining useful life is more meaningful in the context of condition-based assessment since the engineering system is under operation. Further, these differences are used to analyze the interpretation of uncertainty in prognostics.

Probability and uncertainty can be interpreted in two ways. The frequentist interpretation of uncertainty is applicable in the presence of true randomness, as is the case in testing-based health management. The Bayesian (subjective) interpretation of uncertainty is applicable even while talking about

events that may not be random, and therefore, this interpretation is applicable for both testing-based health management and condition-based health management. In fact, only the Bayesian interpretation of uncertainty is applicable in condition-based health management. Techniques such as Kalman filtering, particle filtering, etc. that are commonly used in condition-based prognostics are collectively known as Bayesian tracking algorithms, not only because they use Bayes theorem but also because they are based on the subjective interpretation probability. Numerical examples were discussed in order to illustrate the effects of uncertainty interpretation on prognostics.

The final goal of this paper was to investigate methods for computation of remaining useful life, in the context of condition-based prognostics. It was illustrated that it is not possible to analytically calculate the uncertainty in the remaining useful life prediction even for certain simple problems involving Gaussian random variables and linear state-prediction models. Therefore, it is necessary to resort to computational methodologies for such uncertainty quantification and compute the probability distribution of remaining useful life prediction. While different types of uncertainty quantification methodologies were discussed, there are still several challenges that exist in this regard (Sankararaman & Goebel, 2014), and further research is necessary to investigate the applicability of these methods to prognostics and health monitoring applications.

ACKNOWLEDGMENT

REFERENCES

Bichon, B., Eldred, M., Swiler, L., Mahadevan, S., & McFarland, J. (2008). Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA journal*, *46*(10), 2459–2468.

Bucher, C. G. (1988). Adaptive samplingan iterative fast monte carlo procedure. *Structural Safety*, *5*(2), 119–126.

Caflisch, R. E. (1998). Monte carlo and quasi-monte carlo methods. *Acta numerica*, *1998*, 1–49.

Celaya, J., Saxena, A., & Goebel, K. (2012). Uncertainty representation and interpretation in model-based prognostics algorithms based on kalman filter estimation. In *Proceedings of the Annual Conference of the PHM Society* (pp. 23–27).

Celaya, J., Saxena, A., Kulkarni, C., Saha, S., & Goebel, K. (2012). Prognostics approach for power MOSFET under thermal-stress aging. In *Reliability and Maintainability Symposium (RAMS), 2012 Proceedings-Annual* (pp. 1–6).

Coppe, A., Haftka, R. T., Kim, N. H., & Yuan, F.-G. (2010). Uncertainty reduction of damage growth properties using structural health monitoring. *Journal of Aircraft*, *47*(6), 2030–2038.

Daigle, M., & Goebel, K. (2010). Model-based prognostics under limited sensing. In *Aerospace Conference, 2010 IEEE* (pp. 1–12).

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* (pp. 326–335).

Der Kiureghian, A., Lin, H.-Z., & Hwang, S.-J. (1987). Second-order reliability approximations. *Journal of Engineering Mechanics*, *113*(8), 1208–1225.

Dolinski, K. (1983). First-order second-moment approximation in reliability of structural systems: critical review and alternative approach. *Structural Safety*, *1*(3), 211–231.

Engel, S. J., Gilmartin, B. J., Bongort, K., & Hess, A. (2000). Prognostics, the real issues involved with predicting life remaining. In *Aerospace Conference Proceedings, 2000 IEEE* (Vol. 6, pp. 457–469).

Farrar, C. R., & Lieven, N. A. (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.

Glynn, P. W., & Iglehart, D. L. (1989). Importance sampling for stochastic simulations. *Management Science*, *35*(11), 1367–1392.

Gu, J., Barker, D., & Pecht, M. (2007). Uncertainty assessment of prognostics of electronics subject to random vibration. In *AAAI fall symposium on artificial intelligence for prognostics* (pp. 50–57).

Haldar, A., & Mahadevan, S. (2000). *Probability, reliability, and statistical methods in engineering design*. John Wiley & Sons, Incorporated.

Hastings, D. and McManus, H. (2004). A framework for understanding uncertainty and its mitigation and exploitation in complex systems. In *Engineering Systems Symposium MIT* (p. 19). Cambridge MA..

Hohenbichler, M., & Rackwitz, R. (1983). First-order concepts in system reliability. *Structural safety*, *1*(3), 177–188.

Kiureghian, A. D. (1989). Measures of structural safety under imperfect states of knowledge. *Journal of Structural Engineering*, *115*(5), 1119–1140.

Liao, H., Zhao, W., & Guo, H. (2006). Predicting remaining useful life of an individual unit using proportional haz-

ards model and logistic regression model. In *Reliability and Maintainability Symposium, 2006. RAMS'06. Annual* (pp. 127–132).

Loh, W.-L. (1996). On latin hypercube sampling. *The annals of statistics*, *24*(5), 2058–2080.

Orchard, M., Kacprzynski, G., Goebel, K., Saha, B., & Vachtsevanos, G. (2008, oct.). Advances in uncertainty representation and management for particle filtering applied to prognostics. In *Prognostics and Health Management, 2008. PHM 2008. International Conference on* (p. 1 -6). doi: 10.1109/PHM.2008.4711433

Saha, B., & Goebel, K. (2008). Uncertainty management for diagnostics and prognostics of batteries using Bayesian techniques. In *Aerospace Conference, 2008 IEEE* (pp. 1–8).

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). A novel computational methodology for uncertainty quantification in prognostics using the most probable point concept. In *Annual conference of the prognostics and health management society.*

Sankararaman, S., & Goebel, K. (2013b). Remaining useful life estimation in prognosis: An uncertainty propagation problem. In *2013 aiaa infotech@ aerospace conference.*

Sankararaman, S., & Goebel, K. (2013c). Why is the remaining useful life prediction uncertain? In *Annual conference of the prognostics and health management society.*

Sankararaman, S., & Goebel, K. (2014). Uncertainty in prognostics: Computational methods and practical challenges. In *Aerospace Conference, 2014 IEEE* (pp. 1–9).

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), 15 pages.

Sankararaman, S., & Mahadevan, S. (2011). Likelihood-based representation of epistemic uncertainty due to sparse point data and/or interval data. *Reliability Engineering & System Safety*, *96*(7), 814–824.

Sankararaman, S., & Mahadevan, S. (2013a). Distribution type uncertainty due to sparse and imprecise data. *Mechanical Systems and Signal Processing*, *37*(1), 182–198.

Sankararaman, S., & Mahadevan, S. (2013b). Separating the contributions of variability and parameter uncertainty in probability distributions. *Reliability Engineering & System Safety*, *112*, 187–199.

Tang, L., Kacprzynski, G., Goebel, K., & Vachtsevanos, G. (2009, march). Methodologies for uncertainty management in prognostics. In *Aerospace conference, 2009 IEEE* (p. 1 -12).

Van Zandt, J. R. (2001). A more robust unscented transform. In *International symposium on optical science and technology* (pp. 371–380).

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, decision-making 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 (AIAA), the Probabilistic Methods Technical Committee (PMC) at the American Society of Civil Engineers (ASCE), the Institute of Electrical and Electronics Engineers (IEEE), 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 Center 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.