Evaluating Prognostics Performance for Algorithms Incorporating Uncertainty Estimates

Abhinav Saxena
Stinger Ghaffarian Technologies, Inc.
650-604-3208
abhinav.saxena@nasa.gov

José Celaya
Stinger Ghaffarian Technologies, Inc.
650-604-4596
jose.r.celaya@nasa.gov

Bhaskar Saha
Mission Critical Technologies, Inc.
650-604-4379
bhaskar.saha@nasa.gov

Sankalita Saha
Mission Critical Technologies, Inc.
650-604-4593
sankalita.saha-1@nasa.gov

Kai Goebel
NASA Ames Research Center
650-604-4204
kai.goebel@nasa.gov

Prognostics Center of Excellence
NASA Ames Research Center, Moffett Field CA 94035

Abstract-Uncertainty Representation and Management (URM) are an integral part of the prognostic system development. As capabilities of prediction algorithms evolve, research in developing newer and more competent methods for URM is gaining momentum. Beyond initial concepts, more sophisticated prediction distributions are obtained that are not limited to assumptions of Normality and unimodal characteristics. Most prediction algorithms yield non-parametric distributions that are then approximated as known ones for analytical simplicity, especially for performance assessment methods. Although applying the prognostic metrics introduced earlier with their simple definitions has proven useful, a lot of information about the distributions gets thrown away. In this paper, several techniques have been suggested for incorporating information available from Remaining Useful Life (RUL) distributions, while applying the prognostic performance metrics. These approaches offer a convenient and intuitive visualization of algorithm performance with respect to metrics like prediction horizon and α-λ performance, and also quantify the corresponding performance while incorporating the uncertainty information. A variety of options have been shortlisted that could be employed depending on whether the distributions can be approximated to some known form or cannot be parameterized. This paper presents a qualitative analysis on how and when these techniques should be used along with a quantitative comparison on a real application scenario. A particle filter based prognostic framework has been chosen as the candidate algorithm on which to evaluate the performance metrics due to its unique advantages in uncertainty management and flexibility in accommodating non-linear models and non-Gaussian noise. We investigate how performance estimates get affected by choosing different options of integrating the uncertainty estimates. This allows us to identify the advantages and limitations of these techniques and their applicability towards a standardized performance evaluation method.

TABLE OF CONTENTS

1. INTRODUCTION

As prognostics matures and moves towards real applications, incorporating various sources of uncertainty into remaining life predictions becomes naturally essential for dependable prognostic health management. Existing concepts in the literature merely provide confidence bounds around the predictions, and more often than not stick to precision measures that are derived from statistics over assumed Gaussian distributions. Consequently, there is a significant push towards developing algorithms that generate more realistic estimates of uncertainty without such strong assumptions. Prognostic performance evaluation metrics (like prediction horizon, α-λ performance, convergence, etc.), introduced in our earlier work, were developed based on the assumption that a point estimate of the predictions can be generated from all algorithms. This was done to develop new concepts for performance evaluation tailored for prognostics.

Prognostics being an emerging research field, most of the published work has naturally been exploratory in nature, consisting mainly of proof-of-concepts and one-off applications. Prognostic Health Management (PHM) has by-

[1]978-1-4244-3888-4/10/$25.00 ©2010 IEEE.
[2] IEEEAC paper #1372, Version 2, Updated December 28, 2009

and-large been accepted by the engineered systems community in general, and the aerospace industry in particular, as the direction of the future. However, for this field to mature, it must make a convincing case in numbers to the decision makers in research and development as well as fielded applications. It is as Prof. Thomas Malone, an eminent management guru, said, “If you don’t keep score, you are only practicing”[1].

2. BACKGROUND

In the past one year the prognostics center of excellence at NASA Ames has closely followed the developments in the area of prognostics performance evaluation. In addition to analyzing performance evaluation methods prevalent in the PHM community, methods from other domains were also considered. A comprehensive study that classified various forecasting applications and compiled a list of a variety of metrics set the stepping stone for this work. It was realized that conventional metrics do not necessarily answer all the questions that a prognostic system needs to resolve from the performance evaluation point of view and hence a number of possible ideas were introduced [2]. These metrics were then implemented and applied to battery aging datasets made available by a comprehensive study conducted at the Idaho National Labs. It was shown that the new metrics cover the intended aspects of prognostic performance and that the conventional metrics fall short in their current form [3]. Four different algorithms were compared and it was discovered that 1. No single algorithm was reportedly the best, 2. These metrics did not resolve several ambiguous situations that were found to occur in the real data, 3. These metrics did not make use of the uncertainty information whenever available, 4. There did not exist any formal guideline on how to extract point estimates from non-parametric distributions. Several organizations were encouraged to use these metrics and provide feedback. Findings from all the sources echoed for the need of incorporating uncertainty estimates in assessing the performance of these algorithms. This led to a the work described in [4] where not only the suggestions to resolve ambiguous situations were provided but also several modifications to these metrics were introduced, specifically with respect to including uncertainty information. Guidelines were laid out to determine how to deal with and subsequently make approximations in cases where distributions resembled some known parametric distributions and also the cases where no parametric form could be identified. This paper extends those concepts by applying them to another set of data posing battery End of Discharge (EoD) prediction problem. The current state-of-the-art prognostics algorithm, namely the Particle Filter, was chosen to generate probabilistic distributions of the Remaining Useful Life (RUL) estimates before a battery loses its charge beyond a pre specified threshold. These results are then analyzed following all possible recommendations and discussed to highlight the differences that result in making different assumptions about the uncertainty distributions.

We first very briefly describe the prognostic performance metrics and the thought process behind them. Then in the next section we expand on the approaches taken for uncertainty management and what may be needed from metrics to incorporate those ideas. The discussions in the subsequent sections would detail the enhancements proposed and illustrate through the chosen case study.

Prognostic Performance Metrics

Traditionally performance metrics are viewed as means of algorithmic performance evaluation. However, it can be argued that algorithmic performance must be guided by top level requirements that are generated from system’s mission objectives. Therefore, as shown in Figure 1, metrics play an important role in connecting the high level system requirements to the low level algorithmic performance.

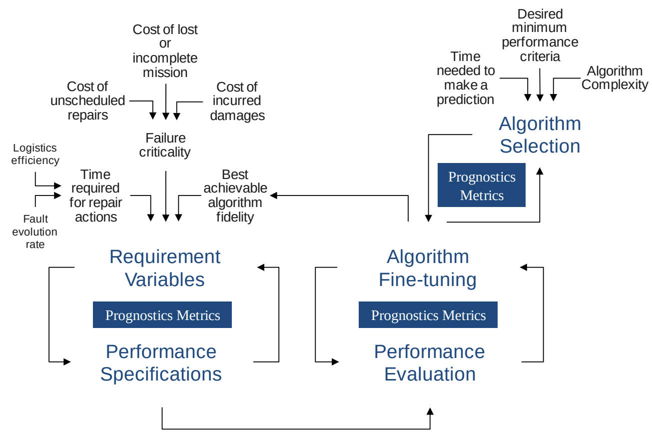


Figure 1: Prognostics metrics facilitate performance evaluation and also help in requirements specification

The new prognostics metrics developed in previous work require a change in thinking about what constitutes a good performance. More importantly the time varying aspect of performance, each time the estimates are updated, differentiates these metrics from other related domains. These metrics offer visual as well as quantitative assessment of performance as it evolves over time. The visual representation allows making several observations about the performance and it is necessary for us, now, to understand the capabilities and the limits of information these new metrics can provide.

It must be noted that these metrics are intended for offline evaluation of prognostics and are not applicable for online cases directly. Prognostic performance evaluation is an acausal problem that requires inputs from the events that are expected to take place in the future. Specifically, one needs to know the true End of Life (EoL) of the system to evaluate prediction accuracy. Online evaluation will have to incorporate methods to deal with uncertainties associated with future operating conditions in particular. This will require significant advancements in uncertainty representation, quantification and management methods, and hence remains a subject for the future work.

3. UNCERTAINTIES IN PROGNOSTICS

Accounting for various uncertainties is of key importance in prognostics. A good prognostics system not only provides accurate and precise estimates for the RUL predictions but also specifies the level of confidence associated with such predictions. Without such information any prognostic estimate is of limited use and cannot be incorporated in mission critical applications [5]. Uncertainties arise from various sources in a PHM system [6-8]. Some of these sources include modeling uncertainties (modeling errors in both system model and fault propagation model), measurement uncertainties (arise from sensor noise, ability of sensor to detect and disambiguate between various fault modes, loss of information due to data preprocessing, approximations and simplifications), operating environment uncertainties, future load uncertainties (arising from unforeseen future and variability in usage history data), input data uncertainties (estimate of initial state of the system, variability in material properties, manufacturing variability), etc. It is often very difficult to assess the levels and characteristics of uncertainties arising from each of these sources. Further, it is even more difficult to assess how these uncertainties that are introduced at different stages of the prognostic process combine and propagate through the system, which in most likelihood has a complex non-linear dynamics. This problem further deepens if the statistical properties do not follow any known parametric distributions thereby eliminating any scope of analytical solutions.

Owing to all of these challenges Uncertainty Representation and Management has become an active area of research in the field of PHM. A conscious effort in this direction is clearly evident from recent developments in prognostics in the past few years [8-12]. These developments must be adequately supported by suitable methods for performance evaluation that can incorporate various expressions of uncertainties in the prognostic outputs.

Although several approaches for uncertainty representation have been explored by researchers in this area, the most popular approach has been probabilistic representation. A well founded Bayesian framework has led to many analytical approaches that have shown promise [13-15]. In these cases a prediction is represented by a corresponding probability density function. In most cases, for the sake of simplicity, an assumption is made about the form of distribution. Our experience however, shows that this is hardly the case ever. In such cases these distributions are non-parametric and are represented by sampled outputs. This paper presents an adaptation of prognostic performance metrics to incorporate all these cases irrespective of their distribution characteristics. In the next section we describe the notions of performance evaluation for prognostics and then extend these ideas to incorporate uncertainty estimates.

4. PROGNOSTIC PERFORMANCE EVALUATION

In our previous works five prognostic metrics were proposed that have been applied to several applications since then and subsequently refined based on feedback. These metrics include Prognostic Horizon (PH), α-λ Performance, Relative Accuracy (RA), Cumulative Relative Accuracy (CRA), and Convergence that can be used for offline performance evaluation of the prognostic performance. To illustrate the concepts while keeping the discussion concise and to the point in thus paper we will expand only two of these metrics namely Prognostic Horizon and α-λ Accuracy. As discussed in [4] these concepts can be easily incorporated in the rest of the metrics. We first discuss the basic definitions of these two metrics.

Prognostic Horizon

Prognostic Horizon is defined as the difference between the time index i when the predictions first meet the specified performance criteria (based on data accumulated until time index i) and EoL. The performance requirement may be specified in terms of allowable error bound (α) around true EoL. The choice of α depends on the estimate of time required to take a corrective action. Depending on the situation this corrective action may correspond to performing maintenance (manufacturing plants) or bringing the system to a safe operating mode (operations in a combat zone).

$$PH = EoL - i \qquad (1)$$

where:

$i = \min\left\{ j \mid \left(j \in \ell \right) \wedge \left(r_* - EoL * \alpha \right) \le r^l(j) \le r_* + EoL * \alpha \right) \right\}$ is the first time index when predictions satisfy α-bounds

ℓ is the set of all time indexes when a prediction is made

l is the index for l^{th} unit under test (UUT)

r_* is the ground truth RUL

$r(j)$ is the predicted RUL at time j

EoL is the ground truth End-of-Life (actual failure)

As shown in Figure 2, the desired level of accuracy with respect to the EoL ground truth is specified as $\pm\alpha$-bounds. RUL values are then plotted against time for various algorithms that are being compared. The PH for an algorithm is declared as soon the corresponding predictions enter the band of desired accuracy. As clearly evident from the illustration, the first algorithm has a longer PH.

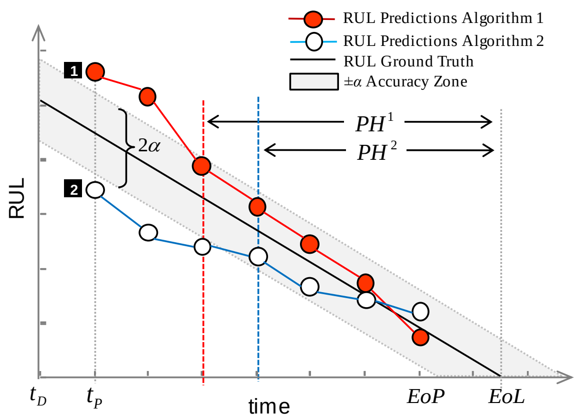


Figure 2: Prognostic Horizon

Prognostic horizon produces a score that depends on length of ailing life of a system and the time scales in the problem at hand. The range of PH is between (t_{EoL}-t_P) and $max\{0, t_{EoL}$-$t_{EoP}\}$. The best score for PH is obtained when an algorithm always predicts within desired accuracy zone and the worst score when it never predicts within the accuracy zone.

α-λ Accuracy

α-λ Accuracy quantifies prediction accuracy by determining whether the prediction falls within specified limits at particular times specified by parameter λ (Figure 3). These time instances may be specified as percentage of total ailing life of the system. In our implementation of *α-λ accuracy* we seek answer to the question whether the prediction accuracy is within α*100% of the actual RUL at specific time instance t_λ, which is expressed as a fraction of time between the point when an algorithm starts predicting and the actual failure. For example, this metric determines whether a prediction falls within 10% accuracy (i.e., $\alpha = 0.1$) halfway to failure from the time the first prediction is made (i.e., $\lambda = 0.5$). Therefore, one needs to evaluate whether the following condition is met.

$$(1-\alpha)\cdot r_*(t) \le r^l(t_\lambda) \le (1+\alpha)\cdot r_*(t) \qquad (3)$$

where:

α is the accuracy modifier

λ is a time window modifier such that $t_\lambda = t_P + \lambda(EoL - t_P)$

The output of this metric is binary (True or False) stating whether the desired condition is met at a given particular time. This is a more stringent requirement as compared to prognostic horizon as it requires predictions to stay within a cone of accuracy i.e. the bounds that shrink as time passes by.

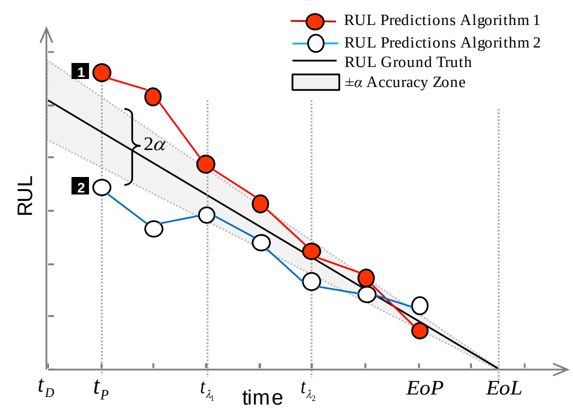


Figure 3: Concept of *α-λ* Accuracy

Improved Metrics to Handle Probabilistic Outputs

We now describe how these definitions may be modified to elegantly incorporate RUL distributions. First we tackle the visual enhancements in these metrics as these metrics convey important information visually through their corresponding RUL versus time plots. It is suggested that other measures of location and variance be used instead of mean and standard deviation, as for Normal cases. For cases were normality cannot be established, one can rely on median as a measure of location and the quartiles or inter quartile range as a measure of spread [16]. We broadly categorized various distributions into four categories (see Table 1) and suggested various options to choose from to compute more appropriate location and spread measures that can be, then, indicated in the metric plots. This subclassification mainly determines the method of computing the total probability, i.e. continuous integration or discrete summation and then how to represent it in the metrics plots. For instance, in cases that involve Normal distribution, including a confidence interval represented by an error bar around the point prediction is useful [17]. For cases with non-Normal single mode distributions this can be done with an inter-quartile plot represented by a box plot [18]. This conveys how a prediction distribution is skewed and whether these skew should be considered while declaring prognostic horizon. Box plot also has provisions to represent outliers that may be useful to keep track of in risk sensitive situations. It is suggested to use box plots along with a dot representing the mean of the distribution, which will allow keeping the visual information in perspective with respect to original plots. For mixture of Gaussians case, it is suggested that a model with few (preferably $n \le 4$) Gaussians is created and corresponding error bars plotted adjacent to each other. The weights for each Gaussian component can then be represented by the thickness of the error bars. We do not recommend multiple box plots in this case as there is no methodical way to differentiate between samples, assign them to particular Gaussian components, and compute the quartile ranges for each of them. Also, to keep things simple we assume a linear additive model while computing the mixture of Gaussians.

$$\phi(x) \cong \omega_1 \cdot N(\mu_1, \sigma_1) + ... + \omega_n \cdot N(\mu_n, \sigma_n);\ n \in I^+ \qquad (2)$$

where:

ω is the weight factor for each Gaussian component

$N(\mu, \sigma)$ is a Gaussian distribution with parameters μ and σ

These enhancements can be analytically incorporated into the numerical aspect of these metrics by computing total probability mass of a prediction falling within the specified α-bounds versus using a point estimate to compute the metric. This concept has been depicted in Figure 4 with original point prediction superimposed on box plots. In this manner a prediction is considered inside α-bounds only if the total probability mass of the corresponding distribution within the α-bounds is more than a predetermined threshold β. This parameter is also linked to the issues of uncertainty management and risk absorbing capacity of the system. In the most simple case we suggest using β = 0.5 that would correspond to making a decision based on the mean value for a Gaussian distribution case, the approach that we had been following with the original definition of PH using means as point estimates. Consequently now the definition of PH is modified by determining the index i in Eq.1 as,

$$i = \min\left\{ j \,|\, (j \in \ell) \wedge \left(\pi[r(j)]_{\alpha^-}^{\alpha^+} \geq \beta \right) \right\},$$

$\pi[r(j)]_{\alpha^-}^{\alpha^+} = \int_{\alpha^-}^{\alpha^+} \phi(x)dx;\quad x \in \Re^+$ is the total probability mass of the prediction pdf within the α-bounds that are given by $\alpha^+ = r_* + \alpha \cdot EoL$ and $\alpha^- = r_* - \alpha \cdot EoL$.

Computing the metric now requires integrating the probability distribution that overlaps with the desired region to compute the total probability. For cases where analytical form of the distribution is available, like for Normal distributions, it can be computed analytically by integrating the area under the prediction pdf between the α-bounds (α^- to α^+). However, for cases where there is no analytical form available, a summation based on histogram obtained from the process/algorithm can be used to compute total probability.

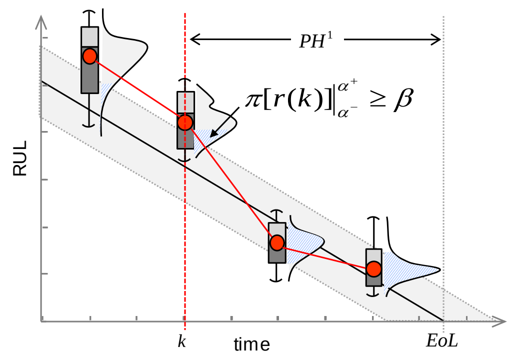
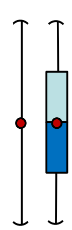
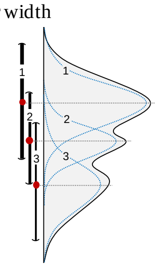
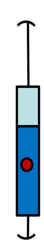


Figure 4: Enhanced representation for prognostic horizon incorporating distribution information

Table 1: Recipe to select location and spread measures along with visualization methods [4]

	Normal Distribution	**Mixture of Gaussians**	**Non-Normal Distribution**	**Multimodal (non-Normal)**
	Parametric		Non-Parametric	
Location (Central tendency)	Mean (μ)	Means: $\mu_1, \mu_2, \ldots, \mu_n$ weights: $\omega_1, \omega_2, \ldots, \omega_n$	Mean, Median, L-estimator, M-estimator	Dominant median, Multiple medians, L-estimator, M-estimator
Spread (variability)	Sample standard deviation (σ), IQR (inter quartile range)	Sample standard deviations: $\sigma_1, \sigma_2, \ldots, \sigma_n$	Mean Absolute Deviation (MAD), Median Absolute Deviation (MdAD), Bootstrap methods, IQR	
Visualization	Confidence Interval (CI), Box plot with mean 	Multiple CIs with varying bar width  Note: here $\omega_1 > \omega_2 > \omega_3$	Box plot with mean 	Box plot with mean 

5. BATTERY HEALTH MANAGEMENT EXAMPLE

To illustrate the concepts and methodology of the new metrics, data from battery health management domain have been used. As batteries age, their charge retention capacity depletes and hence the time taken to discharge a battery reduces. This change in charge retention capacity is an effect of a complex interplay between several dynamical processes as described in [15]. The paper further described how these effects can be modeled and then using a Particle filtering approach unknown parameters of the model can be learned. Particle filters have been shown to incorporate various essential characteristics that are key for a successful prognostics system. For instance their ability to incorporate non-linear fault propagation models and non-Gaussian noise have provided a general framework to represent uncertainties and propagate them to RUL predictions in near real time. Particle filters manage uncertainties though an importance re-sampling technique that does not let the uncertainty bounds grow while making long term predictions and provides a non-parametric distribution of the predicted values that can then be processed and used as desired. Further details on the mechanics of particle filters can be found in [14, 19]. This paper uses the results from [15] and extends their performance assessment that was based on point estimates to using probabilistic estimates as facilitated by the enhancements. A brief description of the battery aging data and corresponding prognostic problem is described next.

Battery Health Management Problem

The data have been collected from a custom built battery prognostics testbed at the Prognostics Center of Excellence (PCoE) at NASA Ames Research Center. Commercially available Li-ion 18650 sized rechargeable batteries were chosen as test article. In this testbed Li-ion batteries were run through 3 different operational profiles (charge, discharge and EIS) at room temperature, 23ºC. Charging was carried out in a constant current (CC) mode at 1.5 A until the battery voltage reached 4.2 V and then continued in a constant voltage (CV) mode until the charge current dropped to 20 mA. Discharge was carried out at a constant current (CC) level of 2 A until the battery voltage fell to 2.7 V. Repeated charge and discharge cycles result in accelerated aging of the batteries. The experiments were stopped when the batteries reached the EoL criteria of 30% fade in rated capacity (from 2 Ah to 1.4 Ah). Due to the differences in depth-of-discharge (DoD), the duration of rest periods and intrinsic variability, no two cells had the same State of Life (SoL) at the same cycle index. The aim was to be able to manage this uncertainty, which is representative of actual usage, and make reliable predictions of RUL in both the EoD and EoL contexts. In this study we present the results from a single battery from those experiments. For reference Figure 5 shows the EoD predictions generated by the PF algorithm for an arbitrarily selected discharge cycle, as reported in [15]. The red solid line shows the measured cell voltage, while the green patch represents the envelope of the PF tracking performance.

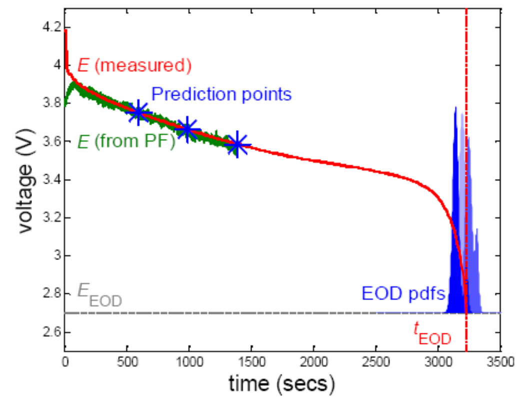


Figure 5: Prediction results from Particle Filter algorithm for the battery dataset [15]

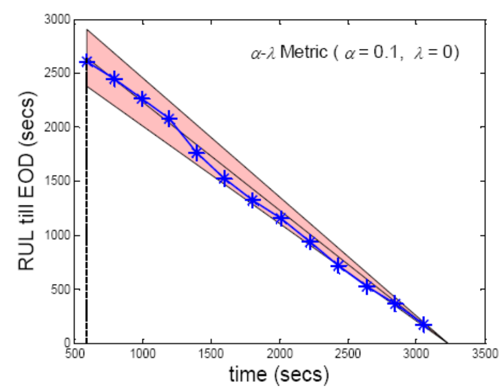


Figure 6: α-λ Accuracy as reported in [15] was based on sample mean as a location measure

Further, as shown in Figure 6, a point estimate was generated from the corresponding probability distribution and then used to show prediction performance in the α-λ accuracy plot. Distribution sample mean was treated as the point estimate with an implicit assumption of the Normality in the distribution. In this paper we take a step further and again present those prediction results in their entirety considering the enhancements described above. For reference, these distributions are plotted here in a RUL versus time plot (Figure 7). In total thirteen predictions were made starting around 591 seconds. Several observations can be made here. It is inconclusive looking at these distributions that they follow Gaussian characteristics. Some distributions show presence of outliers. In a risk sensitive uncertainty management context outliers represent low probability events that may not necessarily be low risk [14]. Taking a sample mean including outliers does not indicate the true mean of the majority of the distribution and otherwise, excluding them may enhance the risk of not catching the low probability events. Furthermore, as expected distributions become narrower as more time passes by and additional information becomes available.

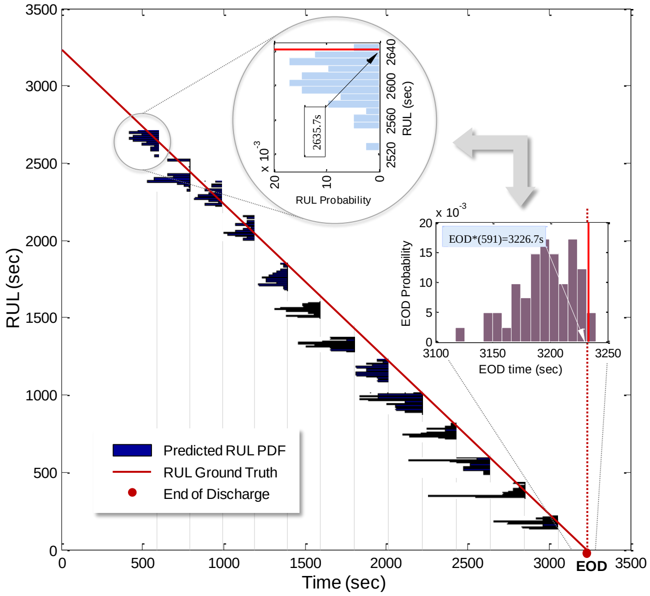


Figure 7: Predicted distributions plotted on the RUL versus time plot show evolution of probability histograms for successive predictions

Another interesting point to note in this example is that compared to the time range in this experiment the variability in the predictions is quite small. While here this is the case due to a good modeling approach in a relatively simple example of predicting end of battery discharge. In other applications similar situation may occur if the lifespan of the system is very large and aging process slow. In such cases it becomes very difficult to show features of interest such as distribution and variability of an individual prediction within a RUL vs. time plot. Therefore, as shown in the Figure 7, EoL distributions can be equivalently used instead of RUL distributions for the purpose of performance visualization in the same framework. This is possible since EoL distribution has a one to one correspondence to its RUL distribution that is shifted by t_p, the time of prediction. Therefore, wherever required we will show results in the context of EoD for the battery example to better illustrate finer details of these distributions and how they are incorporated in visualization of these metrics.

Application of Prognostic Metrics

Following various steps shown in Figure 10 in [4], in this section we will show how assuming different forms for the RUL distribution results in different outcomes for the metrics. Also, as mentioned earlier for the sake of clarity in the plots we will use EoD versus time plots and only show six out of all thirteen predictions wherever necessary.

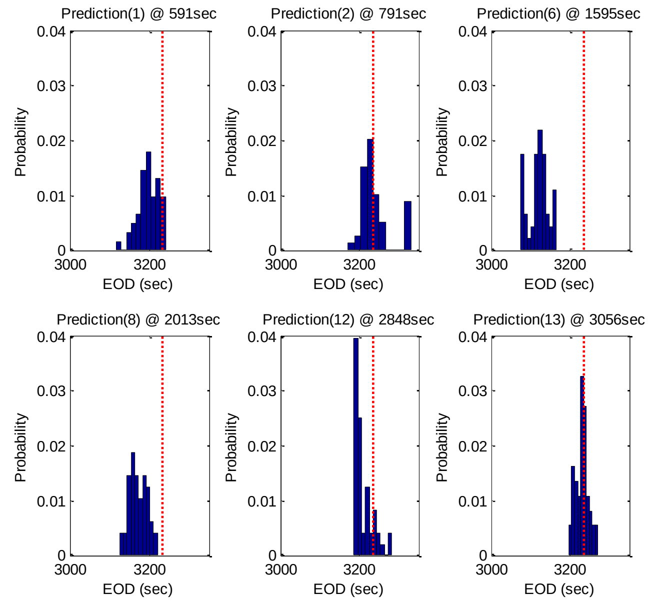


Figure 8: Original normalized histograms for six predictions

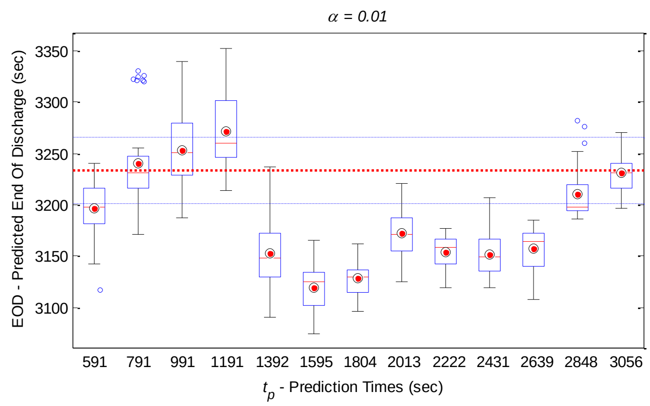


Figure 9: EoL vs. time plot with RUL box plot and sample mean.

We first approximate these histograms as unimodal Gaussian and compute corresponding statistics. Figure 10 shows approximated probability density functions (PDFs) superimposed on the obtained histograms. As can be clearly seen, these distributions do not reflect the true statistics very well. Specifically for the second prediction, presence of outliers biases the mean. The red vertical line indicates the true end of discharge from the experiments. While this figure is for visual reference, Figure 11 shows box plots as intended for the prognostics metrics. Sample mean (indicated by white dot) can be compared with the median (white line in the box). Skew within individual predictions is easily visible in these plots. It must be noted that a unimodal assumption often leads to ignoring outliers that may not be a preferred way to handle uncertainty information in high risk low probability situations.

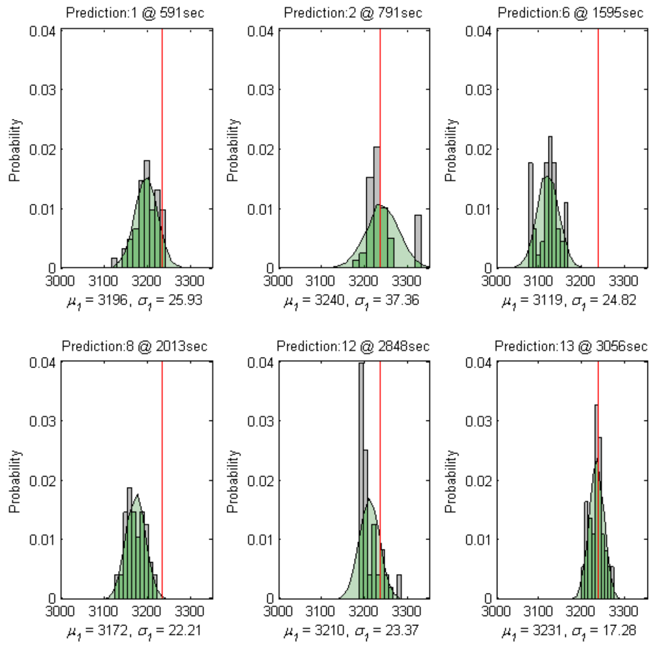


Figure 10: RUL probability density functions if approximated as unimodal Gaussian

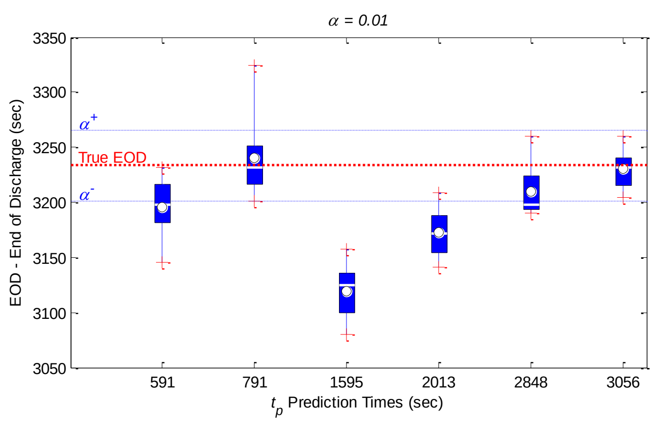


Figure 11: Box plots including sample means show deviations from Gaussian characteristics

Now we approximate these results with a bimodal Gaussian mixture and show corresponding fit in Figure 12. Corresponding means and standard deviations are also indicated in the figure. These modes are generated and correspondingly weighted using an optimization routine. Figure 13 represents these modes using two error bounds corresponding to each mode and their thickness is representative of their weights. The information presented in this plot is far more clear and suitable for the performance assessment purposes. It also shows which modes contribute to good PH performance. Outliers in these cases are treated as separate modes and may be easier to keep track of. Relative presence of different modes, as indicated by width of the error bars, provides a sense of priority while interpreting results and expressing prognostic confidence.

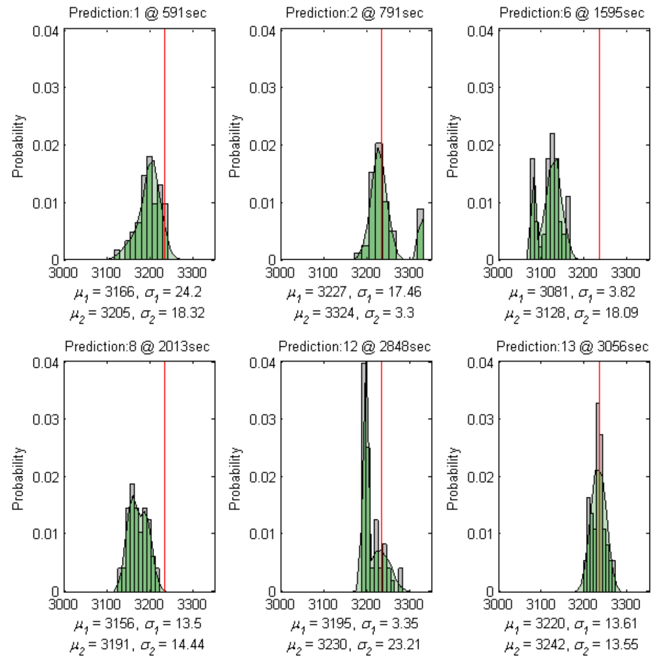


Figure 12: RUL predictions when approximated with bimodal Gaussian. Corresponding moments of individual modes are indicated

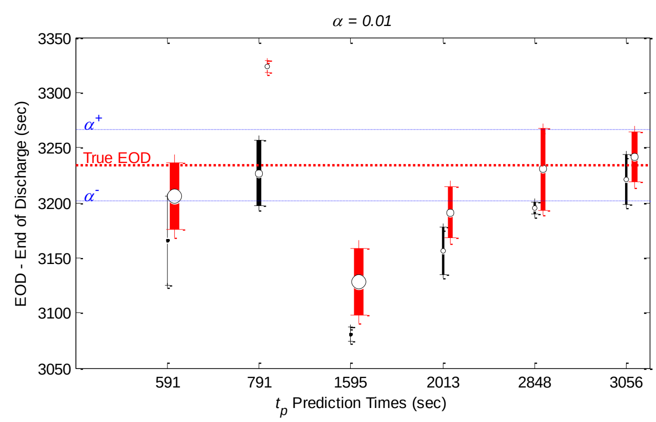


Figure 13: Two confidence intervals per prediction are used to show statistics for bimodal assumption. Width of each bar is proportional to the weight of corresponding mode in the mixture

In a similar fashion, Figure 14 and Figure 15 show a trimodal assumption case. In our experience approximating a distribution with more than 3 or 4 modes tends to overfit and does not provide any useful information. Hence, we do not recommend using multiple modes unless such indicators are available from the process itself. it must be kept in mind that breaking a distribution into several modes just helps analyze a more complex distribution but may not be a natural decomposition based on process characteristics. It may help keep track of outliers but also runs the risk of identifying features that do not really exist in the process.

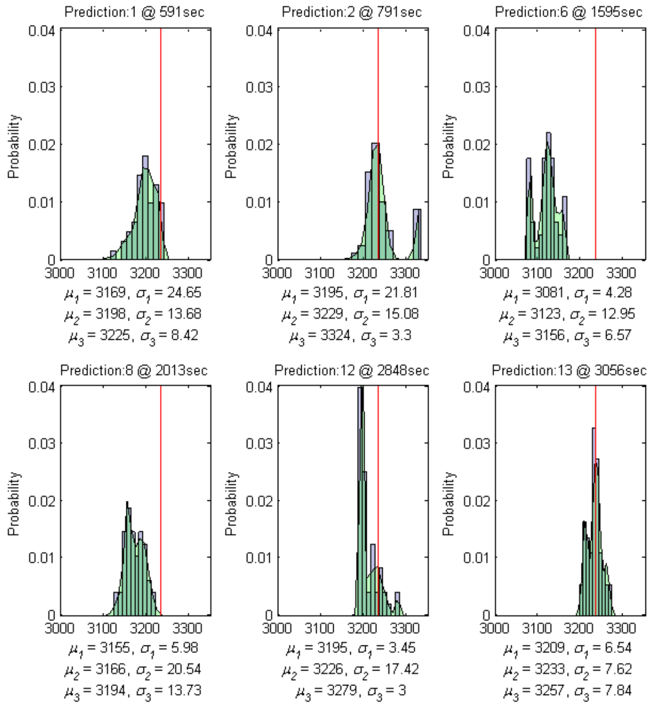


Figure 14: Trimodal approximation for RUL predictions

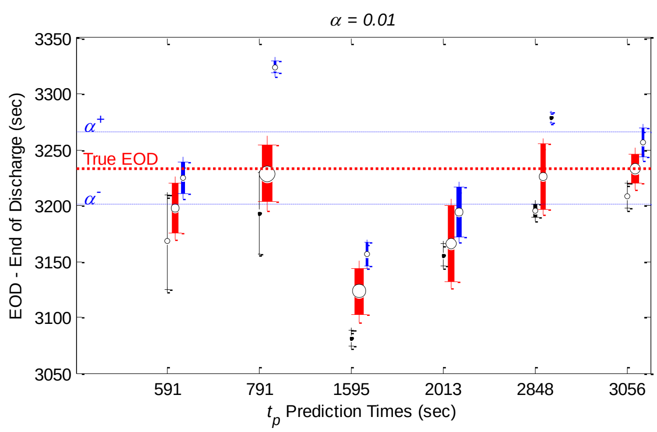


Figure 15: Three bars represent the statistics of each of the three modes in the Gaussian mixture

6. Conclusions & Future Work

In this study we have shown how prognostics metrics can be modified to incorporate uncertainty information available from various algorithms. A Particle Filter based prediction algorithm was used to generate RUL predictions. These results were then analyzed using the enhanced performance metrics. It was shown that an arbitrary assumption of PDF into Gaussian may not necessarily be the best choice. Specifically for safety critical applications, where the risks arising from uncertainties need to be managed carefully for decision making, a more detailed characterization of probability distributions may be desired. The new metrics facilitate such characterization in a conducive manner such that they can be easily implemented in an automated fashion without loss of intuitiveness. Furthermore, computing probability mass presents a more robust methodology for using prognostic metrics as far as outliers or non-normal distributions is concerned. A suitable way of graphically representing these metrics is also shown to effectively convey distribution characteristics vis-à-vis prognostics metrics plots and corresponding desired error bounds.

While these metrics are suitable for offline performance evaluation of prognostics algorithms future work will incorporate methods to accommodate online performance tracking. So far, the performance evaluation assumes that future loading conditions do not change or at least do not change the rate of fault growth. For offline studies this may be reasonable as we know the actual EoL index and can linearly extrapolate true RUL for all previous time indices to draw a straight line. However, for real-time applications this would not hold true as changes in operating conditions do affect the rate of fault evolution. Hence, we would also like to investigate how to incorporate effects of changes in the loading conditions that alter the RUL slope by changing the rate of remaining life consumption. Similar description will also support cases where maintenance actions prolong the lives of the system or the systems with self-healing characteristics. Furthermore, we would like to extend this study and connect high level requirements to low level performance specification from a post-prognostic reasoning point of view.

Acknowledgements

The authors would like to express their gratitude to colleagues at the Prognostic Center of Excellence (NASA Ames Research Center) and external partners for participating in research discussions, evaluating these metrics in their respective applications, and providing a valuable feedback. This work was funded by the Integrated Vehicle Health Management (IVHM) Project under NASA Aviation Safety Program.

REFERENCES

[1] S. A. Melnyk, "Metrics – The Missing Piece in Operations Management Research," in *Decision Line – Research Issues*. vol. 30, 1999.

[2] A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for Evaluating Performance of Prognostic Techniques," in *Prognostics and Health Management (PHM08), International Conference on* Denver CO, 2008, pp. 1-17.

[3] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "Evaluating Algorithmic Performance Metrics Tailored for Prognostics," in *IEEE Aerospace Conference*, Big Sky, MT, 2009.

[4] A. Saxena, J. Celaya, B. Saha, S. Saha, and K. Goebel, "On Applying the Prognostics Performance Metrics," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 16.

[5] S. Uckun, K. Goebel, and P. J. F. Lucas, "Standardizing Research Methods for Prognostics," in *International Conference on Prognostics and Health Management (PHM08)*. Denver, CO, 2008, p. 10.

[6] A. Coppe, R. T. Haftka, N. Kim, and F. Yuan, "Reducing Uncertainty in Damage Growth Properties by Structural Health Monitoring," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 10.

[7] D. Hastings and H. McManus, "A Framework for Understanding Uncertainty and its Mitigation and Exploitation in Complex Systems," in *Engineering Systems Symposium* MIT, Cambridge MA, 2004, p. 19.

[8] M. E. Orchard, G. Kacprzynski, K. Goebel, B. Saha, and G. Vachtsevanos, "Advances in Uncertainty Representation and Management for Particle Filtering Applied to Prognostics," in *Applications of Intelligent Control to Engineering Systems*, K. P. Valavanis, Ed. Netherlands: Springer, 2009, pp. 23-35.

[9] R. deNeufville, "Uncertainty Management for Engineering Systems Planning and Design," in *Engineering Systems Symposium* MIT, Cambridge, MA, 2004.

[10] K.-C. Ng and B. Abramson, "Uncertainty Management in Expert Systems," *IEEE Expert Systems,* vol. 5, p. 20, 1990.

[11] S. Sankararaman, Y. Ling, C. Shantz, and S. Mahadevan, "Uncertainty Quantification in Fatigue Damage Prognosis," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 13.

[12] L. Tang, G. J. Kacprzynski, K. Goebel, and G. Vachtsevanos, "Methodologies for Uncertainty Management in Prognostics," in *IEEE Aerospace Conference* Big Sky, MT, 2009, p. 12.

[13] X. Guan, Y. Liu, A. Saxena, J. Celaya, and K. Goebel, "Entropy-based Probabilistic Fatigue Damage Prognosis and Algorithmic Performance Comparison," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 11.

[14] M. E. Orchard, L. Tang, K. Goebel, and G. Vachtsevanos, "A Novel RSPF Approach to Prediction of High-Risk, Low-Probability Failure Events," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 9.

[15] B. Saha and K. Goebel, "Modeling Li-ion Battery Capacity Depletion in a Particle Filtering Framework," in *Annual Conference of the Prognostics and Health Management Society (PHM09)* San Diego, CA, 2009, p. 10.

[16] D. C. Hoaglin, F. Mosteller, and J. W. Tukey, *Understanding Robust and Exploratory Data Analysis*: Wiley, 1983.

[17] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 6th ed.: Thomson, 2004.

[18] A. R. Martinez, "Exploratory Data Analysis with MATLAB," A. R. Martinez, Ed.: CRC Press, 2004.

[19] M. E. Orchard and G. J. Vachtsevanos, "A Particle-Filtering Approach for On-line Fault Diagnosis and Failure Prognosis," *Transactions of the Institute of Measurement and Control,* vol. 31, pp. 221-246, 2009.

BIOGRAPHIES

Abhinav Saxena *is a Research Scientist with SGT Inc. at the Prognostics Center of Excellence NASA Ames Research Center, Moffett Field, CA. His research involves developing prognostic algorithms and methodologies to standardize prognostics that include performance evaluation and requirement specification for prognostics of engineering systems. He is a PhD in Electrical and Computer Engineering from Georgia Institute of Technology, Atlanta. He earned his B.Tech in 2001 from Indian Institute of Technology (IIT) Delhi, and Masters Degree in 2003 from Georgia Tech. Abhinav has been a GM manufacturing scholar and is also a member of IEEE, AAAI and ASME.*

Jose R. Celaya *is a staff scientist with Stinger Ghaffarian Technologies at the Prognostics Center of Excellence, NASA Ames Research Center. He received a Ph.D. degree in Decision Sciences and Engineering Systems in 2008, a M. E. degree in Operations Research and Statistics in 2008, a M. S. degree in Electrical Engineering in 2003, all from Rensselaer Polytechnic Institute, Troy New York; and a B.S. in Cybernetics Engineering in 2001 from CETYS University, Mexico.*

Bhaskar Saha *is a Research Scientist with Mission Critical Technologies at the Prognostics Center of Excellence NASA Ames Research Center. His research is focused on applying various classification, regression and state estimation techniques for predicting remaining useful life of systems and their components. He has also published a fair number of papers on these topics. Bhaskar completed his PhD from the School of Electrical and Computer Engineering at Georgia Institute of Technology in 2008. He received his MS from the same school and his B. Tech. (Bachelor of Technology) degree from the Department of Electrical Engineering, Indian Institute of Technology, Kharagpur.*

Sankalita Saha *received her B.Tech (Bachelor of Technology) degree in Electronics and Electrical Communication Engineering from Indian Institute of Technology, Kharagpur, India in 2002 and Ph.D. degree in Electrical and Computer Engineering from University of Maryland, College Park in 2007. She is currently a Research scientist working with Mission Critical Technologies at NASA Ames Research Center, Moffett Field, CA. Her research interests are in prognostics algorithms and architectures, distributed systems, and system synthesis.*

Kai Goebel *received the degree of Diplom-Ingenieur from the Technische Universität München, Germany in 1990. He received the M.S. and Ph.D. from the University of California at Berkeley in 1993 and 1996, respectively. Dr. Goebel is a senior scientist at NASA Ames Research Center where he leads the Diagnostics & Prognostics groups in the Intelligent Systems division. In addition, he directs the Prognostics Center of Excellence and he is the Associate Principal Investigator for Prognostics of NASA's Integrated Vehicle Health Management Program. He worked at General Electric's Corporate Research Center in Niskayuna, NY from 1997 to 2006 as a senior research scientist. He has carried out applied research in the areas of artificial intelligence, soft computing, and information fusion. His research interest lies in advancing these techniques for real time monitoring, diagnostics, and prognostics. He holds eleven patents and has published more than 100 papers in the area of systems health management.*