**Conference paper**
Zaidan, Martha Arbayani, Mills, Andy R and Harrison, Robert F (2011) *Towards Enhanced Prognostics with Advanced Data-Driven Modelling.* In: Proceedings of The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies. The Eighth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, 20-22 June 2011, Cardiff.

# Towards Enhanced Prognostics with Advanced Data-Driven Modelling

M A Zaidan, A R Mills and R F Harrison
Dept. of Automatic Control and Systems Engineering
University of Sheffield
Sheffield, S10 2TN, UK
+44 114 2225663
m.a.zaidan@sheffield.ac.uk

**Abstract**

A considerable amount of prognostics research has been conducted to improve the remaining useful life prediction of engineering assets. Advantages such as lowering sustainment costs and improving maintenance decision making, are significant motivations to enhance the prognostics capability. Sensor selection, data pre-processing, knowledge elicitation and the mathematical techniques are some of the elements required of prognostics research to enhance capability.

This paper takes a broad view of prognostics and explores techniques available from a variety of research and application disciplines. A prognostics dataflow diagram illustrates the complete prognostics process and the paper discusses the impact of improvements in each process step to enhance the prognostics performance. The mathematical approach to prognostics is a crucial issue. Exploring cross-disciplinary prognostic approaches is helpful to extract useful techniques from different domains and to fuse the strengths of each discipline.

A case study of fatigue induced crack-growth using Bayesian approaches is used to illustrate that data-driven prognostics can deliver benefits to the industry.

## 1. Introduction

### *1.1 The importance and definition of prognostics*

The operational life of vital assets is traditionally estimated using a conservative statistic, the so called “safe life removal interval”. The use of on-line health monitoring has created opportunities to increase the fidelity of the predictions, potentially reducing conservatism and operational disruption.

Statistical reliability distributions are based on the collective behaviour of a population of individuals acting in an assumed environment. The reliability distribution must capture the spread of failure behaviour resulting from each individual having its own sources of durability variation caused by manufacturing, material or maintenance. In addition, Byington et al. [1] claim that based on historical evidence, the actual usage of components/systems such as military aircraft is often significantly different from the intended usage and the operating environment. For example, the usage will depend on the pilot and flying style in aviation systems. Moreover, major causes of unscheduled maintenance events are unanticipated and extreme operating scenarios, which lead to serious operational issues, such as failure of missions and disruption costs. Therefore, it is advantageous to deploy condition-based maintenance, such that maintenance operations are based upon detection of faults (diagnostics) and the prediction of failure times (prognostics). Benefits from improved life prediction are given in [2,3].

A number of definitions of prognostics have been cited that seem to mix prognostic and diagnostic activities. For example, Schwabacher and Goebel [4] define prognostics as detecting the precursors of a failure, and predicting how much time remains before a likely failure. We would suggest that this refers to both diagnostics and prognostics and prefer a definition closer to that suggested by Saxena [5] that prognostics is the estimation of remaining useful life (RUL), where we would define RUL as the time until the functional requirements can no longer be met. Sikorska et al. [6] introduce a simple delineation regarding the relationship between diagnostics and prognostics: "diagnostics involves identifying and quantifying damage that has occurred, whilst prognostics are concerned with trying to predict the damage that is yet to occur". Therefore, prognostics can be considered as an extension to diagnostics and defined as: prediction, with quantified certainty, of residual functional capability.

Application of on-line health monitoring to enable prognostics is a relatively immature activity within the aviation industry. It is seen as advisable to explore other fields in search of experience. In this paper we draw on examples from techniques used in the health care of patients and econometric price modelling, exploring present and potential interaction enabled by parallels in these fields. Analogies can be made between the terms used in prognostics and in other disciplines, such as biomedicine and econometrics. For example, assets correspond to patients and commodities, failure modes map to disease and market events and operational conditions are comparable with smoking activity and market sentiment.

From the ideas explored in our cross-disciplinary review, we illustrate the potential advantages of fusing condition monitoring information with statistical reliability distributions. The Bayesian framework adopted allows the concepts and advantages to be illustrated graphically.

### *1.2 Background: cross-disciplinary prognostics*

A wide viewpoint has been adopted whilst reviewing the literature related to prognosis. This macroscopic view has uncovered three different classes of technique related to prognostics: reliability, survival analysis and forecasting. These fields have been explored in the context of the disciplines of health care and finance, as well as in engineering.

Reliability is the ability of a system or component to perform its required functions under stated conditions for a specified period of time [7]. Reliability theory has been widely used in the aviation industry to estimate the failure time probability distributions for aircraft fleets. Examples of reliability theory include: the use of the Bernstein reliability model to model the life characteristics of machine components by Ahmed and Sheikh [8], and the development of statistical methods for using degradation measures to estimate a time-to-failure distribution in Lu and Meeker [9]. The reliability failure distribution can be used as initial (prior) probability distributions (the knowledge base) in the prognostics process to reduce uncertainty. This field is concerned with inferences based upon a population of assets.

Survival Analysis is a collection of statistical procedures for data analysis for which the outcome variable of interest is the time until an event occurs [10]. In contrast to the population approach prevalent in reliability theory, the techniques described as survival analysis are often more individual centred, especially in biomedical applications. For instance, research into the survival of multiple myeloma patients to examine the association between the values of certain explanatory variables or covariates and the survival time of patients, conducted by Medical Centre of the University of West Virginia, USA [11]. Another field related to survival analysis is named competing risks analysis. Competing risks analysis is a field of applied statistics that has the ability to handle dependent failure. The experts in this area claim that “if something can fail, it can often fail in one of several ways and in more than one way at a time”. The unique capability of competing risks analysis in handling information censoring and dependent failures, makes this approach potentially valuable for analysing and modelling time-to-event data [12].

In analysing survival data, information censoring, such as the incomplete observation of survival times, is a particular challenge. Survival analysis has already tackled some of these problems and there is a portfolio of mathematical models and methodologies that have been developed to extract information from the censored observations maximally [13]. There is clearly a strong potential for survival analysis to provide new tools for prognostics.

Forecasting is the construction of a suitable model based upon analysis of the historical development of a data series and utilisation of information relevant to its likely future development [14]. This discipline is very common in econometrics, such as commodity price modelling and consumption predictions. In addition, other fields of science, e.g. metrology, have obtained benefit from this theory [15]. Evidence of numerous forecasting methods have been applied to prognostics, for example, the use of autoregressive integrated moving average (ARIMA) models to predict future machine health [16] and prognosis of remaining bearing life using neural networks [17]. Forecasting methods, through the use of sensor data, are adept at capturing the idiosyncratic failure behaviour of real world components.

Because each discipline has unique capabilities and past successes in research, it is believed that these advantages should be taken into consideration to develop optimal prognostic methods. Figure 1 shows the interaction of prognostics with these fields. Prognostics is a subset of the combination of these disciplines and therefore this representation is a powerful illustration of approaches that might be applied in prognostics.

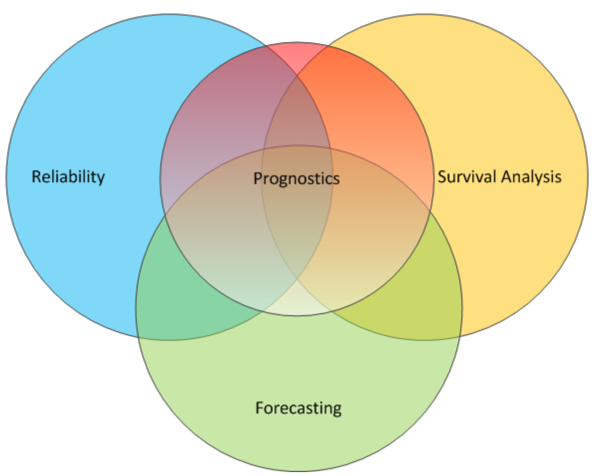


**Figure 1. The interaction of prognostics with different disciplines**

An example of the combination of these theories for prognostics is the research of Gebraeel et al. [18] in bearing prognostics. Reliability theory is used to create knowledge based on a population of bearings. A probabilistic forecasting method, in this case based on a Bayesian approach, is used to compute the update of the posterior parameters in order to predict remaining useful life of bearing.

The focus of our research work is to extract useful techniques from different domains and to fuse these in order to take the strength of each discipline.

## 2. The factors influencing prognostics accuracy

Much effort in prognostics research focuses on developing algorithms that can provide a precise RUL estimate with a small uncertainty bound around the prediction. In fact several factors influence the performance of prognostic tasks: each process step has specific functions and contributions to overall prognostic accuracy.

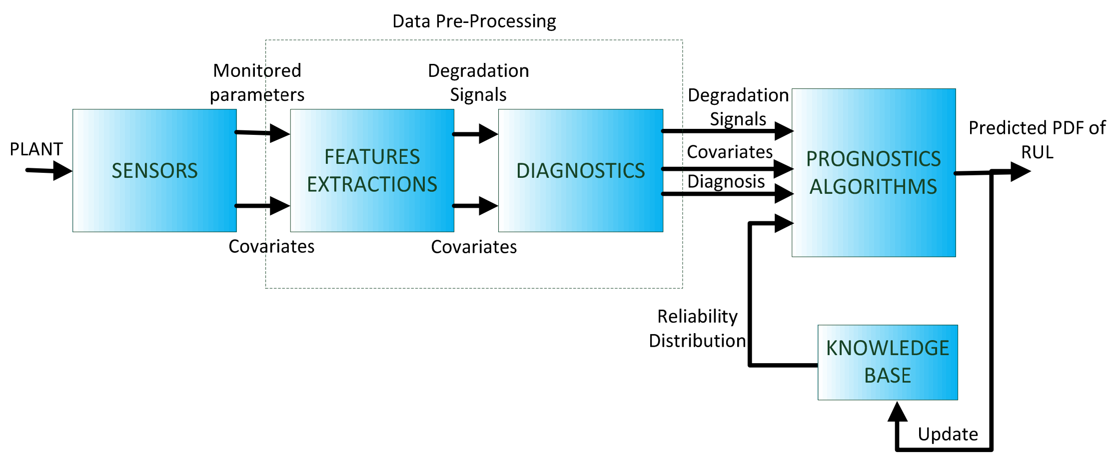


**Figure 2. The prognostics process**

Figure 2 shows a diagram of the prognostics process. Data collection is a fundamental part of prognostics and often requires the use of sensor systems to measure environmental and operational parameters. Sensor selection is one of the factors that have an effect on prognostic accuracy, which is dependent on the sensor accuracy, sensitivity, precision, resolution and measurement ranges [19].

Data pre-processing includes feature extraction and diagnostics.  These tasks detect abnormalities, identify failure modes and are potentially used to select a suitable prognostic model [6].

The knowledge base contains prior information about the pre-supposed reliability and failure behaviour of components/systems. Physics of failure analysis, expert opinions, previous qualification tests and in-service data are examples of sources for the knowledge base. The first two of these sources can be used for model-based approaches whereas the last two can be employed for data-driven prognostics. In-service collected data offers potential to account for real component and environmental variability and may provide a superior alternative to acceleration test data for which identifying an actual operational relationship can be difficult [7].

Finally, the prognostics algorithms play a very important task. Schwabacher and Goebel [4] surveyed numerous prognostics approaches and classify them into two main categories: model-based and data-driven approaches. In general, it can be said that the use of data-driven approaches is more promising because of their ability to deal with complex systems, which are difficult to capture with physical modelling. Techniques, such as those used in reliability theory allow the inclusion of covariates, such as environmental effects. Another reason for using the data driven approach is the abundance of normal degradation data because of sensor technology improvements. However, if it is possible to model degradation of components from physical considerations, both of these methods can be fused to take the strengths of each approach.

The most popular approaches to data-driven prognostics are artificial neural networks and fuzzy logic, as detailed in [4] and [6]. Bayesian techniques might be the most promising approach because of their ability to update probabilities of future observations by incorporating evidence from previous experience and experiment into the overall conclusion [20]. Numerous researchers have implemented this approach in prognostic models such as [18,21-23]. A prognostics case study using a Bayesian approach will be discussed in the next section.

It can be concluded that each process in prognostics has a different function and affects the accuracy and performance of the overall system. This can be illustrated as shown in Figure 3, where inputs, approaches and the knowledge base are key determinants of prognostic accuracy.

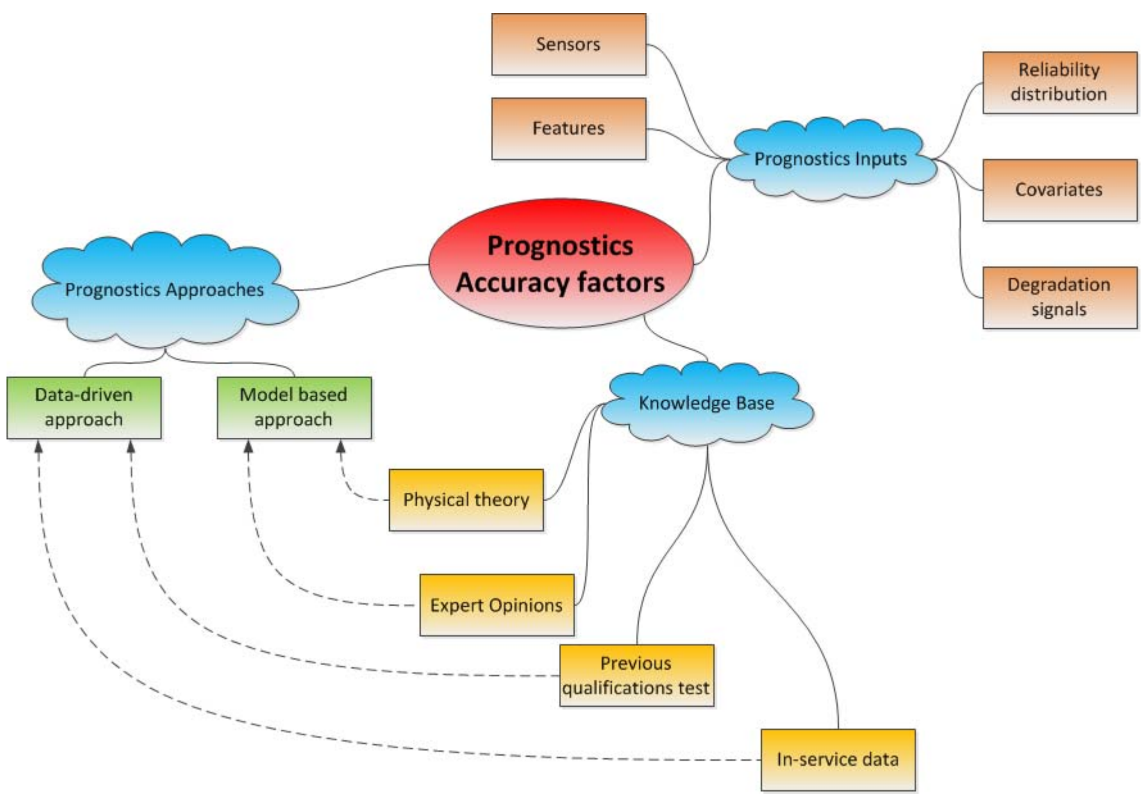


**Figure 3. The factors influencing prognostic performance**

## 3. A case study

In this section, an example of a prognostics algorithm is discussed. The approach implemented uses a Bayesian method extracted from the intersection of reliability, forecasting and prognostics perspectives (Figure 1). Fatigue crack growth is selected as a case study. This process is affected by many sources of variability, such as loading, material properties, geometry and boundary conditions. Therefore it is essential to express the crack size after a certain number of load cycles through a probability distribution. Probabilistic fracture mechanics is an active research area and numerous studies have addressed both model-based and data-driven techniques for probabilistic crack growth and life prediction. The detail and the latest progress in the topic of fatigue crack growth can be found in [24]. The assumption has been made is that periodic measurements are made of the crack length.

In this paper, the fatigue-crack-growth data obtained from [9] is used for testing the algorithm. There are 21 sample degradation paths, one for each of 21 test units. A critical crack length of 1.5 inches is defined as a "failure", a condition reached by half of the units. The sample paths contain useful information to estimate the time-to-failure distribution.

A Bayesian updating method developed by [18] is used to predict the degradation path of fatigue-crack-growth data in order to estimate remaining useful life. Readers are encouraged to refer to this reference for detail of the method. This method uses real-time condition monitoring information to update the stochastic parameters of exponential models. The natural logarithm of the degradation signal is modelled at time $t_k$ as follows:

$$L_k = \theta' + \beta t_k + \epsilon(t_k) \qquad (1)$$

where $\epsilon(t_k)$ is a random estimation error term that follows a normal distribution with mean zero and variance $\sigma^2$. $\beta$, $\theta'$ are normal random variables acting as tuning parameters for the estimated degradation model.

A joint posterior distribution for the tuning parameters can be calculated using the Bayesian updating method and is given by:

$$p(\theta', \beta | L_1, L_2, \dots, L_k) \propto f(L_1, L_2, \dots, L_k \middle| \theta', \beta)\pi(\theta')\pi(\beta) \qquad (2)$$

where, $\pi(\theta')$ and $\pi(\beta)$ are the prior distributions of the signal model parameters determined from the estimation of the reliability characteristics of the population of units and $f(L_1, L_2, \dots, L_k | \theta', \beta)$ is the likelihood function given the observed degradation signal $L_1, L_2, \dots, L_k$ at times $t_1, t_2, \dots, t_k$.

Once the joint posterior is obtained, the predictive distributions can be calculated and therefore the failure time distribution can also be determined.

Figure 4 shows the degradation signal of the fatigue crack growth of 21 units from which can be seen that about half of the data are right censored. Figure 5 illustrates the evolution of posterior means for $\theta'$ and $\beta$, as increased numbers of degradation signal measurements are incorporated.

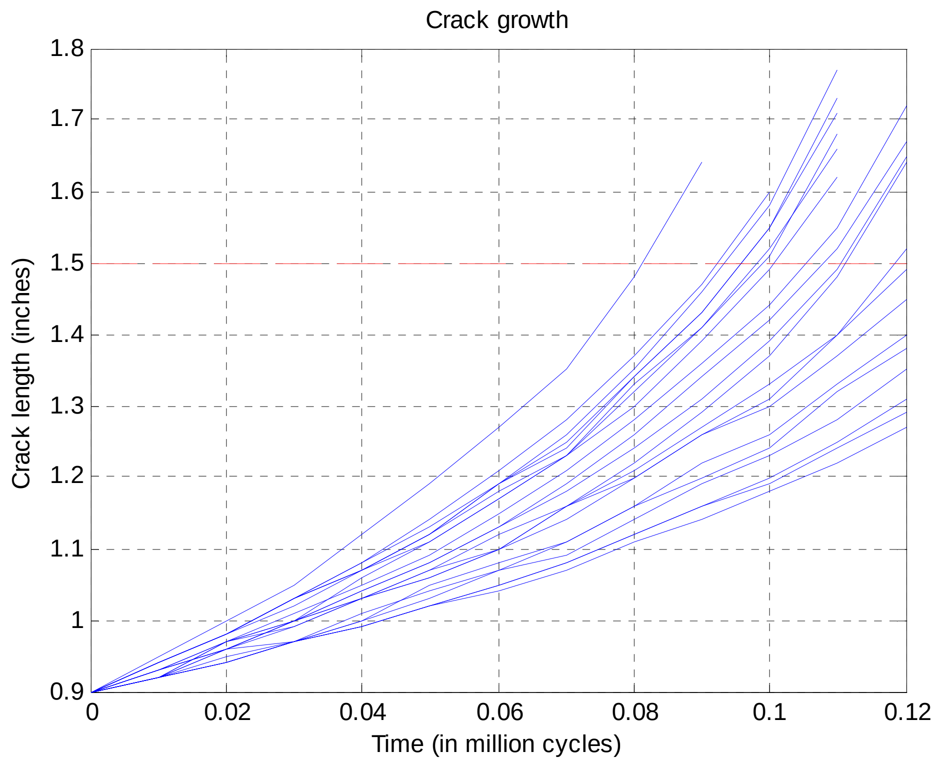


**Figure 4. Crack-length measurements versus time (in million cycles)**

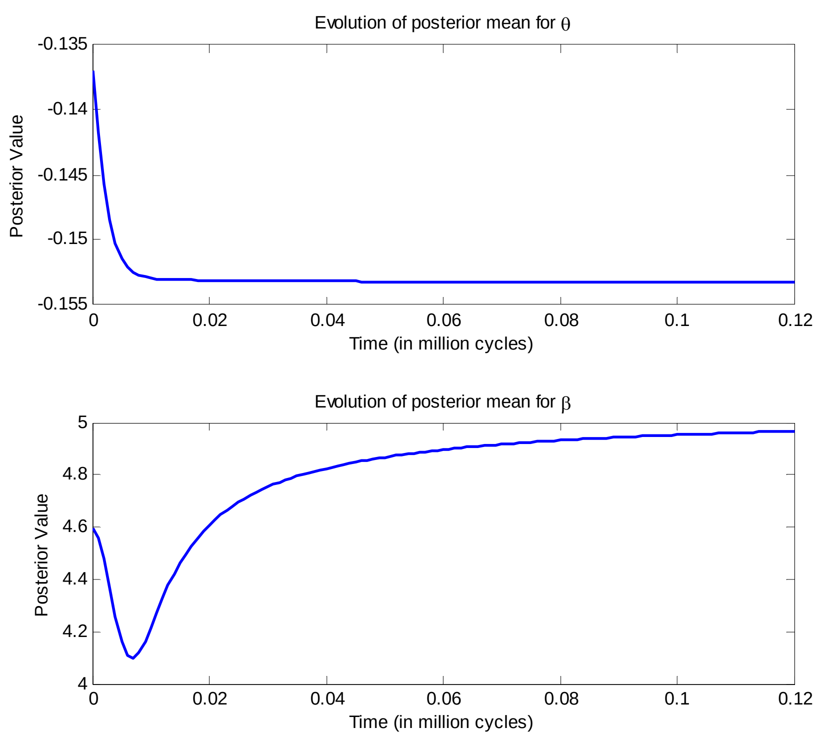


**Figure 5. The evolution of posterior means for $\theta'$ and $\beta$ for unit 10**

Figures 6 and 7 show two situations where benefit can be realised from the Bayesian approach. The first, in Figure 6, illustrates the potential for disruption avoidance. Before any degradation data is collected (at 0 cycles of use) the best estimate of predicted removal time is based solely on the prior distribution's 95% confidence limit (i.e. a time-based schedule). This distribution is the furthest right normal curve on the figure. Were this schedule used, the component planned removal would be at 0.1 million cycles, clearly this is after the actual failure event and would incur disruption cost. Collection of data over time and use of the Bayesian approach shows updated predictions can be made. A prediction after 0.07 million cycles produces a probability distribution curve (solid line in figure) with a 95% confidence limit marked by an arrow. This instructs the operator to remove the unit at 0.075 million cycles, which avoids the functional failure event and consequent disruption.

Figure 7 demonstrates how the unit which has a high durability can incur lost useful life if the maintenance decision was based on the 95% confidence limit of prior distribution, again removal at 0.1 million cycles. This is obviously before the functional failure event and would waste a healthy component which will impact on profit. Bayesian predictions at 0.07 million cycles of operation show that the safe removal limit may be extended to 0.15 million cycles. The uncertainty around the prediction increases because the predicted failure-time distribution moves further from the latest observation time (0.12 million cycles) and the mean prior failure time.

These two figures explain that the Bayesian updating method improve the failure time distribution estimations. In addition, these scenarios illustrate well how prognostics change the business, especially in aerospace industry that has high cost components.

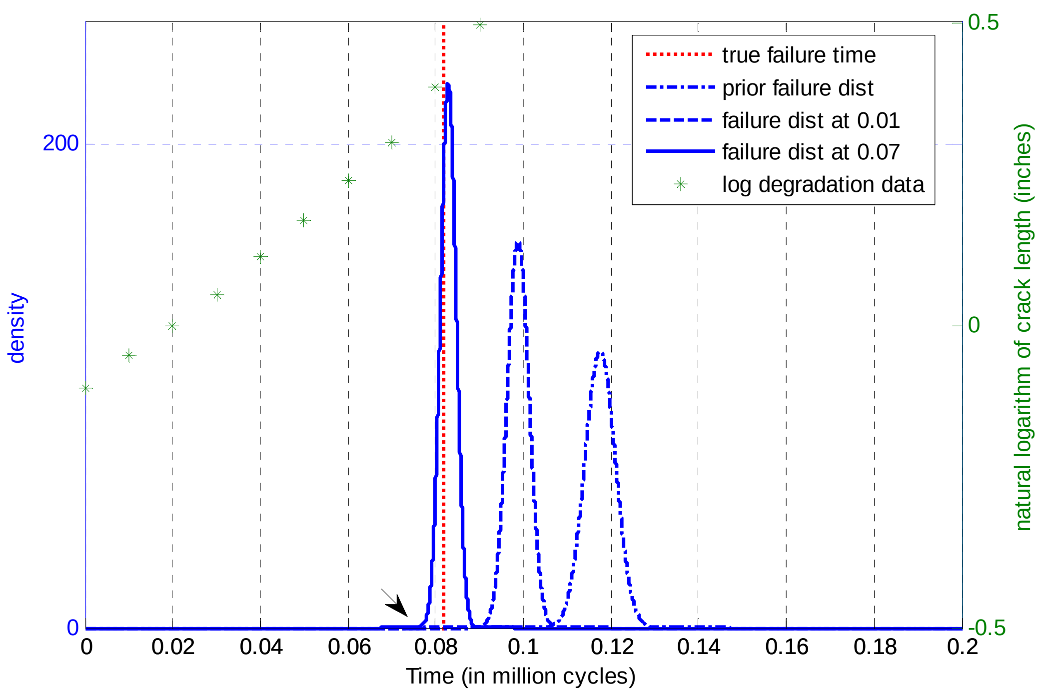


**Figure 6. Failure time distribution for unit 1**

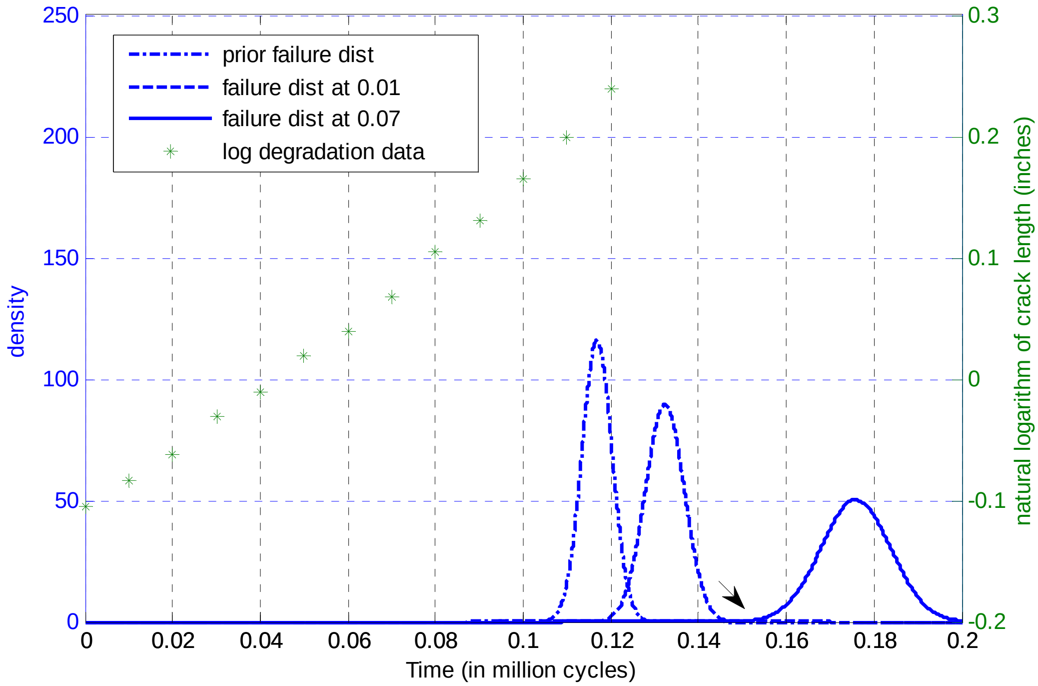


**Figure 7. Failure time distribution for unit 21**

# 4. Conclusion

In this paper, prognostics are explored from different viewpoints. This perspective is valuable to understand the possible approaches, the factors that impact performance and the potential cost benefits. Through a case study, the paper illustrates several advantages of taking the Bayesian route.

The paper has presented an overview highlighting similarities between several disciplines and has identified common connections to prognostics. The interaction diagram is helpful to classify prior art from different domains and to fuse the strengths of each discipline.

The explanation of the factors influenced prognostics accuracy is provided to understand the prognostics process. The prognostic process represents a complete prognostics system needed to be deployed in field applications. One of the most influential accuracy factors is the prognostics approach, which is the main focus of the on-going research. The Bayesian methods have been identified to possess many appropriate traits, therefore this technique is used to analyse in a case study.

The case study used fatigue crack growth data as a population of degradation signals which contains useful information to estimate the time-to-failure distribution. The approach demonstrated is extracted from the prior research in the intersection area of forecasting and reliability disciplines. The results from application to crack growth data show scenarios where a Bayesian treatment of prognostics has the potential to add value for industry over traditional time-based approaches.

In complex situations, when the environmental conditions are variable and the degradation signals are highly stochastic, the prediction becomes extremely difficult. Therefore, future research direction includes the use of covariates (explanatory parameters) to obtain precise results in more complex problems. Applying different models and more advanced approaches will also be taken into consideration when prognosing complex degradation characteristics.

**Acknowledgements**

The authors would like to thank Rolls-Royce plc for their support in this research.

**References**

[1] C.S. Byington, P. Sloelting, M. Watson, and D. Edwards, "A model-based approach to prognostics and health management for flight control actuators," *2004 IEEE Aerospace Conference Proceedings (IEEE Cat. No.04TH8720)*, 2004, pp. 3551-3562.

[2] J. Banks and J. Merenich, "Cost Benefit Analysis for Asset Health Management Technology," *2007 Proceedings - Annual Reliability and Maintainability Sympsoium*, 2007, pp. 95-100.

[3] K. Feldman, T. Jazouli, and P. a Sandborn, "A Methodology for Determining the Return on Investment Associated With Prognostics and Health Management," *IEEE Transactions on Reliability*, vol. 58, Jun. 2009, pp. 305-316.

[4] M. Schwabacher and K. Goebel, "A Survey of Artificial Intelligence for Prognostics," *Working Notes of 2007 AAAI Fall Symposium: AI for Prognostics*, 2007.

[5] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," *2008 International Conference on Prognostics and Health Management*, Oct. 2008, pp. 1-9.

[6] J.Z. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, Dec. 2010.

[7] M.S. Hamada, A.G. Wilson, C.S. Reese, and H.F. Martz, *Bayesian Reliability*, New York: Springer, 2008.

[8] M. Ahmad and a Sheikh, "Bernstein reliability model: Derivation and estimation of parameters," *Reliability Engineering*, vol. 8, 1984, pp. 131-148.

[9] C.J. Lu and W.Q. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, 1993, p. 161–174.

[10] D.G. Kleinbaum and M. Klein, *Survival analysis: a self-learning text*, New York: Springer Verlag, 2005.

[11] D. Collett, *Modelling survival data in medical research*, London: Chapman & Hall, 1994.

[12] Z. Ma and A.W. Krings, “Competing Risks Analysis of Reliability, Survivability, and Prognostics and Health Management (PHM),” *2008 IEEE Aerospace Conference*, Mar. 2008, pp. 1-21.

[13] Z. Ma and A.W. Krings, “Survival Analysis Approach to Reliability, Survivability and Prognostics and Health Management (PHM),” *2008 IEEE Aerospace Conference*, Mar. 2008, pp. 1-20.

[14] A. Pole, M. West, and J. Harrison, *Applied Bayesian Forecasting and Time Series Analysis*, New York: Chapman & Hall, 1994.

[15] S. Makridakis, S.C. Wheelwright, and R.J. Hyndman, *Forecasting Methods and Applications*, New York: John Wiley & Sons, Inc, 1998.

[16] W. Wu, J. Hu, and J. Zhang, “Prognostics of Machine Health Condition using an Improved ARIMA-based Prediction method,” *Industrial Electronics and Applications, 2007. ICIEA 2007. 2nd IEEE Conference on*, IEEE, 2007, p. 1062–1067.

[17] Y. Shao and K. Nezu, “Prognosis of remaining bearing life using neural networks,” *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 214, Jan. 2000, pp. 217-230.

[18] N. Gebraeel, M. Lawley, R. Li, and J. Ryan, “Residual-life distributions from component degradation signals: A Bayesian approach,” *IIE Transactions*, vol. 37, Jun. 2005, pp. 543-557.

[19] M.G. Pecht, *Prognostics and Health Management of Electronics*, New Jersey: John Wiley & Sons, Inc, 2008.

[20] D.A. Berry, *Statistics: A Bayesian perspective*, New York: Duxbury Press, 1996.

[21] K.W. Przytula and A. Choi, “Reasoning Framework for Diagnosis and Prognosis,” *2007 IEEE Aerospace Conference*, 2007, pp. 1-10.

[22] B. Saha, S. Poll, K. Goebel, and J. Christophersen, “An integrated approach to battery health monitoring using bayesian regression and state estimation,” *2007 IEEE Autotestcon*, Sep. 2007, pp. 646-653.

[23] B. Saha and K. Goebel, “Uncertainty Management for Diagnostics and Prognostics of Batteries using Bayesian Techniques,” *2008 IEEE Aerospace Conference*, Mar. 2008, pp. 1-8.

[24] S. Sankararaman, Y. Ling, and S. Mahadevan, “Uncertainty quantification and model validation of fatigue crack growth prediction,” *Engineering Fracture Mechanics*, Mar. 2011.