Bayesian-based Methods for Transitioning Between Prognostic Estimates to Leverage Available Data

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Alan Y. Nam

May 2016

Copyright © 2016 by Alan Y Nam
All rights reserved.

ABSTRACT

The availability of failure data directly impacts the empirical prognostic models that can be built. In turn, these models impact the accuracy and uncertainty of remaining useful life (RUL) estimates of systems and components. While ideally a large amount of data of previous failure modes can be collected, the difficulty in obtaining such data can present a significant hurdle. To alleviate the constraints of limited data, Bayesian-based methods of transitioning between different prognostic models were developed. This updating scheme leverages existing data in order to create a unified estimate.

Two novel methods of transitioning are proposed to augment existing prognostic models. The first is the RUL update. This method combines two or more RUL distributions into a posterior RUL using Bayes formula. The RUL regression transition can be used with the general path model (GPM). The GPM uses linear regression to extrapolate to future states of a degrading system. The use of transitions are variations on this regression theme. The RUL regression model is a weighted total least squares regression model that accounts for observation errors in both the RUL and degradation threshold uncertainties.

A third method, while not a transition, improves the basic GPM. The coefficient update applies Bayes rule on the linear regression coefficient estimates using a prior population of coefficients.

These three methods were validated on two datasets: a simulated set of 24 signals and data from a heat exchanger test bed. The best models were found to decrease the root mean squared error by 76% and 39%. The use of any transition lowered the prediction errors over the lifecycle of each test case.

TABLE OF CONTENTS

1. Introduction 1
1.1 Background 1
1.2 Problem Statement 5
1.3 Original Contributions 6
1.4 Organization of Dissertation 7
2. Lifecycle Framework 8
2.1 Type I - Traditional Reliability 9
2.2 Type II - Stressor Based 14
2.3 Type III - Degradation Based 16
2.3.1 Monitoring and Fault Detection 17
2.3.2 Prognostics Parameter Generation 21
2.3.3 General Path Model 23
2.3.4 Alternatives to GPM 26
2.4 Previous Methods for Model Fusion 29
2.5 Bayesian Applications to Prognostics 30
2.5.1 Updating Degradation Path Parameters [Coble and Hines 2011] 31
2.5.3 Other Applications of Bayesian Theorem 35
2.6 Summary of Lifecycle Prognostics Methodology 36
3. Lifecycle Transition Methodology 38
3.1 Bayesian Statistics 42
3.1.1 Bayes Formula 42
3.1.2 Analytical and Conjugate 43
3.1.3 Numerical 47
3.1.4 Metropolis-Hastings Algorithm 49
3.1.5 Summary of Bayes Formula Computation 52

3.1.6 Kernel Density Smoother and Empirical Fits 54
3.2 Weighted Total Least Squares Regression 57
3.3 Transitions 60
3.3.1 RUL Update 60
3.3.2 Type III GPM Coefficient Update 63
3.3.3 Type III GPM WTLS Regression 67
3.4 Summary of Transitioning Methods 71
4. Application to Data sets 73
4.1 PHM Challenge Data 73
4.1.1 Type I RUL Update of PHM Data 82
4.1.2 Coefficient Transition of PHM Data 84
4.1.3 Combined Coefficient and RUL Update on PHM Data 87
4.1.4 RUL regression on PHM Data 89
4.1.5 Error Analysis of PHM Data 90
4.2 Heat Exchanger Test Bed 92
4.2.1 Type I RUL Update of HX Data 96
4.2.2 Coefficient Updates of HX Data 99
4.2.3 RUL Regression of HX Data 103
4.2.4 Error Analysis of HX Data 104
4.3 Summary of Application Results 105
5. Conclusions and Recommendations 108
5.1 Conclusions 108
5.2 Recommendations for Future Work 112
List of References 115
Appendix 124
Vita 130

LIST OF TABLES

Table 2-1: RUL Estimates for Type I 13

Table 3-1: A comparison of Bayes calculation methods 53

LIST OF FIGURES

Figure 1-1: Cost Optimization of Maintenance Strategies [Tchakoua et al., 2014] 2
Figure 1-2: The P-F Curve 3
Figure 2-1: Lifecycle Prognostics 10
Figure 2-3: Prognostic Parameter 17
Figure 2-5: Single Hidden Layer Feedforward Neural Network 27
Figure 2-6: Test Case Censored at Time 20 34
Figure 2-7: Censored Case Using Coefficient Priors 34
Figure 3-1: Information Flow for Transitions Among Prognostic Models 40
Figure 3-2: Gaussian Conjugate Family Updating for $\eta = 100$, $\tau = 7$, $\mu = 110$, $\sigma = 5$ Days . 46
Figure 3-3: Gaussian to Weibull Using Numerical for $\eta = 100$, $\tau = 7$, $\beta = 7$, $\theta = 150$ Days . 48
Figure 3-4: Gaussian to Weibull Using MH for $\eta = 100$, $\tau = 7$, $\beta = 7$, and $\theta = 150$ 51
Figure 3-5: Kernel Smoother on Data 55
Figure 3-6: RUL Update 62
Figure 3-7: Previous Method for Including Prior Coefficients 65
Figure 3-8: Proposed Coefficients Update 65
Figure 3-11: RUL Regression Example 70
Figure 4-2: PHM Correlation Coefficients 76
Figure 4-3: AAKR RMSE 78
Figure 4-5: Testing Prognostics Path 80
Figure 4-6: Degradation Path of an Example Case 81
Figure 4-7: RUL Distributions on Run 1 at Time 77 83
Figure 4-8: Test #1 with Type I RUL Update 83
Figure 4-9: Prior Quadratic and Linear Coefficients 85
Figure 4-10: Prior Intercept Coefficient 85
Figure 4-11: Coefficient Update 86

Figure 4-12: Combined Coefficient and RUL Updates 88
Figure 4-14: RMSE of Models for PHM Data 91
Figure 4-15: Heat Exchanger Schematics 94
Figure 4-16: Temperature Residuals over Lifecycle 95
Figure 4-17: Test Run 9 with Self-Healing 97
Figure 4-18: Heat Exchanger Time of Failure Distribution 97
Figure 4-19: Test HX RUL Update 98
Figure 4-20: Histogram of Quadratic Prior 100
Figure 4-21: Histogram of Linear Prior 100
Figure 4-22: Histogram of Intercept Prior 101
Figure 4-23: HX Coefficient Transition 101
Figure 4-24: HX Combined Transition (Zoomed) 102
Figure 4-25: RUL Regression 103
Figure 4-26: RMSE Over All HX 104

Abbreviations and Symbols

AAKR	Auto-Associative Kernel Regression
BOL	Beginning of Life
EOL	End of Life
GLS	Generalized Least Squares
GPM	General Path Model
HX	Heat Exchanger
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
ML	Maximum likelihood
NN	Neural Network
OLS	Ordinary Least Squares
PDF	Probability Density Function
PHM	Prognostics and Health Management
POF	Point of Failure
RMSE	Root Mean Squared Error
RUL	Remaining Useful Life
SPRT	Sequential Probability Ratio Test
SSR	Sum of Squared Residuals
TOF	Time of Failure
WTLS	WTLS

1. INTRODUCTION

1.1 Background

The reliability and safety of the 99 nuclear power reactors operating in the USA are key issues facing the nuclear industry today. Most of these plants have completed the necessary license renewal process with the NRC to extend their original 30-40 year operating design lives by 20 years. However, some consideration is being given to extending the life another 20 years to beyond 60 years [Bond et al., 2011].

Much has been invested into safely operating these valuable plants that produce 20% of the nation's energy. These reactors' operations and maintenance costs are estimated at around 60-70% of the overall generating cost, compared to the cost of fuel at 15-30%. Of the total maintenance cost, about 80% is labor related. To mitigate these costs, it's been suggested that the nuclear industry can save an estimated $1 billion a year by implementing on-line monitoring with diagnostics and prognostics on all key equipment [Bond et al., 2007].

Some inefficiencies are introduced in the way maintenance is carried out. This maintenance is often time-based. Each system or component is scheduled for maintenance at specified time intervals, regardless of any indication of abnormal behavior. To maintain high reliability, a high-frequency replacement period can be imposed to replace components long before their time of failure (TOF). Failure is defined as the point at which a component no longer meets its design function. The cost of this preventative maintenance scheme of replacing healthy parts before they fail is shown in Figure 1-1.

On the other extreme, reactive maintenance is done after a failure has already occurred. Each instance of unplanned downtime is typically much more costly to plant

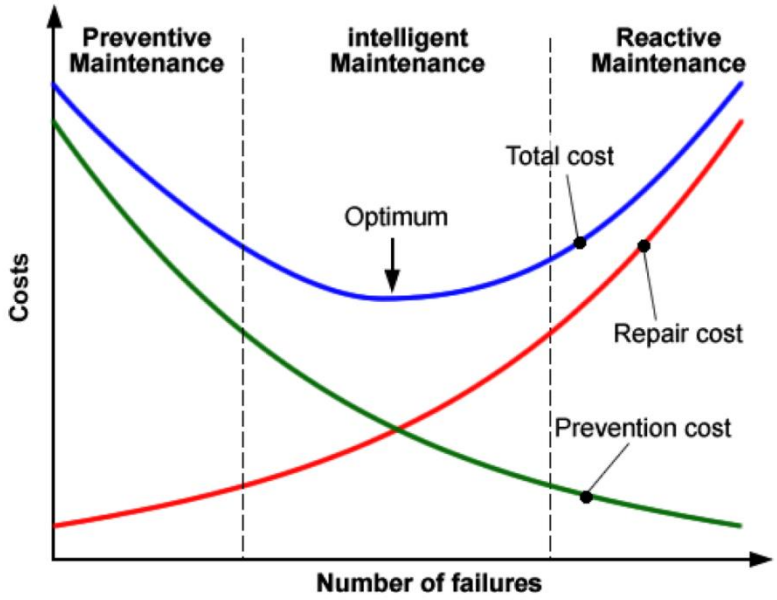


Figure 1-1: Cost Optimization of Maintenance Strategies [Tchakoua et al., 2014]

production than planned downtime due to increased repair time and potential damage to other parts. A nuclear power plant's unplanned shutdown has a large impact on production and profitability. Between the two extremes lies a minimum maintenance cost. While Figure 1-1 shows a general case for operation in any industry, the nuclear industry strives to be as reliable as possible, shifting the optimum to the left. The overall point still stands. Too much preventive maintenance results in increased cost that could be avoided if a smart maintenance replacement strategy were implemented. If the exact TOF were known, a piece of equipment could be run to the full extent of its useful life before being replaced.

Some pieces of equipment do not break immediately, but gradually decline in system health. The well-known P-F curve, illustrated in Figure 1-2, indicates the health of a single component as it approaches its TOF. After a component operates within its design parameters, a fault can occur, precipitating a decrease in equipment condition. At this stage, the fault may be undetectable. As it degrades, however, inspection or a diagnostic system can detect symptoms, such as increased vibration or heat.

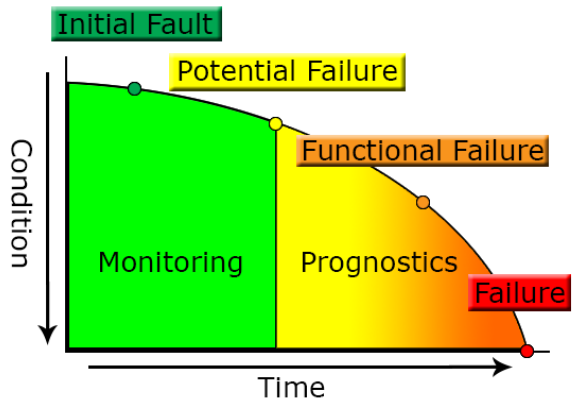


Figure 1-2: The P-F Curve

Without maintenance, the equipment may not operate at its full functionality or can reach the end of its allowable tolerance. Left untreated, the component will reach its failure, at which point it no longer meets its design functions. Performing maintenance after a fault is detected, but before failure occurs, is known as condition-based maintenance. This strategy avoids unnecessarily replacing healthy parts that is inevitable in scheduled maintenance.

Prognostics is the science behind accurately estimating a RUL. The general approach is to predict the system health's future state, until reaching a high probability of failure. Broadly speaking, there are two distinct approaches to prognostics [Goebel et al., 2005]. The first is based on first principle, or physics-based, models, which use the science of underlying interactions among materials. This approach can also involve using process simulation or other means of finding underlying relationships among critical components [Shen et al., 2010].

The other main approach, which is the focus of this paper, is empirical or data-driven prognostics [Schwabacher and Goebel, 2007]. This methodology uses data taken from the system to determine the amount and type of degradation accumulated in the system over its lifecycle [Lu and Meeker, 1993]. Data collected when the system is known to be healthy can be used as a benchmark for how the system should behave in the absence of degradation. This data contrasts with data collected before failure, when signs of degradation can be identified. Using an automated empirical prognostics approach can provide maintenance personnel fast and cost-effective information. It can also reduce the need for a large spare-parts inventory by facilitating ordering when necessary.

1.2 Problem Statement

One of the main bottlenecks to empirical prognostics is the availability of data. Data of systems or components leading towards failure can be difficult to obtain for many reasons. When a failure mode is discovered, the system is often redesigned to avoid that failure mode in the future. When a failure mode cannot be designed out, a maintenance strategy can be adopted to prevent the failure mode from manifesting. Sudden failures while the system is in operation can be very costly and are avoided when faults are detected. These faults are not allowed to precipitate towards failure.

The high cost of unplanned maintenance makes examples of systems heading towards failure costly to reproduce. In addition, the type, organization and compilation of data can vary. Different types of data can lead to different prognostic models of the system; each has its own set of assumptions, advantages, and characterization of uncertainty.

To leverage available data, multiple prognostic methodologies can be applied to produce a combined RUL estimate. In this research study, these models are fused by using a Bayesian-based updating methodology. Bayesian statistics naturally combines two sets of beliefs weighted by their uncertainties. First, a prior distribution for a specific random variable, usually RUL, must be defined based on prior data or analysis of the system. This prior is combined with newly observed data to calculate the posterior distribution. This data fusion can use all data available, not just specific sets required for each method. If two models differ, their posterior estimate is calculated using the prior and likelihood estimates, weighted by their relative uncertainties.

1.3 Original Contributions

While different empirical prognostic methods exist in literature, this dissertation focuses on transitioning among methods in order to produce combined RUL estimates. This combined estimate incorporates information taken from different sources of data and may increase the accuracy of RUL prediction. These methods were developed and applied to two test datasets: a publically available dataset of simulated sensors and a heat exchanger test bed. The original contributions of this research can be summarized as follows:

- Development of a framework to incorporate different sources of data across multiple prognostic models.
- Development of two transitioning methods between different prognostic methods.
    - RUL update applies Bayes updating on RUL estimates to form a posterior estimate.
    - RUL regression uses prior prognostic models in weighted total least squares regression.
- A coefficient update method for improving the GPM using Bayes formula on the prior distribution and regression estimate
- Development of MATLAB functions that can be applied to prognostic datasets. These include functions that codify Bayesian algorithms and adaptations of the PEP toolbox to accommodate transitions.
- Demonstration of error reduction on two test datasets: a publically available dataset of simulated faults and a heat exchanger test bed. An analysis of the error reduction is discussed and key areas of benefits are identified.

1.4 Organization of Dissertation

The following section, Chapter 2, outlines different types of prognostic methods, which are categorized into three types based on the information they use to generate RUL estimates. This chapter lays the foundations of prognostic analysis. It shows how to use different sources of data to generate RUL distribution estimates.

With the prognostic methods introduced, Chapter 3 focuses on how to transition between those methods. Beginning with a brief introduction to Bayesian statistics, Bayes' formula is presented as well as methodologies for its use. The weighted total least squares regression (WTLS) model is also introduced. With the mathematical background in place, the three transitioning methods, RUL update, coefficient update, and RUL regression, are detailed.

Chapter 4 presents two case studies applying the described transitioning methods to two different data sets. The first data set is a simulated set of hundreds of training and test cases of 24 unlabeled signals. This dataset is available from the NASA data repository [Saxena and Goebel, 2008]. The RUL estimates are analyzed in the context of seeing how the RUL evolves throughout an individual case's lifecycle. For both datasets, applying a transition reduces the reported root mean squared error by significant percentages.

The final chapter summarizes the research's conclusions and suggests methods to improve analysis and the developed methods.

2. LIFECYCLE FRAMEWORK

While the novel work in this study addresses the transitions between prognostic models, the models themselves must first be defined and presented in a lifecycle prognostics framework. This lifecycle framework assumes that at different stages during the life of an individual piece of equipment, different prognostic data are available. Before installation, nothing is known about the individual's behavior. Possible data could include historic TOF information gathered either on-site or from the parts manufacturer. During operation, the operating conditions can be monitored. An operating condition is any measure of the severity of stress put on the individual component. Measures include temperature and operating capacity. Increased stress reduces equipment lifetime. When an expert monitoring system detects a fault, deviations from normal behavior can be monitored. These deviations, such as increased temperature or vibration, provide valuable information about the system's state of health. The degradation is an expression of how far from a healthy condition the faulted system has strayed. This degradation can be tracked until failure is reached.

For each of these stages of data collection, there are types of models to calculate RUL [Hines and Usynin, 2008]. Type I models encompasses traditional reliability and survival analysis methods based on statistical analysis of historical TOF data. Type II models consider the individual component's operating conditions. Finally, type III models use system-degradation information to extrapolate to the system's future states. The degradation's quantification into a variable can be referred to as the prognostics parameter. A slight distinction is made here in that the actual amount of degradation is usually hidden. The prognostics parameter is a degradation estimate that is used in modeling.

Not only can these types of models be separated by the type of data used, but also when in the lifecycle they are available. Information about an individual component

at the beginning of life (BOL) is limited. However, at the end of life (EOL) more data can be gathered that describes the individual case with a higher degree of specificity. Therefore, it is expected that, as an individual ages, different prognostic models can be applied with increasing sophistication and the pool of data from which to draw knowledge also increases. A combined lifecycle framework should incorporate all measurable information to achieve higher RUL prediction accuracies.

Figure 2-1 shows this framework starting with data collection from equipment and ending in RUL estimation. The equipment moves through time from before installation, to operation, to fault detection to failure. From these milestones, the historical TOF, operating conditions and tracked degradation datasets, presented in yellow, can be collected. Different prognostic types, in green, are limited by the dataset that is available. TOF data can be used to build type I models, but type II models depend on operating conditions. Likewise, type III models require tracked degradation. Without transitions among prognostic types, each model produces an RUL estimate. It is assumed that the more specific the information available, the better the RUL estimate; if type I, II and III models are available, the type III would be favored. However, with transitions, the types I and II can be used to augment the type III to form a combined RUL estimate.

2.1 Type I - Traditional Reliability

As previously mentioned, these methods are based on widely known traditional reliability analysis [Weibull 1951]. By gathering data on previously known TOF, a distribution can be fit to these failure times. When a TOF distribution is known, it can be characterized using such functions as Weibull, Gaussian/normal, and lognormal. Depending on the distribution used, the corresponding conditional probability of failure

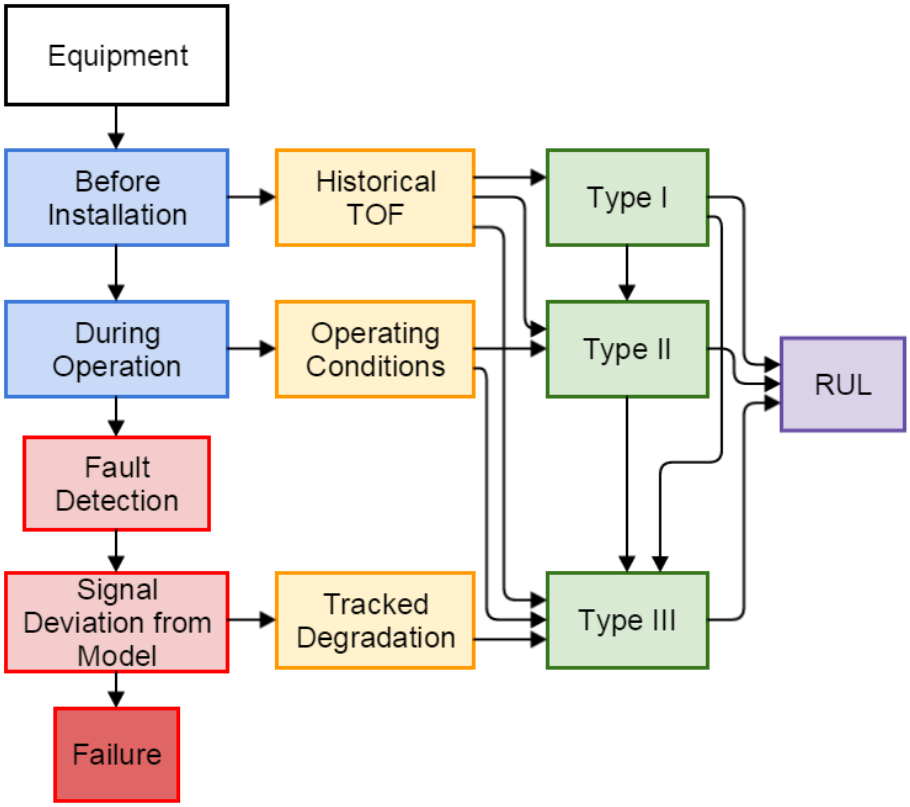


Figure 2-1: Lifecycle Prognostics

can be found based on the current time. This probability of failure is an expression of the future reliability, given that it survived a certain length of time. In addition, by setting the reliability percentile, an RUL can be found at a given confidence interval. The Weibull distribution is presented to illustrate a RUL distribution's calculation.

The Weibull probability density function (PDF) is described by two parameters in the equation

$$f(t) = \frac{\beta}{\theta}\left(\frac{t}{\theta}\right)^{\beta-1} \exp\,(-t/\theta)^{\beta} \qquad 2\text{-}1$$

where β is the shape parameter and θ is the scale parameter. The shape parameter determines the distribution's overall shape. The Weibull distribution is the most widely-used fit because of the flexibility in shape. It can approximate a wide range of distributions, encompassing the exponential and approximating the normal and lognormal. The scale parameter sets the distribution's magnitude. The parameter's value is equivalent to the time at which the cumulative failure is approximately 62.3%.

Figure 2-2 shows an example TOF distribution with a Weibull fit. The TOF PDF is approximated using a Weibull, shown in red.

The conditional Weibull can be derived from a ratio of probability of failures expressed in the following equation.

$$P(T \leq t_0 + t) = \frac{P(t_0 < T \leq t_0 + t)}{P(T > t_0)} \qquad 2\text{-}2$$

The numerator is the probability of failure between the current time t_0 and some projected future time t. T denotes the TOF random variable. The probability of survival normalizes the numerator to the current time. When the Weibull is substituted in and solved in terms of reliability R, the conditional Weibull is found.

The following equation finds the reliability after a period of survival and is used to establish the PDF over time.

$$R(t|t_0) = \exp\,-[(\frac{t_0+t}{\theta})^{\beta} - (\frac{t_0}{\theta})^{\beta}] \qquad 2\text{-}3$$

By rearranging terms, the reliability can be found over a 95% confidence interval by

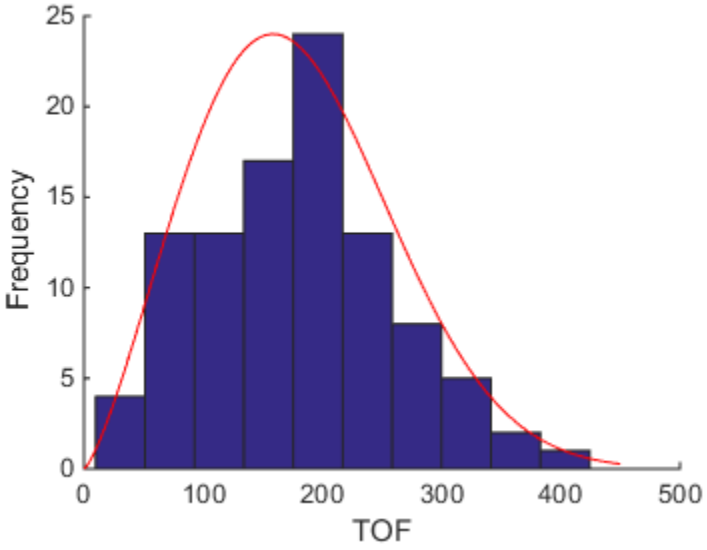


Figure 2-2: Example TOF Distribution with Weibull Fit

inputting the upper and lower bounds and solving for time.

The following equation is the calculation of the RUL, based on the reliability.

$$t = \theta\left(\left(\frac{t_0}{\theta}\right)^{\beta} - lnR\right)^{1/\beta} - t_0 \qquad 2\text{-}4$$

Table 2-1 shows the results of type I prognostics on the sample distribution. These results were calculated from mean residual life calculations based on a 95% reliability. The RULs shown are the amount of time away from the current time that 95% of the remaining population are still in operation. This early safe calculation conservatively estimates the RUL. As the current time increases, the spread of the 95% CI of the RUL decreases. This decrease expresses a tighter range between which a component is expected to fail. At the EOL extreme, the life is expected to end in the near future. The 95% CI was found by setting the reliability to 2.5% and 97.5%.

Table 2-1: RUL Estimates for Type I

Current Time	RUL at 95% Reliability	95% Confidence Interval of RUL	
0	56.3	41.6	354
25	34.8	21.6	329
50	21.8	12.0	306
75	14.6	7.65	283
100	10.6	5.39	262
125	8.06	4.06	242

2.2 Type II - Stressor Based

Type II models consider a component's operating conditions when calculating RUL. For example, it considers the idea that components under different levels of stress degrade at different rates. Components under high stress are expected to have shorter lifespans than those under low stress.

One such model is the Cox proportional hazards model [Cox 1972], which can be applied to predict RUL [Liao et al., 2006]. This model uses operating conditions as covariates, z, which can indicate the level of stress imposed on the system, such as temperature or material composition. These covariates are assumed to be multiplicative factors that affect the hazard rate in those operating conditions. This model is a statistical survival technique often used with accelerated life testing because of its potential to model a range of operating conditions, even those not tested directly. Accelerated life testing has many benefits. A small group of failed units under scrutiny can potentially provide useful information about the whole fleet. This test group's accelerated life-spans (setting aside the long life-spans that some components can have) saves money and produces results in a timely fashion.

To create a proportional hazard model, a baseline hazard rate is set, λ_0. This function represents the probability of failure over time in the absence of covariates' effects or at covariates' baseline reference values. A specific hazard rate can be found using Equation 2-5, where k is the number of covariates, and β_i is a scaling parameter for each covariate z_i. This model assumes that the covariates have a multiplicative effect on the hazard rate, with an acceleration factor exp(βz).

$$\lambda(t|z) = \lambda_0(t) * exp\left(\sum_{i=1}^{k} \beta_i z_i\right) \quad 2\text{-}5$$

The acceleration factor is generally assumed to have a known functional form. Using the hazard rate and basic conversion to the reliability distribution, an RUL distribution can be estimated.

Another approach to stressor-based prognostics is to use Markov Chain Monte Carlo (MCMC) simulations to estimate RUL [Kharoufeh and Cox, 2005]. These MCMC models assume that the operating condition sustained over time will have a negative effect on the system's health. A system operating at a higher capacity, or stress, should contribute a larger amount of degradation, which is tracked and accumulated over time. This degradation is the multiplication of the degradation rate assigned for a given operating condition by the amount of time spent in that operating condition. Some important assumptions limit the proposed model's use. One is that the accumulated degradation results in failure at around the same point for all components. Another assumption is that the degradation paths monotonically increase or decrease.

At a MCMC algorithm's core is a transition probability matrix, shown in equation 2-6. Each element p_{ij} represents the probability of transitioning to the next state based on the current state, from i to j. The sum of a row should be equal to unity, as this probability encompasses all the subsequent state's possibilities. For example, a unit in state 1 has the potential to stay in state 1 or transition to any other state. For each state some predictable amount of damage, z_k, can be assigned. Previous failed cases are used to determine the average accumulated damage threshold at which failure occurs, along with assigning damage values to each state.

$$Q = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \cdots & \cdots & \cdots \\ p_{m1} & \cdots & p_{mn} \end{bmatrix} \quad 2\text{-}6$$

As the component passes from state to state, damage is accumulated, as shown in the following equation.

$$Y(t_k) = \sum_{i=1}^{k} g(E(t_i, t_i + \Delta t))\Delta t \quad 2\text{-}7$$

where g is rate of damage as a function of environmental stressors E over the interval $[t_i, t_i + \Delta t]$.

For a hard threshold model, when Y reaches a predetermined amount, it is assumed that the component will fail. In contrast, a soft threshold model attempts to

include variation based on the uncertainty of the point of failure. Regardless of the model, because of the transition matrix's probabilistic nature, chains over repeated trials yield different RUL estimates. MCMC methods using repeated sampling of the Markov Chain process yield an RUL distribution.

2.3 Type III - Degradation Based

Type III methods use direct or inferred observation of degradation to predict RUL. The measureable degradation can be found in a monitored system variable, such as temperature, vibration, efficiency, or a combination of different variables. For the purposes of prognostics, it's useful to quantify degradation as a prognostics parameter that can be tracked and extrapolated over time.

A monitoring system is used to track the system's health over time. When the system operates under normal conditions, the residuals are expected to stay at low values with an approximate zero mean.

A fault is the earliest point at which something abnormal can be detected. A dedicated fault-detection system can be designed to alert plant operators that the system has moved from a normal to a faulted condition. When the fault precipitates towards failure, the degradation can be tracked until the point of failure (POF), when the system is assumed to fail.

The failure should be clearly defined for the specific application. While failure usually refers to when the component no longer fulfills its design function, it does not necessarily have to be so for prognostic purposes. For example, a pump failure can be a flow rate of zero. Alternatively, the failure can be some level of tolerance such that it no longer meets operational requirements or safety standards. The pump is performing so poorly it has effectively failed. Another definition could be when the pump reaches

some tolerance level that alerts an operator it needs to be replaced. The flow rate may be enough for safe operation, but should be repaired to avoid future problems. Once the failure is defined, it should be kept consistent between the training and testing cases.

Figure 2-3 shows an example of a prognostic parameter for a single case over a lifecycle. For this case, the system operates under normal conditions until time 160. The nonzero noise is a natural result of sensor and process variation over time. At this point a fault occurs and the degradation increases up to the point of failure around a degradation of 15.

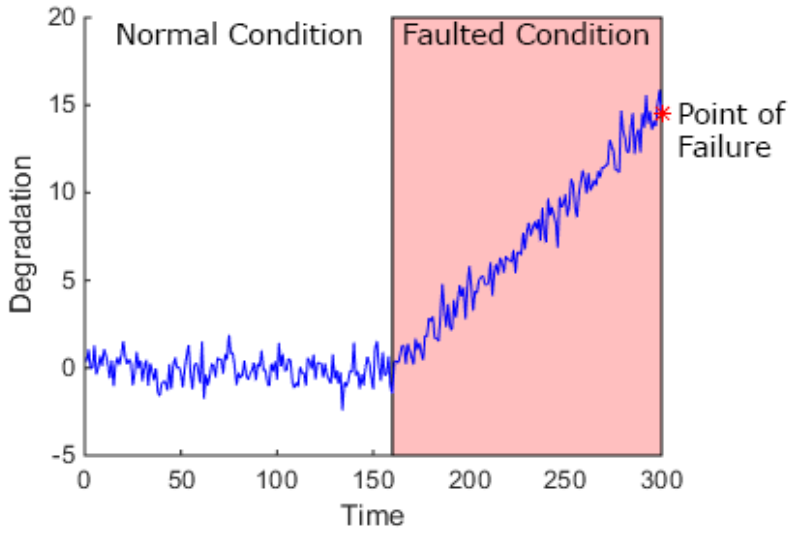


Figure 2-3: Prognostic Parameter

2.3.1 Monitoring and Fault Detection

A condition monitoring system is one that measures and tracks the condition of an operating piece of equipment in order to assess the current state of health. It has many applications in safety, reliability, and maintenance. Many critical systems across

all industries apply some form of condition monitoring. Hence, the addition of prognostic analysis to those monitoring systems takes advantage of existing infrastructure. Condition monitoring tracks measured signals over time. Typically, these signals hold at constant values with small fluctuations under steady-state conditions. These constants can change if there is a change in the operating condition, such as system transients during start up and shut down or during the onset of failure. By comparing the current measured system with all known instances of normal operations, a change in the system can be checked for normal or abnormal behavior. If the system is outside the range of normal behavior, a fault-detection system can signal to the operator a potential fault.

One monitoring method is the auto-associative kernel regression model (AAKR) [Hines and Garvey 2006]. It's an error correction method that predict corrected values for all inputs at once. The model stores vectors of signal data taken under unfaulted operation conditions into a memory matrix (MM), X, of n memory vectors and p signals. Each row is an observation of normal behavior.

$$X = \begin{bmatrix} X_{1,1} & \cdots & X_{1,p} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \cdots & X_{n,p} \end{bmatrix} \quad \text{2-8}$$

When an observation state or vector is input to the model, the distance between the input and each memory vector can be calculated. The weight of each memory vector is based on the distance, in this case Euclidean. Each weight is found using a Gaussian kernel of the distance, where d is distance and h is the kernel bandwidth. The farther a memory vector is from the input, the less weight it has.

$$w = \frac{1}{\sqrt{2\pi h^2}} \exp\left(-d^2/h^2\right) \quad \text{2-9}$$

The weights are normalized over all memory vectors, and the product of the weights and memory vectors provides the model prediction estimate.

$$\hat{x} = \frac{\sum w_i X_i}{\sum w_i} \quad \text{2-10}$$

The predictions are weighted averages of the MM that weigh more heavily those memory vectors that are more similar to the input vector. One of the model's key parameter is the kernel bandwidth, h, which controls the distance from the kernel center that the model interpolates between vectors. With a large bandwidth, the AAKR may take longer to detect a fault, but it interpolates within the MM. With a bandwidth too small, the AAKR can be overfit to the MM and magnify the noise's importance. This bandwidth can be optimized by calculating the test error over a range of bandwidths to minimize the prediction error.

Because the AAKR is a weighted average of the MM, the model should correct the inputs' values, even if a fault is in the system. The residuals, the difference between the model and the system, should stay near zero under unfaulted conditions. However, under faulted conditions, the residuals should change over time. A fault-detection system can detect this change.

A fault detection algorithm can use a variety of statistical approaches to analyse residuals. Ideally, a fault-detection system should serve two functions. It should alarm every time there is a fault and if and only if there truly is a fault. A false positive, or false alarm, is an alarm when there is no fault. A false negative, missed alarm, occurs when the detection fails to alarm in the presence of a fault. Fine tuning by reducing the instances of both errors is a persistent focus in designing any fault-detection system.

One simple example of a fault-detection system is triggering an alarm when the residuals exceed some predetermined tolerance. If the residuals' windowed mean crosses the set threshold, it may alarm. It can also alarm if the noise variance exceeds a noise tolerance. These thresholds can be set fairly high to avoid false positives. However, with greater a higher threshold, the alarm may take longer to trigger.

A more sensitive approach is the Sequential Probability Ratio Test (SPRT), a statistical method developed by Wald [Wald 1947], that uses the likelihood of being in one condition over the other as shown below.

$$L_n = \frac{probability\ of\ observing\ x_n\ given\ H_1\ is\ true}{probability\ of\ observing\ x_n\ given\ H_0\ is\ true} = \frac{p(x_n|H_1)}{p(x_n|H_0)} \qquad 2\text{-}11$$

This expression is the likelihood ratio that the data exists in the unfaulted mode H_0, rather than the faulted mode H_1. If the state x is monitored for every time step n, the likelihood's history is kept. If the likelihood's history is summed, S, over each n, then a dynamic estimate of system diagnosis can be developed. From this estimate, a decision criteria can be imposed based on setting thresholds.

if $S \geq B$: accept H_1

if $S \leq A$: accept H_0

$$A = \ln\left(\frac{\beta}{1-\alpha}\right) \qquad \text{and} \qquad B = \ln\left(\frac{1-\beta}{\alpha}\right) \qquad 2\text{-}12$$

The A and B boundaries are based on the set expected false alarm, α, and missed alarm, β probabilities. By adjusting these two parameters, a suitable SPRT detection can be fine-tuned.

At the boundary between unfaulted and faulted conditions, false and missed alarms are unavoidable. False alarms can be a nuisance to plant operators. To eliminate false alarms, a consolidated fault hypothesis selection rule is added as a buffer. The fault hypothesis observes a window of multiple observations. If one alarm triggers within the window, the fault hypothesis can be designed to not identify the fault. However, if the number of alarms exceeds the fault- hypothesis selection threshold, then the faulted condition can be assumed. For example, if the fault hypothesis is set to "3 out of 5," then 3 alarms within a window of 5 observations would trigger the alarm. In other words, isolated instances of spurious alarms are an expected part of detection. However, consecutive alarms within a window would trigger the fault hypothesis decision-making.

2.3.2 Prognostics Parameter Generation

Once a fault is identified, the residuals can be used to quantify the state of system health, the prognostic parameter. It's a useful estimate that can be extrapolated to predict future degradation. While there are other ways of producing this parameter from different sources of prognostics information, the default method used in this study is a system-monitoring model to collect residuals after a fault is detected. These residuals are linearly combined to yield a single parameter degradation estimate.

To deploy this model, the weights of each source can be selected using optimization algorithms. The performance metrics of a particular set of weights can be compared to produce prognostic parameters that are useful, predictable, and reliable. This overall fitness can be a linear combination of three suitability metrics: monotonicity, prognosability, and trendability [Coble and Hines, 2009].

$$Monotonicity = mean\left(\left|\frac{\#\,pos\ {}^{d}/_{dx}}{n-1} - \frac{\#\,neg\ {}^{d}/_{dx}}{n-1}\right|\right) \quad 2\text{-}13$$

$$Prognosability = exp\left(-\frac{std(failurevalues)}{mean(|failurevalue - startingvalue|)}\right) \quad 2\text{-}14$$

$$Trendability = \ min_{i=1:n,j=1:n}\left(\left|corrcoef_{i,j}\right|\right) \quad 2\text{-}15$$

Monotonicity is a measure that the prognostics parameter moves in one direction, up or down. Prognosability is a measure that the prognostics parameter for each case ends in values close to each other, compared to the distance away from the starting value. A population with a tightly clustered end-of-life parameter that's high in value has good prognosability. Trendability is how much each case shares the same underlying shape, or functional fit.

Determining an optimal set of weights to construct the parameter can be challenging. One method uses a regression model called ordinary least squares (OLS), which creates a set of weights that transforms the signals into the univariate prognostics parameter. When the OLS is used, the prognostics parameter is assumed to have the general linear form

$$Y_i = X_i\beta + \varepsilon_i \quad 2\text{-}16$$

where Y is the prognostics parameter, X is the input data matrix of signals, β is the vector of weights, and ε is the noise. This model also assumes normally distributed errors of variance that are identical and independent. The equation expressing the expected value of y can be estimated by a linear combination of input X. The OLS solution can be found using the pseudoinverse equation

$$\beta = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T Y \qquad 2\text{-}17$$

In this application, X is populated by the concatenated signal residuals of every run to failure. The response Y for each run is a line for the run's duration, from 0 to 1. This OLS is then optimized to a linear combination of residuals such that the prognostic parameter is as close as possible to a straight line from 0 to 1, based on the sum of squared error. A straight line was chosen as a simple representation of a degradation path's desirable metrics. It is completely monotonic. The fact that each run ends at around 1 is an attempt to force the POF to end in a tight cluster around 1, increasing the prognosability. Each run's trends are also kept consistent to approximate a straight line. This OLS method, however, does not necessarily end in a completely linear shape of the prognostics parameter. In most cases, the line can still have a pronounced curve, leading to quadratic fits, when the prognostics parameter is evaluated.

Alternatively, a genetic algorithm approach can be used to optimize the signals' weights to form the prognostics parameter [Coble and Hines, 2009], using the suitability metrics as fitness. A fitness function compares the performance of individuals. In this case the fitness is a linear combination of the monotonicity, prognosability and trendability. A default approach is to take the three metrics as equally weighted. Thus the fitness function can be a sum of the three metrics. The genetic algorithm starts with a population of potential sets of weights. Based on the fitness, the algorithm selects the fittest individuals to pass on their characteristics to a new generation. The breeding and the injection of random mutations introduce variations within the population. Over successive generations of breeding, the individuals become more fit and better

combinations of weights can be identified. A suitable stopping criterion is imposed when there is no longer a sufficient change in performance between generations, when a set number of generations is exceeded, or when a sufficiently fit component is found. This algorithm intelligently searches through very large potential spaces and can find the global maximum of fitness. However, some minor drawbacks exist. If the algorithm is not properly tuned, it can be subject to staying in local maximums. This drawback is evident when different starting populations result in different individuals selected. Another minor drawback is the algorithm's computational time and the investment in testing multiple trials.

To alleviate some of these limitations, candidate individuals can be considered in the starting population of searched individuals. If an individual, in our case defined by the linear set of weights of input signal residuals, is estimated by other means, it's likely to be more fit than the other randomly selected individuals. Increasing the fitness of the first generation may reduce convergence time. Even if one or two initial estimates are bad, they should be filtered out by the next generation. One candidate that can be considered is the OLS estimate given in equation 2-17.

2.3.3 General Path Model

A common means of using this prognostic parameter to predict a system's future degradation is through the General Path Model (GPM) approach [Lu and Meeker, 1993]. When using the GPM, an OLS is fit to the degradation parameter and extrapolated until it crosses the failure threshold. Typically, this failure threshold is based on historical failures, but need not directly indicate a point of catastrophic failure. The failure threshold can be set as any point where a system no longer conforms to the necessary specifications and demands placed upon it.

While different methods exist, an OLS regression model was used. Time, or a function of time, is used to estimate the prognostic parameter. As stated previously, this model assumes that the error term, ε, of the parameter is normally distributed around

zero. The OLS regression on a joint Gaussian distribution of parameters gives the maximum likelihood (ML) estimate, the point at which the likelihood distribution is at a maximum. By applying OLS to a degradation path, an approximate fit can be found.

For a certain failure mode, the degradation paths are assumed to follow similar fits. Therefore, a suitable fit (e.g. linear, quadratic, and exponential) can be chosen by examining the errors between the fit and the data. For example, linear and quadratic fits can be compared for one path. The difference between the fit and the data can be squared and summed to give the squared error's sum. The fit with the lower sum of squared error is selected as more accurately describing the path towards failure.

The prognostics parameter is attained using monitoring and prognostics-parameter generation, which are described in the previous two sections. In summary, a model of the system is created using an AAKR. The signal residuals of a system heading towards failure can be found and combined to form the prognostics parameter.

An example of the GPM is given in Figure 2-4. At the beginning of life, no fault has been detected, and the blue prognostics parameter stays near zero. However, as degradation accumulates, the parameter moves towards the failure threshold, the horizontal yellow line at above a parameter of 1. This path is measured until time 150, when the data is censored. Based on the current path, a quadratic fit, red parabola, is given to the path, which is extrapolated to the degradation threshold. As shown in Figure 2-4, the path crosses the threshold at about time 280, giving the estimated TOF.

For this research, the prognostics and monitoring methodologies were performed using the MATLAB PEP and PEM toolboxes developed at the University of Tennessee by Dr. Jamie Coble [Coble and Hines 2009] and Dr. Dustin Garvey [Garvey and Hines 2007]. The PEM toolbox was used to create the AAKR models, fault detection, and residuals. The PEP toolbox was used to calculate the estimated RULs. In addition, modified PEP type III GPM functions were created to allow for the Bayesian transitions. The modifications included Bayes updating functions, expanded function inputs, and

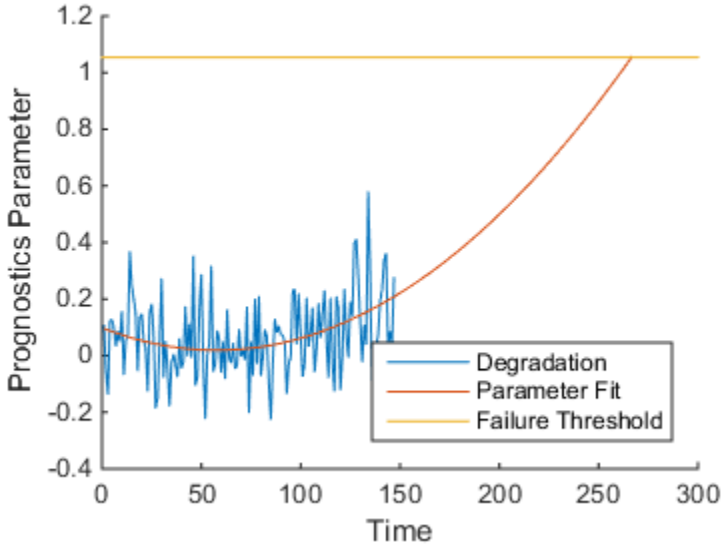


Figure 2-4: Test Case Heading Towards Failure

alternative executions of RUL estimation and uncertainty calculations.

2.3.4 Alternatives to GPM

It's important to note that the methods described in the previous sections are not the only degradation-based models. While the GPM depends on linear regression models, other models have been applied to find suitable solutions for quantifying and extrapolating degradation. Nonlinear paradigms include pattern recognition, machine learning algorithms, and adaptive regression splines. With respect to transitioning between the types, the specific type III model does not matter as long as an RUL distribution can be found. The GPM was selected as an easily optimized and flexible approach to calculating predictions. Most aspects of the GPM can be optimized with limited user intervention, reducing arbitrary decision making and trial-and-error testing. Advanced machine learning algorithms can require fine tuning before they produce proper results.

There has been interest in using various Neural Network (NN) methods [Connor and Martin 1994; Upadhyaya et al., 1994; Chinnam 1999; Jaw 1999] and neuro-fuzzy inference [Husmeier 1999; Liang and Liang 2006]. The types of NN frameworks are feed-forward and recurrent models. In a feed forward NN (illustrated in Figure 2-5), with back propagation, the inputs are fed into the NN. Each node in the input layer sends information to each node in the hidden layer, with a weight w_{hj}. The notation represents the weight of each input node j into each hidden node h. An additional input bias node, usually of value 1, is included in the figure. The hidden nodes then move information forward to the output layer in a similar fashion; each hidden node z_k, weighted by v_{ik}, is input into each output node.

The weights between nodes are optimized by back propagating the error from the output through the hidden layer and back to the inputs. In a recurrent scheme, the connection between nodes forms closed directed cycles, creating an internal NN state

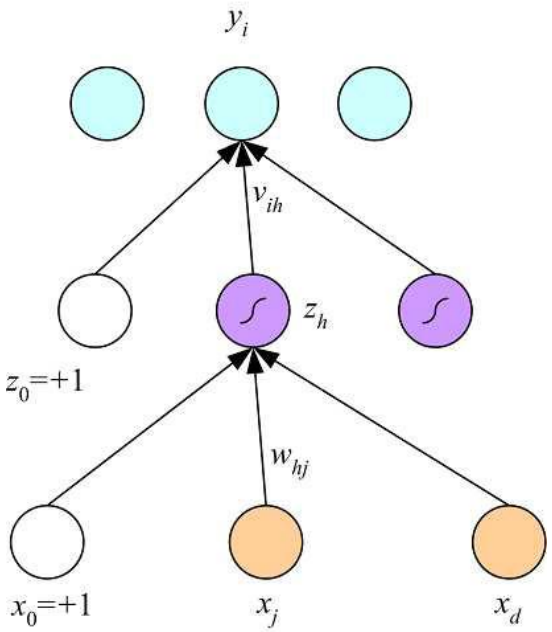


Figure 2-5: Single Hidden Layer Feedforward Neural Network

that can change with time. A neuro-fuzzy system is a combination of NN and fuzzy logic, which assigns weighted true values between 0 and 1, in contrast to Boolean logic, which assigns 0 or 1.

An alternative to both the GPM and NN for type III prognostics is particle filters [Orchard et al., 2008; Saha et al., 2009]. This state-representation scheme uses discrete data-modeling techniques to calculate a future state's probable distribution. By subsequently predicting states, the current degradation path can be extrapolated towards failure.

The history of particle filters has its roots in molecular chemistry [Rosenbluth and Rosenbluth, 1955]. The terms particle filter [Moral et al., 1996] and sequential Monte Carlo [Liu and Chen 1998] were coined later. While the groundwork for what we know today as particle filters was laid by Gordon's paper [Gordon et al., 1993], the following derivation and computation is taken from Arulampalam's paper [Arulampalam et al., 2002], one of the most widely cited and accessible representations on the matter.

These particle filters can vary based on their exact implementation, though usually a sequential importance sampling algorithm is included. The importance sampling helps to draw future predictions based on a prior idea of where the states are headed. This process is described as a hidden Markov model [Baum and Petrie, 1966] because the system state is assumed to be unobserved. The measurement equation can then be expressed as a function of the hidden state and measurement error.

To solve this hidden Markov problem, i.e. predict likely future states, a population of particles are used to estimate the posterior distribution based on noisy measurements. The importance sampling draws a population of particles proportional to a prior likelihood distribution, which can be obtained from constructing the state space and measurement equations. When a new measurement z is drawn, the likelihood and previous generation of weights produces the new set of weights, which are then normalized.

One issue that arises in particle filtering is degeneracy, where after enough iterations have passed, very few particles contain most of the weight. It can also be referred to as the effective sample size. The effective sample size refers to the outliers' decreased importance as a natural occurrence of this algorithm. If degeneracy falls too far, then only a few particles will hold the weight. If a degeneracy threshold is set, then a resampling algorithm can be implemented when the effective sample size drops below the threshold.

Once the particle filter is defined, the RUL calculation can also vary. The simplest methods extrapolate out subsequent states until the degradation threshold is exceeded. Other methods use particle filters to sample degradation path coefficients. Similar to the GPM, the coefficients are used to extend the path to failure.

2.4 Previous Methods for Model Fusion

The idea of fusing different sources of information either directly or through multiple prognostic methods is not a new concept. At almost every juncture of analysis, starting with redundant sensors [Orsagh et al., 2004] and ending with RUL estimates, different data sources can be merged into a holistic approach. However, while this research's goal is to fuse multiple prognostic methods to obtain one combined estimate, other work has focused on fusing information at other stages leading up to prognostics, including signals analysis, monitoring, and diagnostics. The potential for data fusion in these areas has been explored, but is not this research's primary concern.

For example, Orsagh et al. [2004] used multiple concepts of fusion when predicting RUL for gas turbines. Fusion of multiple sensors happens at the lowest level. At a higher level, extracted features are combined into a degradation model. Using different prognostic methods before and after a fault is detected is another suggested

fusion method. These concepts are assumed when applying system monitoring, as discussed in section 2.3.1.

In contrast, attempts have been made to fuse different prognostic methods or to apply data fusion in a way that enhances prognostics. Garga et al., [2001] combined the automated reasoning method with neural networks. While a neural network produces prognostic results, it also uses information from a diagnostic system. Goebel et al., [2005] described several methods of RUL prediction, including weighted averaging, adaptive neuro-fuzzy inference, and Dempster-Shafer to fuse different measures of uncertainty. The use of weighted averaging to produce RUL estimates is an obvious transition. The errors associated with each prediction have the potential to cancel each other out, and including both methods stabilizes the combined result. By using the total error, the weights can be assigned to give greater importance to the better performing models. These weights can be found by minimizing the prediction errors using OLS regression.

These are similar to weighted sum models, based on those originally introduced by Bates and Granger [Bates and Granger, 1969; Rao 2000; Menezes et al., 2000], who proposed that the combined sum of two prognostic models produces more stable results than separately. These methods can be used to fuse future states' predictions, which are extrapolated to find an RUL estimate based on multiple-state predictions.

2.5 Bayesian Applications to Prognostics

This section summarizes Bayesian applications to prognostics in other than the proposed Bayesian transition methods in the next chapter. The first method improves the basic GPM model using a Bayesian regression scheme. Prior observations of

regression coefficients are used to update the OLS model. Section 2.5.2 gives a brief overview of other Bayesian applications, but do not involve prognostic model fusion.

2.5.1 Updating Degradation Path Parameters [Coble and Hines 2011]

Previous research has shown that Bayesian regression models can be applied to the prognostic parameters in the GPM. This method uses information from previously observed failure paths to form beliefs. Instead of estimating the path regression coefficients according to an OLS with equal variance, equation 2-16, an analogous version using a covariance matrix can be applied. This application models different measures of uncertainty for each data point.

If the variance of the OLS with equal variance is known, then the conditional posterior distribution of the coefficients β can be defined as a normal distribution [Gelman et al., 2004]. Variable V multiplied by the variance yields the coefficient covariances.

$$\beta|\sigma^2, y \sim N(\hat{\beta}, V\sigma^2) \quad \text{2-18}$$

$$V = (X^T X)^{-1} \quad \text{2-19}$$

The parameter estimate is found by minimizing the sum of squared residuals (SSR) between the fit and data. These residuals can be simplified to

$$SSR(\beta) = \sum(y_i - y_i^{pred})^2 = (y - X\beta)^T(y - X\beta) \quad \text{2-20}$$

The common solution, solving for regression coefficients by minimizing the SSR, is the Penrose-Moore pseudoinverse [Moore 1920; Penrose 1955]. By taking the derivatives of both sides with respect to β and setting them equal to zero, the squared residuals' sum is minimized. The left-hand side was found using product rule:

$$-2X^T(y - X\beta) = 0 \quad \text{2-21}$$

This can be rearranged and solved for β as shown below:

$$\hat{\beta} = (X^T X)^{-1} X^T y \quad \text{2-22}$$

A classical statistics approach can be used to estimate the model's variance:

$$\sigma^2 = \frac{1}{n-k}(y - X\hat{\beta})^T(y - X\hat{\beta}) \quad \text{2-23}$$

A data covariance matrix Σ can be introduced to incorporate unequal variances in a generalized least squares model (GLS) [Aitken 1934]. Instead of assuming equally distributed errors, $\sigma^2 I$, the covariance matrix is an n x n symmetric positive matrix. The minimization function attempts to minimize the sum of squared residuals, subject to weighting by the covariance matrix:

$$SSR(\beta) = (y - X\beta)^T\Sigma^{-1}(y - X\beta) \quad \text{2-24}$$

Repeating the process of deriving the expression and setting equal to zero, the previous equations are then replaced by their analogous versions:

$$\hat{\beta} = (X^T\Sigma^{-1}X)^{-1}X^T\Sigma^{-1}y \quad \text{2-25}$$

$$V\sigma^2 = (X^T\Sigma^{-1}X)^{-1} \quad \text{2-26}$$

$$\sigma^2 = \frac{1}{n-k}(y - X\hat{\beta})^T\Sigma^{-1}(y - X\hat{\beta}) \quad \text{2-27}$$

This weighted regression model gives estimates based on each data point's uncertainty.

The X, y, and Σ matrices can be appended using prior information weighted by its uncertainty as shown below:

$$y_* = \begin{bmatrix} y \\ \beta_{j0} \end{bmatrix}, X_* = \begin{bmatrix} X \\ I_p \end{bmatrix}, \Sigma_* = \begin{bmatrix} \Sigma_y & 0 \\ 0 & \sigma^2_{\beta_{j0}} \end{bmatrix} \quad \text{2-28}$$

The y is appended with the estimated j^{th} coefficient based on previous failures. The X is appended with an identity matrix the size of the number of parameters. The covariance matrix weights the prior estimates' uncertainty against the prognostic parameter's noise.

The benefit of this method is that the path parameter's prior expectation can be estimated from data used to build a GPM model in the first place. The GPM requires previously failed cases tracked from fault to failure. Each training case of failure

provides one example of a set of path parameters. With a population of cases, the distribution of those parameters can be characterized.

Including these data points can have a beneficial effect on the stability of path parameter estimation, in turn leading to more stable RUL predictions. Soon after a fault is detected, the prognostic parameter data are limited. At these early stages, the OLS regression can be susceptible to the prognostic parameter noise.

To illustrate the prior information's benefits, Figure 2-6 shows a GPM applied to the same test case shown in Figure 2-4, except censored at time 20. As before, the extrapolated path is assumed to fail when it crosses the horizontal failure threshold. However, because it was censored early in life, the parameter's noise causes the OLS to fit a concave downward quadratic function. The result is a nonsensical imaginary RUL estimate, as the parabola never crosses the threshold due to lack of prognostic information obtained through sensor residuals alone. This issue can be remedied by including prior information about the regression coefficients themselves, shown in Figure 2-7. If a population of training cases were used to create the GPM, the sets of regression parameters could be stored and used to create PDFs of each regression parameter. This prior distribution is included in the OLS regression model to influence the path towards what is expected, based on prior knowledge.

Because the regression parameter estimates for the censored case deviate far from expected behavior over a small number of recorded data points, the prior distributions of the parameters heavily influence the projected prognostic parameter's shape at the beginning of life.

This update, including prior coefficient information into the regression, shows a clear improvement over the GPM alone. The data used to create the GPM is sufficient to provide the necessary information needed to apply this regression update. Therefore, if situations arise in which including the update greatly improves RUL estimates, there is little reason not to implement the update. In situations where both models' RUL

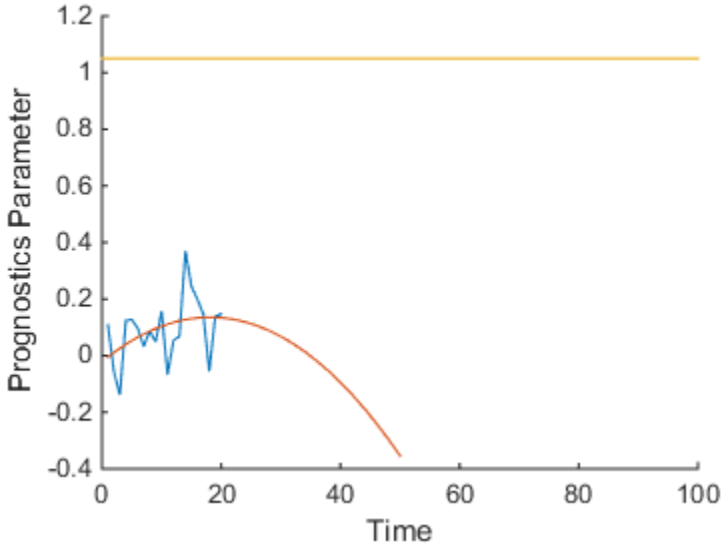


Figure 2-6: Test Case Censored at Time 20

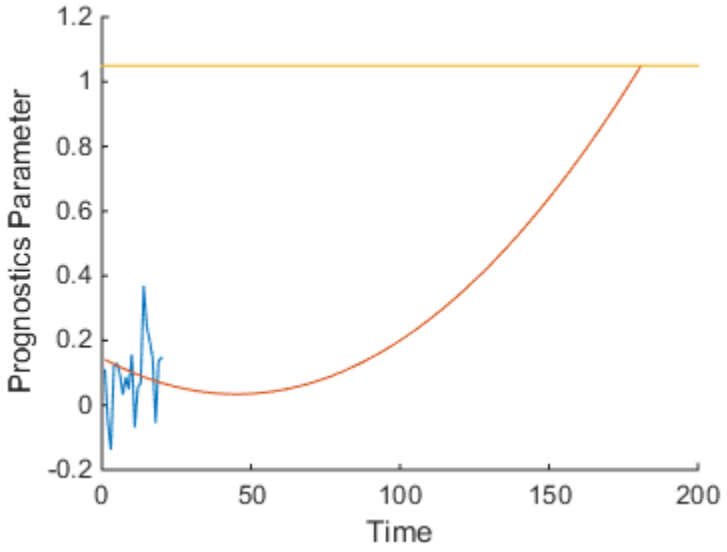


Figure 2-7: Censored Case Using Coefficient Priors

estimates are similar, the only lost resource is in the computation time of storing and applying the updates based on the prior parameter's PDFs. This increased time is negligible in most situations.

2.5.3 Other Applications of Bayesian Theorem

This section provides an overview of methods that incorporate Bayesian statistics, but not for the purposes of prognostic model fusion. These methods were suggested in the discussion of particle filters, in section 2.3.4. For example, Bayesian model averaging has been used to select variables in additive Cox PHM [Volinksky et al., 1997]. Bayesian updating has also been applied to type I Weibull survival analysis [Guure and Ibrahim, 2012].

In one analysis, Bayesian priors were used for degradation path projection, but in a slightly different framework from the OLS model presented here [Karandikar et al., 2013]. The degradation was quantified and extrapolated using OLS regression with a quadratic fit. Wear data of tools at certain times were drawn from defined distributions to create samples of degradation growth curves. Therefore, when calculating the RUL along the path, the future state was estimated using the defined distributions as priors. This method is similar to the research presented in this paper, in that Bayesian updating directly affects OLS regression models. However, the specific variable on which Bayesian updating was used differs.

Particle filtering algorithms have been widely used in prognostics [Orchard et al., 2008; Zio and Peloni, 2011], as previously mentioned in section 2.3.4. These applications estimate future states, usually of degradation, using particle filters. One modification of the algorithm predicts lithium-ion battery failure using a particle filtering algorithm to update regression path parameters [He et al., 2011]. The parameters were first initialized using Dempster-Shafer theory, which assigns masses with belief weights. When certain parameter estimates agree based on their confidence intervals, those estimates have a combined belief and are given more weight when calculating the

combined estimate. It's then assumed that the iterative parameters follow a known Gaussian distribution. The parameters are then updated at every new observation according to the particle filtering algorithm. Particle filtering is used to estimate the parameters at future instances.

2.6 Summary of Lifecycle Prognostics Methodology

RUL's accurate estimation is crucial for optimizing maintenance strategies. Often one of prognostics' bottlenecks is the availability of data. To get the most use out of existing data, multiple prognostic models can be applied. These are categorized into three types depending on the data they use and at what point in a component's lifecycle the data becomes available.

Type I prognostics uses previous instances' TOF information to predict the unfailed component's expected failure time. Type II methods consider operating conditions. The more severe the operating conditions, the sooner components are likely to fail. Type III uses directly monitored signals to estimate the overall degradation, which is extrapolated to a failure threshold, at which point the component is assumed to fail.

To monitor degradation, a system model can first be made by using an AAKR model based on unfaulted data. Memory vectors are stored in a memory matrix. Signal sets input into the AAKR are measured as a Euclidean distance away from each memory vector to find each vector's weight. The memory vector's weighted sum outputs the corrected signal. If no anomalies are in the input data, the residuals should resemble Gaussian noise centered at zero.

If a fault occurs, the output should return the corrected signals of how the model behaves without a fault. The difference between the input and the model are the

residuals, which can then be linearly combined to form a single prognostics parameter. The three metrics by which to judge a desired prognostics parameter are monotonicity, prognosability and trendability.

Other methods for data fusion for prognostics have been proposed. However, most do not deal with fusion between prognostic models. In many instances data from different sources can be fused into an intelligent prognostic framework, including fusion in the monitoring and diagnostics analysis before the prognostics begins. A fusion method can involve an application of Bayes' theorem to a single prognostic model, such as in the particle filter algorithm and degradation path parameter update.

Benefits of fusing prognostic models have been seen in other applications. In contrast to these other methods, the novel research presented in the following chapter involves a Bayesian approach to merging model types. Bayesian statistics deals with the uncertainty of random variables, and not just one point of expected value. It naturally combines measures of uncertainty for different information sources. The combination of models, with their analysis and data, can produce better results than each of the models separately.

Up to now, the necessary background of applying prognostic methods to failure data in order to calculate RUL estimates has been introduced. Using these prognostic models, the following chapter demonstrates how each model can be applied over the lifecycle of equipment. Using different available sources of data, the models can be tied together using different Bayesian transitions.

3. LIFECYCLE TRANSITION METHODOLOGY

The transition models' mathematical bases are presented in this chapter to ensure the new methods' validity when applied to a lifecycle prognostics framework. To develop these new methods, a Bayesian approach was used as a method for quantifying beliefs. A belief is a PDF of what we expect the parameter of interest to be, based on prior knowledge. An a priori belief is the belief in a parameter's distribution before new information is observed. It is updated with new sampled observations to form a posterior distribution. While many branches of Bayesian statistics derive from the Bayes theorem, the theorem itself can be used directly. It weighs the prior belief with new information taking into consideration the PDF's shapes.

Bayes theorem is useful in many ways. It differs from classical statistics by primarily addressing the PDFs of a parameter of interests, instead of point mass estimates with confidence bounds. For instance, one classical approach to quantifying the RUL estimate is to calculate the mean and median with an upper and lower 95% confidence interval. This approach is in contrast to representing the RUL as a PDF. For situations like the type I survival analysis model, the RUL's PDF is often a known functional fit, such as the conditional Weibull model. This representation conveys more information about the RUL prediction, as it captures the expected value and confidence interval while including information about the shape. The theorem is a more detailed approach to uncertainty.

The specific method for computing the posterior can vary depending on the nature of the PDFs. The main issue involves the PDFs' representation. Because empirical methods use discrete analysis of continuous space, the PDFs can be expressed as functional fits or as numerical approximations. For example, a RUL distribution can be binned over small intervals, or a Weibull fit can be found. Based on the data's nature, several computational methods are introduced, including numerical approximation,

conjugate distributions, and a Metropolis-Hastings (MH) algorithm. Each method should produce very similar results, but can be thrown off for various reasons. If proper care to statistical fluctuations, computation, and suitability of functional fits is considered, the results should be nearly the same.

Before the transitions are introduced, the flow of information starting from data collection and ending in combined RUL estimates can be tracked, as seen in Figure 3-1. To reiterate several points made in the last chapter, the input data can come from different sources. The prior RUL can be collected from maintenance logs, isolating the TOF, or supplied by the manufacturer. The conditional priors contain information about the system's operating conditions. That information is superior data to the prior RUL because it is more individual-specific. The idea is that high-stress conditions wear out components more quickly than low-stress conditions. Measures of operating conditions can include temperature, capacity, vibration, and exposure to moisture. Degradation priors are collected when a system is tracked towards failure. The deviations from normal behavior can be collected into useful health information, such as a prognostic parameter.

In Figure 3-1, the sources of information indicated in blue are used to create different prognostic models. The TOF and prior conditions are input into type I and II models, respectively. The choice of each model does not have a large impact on which transition to use. For the type III model, the GPM is the main focus of this dissertation. While other type III models can be used, they can't take advantage of all the methods developed here.

To transition among prognostic models, three different transition methods are proposed. The three red rectangles in Figure 3-1 represent different applications of Bayes' formula depending on the nature of the prognostic model applied. The first and most general case of the transition is referred to as a RUL update. For any two RUL PDFs, Bayes' formula can be used to form the posterior distribution. The RUL regression

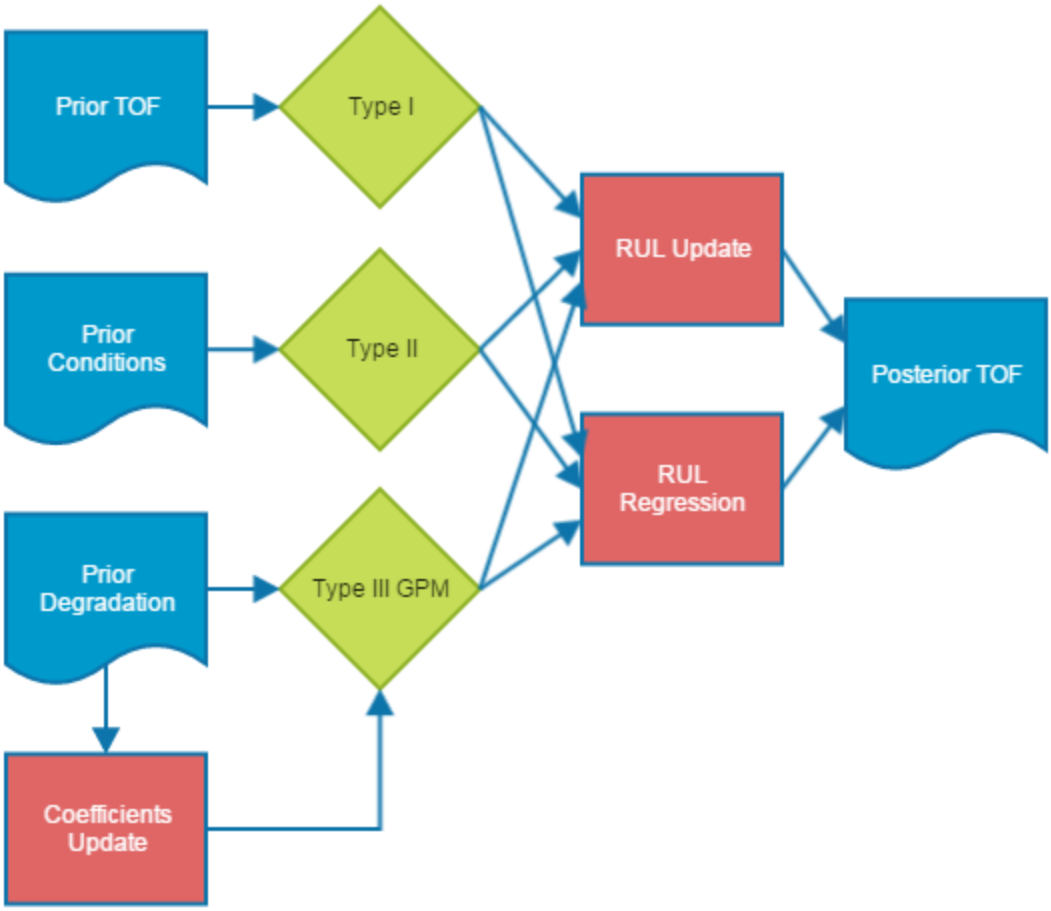


Figure 3-1: Information Flow for Transitions Among Prognostic Models

is used to transition from either type I or II to a type III GPM. In addition, the coefficients update can be used as part of the GPM. The posterior TOF is found using either the RUL update or RUL regression transitions. The figure illustrates that this posterior TOF benefits from more than one source of degradation information, and combines multiple prognostic models.

The RUL update is a direct application of Bayes' formula on the RUL distributions of each model. This transition is the most flexible and broadly applicable as it can be used to transition between any two types of prognostic models, as it only deals with the results of both. For the type III, the GPM can be substituted for any other type III model. This is not the case for the next transition.

The RUL regression is applied when the type III model was developed as a modification to the GPM, which estimates regression coefficients in order to extrapolate future states' degradation. It includes additional information about the expectation of future degradation into the regression. In this case, the prior information is a direct result of a previous model. A type I RUL distribution output contains useful information for a type III GPM model. The type I model estimates the time; and based on the GPM, a case is assumed to fail when the degradation exceeds the failure threshold. By combining these two estimates into one data point, the prior can be included in regression models. However, one issue that arises is treating the uncertainty. The GPM uses OLS regression to solve for the coefficients. This OLS assumes an error only in the dependent variable, degradation. However the additional prior data point supplied by the prior RUL has uncertainty in the independent variable, time, in addition to degradation. The OLS can't model both sources of uncertainty. To solve this issue, a weighted total least squares (WTLS) regression model is introduced that models noise variance in both dimensions.

Finally, a coefficient update model is presented. While this transition does not involve the transition among prognostic models, it is included as a novel and useful

implementation of Bayes statistics to form improved predictive models. If the population of previous degradation paths is regressed, the distribution of regression coefficients can be used to form priors for a new test case. The GPM's coefficient estimates are updated with the prior distributions of expected path coefficients to form a posterior degradation path. While section 2.5.1 shows a similar method, this one differs in the way the coefficient priors are updated. The previous method included the priors within a modified OLS framework to produce coefficient estimates. This method uses the OLS to produce the coefficient estimate. It updates this estimate using the priors and Bayes rule.

In summary, the available data is used to create multiple prognostic models. Once created, the RUL update or Bayesian regression transitions can be used to produce combined RUL estimates that incorporate all the known information. The type III GPM can be used in both transition schemes. If substituted, the RUL update can still be used. In addition, the GPM can use a coefficient update transition to supplement the GPM algorithm.

3.1 Bayesian Statistics

3.1.1 Bayes Formula

Bayesian statistics is a branch of statistics, distinct from classical statistics, based primarily on Bayes formula [Ghosh et al., 2006].

$$\pi(A|B) = \frac{\pi(A)f(B|A)}{\int_A \pi(A')f(B|A')dA'} \qquad 3\text{-}1$$

Equation 3-1 calculates the conditional PDF of A, the parameters of interest, given new data B. The prior density function, π(A), is a prior belief or estimate, with some characteristic distribution, about the parameters. The f(B|A) is the likelihood of B, the data, interpreted as the conditional density of B given A. In other words, f(B|A)

represents the likelihood of observing the data B, if A were true. The numerator is the joint density of A and B, while the denominator is the marginal density of B, or prior predictive distribution. It is considered prior because it does not depend on the data B and predictive because it describes a quantity that is observable. The denominator is essentially to normalize the probability. If the parameter of interest is discrete then the integral is replaced with a summation. Both the sum and integral are integrated over all possible values of A.

Bayes formula, when solved, gives the posterior density, a quantification of uncertainty about A in light of new data B. The transition from $\pi(A)$ to $\pi(A|B)$ is what is learned from the data.

There are several ways to evaluate Bayes formula. In practice the direct application of the formula is rarely used. Instead, several alternate methods are available that approximate the posterior to very high accuracies. The methods suggested in this paper can be categorized: conjugate families, numerical approximation, and the Metropolis-Hastings (MH) algorithm. An empirical fit can be made using a kernel density smoother when a proper distribution fit can't be found.

3.1.2 Analytical and Conjugate

To introduce why multiple algorithms may be necessary, it's important to clearly state that a direct computation of Bayes formula to arrive at an analytical solution is in most cases not necessary or desirable. This is because an analytical integral of the denominator can be complex or at worst unsolvable. It can be especially difficult for multidimensional fits. Solved instances of the analytical solution that use convenient distributions are called conjugate families.

To contrast between the analytical and conjugate, suppose that a type I survival RUL distribution can be approximated using a Gaussian fit [Gauss 1809]. This prior can be quantified by the mean and standard deviation.

$$\pi(\mu,\tau) = \frac{1}{\sqrt{2\pi\tau^2}}\exp\left(\frac{-(\mu-\eta)^2}{2\tau^2}\right) \quad 3\text{-}2$$

The μ represents the mean of the expected value of the RUL. The standard deviation τ is a quantification of both the uncertainty of the prior and how much the mean is expected to change based on future observations.

The prior mean and standard deviation η and τ, is updated with a dataset X. If X is the RUL estimate from a type II or III model and can be approximated normally, it's represented as another Gaussian.

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(X-\mu)^2}{2\sigma^2}\right) \quad 3\text{-}3$$

According to Bayes formula, the numerator is the product of the prior and likelihood. The denominator is the integral of that product, in order to normalize the posterior PDF.

$$\int \frac{1}{\sqrt{2\pi\tau^2}}\exp\left(\frac{-(\mu-\eta)^2}{2\tau^2}\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(\frac{-(X-\mu)^2}{2\sigma^2}\right)dx \quad 3\text{-}4$$

Bayesian mathematicians have long figured out this particular transition, and have labeled such solved cases conjugate distributions. These distributions form conjugate families. If the likelihood and the posterior are known and fall within a family, then the posterior function can easily be referenced. In this case, the prior and likelihood PDFs form the Gaussian conjugate family. The posterior happens to be predicted using a Gaussian distribution with the following parameters [Raiffa and Schlaifer, 1961].

$$E(\mu|x) = \frac{\frac{\eta}{\tau^2}+\frac{\sum x_i}{\sigma^2}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} = \frac{\frac{\eta}{\tau^2}+\frac{n}{\sigma^2}\bar{X}}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} \quad 3\text{-}5$$

$$var(\mu|x) = \frac{1}{\frac{1}{\tau^2}+\frac{n}{\sigma^2}} \quad 3\text{-}6$$

Equation 3-5 gives the posterior ML estimate of the mean of the distribution, while equation 3-6 gives the variance of the mean, or posterior variance. Though all the variables have been somewhat defined in the context of updating two RUL distributions, they can also be viewed in more abstract terms. For one, it's useful to

distinguish between the mean and variance of each of the Gaussian distributions in the context of updating. The prior hyper-parameters refer to the prior mean and variance, η and τ^2. The prior variance τ^2 measures the strength of the belief in the uncertainty of the prior distribution. The smaller the variance, the more precise the estimate is, as less likely it is to change in future sampling [Ghosh et. al, 2006]. A larger variance gives the prior less weight. In this sense $1/\tau^2$ is the precision of the prior, while n/σ^2 is the precision of n data points. The posterior hyper-parameters can become priors when new data become available. The posterior mean is the expected value of the mean of the likelihood. For the Gaussian, this means that the posterior mean is the weighted average of the prior and sample means with the precisions of each as weights. It also means that with more data, the prior is overcome, and eventually the posterior approaches the sampling data.

Conjugate families can be useful in certain situations but are often limited in application, based on the assumptions made. The usefulness comes from the fact that if all the assumptions hold, then the posterior function can be referenced from widely available sources. These referenced functions lead to straightforward calculations. In the ongoing example, the prior happens to be represented by a Gaussian. The likelihood also happens to be Gaussian. Therefore the Gaussian conjugate can be applied. If however the likelihood is instead a Weibull, then the assumptions don't hold. In this situation, a Weibull with Gaussian does not form a conjugate. Therefore it is shown that only strict sets of distributions can be used, and may not apply to the current situation. If the conjugate is desired enough, the expert can determine whether a less accurate distribution, or a transformation of the distribution, is good enough as to not misrepresent the distribution so that the conjugate can be applied.

An example of a Gaussian conjugate update on RUL estimates is shown, in Figure 3-2. The prior RUL is centered around 100 days. The likelihood, around 110 days. It can

be seen that the posterior is a compromise between the prior and the likelihood. The peak is sharper because the variance of the posterior lies in the interval where both the

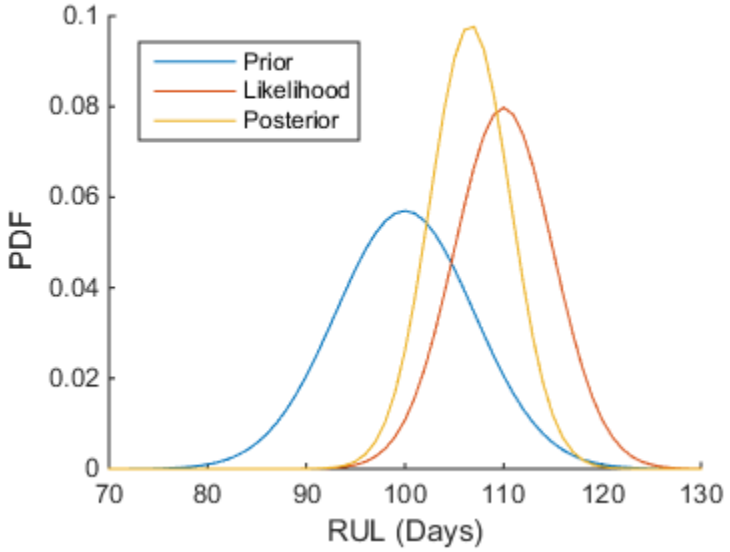


Figure 3-2: Gaussian Conjugate Family Updating for η = 100, τ = 7, μ = 110, σ = 5 Days

prior and posterior are contributing. At the ends of the posterior tails, below 90 and above 120, only one of the two distributions have substantial contributions. The ML of the posterior lies between the two PDFs. Furthermore, this peak is closer to the likelihood than the prior. This is due to the variance of the likelihood being smaller than the variance of the prior.

While the Gaussian conjugate distribution is very simple to implement using conjugate families, when combining RUL estimates, those distributions may not fit the data. If instead the likelihood can be approximated with a Weibull distribution, then the following integral must be evaluated to form the posterior.

$$\int \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(\frac{-(x-\eta)^2}{2\tau^2}\right) \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp\left(-x/\theta\right)^{\beta} dx \qquad 3\text{-}7$$

This integration could pose a challenging calculus problem. It also illustrates the difficulty in integrating the product of any pair of viable distributions. It's even more difficult for multidimensional distributions.

3.1.3 Numerical

Numerical integration approximation methods can solve the problems encountered by the analytical and conjugate. They are flexible; they easily accomodate a wide range of distributions, while retaining a high degree of accuracy. While this method is not very widely practiced, it is theoretically an accurate and straightforward method that can be useful in benchmarking the other algorithms.

If appropriate functional fits can be found for a data set, then the integral can be evalualted using area approximation. Classic examples include trapezoidal and Simpson's rule. These methods estimate the area under the curve using trapezoids and parabolas. By dividing the range over which the distribution is integrated into smaller segments, they become increasingly accurate. By using a large number of segments over a sufficiently large range, the error can be dwarfed by other sources of uncertainty.

Figure 3-3 shows an example updating a Gaussian prior with a Weibull likelihood.

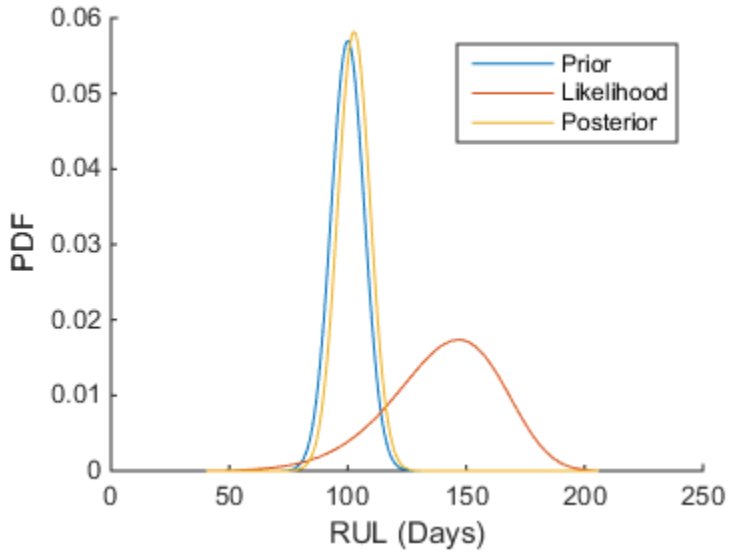


Figure 3-3: Gaussian to Weibull Using Numerical for $\eta = 100$, $\tau = 7$, $\beta = 7$, $\theta = 150$ Days

In this case, the posterior stays within the region that the prior and likelihood both take significant values. Because the likelihood has a larger uncertainty than the previous Gaussian conjugate example, the posterior is much closer to the prior.

3.1.4 Metropolis-Hastings Algorithm

Monte Carlo approximation sampling using the Metropolis-Hasting algorithm [Metropolis 1953, Hastings 1970], is a popular approach [Chib and Greenberg, 1995; Gelman et al., 2004; Ghosh et al., 2006] to sample from multivariate distributions. It's similar to acceptance-rejection sampling, while being closely related to Gibbs sampling.

The algorithm starts by selecting an initial candidate. This starting value should have a non-zero posterior likelihood. This means it should be defined for both the prior and the likelihood. It does not necessarily need to be close to the ML, and usually a burn in period is implemented, in which some beginning samples are ignored. This exclusion has a higher chance of starting the record in a region with higher probability density.

In the Metropolis algorithm, a symmetric proposal (jumping) distribution is applied. This proposal selects the next candidate in a sequence of samples. The exact choice of the proposal distribution may not have a large impact on the performance of the algorithm. The biggest concern is convergence time. A bad proposal distribution requires more samples to estimate accurately. If computation time is not an issue, any range of reasonable distributions should produce a sufficient approximation, given enough samples.

When the candidate is selected, the posterior likelihood of both the candidate and the previous sample are computed. The ratio of the likelihoods is used to determine whether or not the sample moves. If the candidate is more likely than the previous sample, then the sample will move. If it's less likely, it will move with a probability proportional to the likelihood ratio.

Hastings modified Metropolis' original algorithm by assuming that the proposal distribution does not have to be symmetric. This extra flexibility opens the repertoire of

proposal distributions to speed up convergence. As an algorithm, the MH can be defined as follows.

Initialize the starting point

$$x_0 \sim q(x)$$

For i = 1 : n

Select candidate from jumping distribution

$$x_{candidate} \sim q(x_i | x_{i-1})$$

Acceptance probability

$$\alpha(x_{cand} | x_{i-1}) = \min\{1, \frac{q(x_{i-1} | x_{cand})\pi(x_{cand})}{q(x_{cand} | x_{i-1})\pi(x_{i-1})}\}$$

$$u \sim Uniform(0,1)$$

If u < α then

$$x_i \leftarrow x_{cand}$$

Else

$$x_i \leftarrow x_{i-1}$$

Such a MH algorithm with a Gaussian proposal distribution was made for the Gaussian to Weibull RUL example, Figure 3-4. The proposal distribution was set as a normal distribution centered on the previous sample, with a variance equal to the prior variance. This gives an appropriate scale to the jumping distribution that will select distances of the same magnitude as the prior distribution. The results are slightly different than that presented in Figure 3-3. The location of the posterior peak slightly shifted. However overall they are very similar.

One advantage of the MH algorithm is that it can be used to sample from multidimensional distributions. There are two ways this multidimensional sampling is implemented. Blockwise updating uses a jumping distribution with the same dimensions as the target distribution. The algorithm is the same as the MH presented above, while keeping in mind q(x) is multidimensional.

A slightly more sophisticated version is the componentwise update. This is

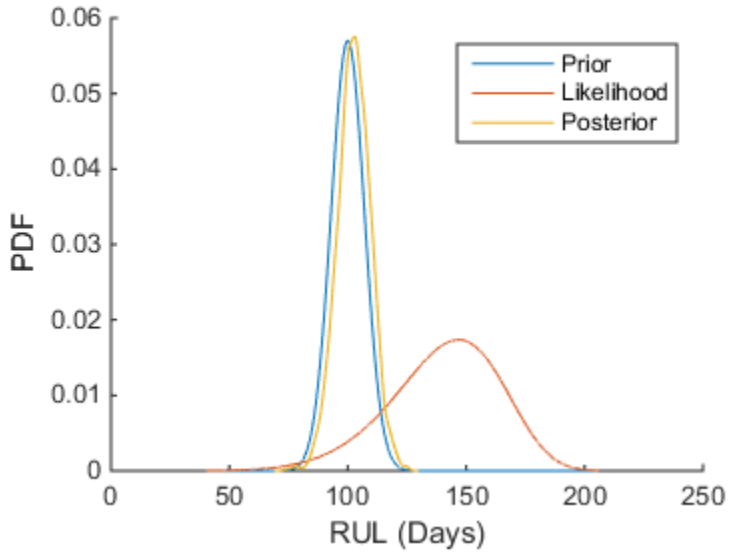


Figure 3-4: Gaussian to Weibull Using MH for η = 100, τ = 7, β = 7, and θ = 150

generally considered better for high dimensional distributions, because it rejects fewer candidates. The algorithm is very similar to the regular MH algorithm, except each component of the candidate is drawn one at a time. For a two dimensional target distribution, the first component is drawn using the previous sample. If it's accepted, that component updates the previous sample. This new previous sample is used to draw the second component of the candidate. If the first component was rejected, then the component is not updated. The second component is drawn from the previous sample, without updating the first component.

3.1.5 Summary of Bayes Formula Computation

summarizes the advantages and disadvantages of each method. While the analytical, aka the direct application of Bayes formula, and conjugate families are the simplest methods conceptually, they are often times not the best choice. The analytical can be used if normalizing the posterior is not necessary. If the posterior does need to be normalized, then the integral of the product of the distributions may pose a problem. The conjugate families are the fastest and most straightforward computationally if the prior and likelihood distribution fits are conjugate. However, there is no guarantee that suitable conjugate fits can be found.

Instead the numerical approach is a straightforward method that is easy to apply and does not depend on using specific distribution fits. In practice this method is seldom used, but it is useful as a comparative benchmarking tool. It's not often implemented as it's considered to be computationally more expensive than MC sampling. It may also be imprecise for high dimensional distributions.

The MH algorithm is a widely used method of fast sampling from a posterior distribution. A good introduction to the algorithm can be found in Chib and Greenberg's paper [Chib and Greenberg 1995]. It's a MCMC method that quickly samples from multidimensional spaces. However, it can also be used for one-dimensional spaces.

There are minor drawbacks to the MH algorithm. The biggest is that the

Table 3-1: A comparison of Bayes calculation methods

Method	How to Compute	Advantages	Disadvantages
Analytical	Direct application of Bayes formula	- Conceptually straightforward	- Integration of distribution products
Conjugate	Reference conjugate family lists	- Computationally straightforward - Easily referenced	- Applied to limited number of distributions - Depends on finding a suitable distribution fit
Numerical	Approximate integrals with areas	- Covers all distribution fits - Easy to implement	- The approximations should appropriately capture distribution
MH Algorithm	Select jumping distribution and iterate	- Covers all distribution fits - Fast sampling of high dimensions	- Some subjectivity in jumping selection, burn in, and sample size

parameters of the algorithm must be defined in a way that results in a suitably accurate posterior distribution within a suitable length of computation time. These parameters include burn in time, number of samples, and selection of a jumping distribution. However the leeway allowed in setting algorithm parameters is fairly generous and can result in reasonably accurate and fast performance. It's good practice to ensure that the algorithm has enough time to converge on the posterior.

3.1.6 Kernel Density Smoother and Empirical Fits

One limitation with every Bayes method presented so far is they assumed that distribution fits can be made to the data. This may not always be the case. While a full introduction to the coefficient update will be given later, it can be assumed for now that instead of RUL distributions, we wish to update regression coefficients instead of RUL. There may be cases where a proper fit to the data can't be found. In those cases a kernel smoother can be applied, as shown in Figure 3-5. Compared to the Gaussian fit, the kernel smoother is arguably a more accurate fit, by capturing the asymmetry.

The kernel smoother, in its current form, is usually attributed to Parzen and Rosenblatt [Parzen 1962; Rosenblatt 1956]. If an independent and identically distributed data matrix x_i has an unknown distribution f, then f is estimated by the sum of kernels

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - x_i) = \frac{1}{nh}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K}(\frac{\mathrm{x-x_i}}{\mathrm{h}}) \qquad 3\text{-}8$$

where K(x) is the kernel function, and K_h is the scaled kernel. While different functions exist, including uniform and triangular, the Gaussian kernel was selected based on its widespread use.

This kernel density estimator also depends on the bandwidth, h. The kernel bandwidth can be adjusted to appropriately interpolate the PDF, changing the width of each kernel. A larger bandwidth will result in a wider and smoother distribution. A widely used bandwidth selection is the normal distribution approximation, also known as Silverman's rule of thumb [Silverman 1986].

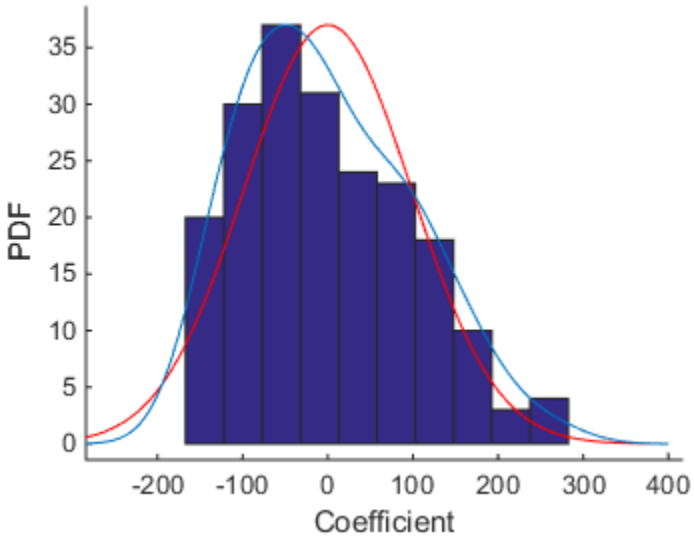


Figure 3-5: Kernel Smoother on Data

$$h = \left(\frac{4\sigma^5}{3n}\right)^{1/5} \quad 3\text{-}9$$

It should be noted that the tails of the kernel smoother fall off differently from the Gaussian fit, depending on the size of the bandwidth. If the bandwidth is smaller than the Gaussian standard deviation, then at a certain point beyond the range of the data, the kernel density will fall off faster than the Gaussian. Therefore the kernel smoother can be difficult to use if the distribution is expected to have heavy tails.

Multivariate kernel density estimation can be applied to multidimensional data to model more than one random variable. This estimation is useful when updating the regression coefficient distribution, which has more than one variable, as opposed to the univariate RUL distribution. For a linear model, the two regression coefficients, slope and intercept, form a two-dimensional distribution.

The d-dimensional kernel density estimator is given by [Simonoff 1996]

$$\hat{f}(x) = \frac{1}{n|H|}\sum_{i=1}^{n} K_d\left(H^{-1}(x - x_i)\right) \quad 3\text{-}10$$

In this equation, $|H|$ is the absolute value of the determinant of H, the bandwidth matrix. Again, by using Silverman's rule of thumb [Silverman 1986], H can be found.

$$H = \left(\frac{4}{d+2}\right)^{1/(d+4)} \Sigma^{1/2} n^{-1/(d+4)} \quad 3\text{-}11$$

It can be seen that the multidimensional H simplifies to the univariate rule of thumb when d equals 1.

An alternative to the kernel estimator is to use the empirical distribution of the data. The method calculates the CDF at each data point. Thus the distribution is the data itself. One problem with this method is that the tails past the minimum and maximum values are undefined. The empirical fit can interpolate, but it can't extrapolate. If the empirical fit is the prior, no matter the likelihood, the posterior will only be defined within the range of the prior. The kernel smoother does not suffer from this issue, because each Gaussian kernel is defined over all values. The kernels will extrapolate

outside the empirical interval to give small but importantly non-zero estimates while keeping the integrity of the overall distribution shape.

3.2 Weighted Total Least Squares Regression

WTLS regression can be used to include prior data into a regression model. For example, a type I RUL prior can be included in the regression step of the type III GPM. It estimates an additional point centered on the expected RUL and at the failure threshold. This way it represents the POF with uncertainties in the RUL and failure threshold. Because the sources of uncertainty occur in two dimensions, time and degradation, the OLS model can't be applied. The OLS assumes that regression errors are only expressed in the dependent variable, degradation, and doesn't account for variation in the independent variables, time.

There is no closed form solution for the computation of a WTLS [Markovsky et al., 2006]. However various iterative models for calculating an estimate exist. One step towards modeling uncertainties in both dimensions is based on the errors-in-variable model [Griliches and Ringstad, 1970; Schaffrin et al., 2006]. It can be expressed as the following [Schaffrin and Wieser, 2008].

$$y = (A - E_A)\beta + e_y \tag{3-12}$$

The variable A has replaced the notation X of independent input variables, in this case time. This A matrix is subject to errors from E_A while the y is subject to e_y. These errors can be sampled from a normal distribution.

$$\begin{bmatrix} e_y \\ e_A \end{bmatrix} = \begin{bmatrix} e_y \\ vec E_A \end{bmatrix} - \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \begin{bmatrix} Q_y & 0 \\ 0 & Q_A \end{bmatrix} \right) \tag{3-13}$$

where vec is the operator that stacks columns of the matrix underneath each other. The Q matrices are symmetric cofactor matrices of e_y and e_A. The homoskedastic case, which assigns an equal variance on all observations, is represented by identity Q matrices of

size n, the number of data points, and nm, the number of data times the number of regressors.

The iterative errors-in-variables algorithm can be summarized as follows [Schaffrin and Wieser, 2008].

First step

$$v^0 = 0$$

$$\beta^0 = N^{-1}c \text{ for } [N, c] = A^T[A, y]$$

Second step

$$v^i = \left(y - A\beta^i\right)^T (y - A\beta^i)/(1 + (\beta^i)^T\beta^i)$$

$$\beta^{i+1} = (N - v^i I_m)^{-1}c$$

Third step

Repeat second step until

$$\|\beta^k - \beta^{k-1}\| < \delta_0$$

for a predetermined stop criterion δ_0

$$\sigma_0^2 = \frac{v^k}{n - m}$$

It is initialized using an OLS estimate, and iterates the parameter estimates until a stopping criterion is reached, based on the change of β from the previous step to current step. This criterion δ_0 is typically assigned a very small positive value. Finally the variance of the model can be calculated.

To model different variances for different points, a heteroskedastic condition is required. The Q matrices must be altered to reflect the different weights of each point. Some useful notations can be made.

$$\begin{aligned} Q_y &= P_y^{-1} \\ Q_A &= Q_0 \otimes Q_x, \\ Q_0 &= P_0^{-1} \\ Q_x &= P_x^{-1} \end{aligned} \qquad \text{3-14}$$

where ⊗ is the Kronecker-Zehfuss product. This is the resultant matrix as each element of the first term is multiplied by the second term. The purpose for this notation is to allow for different effects of each input regressor. By setting the diagonals of Q_0 to either zero or one the product Q_A determines which regressors are included. The diagonals of Q_x can contain the variance for each data point.

According to Schaffrin, the algorithm attempts to minimize the weighted sum of the contribution of errors from the y and A.

$$e_y^T P_y e_y + e_A^T (P_0 \otimes P_x) e_A = min \qquad 3\text{-}15$$

The algorithm itself can be summarized as follows.

First step

$$v^0 = 0$$

$$\beta^0 = N^{-1}c \text{ for } [N, c] = A^T P_y [A, y]$$

$$\beta^1 = (\mathrm{A^T}(\mathrm{Q_y} + ((\beta^0)^{\mathrm{T}}\mathrm{Q_0}\beta^0)\mathrm{Q_x})^{-1}\mathrm{A})^{-1}\mathrm{A^T}(\mathrm{Q_y} + ((\beta^0)^{\mathrm{T}}\mathrm{Q_0}\beta^0)\mathrm{Q_x})^{-1}\mathrm{y}$$

Second step

$$\lambda^i = (Q_y + ((\beta^0)^T Q_0 \beta^0) Q_x)^{-1}(y - A\beta^i)$$

$$v^i = (\lambda^i)^T Q_x \lambda^i$$

$$\beta^i = (A^T(Q_y + \left((\beta^i)^T Q_0 \beta^i\right) Q_x)^{-1} A - v^i Q_0)^{-1} A^T (Q_y + \left((\beta^i)^T Q_0 \beta^i\right) Q_x)^{-1} y$$

Third step

Repeat second step until

$$\|\beta^k - \beta^{k-1}\| < \delta_0$$

for a predetermined stop criterion δ_0

$$r = n - m$$

$$\sigma_0^2 = \frac{\lambda^T (y - A\beta)}{r}$$

The algorithm is initialized by using an OLS model in the first step. The second step iteratively adjusts the estimated β using the "Lagrange multiplier" λ. It iterates until

a stopping criterion is reached, based on the change of β from the previous step to current step. This criterion δ_0 is typically assigned a very small positive value. Then, using the degrees of freedom r, the variance component can be estimated.

3.3 Transitions

The primary focus of the current investigation is the use of Bayes to transition between the prognostic types, as presented in Figure 2-1. These were previously explored [Jeffries et al., 2014; Welz et al., 2014; Nam et al., 2013; Nam et al., 2012] on simulated data, heat exchangers, and pump impellers with some positive results. Because each type can be categorized within the lifecycle of an individual, the methods can be combined based on the data available. A full lifecycle prognostics methodology can apply the type I at BOL. When operating conditions are measured, not only can the type II methods be applied, but using Bayes, the type I information can also be included. When degradation data is available, either a type I or a type II or a combined type I and II result can be used as a prior for the type III model.

The following sections will go through an example of each transition that has been developed. With an understanding of how and when to use the various model transitions, a discussion of which methods can be applied if all sources of data were available is given. Rather than applying each transition in isolation, multiple transitions can be merged.

3.3.1 RUL Update

For transitions that update the RUL from one type to a latter, a direct application of Bayes rule is possible [Sharp et al., 2014]. For all RUL transitions between different types of model, the methodology is generally the same.

Suppose that for a population of some monitored system, the TOF and sensor data were available for previously failed individuals. In this case, a type I survival model can be updated using the type III GPM.

The most widely used survivability model is the Weibull distribution, introduced in section 2.1. A Weibull fit is given to the TOF. Then for a test case that has been operating for some time, the conditional Weibull can be derived conditional on the fact that the case has survived to the current time. The conditional Weibull reliability function is reproduced here.

$$R(t|T) = \exp - \left[\left(\frac{t_0 + t}{\theta} \right)^{\beta} - \left(\frac{t_0}{\theta} \right)^{\beta} \right] \quad 3\text{-}16$$

Using the definitions of reliability, failure cumulative distribution functions, F, and the failure PDF, f, the following relationships can be used to calculate the failure PDF as a function of time.

$$F = 1 - R \quad 3\text{-}17$$

$$f(t|T) = -\frac{dR}{dt} = \frac{\beta \exp\left[\left(\frac{T}{\theta} \right)^{\beta} - \left(\frac{T+t}{\theta} \right)^{\beta} \right] * \left(\frac{T+t}{\theta} \right)^{\beta}}{T+t} \quad 3\text{-}18$$

The conditional failure PDF is the negative derivative of the reliability function. This PDF should be normalized to an area of 1 after the shape is found.

After the type I model is applied, a fault occurs and degradation tracking is now possible. Using the degradation data, a type III GPM model is selected to create another RUL estimate. Based on the training population, a linear fit can be selected, with a threshold near 1. Then for a single censored test case, the RUL estimate is found using linear OLS regression. This estimate can be made into a distribution by repeated sampling of path coefficients using MC.

An example of transitioning between the type I and GPM is shown in Figure 3-6. The conditional Weibull was created after estimating the shape and scale parameters of the RUL to be 4.7 and 100 for an arbitrary unit of time. This conditional Weibull is the left broad peak of the three PDFs at the top of the figure, labeled "Prior RUL".

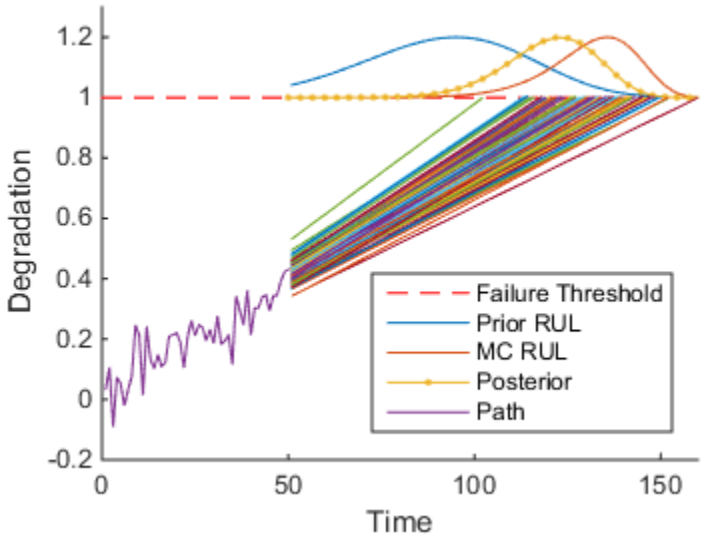


Figure 3-6: RUL Update

The degradation path, the noisy path from time 0 to 50, was regressed to an estimated TOF of 130. While a mean estimate of the path coefficients was found, 100 MC samples were drawn to vary the coefficients to produce MC paths. Each of the lines from time 50 to the failure threshold represents these MC paths. The different times when they cross the failure threshold form the TOF distribution of the GPM. A functional fit was then used to approximate the MC TOF distribution. A suitable Weibull fit was found with shape and scale parameters of 13 and 136. This likelihood peak is the right purple PDF above the MC POFs. Using the conditional Weibull prior, and the GPM Weibull likelihood, a MH sampling was used to calculate the posterior distribution, the middle dotted yellow peak.

The main steps of the RUL update are as follows. The first is to choose prognostic models based on the information available. Calculate the RULs. Use any Bayes method, i.e. MH, to calculate posterior RUL.

Now let's consider a different situation. Suppose that instead of two RUL estimated distributions, there is enough data to apply all three types of prognostic models. To combine all three RUL, each PDF is updated sequentially. For instance, with a type I, type II PHM, and type III GPM, two RUL updates can be made. One is from type I to type II. Then, using this combined RUL as a prior, the type III GPM can be updated. This results in one RUL distribution that is made from all three sources of data, and all prognostic models made.

3.3.2 Type III GPM Coefficient Update

This update is an adaptation to the GPM. By recording the OLS regression coefficients for previous failures, a prior distribution can be identified. This is updated with the OLS estimate of the coefficients for an individual unfailed case. Using the prior distribution, and the OLS estimate, the posterior is found using a direct application of Bayes formula.

The differences between the previously developed method presented in section 2.5.1. and the proposed coefficient update are illustrated in Figure 3-7 and Figure 3-8. In the first figure, it's assumed that the two sources of data, the prior coefficients and the prognostic parameters for the individual case, are known. These data are input into the green GLS, as part of the GPM. The output of the GLS gives the purple posterior distribution estimate of the coefficients. The expected values and covariance matrix are given by equations $\hat{\beta} = (X^T \Sigma^{-1} X)^{-1} X^T \Sigma^{-1} y$ 2-25 and $V\sigma^2 = (X^T \Sigma^{-1} X)^{-1}$ 2-26.

In the next figure, the prognostic parameters are input into the OLS to produce the blue coefficient estimates. These estimates are updated using Bayes formula, in green, and the prior coefficients distribution. The result is the purple posterior distribution estimate.

One advantage of the proposed coefficients update method is the relaxation of assumptions made by the previous method. In particular, the GLS models the variances and covariance of the prior. Thus, the GLS restricts the prior to a multivariate normal distribution. In the proposed method, because Bayes formula is used, any representation of the priors can be used, including the multivariate kernel.

To illustrate this transition, a two-dimensional prior population of regression coefficients was created for a linear GPM. It was made by sampling from two Weibull distributions to ensure some amount of covariance between the variables. To limit assumptions about assigning a distribution fit to the prior, a multivariate kernel density estimator was applied, using Silverman's rule of thumb for the bandwidth. The two-dimensional density is shown in Figure 3-9. The first coefficient is the slope, or linear, coefficient, and the second is the intercept. The density shows the MLE at around (1, 8). The density is also left tailed with respect to the first coefficient, and right tailed for the second.

After the prior was created, an individual degradation path was simulated by sampling coefficients from the same distribution as the prior and adding Gaussian noise.

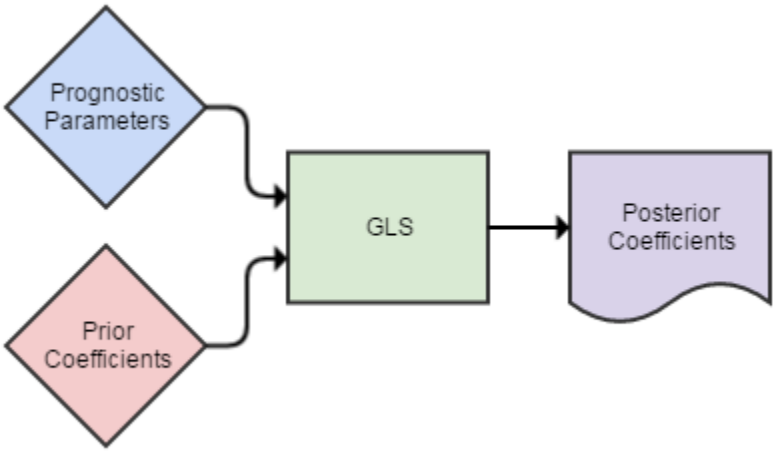


Figure 3-7: Previous Method for Including Prior Coefficients

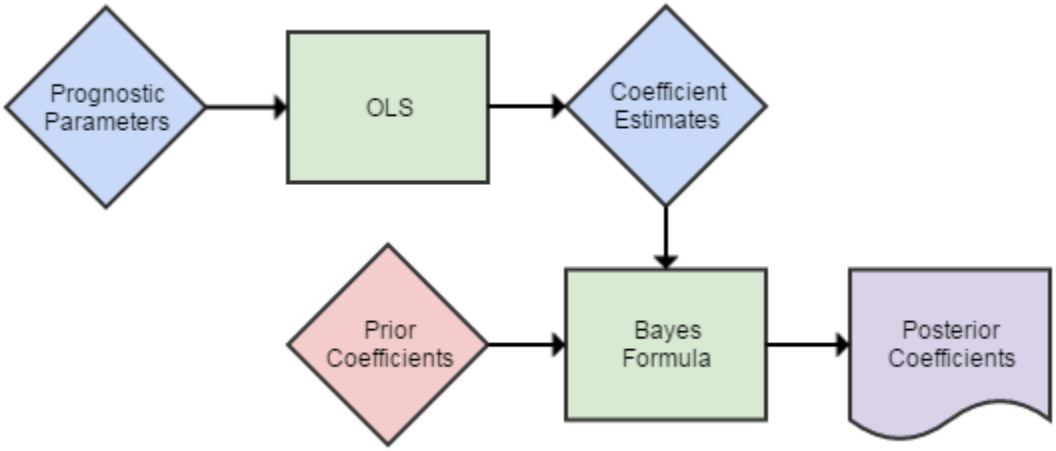


Figure 3-8: Proposed Coefficients Update

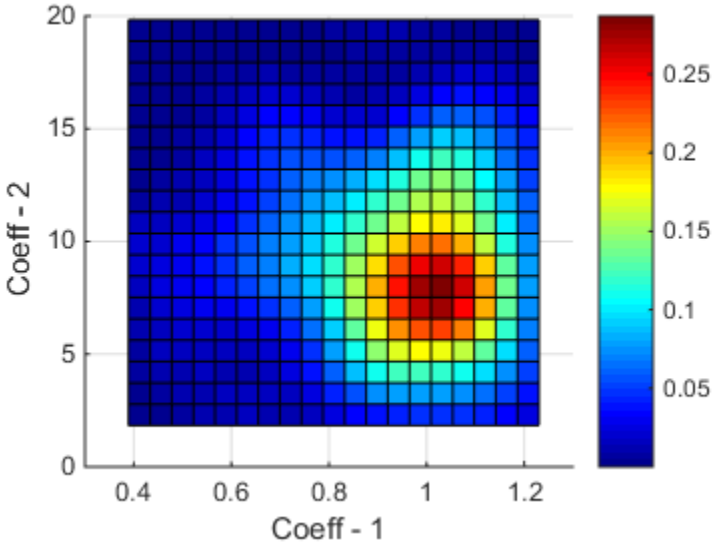


Figure 3-9: Multivariate Kernel Density Estimator on Two Coefficients

This path reached time 50, when it was censored and regressed. The results were a mean estimate of the coefficients as [1.10, 10.5], with a covariance matrix of [5.76e-4, -0.015; -0.015, 0.495]. The mean puts this new sample near the MLE shown in the previous figure. The variance of this sample is smaller than the variance of the prior. Together, the mean and variance of the new sample had a large impact on the posterior, as shown in Figure 3-10. The blue circles are each of the 100 samples that make up the prior distribution. The likelihood estimate is shown as the red X nestled in the green posterior samples made by the MH. The space occupied by the posterior is much smaller than the prior, due to the small covariance matrix of the likelihood.

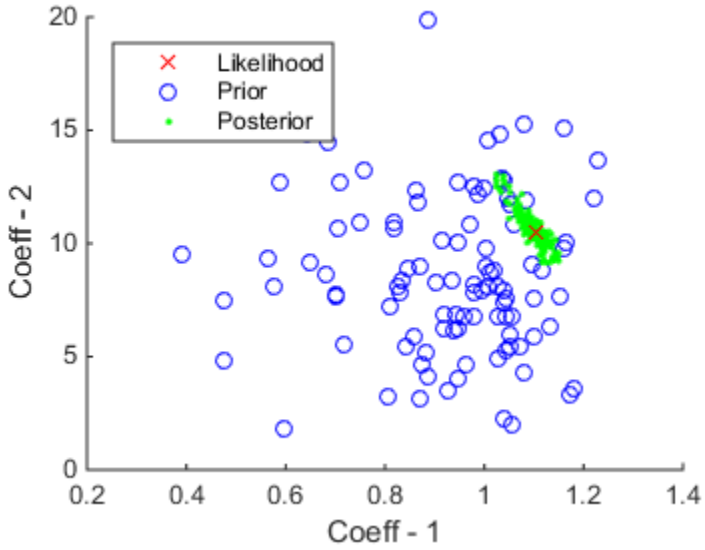


Figure 3-10: Componentwise MH on Coefficient Update

3.3.3 Type III GPM WTLS Regression

Like the previous transition, the RUL regression procedure requires the application of a type III GPM. If a prior RUL distribution, such as a type I, is

approximately normal, the regression can include the additional data point, weighted by the uncertainty. This additional data point is centered at the type I RUL estimate and degradation threshold. However, because of the uncertainty in the independent variables, an OLS model can't be used. The OLS assumes that errors only affect the prognostic parameter. In addition, the errors in both dimensions must also be weighted to allow for the appropriate uncertainty of the prior, relative to the degradation path. To properly include such weights, a WTLS regression algorithm can be implemented, as discussed in section 3.3.3.

For the same situation as the previous two sections, the type I RUL can be included in the WTLS. One limit to this regression method is that while the WTLS regression can model different uncertainties, it assumes that all errors are sampled from Gaussian distributions. This may not hold true, especially when considering that a type I Weibull is much more likely to be used than a Gaussian. There are not a lot of options to avoid having this problem, other than to check whether the prior TOF is approximately normal. This assumption holds best if the type I TOF distribution is approximately normal, and the posterior is calculated early in life. The early life conditional normal is similar to the un-conditional normal; by excluding a small portion of the estimated failure distribution from the condition, the distribution does not change much from the un-conditional.

To include the type I prior, the prognostic parameter matrix is appended by the threshold, while the A input matrix is appended with the appropriate values in each column, given the estimated type I TOF. Although the nomenclature here is to use TOF, this is to clarify the precise value that is appended. This TOF is derived from the type I RUL results. Using these values, the appended data approximates the expected future POF.

$$y_* = \begin{bmatrix} y \\ thresh \end{bmatrix}, A_* = \begin{bmatrix} A \\ typeITOF \end{bmatrix} \quad \text{3-19}$$

The uncertainties in both y and A dimensions are included in the model using the Q covariance matrices. The y covariance is appended with the threshold variance. The Q_0 was set to having a single 1 in the first diagonal. This means the type I TOF variance contained in Q_x only impacts the time input, and not the y intercept.

$$Q_y = \begin{bmatrix} \sigma_y^2 & 0 \\ 0 & threshVar \end{bmatrix} Q_0 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} Q_x = \begin{bmatrix} \sigma_x^2 & 0 \\ 0 & typeITOFVar \end{bmatrix} \quad 3\text{-}20$$

The y and x variances are the uncertainty of the regression model. Recall that for the homoskedastic error-in-variable model, the Q_y and Q_A matrices were identities. This means that sampling the errors in y and A would both have the variance σ^2_0. Therefore the σ_x and σ_y can both be estimated if the training paths were used to create error-in-variable regressions. The mean value of the variance over all paths is used to estimate the y and x variances.

After all the matrices are set up, the WTLS algorithm can find the updated coefficient estimates. This update can be very beneficial at the BOL, Figure 3-11. In this instance, a test case of 3 observations was created, along with a population of prior paths all ending in failure. The blue path starts near zero, dips down, then rapidly increases. An OLS on this path results in the steep yellow path fit. As a result, with a GPM with no updating, the RUL estimate was very short, around time 26. However, the inclusion of the type I RUL shown as the singular red peak, helps the GPM. Instead of the steep yellow path, the more moderate dotted purple path estimate crosses the horizontal red failure threshold near the center of the type I, time 97.

If a type II model is available in addition to the type I, then both RUL can be appended to the data matrices. The algorithm would include the contributions from both at once. It's also cautioned that this Bayesian regression transition is used separately from the RUL update. While it is possible to include two priors using two different transitions, the same model, or same source of data, should not be used more than once to calculate the posterior. For instance, while it's mathematically possible to use a type I prior in both the RUL update and RUL regression, the same information is

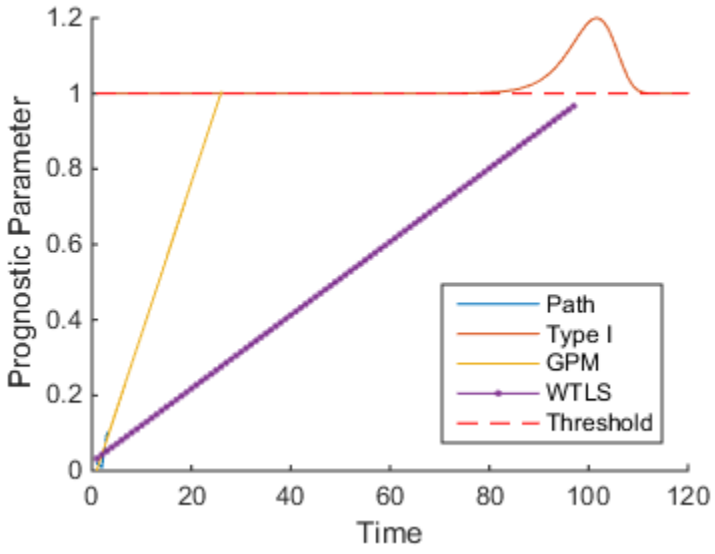


Figure 3-11: RUL Regression Example

included twice, and would strongly weigh the posterior in favor of the prior.

3.4 Summary of Transitioning Methods

Bayes formula was introduced as a method of combining a prior estimate of a parameter with new sampled data to form a posterior estimate. The prior is the belief about a certain parameter of interest, before new data is seen. The likelihood is the probability of observing the data, given the parameters.

The formula can be solved in several ways, including the use of conjugate families, numerical approximation, or the Metropolis-Hastings algorithm. Each algorithm has its own advantages and disadvantages, and the selection should take into account the situation. In the best case scenario, the conjugate family is the easiest to apply. However, this imposes strict requirements on what the shape the prior and likelihood have to be. A Gaussian prior with Gaussian likelihood form a conjugate family, but a Gaussian and Weibull do not. For any pair of distributions, the MH algorithm is a widely recognized fast sampling technique. A kernel density smoother can be used with any of the posterior calculation methods when a acceptable fits for any of the distributions can't be found.

By combining different sources of information, gathered over the lifecycle of a component, a better RUL estimate is hoped to be achieved. Three transition methods have been developed and presented. First, a RUL update can be used to transition between any prognostic models, as long as the RUL can be expressed as a distribution and not a simple estimate. If the type III GPM is used, then RUL regression can be applied instead. Finally, a method for updating the path coefficients within the GPM is also presented, and provides some potential benefit.

Multiple transitions can be used in tandem depending on the availability of data. If all three sources of data are available, then three prognostic models can be made. The

main rule when using multiple transitions, is that a particular set of data should not be incorporated into the overall RUL more than once. A type I and type II can be transitioned using RUL update. This combined RUL can be used as a prior in another RUL update to the type III GPM. Alternatively the posterior type I and II RUL can instead be transitioned using RUL regression. This case is also valid, though it uses two different updating methods, because the main rule is not violated. If the type I and II combined RUL is used in RUL regression, and that RUL output is transitioned using the I and II RUL as a prior in a RUL update, then both the type I and II model are included in the final RUL twice. This has the effect over giving more weight to the TOF and condition data sets, and less on the degradation data.

While this chapter summarized the different transition methods, and how to apply and calculate them, the next chapter uses each transition to compare RUL estimates on two different datasets.

4. APPLICATION TO DATA SETS

4.1 PHM Challenge Data

The Prognostics and Health Management (PHM) society hosts an annual prognostics competition to test emerging prognostic methodologies. In 2008, a large dataset was made publically available with no information about what the signals were modeling to encourage using methods that did not depend on the system's nature to give informed RUL estimates. As a result, the techniques applied to the dataset are not specific to any single situation but instead can be generalized for different groups of sensors as long as the data tracks multiple instances of operation, faults, and failures. In practice, understanding the process equipment beyond just the measured values should provide better insight than data analysis alone.

The dataset, which can be found in NASA's online prognostic data repository [Saxena and Goebel, 2008], consists of a training set, test set, and the test set's actual RUL. The training set consists of 260 cases of 24 sensors. Each case is assumed to start with no detectable degradation. After an average of 200 time cycles of data collection, the run ends at the TOF. The test set consists of 259 cases of the same 24 sensors except they are censored before failure. These test cases' true TOF is provided to validate prognostic models.

Figure 4-1 shows the historical TOF for the training cases, with a Weibull distribution fit. As part of distribution analysis, the negative log-likelihood was found for both normal and Weibull fits: 1.37e3 and 1.38e3, respectively. Minimizing the negative log-likelihood is equivalent to maximizing the likelihood estimate. Although the negative log-likelihood slightly favors the normal fit, the Weibull was applied because this distribution is generally favored in reliability analysis and the difference is small. The TOF distribution fit's shape and scale parameters were 4.39 and 226 time cycles

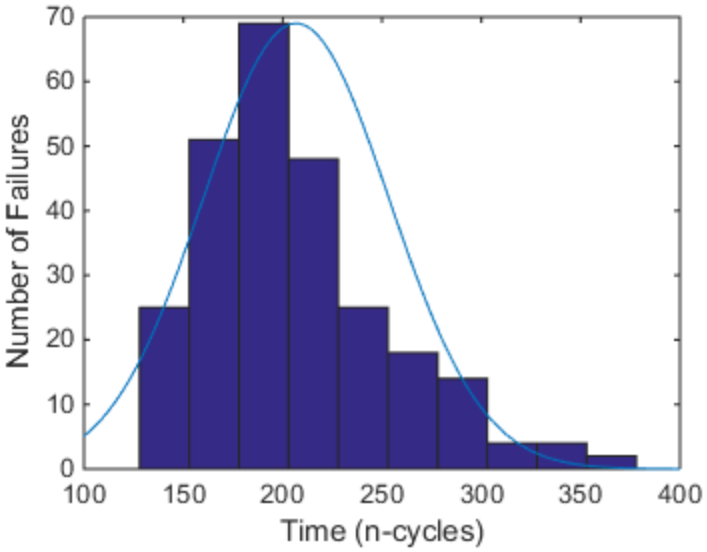


Figure 4-1: Historical Time of Failures Fit

respectively.

A Weibull fit was applied to the TOF distribution results in a type I survival model that follows a conditional Weibull PDF. As mentioned in section 2.1, the model calculates the RUL conditional on the fact that it has survived to the current time.

As previously mentioned, some analysis had to be done before the type III GPM could be applied. First, an AAKR model was created to describe the system's normal operating behavior without a fault. To create this model, the first 5% of each training case was concatenated to create one fault-free matrix. This assumed that there was no degradation for the first 5% of usage.

Variable selection was then considered. While all 24 signals can be included, some may not be useful and would only add noise to the model. One way to select useful signals is to compare the correlation coefficients, r, also known as the Pearson's correlation coefficient. These coefficients are a measure of a linear relationship between two signals from -1 to 1. The correlation can also be represented by r^2 to confine the value between 0 and 1. A value closer to 1 represents a strong correlation between the signals. To consider which signals are strongly correlated, a cutoff of 0.7 is imposed. Any pair of signals above the cutoff are considered to be strongly correlated. If groups of signals are strongly correlated with each other, they can be modeled.

Figure 4-2 shows the correlation coefficients on the 24 signals. The diagonals are the correlations of each signal with itself and, therefore, are equal to one. The off-diagonal elements show the correlation between two different signals. The red and orange squares show strong correlations, while the blue squares indicate weak ones. Some positive correlation is always expected between two random variables, but can usually be regarded as noise. Based on the cutoff, a group of 21 signals with strong intra-group relationships were found.

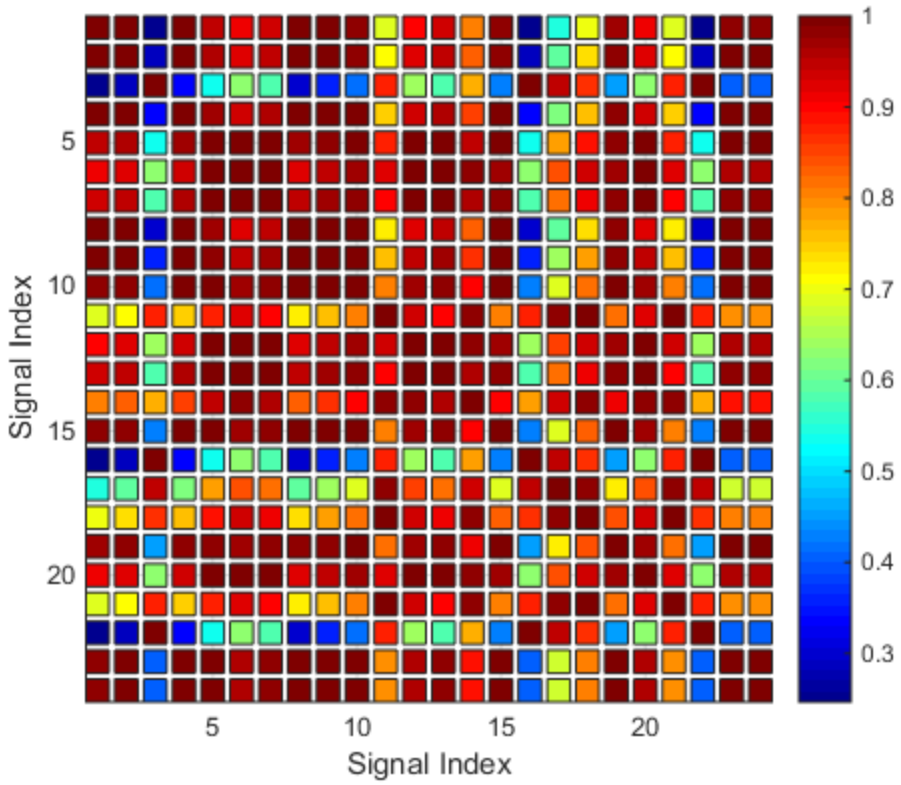


Figure 4-2: PHM Correlation Coefficients

The fault-free observation matrix is further divided into three different matrices. The "training" matrix (not to be confused with the train and test fault cases) is used as the pool from which AAKR memory vectors can be selected. This matrix is used to build the model. The testing data is used to optimize the AAKR's kernel bandwidth. The validation is used as a benchmark to calculate the model's performance measures, such as cross-validation error.

The AAKR's memory matrix is a subset of the training matrix. While the number of memory vectors desired can vary depending on the system being monitored, some care should be exercised when selecting the number. If the number of vectors is too small, the MM may not capture the total space of unfaulted operation. With too many vectors, the marginal benefit of including more memory diminishes, and leads to an increase in computation time. Therefore, 500 vectors were selected.

This model also transforms the data to standardize each signal's distribution. This transformation is sometimes referred to as a z-score. By standardizing the data, each signal's mean value is zero with a standard deviation of one. This method is used to scale the signals' nominal values so they can be comparable. For instance, a temperature recorded in Celsius has lower value than the same temperature in Kelvin, even though the same quantity is described. The changes within the observed operating range, not the signals' mean values, should be considered important.

To find a suitable bandwidth for the kernels, a range of bandwidths was tested. The resulting mean squared error between the model and the testing data for each bandwidth was compared, and the lowest one was selected.

After the AAKR was built and optimized, the performance metrics were found. The first metric was accuracy, as represented by a root mean squared error (RMSE). For each signal of each observation in the validation data, the difference between the input and the output was considered the error, which can also be referred to as the residual.

The mean squared error over all observations is found, and the square root is calculated.

$$RMSE = \sqrt{mean((x_i^{pred} - x_i)^2)} * 100 \quad \text{4-1}$$

shows the RMSE for each selected signal. Some signals, such as 3 and 7, are so small that they are not visible on the plot. On the other hand, the prediction of signals 20 and 21 is less accurate. The AAKR can be further validated by observing the validation dataset's residuals. For each signal, the residuals should be a randomly distributed variable with a mean of zero.

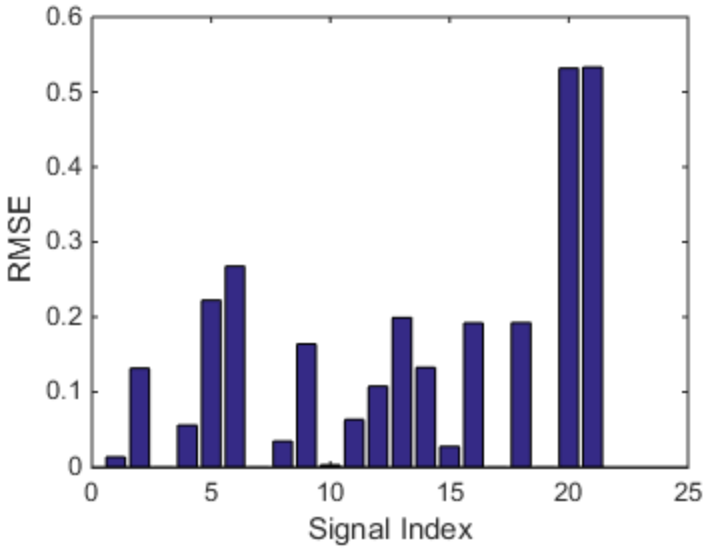


Figure 4-3: AAKR RMSE

After the AAKR is built, each training case can be separated at the point of fault detection. The observations before the fault should encompass the amount of time the case is unfaulted. After the fault is detected, the residuals should be nonzero and

contain information about the degradation. The SPRT with consolidated fault hypothesis was used to detect faults with default values of $\alpha = 0.01$ and $\beta = 0.1$. For the training cases, the faults were usually detected between times 50 and 100, with some faults extending to time 200. For cases in which the SPRT did not trigger any fault alarms, the entire run data was kept. Keeping all the data is the same as assuming that the fault was detected at time 0, and no data was cut. The goal of separating the data at fault detection is to apply the GPM only for the degraded portion of the lifecycle.

These residuals with similar trends over multiple cases were collected and linearly combined to form the prognostics parameter, a single quantification of the degradation. This process was achieved using the regression model outlined in section 2.3.2. The training and testing prognostic paths are shown in Figure 4-4 and Figure 4-5. While many paths overlap, the population's overall shape as it runs towards failure can be seen. The paths start near or slightly above 0 and follow an increasing trend upwards to fail around 1. The monotonicity, prognosability, and trendability for the training cases were found to be 0.589, 0.782 and 0.328, respectively. The monotonicity was limited by the summed process noise seen in the path. The prognosability showed a clear differentiation between each path's beginning and end. The trendability was by far the lowest metric, contrasting with the training paths' trendability before being cut at fault detection. The uncut trendability was 0.779 with similar monotonicity and prognosability. This reduction may be due to the fact that the degradation paths themselves are more varied than the case's overall lifecycle. While some paths are linear and others have an increasing slope, the lifecycle path's overall shape would generally have an early low slope period before the fault, followed by acceleration towards failure.

The degradation path for a single test case is shown in Figure 4-6. The horizontal PDF's baseline is set at the failure threshold and represents the historical TOF fit. The vertical PDF is based on the average TOF, with the distribution showing the threshold's

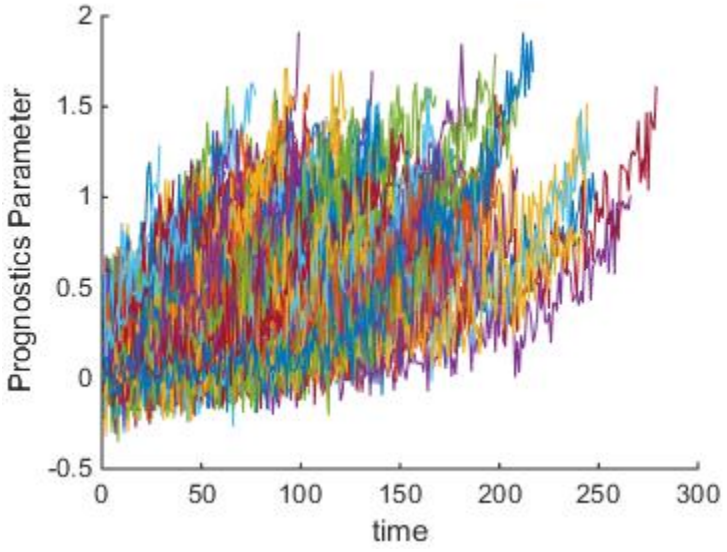


Figure 4-4: Training Prognostics Path

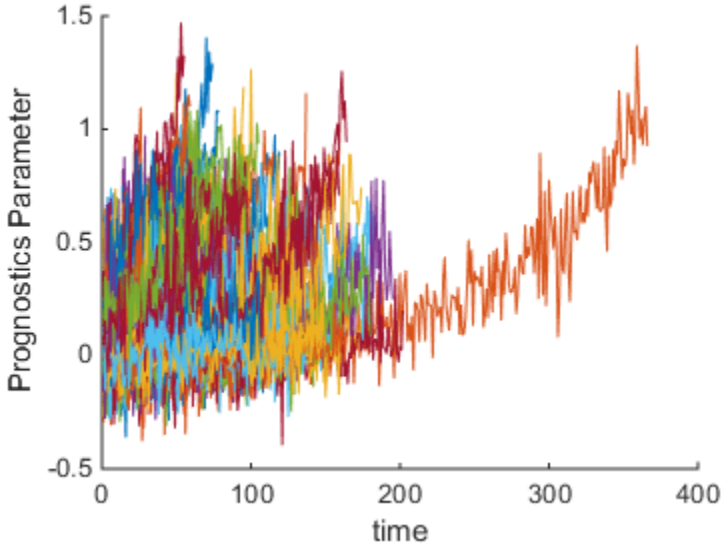


Figure 4-5: Testing Prognostics Path

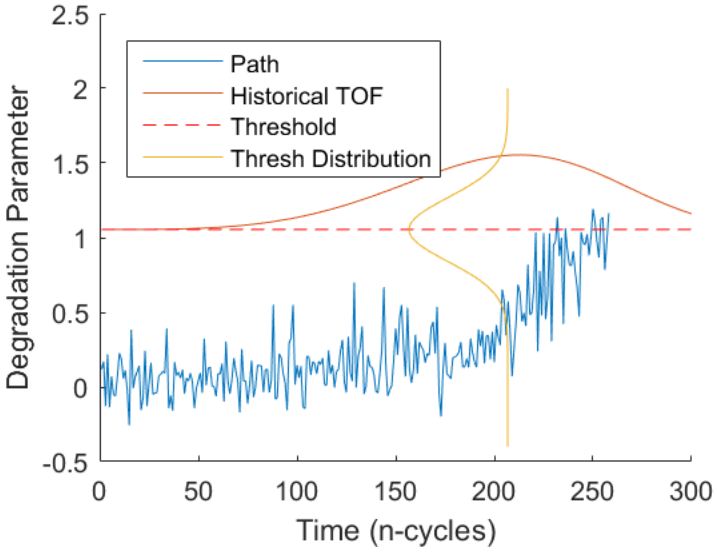


Figure 4-6: Degradation Path of an Example Case

variance. These two distributions help determine the average expected POF and the uncertainty in both time and degradation.

This individual case test has a longer than average TOF with an above average point of failure. For each observation through time, after a small minimum cutoff number of observations, the estimated RUL can be found using types I and III until it's censored. A minimum cutoff number was chosen to account for the degrees-of-freedom problem, when regressing too few data points for the number of coefficients. In an extreme case, fitting a quadratic function to one point is an arbitrary task. For the entire lifecycle, the RUL estimates can be observed if the path is censored at any point.

4.1.1 Type I RUL Update of PHM Data

At each observation, two predictions were made using the type I conditional Weibull and the type III GPM. The type III RUL distribution was made using a MC sampling of path coefficients. A Weibull fit to the MC sampling could have been made. However, upon inspection, the Weibull's shapes would tend towards exponential, resulting in a large weighting at very low RUL. Instead of trying to find appropriate fits at every observation, kernel smoothing was used.

An example of the RUL update for test case 1 at time 77 is shown in Figure 4-7. The histogram is the result of the GPM MC sampling to form a distribution estimate. The blue fit to the MC samples was found using a kernel smoother. In this instance, it interpolates the missing RUL predictions before time 300. The type I prior is the conditional Weibull distribution.

Figure 4-8 shows the RUL estimates of the type I, GPM, and the GPM with type I RUL update over the lifecycle. The RUL estimates stop at about time 250, when the individual is censored. The dataset included the true TOF, from which the true RUL was found. Towards the beginning of life (BOL), the red GPM stays at 0 while occasionally moving upwards in sharp peaks. With only a small amount of data used, the prognostic parameter noise dominated, resulting in an inaccurate RUL. As noted previously, this

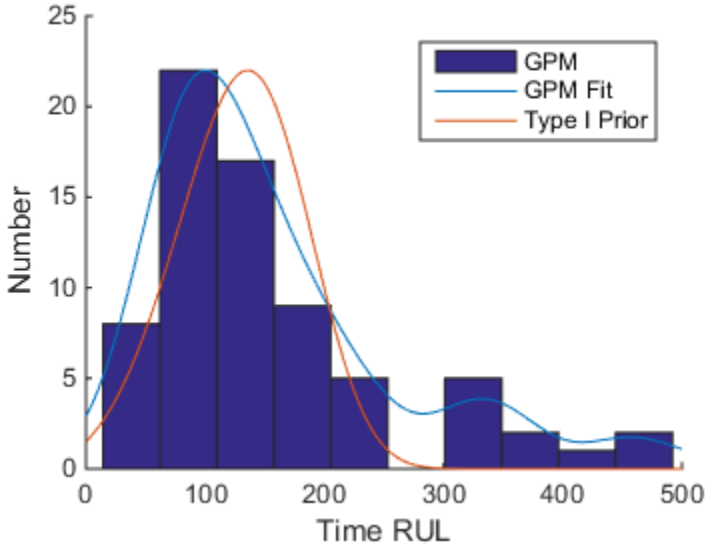


Figure 4-7: RUL Distributions on Run 1 at Time 77

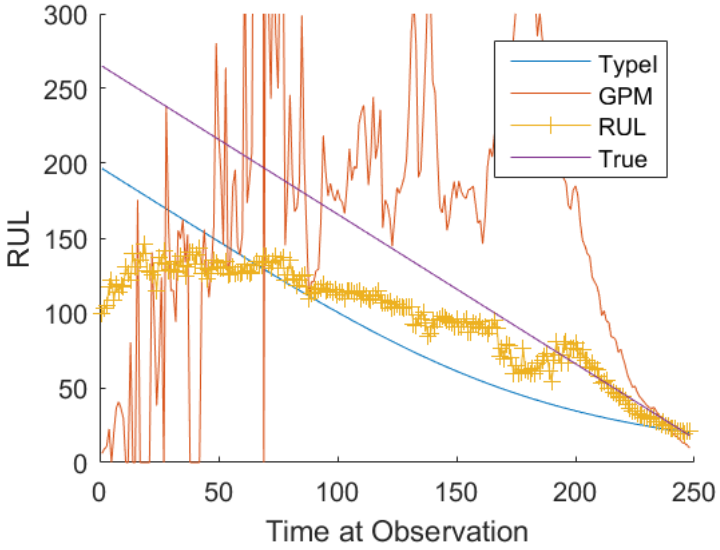


Figure 4-8: Test #1 with Type I RUL Update

test case has a very large TOF relative to the training distribution. Therefore, the type I consistently under-predicts. The conditional survival model causes the RUL estimate to increase toward the true RUL near the end of life (EOL). Using this type I as a moderating effect, the RUL update improves on the GPM throughout the lifecycle. After time 50 and before 200, the GPM suddenly peaks to over 200; however, the RUL update is not affected.

This case exemplifies several trends seen in other test cases. While the RUL estimate fluctuates rapidly at the BOL, this case is an improvement over the GPM alone because it includes the stable type I. By the middle of life, the RUL update stabilizes to the point that large errors in the GPM no longer affect it. Because of the type I's conditional nature, the type I's variance decreases as the system reaches the EOL. This decrease increases its weight on the posterior, moving the posterior estimate closer to the type I.

4.1.2 Coefficient Transition of PHM Data

For the coefficient transition, the population of coefficients from all training cases was used to form the prior. The prior coefficients (quadratic, linear and intercept) are shown in Figure 4-9 and Figure 4-10. The first figure shows the quadratic against the linear to show the correlation, and the need to sample from a single distribution, as opposed to sampling each coefficient independently. If sampled independently, the correlations will be ignored, and the sampled space will lie outside the prior. Therefore a multivariate kernel smoothing method was used on the empirical PDF. This method allows for a non-specified fit, i.e. multivariate Gaussian. If a Gaussian were an appropriate fit, which in this case it isn't due to the visible skew of both the quadratic and linear coefficients, the multivariate Gaussian would be preferred due to its computational speed. The intercept coefficient in Figure 4-10 was shown independently, as it does not have a large effect on path extrapolation.

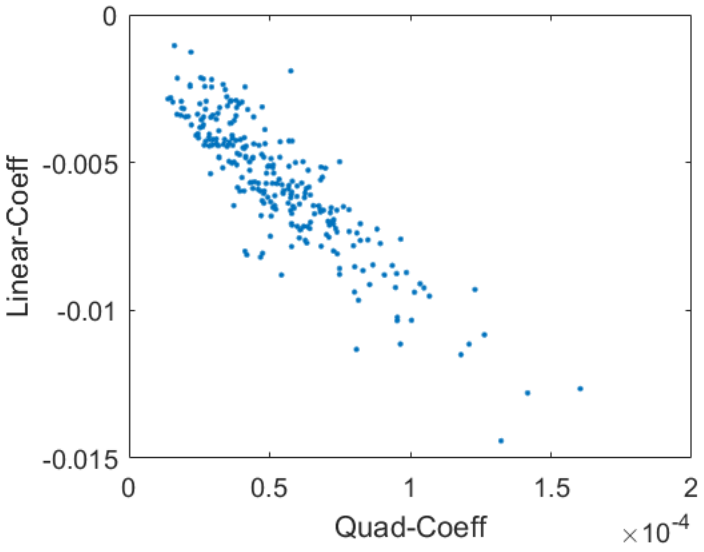


Figure 4-9: Prior Quadratic and Linear Coefficients

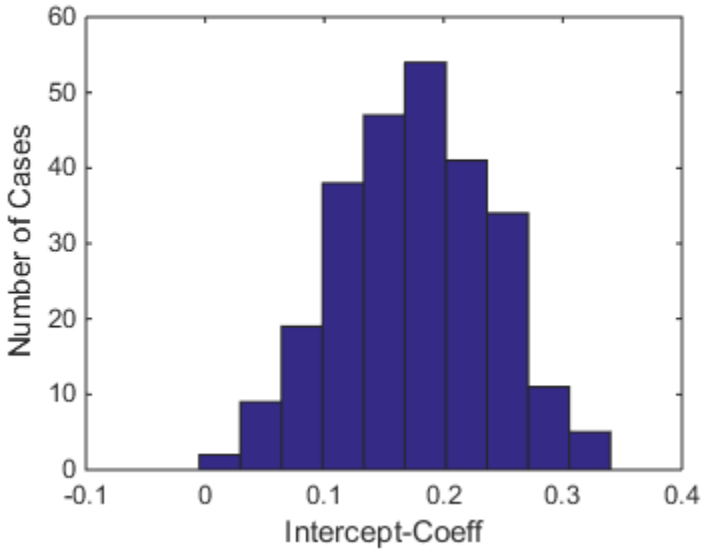


Figure 4-10: Prior Intercept Coefficient

At each observation, the GPM estimates the regression coefficients. This estimate is updated using the priors. The updated coefficient distribution is then used to extrapolate the prognostics parameter to calculate the RUL. This coefficient update is done over the lifecycle as shown in

Figure 4-11. At the BOL, the coefficient update behaves very similarly to the type I. Throughout the lifecycle, the transition compromises between the type I and resists the large errors of the GPM. By the EOL the transition is more of a compromise between the type I and the GPM.

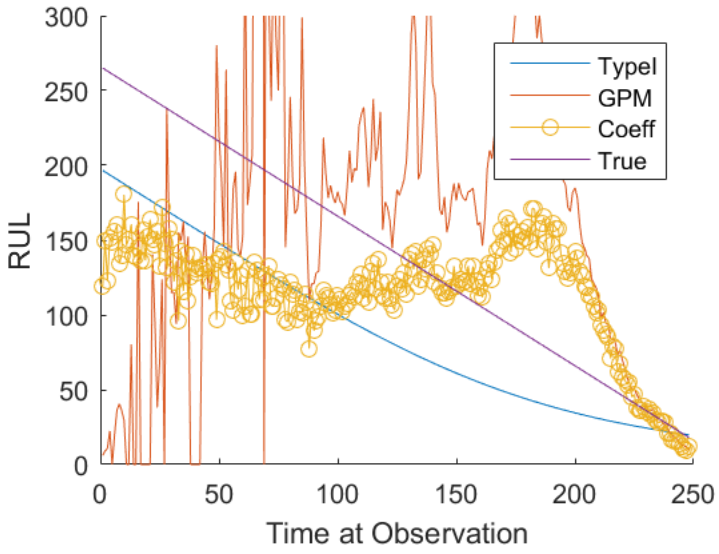


Figure 4-11: Coefficient Update

As time passes, the regression's variance is reduced. The variance reduction is caused by the degrees of freedom overcoming the noise in the prognostics parameter, thus producing more stable fits to the path. This reduction causes the coefficient update to be influenced by the GPM after about time 100. Towards the EOL, the coefficient

estimates' variance is further reduced. The effect is clearly demonstrated at time 200, when the coefficient update clearly follows the GPM.

To summarize this update, the effect of including the coefficient update is to move the RUL towards an average training case. At the BOL, this movement results in very stable RUL estimates, compared to the GPM. This stability holds over most of the lifecycle. At the EOL it follows the GPM more closely.

4.1.3 Combined Coefficient and RUL Update on PHM Data

The GPM may suffer from some unpredictability, especially at the BOL. This unpredictability is partially because the GPM does not take advantage of all the degradation data available. The data required for a GPM can also be used to produce the prior distribution of path coefficients. Therefore, there are few reasons not to augment the GPM with either the transition Coble [Coble and Hines 2009] developed or the coefficient update outlined in this research. Both methods use the same data but adapt it to the GPM differently.

To take advantage of all data available, a combined RUL update is found using a type I prior and a coefficient-update likelihood. The lifecycle estimates are shown in Figure 4-12.

The combined updates exhibit similar patterns to each update shown previously. At the BOL, the combined updates follow the coefficient, and, thus type I closely because of the large variance in the GPM likelihood. Notably, this transition stays closer to the type I before time 50. At this point it continues to stay very close to the RUL update. It deviates at the GPM peak before time 200, near time 180. While the RUL update dips down towards the type I prior, due to the large uncertainty associated with the GPM peak, the combined transition compromises more with the GPM, and happens to stay near zero.

This combined transition shows potential to take advantage of each of the individual transitions' better attributes. At the BOL, this transition follows the stables

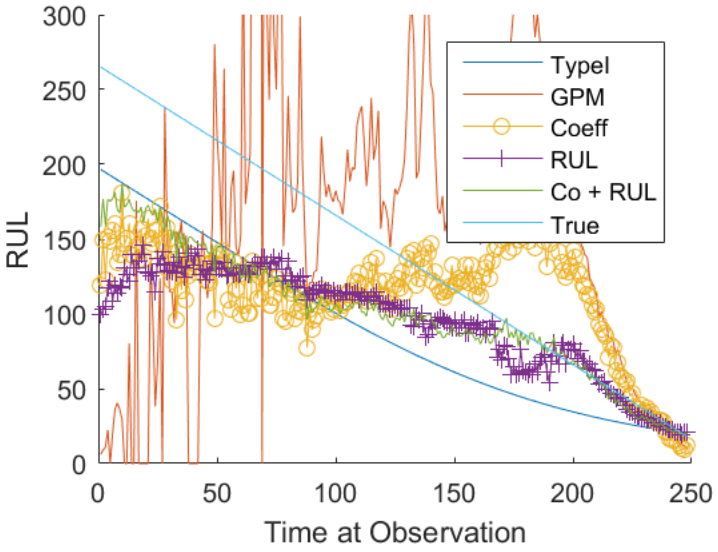


Figure 4-12: Combined Coefficient and RUL Updates

coefficient update. For the middle and EOL, it is robust to large errors. Then towards the EOL, it follows the RUL update more closely, which resists errors better than the coefficient update.

4.1.4 RUL regression on PHM Data

For this test case, the type I RUL is used as an additional data point when applying a GPM regression. With a WTLS iterative solution, the point is weighted by the uncertainty of the type I and failure threshold, meaning that the data point lies outside the prognostic path and is centered on the type I RUL estimate and the failure threshold. This method's results are shown in Figure 4-13.

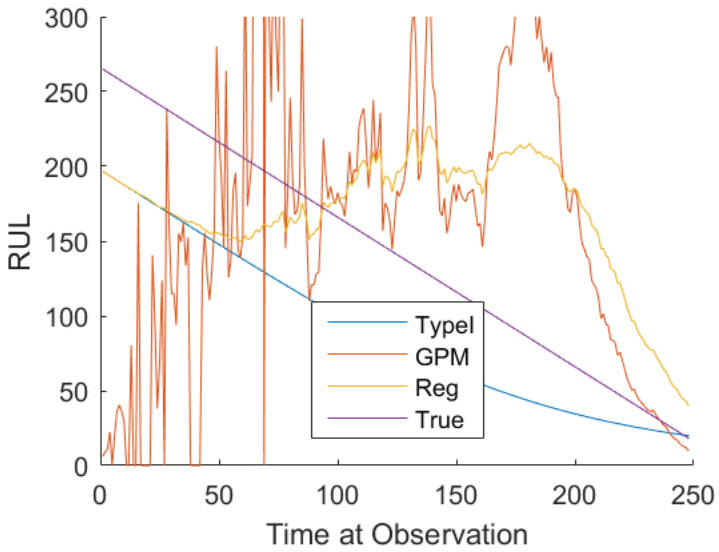


Figure 4-13: RUL Regression

At the BOL, the regression is very similar to the type I. At around time 50, the regression stops following the type I and begins to rise. After this period it seems to

follow the GPM more closely than the type I. At the peak in the GPM at time 200, while the regression transition partially resists the GPM, the estimate goes above time 200, showing it resists large errors near the EOL to a less extent than the other transitions previously shown. At the very EOL it also has an error not seen in the GPM or transitions. This error is due to the effect of a decreasing prior uncertainty at the EOL.

4.1.5 Error Analysis of PHM Data

To compare the different transitions across the testing population, the error is calculated as the RMSE for each set of predictions by using the entire length of the prognostic path until it is censored. This selection should capture snapshots of paths at different points in their lives. While some are censored early and late in life, the training population is approximated to cover a wide range of the lifecycle.

The RMSE is a useful error measurement with the same dimension as the RUL estimates. It starts by calculating the squared error between the true and predicted RUL. The squared error is averaged over all test cases i. Finally, it is square rooted to return the dimensions back to time, instead of time squared:

$$RMSE = \sqrt{mean((RUL_i^{pred} - RUL_i^{true})^2)} \qquad 4\text{-}2$$

Using the RMSE, the transition methods can be compared, as shown in Figure 4-14. The methods are, in order, the plain GPM, the RUL update, the coefficient update, the combined coefficient and RUL, and the RUL regression. As expected, the GPM performed the worst, because it does not benefit from all the information available. It's clear that the use of any of the transitions decreases the RMSE considerably.

While the RUL and coefficient updates reduce the RMSE from the GPM by 68% and 72% respectively, the combined transition further reduces the RMSE by 76%. The RUL regression had the largest error among the transitions, reducing the RMSE by 61%.

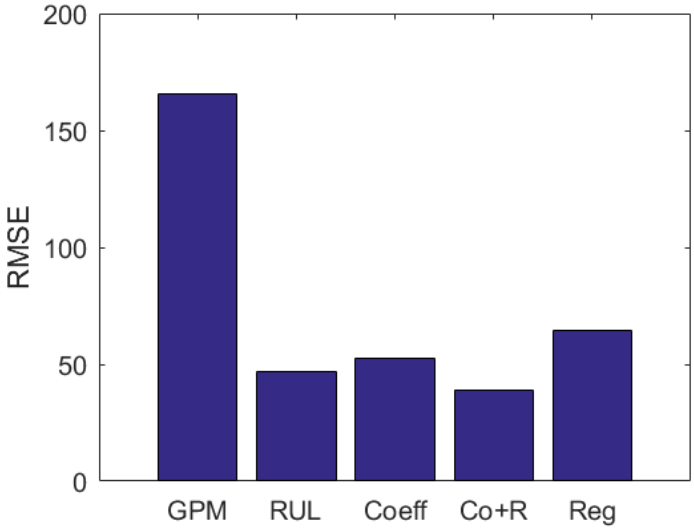


Figure 4-14: RMSE of Models for PHM Data

Providing more insight into the errors during the lifecycle, Appendix A contains more examples of the RUL evaluated at each observation. Most of the features noted in this chapter are consistent across test cases. At the BOL, the GPM is inaccurate, having a detrimental effect on the RUL prior. The RUL update, coefficient update, and combined transitions have stable predictions throughout the lifecycle. The RUL regression is also stable, but is prone to slightly higher errors at the EOL.

4.2 Heat Exchanger Test Bed

At the University of Tennessee's Knoxville campus, a heat exchanger (HX) test bed was designed to test prognostic methodologies and performance metrics on fouling degradation, with the goal of reducing uncertainty in RUL calculations [Welz et al., 2014]. Each heat exchanger was monitored as clay built up in the inner hot leg side, resulting in a loss of heat transfer efficiency.

The following describes the experimental method. The 64-tube shell and tube BASCO heat exchanger was fouled in regular intervals over two-week periods. The purpose was to model the heat exchangers' degradation due to surface deposition, a common problem with all heat exchangers, including those used in nuclear reactors and spent fuel pools. As mass accumulates on the inner walls, the heat transfer resistance increases; and the heat transfer between the hot and cold legs is reduced. To simulate this failure mode in the experimental setup, 105g of kaolin clay was deposited at the start of each cycle, with 75g of additional clay every 48 hours. Regularly adding clay kept the suspended clay's concentration high enough to continually deposit on the heat exchanger's surface, rather than aggregate and settle in different parts of the experimental setup.

The HX physical setup can be seen in Figure 4-15. The HX was attached to a 15-gallon heated tank. The hot and cold legs' inlet and outlet temperatures are critical in defining the health state along with the mass flow rates. Using these six signals, the HX's energy balance can be calculated.

For the 12 runs included in the analysis, a leave-one-out cross-validation strategy was adopted to conduct prognostics on each run. For example, when testing run 1, the monitoring and prognostic models were built using runs 2 through 12. This bootstrapping technique kept a priori knowledge separate from each test run.

The collection of signals into a prognostics parameter follows the same general analysis outlined for the PHM data. The unfaulted data was extracted from each run's BOL. By sorting and selecting fault-free observation vectors, an AAKR model was created. Fault detection for this case study was not used because the experimental outline calls for an initial amount of clay in the hot side loop. All runs start with some degradation. As represented in the AAKR residuals, this degradation was apparent soon after data collection.

The residuals were combined using the OLS technique to form the prognostic parameter. Based on multiple trials with different signals for the AAKR, the four temperature signals (hot leg 1 and 2 and cold leg 1 and 2) and the cold side heat transfer were chosen. Disregarding degradation, each test case was fouled over a period of roughly two weeks. (For more detailed information regarding the experimental set up, see Welz's et al. paper [2015].)

The temperature residuals for test case 2 is shown in Figure 4-16. It's seen that the yellow and purple cold leg temperatures slightly decrease to about 0.25 C below what is expected. On the other hand, the hot leg temperatures increase by a larger amount over the course of fouling. The orange hot leg out temperature reaches above 1 C. This increase is due to the reduction in heat transfer. The cold leg does not absorb as much heat, and the change in temperature for the cold leg is reduced. Therefore the

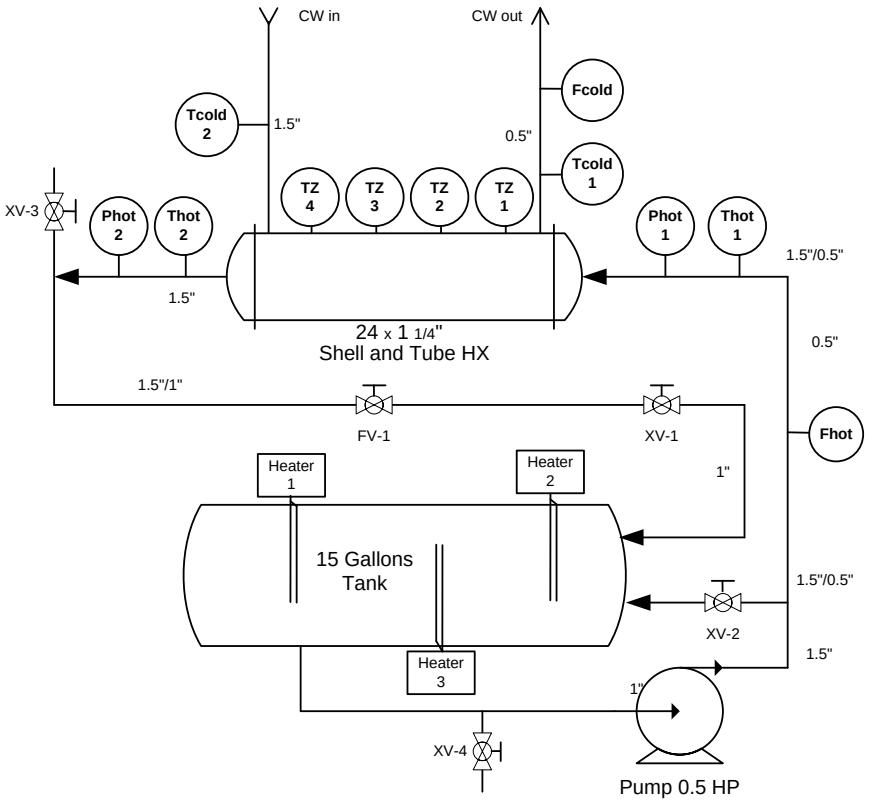


Figure 4-15: Heat Exchanger Schematics

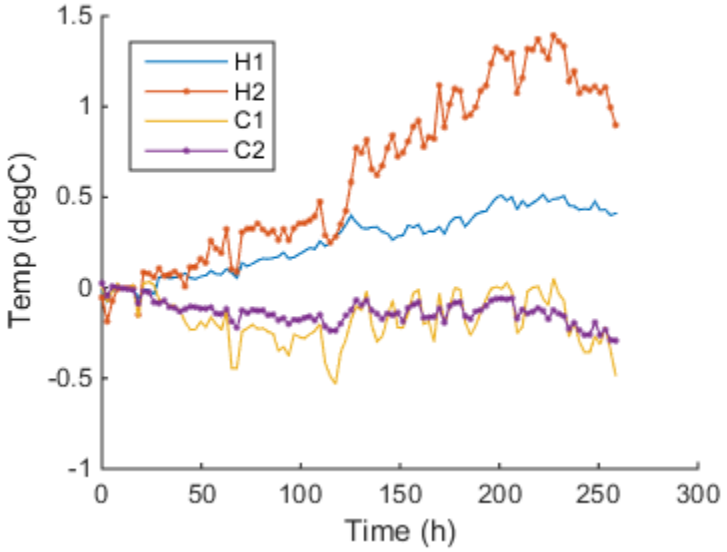


Figure 4-16: Temperature Residuals over Lifecycle

cold out temperature is also reduced. Conversely, the hot leg does not dissipate as much heat. Since the hot leg is in a close system, the excess heat gets recycled, which warms up the hot leg in even further, while the hot leg out drifts further from thermal equilibrium with the cold leg.

This particular test setup had no clear definition of failure; therefore, an artificial failure definition was adopted. A failure threshold was given to each run, based on the total observed degradation at the end of data collection. Taking a fraction of the total degradation would ensure that the failure threshold was crossed before the end of the run. This selection can be substituted for adjustable operating tolerances. The faulted tolerance can be selected and imposed onto the models.

Once the prognostic parameters were generated, the prognostic performance metrics were determined. The monotonicity, prognosability, and trendability were found to be 0.0296, 0.710, and 0.446, respectively. The monotonicity was very low, because of not only the process noise, but also some heat exchangers' self-healing.

During data collection, large downward dips in degradation occurred, as shown in Figure 4-17. This process of self-healing returns the HX to a less fouled state, likely caused by chunks of clay being pushed off the walls due to the circulating water or other clay chunks. As a result, the heat transfer efficiency is slightly restored until more clay is deposited.

4.2.1 Type I RUL Update of HX Data

Type I survival models were created using a Weibull distribution fit on the TOF, resulting in the RUL calculated as a conditional Weibull. Using all test cases, the Weibull distribution had shape and scale parameters of 2.57 and 267 hours, respectively, as shown in Figure 4-18. Because the test cases are run by bootstrapping, the individual type I distributions will vary slightly, depending on which test case is being analyzed. The type I model should be built by excluding the current test case's TOF.

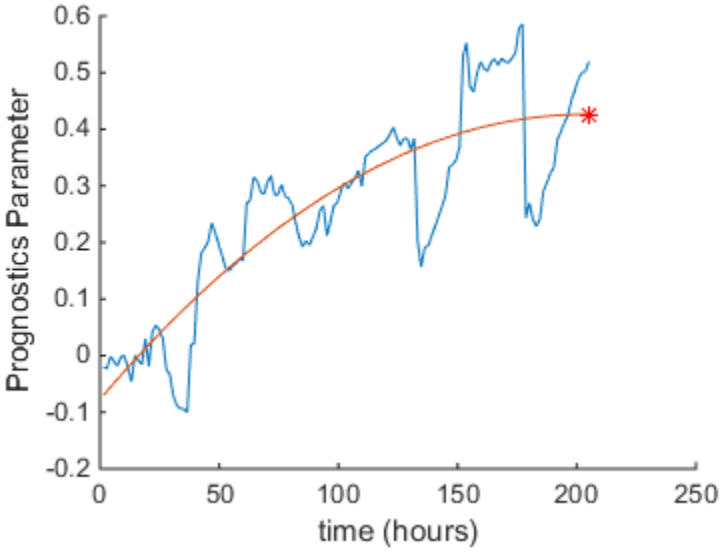


Figure 4-17: Test Run 9 with Self-Healing

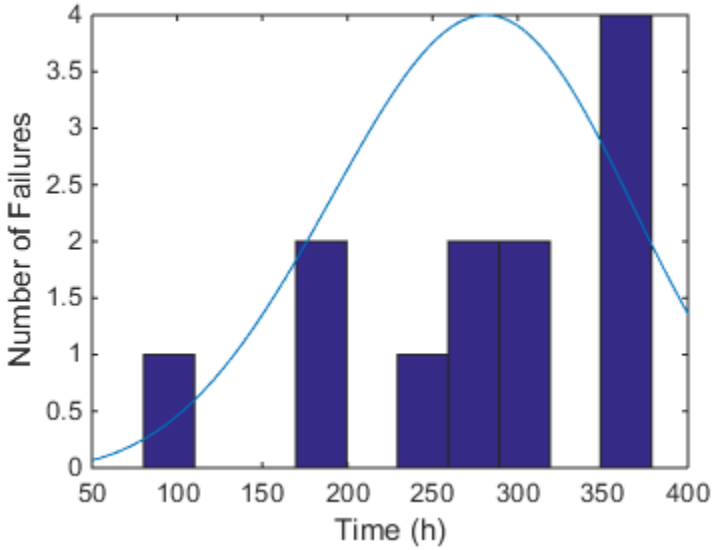


Figure 4-18: Heat Exchanger Time of Failure Distribution

For the type I, Weibull distribution, instead of a normal distribution, was chosen. While neither distribution matches the time of failure data exactly, the Weibull was considered to capture more of the asymmetry and the increased likelihood near 400 hours. As with the previous dataset, the prior is set as the type I output at each observation, and it is updated using the GPM likelihood based on MC sampling.

Figure 4-19 shows the RUL estimates of the GPM and RUL update. During the first few observations at the BOL, the GPM under-predicts a large amount, while the RUL update stays near the type I. At around 25 hours, the RUL update then compromises between the GPM and the type I. Just as for the previous dataset, the spikes in the GPM near 80 and 120 hours do not seem to have a big effect on the posterior. Towards the EOL, both models converge towards the true RUL.

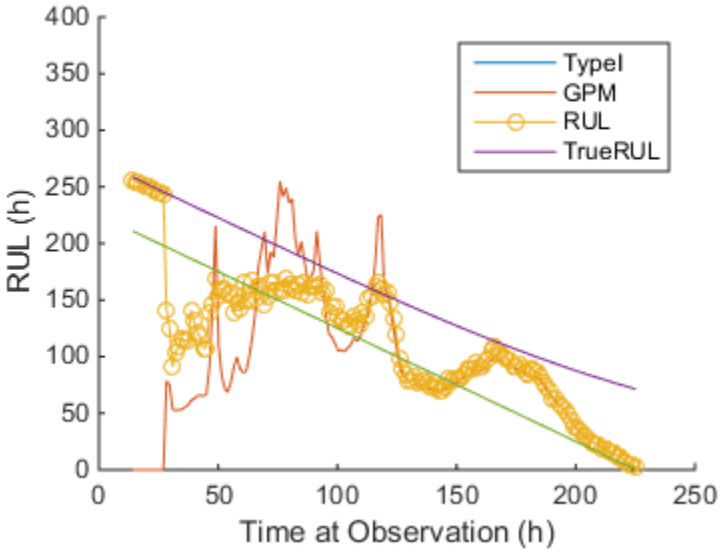


Figure 4-19: Test HX RUL Update

4.2.2 Coefficient Updates of HX Data

To apply the coefficient transition, the prior coefficient distributions must be made. As shown in Figure 4-20 through Figure 4-22, a functional fit would be difficult to find, given the small number of training cases.

As with the PHM analysis, these coefficients were updated using a multivariate kernel smoother, which quickly estimates the posterior while keeping the empirical distribution's shape with no need for a functional fit. For Figure 4-20, the gap at -1e-5 could be partially filled using kernel smoothing.

When the coefficient transition was applied to a test case, shown in Figure 4-23, the results were found over the lifecycle. Based on the over prediction of the type I, this test case has a shorter-than-average life, causing the coefficient update to over predict throughout the entire lifecycle. The large errors at the BOL for the coefficient update are due to the posterior sampling quadratic and linear coefficients near zero. Looking at only the quadratic, the sampling is a result of the prior distribution encompassing zero, and a weak GPM with an estimate far from the prior. The near zero sampling results in abnormally high RUL estimates.

The combined coefficient and TOF transitions were very similar to those found for the PHM dataset shown in Figure 4-24. One advantage is that the BOL estimates are stable, despite the large coefficient update peak at 20 hours. This trend was also seen in the PHM set because both of its constituent transitions' effects were to move the test case towards an average training prior. Thus, two combined transitions give more weight to the average prior assumption. The combined transition RULs are very stable at the BOL compared to the other transitions and GPM. At the EOL, the three transitions are very similar to each other. Note that the figure excludes the EOL for visual clarity of the beginning and middle.

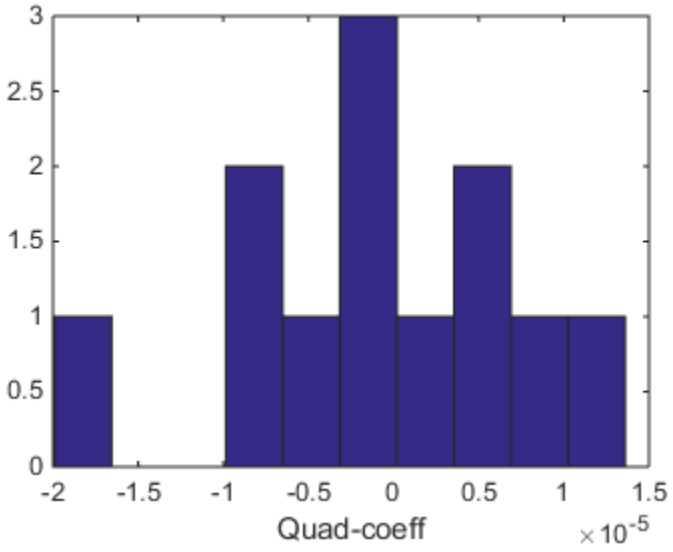


Figure 4-20: Histogram of Quadratic Prior

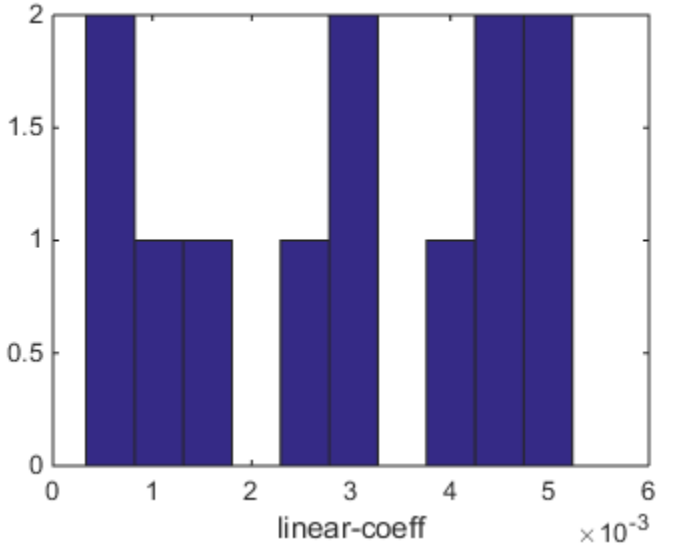


Figure 4-21: Histogram of Linear Prior

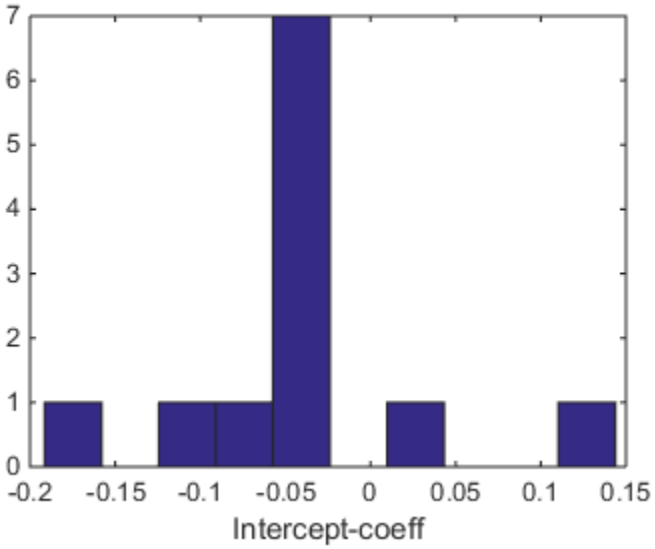


Figure 4-22: Histogram of Intercept Prior

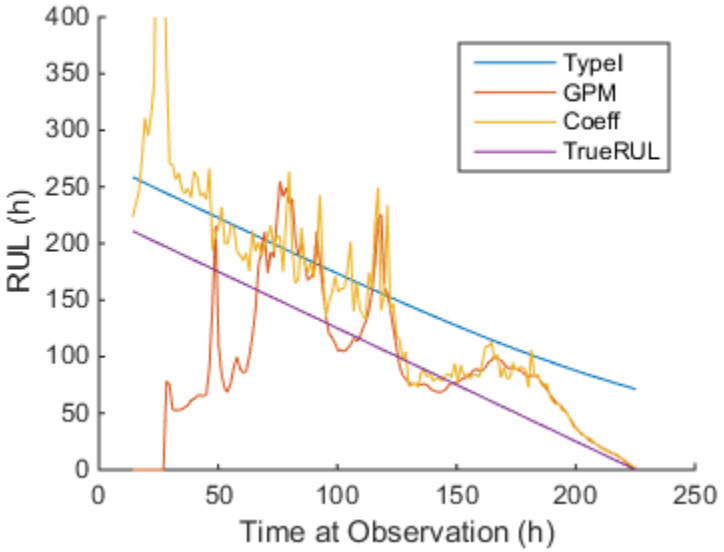


Figure 4-23: HX Coefficient Transition

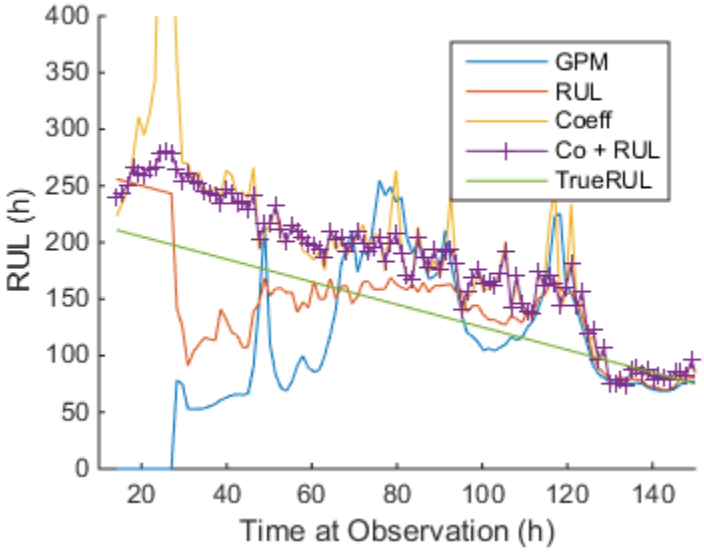


Figure 4-24: HX Combined Transition (Zoomed)

The combined transition accurately estimates the RUL at the BOL and holds consistently throughout the lifecycle. It resists the large peaks from the coefficient update. After the BOL it's generally a compromise between the RUL and coefficient updates.

4.2.3 RUL Regression of HX Data

Using the type I conditional Weibull as an additional point in a WTLS regression, the RUL regression transition was applied, as shown in Figure 4-25. The results are similar to those seen previously for this transition. Because the RUL considers type I information, it compromises the GPM with the type I. Over the entire lifecycle, the RUL regression generally follows the GPM, while staying close to the type I, which can be beneficial for an average case. However, because this case has a short life, the predictions are offset by the difference between the average TOF prior and the true TOF.

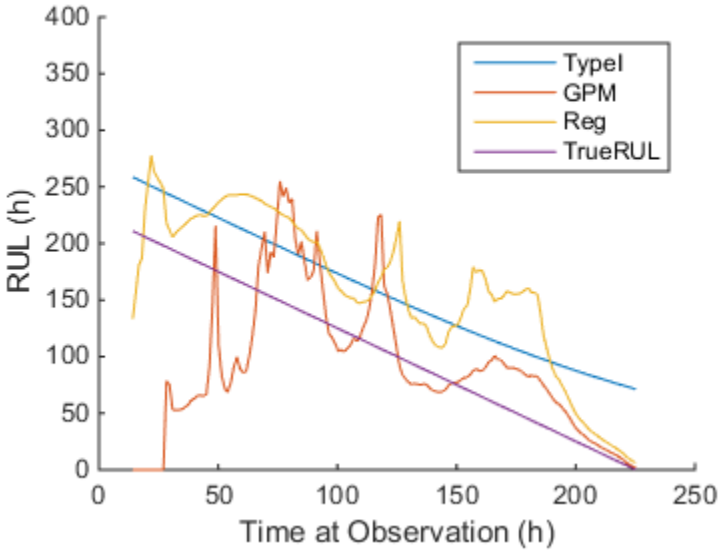


Figure 4-25: RUL Regression

4.2.4 Error Analysis of HX Data

To compare the transition models' overall errors, the RMSE for each test case and model were averaged over the lifecycle. This RMSE contains the model's error as an average over the observations in that case. Then for each of the 12 cases, if 5 transitions were tested, 60 RMSE estimates were found. To further compare each model, these RMSE were averaged over all testing cases, as shown in Figure 4-26. The methods presented are the plain GPM, the RUL update, the coefficient update, the combined coefficient and RUL, and the RUL regression.

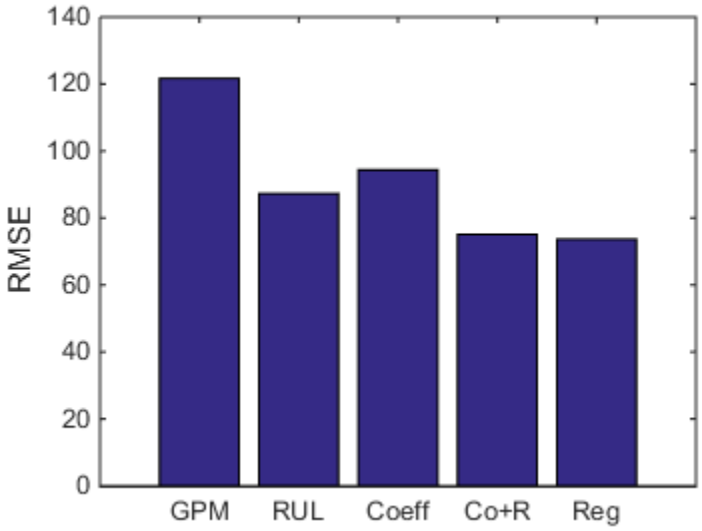


Figure 4-26: RMSE Over All HX

Just as in the PHM dataset's error analysis, the use of any transition reduces the RMSE. The reductions for the RUL, coefficient, combined coefficient and RUL, and regression transitions are 28%, 22%, 38%, and 39% respectively. The combined transition lowered the error further than either of its two constituents.

The regression transition outperformed the others for this dataset. This relative performance contrasts with the PHM dataset results. This can be the result of two simultaneous effects. The regression transition may have performed better for this dataset, due to the large prior uncertainty not impacting the EOL estimates as much. At the EOL, the GPM is preferred over the type I, as it has more information on the individual. The second effect is that the coefficient update was more prone to large errors at few specific points. While these large peaks are not very common, the large associated errors inflate the RMSE, relative to the other transitions.

4.3 Summary of Application Results

Two data sets were used to validate the application of five different models. For the PHM dataset, the combined coefficient and RUL update reduced the GPM RMSE by 76%. The RUL regression minimized the RMSE for the HX at a reduction of 39%. All five candidate models were analyzed by viewing the RUL outputs at each of the prognostic parameter's observations. This application sought to understand the prediction results as a function of life consumed. By comparing these RUL profiles, the advantages and disadvantages of using the transitions can be examined.

For the PHM dataset, a type I Weibull fit was made to the historical TOF. The survival probability's calculation was then based on the conditional Weibull. Such a prediction is not plagued by signal noise and does not fluctuate from observation to observation. It calculates the RUL based on time, rather than on information about system health or predicted future states.

The GPM errors were shown to be unpredictable, particularly at the BOL when few data points are available to support regression. The GPM is sometimes prone to sudden high estimates or zero values. By the middle to EOL, it is often comparable to

the other models. This outcome makes sense when considering the amount of information available. With much data, better predictions of the future prognostics parameter are possible. The GPM should perform at its best at the EOL. This model was included as a benchmark for comparing different transitioning methods.

The first transition introduced was the RUL update, which had some clear advantages. For example, it is flexible in terms of which prognostic models could be used to transition between, and it was easy to implement. While its BOL estimates were improved over the GPM, it was still somewhat susceptible to the large peaks that plagued the GPM. Towards the EOL, it matched the GPM closely.

Next, the coefficients update was used to augment the GPM. While this update generally improved the GPM's accuracy, it sometimes introduced errors by itself, due to sampling coefficients near zero. These few cases of high error negatively impacted the overall performance.

The coefficients update was used in conjunction with the RUL update to create a transition that uses all the information available. This combination had both models' advantages. At the BOL, the RUL was accurately estimated around the type I prediction. Throughout the lifecycle, it would be resilient to sharp changes in the prognostic parameter, which yields unreasonable RUL estimates. At the EOL it matched the RUL update more than it matched the GPM. This EOL behavior could sometimes introduce more errors, as the GPM is considered to be the most accurate.

Finally, the Bayesian regression was found to improve the RUL's accuracy over the entire lifecycle. Like the combined transition, the RUL regression's BOL estimates matched closely with the type I prior. However, the Bayesian regression was occasionally susceptible to fluctuations in the middle of life, independent from the GPM fluctuations. These large errors could be caused by the breakdown of one of the WTLS's assumptions towards the EOL. While the WTLS can model uncertainty in both the dependent and independent variables, it does so by using a normal distribution of mean

zero and some variance. In the two datasets observed, the Weibull's shape can loosely be described as "Gaussian-like" with two tails and an increased likelihood in the center. Although not fully symmetric, it approximates the sampling variance adequately enough. However, as the conditional Weibull was used as the RUL prior, the shape changed from Gaussian-like to a shape that more resembles a Gaussian cut in half. Thus, the assumption's breakdown may be an important source of error.

To understand why the posterior estimates change through time, knowing how the uncertainties, as well as the estimates, change over time is important. The uncertainty can be changed in both shape and magnitude. At later times, the ML is near zero, with a tail extending outward, such as in Figure 4-7. In this case, the mean estimate not only heads towards zero, but the variance is also reduced. If the original Weibull fit has a shape parameter of less than 1, then the opposite would be true with the variance increasing as the case survives.

To compare the models using a standard error weighting, the RMSE was presented as an average over all test cases. For the PHM dataset, this RMSE was calculated at the censored point, using the prognostics parameter's full length after fault detection. For the HX dataset, the RMSE was calculated at each point of observation over the lifecycle of all cases. In both of these comparisons, the combined coefficient and RUL update performed very well. For the HX, the RUL regression nearly beat the combined transition.

5. CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

This research developed methods for transitioning between prognostic types with the goal of utilizing all information available to improve RUL estimates. Within the lifecycle prognostics framework, important sources of information about the health of the system can be gathered over the lifecycle. Before the system is put into operation, historical TOF information can be gathered, and a type I reliability model can be applied. While the system is operating, the history of operating conditions can be tracked, to estimate the total stress given to the system. Finally, after a fault is detected, the measurable deviations from normal behavior can be tracked to give an overall estimate of system degradation. The unified quantification of that degradation is the prognostics parameter. Different prognostic methods can be applied at different times in the lifecycle, based on the availability of data. These different methods can be merged, improving the posterior estimates and reducing the uncertainty.

The type III GPM was especially used as an opportunity for applying transitions. After the prognostics parameter is obtained, an OLS model is applied to estimate the degradation path. This OLS fit can be any linear-in-terms fit, such as a linear, quadratic, exponential, etc. The fit is characterized by the estimated regression coefficients. The path can then be extrapolated to a failure threshold. A failure is assumed to occur when the parameter exceeds the threshold.

Transitions between prognostic types can be made using Bayesian statistics. This branch of statistics quantifies the expectation of parameters of interest as probability distributions instead of point mass estimates. This involves updating a prior belief with a

likelihood distribution given by a newly observed set of data. Updating in this way combines each estimate and the shape of its uncertainty as a posterior distribution.

There are several methods for the actual computation of the posterior. They differ in the necessary conditions and assumptions made to apply those transitions. Certain priors and likelihoods make up a conjugate family, for which an analytical solution can easily be referenced. However, this assumes that an appropriate fit can be made, and that the prior and likelihood both happen to fall within a conjugate family. Otherwise, numerical integration techniques can be used both to straightforwardly evaluate known fits and to estimate empirical probability densities, but are mostly used as a benchmarking technique. A trapezoidal approximation of the posterior integral was selected as an accurate and efficient algorithm. Otherwise a Simpson's rule, or other numerical integration approximations can be used. A MH algorithm can be applied for a known distribution regardless of its shape. While the MH algorithm requires some manual inputs, such as the selection of a jumping distribution, it can quickly sample from high dimensional spaces, and is a widely used technique. If smoothing an empirical distribution is desired, when a distribution fit can't be assigned, a kernel density smoother can be applied.

Three novel methods for incorporating different sources of information were presented and validated on two different data sets. The first transition, the RUL update, is the most flexible update. It uses the RUL estimates of any prognostic model as the prior and likelihood. The only assumption, other than those required to build the prognostic models, is that the RUL can be expressed as a PDF, instead of a point estimate. This is not a hard assumption to make, and can easily be fulfilled by any sophisticated prognostic model.

Two other transitioning methods were developed exclusively for transitions involving the GPM. The coefficient update involves updating the prognostics path regression coefficients using Bayes rule and prior coefficient information. If a quadratic

fit is used to extrapolate the prognostics parameter, then three prior populations of coefficients - quadratic, linear and intercept - can be recorded. This update is a simple application of Bayes rule, using any suitable method for calculating the posterior.

The last transition is termed RUL regression, and is a method of incorporating a prior RUL distribution into the GPM in a WTLS regression. The basic GPM regression is an OLS that assumes an equal and identical variance error that is added to predictors. This error is a normally distributed uncertainty around a mean of 0. The total least squares regression refers to error being modeled in both the independent and dependent variables. This contrasts with the OLS, which only assumes errors in the independent variable. The WTLS also models unequal variance in both sets of variables. In this way, each point of data in the regression can have different amounts of variance. The prior RUL is included as one additional data point in the GPM regression with uncertainty contributions in the independent and dependent variables. This point of data is centered around the mean RUL prior value, and the mean failure threshold. The variance comes from the prior RUL uncertainty and failure threshold uncertainty. While there is no closed form solution to the WTLS, an iterative algorithm can be computed until the coefficient estimates no longer appreciably change.

These transitions were applied to two different data sets, with similar conclusions. The GPM was used as a benchmark for how much the inclusion of the transition improves RUL estimation. It had the highest RMSE, and gave erratic RUL estimates especially in the BOL. It occasionally peaked and dipped between extreme values, before converging towards a more accurate RUL at the EOL.

The RUL update from a type I conditional Weibull was applied to great success. While it was still susceptible to the large peaks in the GPM, the RUL estimation was greatly improved over the course of a lifecycle. It was much less prone to jumping to extreme values, and consistently stayed near both the type I prior, and the GPM likelihood.

The coefficient update had some advantages and disadvantages. At the BOL the erratic GPM was replaced with very stable RUL estimates based around the type I prior. However, the coefficient update would sporadically jump at very high rates moving through the lifecycle. The large errors at the BOL for the coefficient update are due to the posterior sampling quadratic and linear coefficients near zero. The near zero sampling results in abnormally high RUL estimates.

The combined transition using the coefficient and RUL updates was found to take advantage of the positive attributes of both updates. When the coefficient error would peak, the combined would largely stay unaffected. It will use both constituent transitions to reduce the error over the lifecycle.

The RUL regression model by itself also produced stables estimate throughout the lifecycle. Notably it improved the BOL estimates, in addition to being more stable than the GPM over the lifecycle.

The use of any transition method was found to reduce the RMSE of the GPM by a significant margin. Each transition uses the stabilizing prior estimate at the beginning and middle of life. Towards the EOL all transitions will follow the GPM, resulting in accurate EOL estimates.

Original contributions to the field, developed in this dissertation, can be summarized as follows:

- Development of a framework to incorporate different sources of data across multiple prognostic models.
- Development of two transitioning methods between different prognostic methods.
    - RUL update applies Bayes updating on RUL estimates to form a posterior estimate.
    - RUL regression uses prior prognostic models in weighted total least squares regression.

- A coefficient update method for improving the GPM using Bayes formula on the prior distribution and regression estimate
- Development of MATLAB functions that can be applied to prognostic datasets. These include functions that codify Bayesian algorithms and adaptations of the PEP toolbox to accommodate transitions.
- Demonstration of error reduction on two test datasets: a publically available dataset of simulated faults and a heat exchanger test bed. An analysis of the error reduction is discussed and key areas of benefits are identified.

The transition methods were carried out with the assistance of MATLAB functions. The monitoring and prognostic models were created and run using the PEM and PEP toolboxes. To accommodate the transitions, changes to existing PEP functions were made, while maintaining the integrity of those functions to perform without transitions. Extensive modifications to the GPM functions were made to allow for optional transitions. These modifications were supplemented by the creation of new functions that calculates the posterior for various situations. These original functions include several conjugate families, numerical approximation, multivariate MH sampling, and multivariate Gaussian kernel smoothing.

5.2 Recommendations for Future Work

While several Bayesian transition methods have been defined and shown to work for the current applications, there is room to explore further methods. One area of interest is obtaining the distribution fits of specific transitions. In some cases, such as the conditional type I, the distribution fit can be derived analytically, based on the prognostic model used.

However, there are situations in which the distribution fit is unknown. These include the coefficient update priors. Some guidelines on the appropriate selection of fits that can be made on this distribution can be explored. For example, the criteria used to determine whether a fit is a good enough to approximate the distribution can be defined, or given guidelines. In addition, different candidate functions can be explored from outside the well known Gaussian, lognormal, exponential, and Weibull distributions. If an empirical kernel function is used, some exploration into different kernels, other than the Gaussian, as well as bandwidth selection, can be explored.

While the RUL update was found to apply in many situations with few assumptions, the coefficients update and the Bayesian regression could be improved. The main shortcoming of the coefficient update is that sometimes the posterior would sample near zero values for the quadratic and linear coefficients. This can be avoided with a better fit to the priors that decreases the probability in the near double zero region.

For the RUL regression, it is assumed that the uncertainty of each point can be modeled using Gaussian noise. At the BOL, for certain Weibull TOF distributions, this assumption can be approximately valid. Thus the uncertainty of the RUL prior is appropriately defined by the variance. However, towards the EOL, this prior distribution was shown to change shape and no longer be a symmetric Gaussian. It's a skewed distribution that estimates a near RUL, with one tail that extends to the future. In addition, this prior should not contain any negative or 0 RUL estimates, or at least should not give weight to such estimates. While the GPM likelihood should ensure the non-negative characteristic of the RUL, this change in shape of the estimated prior point is a violation of the WTLS assumption.

Additional methods of transitioning can also be developed for specific prognostic models. This is similar to how the coefficients update and RUL regression were tailored to be used in the GPM. In the same way specific transitions into prognostic models can

be made, with each new transition, the diversity of combinations of transitions can be increased.

And finally, although the methods were applied to two datasets, a more ideal dataset can be used to validate the models. The main drawback to working with the datasets presented is that all the data used is drawn from a single source. For a more realistic application it should be expected that the TOF information is obtained separately from the degradation. Few examples of a system degrading towards failure can exist due to the high cost of obtaining such information. However, it's expected that TOF information is more readily available, either from the manufacturer or from the operator's own maintenance logs. This TOF information should then be used in conjunction with the prognostic model of choice.

LIST OF REFERENCES

A. C. Aitken, "On Least-squares and Linear Combinations of Observations," *Proceedings of the Royal Society of Edinburgh*, vol. 55, pp. 42–48, 1934.

M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174 - 188, 2002.

J. Bates and C. Granger, "The combination of forecasts," *Operations Research Quarterly,* vol. 20, pp. 451 - 468, 1969.

L. E. Baum and T. Petrie, "Statistical Inference for Probabilistic Functions of Finite State Markov Chains," The Annals of Mathematical Statistics, vol. 37, no. 6, pp. 1554–1563, 1966.

L. J. Bond, S. R. Doctor, D. B. Jarrell, J. W. D. Bond, "Improved economics of nuclear plant life management," PNNL, Richland, WA, Rep. CN-155-008KS, 2007.

L. J. Bond, P. Ramuhalli, M. S. Tawfik and N. J. Lybeck, "Prognostics and life beyond 60 years for nuclear power plants," *IEEE Conference on Prognostics and Health Management,* Montreal, QC, pp. 1 - 7, 2011.

G. Casella and R.L. Berger, *Statistical Inference* (2nd ed.), Duxbury Press, 2001.

S. Chib and E. Greenberg, "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, vol.49, no. 4, pp. 327 - 335, 1995.

R. B. Chinnam, "On-line reliability estimation of individual components, using degradation signals," *IEEE Transactions on Reliability*, vol. 48, no. 4, pp. 403 - 412, 1999.

J. Coble and J. W. Hines, "Applying the general path model to estimation of remaining useful life," *International Journal of Prognostics and Health Management*, vol. 2, 2011.

J. Coble and J. W. Hines, "Identifying optimal prognostic parameters from data: a genetic algorithms approach," *Annual Conference of the Prognostics and Health Management Society,* San Diego, CA, 2009.

J. Coble and J.W. Hines, "Development of a prognostics toolbox with application to GPS degradation data," *Conference of the Society for Machinery Failure Prevention Technology*, Dayton, OH, Apr, 2009.

J. T. Connor and R. D. Martin, "Recurrent neural networks and robust time series prediction," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 240 - 254, 1994

D. R. Cox, "Regression models and life tables (with discussion)," *Journal of the Royal Statistical Society*, *Series B*, 34, pp. 187 - 220, 1972.

C. E. Ebeling, *An Introduction to Reliability and Maintainability Engineering*, Long Grove, Il, Waveland Press INC, 2010.

A. K. Garga. K. T. McClintic, R. L. Campbell, C. Yang, M. S. Lebold, "Hybrid reasoning for prognostic learning in CBM systems," *IEEE Aerospace Conference*, Big Sky, MT, pp. 2957 - 2969, 2001.

D.R. Garvey and J.W. Hines, "Process and equipment monitoring methodologies applied to sensor calibration monitoring," *Quality and Reliability Engineering International*, vol. 23, no. 1, pp. 123 - 135, 2007.

J. C. F. Gauss, *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*, 1809.

A. Gelman, J. Carlin, H. Stern, and D. Rubin, *Bayesian Data Analysis,* 2nd ed. Boca Raton: Chapman and Hall/CRC, Boca Raton, USA, 2004.

J. Ghosh, M. Delampady, and T. Samanta, *An Introduction to Bayesian Analysis*, Springer, New York, 2006.

Z. Griliches and V. Ringstad, "Errors-in-the-variables bias in nonlinear contexts," *Econometrica*, vol. 38, no. 2, pp. 368 - 370, 1970.

K. Goebel, N. Eklund, and P. Bonanni, "Fusing competing prediction algorithms for prognostics," *IEEE Aerospace Conference*, Big Sky, MT, 2005.

N. J. Gordon, D. J. Salmond and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEE-Proceedings-F*, vol. 140, pp. 107–113, 1993.

C. B. Guure and N. A. Ibrahim, "Bayesian analysis of the survival function and failure rate of Weibull distribution with censored data," *Mathematical Problems in Engineering*, 2012.

W. K. Hastings, "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97 - 109, 1970.

W. He, N. Williard, M. Osterman, and M. Pecht, "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method," *Journal of Power Sources*, vol. 196, pp. 10314 - 10321, 2011.

J. W. Hines and A. Usynin, "Current computational trends in equipment prognostics," *International Journal of Computational Intelligence Systems*, vol. 1, no. 1, pp. 94 - 102, 2008.

J. W. Hines and D. R. Garvey, "Development and application of fault detectability performance metrics for instrument calibration verification and anomaly detection," *Journal of Pattern Recognition Research*, vol. 1, no. 1, pp. 2 - 15, 2006.

D. Husmeier, *Neural Networks for Conditional Probability Estimation*, Springer-Verlag, London, 1999.

L. C. Jaw, "Neural networks for model-based prognostics," *IEEE Aerospace Conference*, vol. 3, pp. 21 - 28, Aspen, CO, 1999.

B. Jeffries, J. W. Hines, A. Nam, M. Sharp and B. Upadhyaya, "Lifecycle prognostic model development and initial application results," *International Symposium on Future I&C for Nuclear Power Plants/International Symposium on Symbiotic Nuclear Power Systems*, Jeju, Korea, 2014.

J. M. Karandikar, A. Abbas, and T. L. Schmitz, "Remaining useful tool life predictions in turning using Bayesian inference," *International Journal of Prognostics and Health Management*, vol. 4, no. 2, July, 2013.

J. P. Kharoufeh and S.M. Cox, “Stochastic models for degradation-based reliability,” IIE Transactions, vol. 37, pp. 533 - 542, 2005.

J. R. Koza, *Genetic Programming*, The MIT Press, Cambridge, MA, 1992.

Y. Liang and X. Liang, "Improving signal prediction performance of neural networks through multiresolution learning approach," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 36, no. 2, pp. 341 - 352, 2006.

H. Liao, W. Zhao and H. Guo, "Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model," Proceedings of the Reliability and Maintainability Symposium, pp. 127 - 132, 2006.

J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032 - 1044, 1998.

C. J. Lu and W. Meeker, “Using degradation measures to estimate a time-to-failure distribution,” *Technometrics*, vol. 35, no. 2, pp. 161 - 174, 1993.

I. Markovsky, M. L. Rastello, A. Premoli, A. Kukush and S. V. Huffel, "The element-wise weighted total least-squares problem," *Computational Statistics & Data Analysis*, vol. 50, no. 1, pp. 181 - 209, 2006.

L. M. de Menezes, D. W. Bunn and J. W. Taylor "Review of guidelines for the use of combined forecasts," vol. 120, no. 1, pp. 190 - 204, 2000.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087 - 1092, 1953.

S. Mishra, S. Ganesan, M. Pecht and J. Xie, "Life consumption monitoring for electronics prognostics," *Proceedings of the IEEE Aerospace Conference*, pp. 3455 - 3467, 2004.

T. M. Mitchell, *Machine Learning*, McGraw-Hill, 1997.

E. H. Moore, "On the reciprocal of the general algebraic matrix," *Bulletin of the American Mathematical Society*, vol. 26, no. 9, pp. 394–395, 1920.

P. D. Moral, A. Doucet and A. Jasra, "Sequential Monte Carlo samplers," *Journal of the Royal Statistical Society*, vol. 68, pp. 411 - 436, 2006.

A. Nam, B. Jeffries, J.W. Hines, B. Upadhyaya, "A Bayesian statistical updating method: error reduction in remaining useful life estimation," *Nuclear Plant Instrumentation, Control & Human Machine Interface Technologies*, Charlotte, NC, 2015

A. Nam, M. Sharp, J.W. Hines, B. Upadhyaya, "Prognostic types: lifecycle prognostics," *Nuclear Plant Instrumentation, Control & Human Machine Interface Technologies*, San Diego, CA, 2012.

A. Nam, M. Sharp, J.W. Hines, B. Upadhyaya, "Lifecycle prognostics for asset management," *Maintenance and Reliability Conference*, Knoxville, TN, 2013.

A. Nam, M. Sharp, J.W. Hines, B. Upadhyaya, "Lifecycle prognostic algorithm development and application to test beds," *Chemical Engineering Transactions,* vol. 33, pp. 901 - 906, 2013.

M. Nelson and K. Mason, "A model-based approach to information fusion," *Proceedings of Information, Decision, and Control*, Adelaide, SA, pp. 395 - 400, 1999.

M. Orchard, G. Kacprzynski, K. Goebel, B. Saha and G. Vachtsevanos, "Advances in uncertainty representation and management for particle filtering applied to prognostics," *International Conference on Prognostics and Health Management,* Denver, CO, pp. 1 - 6, 2008.

R. Orsagh, M. Roemer, J. Sheldon and C. J. Klenke, "A comprehensive prognostics approach for predicting gas turbine engine bearing life," *Proceedings of IGTI TurboExpo*, Vienna, 2004.

E. Parzen, "On Estimation of a Probability Density Function and Mode". The Annals of Mathematical Statistics, vol. 33, no. 3, pp. 1065, 1962.

M. Pecht, D. Das, and A. Ramakrishnan, "The IEEE standards on reliability program and reliability prediction methods for electronic equipment," *Microelectronic Reliability*, vol. 42, pp. 1259 - 1266, 2002.

R. Penrose, "A generalized inverse for matrices," *Proceedings of the Cambridge Philosophical Society*, vol. 51, pp. 406–413, 1955.

H. Raiffa and R. Schlaifer, "Applied Statistical Decision Theory," Graduate School of Business Administration, Harvard University, Cambridge, MA, 1961.

A. Ramakrishnan and M. G. Pecht, "A life consumption monitoring methodology for electronic systems," *IEEE Transactions on Components and Packaging Technologies*, vol. 26, no. 3, pp. 625 - 634, 2003.

N. S. V. Rao, "Finite sample performance guarantees of fusers for function estimators," *Information Fusion*, vol. 1, no. 1, pp. 35-44, 2000.

M. Rosenblatt, "Remarks on Some Nonparametric Estimates of a Density Function". The Annals of Mathematical Statistics, vol. 27, no. 3, pp. 832, 1956.

M. N. Rosenbluth, A. W. Rosenbluth, "Monte-Carlo calculations of the average extension of macromolecular chains," *Journal of Chemical Physics*, vol. 23, pp. 356–359, 1955.

S. Saha, Y. Boers, J. N. Driessen, P. K. Mandal, and A. Bagchi, "Particle filter based MAP state estimation: a comparison," *12th International Conference on Information Fusion,* Seattle, WA, pp. 278 - 283, 2009.

A. Saxena and K. Goebel, "PHM08 Challenge Data Set", NASA Ames Prognostics Data Repository (http://ti.arc.nasa.gov/project/prognostic-data-repository), NASA Ames Research Center, Moffett Field, CA, 2008, accessed April 2016.

A. Saxena, J. Celaya, E. Balaban, K. Goebel, B. Saha, S. Saha, and M. Schwabacher, "Metrics for evaluating performance of prognostic techniques," *International*

Conference on Prognostics and Health Management, Denver, CO, pp. 1 - 17, 2008.

B. Schaffrin, I. P. Lee, Y. Felus and Y. S. Choi, "Total least-squares for geodetic straight-line and plane adjustment," *Bollettino di Geodesia e Scienze Affini*, vol. 65, pp. 141 - 168, 2006.

B. Schaffrin and A. Wieser, "On weighted total least-squares adjustment for linear regression," *Journal of Geodesy*, vol. 82, no. 7, pp. 415 - 421, 2008.

M. Schwabacher and K. Goebel, "A survey of artifical intelligence for prognostics," NASA Ames Research Center, Moffett Field, CA, Rep. MS 269-3, 2007.

M. Sharp, J. Coble, A. Nam, J. W. Hines, and B. Upadhyaya, "Lifecycle prognostics: transitioning between information types," *Institution of Mechanical Engineers Part O: Journal of Risk and Reliability*, vol. 229, no. 4, pp. 279 - 290, 2014.

T. Shen, F. Wan, W. Cui, B. Song, "Application of prognostic and health management Technology on aircraft fuel system," *Prognostics and Health Management Conference*, Macao, Jan, 2010.

B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC. London, 1986.

J. S. Simonoff, *Smoothing Methods in Statistics*. Springer, New York, 1996.

P. Tchakoua, R. Wamkeue, M. Ouhrouche, F. Slaoui-Hasnaoui, T. A. Tameghe and G. Ekemb, "Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges," *Energies*, vol. 7, no. 4, pp. 2595 - 2630, 2014.

B. R. Upadhyaya, M. Naghedolfeizi and B. Raychaudhuri, "Residual life estimation of plant components," *P/PM Technology*, vol. 7, no. 3, pp. 22 - 29, 1994.

P. V. Varde and M. G. Pecht, "Role of prognostics in support of integrated risk-based engineering in nuclear power plant safety," *International Journal of Prognostics and Health Management*, vol. 3, no. 1, April, 2013.

C. T. Volinsky, D. Madigan, A. E. Raftery, and R. A. Kronmal, "Bayesian model averaging in proportional hazards models: assessing the risk of a stroke," *JSTOR Applied Statistics,* vol. 46, no. 4, pp. 433 - 448, 1997.

A. Wald, *Sequential Analysis*, John Wiley & Sons, New York, 1947.

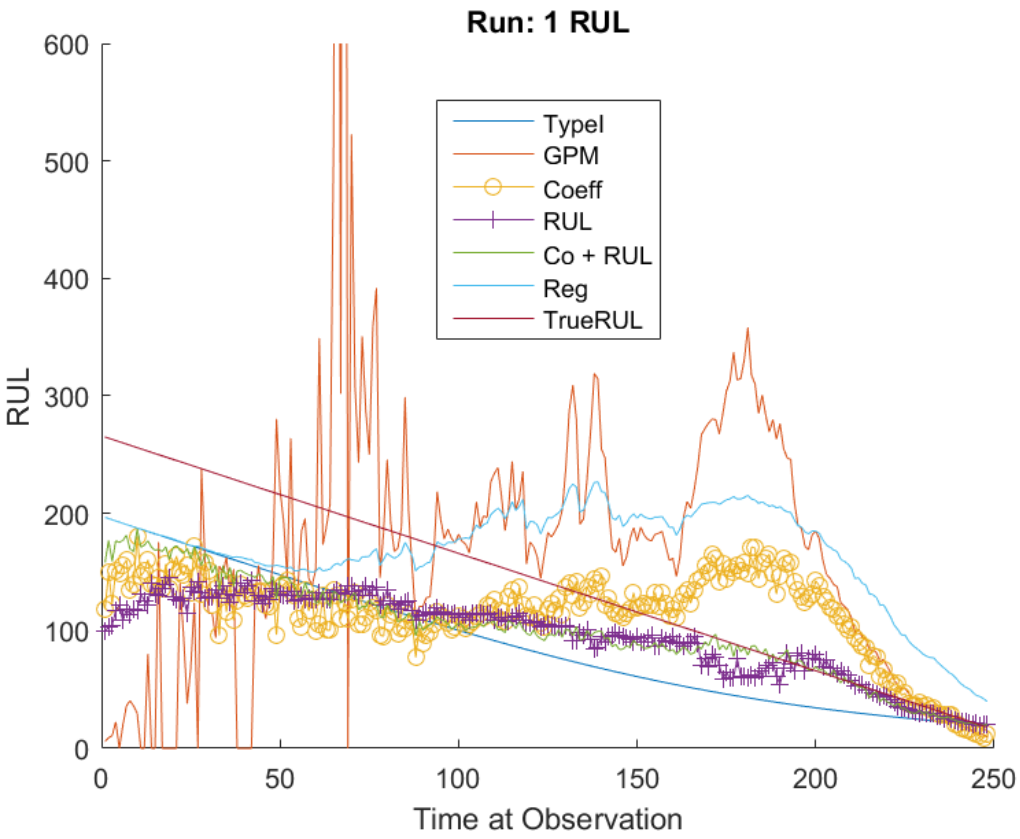
W. Weibull, "A statistical distribution function of wide applicability," *Journal of Applied Mechanics Transactions ASME,* vol. 18, no. 3, pp. 293 - 297, 1951.

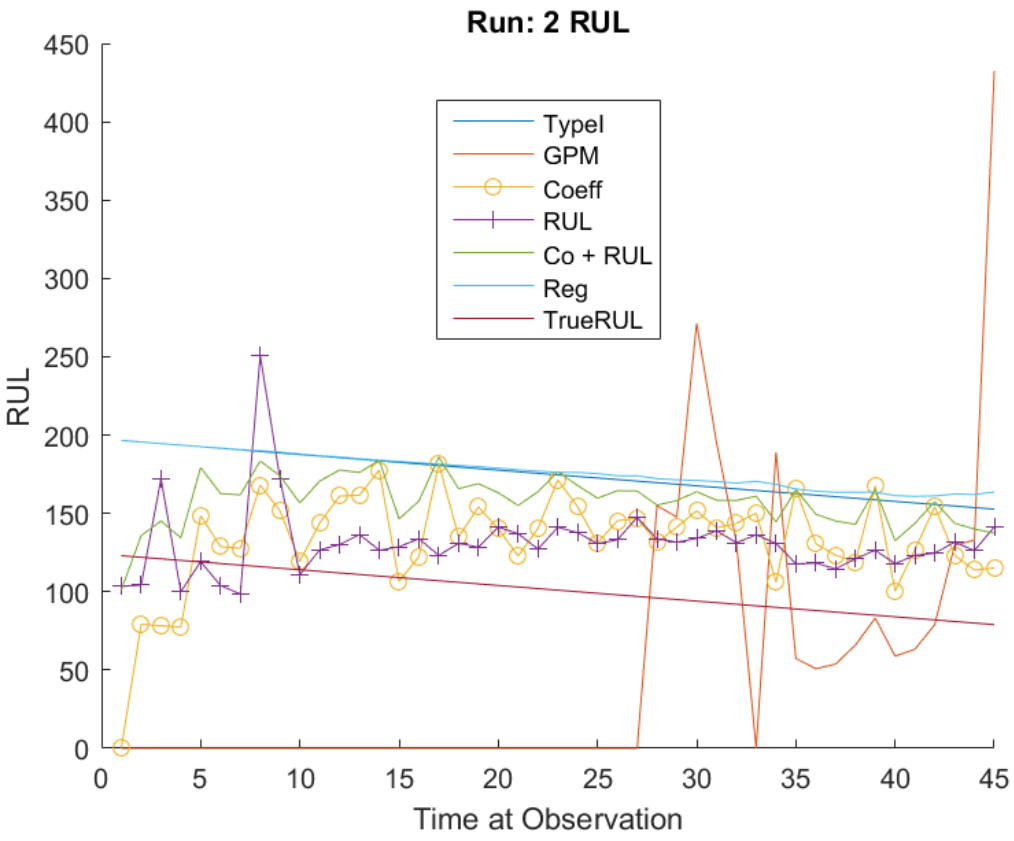
Z. Welz, A. Nam, M. Sharp, J. W. Hines, and B. R. Upadhyaya, "Prognostics for light water reactor sustainability: empirical methods for heat exchanger prognostic lifetime predictions," *Prognostics and Health Management Europe*, Milan, Italy, 2014.

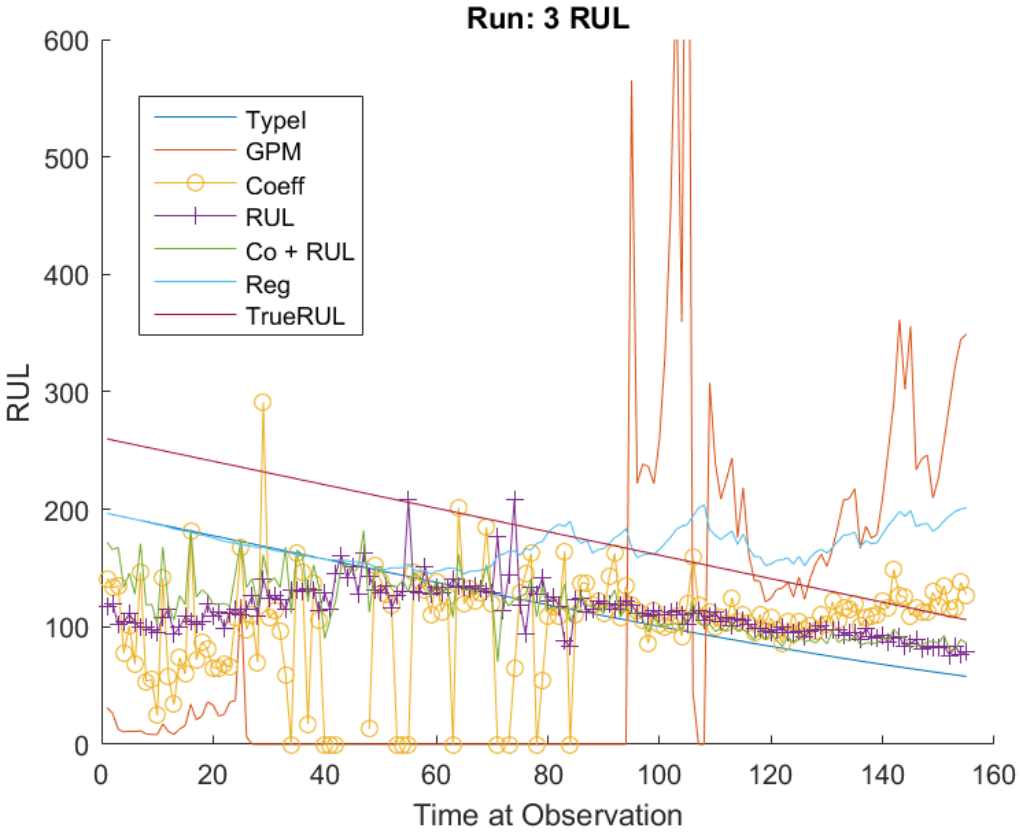
Z. Welz, A. Nam, M. Sharp, J.W. Hines, B. Upadhyaya, “Improved heat exchanger lifecycle prognostic methods for enhanced light water reactor sustainability," *International Journal of Prognostics and Health Management,* vol. 6, 2015.

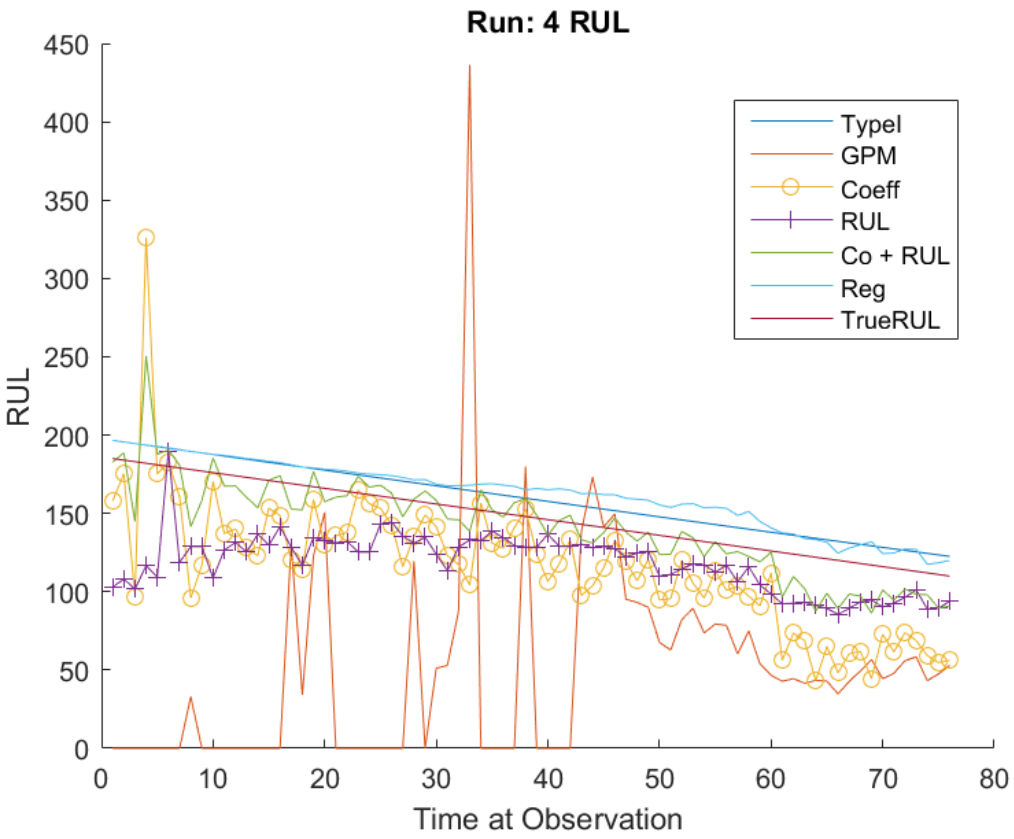
E. Zio and G. Peloni, "Particle filtering prognostic estimation of the remaining useful life of nonlinear components," *Reliability Engineering and System Safety*, vol. 96, pp. 403-409, 2011.

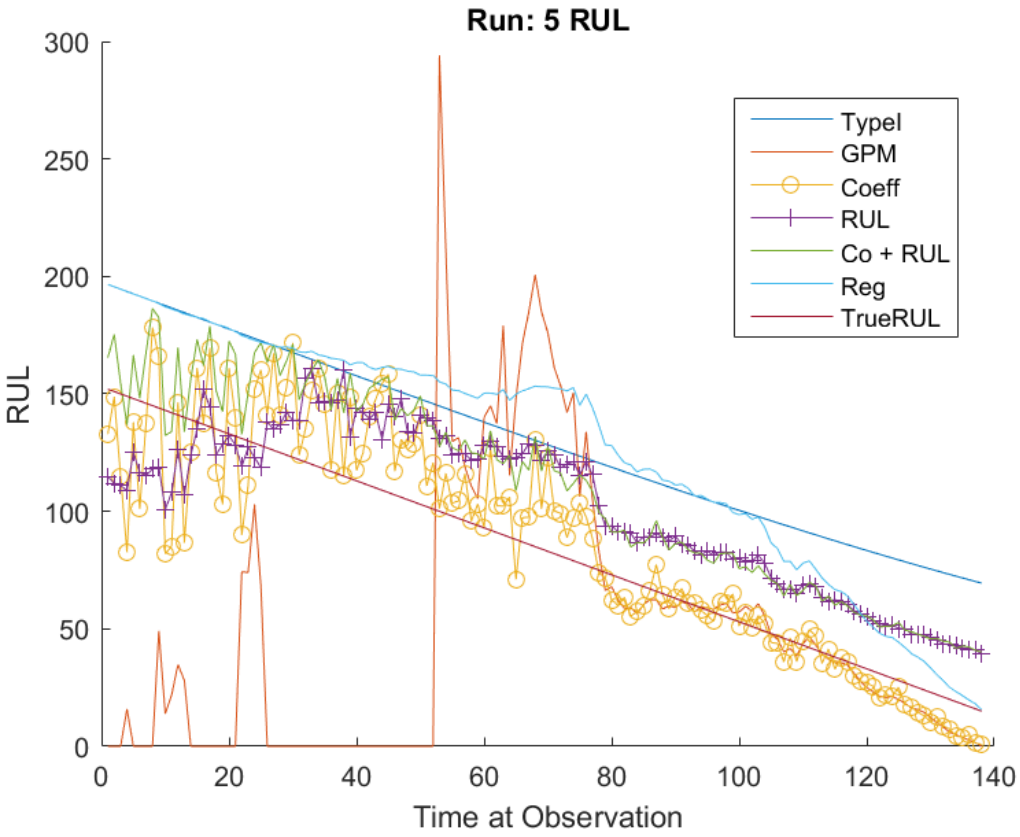
APPENDIX


Run: 1 RUL
TypeI
GPM
Coeff
RUL
Co + RUL
Reg
TrueRUL
RUL
Time at Observation
0
100
200
300
400
500
600
0
50
100
150
200
250


Run: 2 RUL
TypeI
GPM
Coeff
RUL
Co + RUL
Reg
TrueRUL
RUL
Time at Observation
0
50
100
150
200
250
300
350
400
450
0
5
10
15
20
25
30
35
40
45


Run: 3 RUL
Typel
GPM
Coeff
RUL
Co + RUL
Reg
TrueRUL
RUL
Time at Observation
0
100
200
300
400
500
600
0
20
40
60
80
100
120
140
160


Run: 4 RUL
RUL
Time at Observation
TypeI
GPM
Coeff
RUL
Co + RUL
Reg
TrueRUL
0
50
100
150
200
250
300
350
400
450
0
10
20
30
40
50
60
70
80


Run: 5 RUL
RUL
Time at Observation
TypeI
GPM
Coeff
RUL
Co + RUL
Reg
TrueRUL
0
50
100
150
200
250
300
0
20
40
60
80
100
120
140

VITA

Alan Nam was born and raised in Los Angeles, California. In high school he competed in science competitions like the Chemistry Olympiad, winning a first place medal, and Science Bowl, for which his team came in fourth.

He studied as an undergrad for 5 years at UCSD, earning his bachelor's in chemical engineering with a minor in management science. He worked part time as a lab assistant in micro RNA research.

In 2013, he received his Master's in Nuclear engineering from UTK, for his work on Bayesian prognostic techniques. Since then his projects have included work on a heat exchanger test bed for which he won a "Best Paper" award at the PHM Europe conference in 2014. He helped to develop high performance radioactive nanoparticles for particle flow tracking. Currently he works at the Reliability and Maintainability Center as a research assistant on reliability and maintainability practices.