

**Integrating Disparate Nuclear Data Sources for
Improved Predictive Maintenance Modeling:
Maintenance-Based Prognostics for Long-Term
Equipment Operation**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Zachary Allen Welz
August 2017

Copyright © 2017 by Zachary A. Welz
The University of Tennessee, Knoxville
All rights reserved.

DEDICATION

To my grandfather, without whom I would have never discovered the secrets of science and discovery.

To my mother, who has continually sacrificed more for me than I could ever begin to fully appreciate.

&

To my amazing wife, who has supported me and given me positivity through many challenging times.

Thank you.

PERSONAL ACKNOWLEDGEMENTS

First, I would like to acknowledge my incredible dissertation committee Dr. J Wesley Hines, Dr. Jamie Coble, Dr. Belle Upadhyaya, Dr. James Ostrowski, and Dr. Richard Wood. Their continual advice and support has been a great comfort through this challenging process. I would like to extend further thanks to Dr. Hines for giving me the opportunities I have had throughout my experience at UT. I would also like to thank a few other faculty members that guided and instructed me through the years: Dr. L Heilbronn, Dr. L Miller, Dr. O Chvála, Dr. E Lukosi, Dr. K Jones, and Dr. G Peterson.

The staff in the UTNE department was always a bright and positive support system within the department. Their help with everything from printing to travel requests was a constant aide. Thanks to those current and past staff members that made me feel like part of the UTNE family: Lydia Salmon, Gisela Goeritz, and Kristin England.

Lastly, a great thanks to all my friends and family who have gone on this adventure with me. I have had nothing but amazing experiences to thank you for. To my peers, you have been as instrumental in my success as most professors. I thank you for the countless all-nighters, and the fun times watching football on Saturdays. To my family, thank you for not giving up on me while it has been radio-silence for 7 years. I look forward to rejoining the family circle.

RESEARCH ACKNOWLEDGEMENTS

I would like to acknowledge the Lloyds Register Foundation and the International Joint Research Center for the Safety of Nuclear Energy for funding of this research. Lloyd's Register Foundation helps to protect life and property by supporting engineering-related education, public engagement, and the application of research.

I would also like to acknowledge the DOE Nuclear Engineering University Program (NEUP) under prime contract number DE-AC07-05ID14517 and Battelle Energy Alliance LLC Standard Contract No. 00118294, for funding of the original heat exchanger research.

Lastly, I would like to thank General Electric's Power Division, formerly Alstom Power, for providing data that was key in validating research efforts.

ABSTRACT

The United States (US) nuclear industry is one of the most heavily regulated businesses in the world, creating a culture of world-class design, operation, and maintenance. In an article published on modern maintenance technologies, Terrence O'Hanlon (past Chief Asset Manager for Reliabilityweb.com) stated, "world class companies often devote up to 50 percent of their entire maintenance resources to condition based monitoring and the planned work that is required as a result of the findings" [1]. One would expect US nuclear power plants to constantly upgrade, improve, and expand their operations and maintenance departments and tactics. Since the early 1990s, US nuclear plant expenses due to operations and maintenance have increased by over 10% and were estimated at \$20.62/MWhr (>16 billion USD) in 2015 [2]. While costs are increasing, and supporting technologies are more readily available than ever, plants commonly rely on reactive and basic preventive maintenance techniques.

This dissertation investigates improved maintenance practices by establishing baseline performance capabilities only possible with advanced maintenance strategies. A method of extracting plant data to facilitate predictive modeling is introduced. This method utilizes information (condition data and maintenance data) from disparate sources within modern nuclear plants to extract failure cycles. These data sources will help transition plants from reactive to preventive maintenance through establishment of maintenance intervals, improvement of existing preventive maintenance intervals using better-quality failure cycle information, and/or transition from preventive to predictive maintenance. To extract this information, digital maintenance records are essential; therefore, a formal discussion of digital maintenance systems and related implementation standards is given. To support the need for maintenance data, a framework for utilization of failure cycles in predictive maintenance models is provided. Three different applications are examined, where maintenance dependent models are compared to traditional models to quantify the capacity for improvement in failure-time predictions.

This work shows that the utilization of process and maintenance data in prognostic modeling results in significantly improved failure predictions. This additional time to respond will help organizations avoid or plan for failures before they occur, which supports more effective maintenance and capital replacement policies.

TABLE OF CONTENTS

1. Introduction	1
1.1. Problem Statement	3
1.2. Original Contributions	4
1.3. Dissertation Structure	6
2. Literature Survey	7
2.1. Prognostics in Health Monitoring	8
2.2. Prognostic Models	8
2.2.1. Type I, II, and III Models	10
2.2.2. Empirical vs First Principles Prognostics	16
2.2.3. Prognostic Input/Output	17
2.3. Disparate Data Sources	18
2.3.1. Computerized Maintenance Management Systems	20
2.3.2. Condition Monitoring Systems and Data Historians	21
2.3.3. Coupling Algorithm for Plant Data Merging	22
2.4. Improved Maintenance Strategies	23
2.4.1. Preventive/Periodic Maintenance	24
2.4.2. Predictive Maintenance	25
2.4.3. E-Maintenance	25
2.4.4. Repairable Systems	26
2.5. Summary	28
3. Survey of Requisite Changes to Support Modern Maintenance	29
3.1. Problem Statement for Maintenance Data Standards	29
3.2. Utilizing Maintenance Data to Improve Asset Management	29
3.3. Future of Prognostics for Asset Management Support	30
3.4. Challenges and Standards	31
3.5. Summary	33
4. Research Methodology	34
4.1. Data Generation	34
4.1.1. Proof of Concept Data Generation Tactics	34
4.1.1.1. Proof of Concept Process Data Simulation	35
4.1.1.2. Proof of Concept Simulated Data Fidelity	37
4.1.1.3. Proof of Concept Maintenance Data Generation	37
4.1.1.4. Analysis of Simulated Data and CMMS Records	38

4.1.2.	Primary Case Study Data Generation Tactics.....	41
4.1.2.1.	Condenser Data Simulation	41
4.1.2.2.	Simulated Data Parameters.....	43
4.1.2.3.	Simulated Data Analysis	45
4.1.2.4.	Augmented Condenser Data Generation	47
4.1.2.5.	Augmented Data Validity.....	51
4.1.3.	Secondary Case Study Data Generation Tactics.....	55
4.2.	Integrating Data Sources.....	61
4.3.	Prognostic Model Development.....	64
4.4.	Metric Representation of Results	65
4.5.	Summary	66
5.	Applications and Results	67
5.1.	Proof of Concept Study.....	68
5.1.1.	Prognostic Model Development.....	68
5.1.2.	Prognostic Model Results	70
5.2.	Primary Validation Case Study	70
5.2.1.	Prognostic Model Development.....	72
5.2.1.1.	Simulated Condenser Data Modeling	72
5.2.1.2.	Augmented Condenser Data Modeling	73
5.2.2.	Primary Case Study Model Results	75
5.2.2.1.	Type I Model Analysis – Simulated Data	75
5.2.2.2.	Type I Model Analysis – Augmented Data.....	75
5.2.2.3.	Type III Model Analysis – Simulated Data	77
5.2.2.4.	Type III Model Analysis – Augmented Data.....	79
5.3.	Secondary Validation Case Study	79
5.3.1.	Prognostic Model Development.....	79
5.3.2.	Secondary Case Study Model Results	81
5.3.2.1.	Data Generation	81
5.3.2.2.	Model Results	81
6.	Conclusions and Recommendations for Future Work	84
6.1.	Conclusions	84
6.2.	Recommendations for Future Work	86
	List of References	87
	Vita.....	95

LIST OF TABLES

Table 1 - Example of CMMS entries from the simulated CMMS code (taken from .csv file).....	39
Table 2 - List of heat exchanger signals	56
Table 3 – Proof of Concept: Prediction errors for 50 simulated cycles using competing maintenance independent and dependent models	71
Table 4 - Proof of Concept: Prediction errors for 300 simulated cycles using competing maintenance independent and dependent models	71
Table 5 – Primary Case Study: Type I model parameter uncertainties in observations (# obs.) for competing models using simulated data.....	76
Table 6 - Primary Case Study: Type I prediction errors (MAPE) and uncertainty (u , standard deviation) in number of observations (# obs.) for competing models using simulated data.....	76
Table 7 – Primary Case Study: Augmented data Type I model prediction error (MAPE) and standardized uncertainty (u , standard deviation)	78
Table 8 - Primary Case Study: Type III prediction errors w/o Bayesian priors for simulated dataset	78
Table 9 - Primary Case Study: Type III model prediction errors with and without Bayesian prior information for simulated dataset	78
Table 10 - Primary Case Study: Type III model prediction errors for augmented dataset	80
Table 11 - Secondary Case Study: Type I prediction results for competing models (heat exchanger data)	82
Table 12 - Secondary Case Study: Type III prediction results for competing models (heat exchanger data)	83

LIST OF FIGURES

Figure 1: Digital elements in modern DPMS for NPPs [6].....	2
Figure 2: Predictive maintenance system flow [5].....	5
Figure 3: Detailed predictive maintenance system flow (adapted from [5])	5
Figure 4: Typical flow in a health monitoring system [12]	9
Figure 5: Example of a TTF distribution (blue) with a Type I model applied (green)	11
Figure 6: Influence of scale parameter (left) and shape parameter (right) on Weibull PDF	11
Figure 7: PHM example with covariates that have strong proportionality [26]	13
Figure 8: Example of results from GPM model applied to test data (red). The extrapolated path (blue) has predicted RUL shown in cyan. The actual path (black) has a larger RUL (purple)	15
Figure 9: Prognostic model transitions based on available input data [48] [49].....	19
Figure 10: Expected pathway for nuclear instrumentation and controls advancement (adapted from [4])	19
Figure 11: The three categories of equipment repair	27
Figure 12: Repair action influence on simulated data parameters	36
Figure 13: Visual depiction of the degradation removed (red) by maintenance action 0 (repair)	36
Figure 14: Simulated cycles (top) undergoing different repair actions (bottom)	39
Figure 15: Maintenance dependent initial cycle degradation levels for feature 1	40
Figure 16: Maintenance action dependent TTF samples from simulated dataset	40
Figure 17: Plate-style heat exchanger [76]	42
Figure 18: Early-life trend for heat exchanger degradation from fouling	42
Figure 19: Simulated maintenance dependent initial degradation levels	44
Figure 20: Simulated degradation parameter (top) and respective Maintenance Actions (MA) (bottom) across several maintenance cycles.....	46
Figure 21: Simulated maintenance action dependent degradation parameters	46
Figure 22: Maintenance dependent TTF for 100 simulated cycles	48
Figure 23: Simplified process for sampling augmented data cycles.....	48
Figure 24: QQ plot for partial repair initial degradation Weibull distribution (red) compared to training data (blue)	49
Figure 25: Distributions of initial degradation for the augmented data	49
Figure 26: Augmented failure cycles using sampled parameters from real data distributions	50
Figure 27: Comparative QQ plot for partial repair initial degradation	52
Figure 28: Comparative QQ plot for full repair initial degradation	54
Figure 29: QQ plot for full repair sub-sample dataset initial degradation.....	54

Figure 30: Small-scale accelerated degradation heat-exchanger system [36]	56
Figure 31: Example of calculated heat transfer coefficient under the influence of fouling degradation.....	58
Figure 32: Sampled heat transfer coefficient features	58
Figure 33: Kernel regression model for heat exchanger system [81]	60
Figure 34: Generated prognostic parameters with applied failure threshold	60
Figure 35: Simplified flow diagram for coupling algorithm design	62
Figure 36: Format of the hierarchical prognostic information database	62
Figure 37: Plot of 100 simulated cycles for a single signal	63
Figure 38: Auto-Associative Kernel Regression model flow diagram (adapted from [81]).....	69
Figure 39: Example of maintenance dependent prognostic parameters with critical failure threshold.....	71
Figure 40: Example of poor GPM extrapolated path.....	74
Figure 41: Comparison of real data (left) and augmented data (right) for heat exchanger system	82

1. Introduction

The landscape of commercial power plants is changing due to upgrades and transitions to digital systems. The first Nuclear Power Plant (NPP) recognized by the Nuclear Regulatory Commission (NRC) was the Shippingport Atomic Power Station, which first reached criticality in 1957 [3]. Around the same time, one of the first single-user computers, the LGP-30 personal computer, was marketed for purchase by General Precision for \$50,000 (approximately \$400,000 today) with a 120 kHz clock rate. This is about the same computing speed as a modern day coffee maker. Needless to say, early plant designs in the US did not rely on advanced microprocessors for data analysis, acquisition, monitoring, or plant control. The tech surge in the late 20th and early 21st centuries greatly improved humanity's ability to manufacture powerful processors for computation and analysis; our cell phones and laptops have more computational power today than supercomputers did in the fifties. This has increased availability and decreased costs of digital systems for nuclear power plants, allowing for advanced signal processing, monitoring, and diagnostics capabilities. With this change in the technology world, there has been a slow change seen in commercial power plants, especially nuclear power. The NRC sets requirements for digital systems, and can make it difficult for existing plants to obtain licensing updates if using digital systems. Many plants now have digital computers for operators, data monitoring and storage, and diagnostics, but there is still much room for advancement.

With the next generation of nuclear reactors being designed and existing plants looking to extend operating permits, it has become an important area of research to identify how to improve plant operations with computers and digital systems. The modern Data Processing and Monitoring System (DPMS) in Figure 1 shows most of the digital systems found in state-of-the-art NPPs. Westinghouse outlines this data processing and monitoring network as a framework for new and retrofit reactor designs. The entire system is commonly referred to as the Nuclear Plant Computer (NPC) and is a large focus area in current reactor design research. While not all reactors consist of elements found in Figure 1, it represents a well-organized digital ecosystem for maintaining plant operations. One important element not found in Figure 1 is the plant maintenance system. This is because most commercial NPPs do not use digital maintenance records, but rather still depend on manual hand-written logs for recordkeeping. The use of analog maintenance records commonly results in human errors, which can cause incorrect findings, repeat maintenance for corrective measures, delays in plant operation, and plant safety issues including the health of safety systems [4]. One method of reducing these error sources is to use a digital maintenance recordkeeping system to track, schedule, and standardize maintenance efforts across the plant. The last few decades have seen surges in software development for digital maintenance. Many commercial manufacturing and operating industries have embraced the Computerized Maintenance Management System (CMMS) as a principal tool for digital-based maintenance strategies, due to its low impact on asset lifecycle budgets and proven substantial benefits [5]. The CMMS is a key digital system for improving nuclear plant performance in future reactor generations.

On the topic of maintenance, traditional efforts in NPPs have been primarily reactive maintenance.

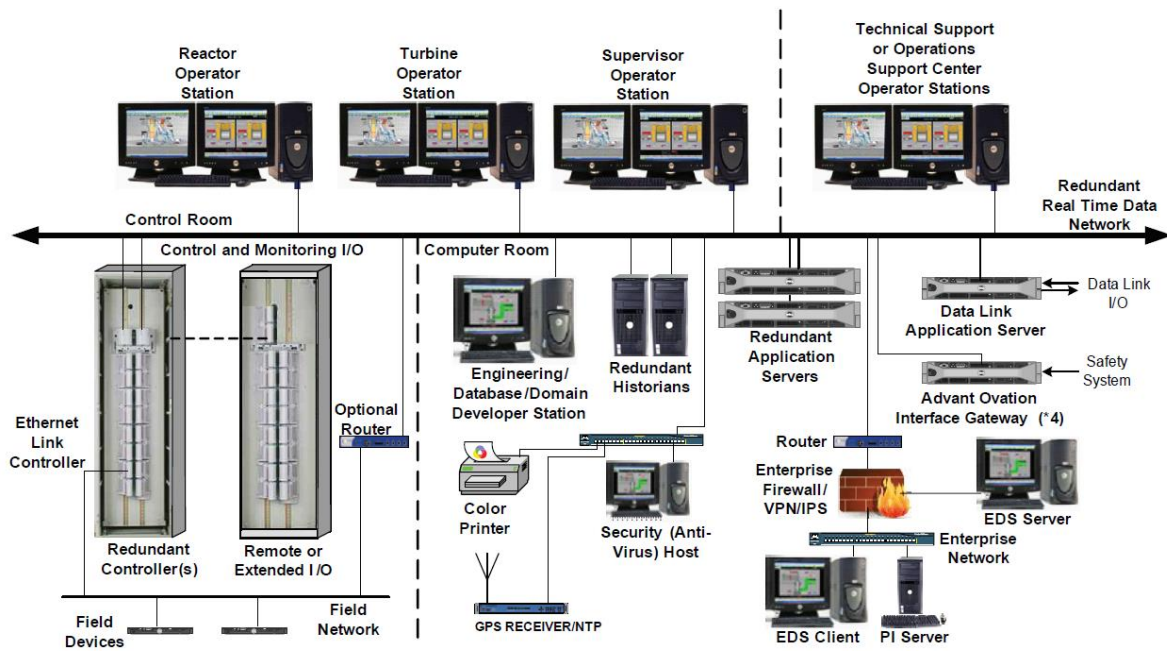


Figure 1: Digital elements in modern DPMS for NPPs [6]

This means that the system has a fault or failure, and maintenance workers *react* to the issue. The results of such practices include unexpected downtime, reduced safety, and low system health understanding among maintenance workers. Unlike reactive maintenance, preventive maintenance is an effort to replace parts that have a high likelihood of failing before the failure event actually occurs. These maintenance strategies are based on historical failure times, and occur at regularly scheduled intervals to prevent critical asset failures. Rather than reacting to failures, maintenance workers are proactively repairing or replacing parts and components with a high statistical probability of failure. Preventive maintenance can help lessen unexpected downtimes, improve plant safety, and increase system understanding for maintenance workers, but at the cost of generalizing systems to expected behaviors; unexpected failures can still occur when the system degrades faster or slower than the historical averages. Preventive maintenance only serves to reduce the erratic nature of reactive maintenance efforts, not tame it. In contrast, a relatively new strategy called predictive maintenance utilizes condition-based measurements to determine the true system degradation and predict when maintenance should be conducted. The predictive maintenance research field has many subdivisions, and is quickly becoming a revolutionary maintenance practice [5]. Utilizing predictive maintenance in nuclear plants would create the opportunity for improvements to safety, increased plant availability, and better operator understanding of system health. The end-goal of predictive maintenance is to support operator decision-making by predicting failure information, and providing effective maintenance strategy opportunities [7].

Amongst published research on industrial applications, there is little discussion of methods to feasibly conduct predictive maintenance in NPPs. The work described throughout this paper is an attempt to illustrate a path towards the future of predictive maintenance in the nuclear industry. The following dissertation presents a literature survey of current practices, and a proposed methodology for coupling disparate nuclear data sources in order to improve prognostic model capabilities and enable future predictive maintenance opportunities.

1.1. Problem Statement

Possibly the most important tool in predictive maintenance is prognostics, which is defined as a series of steps made to predict failure time or cycle degradation prior to failure [8]. In simpler terms, prognostics is a method of applied predictive modeling. The steps involved in the development of a prognostic model turn a relatively basic concept into a complicated network of optimization, decisions, categorization, and interpretation. Even rudimentary prognostic models require significant understanding of the system, and how data from the system is related to its degradation. Unlike raw plant data, prognostic data (that is, prognostic model inputs) must be divided into operating cycles in order to establish normal operating limits. Each operating cycle begins with system startup and ends at system “failure”. The prognostic data is categorized by the system it belongs to, using relationships between specific equipment and signals.

Data from the plant computer or data historian is rarely stored in a format that is directly usable in prognostic models. The amount of data produced by a nuclear plant is often overwhelming, which prevents effective utilization and analysis by data scientists [5]. In typical nuclear plants, uncompressed raw data can amount to as much as 100GB per year of text based numerical data. Traditionally, transforming raw data into a suitable format for prognostic model development would require hundreds of man-hours, sifting through datasets and extract operating cycles for individual systems. In a review of Condition-Based Maintenance (CBM) by Jardine et.

al, the author states that the development of tools for extracting and processing data for use in prognostic models is a requirement for next-generation predictive maintenance efforts [7].

Programs that utilize CBM consist of two critical parts: diagnostic and prognostics [7]. Both diagnostics and prognostics rely on Condition Monitoring (CM) in order to detect significant changes in asset performance, and CM relies on sensor data from monitored assets. A flow diagram for a typical predictive maintenance system is shown in Figure 2, which depicts the steps in predictive maintenance in systems where CM data is configured for model development. However, in most applications additional information from the CMMS is required to properly transform raw output from the Data Acquisition (DAQ) system into CM data for diagnostic and prognostic modeling. The CMMS records data on part/component failures and subsequent maintenance actions, which can provide necessary information for separating large data files into individual operating cycles (startup to failure). CMMS data also contains relationships between systems and their respective sensors, which is critical for data extraction and categorization. Another important piece of information stored within the CMMS is the maintenance action taken to complete each scheduled work order. These data are related to the specific repair actions taken, and may help inform operators about the degree of repair a system receives depending on the effectiveness of applied maintenance efforts.

To efficiently extract prognostic data from CM data for failure predictions, it is necessary to integrate maintenance data from the CMMS as well as process data from the CM/DAQ systems [8]. While either a CM system or CMMS alone may be used to improve current maintenance practices in nuclear applications, combining the systems (or data from each) will result in more effective maintenance strategies than either system could achieve independently. Progress related to methods of data source merging is slow due to relatively few studies and immature CM-CMMS relationship definitions in the scope of predictive maintenance [5] [8]. Developing a method for coupling data from the CM system and CMMS is a major focus of this research. This is a challenge that requires a definition of the relationship between each system, as well as methods for predicting failures using the available information from each system. Perhaps the largest challenge throughout this development, is determining how to effectively communicate prediction results to maintenance engineers in a way that improves maintenance practices [8]. A detailed predictive maintenance flow diagram for this research is shown in Figure 3. Data from both the CM system and CMMS is combined using a coupling algorithm, which extracts operating data and processes it into a model-ready format. The extracted and categorized cycles are passed to a predictive model for determining future system degradation from sensor data. As data are passed to the predictive model, Remaining Useful Life (RUL) estimates for a system can be made. The RUL represents the expected time until the system can no longer operate according to the defined limits. This quantitative RUL value must be translated into a qualitative risk mitigation strategy to improve maintenance planning. Once a maintenance decision has been made using the predicted failure time, it can be fed back into the CMMS system to automatically generate a new work order.

1.2. Original Contributions

The research discussed in this work contains original contributions to nuclear data mining and predictive maintenance fields, with focus on viable commercial nuclear predictive modeling. The following are selections of specific topics that contribute to advancement of the aforementioned fields:

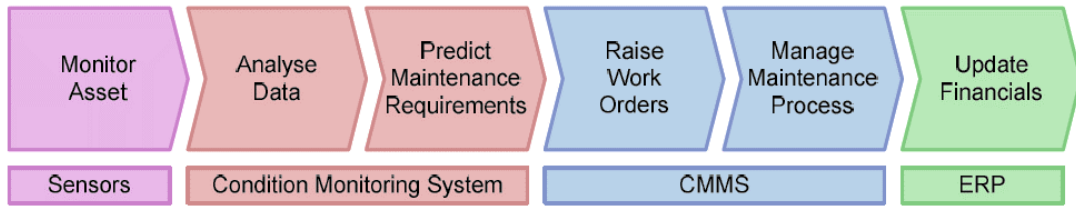


Figure 2: Predictive maintenance system flow [5]

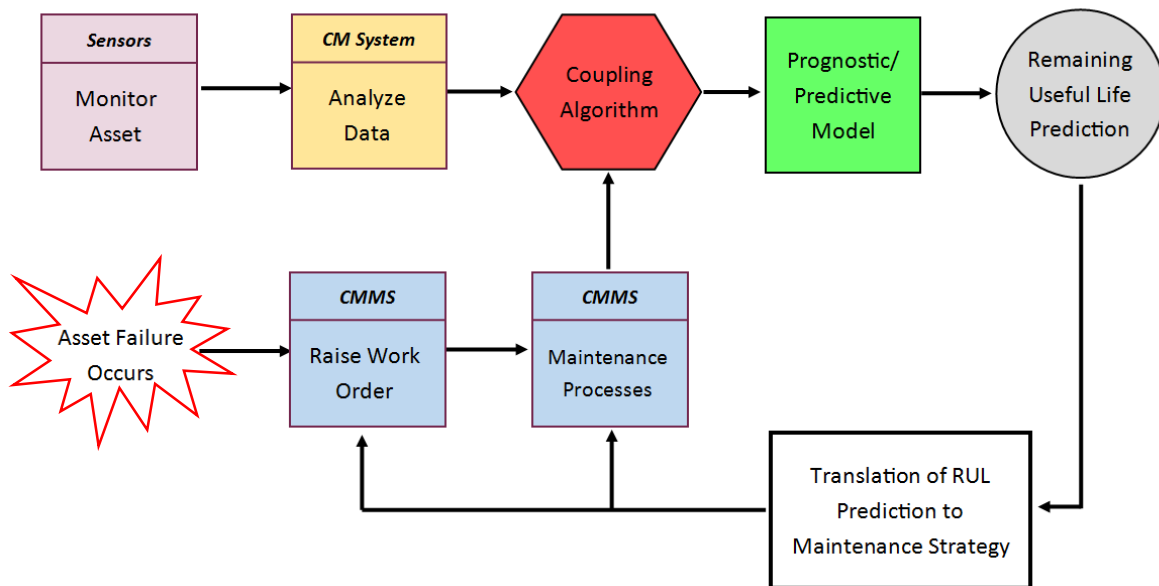


Figure 3: Detailed predictive maintenance system flow (adapted from [5])

1. Development of a formal overview for the state of industry related to advanced maintenance strategies, including: problems related to current industry maintenance practices, the need for digital maintenance systems in any plant that utilizes proactive maintenance, the future of industrial asset management, and challenges related to advanced maintenance strategy implementation. Recommended standards for organizations adopting new maintenance practices are provided to ensure the effective transition from analog to digital work orders in applications that may utilize predictive models, thereby preventing the need for major standards revisions later on.
2. Development of a preliminary algorithm for combining raw plant data with CMMS data with the intent to extract operating cycles for use in diagnostic/prognostic models. This improves the utility of recorded plant data for proactive maintenance.
3. Development of a framework for utilization of extracted operating cycles in predictive models, with focus on the development of generalized baseline prognostic models in nuclear plant applications for real-time predictive maintenance strategies. This is the key method developed in this dissertation in terms of contributory work.
4. Development of a proof of concept and two validation case studies to define the influence of maintenance actions on system degradation, proactive model capabilities, and predictive model accuracy. Demonstration of RUL prediction improvements using developed algorithms for each application.

These contributions will assist industry as predictive maintenance programs are adopted. The formal presentation of needs for digital maintenance standards, as well as cases presented that validate performance improvement when maintenance data is available, establish existing work that maximizes the effectiveness of predictive maintenance model development for industrial equipment.

1.3. Dissertation Structure

Section 2 of this dissertation reviews literature pertaining to the use of prognostics in commercial plants, existing prognostic model research, and data sources related to proactive maintenance. The literature survey also discusses modern maintenance strategies including predictive maintenance and e-maintenance concepts. Following sections discuss methods of achieving the contributions highlighted in Section 1.2 including coupling algorithm design and model development, as well as information pertaining to disparate data sources in nuclear plant settings. General discussion of methodologies for research efforts are followed by results of three research applications, including outcomes of data simulation, coupling algorithm development, and predictive maintenance dependent modeling. Lastly, Conclusions drawn from the research are conferred.

2. Literature Survey

Prognostics is a relatively young concentration within reliability engineering that focuses on predicting the amount of life a system/component has left until it will not meet a defined performance level. The predicted time remaining for the component to operate until it reaches the defined failure threshold is the RUL. There are many different methods of predicting RUL, but they all share a common goal of failure mitigation through optimization of maintenance practices.

Prognostic models come in many shapes and sizes, from simple parameterized distribution models, to complex multi-stage models. Their complexity depends on their application, and how resultant failure predictions will be used. The field of prognostics focuses on relationships between signal-level deviations from normal behavior, and system-level degradation. As components begin to degrade, changes in operating data may be seen in certain signals. The measure of signal deviation from normal behavior is commonly called a degradation parameter. These degradation parameters can be combined into a single parameter that represents the amount of degradation the entire system has experienced. This singular system health measure is referred to as the prognostic parameter, and can be used to quantify system degradation in real-time, which is the key to predicting the RUL in advanced prognostic models.

In general, prognostics has not been widely adopted in any industry including NPPs. The NRC governs plant licensing based on the Code of Federal Regulations 10 CFR 50, and is strict when it comes to plant modernization. The average nuclear plant in the US is approximately 35 years old. This means that outfitting existing plants with the digital technologies necessary to incorporate on-line monitoring and prognostic tools would require major license amendments, which are both time and cost expensive. As the NRC has not officially endorsed standards for prognostics (standards are still in development), it would be very difficult to get prognostic tools for plant control licensed within the US. This is similar to issues that have arisen in the past decade associated with updating plants to digital instrumentation and control. The NRC was only recently introduced to advanced diagnostics and prognostics as they would pertain to NPPs in a formal setting [9], making it even more difficult to get approval of advanced prognostic tools for plant management. However, prognostics is still useful as supporting information for plant operators, as well as applications other than plant control within the main Nuclear Steam Supply System (NSSS). One of the largest expenses for a NPP is the cost related to Operations and Maintenance (O&M). Keeping a plant running without unexpected shutdowns and failures is a constant effort. Prognostic models may help improve maintenance efforts by acting as a support system for O&M. These models provide unique information to the operators about system failures, and can be used to optimize maintenance scheduling and practices. Applying prognostic tools to NPP operation areas that are not safety critical will lay the framework for future adoption by NRC for advanced plant support.

The following literature survey discusses the basics of prognostics in health monitoring applications, and the types of prognostic models. Differences in the complexity of the model types are explained, and model applications are provided. For more complex prognostic models, the strengths and weaknesses between empirical and first principles models are discussed with focus on NPP applicability. To introduce the improvements to prognostic methods from this dissertation, existing approaches to data source merging are discussed; data sources in NPPs are outlined, and methods of improving prognostic modeling through extracted data are explained in order to provide clear reasoning for implementation. The developed prognostic techniques and

improvements directly enable maintenance optimization strategies; therefore, potential improvements to NPP maintenance compared to current practices are highlighted. These changes focus on the end-goal improvements to NPP O&M by transitioning to advanced strategies for equipment maintenance.

2.1. Prognostics in Health Monitoring

The prognostics field is a key area in health monitoring. System monitoring utilizes tools for identifying and measuring system states of operational data, which are analyzed by fault detection and diagnostic systems. The prognostics system relies on information and analysis from the other tools in order to effectively predict future system degradation levels and failure times. The general steps in health monitoring are shown in Figure 4. Data is most commonly collected from a Data Acquisition (DAQ) system, but can also come from disparate sources such as plant data historians, and maintenance logs. Raw data is passed through a monitoring system to determine if there are faults in the signal. If a fault is detected, the data is passed to a diagnostic system in order to quantify the fault within the data, and characterize the failure mode for the part/component. Data for specific systems and/or failure modes can be used to develop prognostic models in order to predict RUL for specific equipment. The RUL and related statistics provide information to O&M for optimizing performance and maintenance scheduling. These actions are referred to as failure/risk mitigation and are the major health monitoring objective.

In maintenance applications, health monitoring systems are considered CBM because they rely on specific conditions that the system/component undergoes during operation [10]. As the field of prognostics in health monitoring is relatively new, there is much discussion on the value of prognostics in industrial applications. Some expected benefits of prognostics implementation include: reduction in total number of unplanned maintenance events, overall improvements to plant safety, increased usable life for parts and components, reduced costs associated with O&M, and reduction in the amount of manual inspection required [10] [11]. These benefits can be generalized into two major goals for prognostics in CBM applications: accurate prediction of remaining useful life for components, and effective utilization of failure information for improved maintenance efforts and failure mitigation strategies. Referring back to Figure 4, prediction of RUL and mitigation of consequences are the final two stages in a health monitoring system.

RUL determination and subsequent failure mitigation require the development of accurate prognostic algorithms. These tools and models use varying amounts of failure data in order to characterize the system under investigation. The following section discusses the types of prognostic algorithms commonly used, including discussion of relevant model input/output. Considerable amount of current research related to prognostics involves the improvement and development of these algorithms.

2.2. Prognostic Models

Prognostic models are groups of algorithms used to predict the RUL, or related measure, for specific components/systems. Once information about equipment fault modes is collected and analyzed within a diagnostic system, failure data from operation can be passed to the prognostic system. There is a popular adage in computer science and related fields that states there is “no free lunch”, meaning that there is an opportunity cost associated with each decision in algorithm and model development. Choosing one model that is more accurate may be more time intensive

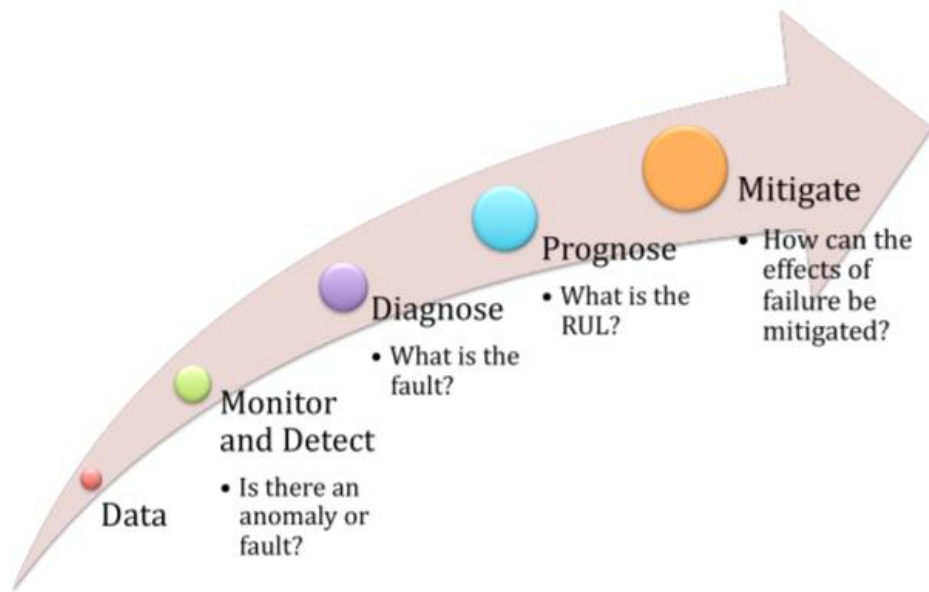


Figure 4: Typical flow in a health monitoring system [12]

than using a simplified model that runs very quickly; each model is useful in its own situations [13]. There are also different models used based on the desired output prediction and data available for modeling. While many of these predictive models calculate the RUL, prognostic models are also used to predict the Time-to-Failure (TTF) and Probability of Failure (POF) [14]. These may be adaptations of existing models or entirely unique algorithms. The following sections discuss the major types of prognostic models for RUL prediction, and the general input/output for related prognostic algorithms.

2.2.1. Type I, II, and III Models

Within the predictive maintenance research group at the University of Tennessee, there is an accepted prognostic model categorization based on the information that each type utilizes. These three model types are outlined in the following subsections.

2.2.1.1. Type I: Reliability based models

Type I models utilize historical failure times to develop characteristic distributions. As cycles occur, the TTF for each is recorded. The failure data is collected in histogram bins, and a parametric fit is applied to the failure time distribution. An example of a Type I distribution is given in Figure 5, which shows historical TTFs sorted into bins (blue). Bin widths are chosen based on the number and diversity of failure times, and dictates the resolution of the data. A generic parametric fit (green) is applied to the histogram. The resulting model characterizes how similar components should fail under similar conditions. This does not mean that condition-related data is used; on the contrary, Type I models assume that each are applied to a population of data statistically representing an average condition for the equipment. The Type I model is used to estimate future TTFs for similar components, and may be updated as additional failure information is collected.

The most common distribution type used for Type I modeling is the two-parameter Weibull fit [15], which is described by Equation 1.

$$PDF = \frac{\theta}{\eta} * \left(\frac{x}{\eta}\right)^{\theta-1} \exp\left[-\left(\frac{x}{\eta}\right)^{\theta}\right] \quad \text{Eq. 1}$$

In Equation 1, θ is the scale parameter. This parameter changes the relative distribution magnitude. The second parameter in the equation is shape parameter (η), which is used to manipulate the distribution slope. The result is the Probability Density Function (PDF) dependent on failure times (x). The effects of each parameter in the Weibull distribution are shown in Figure 6. The Weibull model is a popular choice for Type I prognostics due to its versatility in modeling components with non-decreasing failure rates through manipulation of these parameters. The applicability of two-parameter Weibull modeling for TTF data has been thoroughly tested and verified [16]. Weibull models have been used in numerous applications such as clinical trials for survival rate analysis [17], maintenance planning for power transformers [18], and commercial lightbulb lifetime predictions [19]. For more information on Weibull analysis, refer to [15] [20].

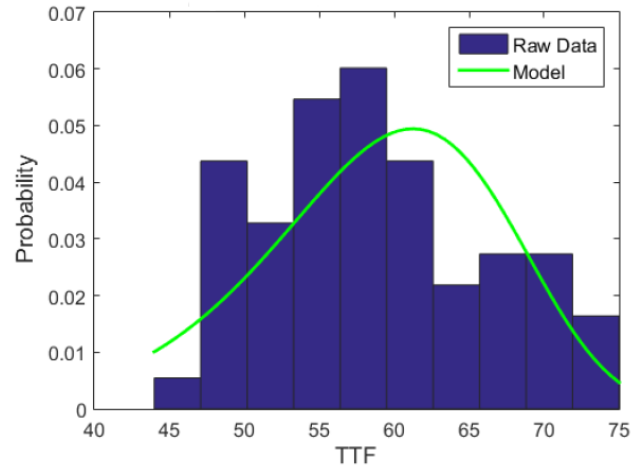
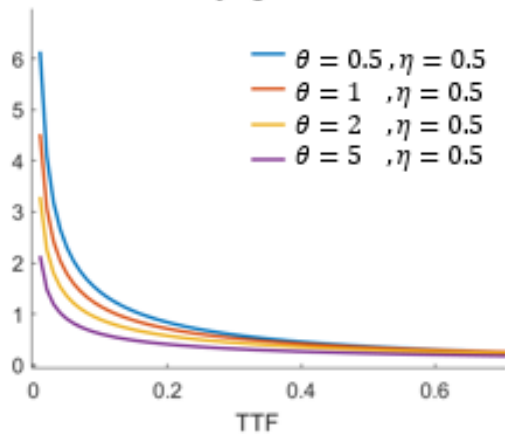


Figure 5: Example of a TTF distribution (blue) with a Type I model applied (green)

Weibull PDF for Varying Scale Parameter Values



Weibull PDF for Varying Shape Parameter Values

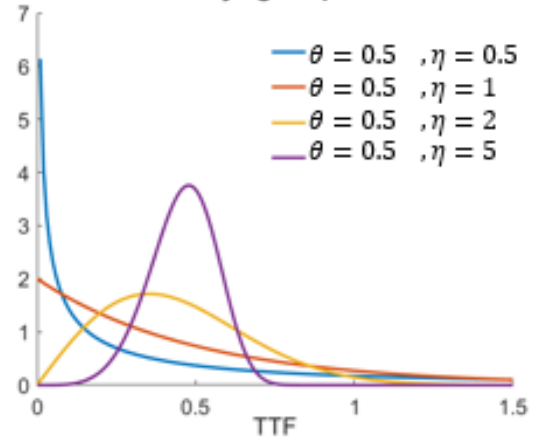


Figure 6: Influence of scale parameter (left) and shape parameter (right) on Weibull PDF

2.2.1.2. Type II: Stressor based models

Unlike Type I models, which characterize similar components under average conditions, Type II prognostic models attempt to utilize operating condition information within failure cycles to characterize systems under specific conditions. Some common stressors (used to define operating conditions) include temperature, load, pressure, and vibration. An example of comparing the difference between Type I and Type II models is tire life predictions. In a Type I analysis, failure times for a single tire brand in all driving conditions (mountainous, city, snow, etc.) are used to develop a model. This example model would determine the average failure times for tires over the scope of the training data. A similar Type II analysis may look at the failure times for a single tire brand, and unlike the Type I model, divide the sample population into 3 age groups for drivers: 16-25, 26-50, and 51+ years old. One would expect the amount of wear would be significantly different when comparing a first time driver to a senior citizen. The TTF distribution would likely be at lower values for the 16-25 group compared to the 51+. In this example, the stressor condition would be driver age. The development of unique models (or equivalent degradation paths) for each operating condition makes this a Type II model example.

There are many different forms of Type II models that may be used for prognostics. One common Type II model for CBM applications is the Proportional Hazards Model (PHM) [21]. The PHM is most commonly used for systems with multiple identical components in order to optimize system reliability [22], and is proven to be an effective model in various survival applications [23] [24] and residual life prediction cases [25]. The PHM is based on the assumption that similar components may have different degradation rates based on the conditions in which the components are operated. The effects of the operating conditions are quantified into ‘covariates’ that are used as multiplicative factors for the component’s baseline hazard rate. A simplified form of the PHM is given in Equation 2.

$$\lambda(t, Z_1, Z_2, \dots, Z_n) = \lambda_0 \exp(\beta_1 Z_1, \beta_2 Z_2, \dots, \beta_n Z_n) \quad \text{Eq. 2}$$

where the hazard rate (λ) at time t is characterized by a baseline hazard rate for the component (λ_0), model parameters (β_n), and the covariates (Z_n) for each operating condition/stressor. The PHM is most effective when individual stressor conditions are well correlated with component degradation, and when the ratio of any two covariates in the PHM is a constant value over time. An example of data with strong covariate proportionality is provided in Figure 7. The figure shows the survival rate for male patients diagnosed at age 50 with different stages (severities) of larynx cancer. The different stages are stressors corresponding to the development of cancer in the body, with later stages meaning a higher cancer/healthy tissue ratio. The proportionality between each stage of cancer is very strong, as indicated by the similarity in the paths. As a result, if the condition is known (in this case the stage of cancer), survival predictions can be improved by utilizing the appropriate degradation rate.

2.2.1.3. Type III: Effects-based models

The last classification of prognostic model is Type III “effects-based” prognostics [14], which utilizes sensor data recorded during component operation to quantitatively approximate system degradation. These models are often multi-stage and consist of several algorithms. The first major step involved in Type III model development is determining degradation parameters that are well correlated with system health.

Larynx Cancer Survival by Stage

Males at Age 50

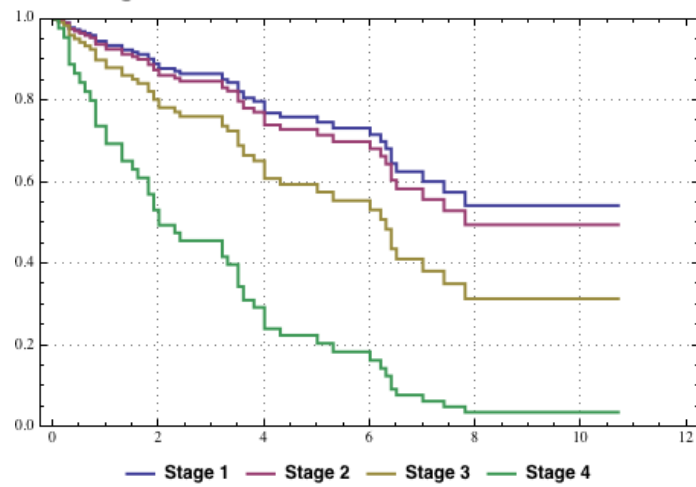


Figure 7: PHM example with covariates that have strong proportionality [26]

These parameters may be as simplistic as one signal, or as complicated as a function of multiple features that are each calculated from several independent signals. An example of a feature for heat exchanger degradation is the Logarithmic Mean Temperature Difference (LMTD), which is calculated using multiple temperature sensors and used in thermal flow applications [27]. This feature is highly correlated to system degradation in heat exchangers that undergo fouling [28].

A common method of utilizing effects-based prognostics is the development of a model describing system degradation as a function of time. As the degradation is extrapolated for future predictions, the amount of cumulative damage in the system is compared to a critical failure threshold. The threshold is predefined to represent the amount of degradation at which point the system will no longer operate within desired limits. After crossing the threshold, the system is considered failed and must undergo maintenance or replacement. The difference between current time and Time Of Failure (TOF) is the predicted RUL for the system. There are several common cumulative damage models that are used in Type III prognostics. For applications relevant to this research the General Path Model will be discussed. For information on other model types, refer to [29].

The General Path Model (GPM) was originally proposed by Lu and Meeker [30] and first used for prognostic estimates by Upadhyaya et. al [31]. GPMs utilize estimates of degradation or prognostic parameters to characterize the general path for a specific component/system. The major assumption of GPMs is that the functional form of a specific fault mode's degradation path will be the same across similar equipment. The functional form of a degradation path is Equation 3 [32].

$$y_i = \eta(t, \phi, \theta_i) + \epsilon_i \quad \text{Eq. 3}$$

where t is a specified time, ϕ is a vector of parameters for all unique components, θ_i is a vector of random-effect parameters for component i , η is the functional path based on the parameters, and ϵ_i is the measurement error within the system [30]. One issue associated with original GPM applications was the ability to identify an optimal prognostic parameter to use for extrapolation. Updates to GPM utilization in prognostic applications are provided in [10]; this work discusses methods of automated prognostic parameter identification. The functional form in Equation 3 allows characteristic models to be developed (for specific fault modes) that represent the probable degradation paths for a system. Once the model has been trained, new data is passed to the GPM, which extrapolates a path (y_i) from the current observation until it passes the failure threshold. The GPM output is a TOF prediction that can be translated to a RUL prediction. An example of GPM results for a failure cycle is shown in Figure 8. Training data points are passed to the GPM (pre-trained), which determines the degradation path for that specific component. The degradation path is extrapolated until it reaches the defined failure threshold (green). The predicted RUL is calculated as the current observation (last value in the test data) subtracted from the predicted TOF (cyan). As additional data is observed, the GPM can recalculate a new RUL prediction, making the GPM an ideal prognostic model for real-time applications. It should be mentioned that the GPM is not 100% accurate. Predictions are accompanied by uncertainty statistics in order to measure the effects of prediction error on failure mitigation decisions. Readers interested in GPM uncertainty analysis and error reduction are referred to the following sources [10] [31] [33]. One industry application of the GPM is the use of Monte Carlo sampled extrapolations to determine a predicted failure distribution for calculating reliabilities in advanced

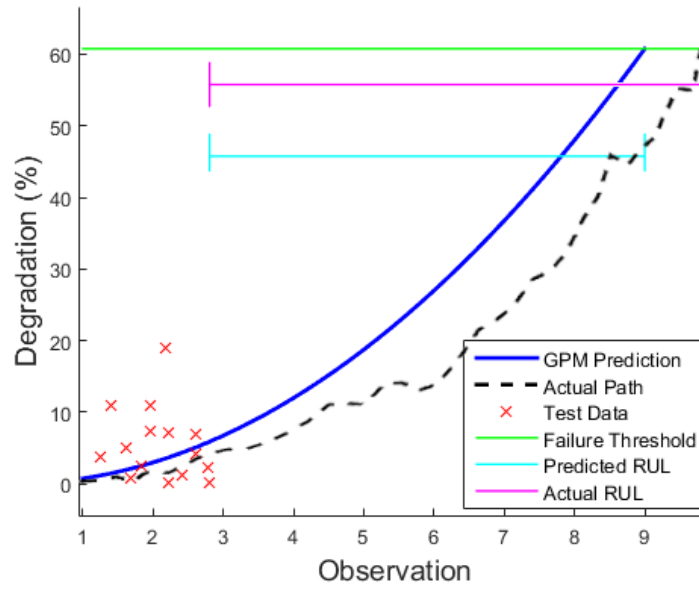


Figure 8: Example of results from GPM model applied to test data (red). The extrapolated path (blue) has predicted RUL shown in cyan. The actual path (black) has a larger RUL (purple)

circuit designs [34]. Another application related to circuits is the use of the GPM to identify common deterioration features from electronic module data in order to extrapolate potential future degradation paths [35]. GPM has recently been used in a nuclear application to predict heat exchanger TOFs due to loss of thermal efficiency through fouling degradation [36].

2.2.2. Empirical vs First Principles Prognostics

Many prognostics models and tools have been developed in recent years. However, each application requires individual modification of applied prognostic techniques to optimize prediction capabilities [8]. There are two major categories of prognostic model: first principles and empirical. Both have unique advantages and disadvantages depending on the application (a.k.a. “no free lunch”). The following paragraphs explain the basic underlying principles of each, compare their advantages and disadvantages, and highlight a few common models and applications.

In many industry applications, exact operating mechanisms related to failure are not known. This prevents the development of physics-based models to directly quantify degradation at a specific time. In these situations, empirical models can offer degradation estimates using data provided from the diagnostic and DAQ systems. These models consist of tools including statistics and pattern recognition to predict system health from available sensor data. Examples of empirical algorithms include Neural Networks (NNs), regression models (e.g. linear regression), Markov chains, and PHM. Empirical methods are uniquely suitable in applications that need dimensionality reduction such as Principal Component Analysis (PCA); empirical models are capable of making accurate predictions by effectively reducing high-dimensionality data sets with relatively few observations [8]. Example applications of empirical prognostic models include prediction of commercial vehicle degradation using an autoregressive Support Vector Machine (SVM) [37], PCA and Principal Component Regression (PCR) to predict remaining ozone concentrations [38], and the use of Hidden Markov Models (HMMs) for drilling equipment degradation [39]. These applications are just a handful of existing empirical models and applications. For further information on common empirical models and related prognostic uses, please refer to [37] [40] [41] [42].

While applications have been able to successfully utilize empirical models, their primary disadvantage is the correlation between model performance and data quantity/quality [43]; better data typically means better model performance. Data quality is adversely affected because of process noise, measurement noise, changing procedures, and many other sources of variance. As empirical models are directly influenced by the uncertainty in their training data, developed models often require larger quantities of data in applications where measurement fidelity is an issue [40] in order to overcome these issues. Description of model architecture and definition of optimum model parameters are two additional challenges associated with empirical methods. An example is the definition of a Neural Network model in terms of the number of hidden layers and nodes within each layer. Each neuron has individual weights and biases that are calculated by optimizing the model using training data. This optimization is driven by processes that utilize random sampling; therefore, each training may result in a different set of optimized parameters. The same can be said for the network architecture, where the number of layers and neurons is optimized for a specific application. It can be difficult to support these decisions and optimization strategies in a physical sense (i.e. related to actual system health rather than model performance). Explanation of empirical modeling approaches may be difficult in applications where correlation between system measurements and component degradation are not well understood.

Unlike empirical approaches, first principle prognostics assumes that the physics of a system's degradation is understood, and can be accurately represented by mathematical equations [43]. First principles techniques use relational metrics to quantify system deviation from normal behavior under the influence of faulty components. These relational metrics are mathematical representations of comparison (distance, deviation, etc.) between the model predictions and system measurements, and are commonly referred to as model residuals. A good example of common metrics is provided by [44]. RUL prediction accuracy is dependent on correlation between the calculated model residuals and real system degradation levels in order to satisfy first principles assumptions. Good correlation exists when model residuals increase in magnitude under fault conditions. In this situation, an advantage of first principles prognostic methods is the ability to easily update the model as system degradation understanding improves [43]. New correction factors may be incorporated to improve accuracy of the models based on new degradation factor evidence. Another advantage is the use of first principles methods in on-line monitoring applications; knowledge of the system degradation in real-time allows for improved system control and risk assessment. The key disadvantage of first principles methods is the challenge discovering an accurate mathematical model for a system. An excellent example of the process for developing a physics-based model for pneumatic valve actuation is given in [45]. Good reference applications of industrial first principles prognostic methods include predicting automotive performance degradation [37], heat exchanger thermal efficiency degradation due to particle fouling [28], and RUL predictions for Lithium-ion batteries under thermal stress [46]. Hybrid models containing both empirical and first principles elements may also be used for failure predictions. However, research on hybrid models for the explicit purpose of degradation prediction in prognostics applications is immature [8] and outside the scope of this dissertation. A good review of hybrid models for prognostic predictions is given by [47].

2.2.3. Prognostic Input/Output

In algorithm development, one of the most important things to establish prior to any modeling is the algorithm Input/Output (I/O). In prognostic model applications, there are a few informal standards for I/O. Input to a prognostic model is usually one or more of the following data sources: diagnostic model results, CM data and/or raw data, and miscellaneous sources such as maintenance records or auxiliary data historians. These inputs influence aspects of prognostic model development including model selection, optimization, application, and prediction results. This means that data available as input to the model directly influences the potential model outputs. The most common prognostic model outputs include TTF/TOF, RUL, risk related probabilities, and degradation level metrics (e.g. residuals).

One result of this relationship between availability of input data and possible model outputs is a limitation on prognostic model applicability (Type I, II, and III) depending on the quantity and quality of information available. For example, in a system where the sole data available for prognostic modeling is diagnostic data (i.e. status of equipment such as online/offline), the only information available to develop a prognostic model may be historical component failure times. In this case it is only possible to effectively apply a Type I prognostic model, because there is no information known about operating conditions/stressors, or available measurement data to analyze. In systems with accessible diagnostic results and maintenance records, information about the operating conditions (e.g. time of year or location within the plant) may be extracted and used to develop a Type II model with stressor-specific distributions for historical TTF data. If

CM data is also available and correlated with system degradation, a Type III model may be utilized to track real-time residuals and predict failure information for specific components. The different prognostic models are capable of transitioning to more complex or system specific types if additional data sources are made available. The entire flow path shown in Figure 9 is referred to as a lifecycle prognostic model. Research into lifecycle prognostics is relatively undeveloped, but is motivated by the concept of data-dependent model transitions. The process starts with a new piece of equipment (e.g. replacement part) before it is installed and introduced to degradation. At this point, the only information available to estimate the RUL for the component is the historical failure times; therefore, a Type I model is applied. After the component has been installed and data related to the equipment's operating condition is assessed, the Type I model is transitioned to a Type II model. As data is recorded for signals related to the component, fault detection models are applied and cycle residuals are calculated. When onset of degradation begins, residual values can be tracked and passed to a prognostic model for failure predictions (e.g. TTF, RUL, etc.). For more detail on lifecycle prognostics please refer to [48].

Understanding the influence of data input on prognostic model performance is a key motivation for this research. Looking at different combinations of input data for prognostic models is an extension of the methods used in lifecycle prognostics. By investigating alternate ways to incorporate disparate data sources into model development, information about system degradation may become more readily available, thereby improving predictive model efforts. The following sections discuss methods of merging disparate data sources for prognostic applications, and improving failure mitigation strategies based on updated model results.

2.3. Disparate Data Sources

One of the unique challenges associated with prognostic input data is extracting information from the different data sources into a usable format. Raw data must be collected from independent sources, and pre-processed into different datasets depending on the desired prognostic model. Merging data from disparate sources is often difficult due to weakly defined relationships between the raw data and prognostic data inputs [8]; therefore, system specific relationships must be developed prior to data merging and extraction. The first step in extracting usable prognostic data is identifying the specific sources of information available. CMMS and CM systems are perhaps two of the most common types of O&M information repositories [8]. While most modern nuclear plants do not generally rely on digital work order systems such as a CMMS, it is expected that transition from analog to digital data systems in future reactor overhauls and design generations will result in the plausible assumption that plants will utilize CMMS-style programs for maintenance data storage in the near future. A technical program plan outline by Idaho National Laboratory prepared for the U.S. Department of Energy shows the expected timeline for nuclear industry advancement of instrumentation and controls in Figure 10. The outline in Figure 10 shows that supporting technologies such as CMMS and other computer based systems will continue to be adopted as the year 2020 approaches. Once a significant number of early adopters have transitioned to the use of CMMS (and other necessary enabling technologies), implementation of automated work activities and advanced outage management will be possible.

Many plants are also making improvements to their maintenance programs as they transition towards CBM. In traditional maintenance settings, Planned Maintenance (PM) intervals may be based on historical failure times, manufacturer's warranty data, and/or expert knowledge.

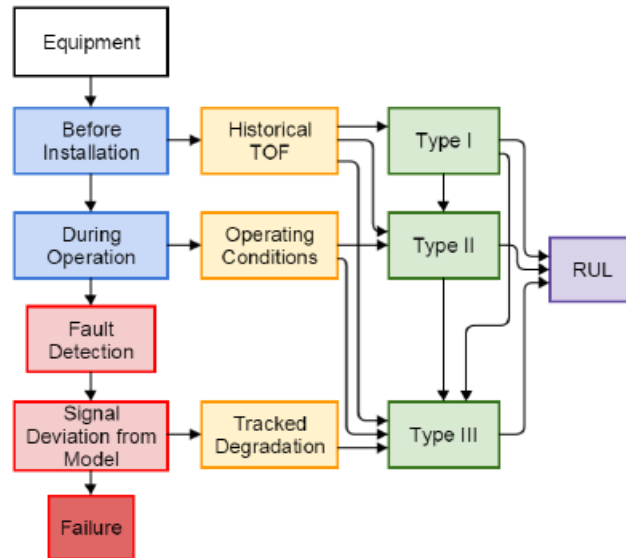


Figure 9: Prognostic model transitions based on available input data [48] [49]

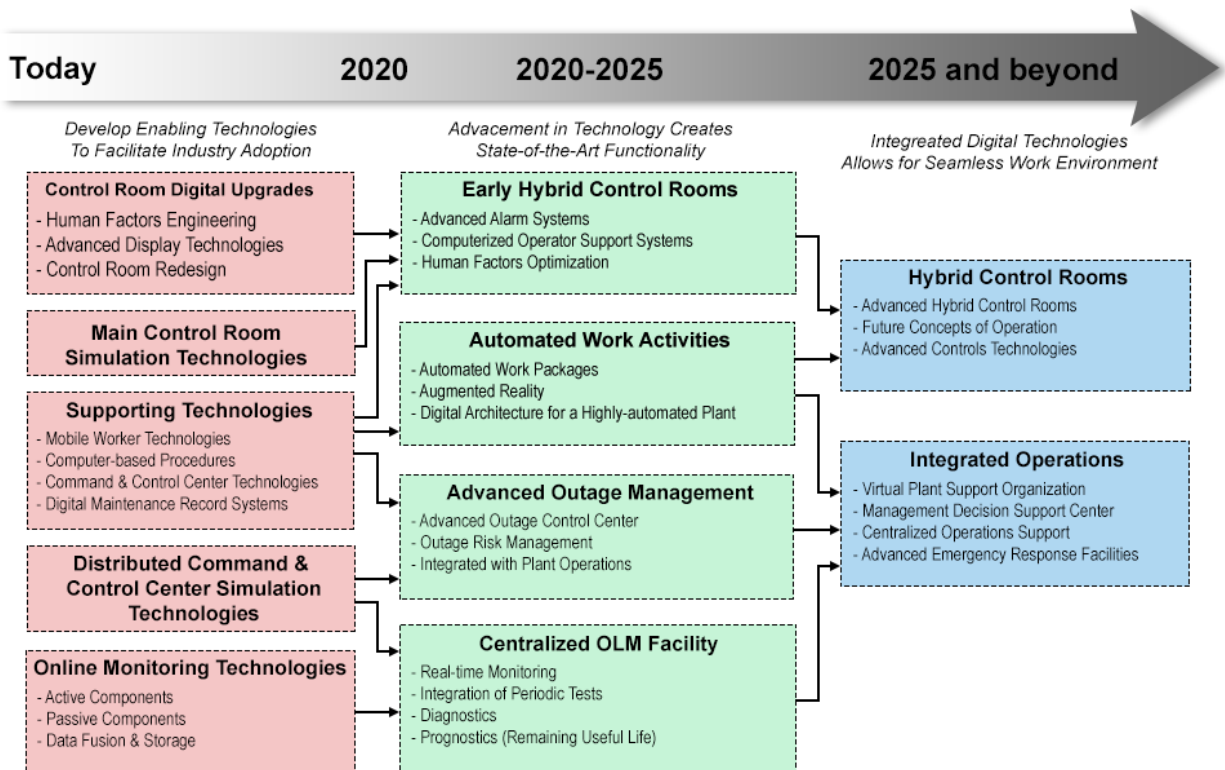


Figure 10: Expected pathway for nuclear instrumentation and controls advancement (adapted from [4])

This data can be taken specifically for the purpose of defining PM schedules, or may be recorded in existing data records and mined out for the purposes of PM. In contrast, CBM requires condition related data, and is often supported with multiple modules that collect, process, and quantify system condition. These modules may already be available in plant settings, or may need to be implemented to ensure a successful CBM program is possible. The primary support module for CBM is the Condition Monitoring (CM) system, which is responsible for monitoring all equipment of interest, and providing preliminary diagnostic capabilities. The CM system tracks equipment signal data and identifies changes in behavior. The system then analyzes the changes, and quantifies equipment deviation from healthy behavior into a degradation estimate. The degradation estimate, or degradation parameter, can later be passed to another module within the CM system; the prognostics module uses degradation parameters to predict future failures before occurring as a result of equipment degradation [50]. However, advanced prognostic and diagnostic support should not be utilized with information within the CM system alone. Often, data related to past equipment failures, operating conditions, and maintenance reports must be additionally collected to assist in developing explanatory failure cycles for CBM support. In state-of-the-art applications, this information is stored within a digital maintenance repository such as a CMMS, which is responsible for maintenance event logging, related data storage, and data availability. Together these data sources, which are unfortunately disparate in nature, can support advanced maintenance practices. Descriptions of the two major plant data repositories are provided below, followed by discussion of a potential method for merging the disparate sources for prognostic data extraction and advanced predictive modeling.

2.3.1. Computerized Maintenance Management Systems

Due to the average age of NPPs in the US, existing work order systems for plant maintenance are commonly paper-based and/or manual procedures. High volume paper records are difficult to work with; paper packets are expensive to print and even dispose of in nuclear applications. If a work order package is carried into primary containment, it is considered contaminated waste and must be disposed of properly [4]. The primary issue with paper and manual work orders is the influence of human performance on information quality. These methods require maintenance workers to correctly obtain plant information, complete necessary maintenance, record plant and maintenance information in the paper packet or computer, and validate the recorded results. The process is often complex and can result in incomplete or missing entries, and poor event descriptions [51]. Consequences of improper work order processing may include additional corrective maintenance work, operating delays, unavailable systems, and other safety related issues [4] [52].

With recent advancements in digital systems, asset management in industrial plants is becoming increasingly important as a means of improving safety and decreasing expenses [5]. This requires plants to evaluate alternative methods of O&M management and recordkeeping. In applications that are transitioning away from reactive maintenance efforts, CMMS are commonly used to accurately collect information for failure model development [7] [8]. A comprehensive assessment of existing CMMS programs is given by [53]. A CMMS, or alternative digital maintenance system, is an effective tool for collecting, organizing, and exporting maintenance related information such as work order logs, as well as overall asset management support [5]. Studies have shown that the use of digital maintenance management tools and maintenance strategy support programs have relatively instantaneous return on investment for large

organizations in operational performance and expenditure capacities [54]. A properly implemented digital maintenance system allows for growing organization complexity by providing improved overall available decision support, independent of the quantity of systems under consideration [55]. There are many providers of CMMS software, but in general, the CMMS is responsible for recording equipment information including part, model, and serial numbers as well as inspection data such as vibration spectra for rotating machinery [8] [51] [56]. Supplemental material on general CMMS information can be found in reference [53]. Successful implementation of a CMMS may improve plant performance, increase worker productivity, and enhance plant safety through reductions in unplanned active and latent component failures [4] [52]. More information on justification of CMMS use in industrial plants can be found in reference [53].

While the exact benefits of CMMS implementation are application specific, strengths include automated work orders, inventory management, and historic maintenance database quality [8]. Availability of a digital maintenance system creates the opportunity for advanced maintenance scheduling such as preventive or predictive maintenance, enhanced data analytics, and detailed event reporting [8] [55] [57]. For organizations where maintenance is primarily reactive, historical TTF data within the site's CMMS can be used to establish PM schedules. Organizations that already utilize PM can analyze CMMS data, including premature failure events, in order to modify and improve established intervals. In either application, CMMS data can assist in identifying problematic equipment and facilitate proactive maintenance through predictive failure modeling [50] [58] [59]. Integrated failure analysis for equipment is an ongoing focus of CMMS systems [56]. An example of an automated work order system is highlighted in references [4] [52]. This dissertation focuses specifically on integration of maintenance information with operating data to improve predictive maintenance efforts in NPPs.

2.3.2. Condition Monitoring Systems and Data Historians

While CMMS systems can store maintenance related information, they are not effective tools for monitoring operating conditions of plant components. Another common data-system is the CM system, which is explicitly designed for equipment monitoring and stressor condition assessment, and is used to quantify equipment health and provide initial-level diagnostics. Analysis of equipment health is facilitated using information from the DAQ system and supplemental diagnostic models (if available). Strengths of expert CM systems include multiple independent methods of component monitoring, degradation analysis tools (including tracking), and advanced diagnostics [8]. The success of programs that utilize CM technology is well-documented [58]. CM systems are useful for identifying changes in operating performance for fault analysis, but in many situations CM produces copious amounts of data, which cannot be effectively utilized. This data, if properly analyzed, would be invaluable for asset health predictions, such as prognostic model development [5]. In order to extract critical information from CM records, maintenance data specific to the CMMS must be integrated [60]. By utilizing machinery condition from the CM system and maintenance data from the CMMS, specific failure cycles can be extracted and categorized by fault mode, maintenance action, equipment type, and operating condition. In order to reduce both the frequency of unexpected failures and excessive maintenance, information from both data sources must be collected [8]. Together, this combination of data can help identify critical maintenance needs and reduce unnecessary O&M activities [60].

There are a number of challenges associated with CM system employment including proper data utilization, availability of expert knowledge, and ability to convert quantitative results to

qualitative failure mitigation strategies [5]. A hypothetical example of these challenges is an application with numerous sensors, where many poorly correlate with system degradation. In this case, there is an excess of data available for processing that has been recorded by the CM system. If an individual attempted to sift through all the recorded data and identify useful sensors for quantifying system health, faults, and degradation, the task would take an exceptionally long time, or may not be possible at all due to the extent of data or time constraints on the application. This is due to the relationship between data availability and data-mining/pre-processing complexity. As the size of the data stream increases, so do the challenges to identify key information (i.e. needle in a haystack metaphor). When essential data related to system health is obfuscated in this manner, the CM system requires additional information that can isolate valuable data, or quickly recognize unusable sensor data. This again points back to the need for data from within the CMMS or comparable digital maintenance database.

Data repositories for CM systems vary for different applications. In some cases, CM systems utilize independent data historians to record high-resolution data for plant equipment. These historians may be comprised of several months of raw plant data and CM data. In other applications, the CM system may only store condition data while relying on the plant historian to record raw signal data. This poses a unique challenge for prognostic efforts. Plant historians such as the OSIsoft PI system [61] often delete and/or compress data after 3-5 months [62]. Once data is removed, prognostic model development can become difficult. Data may also be divided across multiple historians, each with a unique purpose. An example of this is the Point Lepreau NPP in Canada. Here, local data for health monitoring is stored using a 3rd party historian similar to the PI system, and a separate historian is used for legacy plant data [62]. These differences in the data collection and storage practices of different organizations is the primary reason that challenges related to data integration, extraction, and processing exist. There is no explicit regulation or standard for the storage of nuclear plant data; in the US, data storage is seldom mentioned in regulatory documentation. One mention can be found in the International Atomic Energy Agency (IAEA) safety guide on Instrumentation and Control (I&C) design for NPPs [63], which discusses the requirement for historical data availability in human-machine interfaces, but does not mention any requirements on the capacity in which the data is stored. Specifics on data storage resolution and format are difficult to obtain for unique plant designs. Advanced methods of data analysis and extraction from multiple disparate data sources must be implemented to overcome these challenges associated with plant data collection.

2.3.3. Coupling Algorithm for Plant Data Merging

As previously mentioned, in order to facilitate modern maintenance practices, data is needed that currently resides across independent or isolated systems. In many cases, the data from these disparate sources would provide higher information quality as an integrated source than either system could provide individually [8] [64]. The challenge of integrating these sources is not simple to solve, as the common data repositories necessary for advanced maintenance are designed to remain isolated. CMMS programs are typically third-party software packages with little or no predesigned capabilities to interface with other plant systems for purposes of data import/export (external to the inherent CMMS environment); however, these software packages do typically include the ability to export data records. Similarly, while a CM system may consist of multiple interconnected sub-systems, CM systems are not typically designed to communicate with other programs or data repositories. This prevents effective work order scheduling and/or automated

maintenance optimization through the use of a combined singular CM/CMMS arrangement. Instead, collecting information from these data sources requires external data merging to facilitate identification of data relationships (condition versus maintenance) and maintenance decision support. These relationships are relatively unstudied and underdeveloped [5] [8].

Recent advances in O&M technologies have made automated methods of plant data merging possible. Past methods of data merging and extraction for similar applications would require advanced natural language processing and individualized pattern recognition algorithms in order to extract comparable information from handwritten maintenance data and traditional plant historians. The field of prognostics requires very specific and detailed information about component operation, degradation, and failure. Extraction of relevant prognostic information may considerably improve plant-operating performance and reduce costs and/or downtime associated with maintenance. In industry, the common practice for failure prediction (TTF, RUL, etc.) is to extract CM data related to diagnostics. However, additional information must be defined by the end-user before predictions can be calculated. While pure CM based methods of prognostic data extraction may result in effective prognostic models, utilizing information from both the CM system and CMMS can result in further improvements to prognostic data extraction performance [8]. While relational connections between CM and CMMS are relatively undefined [5], advanced methods of diagnostics and prognostics will depend on the development and progress of data extraction tools for plant data processing [7]. An example of reducing maintenance costs based on such a potential data integration method is given in reference [5]. This research investigates the translation of improved data extraction performance to model prediction accuracy through the use of a data-coupling algorithm. The primary goal of this algorithm is to increase prognostic data availability in nuclear plant applications in order to improve safety, maintenance planning, and plant uptime.

The goal of integrating condition data and maintenance data is to implement optimum CBM strategies. Studies of properly assimilated CBM systems have been shown to have a capacity for improvement of as much as 25% reduction in purchasing expenses, 75% reduction in emergency maintenance, 95% improvement in inventory management accuracy, and up to 200% in proactive maintenance capabilities [65] [66]. While this type of implementation requires a strong understanding of relationships between different data sources, developmental research related to identifying these relationships is being conducted. Mathew et al. provides an in-depth study of similar data integration, and suggests a possible framework for executing a successful CBM program [5].

2.4. Improved Maintenance Strategies

Merging data sources is a crucial step in improving failure prediction capabilities. By cultivating knowledge of system degradation and failure, innovative maintenance strategies can be implemented to optimize replacement scheduling and reduce unplanned downtime. The classical approach to maintenance is to fix items when they break down [56], which is referred to as reactive or corrective maintenance because it occurs *after* failure. Modern maintenance strategies are divided into two groups: reactive and proactive. Unlike reactive maintenance, the goal of applying proactive maintenance *before* failure is to reduce the need for unplanned maintenance. Utilizing proactive methods for plant maintenance results in reduced downtime and improvements to equipment reliability, safety, and quality through prevented equipment failures

[56]. The two methods of proactive maintenance are preventive and predictive maintenance, which are outlined in the following subsections.

2.4.1. Preventive/Periodic Maintenance

In nuclear plants, reactive maintenance is often not a viable option. For safety and mission critical equipment, reactive maintenance efforts may result in unplanned plant shutdowns and delays. For other systems (such as balance of plant), reactive efforts often result in increased expenses, reduced system reliability, and increased operator uncertainty. It is important for NPPs to improve maintenance capabilities so that operator confidence and system understanding is maximized. Improved maintenance strategies are also important for increased economic sustainability of nuclear power. Traditional NPPs operate near full power, and are not designed for load following (i.e. changing the reactor power to meet energy demands). This means that nuclear reactors are most profitable at peak power usage hours (when power is most expensive for the consumer). In order to ensure plant availability during these times, advanced methods of maintenance planning must be utilized. A good reference on the fundamentals of equipment replacement strategies to capitalize on uptime is provided [67].

Preventive maintenance is a comprehensive set of scheduled maintenance actions that are applied to a system to improve system reliability and reduce unanticipated equipment failure. Steps to prevent failures include: installation quality inspection, precise recordkeeping of both equipment maintenance and performance (e.g. vibration analysis), system inspections and minor repairs, and scheduled major repairs/replacements [56]. It should be noted that preventive maintenance is not appropriate in all applications. Jardine and Tsang [68] summarize the two major requirements for equipment to benefit from preventive maintenance: the total cost (including opportunity costs) associated with a specific equipment repair/replacement must be larger if conducted post-failure when compared to pre-failure maintenance, and the equipment hazard rate must be non-decreasing. In NPPs, most equipment related to primary containment and balance of plant is more cost effective to maintain prior to failure due to the mission-critical nature of plant operations [68]. In addition, basic preventive methods are often simple to implement and easy for operators/staff to manage [56]. It is important to mention that, while preventive maintenance may use failure data to establish replacement schedules, failure data alone cannot support establishment or adjustment of inspection intervals. This requires condition data related to the onset of degradation for equipment as well as failure time information.

Weibull analysis is an example of a possible method for implementing preventive maintenance. First, Weibull analysis can be used to determine the hazard rate of equipment and verify that preventive maintenance is advantageous. Type I prognostic models using Weibull fit are a common implementation of equipment specific modeling, and are based on historical failure information. The resultant model is used to characterize the expected life of similar components. Future equipment failure is estimated using confidence in the Type I model, and preventive replacement schedules are established accordingly. Cycles that are terminated prior to failure (e.g. cycles after implementation of a preventive plan) can be corrected or censored, which allows for model updating as more information is collected. Due to the reliance on historical degradation cycles, preventive maintenance methods cannot be effectively utilized without the availability of accurate equipment histories [56]. An example case study for preventive maintenance optimization of major and minor vehicle components is given by [68]. It should be noted that

equipment with redundancies can also benefit from preventive methods when multiple maintenance options are available (e.g. repair and replacement).

2.4.2. Predictive Maintenance

While preventive maintenance is decisively time-based maintenance, predictive maintenance utilizes the condition of equipment to form decisions on when to conduct maintenance. The field of CBM and related research has been well established and studied. Though most literature on CBM is driven by cost-based optimization of maintenance strategies, alternative methods such as safety and reliability motivated CBM have been shown to be equally capable of improving maintenance performance such as increased safety, production quality, and equipment availability [57]. Unlike standard methods of CBM, which use only current or past values of system condition, predictive maintenance attempts to extrapolate equipment degradation and predict when a future failure will occur [7] [8]. This allows O&M to effectively plan maintenance or schedule future outages.

More formally, while preventive maintenance utilizes precautionary methods such as inspections and repairs to avoid system breakdowns, predictive maintenance attempts to quantify equipment condition by modeling degradation trends from operating data [56]. Once a model representing historic failure cycles is developed, future data can be extrapolated to failure in order to predict the RUL of a component/system. Due to the ability to predict the RUL for specific components under specific conditions, predictive maintenance may be significantly more effective than reactive maintenance, and has become the state-of-the-art maintenance strategy in industrial plant applications [5]. The category of predictive maintenance consists of diverse algorithms and tools used to measure, monitor, characterize, and predict equipment operating conditions and signal data. Type III prognostic models are an example of tools used for predictive maintenance. The effectiveness of predictive maintenance depends on expert knowledge applied to failure thresholds and definitions of relationships between signals and system degradation [8]. Signals must be evaluated as degradation parameters before applying a predictive model, which is accomplished using a diagnostic model to determine signal residuals and fault modes. Common examples of diagnostic signals include temperature, pressure, current, vibration, fluid composition, and thermography [5]. These degradation parameters are used directly, or combined into a prognostic parameter, to quantify system/component health. The parameters of historic cycles are used to train a prognostic model, so that future data can be passed to the model to calculate failure predictions. Using resultant predictions such as RUL, proactive maintenance actions may be scheduled and performed so that plant downtime can be minimized or avoided. The supplementary information related to equipment degradation and failure, made available through predictive modeling, improves operator confidence and system understanding. An example of optimizing maintenance decisions using predictive tools with several applications is given by [68]. While predictive maintenance may be very effective in NPP applications, substantial research is still needed to fully understand the benefits of applying these tools, and the effect they will have on plant safety, productivity, and resource optimization [56].

2.4.3. E-Maintenance

There is a common theme in industry to focus efforts on enterprise-driven methods. The last decade has seen the rise of e-commerce, e-operations, e-innovations, etc. These 'e' methods are enterprise-focused improvements to existing practices used to make advancements in production

efficiency, quality, safety, and other metrics of productivity. There is currently a rise in demand for production quality, equipment performance, and system availability that has resulted in a new type of maintenance for industry termed 'e-maintenance'. This enterprise focused ideology is used to define maintenance systems that are efficient and effective, above existing maintenance functions currently used [60] [69]. An in-depth description of e-maintenance and its related concepts and methods is provided by Muller [69].

The primary justification for e-maintenance is non-safety related performance topics. Equipment faults and failures will not always result in significant damage to components or injury to workers or public; however, unexpected failures may cause plant shutdowns, production delays, and unanticipated maintenance requirements. These consequences of reactive maintenance are very expensive when compared to proactive maintenance activities, and reinforce the need for e-maintenance tools [56]. Proactive methods facilitate improved identification of incipient failures and predictions of lifetime estimates compared to reactive efforts, but require the additional effort of translating failure predictions to appropriate maintenance strategies [8]. All means of improving maintenance including diagnostics, prognostics, and preventive actions fall into the category of e-maintenance tools. Organizational benefits of implementing successful e-maintenance strategies include reduced shutdowns, increased equipment utilization efficiency, improved production quality, and a general improvement in employee satisfaction [56]. While e-maintenance is not the focus of this work, it is important to be aware of these efforts when presenting new methods of maintenance to industry leaders.

2.4.4. Repairable Systems

A common assumption in proactive maintenance is that the actions taken during maintenance will return the system to an 'as good as *new*' state. This means that component failure distributions for a replaced item do not change from cycle to cycle. This concept of identical and independently distributed failures is an assumption that is often not appropriate in situations that utilize predictive based maintenance; instead, the idea of repairable systems must be introduced. With repairable systems, components that undergo maintenance are either renewed (returned to an as good as *new* condition), or the system is repaired (returned to an 'as good as *used*' state). One of the original books on repairable systems [70] discusses common problems associated with repairable system applications. A good generalized introduction to repairable systems and the effect that repair definitions have on maintenance optimization is given by [68]. In order to improve predictive maintenance efforts, systems that undergo alternatives to renewal must be investigated to better understand how equipment degradation is influenced by maintenance. In such an examination, definition of competing maintenance actions must be established and used.

All maintenance actions can be characterized by the effect each have on system degradation using one of three repair categories: perfect repair (renewal), minimal repair, or general repair. A simple example of each is given in Figure 11. The importance of characterizing maintenance actions by degree of repair is the effect they have on initial levels of residual system degradation (post-maintenance). For classical renewal or perfect repair, the system is returned to an as good as new condition, and there is no residual system degradation remaining after maintenance. An example of perfect repair is replacing the tires on a vehicle. Any degradation accumulated by the old set of tires is removed, and the component is returned to an as good as new state. Minimal repair is used to classify maintenance actions that return a system to running condition, but remove little or no degradation from the system.

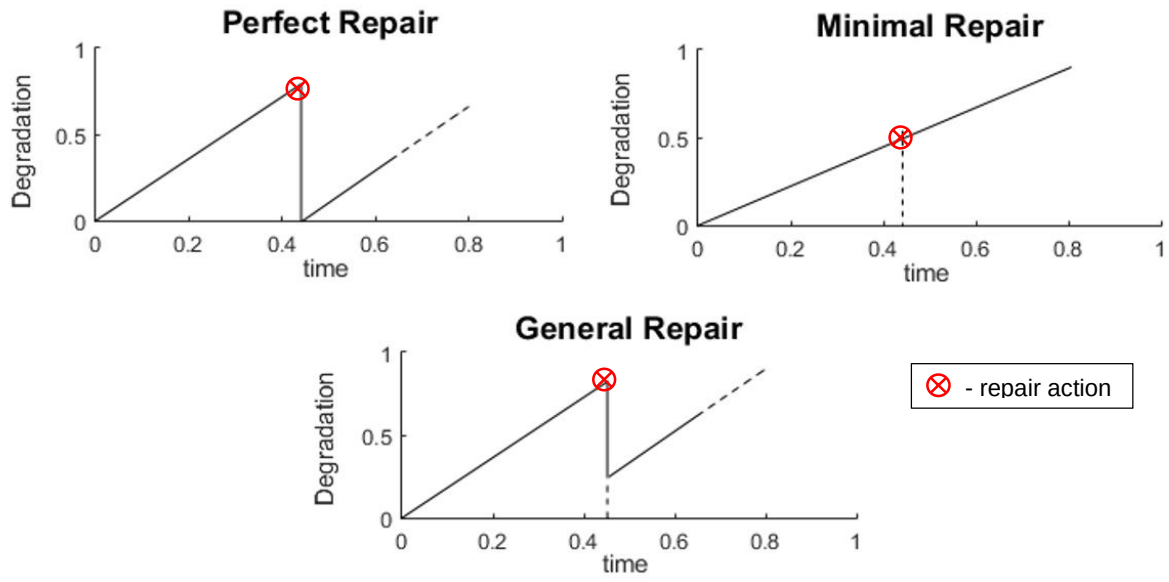


Figure 11: The three categories of equipment repair

An example of minimal repair is using a plug to fix a tire that has been punctured. The vehicle may continue to be driven after the repair action, but other faults such as tread degradation on the tire remain. A less explicit category of maintenance is general repair. Under general repair, equipment is returned to an as good as used state, and a significant amount of degradation is removed from the component/system. An example of general repair is tire resurfacing: the tire is not as good as new because part of the tire retains existing degradation (such as the tire walls), but the tread has been replaced thereby removing associated tread degradation from the system.

These three categories of equipment repair can be simplified into two types of maintenance within commercial plants: repair and replacement. Repair is any non-renewal action that returns the system to working condition and is therefore as good as used. Replacement is an action that returns the system to as good as new, for example replacing a faulty part. Differentiating between maintenance types is proven to improve maintenance optimization efforts [68], and is expected to have significant influence on failure related prediction accuracy. Investigation into how maintenance actions effect system degradation and model development is a major focus area of this dissertation. It is expected that next generation diagnostics systems and future implementation of prognostics programs will develop condition-based models that incorporate specific maintenance action data to better explain equipment degradation [7].

2.5. Summary

This chapter introduced the important concepts related to prognostics and health management in an evolving industry. Section 2.1 discussed the scope of prognostics for equipment management and failure prediction. Section 2.2 introduced the major categories of prognostic model based on availability of data (Type I, II, and III models), and the understanding of underlying system responses to degradation (first principles vs. empirical prognostic models). Discussion of prognostic model input/output was also provided to promote the importance of data related to equipment degradation in order to improve prediction capabilities. Section 2.3 introduced the major sources of condition-related data in nuclear applications, and highlighted the disparate nature of these data repositories. An introduction to a possible data merging solution was presented to support future sections. Section 2.4 focused on the different categories of maintenance strategy (reactive, preventive, and predictive maintenance). To relate the importance of economical/efficient maintenance strategies in industry, a short review of e-maintenance was provided. Lastly, the idea of repairable systems was introduced, which included an overview of the three major categories of repair action. These competing repair actions are critical to understanding the influence of maintenance on prognostic modeling.

The following chapter presents the vital need for advanced maintenance recordkeeping systems in industry. Without the availability of digital maintenance records that can be extracted and processed in (nearly) real-time, the methods presented in future sections of this work will not be possible; therefore, the state of industry, as well as the need to industry to adopt new maintenance practices is formally outlined below.

3. Survey of Requisite Changes to Support Modern Maintenance

The following sections are taken from published work [71] and serve as a formal presentation of the future of asset management and maintenance, as well as a discussion of related challenges and the need for new industry standards.

3.1. Problem Statement for Maintenance Data Standards

Nuclear plants, and most other industrial organizations, do not optimally record, store, or utilize information related to equipment maintenance. Methods of facilitating maintenance information collection and usage are not well defined or standardized across industry, or in many cases even within specific organizations. This creates somewhat of a challenge when attempting to utilize maintenance information to gain understanding of operating behavior, failure patterns, and maintenance effectiveness.

Before diving into the specifics of how maintenance is controlled and information is collected, there needs to be a common understanding of the general maintenance objective. In industrial settings, the naivest purpose of maintenance is to repair equipment so that it can function properly. While this is a very basic way to interpret the *purpose* of maintenance, there is a much more elaborate maintenance *objective*. When one begins looking at the overall intent of continued maintenance, the objective shifts from a position of reaction to a proactive stance. Rather than identifying issues with equipment and conducting maintenance after issues have presented themselves, the proactive maintenance objective is to maximize equipment performance so that systems operate effectively and efficiently. This is done by optimizing maintenance strategies so that component faults and failures are avoided, and/or the resulting losses from these events are minimized [51]. In simplest terms, quality maintenance attempts to operate in a more knowledgeable manner than traditional maintenance.

3.2. Utilizing Maintenance Data to Improve Asset Management

While there may be inadequacies in the way that maintenance data is collected and stored in industry, it is critical that the raw utility of maintenance information is described in order to validate the need for improvements in maintenance recordkeeping practices. When discussing the need for more or better quality information, it is usually a voyage towards increased knowledge and understanding. In terms of overall asset management, five critical questions define the amount of knowledge possessed:

1. What do you own?
2. What is it worth?
3. What is the condition?
4. What is the Remaining Useful Life (RUL)?
5. What should be fixed before time t ?

While questions 1 and 2 are often well understood, most organizations cannot adequately answer the remaining three [72]. This research investigates a claim that improving maintenance record practices will facilitate solutions to these unanswered questions by improving the availability of maintenance data for use in pioneering solutions.

As the landscape of business strategy changes, the importance of effective asset management is increasing to support organizations' competitive capability [5]. Many industrial

assets have substantial influence on operations, and the health of assets is directly related to their performance and availability; therefore, the relationships between maintenance practices, equipment availability, and an organizations performance are interconnected [54] [55]. With this in mind, the future of asset management involves changing existing maintenance strategies in order to improve the inclusive system reliability over time, which requires a shift in focus away from short-term equipment issues towards long term goals of optimized decision support and efficient maintenance (e-maintenance) [51] [69] [73].

When industrial companies adopt e-maintenance plans, the goal is often to provide maximum performance (output and quality) with the minimum number of failures at the lowest required cost. Intermediate improvements of adopting e-maintenance include reduced outage time during repair, minimized spare inventory, overall increase in quality and availability of information, innovative lifecycle support, and advanced fault detection, diagnostics and prognostics [69]. This e-maintenance ideology is part of a World Class Maintenance (WCM) philosophy that includes the use of digital solutions such as CMMS and CM systems, effective implementation of preventive/predictive maintenance strategies, and improved availability of information [51]. A study by Dr. Swanson, a professor in management and production economics, says “the link found between preventive and predictive maintenance, CMMS, lateral relations and maintenance performance... supports the importance of [these] *world class* maintenance practices” [55]. In order to facilitate WCM in existing organizations, a connection between predictive tools and CMMS must be established to support automated scheduling of maintenance. This requires implementation of a CM system, development of predictive models, and installation of a digital maintenance system. Together, this Predictive Maintenance Program (PMP) is integrated into the existing maintenance policies in order to improve operations [73]. This is one method of enabling WCM practices, and can be made more or less complex given the specific needs of the organization. An example of a recent PMP implementation is the F-35 aircraft, which relies on predictive models and information from an advanced CMMS to provide health assessment and decision support [74]. An example of WCM with a simplified maintenance policy is provided by [51], where only a CMMS is installed. The results from improving the organization’s maintenance records include increased equipment availability, reduced spare part inventory, increased employee optimism, reduced number of unanticipated failures, and significant O&M savings. The cost of the CMMS implementation for this application was estimated at 122,000£ (\$176,000 USD), and the savings after (less than) one year was estimated to be over 250,000£ (\$360,000 USD).

3.3. Future of Prognostics for Asset Management Support

Diagnostics is a well-understood discipline of health monitoring, but prognostics is less universally comprehended. Both diagnostics and prognostics are based on assessing the condition of a system, but unlike diagnostics, prognostics attempts to predict future behavior and degradation trends. Both diagnostics and prognostics are tools to aid in asset management support, but must be properly applied to result in significant improvements to O&M. If prognostics is the tool, predictive maintenance is the solution. Predictive maintenance provides improvement to operational safety, quality, and equipment availability across industrial plants by allowing for dynamic maintenance scheduling [8] [73].

As previously discussed, the complexity of an organizations maintenance program (such as the use of predictive maintenance methods and CMMS) follows the complexity of the industrial operations environment. Companies with large or multifaceted processes should investigate the

potential benefits of implementing advanced diagnostics, prognostics, and maintenance management programs in order to improve operations and maintain competitive capabilities across individual trades [8] [55] [73]. Similar to the shift from reactive to proactive maintenance, companies will be faced with one of two futures; either the company will proactively innovate with the pace of progress within industry, or the company will have to react to competitors and adopt these plans to prevent loss of investment/production/profit afterwards. There are many different ways in which an industrial organization can choose to innovate and retain their competitive ability, but each will have varying strengths and weaknesses in terms of return on investment. It is believed that the future of competition will be improved O&M through the use of technologies that enable decision support. Jardine et al. states that the next generation diagnostics and prognostics systems will have advanced capabilities for continuous monitoring and automated decision support, but are dependent on development in the following exploration areas in order to facilitate mass implementation across industry [7]:

- Utilization and improvement of CMMS systems
- Utilization of CBM with diverse maintenance action classification
- Packages and tools for industrial data analytics

The field of research for industrial analytics is developing and evolving at a rapid pace. Tools for industrial data analytics are quickly becoming available, as well as platforms to implement them as effective decision support solutions such as IBM Watson and GE Predix. With the implementation of CMMS systems across industry, these tools could be utilized in the very near future for virtually instantaneous return on investment. The future of diagnostics and prognostics within asset management depends on vital information stored within CMMS. Data such as maintenance conducted, maintenance/failure times, and periodic inspection is invaluable, and in many cases required, for effective model development and validation. This is closely related to the utilization of CBM; information about specific maintenance conducted within the CMMS will improve CBM efforts (e.g. diagnostic and prognostic models). While characteristic CBM utilizes either CM systems or CMMS, employment of CBM by integrating both systems can result in an exponential improvement [8].

3.4. Challenges and Standards

As industry evolves and improves, there is a constant demand on maintenance managers to improve work and preserve efficiency, while reducing or retaining competitive operational expenses [51]. This is a common paradigm of maintenance throughout industrial history, and has increased the need for cutting-edge methods of asset management and development of supporting tools. In terms of the future of asset management, proactive (specifically predictive) maintenance is revolutionizing organizations' ability to properly maintain their equipment, improve decision support, and optimize operations. The greatest challenge associated with implementing these innovative plant asset performance solutions is the disparate nature of the data required to define, create, develop, validate, and evolve advanced diagnostic and prognostic tools [8]. Focusing on CMMS and CM systems, the connection between these data sources is still underdeveloped and weakly understood, which has resulted in issues such as obstructed information flow channels, incomplete maintenance communication strategies, and improper data handling and processing [5] [69]. Another challenge associated with implementing advanced maintenance management tools is their incorporation into existing legacy systems/programs [72].

Links between the CMMS, CM system, Supervisory Control and Data Acquisition (SCADA), and legacy systems must all be identified and defined before effective improvements to asset management can be expected. Once this is complete, there is still the challenge of defining how maintenance workers should communicate and act on the results of predictive models and decision support tools [8]. All of these challenges are significant and should not be overlooked when planning any implementation of an advanced asset management program.

One of the major contributing reasons to the challenges listed above is the lack of standards across maintenance management, diagnostics, prognostics, and decision support. Investigating the scope of existing standards results in limited guiding principles for sensor modules, low-level condition monitoring frameworks, and diagnostic programs. These “standards” are partially developed and are not robust to integration with predictive systems or digital maintenance management programs. Both CMMS and prognostic systems are immature in terms of standards and technique guidelines for effective development/implementation. Focus must be applied to researching and establishing standards in these areas, as well as standards for future decision support modules [69]. If one considers these systems together as part of a PMP, there are some specific needs for standardization of the entire program.

Implementation of a CMMS is considered a requirement for an industrial PMP [73]; however, standards on information within the CMMS must be defined and developed. If data from a CMMS and CM system is to be integrated for predictive capability, then there must be information on time between maintenance events, specific maintenance conducted, explicit identification of systems and equipment, and availability of historical maintenance records. While standardizing the specific format of CMMS entries is not inherently necessary, there should be capabilities for data export and import, so that maintenance information can be integrated with other data sources, and decision support modules can feed back into the CMMS to improve/automate maintenance planning and strategies. Within the individual CMMS records, there should be standardized options for maintenance workers to choose. This is a common issue, as maintenance programs where failure type and/or maintenance action taken is not standardized in some fashion, CMMS entries for the same system, failure, and repair event may be vastly different due to different workers submitting the entries. These systems require advanced natural language processing to extract any useful information to improve maintenance strategies, and are inherently problematic for organizations that wish to simplify and optimize operations. To this end, in organizations that implement CMMS systems for the end goal of effective PMPs, standards on event entries should be well defined to aid in future data extraction, processing, and analysis.

Once information from CMMS and CM systems has been extracted and collected, predictive models must be developed to characterize system degradation and make predictions of RUL. These prognostic models are typically developed by applying thresholds, degradation trend analysis, and advanced statistical methods to failure data, in order to accurately quantify system health and extrapolate current conditions to failure. There are no widely accepted standard models or frameworks for developing these prognostic models in practice, which results in unique model development and results for each application. Two individuals developing a predictive model using the same information will likely make starkly different models with significant differences in final prediction results. Industry is currently focused on improving diagnostic “standards”, and have deferred predictive model framework decisions to expert knowledge and end-user historical experience [8]. While the exact disadvantages to the lack of prognostic modeling standards is impossible to quantify, it does pose a challenge for resulting decision

support methods. Without a standardized way of developing prognostic models or quantifying predictive results, each model requires different interpretation within the subsequent decision support system. This increases the complexity of the model development process, and requires additional expert knowledge to define and explicate. At this early age of prognostics and decision support, it is not reasonable to expect perfect standards to be developed or defined. Instead, as prognostic systems are implemented within PMPs, care should be given to carefully identify steps taken, and the purpose of specific modeling techniques, in order to retain the value of results. Without this, additional system health information may be available, but the utility of the information and ability to improve decision-making will be lost or tarnished.

3.5. Summary

This chapter was intended to present a formal discussion of modern maintenance strategies, and to identify the need for better industry standards related to maintenance recordkeeping and data storage. Section 3.1 discusses issues currently pertaining to maintenance practices, and how these issues are caused by a general lack of technology and standards for maintenance across industry. Section 3.2 provides an overview of the ways that maintenance data (if properly recorded and stored) can improve asset management. Section 3.3 build off the previous sections, and discusses the future of proactive maintenance efforts for asset management and decision support. Section 3.4 highlights the challenges in reaching an advanced maintenance-focused industry due to improper data pathways, handling, and utilization. Specific mention of critical standards needed to facilitate advanced maintenance practices is also provided.

The following chapter discusses the methodology used in this research to investigate the theory that maintenance information will improve predictive maintenance capabilities. Specific methods to generate usable datasets are provided. Discussion of data integration is given with focus on a framework for extracting operating cycles with the availability of maintenance records. A brief description of the overall framework of prognostic model development is provided with discussion on the metric evaluation of prediction accuracy in order to compare results across competing models.

4. Research Methodology

The following sections outline methods used to evaluate the theory that incorporating maintenance information in predictive maintenance models will improve prediction accuracy. Discussion of the need for additional data and the methods used to integrate maintenance and process data is provided with focus on a proposed framework for extracting usable prognostic cycles. A general outline for advanced (Type III – Condition-based) prognostics is given that transforms operating cycle data into predictions of RUL for associated equipment. Lastly, a discussion of the prediction metric used to quantify model error and compare prediction accuracies across competing models is given.

4.1. Data Generation

Collecting process data and maintenance data for equipment and systems is a challenge. Many organizations do not collect maintenance data in a format that is directly usable for data extraction purposes. Due to the large amounts of process data collected, most plants compress data archives for historical data over a few months old. Most organizations also treat data related to system failures as sensitive information, and are hesitant to share this data. In each of the applications investigated in this research, data is simulated or real data is augmented to allow for investigation of the influence of maintenance information on prognostic predictions. The following subsections outline the methodology used in each study to generate process data and maintenance information in lieu of directly usable data from industry. The purpose of generating cycle data in each study is to develop competing prognostic models and evaluate the influence of maintenance data on each model's prediction accuracy. Discussion of the specific data generation methodology is given prior to the application and results in order to simplify and clarify the findings of each study. In all of the following sections, the end result is appropriate process data cycles with respective maintenance action information. These resultant operating cycles are used in Section 5.

4.1.1. Proof of Concept Data Generation Tactics

Due to the specific needs of this research, real data is not available for use in model development or algorithm validation. In order to develop a proof of concept, data is simulated to represent existing partial datasets provided by GE (Alstom Power). The exact specifications of the system used to collect data are protected under non-disclosure. Data is simulated for a semi-passive component that undergoes a general form of thermal efficiency degradation. The system is comprised of two arbitrary features that represent degradation parameters for the component. Each feature can be considered a function of two or more signals correlated with thermal efficiency degradation (for this system). Data is simulated to affect degradation levels based on repair methods taken at failure. Repair actions are simulated for the two major groups of maintenance strategy (repair and replacement), which alter the functional path for the degradation parameters. Maintenance database entries are generated to store repair actions and system information, in order to validate algorithm designs.

Based on relevant literature discussed in Section 2.3, the two major sources of disparate data in industrial plant applications are CM system (or historian) data and CMMS data. To determine their influence on prognostic model improvement, data from both sources is simulated to

represent a repairable system. The methods of designing the data simulation for process and maintenance data are discussed below.

4.1.1.1. Proof of Concept Process Data Simulation

The simulated data is designed to represent a real system that undergoes thermal degradation. A real dataset was used to aid in the design of certain simulation aspects. The real data consists of approximately 10 cycles that undergo two unique categories of repair: minor repair and overhaul (major repair). Unfortunately, the available data does not consist of failure cycles, and there is no predefined definition of soft failure (cycle termination criteria); therefore, the data is not suitable for algorithm or model development. One possible feature of the real data to analyze was the difference in degradation removed from the signals based on the maintenance action conducted. The analysis showed that almost twice as much degradation is removed from the signal during an overhaul compared to minor repair. The following assumptions are based on the real data analysis and previous experience with similar systems:

1. Degradation removed from system/equipment and visualized in the signal response (more appropriately, each degradation parameter response) is dependent on the maintenance action conducted.
2. Perfect repair returns the system to an “as good as new” state, while consecutive repair actions may result in compounding residual degradation after maintenance.
3. Degradation rate and hazard rate are larger for cycles that undergo a lesser degree of repair.
4. The degree of repair is proportional to the TTF distribution mean value, and inversely proportional to its variance.

An example of a single simulated cycle after repair is shown in Figure 12 to illustrate assumptions 1 and 4. Regardless of the level of degradation prior to observation ‘0’, if the system undergoes perfect renewal it is returned to nominal levels. As a result, if we assume that a cycle similar to cycle 2 (C2) occurred prior to cycle 1 (C1), then the degradation removed due to perfect repair is larger than general repair, and the result is a difference in the initial degradation level of each cycle (Δ). This value is non-zero, which represents assumption 1 above. While degradation rate is also influenced (assumption 3), observation of the effect is difficult due to added noise. Figure 12 also shows that the length of C1 is longer than C2 (TTF about 20 observations larger); this represents the effect of assumption 4. Figure 13 illustrates the degradation removed by comparing Δ to the final level of degradation in the cycle prior to failure. This difference (D_r) for repair actions is one of the ways that maintenance changes how the data is simulated. The amount of degradation removed from the system during repair is a function of the final degradation for the previous cycle, and the maintenance action decision (simulated). This relationship between the final degradation of the prior cycle and the initial degradation of a subsequent cycle is due to the capacity for maintenance repair quality based on the amount of present degradation.

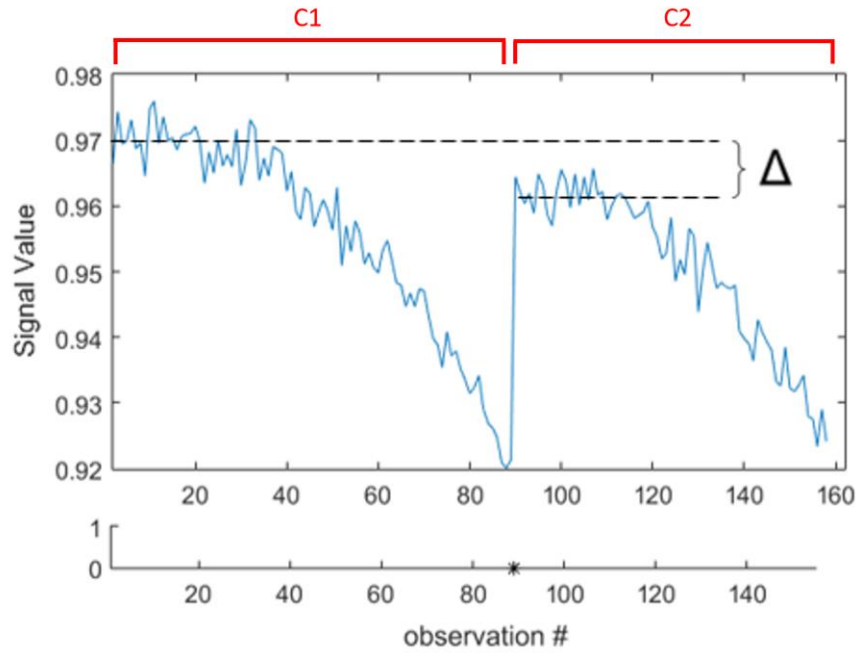


Figure 12: Repair action influence on simulated data parameters

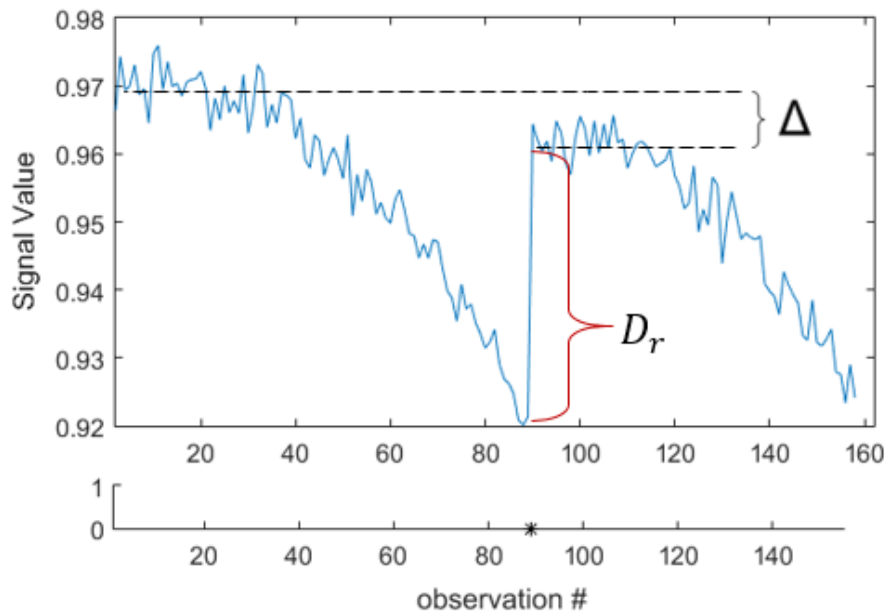


Figure 13: Visual depiction of the degradation removed (red) by maintenance action 0 (repair)

4.1.1.2. Proof of Concept Simulated Data Fidelity

There is much cause for concern when using simulated data for model development. It is easy to overdesign the simulation parameters in favor of the model assumptions. In traditional model development, inputs and outputs are typically known, and the goal is to develop a model that relates the input to the correct output. When designing a model with simulated inputs, the accuracy of the model is entirely dependent on the assumptions made about the input data, and the degree of simulated data “realness”. The following tweaks are all methods used during data simulation to improve real data representation:

- No direct numbers: all values are taken from a generated sample distribution. This increases variability between cycles.
 - Distributions are loosely modeled on real data of a similar system with unique repair strategies.
 - Degradation trends, signal values, onset of degradation, TTF, and signal correlation are all parameters of the simulation that are sampled from distributions or taken from real data.
- Overlapping distributions: all repair and replace specific sampling distributions are designed with overlap to account for uncertainties in assumptions about maintenance dependent parameters.
- Installation error: both replaced and repaired cycles are randomly subjected to installation error resulting in small additions to initial degradation. This tweak improves simulation diversity and represents a common occurrence in plants (not specifically this application).
- Continuous random and periodic noise: both forms of noise used to represent real systems.

These efforts are all targeted to avoid over-design and are expectedly robust for proof of concept utility.

4.1.1.3. Proof of Concept Maintenance Data Generation

A CMMS is intended to record dozens of elements related to equipment condition. The *simulated* CMMS used in this study focuses on the following eight characteristics.

Σ — {	• Location	-Location within the plant (e.g. primary containment, etc.)
	• Operating Status	-Active if online component, Passive if redundant component
	• System #	-i.e. Primary heat exchanger system
	• Part Identifier	-Number associated with the specific part
	• Failure Mode	-Failure mode classifier based on inspection (if available)
	• Maintenance Action	-Type of maintenance conducted (i.e. repair or replace)
	• Time of Maintenance	-Date-time object with time of maintenance scheduled

Each of these pieces of information is critical for facilitating predictive maintenance. The characteristics marked Σ are explicitly used to extract prognostic cycles within the coupling algorithm. The values of system number and part ID are set for the specific simulation being run (i.e. System 1, Part 1). The simulated data is limited to only one failure mode; however, data could be generated with multiple failure modes. Maintenance actions are limited to repair or replace within the scope of this project, but can easily be expanded for other applications. The CMMS simulation code back-calculates failure time based on the sampling rate and TTF of the simulated process data. In the absence of recorded failure times, time of maintenance can be used to represent the limits of the failure cycle. However, additional processing may need to be applied

in order to account for time between when failure occurs and when the work order is scheduled. All of this information related to the simulated data is generated and collected; a reduced example of the resulting CMMS simulation entries can be seen in Table 1. Current data extraction efforts utilize all information apart from location and status, which may be used in other applications. Additional discussion of simulated CMMS usage is provided in Section 4.2. Once process data is simulated and respective CMMS entries are generated, methods can be applied to extract failure cycles using data from both sources.

4.1.1.4. Analysis of Simulated Data and CMMS Records

The output of the data simulation code is a long concatenated dataset with multiple cycles in a single data file. The results of simulating 7 cycles are shown in Figure 14; the system starts off new and degrades until failure, which occurs around observation 80. During maintenance, the system is renewed ($y=1$), therefore the amount of degradation removed is approximately 100% (some initial degradation present in replacement cycles included to represent installation error). The second cycle begins, degrades, and fails around observation 165, where it is repaired back to working condition. It can be noted at this point that the initial signal level for the 3rd cycle is noticeable lower than for the previous cycles. This is because the calculated degradation to remove (D_r) was less than the previously calculated values for renewal. Looking at the degradation trends for the 3rd cycle, the values begin to decrease more rapidly than those of the renewal cycles do. The cycle appears to fail around observation 225. That makes the life of renewal cycles 1 and 2 approximately 80 and 85, whereas the life of the repair cycle was only 75. This is a result of repair actions having a statistically lower TTF compared to renewal. Cycle 3 fails and is again repaired; this causes a compounding of remaining degradation after maintenance. The degradation removed is less than 100% of the degradation remaining in Cycle 3. This means that Cycle 4 signal values are even lower than in the beginning of Cycle 3. This is the result of the compounding degradation feature that was implemented to improve simulation fidelity.

The simulation can be run very quickly to generate a significant number of cycles. Large population statistics can help better explain the diversity of the simulation parameters. A binned histogram of the initial degradation values for 50 simulated cycles (single signal) is shown in Figure 15. The broadening of the distribution for repaired cycles is a result of compounding initial degradation levels. There is overlap between the two distributions, which can be increased or decreased to test the effects on model results. This overlap is used to represent uncertainty in the effect maintenance has on cycle degradation, while still supporting the assumption that degradation levels are a function of the degree of maintenance performed.

Another method of examining the data is by comparing the distributions for maintenance based TTFs. Rather than simulating degradation path parameters and a critical failure threshold in order to calculate TTF, this approach samples a TTF from a normal distribution. Results of this method can be seen in Figure 16 for the 50 cycles. Again, distributions were designed with overlap to compensate for uncertainties in the effect that maintenance action has on subsequent cycle TTF. This is a simple method of generating usable TTFs for proof of concept. Once the process data and maintenance actions have been simulated, the information can be used to generate a CMMS database. An example of the generated work order entries was provided in Table 1.

Table 1 - Example of CMMS entries from the simulated CMMS code (taken from .csv file)

System #	Part ID	Location	Status	Fail Mode	Maintenance Action	Time of Maintenance
0001	0001	SimSystem	Active	1	Replaced	6_11_2016 15:08:33
0001	0001	SimSystem	Active	1	Repaired	8_22_2016 23:50:01

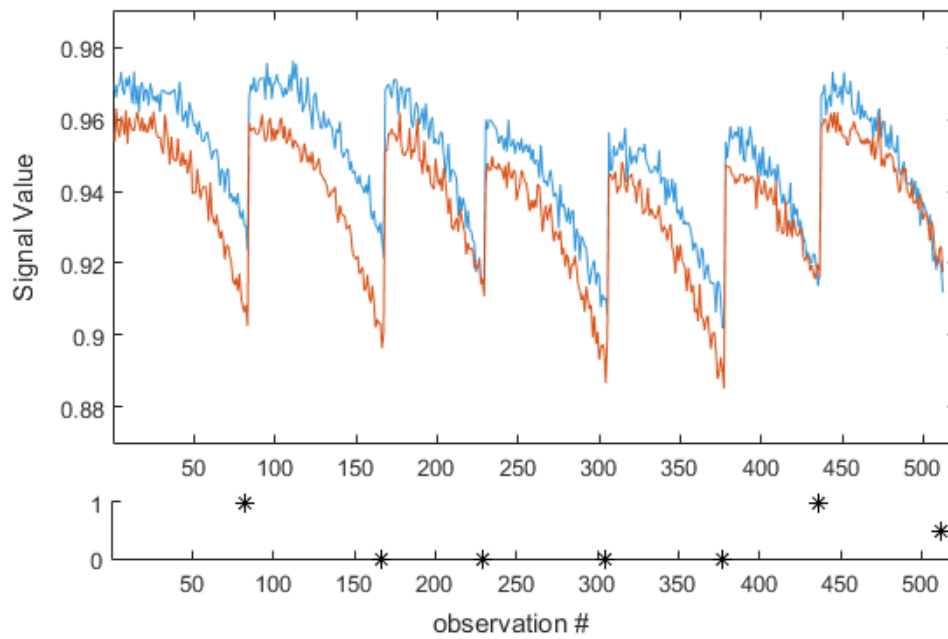


Figure 14: Simulated cycles (top) undergoing different repair actions (bottom)

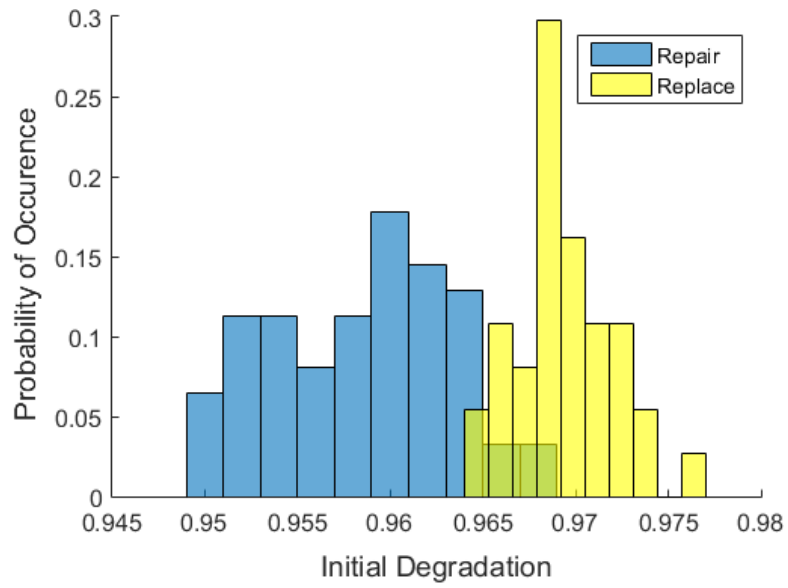


Figure 15: Maintenance dependent initial cycle degradation levels for feature 1

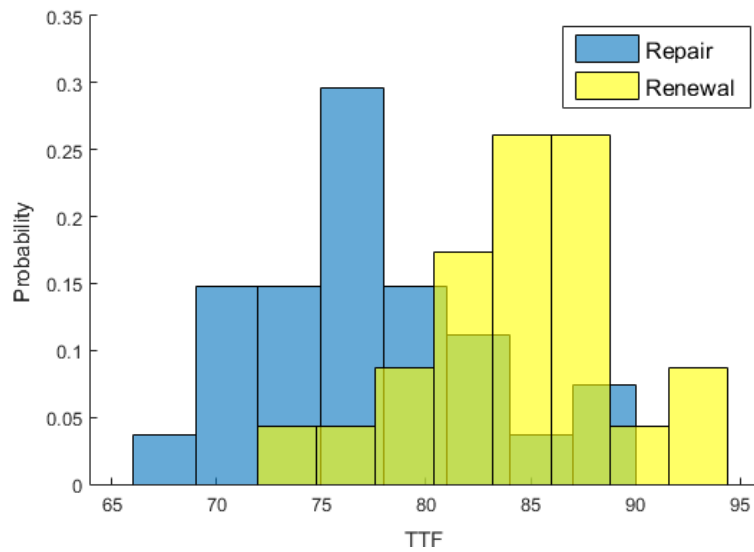


Figure 16: Maintenance action dependent TTF samples from simulated dataset

4.1.2. Primary Case Study Data Generation Tactics

The primary case study investigates a condenser system that undergoes thermal efficiency degradation. The real data is marginally impacted by maintenance action, therefore a simulated dataset that resembles the system but is significantly influenced by maintenance action is created. The studies original dataset consists of cycles that undergo two types of maintenance action denoted as major overhaul and general repair, and are henceforth be referred to as repair and replace respectively. The dataset contains 8 replace and 38 repair cycles. Preliminary analysis showed that the system was only influenced by maintenance action through changes in initial degradation levels, which were marginal. When modeling of the system was attempted, results were unsuccessful due to high variability in cycles. This variability, and the low number of available replacement maintenance cycles, resulted in low coverage of the training space. To overcome data sparsity and maintenance dependency related issues in the study, two datasets were generated:

Simulated Data: dataset designed to resemble the available real data, but designed with maintenance dependent parameters: initial degradation, degradation rate, and time to onset of degradation.

Augmented Data: dataset expanded from the available real data. Cycles created by sampling parameters from the historical data, and using sampled parameters to generate additional failure cycles to improve training space coverage and increase number of cycles available for training.

The two datasets are referred to as simulated and augmented datasets respectively for the remainder of this primary case study discussion. The simulated data is used explicitly to evaluate the capacity of improvement, and is designed with maintenance influence on several operating parameters. This allows for a conceptual investigation into the capacity for predictive model improvements for maintenance dependent versus independent models. To validate the existence of improvement to prediction accuracy for maintenance-based prognostic models, the augmented data is used. Additional discussion on specific data generation tactics for the primary case study is provided in the following sections.

4.1.2.1. Condenser Data Simulation

The condenser simulation was designed to generate data similar to that of a plate-style heat exchanger like the one shown in Figure 17. In the heat exchanger system, alternating hot and cool liquid flows within thin channels. Heat transfer between the plates occurs, and can be calculated from recorded process signals. As the system runs, degradation occurs as a result of particulate buildup collecting on the heat transfer surfaces of the hot and cool channels. This process reduces heat transfer efficiency, and is a common failure mode for many heat exchangers referred to as fouling. Fouling's influence on thermal efficiency degradation can be quantified using deviations in signal values between normal and faulty behavior. This degradation often follows the path shown in Figure 18 shortly after the onset of degradation. While common heat exchanger degradation follows an asymptotic decay towards equipment End-of-Life (EOL), the path of degradation in early-life is roughly polynomial [75]. Due to its similarity to service data, the early-life profile shown in Figure 18 is used to design the heat exchanger simulation.

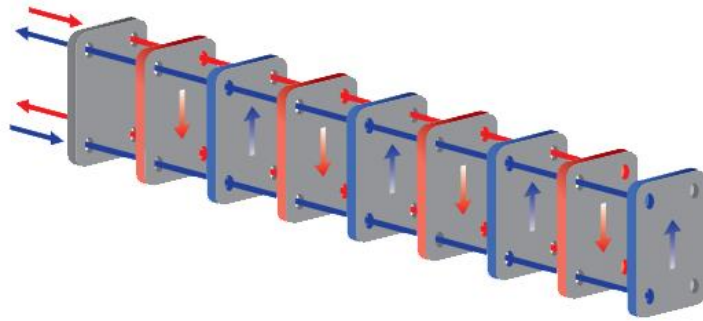


Figure 17: Plate-style heat exchanger [76]

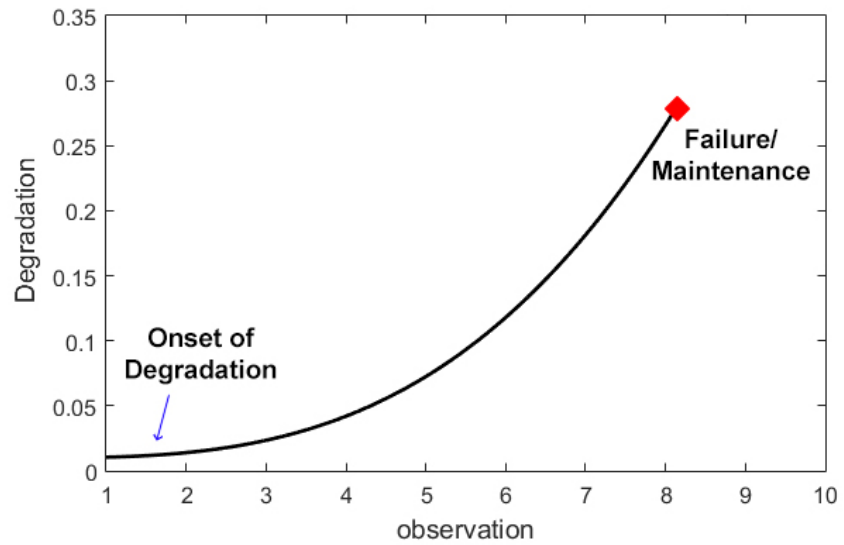


Figure 18: Early-life trend for heat exchanger degradation from fouling

The goal of the simulation is to develop additional maintenance cycles that are suitable for use in prognostic models. The method used to generate multiple runs with maintenance dependent process values is pseudo-random distribution sampling. Each critical cycle parameter has a representative distribution. These parameters include initial degradation level (amount of degradation removed per maintenance action), time before onset of degradation, and degradation rate (slope of path). Once each of the parameters has been sampled, they are used to generate a new cycle with arbitrary length (to be truncated by desired failure threshold). In order to simulate specific maintenance events for each cycle, two uniform distributions are used with the following domains:

<u>Domain</u>	<u>Sampling Criteria</u>
≥ 3 consecutive repairs	15% repair chance
< 3 consecutive repairs	65% repair chance

The result is approximately 70% cycles that undergo repair and 30% replacement. The TTF for repair and replacement cycles is calculated by applying noise to the generated degradation path and truncating at the defined threshold (in this study, the threshold is set arbitrarily). The simulation is designed so that the resulting TTF distributions would reflect the maintenance action dependency seen in the operating parameters.

4.1.2.2. Simulated Data Parameters

As mentioned, the three major cycle parameters are the initial level of cycle degradation post-maintenance action, rate of degradation, and time to degradation onset. Each of the parameters are highly correlated with maintenance action for the simulated dataset, and are therefore representative of the effectiveness of repair and replacement comparatively. The maintenance dependent parameters also influence cycle TTF. This is an important consequence, as the TTF distribution is used for prior information in some advanced Type III prognostic models.

The initial degradation parameter represents how much degradation is removed depending on the maintenance action conducted. It is assumed that any maintenance action that returns equipment to an as good as new condition, will remove nearly 100% of any degradation present after the previous cycle. In contrast, any maintenance action that returns equipment to as good as used condition will remove degradation less than 100%. The application of each simulated maintenance action influences the resulting initial degradation levels, which can be compared by examining the cycle values during early life observations. Initial degradation is calculated in this research by taking the median value of the first 5% of cycle observations. An example of distributions for simulated data initial degradation for each maintenance action is shown in Figure 19. From the figure, the values of the initial degradation parameter for repair cycles are larger on average than those of replacement cycles. There is some overlap where the worst of the replacement cycles are comparable to the best of the repair cycles. This overall relational effect between maintenance actions is appropriate since we assume that equipment that is as good as new should typically contain less degradation than equipment repaired to as good as used. One effect seen in data from real systems is the persistence of degradation across repair cycles, which leads to compounding degradation. For example, a cycle that runs to failure with a final degradation level of 0.8 may be repaired. If the repair action results in 50% degradation removed from the equipment, then the degradation present in the following cycle will have an initial value

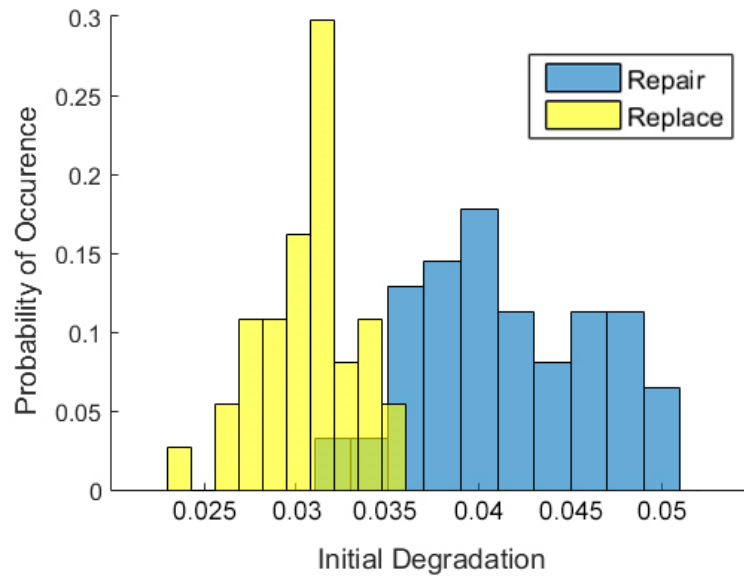


Figure 19: Simulated maintenance dependent initial degradation levels

of approximately 0.4. To create this effect, rather than directly sampling initial degradation from a distribution, each repair cycle has an initial degradation value calculated using Equation 4.

$$ID = ED * [1 - n(\mu, \sigma)] \quad \text{Eq. 4}$$

In Equation 4, ID is the initial degradation level of the current cycle, ED is the end degradation level of the previous cycle, and $n(\mu, \sigma)$ is a normal sample distribution used to select fractional degradation removed from the system with μ mean and σ standard deviation. This distribution can be manipulated to create variance in initial degradation levels for subsequent repair actions.

In contrast, the initial degradation level for replacement cycles is directly sampled from a representative distribution. The influence of maintenance actions on all initial degradation values is not designed to represent a specific physical system. Instead, these sampling methods are used to create the influence of maintenance action on the initial degradation parameter in order to study the resulting impact on maintenance dependent prognostic modeling. One consequence of this method is that subsequent repair actions will have varying initial degradation levels similar to the difference in levels between repair and replacement. As a result, it would equally make sense to separate repair maintenance actions based on their location in a series of repeated repairs (e.g. first repair, second repair, third repair). Without dividing repair cycles in this manner, variance in the cycles is expected to result in larger prediction errors compared to predictions for replacement cycle models. While division of repair actions is not investigated in this research, it is important to mention, so that differences in prediction error improvement between repair and replacement models are better understood.

The method of generating initial degradation parameters for each maintenance action described above is a simple example of the process used to develop usable distributions. The process for generating a maintenance dependent degradation rate and sampling a degradation onset time is similar, but do not require functional manipulation such as Equation 4. To simplify results, explicit descriptions and visual examples for degradation rate and onset of degradation are not provided, but look similar to that shown in Figure 19.

4.1.2.3. Simulated Data Analysis

The primary purpose in creating the described simulated dataset is to develop competing prognostic models. By utilizing the data to train and test maintenance dependent models, a proof of concept can be established. This defines the capacity of prediction accuracy improvement for maintenance dependent modeling. An example of the simulated data across multiple maintenance cycles is shown in Figure 20. The figure shows a degradation parameter for the simulated heat exchanger system across seven maintenance cycles, and provides the corresponding maintenance action for each. The lower subplot is a binary representation of the maintenance conducted, where 0 represents a repair action and 1 represents replacement of the equipment. One noteworthy observation is the non-decreasing initial degradation level during the period of four consecutive repair actions. This is the result of each repair inheriting partial degradation from previous cycles defined in Equation 4. By separating the cycles based on maintenance action, the relative differences between degradation paths for each can be visualized in Figure 21. The figure highlights the differences between the degradation paths as a result of maintenance action influence; these differences stem from imposing different amounts of random fluctuations in degradation parameters. Repair cycles begin with more initial

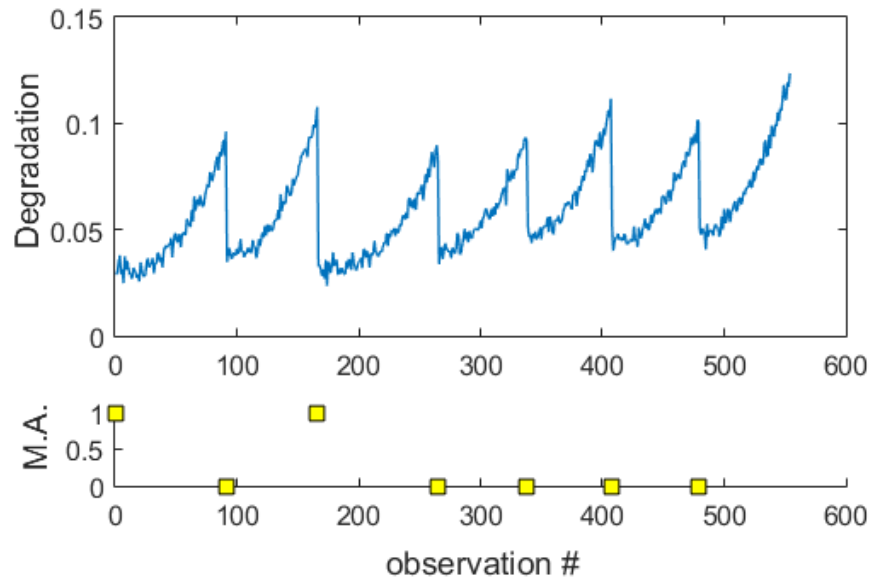


Figure 20: Simulated degradation parameter (top) and respective Maintenance Actions (MA) (bottom) across several maintenance cycles

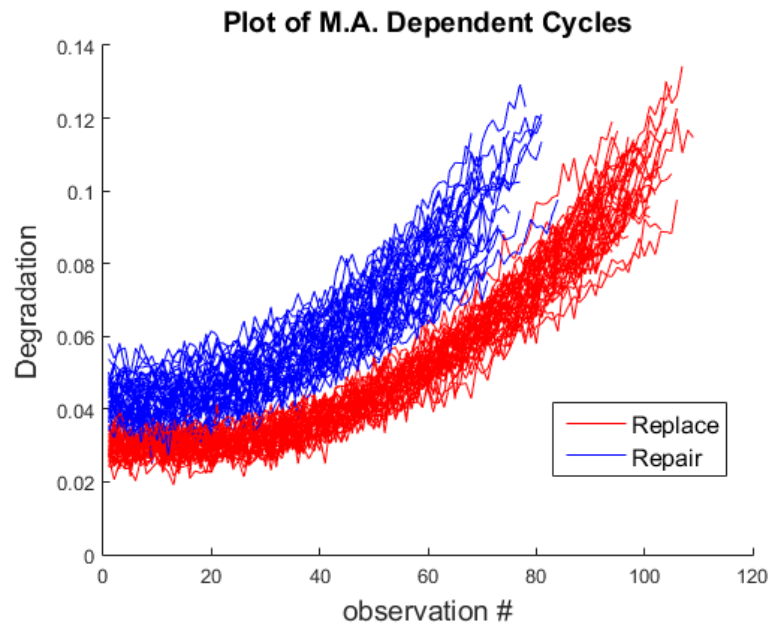


Figure 21: Simulated maintenance action dependent degradation parameters

degradation, have a larger degradation rates, and begin degrading much faster (degradation onset) than replacement cycles comparatively. Together, the influence of maintenance action on these parameters creates a resultant influence on the simulated cycle TTF, which is shown in Figure 22. The units of TTF are arbitrary, but are modeled to resemble the scale of operating days for an industrial heat exchanger. From Figure 22, a significant difference between expected failure times for cycles based on maintenance action can be seen. Typically, repaired equipment lasts 75 days, while replacements last around 90 days on average. This information is suitable for Type I modeling, or as prior information in Type III prognostic models that utilize Bayesian methods.

4.1.2.4. Augmented Condenser Data Generation

The simulated dataset is used to establish a capacity of improvement for maintenance dependent modeling. Similar to the simulated dataset, the augmented dataset is created to represent a thermal degradation system with two distinct maintenance actions (partial repair and full repair). The primary difference between the two datasets is the utilization of real data collected from industrial equipment to directly generate the augmented data. From the available real data, representative cycles are randomly sampled to use as training data, while the remaining cycles are saved for future validation. The available training data is analyzed and processed into distributions representing each cycle degradation parameter, which allows additional cycles to be generated as needed. This process is used to increase coverage of the training data space, and was selected due to high uncertainty present in the real data cycles discovered during early analysis efforts. Figure 23 provides a simplified visualization of the process. In Figure 23a, the training data is collected and used to measure degradation path parameters for each cycle. The histogram for a specific parameter across each cycle is created (Figure 23b), and a representative distribution is applied to the data (red). New parameters are randomly sampled from the resulting distribution based on the number of cycles being created (Figure 23c – green). Lastly, the sampled parameters are used to create additional cycles that lie within the original real data training space (Figure 23d). Since each of the samples lies within the scope of the real data, the augmented data cycles are treated as such. To visually compare the similarities between the real data and generated distributions, a QQ plot is used. Standard MATLAB software packages were used to make calculations and generate plots. Data in a QQ plot is represented using quantiles of the sample compared to quantiles of the applied Weibull distribution. The accurateness of the Weibull fit for the training data can be visually assessed by comparing the data samples (discrete values) to the continuous distribution. A QQ plot for the partial repair initial degradation distribution is given in Figure 24. Distributions are generated for each parameter and validated using QQ plots for basic representation accuracy. As previously mentioned, the resulting distributions are used to sample new cycle parameters, which can be used to generate new cycles. These parameters are initial degradation, degradation rate, and time until onset of degradation. An example of the resulting maintenance dependent initial degradation distributions is shown in Figure 25. The distributions reveal that partial repair cycles start at a higher average initial degradation level compared to full repair cycles. However, there is significant overlap across operating space. This relationship is similar to the simulation methods used earlier in the case study.

Aside from the three major maintenance data parameters, cycle noise is also sampled from the available real data. Together, each sampled parameter is used to generate new cycles as needed, and a realistic noise level is added to improve realism. An example of several augmented degradation cycles is provided in Figure 26. One major observation from the augmented cycles

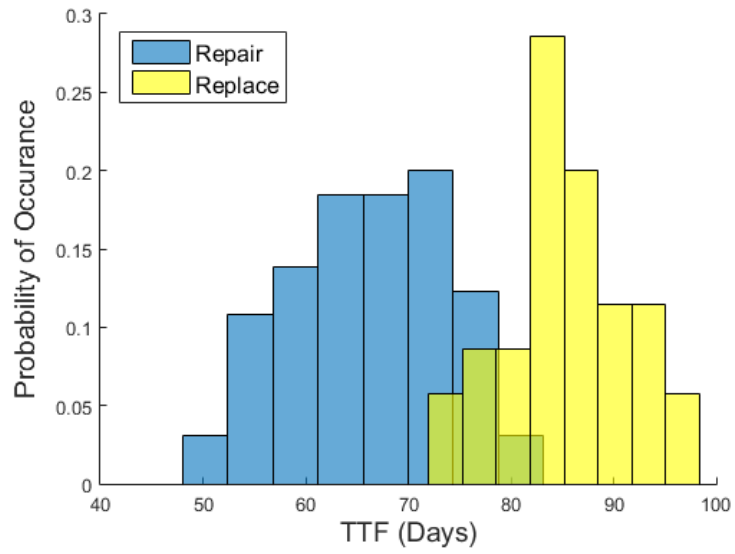


Figure 22: Maintenance dependent TTF for 100 simulated cycles

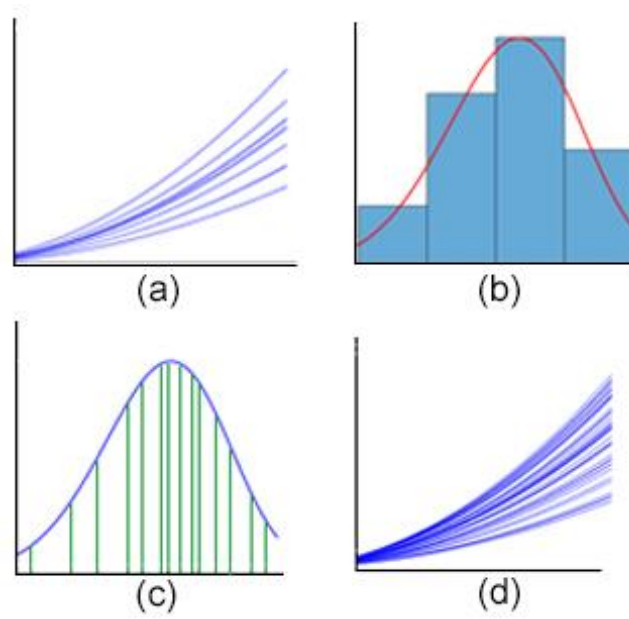


Figure 23: Simplified process for sampling augmented data cycles

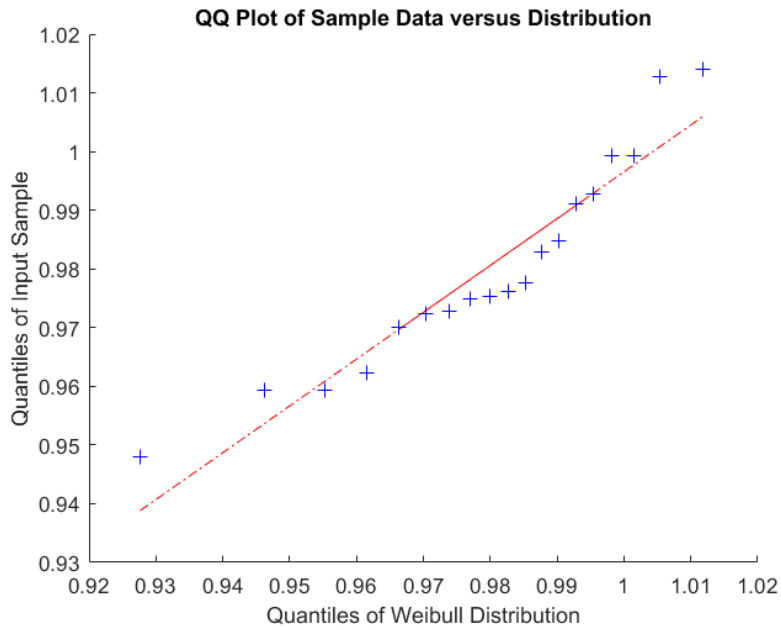


Figure 24: QQ plot for partial repair initial degradation Weibull distribution (red) compared to training data (blue)

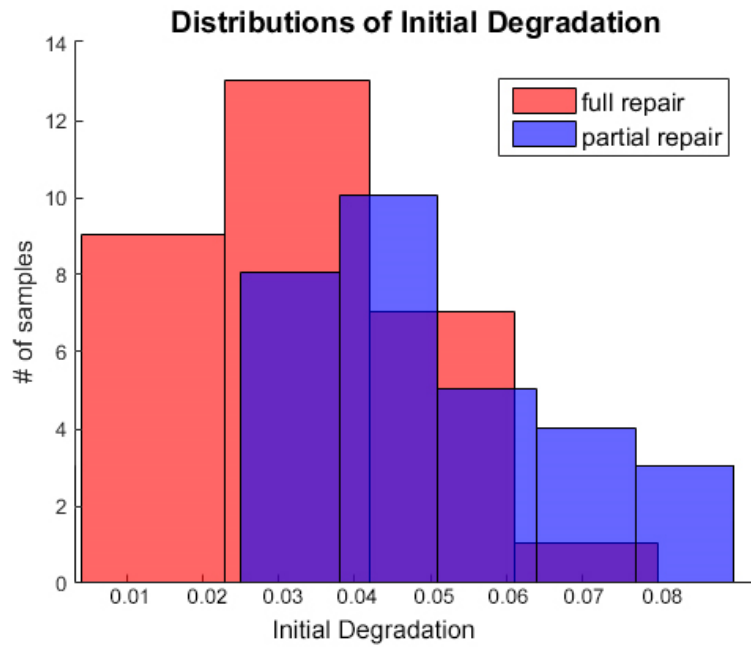


Figure 25: Distributions of initial degradation for the augmented data

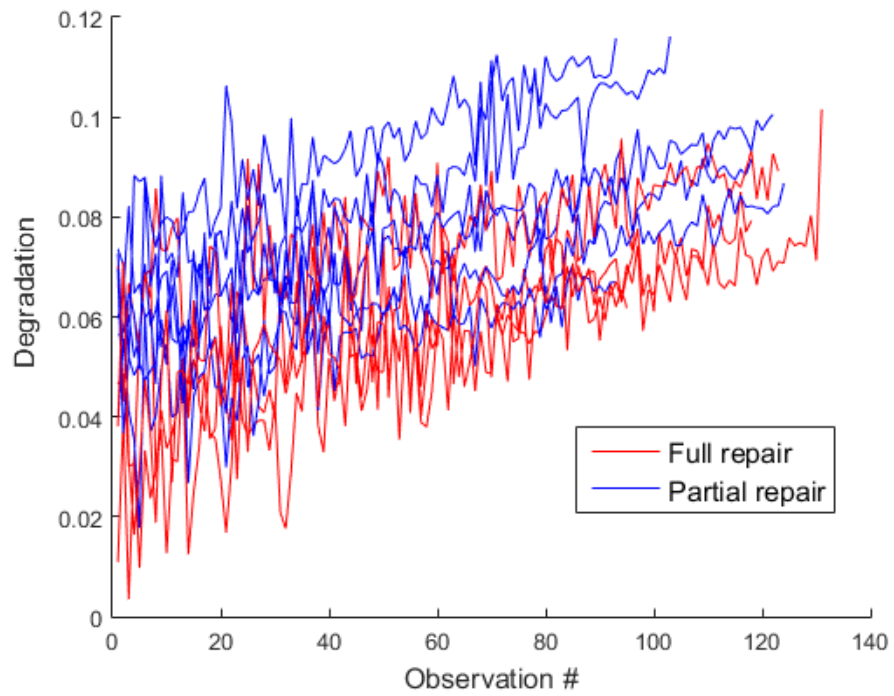


Figure 26: Augmented failure cycles using sampled parameters from real data distributions

is the difficulty identifying maintenance influence compared to the results of the simulation in Figure 21. This effect is caused by high variance present in the real data samples, as well as a minimal impact on cycle parameters due to maintenance action. From the cycles in Figure 26, the initial degradation appears to be somewhat dependent on maintenance action, in agreement with the results in Figure 25. However, there does not appear to be any strong correlation between maintenance action and degradation rate, and it is not possible to tell when onset of degradation occurs. While this may make analysis difficult and will likely affect modeling, it is similar to the nature of the sampled data. As a result, the augmented data is not inherently suitable for supporting the maintenance dependent modeling theory. This is the primary reason for utilizing both the simulated data and the augmented data. The simulated data is used to show the capacity for predictive model improvement when utilizing maintenance-based modeling, whereas the augmented data is used to show how effective maintenance dependent modeling is for a system that is relatively mediocre in terms of maintenance action influence.

4.1.2.5. Augmented Data Validity

When using simulated and/or augmented data, one common question that arises is the validity and integrity of the data. One could assume that data is overdesigned to fit the application, and does not produce results that can be validated for general conclusions. For this case study, it was mentioned that the available real data was sampled into representative distributions for each cycle parameter. The available data was also sub-sampled for training and testing so that the data was not over-fit to the operating space. Another method of improving the validity and integrity of the augmented data was to restrict all parameter sampling to within the domain of the observed historical values. This method can be explained using the bounded Equation 5.

$$f(x | \beta, \eta) = \begin{cases} \frac{\beta}{\eta} \left(\frac{x}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{x}{\eta}\right)^{\beta}\right) & \text{Eq. 5} \\ 0, & \text{for } x > u \\ 0, & \text{for } x < l \end{cases}$$

Equation 5 defines the probability $f(x)$ that a value x will be sampled from a Weibull distribution. If x is outside of the upper and lower bounds (u and l respectively), then the probability of sampling that value is zero. This method of restricting sampled values prevents augmented cycles and observed cycles from differing due to low-probability extremes at the tail-ends of a distribution.

Two methods of verifying the augmented data integrity are discussed formally in the following paragraphs. In Figure 24, a QQ plot is used to visualize the effectiveness of applying a Weibull fit to the generated data parameters. While the QQ plot is useful for comparing a distribution fit to training data, it can also be used to compare the difference in fit effectiveness for two different sub-samples of the same dataset. In the authentic real data sample that was provided, there are 8 full and 38 partial repair cycles available. If only half of the data is used to generate representative distributions it arises a question of how well the sub-sample represents the true dataset. In this scenario, a QQ plot comparing the fit lines to the data for distributions developed on the full dataset and sub-sampled training set can be compared for similarities. In Figure 27, a QQ plot for the initial degradation parameter datasets of partial repair cycles is shown. The distributions (in red) are very similar for the full and sub-sampled datasets. There is some separation towards the lower quantiles due to fewer observations, but overall the sub-sample distribution does a visually acceptable job at representing the data space.

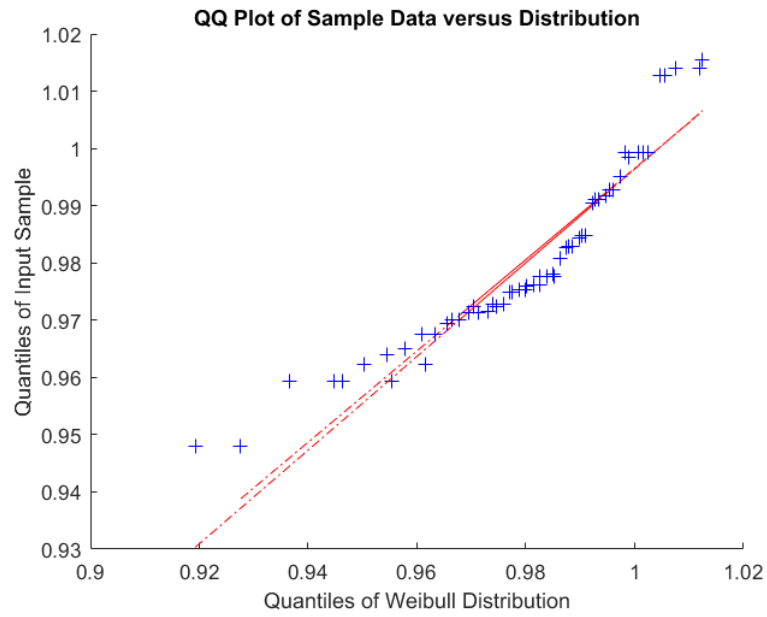


Figure 27: Comparative QQ plot for partial repair initial degradation

The QQ plot is effective at visualizing the similarity for partial repair distributions, but is less effective for the same information (initial degradation parameter) using the full repair data. This effect is shown in Figure 28. Compared to the results for partial repair, the QQ plot in Figure 28 shows differences in the distributions (in red) that are significantly larger. This effect is likely due to the lack of data available in the sub-sample dataset for full repair (4 cycles). This lack of data becomes apparent in the sub-sample specific QQ plot shown in Figure 29. With only four available data points, the lack of information becomes apparent, resulting in a weak definition of a valid representative distribution with significant uncertainty present. Due to its limitations as a visual test, the QQ plot is not effective for small sample datasets. This is primary motive for utilizing the second method of verifying the data generation integrity, the Wilcoxon rank-sum test, which excels at hypothesis testing for small sample datasets [77].

The Wilcoxon-rank sum test, also referred to as the Mann-Whitney U test, is a non-parametric hypothesis test used to compare data from independent samples and determine if they have uniform means [78]. A more widely known statistical test, the two-sample t-test, is similar, but requires that the samples are from normal distributions. This requirement is not necessary, and together with its effectiveness for small sample datasets is why the Wilcoxon rank-sum test was chosen for this case study. Manipulating the rank-sum test statistics allows for definition of the null hypothesis where rejection corresponds to two distributions with similar shapes and offset mean values. Referring to Figure 25, the distributions for each repair action initial degradation parameter are similar, but the mean value for partial repair cycles is generally lower than that of full repair. In this case, the Wilcoxon rank-sum test is applied to the data in order to determine whether or not each maintenance action belongs in its own unique distribution, or if the dataset has insufficient evidence to explicitly characterize each sample as independent. The test statistic hypothesis criterion for this application is provided by Equation 6.

$$\begin{aligned} H_0 : pdf_f(x) &= pdf_p(x) \\ H_1 : pdf_f(x) &\geq pdf_p(x) \end{aligned} \quad \text{Eq. 6}$$

where pdf is the distribution for full repair (f) and partial repair (p). When applied to the sample dataset for the initial degradation parameter, the null hypothesis is rejected; this means that the initial degradation data for full and partial repair belong in independent distributions and do not have a statistically equal mean value. More information on specifics of the Wilcoxon rank-sum statistic is given in [78]. To reiterate, the Wilcoxon rank-sum test is applied to the sub-sample of the real dataset. To solve the issues mentioned above and related to the QQ plots, a similar rank-sum test can be applied to the sub-sample and full datasets. With this test, the goal is to validate that the sample belongs to a distribution with a statistically equivalent mean value, thus confirming the acceptable use of the sub-sample for generating augmented cycles. This translates to a failure to reject the null hypothesis, meaning that there is not statistical evidence that the two datasets belong in unique distributions. The results of applying the rank-sum test in this manner were a failure to reject the null for both maintenance actions. The results of all three tests are clearly outlined below.

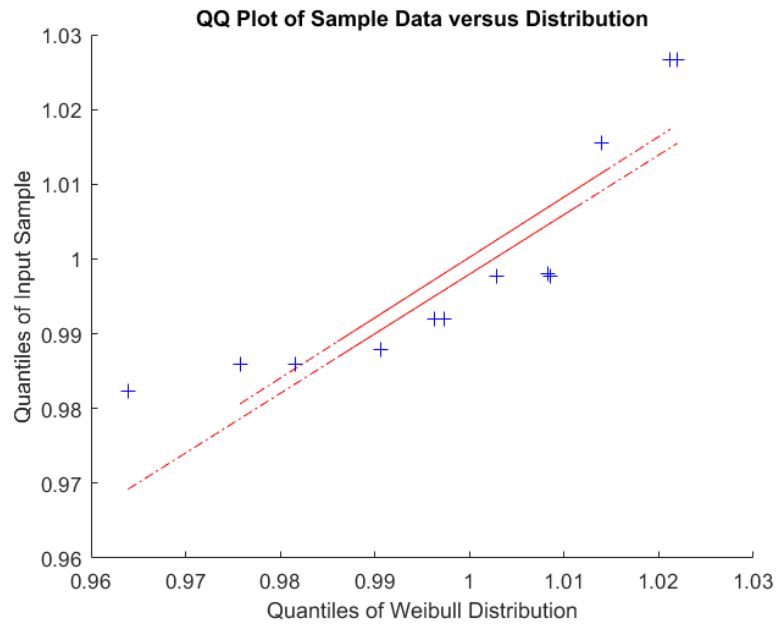


Figure 28: Comparative QQ plot for full repair initial degradation

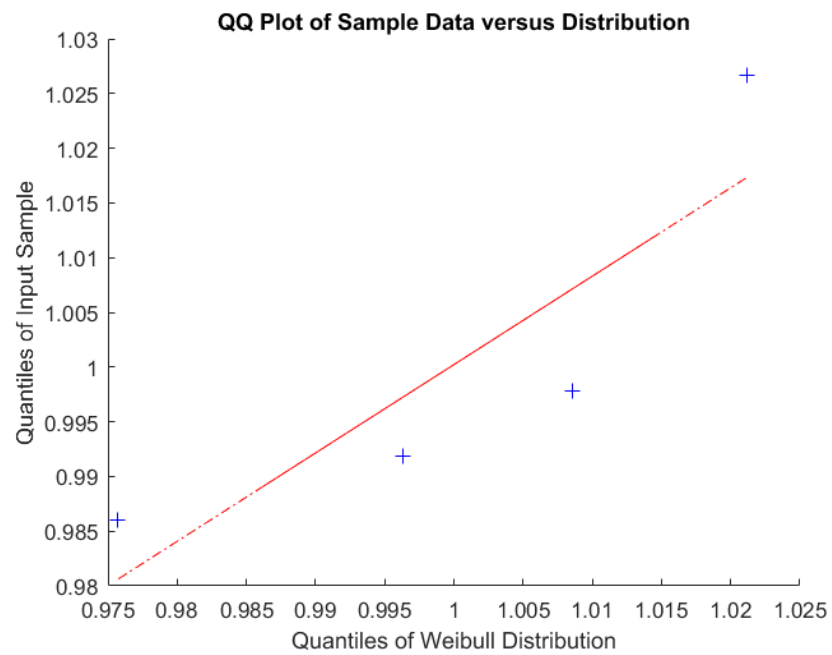


Figure 29: QQ plot for full repair sub-sample dataset initial degradation

<u>Test</u>	<u>Outcome</u>
Full vs Partial Repair	Reject Null
Full Repair Sub-Sample vs Full Dataset	Fail to Reject Null
Partial Repair Sub-Sample vs Full Dataset	Fail to Reject Null

It should be reiterated that the aforementioned methods are all applied to the initial degradation parameter due to its dependence on maintenance action. Since degradation rate and onset of degradation are not significantly influenced by maintenance for the observed real data, these parameters are sampled from maintenance independent distributions.

4.1.3. Secondary Case Study Data Generation Tactics

The secondary case study data generation follows a process similar to the augmented real data from the primary case study, but is applied to a specific shell and tube counter-flow heat exchanger system. This system is a small-scale accelerated degradation test bed, which can be represented by the diagram in Figure 30 with corresponding signals listed in Table 2. All work presented within this section is published [74].

The test bed system is used to generate data representing a heat exchanger that fails by an ambiguous fouling mechanism. In order to incite fouling, clay powder was added to the hot leg channel every other day during a 14-day cycle. As the clay passes through the heat-exchanger tubes, particles begin to accumulate along the walls, which progressively reduce the heat transfer coefficient between the hot and cold legs. Earlier research was conducted and published that develops prognostic models for this specific heat-exchanger system [35]. While early efforts were successful, the research was not concerned with maintenance dependent modeling or maintenance actions at all. As a result, only one generic maintenance action was conducted for the heat exchanger; this involved draining the heat exchanger, pressure-washing the walls of the hot leg, and replacing the muddied clay-water with clean water. The resulting degradation cycles are all considered to have undergone a form of general repair. Since the heat exchanger was tested over 14 day cycles, few degradation cycles were recorded with acceptable quality for this research. Therefore, additional data for the heat exchanger system is generated from a simulation that expands the existing dataset, and incorporates two distinct maintenance actions within the system that are related to real work done to restore heat-exchangers during maintenance: flush and clean. A flush is a simple repair action where pressurized water is forced through each heat-exchanger tube so that any excess fouling is removed from the hot leg. A more complete form of maintenance is a cleaning, in which the heat exchanger system is disassembled, the inner tubes are mechanically cleaned with wire brushes, and high pressure water is used to rinse any remaining particulates from the system. For the purpose of comparison to previously discussed maintenance activities, flush and clean can be thought of as repair and replace respectively.

To reduce development time and increase focus on maintenance dependent modeling, historical data previously collected from the test bed is used to quickly simulate heat exchanger cycles that resemble the real data. This process is broken into six steps:

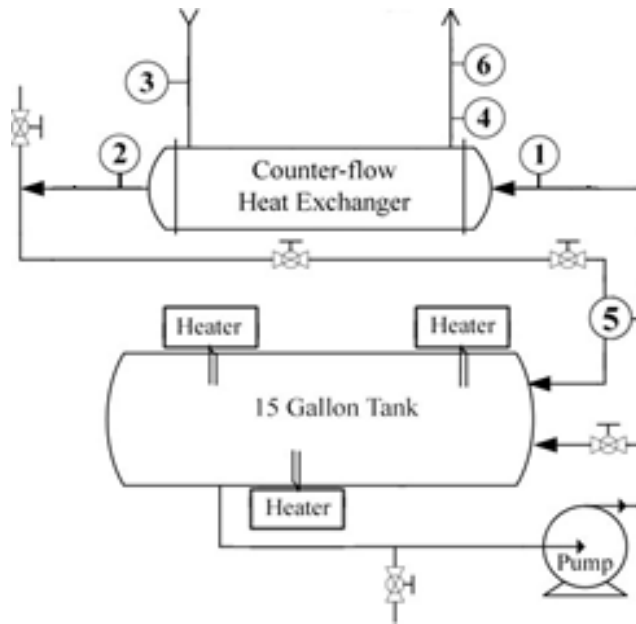


Figure 30: Small-scale accelerated degradation heat-exchanger system [36]

Table 2 - List of heat exchanger signals

Signal Index	Signal/Feature	Abbreviation
1	Hot Leg Inlet Temperature	Thin
2	Hot Leg Outlet Temperature	Thout
3	Cold Leg Inlet Temperature	Tcin
4	Cold Leg Outlet Temperature	Tcout
5	Hot Leg Flow Rate	Mh
6	Cold Leg Flow Rate	Mc

- 1 Sample parameter values from historical data
- 2 Calculate degradation parameters using historical data
- 3 Invoke maintenance influence on cycle degradation parameters
- 4 Generate simulated cycles and add noise
- 5 Predict outlet temperatures using inputs and degradation parameters
- 6 Calculate prognostic parameter and apply arbitrary threshold to all cycles

Similar to the previous case study and research, available real data is processed into distributions for each degradation parameter. Unlike previous research, the heat exchanger has multiple signals available. As a result, the initial degradation values and degradation rate are calculated for each signal. In order to generate additional cycles that resemble the real data as closely as possible, signal noise and the range of signal values through each cycle are additionally recorded. Once distributions are generated for each parameter, the second step is to calculate the heat transfer coefficient of the real data. The primary degradation cause for most heat exchangers is fouling, which reduces heat transfer efficiency over time. This makes the heat transfer an ideal parameter to produce maintenance action influence for the augmented data. The equations for calculating heat transfer for counter-flow heat exchangers are given in Equation 7 and 8 [79].

$$LMTD = \frac{(T_{h1} - T_{c2}) - (T_{h2} - T_{c1})}{\log \left(\frac{T_{h1} - T_{c2}}{T_{h2} - T_{c1}} \right)} \quad \text{Eq. 7}$$

$$U_{h/c} = \frac{\dot{Q}_{h/c}}{LMTD * A} \quad \text{Eq. 8}$$

In the equations, *LMTD* is the Log-Mean Temperature Difference, *T* is the temperature for the hot leg inlet (*h1*), hot leg outlet (*h2*), cold leg inlet (*c1*), and cold leg outlet (*c2*), *Q* is the heat rate, *A* is the surface area of heat transfer, and *U* is the heat transfer coefficient. These features calculated from the real data signals reflect degradation present across each of the signals used to calculate them, making each an ideal degradation parameter or prognostic parameter. An example of the calculated heat transfer coefficient (*U*) is shown in Figure 31.

Parameters for the calculated heat transfer coefficient are sampled and used to develop representative distributions for initial degradation, degradation rate, and noise. While maintenance can influence the time before onset of degradation, the calculated features are too noisy to accurately identify a specific location of onset, and degradation may begin accumulating immediately. As a result, this parameter is not sampled or used to invoke maintenance dependency. In order to produce maintenance action influence on the other parameters, the values sampled from the heat transfer feature are divided into two groups based on value. Parameters can then be sampled from the different groups based on the desired maintenance action; i.e. a larger initial degradation and degradation rate may be sampled for a flush action compared to a clean action. Once the parameters are sampled and a new cycle path is generated, noise is added proportional to the noise present in historical cycles. Examples of augmented cycles for the heat exchanger's heat transfer coefficient are shown in Figure 32.

The process of generating a maintenance dependent heat transfer path could be used as a prognostic parameter for use in model development. However, since the heat transfer path is

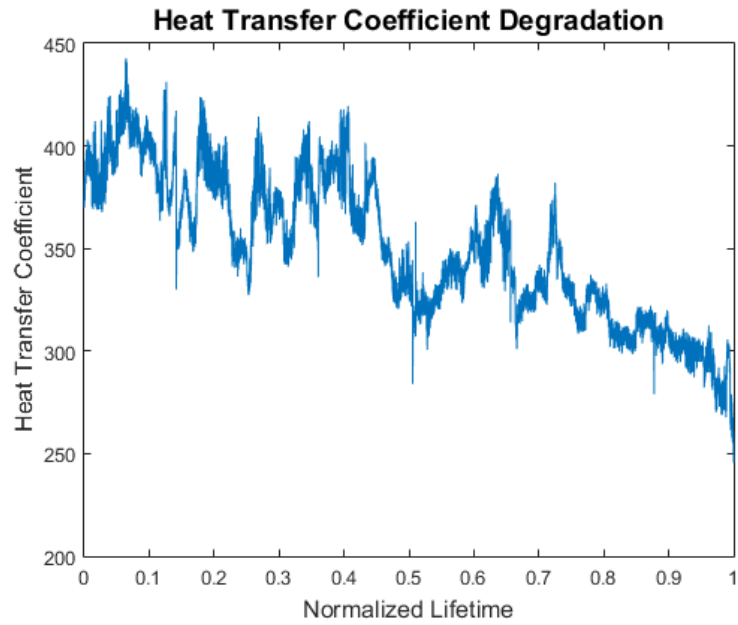


Figure 31: Example of calculated heat transfer coefficient under the influence of fouling degradation

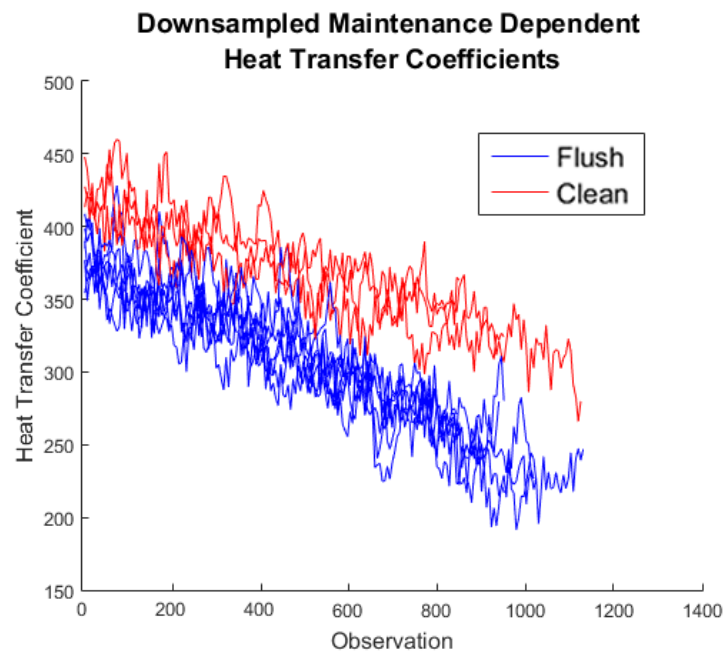


Figure 32: Sampled heat transfer coefficient features

sampled from the distributions and not calculated from the simulated temperature signals, the heat transfer coefficient will not accurately represent any changes seen in the temperature signals. To improve correlation between a prognostic parameter and the simulated signals, the inlet hot and cold leg temperatures and the heat transfer coefficients are passed through a kernel regression model with outlet temperature predictors, as shown in Figure 33. The model in Figure 33 is a multivariate error correction kernel model used to capture the relationships between the input temperature signals and heat transfer features, and the output temperature signals. This type of hetero-associative regression is similar to a General Regression Neural Network (GNRR) [80]. GNRR is non-parametric, making it flexible for many uses. The simplicity of the model also helps increase understanding and distinction between uncertainty in the model and uncertainty in the data, which makes this application of kernel regression extremely useful for exploratory research with specific interest in relative model performances.

For this case study, the kernel regression model captures relationships between the model input and output known as predictor and response variables respectively. This allows typical equipment response to be recreated as a result of the input signal and feature degradation. By training the model on real data, the kernel maps relationships from historical degradation cycles. The model is then evaluated on the generated signals and features, resulting in output temperature signals that retain the real data relationships. In terms of the validity of this method, the input temperatures are not influenced by degradation within the heat exchanger; therefore, it is appropriate to directly generate cycles from the temperature parameters and use them as model inputs. The heat transfer used as model input is influenced by maintenance action, and as a result, the model output predictions reflect the influence of maintenance action on the heat transfer coefficient, as well as the relationships seen in the historical data. At this point, the model output temperature signals and input temperatures could be used to recalculate the heat transfer coefficient for use as a prognostic parameter. However, this would require a simulated or stationary flow rate signal. Instead, the LMTD is calculated from the temperature signals and directly used as the prognostic parameter for model development. This means that the prognostic parameter used to predict the equipment TTF is solely dependent on the generated inlet temperatures and model outlet temperatures, and does not directly utilize the generated heat transfer coefficient or any flow rate signal.

The last step in the data generation process requires a threshold to be applied to the prognostic parameters. Heat exchanger degradation due to fouling does not fail in the traditional sense, which typically results in inability to operate. Instead, fouling causes reduction in heat transfer efficiency and failure is defined as a drop in efficiency below an acceptable level. To mirror this in the augmented dataset, all cycles are normalized to the maximum parameter value so that the range of all cycles is between zero and one. An arbitrary threshold is chosen at 0.45 degradation level, independent of the maintenance action conducted. Any observations past the point of failure are removed from the model and assigned a RUL of zero. Any cycles that do not reach the threshold are labeled as censored cases and treated accordingly. Examples of generated prognostic parameters with the applied failure threshold are shown in Figure 34. Large-scale industrial heat exchangers are operated on long maintenance cycles, which means degradation for these systems is primarily focused on macro-changes on the order of days. This means that the relatively micro changes in the generated prognostic parameters seen in Figure 34 could be appropriately smoothed, as it would not make physical sense for the heat exchanger efficiency to fluctuate as rapidly as it appears.

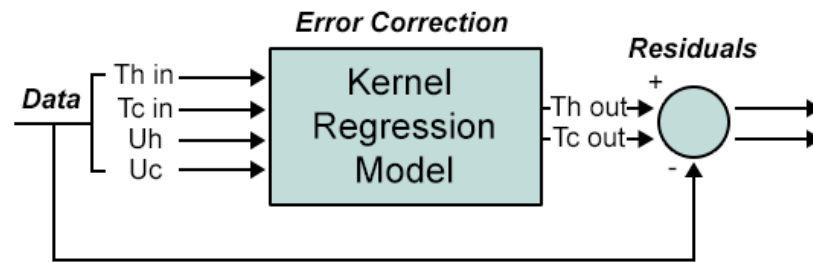


Figure 33: Kernel regression model for heat exchanger system [81]

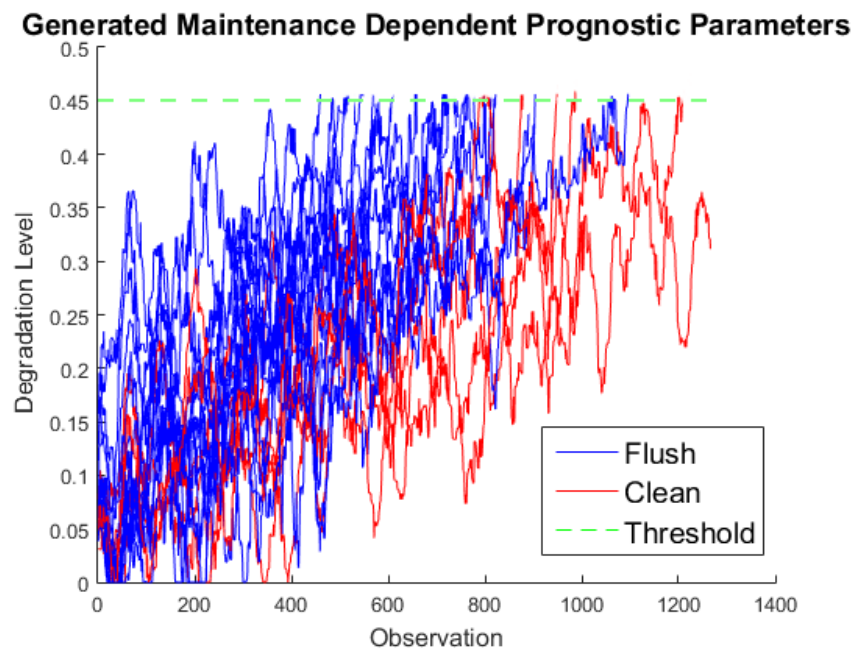


Figure 34: Generated prognostic parameters with applied failure threshold

4.2. Integrating Data Sources

The disparate nature of data sources in a nuclear power plant requires special codes to be developed in order to facilitate failure cycle extraction. The following sections outline a simple algorithm for merging sources together and mining out specific pieces of information from each source to define locations of failures in raw plant data. Failure cycle data is then extracted, categorized, post-processed, and stored within a Prognostic Information Database (PID). The following sections discuss the current coupling algorithm version, and analyze its output. The prototype algorithm is decidedly simplified to allow for research focus beyond data extraction. The goal of the coupling algorithm is to utilize key pieces of information from CMMS records (Table 1) coupled with sensor data stored within the plant computer in order to extract failure cycles. These cycles are key in the development of proactive maintenance models. The current version of the coupling algorithm prototype is depicted in Figure 35. The left side of Figure 35 depicts the overall data flow and I/O for the coupling algorithm. The right side of Figure 35 identifies the basic internal steps in extracting failure cycles. The first module of the coupling algorithm determines the most recent equipment failure from the simulated CMMS (Table 1). Using the extracted CMMS entry for the failure, previous work order entries can be extracted for the same system number and part ID.

In Equation 9, the number of observations for each signal during the failure cycle is calculated using the time between maintenance actions and the sampling rate of the respective sensors.

$$\#obs = f_{dns}(dn_1 - dn_0) * SR \quad \text{Eq. 9}$$

where f_{dns} is a function to convert datenumber objects to seconds, dn_i is the datenumber associated with the i th CMMS entry (Table 1), and SR is the sampling rate for the sensor. When the time of the most recent failure and the number of observations in the failure cycle are both known, the latest failure cycle can be cut from the extracted plant computer data. Out of all the nuclear plant process data, only sensor data related to the failed system needs to be extracted. This requires information from either a detailed plant historian file, or an operator defined file that explicitly relates signals to individual systems. Once the system-sensor relationships are defined, they are stored in a query file. When the CMMS entries are processed and the number of observations is calculated, the system-sensor query tells the algorithm exactly which rows to extract from the raw plant data. There may be one or more sensors that are related to a given system. The resulting failure cycle for the specific system is categorized and stored within the PID, which is a hierarchical database depicted in Figure 36. Data related to each type of prognostic model is the key output from the coupling algorithm, and is stored so that separate models for each system can be generated. The PID is also designed so that historical data can be easily retrieved for future assessment. These design specifications are implemented in order to assist other enabling technologies, outline previously in Figure 10.

The availability of prognostic data from the coupling algorithm makes it possible to compare data across all cycles for related groups within a system. For example, 100 simulated failure cycles grouped by maintenance action are shown in Figure 37. The cycles shown are maintenance dependent. Identifying cycles that were replaced in red and repaired cycles in blue allows for several immediate observations:

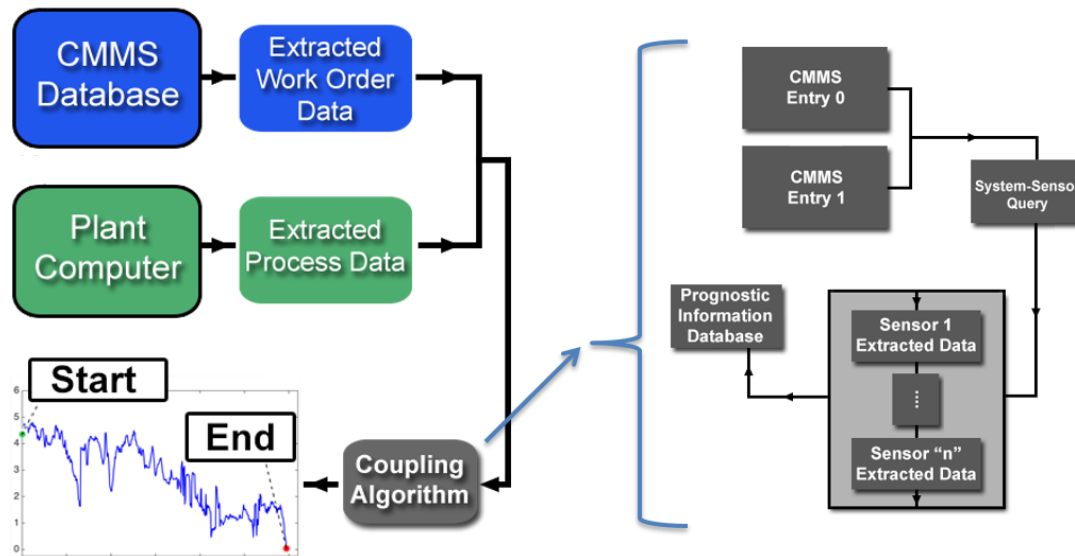


Figure 35: Simplified flow diagram for coupling algorithm design

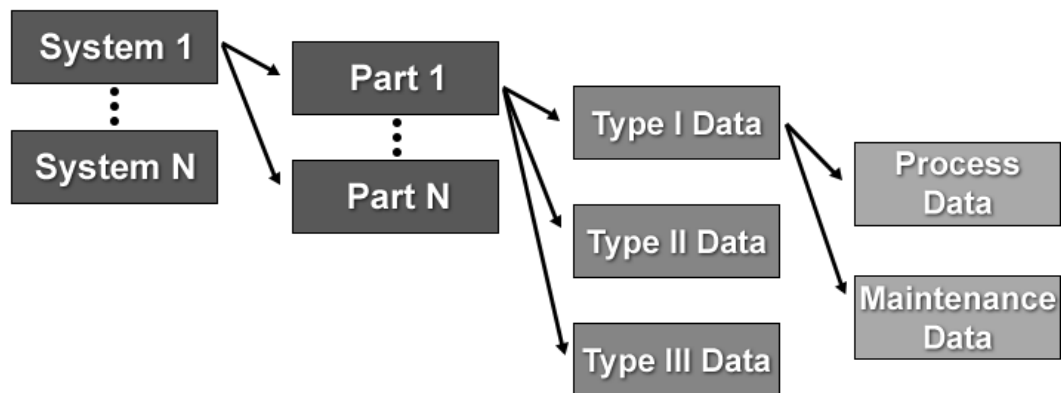


Figure 36: Format of the hierarchical prognostic information database

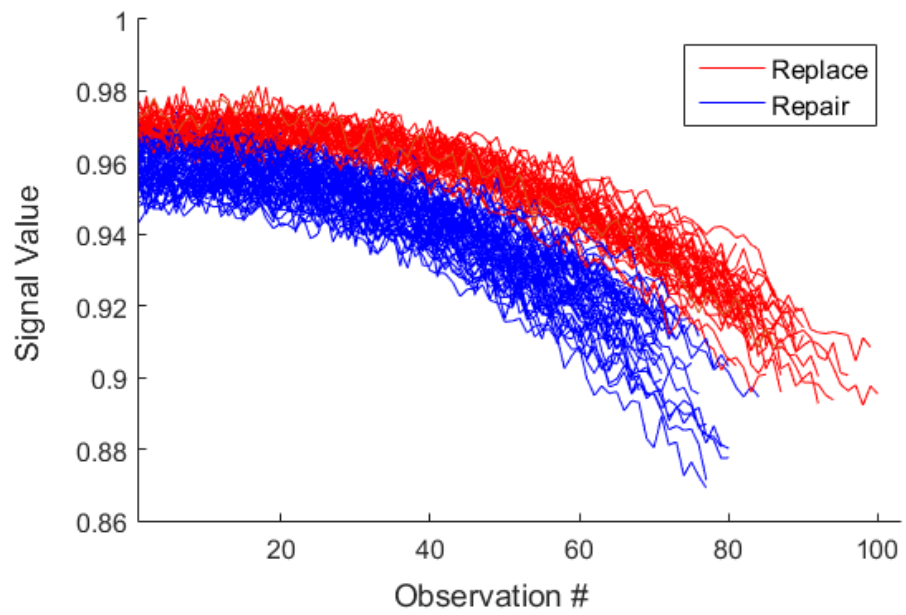


Figure 37: Plot of 100 simulated cycles for a single signal

- Distribution of early signal values for repaired cycles appears to be significantly less than that of cycles after replacement.
- Average failure time for cycles after repair appears to be almost 20 observations less than renewed cycles.
- Rate of degradation appears larger for cycles that have been repaired. The cycles are not as long, yet they reach the same signal value or lower.
- Repaired cycles appear to have more variance in final value than replaced cycles. This is likely a result of all the tweaks made to the data simulation code.

Without any modeling, the coupling algorithm has already provided improved system understanding. More information can be gathered from the coupling algorithm output data, but is not within the scope of this research.

4.3. Prognostic Model Development

One of the original contributions of this research is the development of a framework for the utilization of extracted operating cycles in predictive models. This means creating a prognostic model for the simulated system based on information within the PID in order to investigate the influence of maintenance data on model development.

Referring back to Figure 4, the major steps in developing a system model include analyzing the data, developing a diagnostic model, and developing a prognostic model. These are expanded for the purposes of this proof of concept into the following stages:

- Data processing – evaluate signals as degradation parameters and pre-processing
- Empirical model development – train model on normal behavior and evaluate on failure cycles in order to measure deviation and quantify signal degradation
- Prognostic parameter formulation – combine residuals from empirical model into a single parameter that effectively represents system health
- Prognostic model development – train model on prognostic parameter degradation paths in order to calculate characteristic system degradation path (Eq. 3)
- Failure threshold - determine criterion for cycle termination and use as threshold in failure predictions
- RUL prediction – pass new data to prognostic model in order to predict future degradation path. Path is extrapolated until system degradation passes the critical failure threshold. Subtracting current time from the predicted TOF results in a prediction of RUL.

Outlining these steps explicitly is important because it is possible to incorporate maintenance information into each in some way. One of the simpler ways to utilize maintenance data is in data processing (i.e. cycle extraction, cleaning, and signal evaluation). If the maintenance conducted prior to each cycle is known (or recorded), extracted cycles can be divided by their respective maintenance. Different empirical models can be used depending on the category of maintenance resulting in increased model specificity. If signals present degradation differently for dissimilar maintenance actions, exclusive prognostic parameters can be developed using different groups of signals. Similar to empirical models, separate prognostic models can be developed for different maintenance actions. If prognostic parameters are normalized (resulting in near zero initial degradation), different failure thresholds can be established for different maintenance actions. When predicting the RUL of future cycles, the maintenance dependent models is used, which

results in maintenance specific predictions. In all of these steps, the availability of maintenance information may result in improved predictions.

4.4. Metric Representation of Results

The predicted RUL at each step can be compared to the known value of RUL in order to examine the accuracy of the developed model at predicting failure. The metric used to calculate the prediction error of each model is the Mean Absolute Percentage Error (MAPE) given by Equation 10.

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{p_i - y_i}{y_i} \right| \quad \text{Eq. 10}$$

where p_i is the prediction and y_i is the actual value at observation i for a cycle of length n . MAPE is used to compare the relative performance between competing models, and was chosen due to the fact that it is a well understood extension of the Mean Absolute Error (MAE) metric and is relatively simple. The two groups of competing models within this project are maintenance dependent and maintenance independent models. The traditional assumption is that each cycle undergoes the same renewal process. By developing individual models for each maintenance action, the failure prediction accuracy dependence on maintenance action can be compared to a combined model that is built using all cycles independent of maintenance information. The MAPE metric is used to compare the relative accuracy of each [82].

When comparing the error improvement between competing models, using the difference between the percent errors can cause difficulties in evaluation. For example, if a model performs with a MAPE of 50% and an improved model performs with a reduced error of 20%, the difference is 30%. If the average cycle length is 100 days this translates to a MAE of 50 days for the first model and MAE of 20 days for the second model. The improvement in prediction is 30 days, but the improvement relative to the first prediction is a 60% error reduction, which means that improvements to prediction accuracy is under-represented when looking at improvement in terms of difference in MAPE. This result should be considered when comparing the relative performance in model prediction accuracy.

One important weakness of MAPE defined in Equation 10 is the loss of prediction error information at different stages of life. In prognostics, prediction accuracy often improves as the available observations for model evaluation reach the EOL. This is due to the amount of data available to represent the true degradation path for specific equipment. In these cases, it is more effective to weight the MAPE at each observation based on the normalized amount of true remaining life. An example of a way to calculate this error is shown in Equation 11 where $MAPE_w$ represents the weighted MAPE.

$$MAPE_w = \frac{100}{n * TTF} \sum_{i=1}^n |p_i - y_i| * \frac{y_i}{\sum y_i} \quad \text{Eq. 4}$$

The variables in Equation 11 are the same as in Equation 10; however, in Equation 11 the importance of each prediction is linearly weighted proportional to the observed fractional life (larger importance near EOL). While this method of calculating MAPE may be more appropriate

for prognostic applications, its usage requires a better understanding of the influence that maintenance information has on equipment RUL and prediction accuracy at each stage of life. This research is an attempt to determine if maintenance affects prognostic prediction accuracy in a noteworthy way; therefore, usage of Equation 11 is beyond the appropriate scope of this study and Equation 10 is used for simplicity.

4.5. Summary

This chapter discussed the methods used to generate operating cycle data for use in predictive maintenance models. For each of the studies, the specifics on how process and maintenance data were presented. A framework for the integration of process and maintenance data sources is given for applications where the data from both disparate sources is available. A brief framework for Type III model development is presented to simplify further descriptions in subsequent sections. Lastly, the metric used to evaluate model accuracy is described.

The following chapter presents each study with a description of the specific prognostic model development strategy and model results. The chapter introduction discusses the major differences between each study.

5. Applications and Results

The studies in this research are focused on supporting the theory that utilization of maintenance information can improve predictive model capabilities. To substantiate this theory three studies are considered. A brief description of each is given:

Proof of Concept Study: Development of Type III prognostic models using either maintenance specific or maintenance independent training data. Simulated degradation cycles are influenced by maintenance in three ways: initial degradation levels, degradation rate, and time to onset of degradation. Prediction errors are compared between the models to establish a proof of concept for maintenance dependent prognostic modeling.

Primary Case Study: Development of multiple competing maintenance-based prognostic models for a simulated and real condenser system (augmented data) that undergo two ambiguous categories of maintenance action. Degradation within the simulated system is affected in three ways: initial degradation levels, degradation rate, and time to onset of degradation. Different simulation and modeling techniques are applied to evaluate how maintenance information influences prediction accuracy under different circumstances. Degradation within the real system is primarily affected by maintenance in only one way (initial degradation levels). The real data is used to compare the results from the simulated data to a sub-optimal system (with less maintenance dependence on parameters).

Secondary Case Study: Evaluation of maintenance dependent prognostic modeling for a small-scale heat exchanger that undergoes two explicit types of maintenance. Degradation is significantly impacted by maintenance in two principal ways: initial degradation levels and degradation rate. The data generated for this study is developed using real data from a test bed. Maintenance action influence is produced by manipulating degradation parameters and generating realistic cycles. Focus of this study is on data realism and expected prediction improvements for future implementation of maintenance-based modeling methods.

While each study utilizes some form of data generation, it is important to note the differences between each, and why both case studies are necessary. The proof of concept focuses on a basic data simulation with significant maintenance action influence to establish the ability of maintenance information to improve prediction accuracy. The primary case study focuses on significantly altering degradation parameters to invoke maintenance influence. While this can reveal the potential capacity of improvement for maintenance dependent methods, the validity of assumptions used to simulate the data is questionable, which is why a higher realism dataset is also used in the study. However, in the real system degradation is only marginally affected by the different maintenance actions. To this end, the need for a dataset with a larger dependency on maintenance is needed to validate results. The application in the second case study is a previous developed accelerated degradation test best used for prognostic model development. While maintenance information for the system was not recorded, different degrees of maintenance activity were applied to the system each cycle, which allows the study to utilize data generated to represent the real data with maintenance influence on multiple parameters of the degradation

data. The data used in maintenance dependent modeling for the secondary case study is considerably more realistic than the data used in the primary case study.

In terms of model development methodology, the case studies differ significantly. The proof of concept models are focused on applying the framework provided in Section 4.3. The primary case study is focused on investigating competing modeling strategies and prediction improvement for simulated data with varying factors including available training cycles and TTF distribution overlap. Type III modeling strategies are evaluated by comparing the influence of Bayesian methods on prediction improvement. In contrast, the secondary case study utilizes a very simple prognostic model strategy to make conservative conclusions about the expected improvement when utilizing maintenance dependent models. More information on the methods used for each study are given in the following sections.

5.1. Proof of Concept Study

5.1.1. Prognostic Model Development

For this study, the simulated system is very basic; therefore, both signals are used in model development. The first 10% of recorded data is cut from each cycle and used to train an Auto-Associative Kernel Regression (AAKR) model. It is assumed that this data is prior to any onset of degradation within the system. A representation of the AAKR model is given in Figure 38. The AAKR is used to measure deviations in data from expected behavior. By training the AAKR using healthy data, the model is able to compare differences between data passed to the model and the expected values. This form of error correction model allows residuals to be calculated as the difference between the expected value and predicted value. Residuals are merged into a single prognostic parameter using the method of Ordinary Least Squares (OLS). This method uses the OLS estimator (Equation 12) to calculate the coefficients β of a general regression model [83].

$$\beta = (X^T X)^{-1} X^T y \quad \text{Eq. 12}$$

where X is an $n \times k$ vector of observations for k degradation parameters, and y is an $n \times 1$ vector of independent observations for the prognostic parameter. This prognostic parameter is used to represent overall degradation for the system. In order to develop a parametric representation of system degradation, a GPM is used as the prognostic model. By training the GPM on the data set using a second-order polynomial fit, a general path can be applied to additional failure cycles. To properly train the GPM, cycles must be event terminated (by failure or performance). In order to make sure that the cycles in the simulated dataset are event terminated, an arbitrary threshold is applied to the prognostic parameters at a value of 0.83 (at which point the cycle is considered failed). Additional cycles are simulated and used as test data. The first 10% of each cycle is passed to the GPM, which extrapolated a path to the failure threshold. Using the predicted TTF from the GPM output, an estimate of the RUL is predicted, and the model is updated with an additional observation. The cycle of extrapolation and prediction continues until all cycle observations are added to the model. This process of applying the GPM for prognostic applications, as well as the AAKR model discussed previously, is outlined in [32]. The predicted RUL at each step can be compared to the known value of RUL in order to examine the accuracy of the developed model at predicting failure.

It is important to mention that the models and methods used to calculate subsequent results were not expertly crafted and optimized for the simulated system. Every step is based on

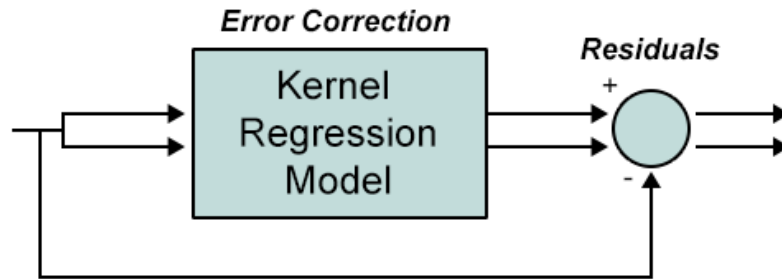


Figure 38: Auto-Associative Kernel Regression model flow diagram (adapted from [81])

common practice and default model parameters. There is likely a more effective way to model the system; however, results are more easily understandable with a less precise model development strategy. Another important note is that the results of these efforts are only meant to serve as a proof of concept. The goal in developing competing prognostic models is to identify the influence of maintenance action on system degradation, and how maintenance dependent predictive models perform relative to pure-renewal models.

5.1.2. Prognostic Model Results

Each stage in model development is done in two ways: maintenance independent modeling where all cycles are in the same group and maintenance dependent modeling where cycles are categorized based on maintenance action. For maintenance dependent modeling, each maintenance action is modeled as a separate stressor condition. The results of maintenance dependent empirical model development and prognostic parameter development are shown in Figure 39. The figure shows that each cycle begins at zero degradation due to the empirical modeling used. This effect causes information related to the initial degradation to be lost. Future methods of maintenance dependent modeling do not use this empirical modeling technique to prevent loss of initial degradation information in the degradation parameters. The differences in TTF and degradation rate for the two maintenance actions are still evident from the figure.

The first test of model performance is the MAPE for RUL predictions using competing maintenance dependent (repair and replace) and independent (combined) models. The results using 50 simulated cycles are given in Table 3. From the results, it appears that both maintenance dependent models have a significantly lower prediction error compared to the maintenance independent model. This is expected because the combined model attempts to represent both data sets simultaneously, resulting in reduced prediction precision and increased uncertainty. The next step is to rerun the simulation with an additional 250 failure cycles. The increased population should improve sampling statistics and reduce model uncertainties. The results for 300 simulated cycles are given in Table 4.

Even with good population statistics, the maintenance dependent models significantly outperform the combined model when applied to the simulated data set. Since little expert knowledge was needed in the development of prognostic models used to obtain these results, it would be expected that these models would not perform well. In fact, the traditional maintenance independent models (combined) perform rather poorly; however, the maintenance dependent models perform almost within accuracy limits of expert models, reaching as low as 5% error. The number of simulated cycles used in the model does not appear to have a severe effect on model performance (3-4%) although additional research should be conducted to determine how lack of data (<50 cycles) affects model performance.

5.2. Primary Validation Case Study

The first case study for validation of maintenance-based prognostic modeling is a comparison of models across simulated and augmented real data for an industrial condenser. In this study, simulated data is generated to resemble a condenser with a high influence from maintenance on degradation parameters. The augmented real dataset is used to compare the results for a condenser system with degradation cycles that are marginally influenced by maintenance. The purpose of this primary case study is to evaluate the extent of prediction improvement associated

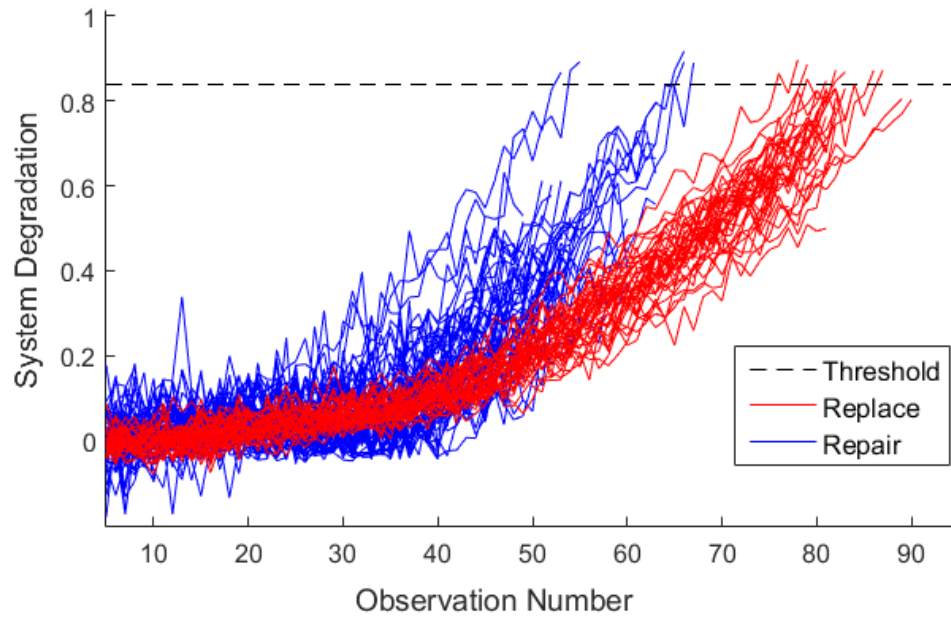


Figure 39: Example of maintenance dependent prognostic parameters with critical failure threshold

Table 3 – Proof of Concept: Prediction errors for 50 simulated cycles using competing maintenance independent and dependent models

	Prediction Error - MAPE (%)		
	Combined	Repair	Replace
50 cycles	20.8	11.3	4.7

Table 4 - Proof of Concept: Prediction errors for 300 simulated cycles using competing maintenance independent and dependent models

	Prediction Error - MAPE (%)		
	Combined	Repair	Replace
300 cycles	16.3	7.7	4.6

with maintenance dependent modeling for similar systems with varying influence of maintenance on degradation levels. The specific data generation methodology was provided in Section 4.1.2.

5.2.1. Prognostic Model Development

Similar to the models in Section 5.1.1, the common assumption is that the use of maintenance information integrated with historical process data will improve overall understanding of system and equipment health. This belief is derived from the result of transitioning between prognostic model types. In prognostic modeling, the accuracy of developed models is often determined by the amount of information available for a given system. Type III models contain data for equipment-related sensors, operating conditions (direct records or derived from process data), and TTF data for past failures (including censored cycles). Type III outperform Type II models for comparable application since Type II models do not typically contain real-time or analogous sensor data. The same is true when comparing Type II models to Type I models, where Type I models lack the knowledge of operating conditions present in Type II models. In general, the trend is that model performance typically increases as more information about the system (related to its degradation) is utilized [13] [84]. Considering maintenance dependent modeling, any information related to how repairable equipment is restored post-failure allows for more specific model design, which will theoretically improve prediction accuracy and/or reduce model-related uncertainties. In this study, maintenance dependent models are developed for both the simulated and augmented data sets. The performance of these models is compared to that of maintenance independent models (benchmark models) to identify the improvement obtained through maintenance data integration.

5.2.1.1. Simulated Condenser Data Modeling

The primary case study is mainly focused on determining the extent of prognostic model improvement (in terms of prediction accuracy) when maintenance data is used to improve model specificity. To test this theory on the simulated dataset Type I and Type III models are used. To test maintenance influence on Type I model prediction accuracy, four different Weibull models are used: two maintenance independent models (combined maintenance actions), one maintenance dependent replacement model, and one maintenance dependent repair model. The combined maintenance models are used as a benchmark to compare the results of maintenance dependent modeling to historically traditional maintenance independent modeling. Since the simulated data is easy to manipulate, two additional methods of parameter sampling are used so that resulting TTF distributions are related in one of two possible ways:

Unique TTF: TTF distributions are distinct and separate; i.e. the largest repair TTF is smaller than that of the smallest replacement TTF. This means that a replacement action will **always** result in a longer expected life than if repaired.

Overlapping TTF: TTF distributions for each maintenance action share inner-extreme values; i.e. the largest repair TTF may be similar to the smallest replacement TTF. This means that we are not certain whether replacement is always qualitatively better than a repair (or significantly better).

The example given in Figure 22 shows an example of overlapping TTF distributions. While this relationship is usually more realistic for repairable equipment than unique TTF distributions, the

method of sampling both allows testing of the diminishing return on performance, depending on the extent to which maintenance influences failure times.

In the case of all four models, a Weibull fit is trained by minimizing the model error in order to calculate the best choice of scale and shape parameters. These optimized parameters are saved for each simulated data set. In order to reduce simulation uncertainty, the same training dataset is used for maintenance independent and dependent models. The model accuracies are also aggregated across 10 independent simulations to reduce uncertainty.

Prognostic parameters are designed to represent system health by quantifying the overall degradation of equipment based on measured/recorded data. For the simulated portion of this case study, the prognostic parameters are generated by normalizing the degradation parameter so that the limits of the prognostic parameters are all within 0 and 1.0. An arbitrary threshold must be applied so that the time-terminated cycles (of arbitrary length) can be transformed to event terminated, which considers equipment to be failed when the prognostic parameter passes the threshold. For the case study, a hard threshold is applied at 84% degraded (e.g. 0.84), which was chosen based on the cumulative number of cycle failures at the threshold value. This is applied to all cycles regardless of the maintenance action conducted, and any observations occurring after a parameter has passed the threshold are considered to have a RUL of zero.

Once the prognostic parameters are generated and the threshold is applied, predictions of RUL at each observation during the cycles are calculated. To estimate the RUL, a unit-specific General Path Model (GPM) [30] is applied, which builds a representative functional fit for the specific equipment's prognostic parameters. By applying the GPM fit to a current cycle, the path can be extrapolated to the failure threshold and used to calculate the RUL prediction, which is defined as the difference between the predicted TTF and the current observation [32]. In order to optimize the models for the case study, training data is chosen by random sampling 80% of failure cycles. The GPM is then trained on those cycles only.

One common problem associated with using GPM is the mischaracterization of degradation path in early life. Due to lack of data, noise, etc., extrapolated paths may poorly represent the true degradation trend, and result in poor RUL predictions or infinite predictions if the path never passes the threshold. An example of this effect is visualized in Figure 40, which shows an extrapolated path that follows the wrong direction due to lack of observations in the model. In this case, the path will never reach the threshold and will produce inaccurate prediction results. To avoid this issue and reduce error in early life predictions (before sufficient data is observed), Bayesian methods can be used to weight predictions towards prior information from within the GPM [32]. Previous research work conducted by Welz et al. [36] uses Bayesian priors to improve heat exchanger modeling, resulting in significant reduction of early-life prediction error. Utilization of Bayesian methods allows for comparison between relative performance with and without Bayesian updating for maintenance dependent and independent models. This paints a picture of how well maintenance information improves prediction accuracy with and without preexisting prediction improvements. For more information on Bayesian information in prognostics, please refer to [10].

5.2.1.2. Augmented Condenser Data Modeling

The purpose of applying maintenance dependent methods to the augmented data is purely to validate the proof of concept established by the simulated models. As a result, Bayesian

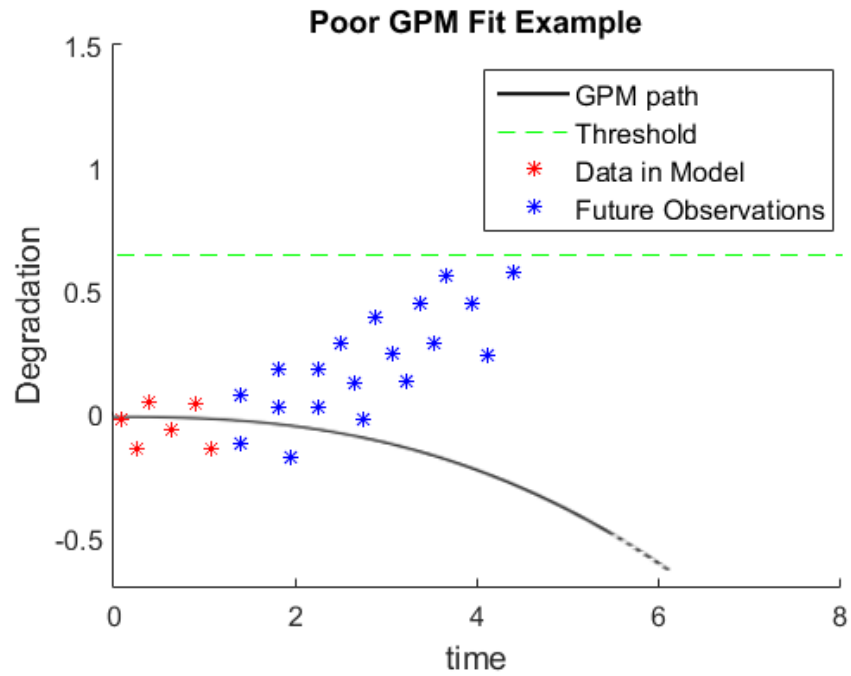


Figure 40: Example of poor GPM extrapolated path

methods are not applied to the augmented data; otherwise, the same methodology for modeling the simulated dataset is used for the augmented data. This includes passing the normalized degradation parameter as the prognostic parameter to the GPM. Since there is little influence of maintenance on the real data operating parameters, there is no expectation that the augmented data modeling will produce outcomes as notable as the simulated data models.

5.2.2. Primary Case Study Model Results

As a reminder, the primary case study involves a simulated and augmented dataset for a condenser system that is only directly influenced by maintenance through varying initial degradation levels. The following sections are first divided by the category of prognostic model (Type I and III), and secondly by the dataset used (simulated and augmented).

5.2.2.1. Type I Model Analysis – Simulated Data

Referring back to 5.1.1, the simulated data for the primary case study was designed so that the TTF distributions for the study maintenance actions (repair and replace) would be unique, or contain overlapping values. Developed Type I distributions are modeled after traditional reliability models, and are fitted with optimized Weibull distributions. These developed models have resultant uncertainties in the model parameters reported as standard deviation (u). The uncertainties for the combined maintenance model and maintenance action dependent models are compared in Table 5 for both overlapping and unique TTF distribution simulations. Comparing the parameter uncertainties of the maintenance independent model (combined) to each of the maintenance dependent models (repair and replace), shows that uncertainty is reduced by almost 50% when maintenance information is utilized for both degrees of effect on TTF distribution overlap. This result is substantial because the uncertainty in Type I model parameters can propagate into additional model results where Type I information is used *a priori*, such as Bayesian methods for Type III prognostic models.

Predictions for the Type I models are different for each cycle. In order to quantify the average performance of each, multiple cycles' TTF are predicted and aggregated to form a singular error metric. The resultant MAPE and uncertainty (u , *standard deviation*) for each model are provided in Table 6 with uncertainty units being in units of observations. The results show that prediction error is marginally reduced for systems where maintenance information is utilized (0-4% approx.). The result is a small amount of additional equipment lifetime for risk mitigation and decision making compared to traditional prognostic modeling efforts. For an average equipment cycle of around 70 days, this would provide up to 3.5 days of additional time for risk mitigation. Regarding the affect maintenance dependent modeling has on uncertainty, for both unique and combined TTF distributions uncertainty in predictions is reduced by more than 30% and as much as 60%. This shows a substantial increase in prediction confidence when maintenance information is used to develop models with increased specificity.

5.2.2.2. Type I Model Analysis – Augmented Data

For the primary case study augmented data set, the average TTF of the augmented cycles is on the order of 100 observations. Due to large variance in the augmented dataset, all reported uncertainty values are standardized to the combined maintenance model and should only be used as a comparative metric between the maintenance independent and dependent models. The

Table 5 – Primary Case Study: Type I model parameter uncertainties in observations (# obs.) for competing models using simulated data

	u Combined (# obs.)	u Repair (# obs.)	u Replace (# obs.)
Unique TTF	4.38	2.86	2.25
Overlapping TTF	5.93	3.10	2.96

Table 6 - Primary Case Study: Type I prediction errors (MAPE) and uncertainty (u , standard deviation) in number of observations (# obs.) for competing models using simulated data

	MAPE (%) Combined	MAPE (%) Repair	MAPE (%) Replace	u Combined (# obs.)	u Repair (# obs.)	u Replace (# obs.)
Unique TTF	7.8	6.2	4.2	53.3	16.5	20.2
Overlapping TTF	6.5	6.6	4.8	41.2	27.8	13.8

results of the Type I models for the augmented data are given in Table 7. These results show that maintenance independent modeling results in an error of over 80%, but the maintenance dependent models are able to predict TTF with less than 60% error. This is significantly more than that of the simulated data, however the capacity for improvement is much larger due to higher prediction error. Similar to the results for the simulated data, uncertainty in predictions is significantly reduced when maintenance information is used to develop dependent models; for the augmented data, this uncertainty reduction is more than 30%. It is important to mention that, while maintenance dependent modeling improves performance, the uncertainties and prediction errors are relatively large compared to typical prognostic models. This is likely a result of the quality and quantity of available data. It is plausible that, in systems where sufficient data is collected for the purpose of maintenance dependent modeling, better prediction results can be expected, similar to the model accuracy results shown in Table 6.

5.2.2.3. Type III Model Analysis – Simulated Data

Type III results for the simulated data are divided into two focuses: Bayesian methods, and no Bayesian prior information. For the Type III model results, simulated data with unique and overlapping TTF distributions are used, but the results given with the use of Bayesian priors only utilize overlapping TTF data to show expected outcomes for more realistic applications. The Type III model results without any use of Bayesian priors are given in Table 8. These results indicate that, for Type III modeling, models that utilize maintenance information see a reduction in prediction error of more than 5% for all cases. For the dataset with overlapped TTF distributions, the MAPE is reduced from 18.3% to 8.9% and 2.5% for the maintenance dependent models. The error reduction for repair was approximately 9%, which corresponds to an accuracy improvement of over 6 days. The comparison between prediction errors for unique and overlapping TTF distributions show the differences in the degree of maintenance dependent model improvement, which highlight the diversity in model improvement for systems in which maintenance affects model parameters to varied degrees. This agrees with the results of the Type I models, in which the overlapping TTF distributions typically result in larger prediction errors and uncertainties than unique distributions.

Bayesian methods are also used; existing research has shown that weighting predictions with Bayesian priors can result in improved early to mid-life prediction for applications with high uncertainty. To compare the results of Type III modeling with and without Bayesian methods, only data cycles developed with overlapping TTF distributions are used. The results of the developed Type III models are given in Table 9. The results of utilizing prior information for Bayesian weighting show an error reduction of more than 9% and as much as 12% for both maintenance dependent models compared to the maintenance independent model. This result was unexpected due to the fact that Bayesian methods incorporating prior information are used for their improvement in early life predictions, which account for a majority of prediction uncertainty and error. It was expected that the use of Bayesian methods would reduce the capacity for improvement when utilizing maintenance dependent models. However, the results show that there is little or no diminishing return when using both Bayesian methods and maintenance dependent modeling. This means that models that achieve high prediction accuracy may still be candidates for error reduction utilizing maintenance dependent methods, further validating the need to incorporate maintenance information in predictive modeling.

Table 7 – Primary Case Study: Augmented data Type I model prediction error (MAPE) and standardized uncertainty (u , standard deviation)

	MAPE (%) Combined	MAPE (%) Repair	MAPE (%) Replace	u Combined	u Repair	u Replace
Type I Model	81.1	57.2	43.9	12.1	8.4	6.1

Table 8 - Primary Case Study: Type III prediction errors w/o Bayesian priors for simulated dataset

	MAPE (%) Combined	MAPE (%) Repair	MAPE (%) Replace
Unique TTF	11.8	6.7	0.5
Overlapping TTF	18.3	8.9	2.5

Table 9 - Primary Case Study: Type III model prediction errors with and without Bayesian prior information for simulated dataset

	MAPE (%) Combined	MAPE (%) Repair	MAPE (%) Replace
w/o Bayes	18.3	8.9	2.5
w/ Bayes	14.9	4.3	2.0

5.2.2.4. Type III Model Analysis – Augmented Data

Models developed on the augmented data are not used to develop proof of concepts or test the capacity for model improvement by utilizing maintenance data in prognostic models. Instead, the results of testing maintenance dependent modeling on the Type III models trained using augmented data are used to validate the proof of concept established with the simulated data that maintenance information can be utilized to improve upon existing state-of-the-art prognostic modeling methods. This includes systems that are sub-optimal, or applications where equipment degradation is weakly influenced by the specific maintenance conducted on the system. The Type III models developed for the augmented data do not utilize Bayesian methods and contain large amounts of noise and cycle variance. Results of Type III models are given in Table 10. As expected, the results show that the model prediction accuracy is improved by more than 5% when maintenance dependent modeling is used; however, the prediction errors are significantly higher than those for the simulated dataset. This is the anticipated outcome due to the noise and uncertainty in the augmented data. One noteworthy result is that the error reduction from the augmented data Type I models to Type III models is over 15%. This shows that there is improvement in increased prognostic model specificity as well as improvement when utilizing maintenance information within the same category of prognostic model. It is expected that for systems with larger maintenance action influence on degradation parameters, increased prediction improvement may be possible in real data applications.

5.3. Secondary Validation Case Study

The second case study for validation of maintenance-based prognostic modeling is a comparison of models across augmented real data for a small-scale accelerated degradation heat exchanger. In this study, augmented data is generated to resemble a heat exchanger fouling degradation with maintenance action influence on initial degradation levels and degradation rates. The purpose of this secondary case study is to assess the improvement of maintenance-based prognostic modeling for realistic data. The specific data generation methodology was provided in Section 4.1.3.

5.3.1. Prognostic Model Development

Modeling for the secondary case study is very similar to the primary case study; the key difference between each is the degree of maintenance influence on cycle factors. As a result, methodology of model development is almost identical. For the heat exchanger system, two categories of prognostic model are developed to evaluate the model performance dependency on maintenance. The first category is a Type I Weibull model, which is typically used to predict TTF for average equipment in average conditions. This is similar to the Weibull fit used in the primary case study. The second prognostic model is the GPM, and belongs to the Type III model category. Again, this is similar to the GPM used for the primary case study, and attempts to build a representative degradation path for the prognostic parameters of historic cycles. This type of model has a higher degree of specificity compared to the Type I Weibull model, and is able to predict TTF for specific heat exchanger equipment under specific conditions. For this specific application, a quadratic GPM fit was used to represent the augmented failure cycles. Again, similar to the primary case study, explicit Type II models are not developed.

Table 10 - Primary Case Study: Type III model prediction errors for augmented dataset

	MAPE (%) Combined	MAPE (%) Partial Repair	MAPE (%) Full Repair
Type III Model	46.2	41.7	26.1

For the secondary case study, three models are developed for each prognostic model category, resulting in six total models. To establish a baseline prediction accuracy, maintenance is not considered and a maintenance independent model is developed that utilizes all failure cycles. This combined maintenance model is used to compare relative performance to the maintenance dependent models. In this case study, only two types of maintenance action are possible; therefore, two maintenance dependent models are used: one for the flush maintenance actions, and one for full cleaning maintenance. While it does not make sense to compare the prediction accuracy between the maintenance dependent models, each can be compared independently to the maintenance independent baseline to assess relative prediction performance improvement.

For all six models, less than half of the augmented cycles are used to train each model, while the remaining cycles are used to evaluate the prediction accuracy. The same MAPE metric that was used in both the research and primary case study is used for this application; see Equation 10. Again, while other metrics for prognostic predictive performance may be used [85] [82], this simple measure is sufficient.

5.3.2. Secondary Case Study Model Results

Referring back to Section 4.1.3, the secondary case study uses data from a counter-flow heat exchanger. Initial degradation and degradation rate parameters are both influenced by maintenance action, as well as subsequent TTF values. All work presented within this section is published [81].

5.3.2.1. Data Generation

The primary purpose of this case study is to investigate the influence of maintenance dependent modeling; however, with the extent of methods used to generate realistic data, it is important to highlight the results of data generation. Figure 41 gives a comparison of the augmented and simulated data for the input signals and generated heat transfer coefficient features. Comparing the datasets, the resulting generated data values and augmented cycles appear to be similar to the real data. One noteworthy difference is the relationship between inlet temperature signals of the simulated data. The reason for the significant difference is due to random sampling of the signal parameters from their respective distributions. In the real system, hot leg inlet temperature is highly correlated with cold leg inlet temperature due to the system design. In the augmented data, signal parameters are randomly sampled, and this correlation is partially lost. While the parameters could be sampled in a correlated manner, the results without such a method provide a conservative result.

5.3.2.2. Model Results

In order to improve confidence in the Type I model results, multiple batch predictions are run and aggregated into a mean prediction score similar to the Type I model results for the primary case study. The results of the Type I model development for maintenance dependent and independent prognostic modeling are given in Table 11. At first glance, the prediction error for the Type I combined model is much larger for the secondary case study. This is likely due to the large amount of noise in the LMTD parameter used to develop the prognostic models. However, when comparing the results of the maintenance dependent Type I models, results are comparable and

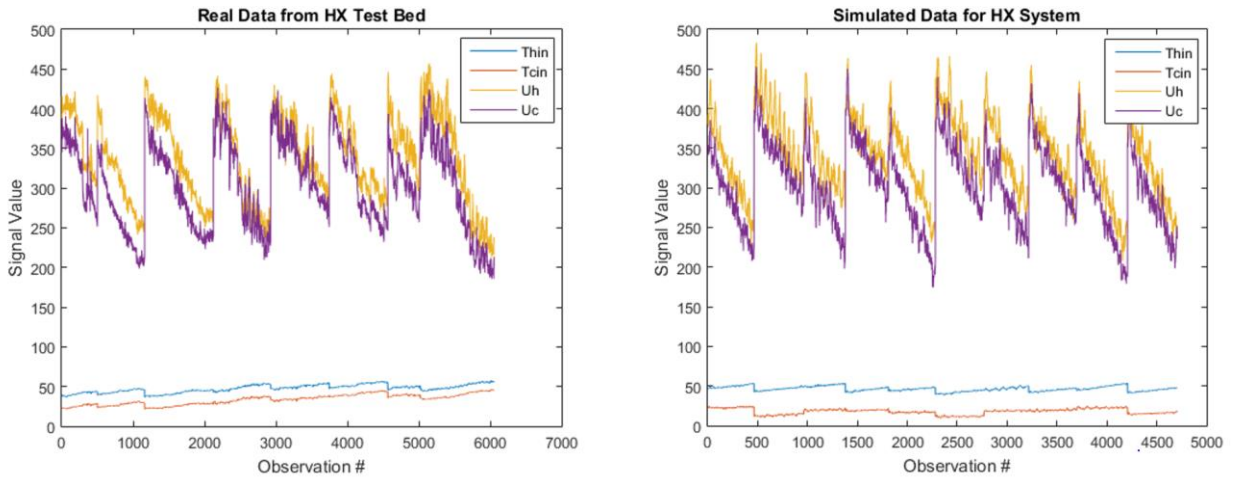


Figure 41: Comparison of real data (left) and augmented data (right) for heat exchanger system

Table 11 - Secondary Case Study: Type I prediction results for competing models (heat exchanger data)

Maintenance Independent	Maintenance Dependent	
MAPE (%) Combined Model	MAPE (%) Flush Model	MAPE (%) Clean Model
96.6	51.3	16.0

in some ways succeed compared to the primary case study results for the augmented condenser data. Comparing the combined model to the maintenance dependent models shows a significant improvement in prediction accuracy, with error reductions from approximately 97% to 51% and 16% for flush and clean models respectively.

Type I models typically experience high uncertainty and poor performance for systems with large variance in failure time data. This is a primary reason that Type III models are necessary for systems with higher complexity. For the heat exchanger system, the results of developed Type III models are given in Table 12. The average cycle length for the heat exchanger system is approximately 500 observations, which results in a prediction error of approximately 36% for the combined maintenance action model. Compared to the Type I combined model, this is a reduction in prediction error of 61%. Comparing the maintenance independent prediction error to those of the maintenance dependent models shows a reduction of more than 13%, with approximate error reductions of 3% and 23% for the flush and clean models respectively. Comparatively, the error reduction for augmented data Type III models in the primary case study was about 5% and 20% for each maintenance action. These results show that the error reduction increases for systems where the degradation of the equipment is increasingly affected by maintenance actions, and are in line with the expected nature of maintenance dependent predictive modeling.

Table 12 - Secondary Case Study: Type III prediction results for competing models (heat exchanger data)

Maintenance Independent	Maintenance Dependent	
MAPE (%) Combined Model	MAPE (%) Flush Model	MAPE (%) Clean Model
35.8	22.5	12.0

6. Conclusions and Recommendations for Future Work

6.1. Conclusions

The world of industry is evolving, and with it, industry and other organizations must adapt and advance towards innovation and enhancement. Across all of commercial business, one of the most important and fastest developing areas is operations and maintenance, especially asset management. Progress in industry has produced advanced data collection, data processing, data analytics, predictive analytics, and decision support technologies with the capability to improve maintenance and operations efforts. The purpose of these innovative methods is to increase the bank of knowledge possessed, related to equipment performance and maintenance management. To this end, in industry two sources of data are the CMMS and CM systems. Either information repository can be used as an effective tool, yet integrating data together from both sources can result in valuable and otherwise inaccessible knowledge related to proactive maintenance efforts and decision support. To support advanced diagnostic and prognostic programs, the need for CMMS implementation and maintenance data utilization in models is validated. Discussion of industry needs for digital maintenance systems to support advanced maintenance strategies, as well as recommended standards for CMMS implementation are provided in this dissertation and support the first major contribution of this work.

Research work is focused on developing supporting algorithms to collect and integrate disparate condition data and maintenance information. A framework for coupling of data sources and mining of failure cycles is presented with the ability to quickly extract historical operating cycles. This algorithm improves the availability of failure cycles for predictive modeling, and offers information related to each cycles maintenance action so that data can be grouped by the degree of repair to the system prior to each operating cycle. This research supports the second contribution of this dissertation, thereby improving the utility of recorded plant data in proactive maintenance applications.

The work conducted in each research study is based on the concept that incorporating operating data from industrial plants with related maintenance data will improve proactive maintenance efforts and predictive maintenance models. Research efforts were used to establish proof-of-principle, and show that utilizing maintenance information to extract failure cycles and predict system RUL results in improvements to failure prediction capabilities. This includes extracting prognostic data from raw historian files, analyzing data for characterization, and developing advanced degradation prediction models. The end-goal of this research was to identify the extent of improvement related to data integration from existing condition data repositories as well as digital maintenance data records, using both simulated data and augmented real data collections. To strategically evaluate prediction improvement through increased data utilization, numerous competing Type I and Type III prognostic models were developed and evaluated for prediction accuracy across unique case study datasets and applications. Type I models were used to establish the ability of maintenance dependent modeling to reduce uncertainty and prediction error compared to traditional maintenance independent methods. For advanced prognostic modeling, i.e. Type III predictive models, prediction improvements resulting from maintenance dependent modeling were compared for standard models as well as Type III models utilizing Bayesian methods. One critical assessment of enhancement associated with maintenance data utilization was to evaluate prediction accuracy improvement on a real industrial dataset with minimal maintenance impact on degradation parameters. This allowed for

comparative analysis between simulated data (high maintenance action to degradation correlation), and augmented real data (large uncertainty between maintenance events and equipment degradation). These models provide a generalized framework for utilizing extracted failure cycles (in similar applications) in-line with the third major contribution of this dissertation.

The research results presented throughout this dissertation show the diverse levels of influence that maintenance events have on prognostic model capabilities depending on equipment conditions and subsequent asset degradation due to specific maintenance actions. While the results of these tests across Type I and Type III models are varied in terms of specific developments, the ability of maintenance dependent modeling to improve prediction capabilities is undisputed across all applications, even when equipment degradation is not considerably impacted in the three major ways: initial degradation levels, degradation rate, and time to onset of degradation. For all case studies, Type I model prediction accuracy was improved by no less than 7% for negligible maintenance impact, and more than 20% for significant maintenance impact. Similarly, for the Type III models, prediction accuracy was improved by more than 45% for simulated data applications, and at least 14% for models developed on augmented data. This improvement was even seen when maintenance dependent models were used to improve upon advanced Bayesian weighted predictions, which further improved prediction accuracy by 25% (45% error reduction w/o Bayes compared to 70% error reduction with Bayes). These case studies and their quantitative results support the fourth contribution of this dissertation.

The original contributions of this work to the field of advanced maintenance strategies can be summarized as follows:

- Development of a comprehensive overview of industrial maintenance practices and the need for improvements to maintenance strategy and supporting technologies.
- Development of a design for extracting operating cycles using process data and maintenance data to support diagnostic and prognostic models.
- Development of a framework for utilizing cycles in predictive maintenance models.
- Implementation of model framework in three applications with varied maintenance action influence to demonstrate prediction improvement.

To put the results of this work into perspective, the nuclear industry spends more than 15 billion dollars (USD) per year on operations and maintenance of nuclear plants (50% of total budget in most cases), and another 5 billion on capital expenses such as replacements and spare inventory. However, most of the industry still relies on traditional preventive or reactive maintenance strategies for asset management. In an industry where so much focus is applied on retaining the highest standards, efficiency, and safety, there is a dire need for change in how the US nuclear industry conducts maintenance. This need for advanced maintenance practices is amplified by the current capabilities of the nuclear industry to retain competitive costs capabilities, in which it is currently failing to the oil and gas industry partially due to high maintenance overhead. A testament to the outdated capabilities of nuclear industry maintenance can be seen in the common use of handwritten paper logs to record maintenance events and failure information. This dissertation includes a formal documentation of the need for digital maintenance records for facilitation of state-of-the-art proactive maintenance efforts in commercial nuclear industry applications. Specific CMMS data requirements for data extraction, failure cycle mining, and maintenance strategy optimization are also discussed. The entirety of results presented validate the theory that by increasing predictive model specificity through integrated process and

maintenance data may result in overall improvements to prognostic model performance, and improve the quality of information available for optimizing maintenance planning.

6.2. Recommendations for Future Work

While the results of this work show that maintenance dependent modeling is constructive in all evaluated case studies, there is still significant research needed to identify and quantify the precise extent of predictive model improvement through maintenance data utilization in industrial applications. One such area of critical research is the influence of maintenance data on predictive capabilities at different stages in equipment life. While the results of this work show that there is improvement in prediction accuracy over the entire life of a system, it is equally important to verify that significant prediction improvement occurs at each stage of life; especially end-of-life when accuracy predictions of failure times are most imperative. This requires a more specific evaluation of prediction errors. One way of implementing this idea is through the division of cycles into periods of life. If prediction accuracies are grouped into two or more stages of life, the error can be averaged for each group and compared. Another solution would be to implement a weighted error metric that uses an additional term to weight prediction importance higher as the cycle approaches failure.

Another chief area of additional research is a study into differences of the degree of maintenance influence on selective cycle parameters, and how these variances affect error reduction in diverse applications. This is essentially a more specific study of the methods presented in this research. With an in-depth strategic manipulation of cycle parameters and maintenance influence, wealthier knowledge related to the explicit influence of each on prediction improvement may be possible. Similarly, it is recommended that future research investigate the diminishing return on prediction improvement by increasing the number of maintenance actions that a system or equipment can undergo. It is expected that the division of failure cycles based on maintenance action will only improve predictions when significant differences between cycles are present. As more maintenance actions are used to categorize cycles, it is expected that prediction improvement will eventually fall of with an inverse relationship to cycle parameter uncertainties. This includes the specification of repeated repair actions (e.g. 1st repair, 2nd repair, etc.).

Lastly, it is recommended that research into other methods of incorporating maintenance information in modeling be investigated. While the results of implementing maintenance dependent modeling show significant improvement in prediction accuracies, there may be more efficient ways to influence model predictions with the use of maintenance data that do not require the training space to be divided to such a degree.

List of References

- [1] T. O'Hanlon, "Good Vibes on the Web," *Maintenance Technology*, 2003.
- [2] Prepared by the Nuclear Energy Institute, "Nuclear Costs in Context (White Paper)," Nuclear Energy Institute (NEI), Washington, DC, 2016.
- [3] W. Beaver, "Duquesne Light and Shippingport: Nuclear Power Is Born in Western Pennsylvania," *The Western Pennsylvania Historical Magazine*, vol. 70, no. 4, pp. pp.339-358, 1987.
- [4] B. Hallbert and K. Thomas, "Light Water Reactor Sustainability Program: Advanced Instrumentation, Information, and Control Systems Technologies Technical Program Plan for FY 2016," Prepared for the U.S. Department of Energy Office of Nuclear Energy, 2015.
- [5] A. Mathew, S. Zhang, L. Ma, T. Earle and D. Hargreaves, "Reducing maintenance cost through effective prediction analysis and process integration," *Advances in Vibration Engineering*, vol. 5, no. 2, pp. pp.87-96, 2006.
- [6] Westinghouse, "Automation: Data Processing and Monitoring System," Westinghouse Electric Company LLC, 2010.
- [7] A. Jardine, D. Lin and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing*, vol. 20, pp. pp.1483-1510, 2006.
- [8] D. Galar, M. Palo, A. Van Horenbeek and L. Pintelon, "Integration of disparate data sources to perform maintenance prognosis and optimal decision making," in *The Ninth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, London, 2012.
- [9] SoHaR Incorporated, "An Introduction to Advanced Diagnostics and Prognostics For NRC Staff," Prepared for U. S. Nuclear Regulatory Commission Office of Nuclear Regulatory Research, Culver City, California, 2014.
- [10] J. B. Coble, *Merging Data Sources to Predict Remaining Useful Life - An Automated Method to Identify Prognostic Parameters*, PhD diss., University of Tennessee, 2010.
- [11] J. Bird and G. Shao, "A View of Standards for Prognostics and Health Management," *International Journal of Prognostics and Health Management*, 2013.
- [12] J. Hines, J. Garvey, J. Preston and A. Usynin, "Tutorial: Empirical Methods for Process and Equipment Prognostics," in *53rd Annual Reliability and Maintainability Symposium*, Las Vegas, Nevada, 2008.
- [13] J. Coble and J. Hines, "Prognostic Algorithm Categorization with Application to the PHM Challenge Problem," in *Prognostics and Health Management Conference PHM 2008*, Denver, 2008.

- [14] J. Hines and A. Usynin, "Current Computational Trends In Equipment Prognostics," *International Journal of Computational Intelligence Systems*, vol. 1, no. 1, pp. pp.94-102, 2008.
- [15] R. Abernathy, J. Breneman, C. Medlin and G. Reinman, "Weibull Analysis Handbook," Pratt & Whitney Aircraft Government Products Division, United Technologies Corporation, West Palm Beach, Florida, 1933.
- [16] A. Razali and A. Al-Wakeel, "Mixture Weibull distributions for fitting failure times data," *Applied Mathematics and Computation*, vol. 219, pp. pp.11358-11364, 2013.
- [17] K. Carroll, "On the use and utility of the Weibull model in the analysis of survival data," *Controlled Clinical Trials*, vol. 24, pp. pp.682-701, 2003.
- [18] A. Barabadi, "Reliability model selection and validation using Weibull probability plot - A case study," *Electric Power Systems Research*, vol. 101, pp. pp.96-101, 2013.
- [19] Lumileds, "Understanding power LED lifetime analysis - Technology White Paper," Phillips, 2014.
- [20] W. Weibull, "A Statistical Distribution Function of Wide Applicability," *ASME Journal of Applied Mechanics*, pp. 293-297, 1951.
- [21] D. Cox, "Regression Models and Life-Tables," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 34, no. 2, pp. pp.187-220, 1972.
- [22] M. You and G. Meng, "Updated proportional hazards model for equipment residual life prediction," *International Journal of Quality & Reliability Management*, vol. 28, no. 7, pp. pp.781-795, 2011.
- [23] A. Jardine, D. Banjevic, M. Wiseman, S. Buck and T. Joseph, "Optimizing a mine haul truck wheel motors' condition monitoring program Use of proportional hazards modeling," *Journal of Quality in Maintenance Engineering*, vol. 7, no. 4, pp. pp.286-302, 2001.
- [24] A. Tsiatis, V. Degruittola and M. Wulfsohn, "Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS," *Journal of the American Statistical Association*, vol. 90, no. 429, pp. pp.27-37, 2012.
- [25] H. Liao, W. Zhao and H. Guo, "Predicting remaining useful life of an individual unit using proportional hazards model and logistic regression model," in *Reliability and Maintainability Symposium (RAMS)*, 2006.
- [26] "Quantify Relative Risk Using a Cox-Proportional Hazards Model," Wolfram Mathematica, [Online]. Available: <https://www.wolfram.com/mathematica/new-in-9/survival-analysis/quantify-relative-risk-using-a-cox-proportional-ha.html>. [Accessed 2016].
- [27] D. Thompson, "Fluid mechanics and transfer processes," *The Canadian Journal of Chemical Engineering*, vol. 65, no. 3, 1985.

- [28] T. Ardsomang, J. Hines and B. Upadhyaya, "Heat Exchanger Fouling and Estimation of Remaining Useful Life," in *Annual Conference of Prognostics and Health Management Society 2013*, New Orleans, LA, 2013.
- [29] N. Lybeck, B. Pham, M. Tawfik, J. Coble, R. Meyer, P. Ramuhalli and L. Bond, "Lifecycle Prognostics Architecture for Selected High-Cost Active Components," Idaho National Laboratory and Pacific Northwest National Laboratory, INL/EXT-11-22915, 2011.
- [30] C. Lu and W. Meeker, "Using Degradation Measures to Estimate a Time-to-Failure Distribution," *Technometrics*, vol. 35, no. 2, pp. pp.161-174, 1993.
- [31] B. Upadhyaya, M. Naghedolfeizi and B. Raychaudhuri, "Residual Life Estimation of Plant Components," *P/PM Technology*, vol. 7, no. 3, pp. pp.22-29, 1994.
- [32] J. Coble and J. Hines, "Applying the General Path Model to Estimation of Remaining Useful Life," *International Journal of Prognostics and Health Management*, vol. 7, 2011.
- [33] J. Hines, "Development and Validation of a Lifecycle-based Prognostics Architecture with Test Bed Validation," U.S. Department of Energy: Nuclear Engineering University Programs (NEUP), 2014 Final Report.
- [34] S. Crowder and J. Lane, "The Use of Degradation Measures to Design Reliability Test Plans," in *Proposed for presentation at the ICMCSSE 2014 : International Conference on Mathematical, Computational and Statistical Sciences*, Rio De Janeiro, Brazil, 2014.
- [35] D. Coit, J. Evans, N. Vogt and J. Thompson, "A Method for Correlating Field Life Degradation with Reliability Prediction for Electronic Modules," *Quality and Reliability Engineering International*, vol. 21, no. 7, pp. pp.715-726, 29 September 2005.
- [36] Z. Welz, A. Nam, M. Sharp, J. Hines and B. Upadhyaya, "Improved Heat Exchanger Lifecycle Prognostic Methods for Enhanced Light Water Reactor Sustainability," *International Journal of Prognostics and Health Management (IJPHM)*, vol. 18, 2015.
- [37] C. Sankavaram, B. Pattipati, A. Kodali, K. Pattipati, M. Azam, S. Kumar and M. Pecht, "Model-based and Data-driven Prognosis of Automotive and Electronic Systems," in *5th Annual IEEE Conference on Automation Science and Engineering*, Bangalore, India, 2009.
- [38] S. Abdul-Wahab, C. Bakheit and S. Al-Alawi, "Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations," *Environmental Modelling & Software*, vol. 20, no. 10, pp. pp.1263-1271, 2005.
- [39] P. Baruah and R. Chinnam, "HMMs for diagnostics and prognostics in machining processes," *International Journal of Production Research*, vol. 43, no. 6, pp. pp.1275-1293, 2005.
- [40] K. Goebel, B. Saha and A. Saxena, "A comparison of three data-driven techniques for prognostics," in *62nd Meeting of the Society for Machinery Failure Prevention Technology (MFPT)*, 2008.

- [41] D. An, N. Kim and J. Choi, "Review: Practical options for selecting data-driven or physics-based prognostics," *Reliability Engineering and System Safety*, vol. 133, pp. pp.223-236, 2015.
- [42] X. Si, W. Wang, C. Hu and D. Zhou, "Remaining useful life estimation – A review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, pp. pp.1-14, 2011.
- [43] J. Luo, M. Namburu, K. Pattipati, L. Qiao, M. Kawamoto and S. Chigusa, "Model-based prognostic techniques [maintenance applications]," in *AUTOTESTCON 2003. IEEE Systems Readiness Technology Conference*, 2003.
- [44] Math.NET, "Distance Metrics," [Online]. Available: <http://numerics.mathdotnet.com/Distance.html>. [Accessed 2016].
- [45] M. Daigle and K. Goebel, "A Model-Based Prognostics Approach Applied to Pneumatic Valves," *International Journal of Prognostics and Health Management*, vol. 8, 2011.
- [46] B. Saha, K. Goebel, S. Poll and J. Christophersen, "Prognostics Methods for Battery Health Monitoring Using a Bayesian Framework," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 2, pp. pp.291-296, 2009.
- [47] L. Liao and F. Köttig, "Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction," *IEEE Transactions on Reliability*, vol. 63, no. 1, pp. pp.191-207, 2014.
- [48] M. Sharp, J. Coble, A. Nam, J. Hines and B. Upadhyaya, "Lifecycle Prognostics: Transitioning between information types," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2014.
- [49] A. Nam, Bayesian-based Methods for Transitioning Between Prognostic Estimates to Leverage Available Data, PhD diss., University of Tennessee, 2016.
- [50] S. Wang and L. Xi, "Condition monitoring system design with one-class and imbalance-data classifier," in *Industrial Engineering and Engineering Management IE&EM 16th International Conference*, China, 2009.
- [51] C. O'Donoghue and J. Prendergast, "Implementation and benefits of introducing a computerized maintenance management system into a textile manufacturing company," *Journal of Materials Processing Technology*, vol. 153, pp. 226-232, 2004.
- [52] K. Thomas and B. Hallbert, "Long-Term Instrumentation, Information, and Control Systems (II&C) Modernization Future Vision and Strategy," Idaho National Laboratory, Idaho Falls, Idaho, 2012.
- [53] W. Cato and R. Mobley, *Computer-Managed Maintenance Systems: A Step-by-Step Guide to Effective Management of Maintenance, Labor, and Inventory*, Boston: Butterworth Heinemann, 2002.

- [54] D. Galar, A. Thaduri, M. Catelani and L. Ciani, "Context awareness for maintenance decision making: A diagnosis and prognosis approach," *Measurement*, vol. 67, pp. 137-150, 2015.
- [55] L. Swanson, "An information-processing model of maintenance management," *International Journal of Production Economics*, vol. 83, pp. 45-64, 2003.
- [56] M. Stephens, *Probability and Reliability-Based Maintenance Management*, New Jersey: Pearson: Prentice Hall, 2004.
- [57] M. Marseguerra, E. Zio and L. Podofillini, "Condition-based maintenance optimisation by means of genetic algorithms and Monte Carlo simulation," *Reliability Engineering & System Safety*, vol. 77, pp. 151-165, 2002.
- [58] D. Travis and L. Casinger, "Five causes of -and remedies for- maintenance manager headaches," *Plant Engineering*, vol. 51, 1997.
- [59] J. Lee, L. Liao, E. Lapira, J. Ni and L. Li, "Informatics platform for designing and deploying e-manufacturing systems," *Collaborative Design and Planning for Digital Manufacturing*, 2009.
- [60] E. Jantunen, E. Gilabert, C. Emmanoulidis and A. Adgar, "E-Maintenance, A Means to High Overall Efficiency," in *Proceedings of the 4th World Congress on Engineering Asset Management*, Athens, Greece, 2009.
- [61] "PI System™: From data to knowledge to transformation," OSIsoft, [Online]. Available: <http://www.osisoft.com/pi-system/>. [Accessed 2016].
- [62] International Atomic Energy Agency, "Implementation Strategies and Tools for Condition Based Maintenance at Nuclear Power Plants," IAEA, Vienna. TECDOC-1551, 2007.
- [63] International Atomic Energy Agency, "Specific Safety Guide No. SSG-39: Design of Instrumentation and Control Systems for Nuclear Power Plants," IAEA, Vienna, 2016.
- [64] D. Galar, U. Kumar, R. Villarejo and C. Johansson, "Hybrid prognosis for railway health assessment: an information fusion approach for PHM deployment," *Chemical Engineering*, vol. 33, 2013.
- [65] S. Pinjala, L. Pintelon and A. Vereecke, "An empirical investigation on the relationship between business and maintenance strategies," *International Journal of Production Economics*, vol. 104, pp. 214-229, 2006.
- [66] R. Sivalingam, "Applying best practices to maintenance: a 12 step programme for moving down the road to recovery," *Plant Engineering*, vol. 51, 1997.
- [67] A. Jardine, *Operational Research in Maintenance*, Manchester: Manchester University Press, 1970.

- [68] A. Jardine and A. Tsang, *Maintenance, Replacement, and Reliability: Theory and Applications*, Boca Raton, Florida: CRC Press. Taylor & Francis Group, 2013.
- [69] A. Muller, A. Marquez and B. lung, "On the concept of e-maintenance: Review and current research," *Reliability Engineering and System Safety*, vol. 93, pp. pp.1165-1187, 2008.
- [70] H. Ascher, "Repairable systems reliability," *Encyclopedia of Statistics in Quality and Reliability*, 2008.
- [71] Z. Welz, J. Coble and W. Hines, "Identifying requisite changes to support modern maintenance practices," in *10th International Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*, San Francisco, 2017.
- [72] D. Vanier, "Why industry needs asset management tools," *Journal of Computing in Civic Engineering*, vol. 15, pp. 35-43, 2001.
- [73] M. Carnero, "An evaluation system of the setting up of predictive maintenance programmes," *Reliability Engineering & Safety System*, vol. 91, pp. 945-963, 2006.
- [74] E. Brown, N. McCollom, E. Moore and A. Hess, "Prognostics and health management: A data-driven approach to supporting the F-35 Lightning II," in *Aerospace Conference*, Montana, 2007.
- [75] B. Upadhyaya and J. Hines, "On-Line Monitoring and Diagnostics of the Integrity of Nuclear Plant Steam Generators and Heat Exchangers," Final Report: Volume 1, Experimental and Hybrid Modeling Approach for Monitoring Heat Exchanger System Performance, Prepared for the DOE-NEER Program, University of Tennessee, Knoxville, 2004.
- [76] HYDAC International, *Plate Heat Exchangers [Brochure]*, Form #D5.815.0/04.14.
- [77] C. Ford, "The Wilcoxon Rank Sum Test," University of Virginia Library: Research Data Services + Sciences, 2017.
- [78] F. Wilcoxon, S. Katti and R. Wilcox, "Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test," *Mathematical Statistics*, vol. 1, pp. 171-259, 1970.
- [79] F. Incropera, D. DeWitt, T. Bergman and A. Lavine, *Fundamentals of Heat and Mass Transfer*, 6th ed. ed., Wiley, 2007.
- [80] D. Specht, "A general regression neural network," *IEEE Transactions on Neural Networks*, vol. 2, no. 6, pp. 568-576, 1991.
- [81] Z. Welz, J. Coble, B. Upadhyaya and W. Hines, "Maintenance-based prognostics of nuclear plant equipment for long-term operation," *Nuclear Engineering and Technology*, 2017.

- [82] M. Sharp, "Simple metrics for evaluating and conveying prognostic model performance to users with varied backgrounds," in *Annual Conference of the Prognostic and Health Management Society*, 2013.
- [83] N. Draper and H. Smith, *Applied regression analysis*, John Wiley & Sons, 2014.
- [84] M. Sharp, J. Coble, A. Nam, J. Hines and B. Upadhyaya, "Lifecycle Prognostics: Transitioning Between Information Types," *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 2014.
- [85] A. Saxena, J. Celaya, B. Saha, S. Saha and K. Goebel, "Metrics for offline evaluation of prognostic performance," *International Journal of Prognostics and Health Management*, vol. 1, pp. 4-23, 2010.
- [86] I. Aulenbacher and J. Jurado, "A data acquisition system for the Laguna Verde nuclear power plant," in *8th WSEAS international conference on Data networks, communications, computers (DNCOCO'09)*, Stevens Point, Wisconsin, 2009.
- [87] A. Nam, M. Sharp, J. Hines and B. Upadhyaya, "Lifecycle Prognostic Algorithm Development and Application to Test Beds," *Chemical Engineering Transactions*, vol. 33, pp. 901-906, 2013.
- [88] "Knowledge Center: US Nuclear Power Plants," Nuclear Energy Institute (NEI), April 2016. [Online]. Available: <http://www.nei.org/Knowledge-Center/Nuclear-Statistics/US-Nuclear-Power-Plants/>. [Accessed 2016].
- [89] A. Mathew, S. Zhang, L. Ma, T. Earle and D. Hargreaves, "Reducing maintenance cost through effective prediction analysis and process integration," *Advances in Vibration Engineering*, vol. 5, pp. 87-96, 2006.
- [90] E. Jantunen, E. Gilibert, C. Emmanoulidis and A. Adgar, "E-maintenance, A means to higher overall efficiency," in *Proceedings of the 4th World Congress on Engineering Asset Management*, Athens, Greece, 2009.
- [91] C. Lu and W. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, no. 2, pp. 161-174, 1993.

Vita

Zachary Welz was born in Ceiba, Puerto Rico to the parents of Pamela Stone and Gregory Welz. He attended Cleveland High School in Cleveland Tennessee, and later attended the University of Tennessee where he attained his Bachelor and Master of Science in Nuclear Engineering as part of the 5-year BS-MS program. He continued to pursue a Ph.D. in Nuclear Engineering from the University of Tennessee while serving as a Graduate Research Assistant and Teaching Assistant under the Department Head of Nuclear Engineering, Dr. J Wesley Hines.

Zachary was awarded the honors of Undergraduate Researcher of the Year and Outstanding Graduate Researcher in 2014 and 2015 respectively. As a member of the American Nuclear Society, Zach was awarded the Robert E. Uhrig graduate scholarship in 2015, and was elected as Student Executive Committee Member of the Human Factors Instrumentation and Control Division in 2016.

Zachary plans on moving to Ohio and working as a data science and analytics engineer for the Department of Defense after graduating.