University of Cincinnati

Date: 11/9/2015

I, Shanhua Yang, hereby submit this original work as part of the requirements for the degree of Doctor of Philosophy in Mechanical Engineering.

It is entitled:
An Adaptive Prognostic Methodology and System Framework for Engineering Systems under Dynamic Working Regimes

Student's name: Shanhua Yang

This work and its defense approved by:

Committee chair: Jay Lee, Ph.D.
Committee member: Rajkumar Roy, Ph.D.
Committee member: Thomas Richard Huston, Ph.D.
Committee member: J. Kim, Ph.D.
Committee member: David Thompson, Ph.D.
An Adaptive Prognostic Methodology and System Framework for Engineering Systems under Dynamic Working Regimes

Author: Shanhu YANG
Committee Chair: Dr. Jay LEE

A dissertation submitted in fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Mechanical & Materials Engineering College of Engineering & Applied Science

February 8, 2016
UNIVERSITY OF CINCINNATI

Abstract

An Adaptive Prognostic Methodology and System Framework for Engineering Systems under Dynamic Working Regimes

by Shanhu Yang

Prognostics and Health Management (PHM) as a research discipline focuses on assessing degradation behavior and predicting time to failure of an engineering system with condition monitoring data collected throughout the lifespan of the system. The information of predicted Remaining Useful Life (RUL) and potential failure modes further enables just-in-time maintenance, reduced operational cost and optimized production. In recent years with the development of information systems such as cloud computing and Internet of Things (IOT), machine data from factory floors can be collected more conveniently with higher speed, volume and variety, which brings about new opportunities and much wider application of PHM technologies. On the other hand, the emerging industrial big data with real world complications also imposes greater challenges to the PHM research community. Data collected from a large amount of machine units under dynamic working regimes requires algorithms to adaptively and autonomously recognize and handle different situations. Autonomous PHM algorithms can further be implemented in centralized computing platforms for more efficient, faster and large scale data mining and analytics, which will eventually lead to more effective handling and exploitation of industrial big data.
PHM algorithms have been developed based on specific applications and datasets. In addition, most of PHM tools are developed based on limited working regimes. In reality, many engineered machinery and systems often work under different dynamic working regimes and as a consequence it is always a challenge to implement PHM in such conditions. This dissertation work presents the development of a systematically designed and implementation-ready methodology for adaptive health assessment and prognostics for real world machine fleets that undergo dynamic working regimes and other complications. Due to limitations in data and knowledge for in-field systems, the approach assumes no prior knowledge or available training data and attempts to extract degradation information only from condition monitoring data streamed in real time. The approach contains a generalized state space model for machine degradation and an adaptive and online methodology for real time degradation assessment and prediction. The degradation model is a generalized yet comprehensive description of the relationships among the three key aspects in the PHM related research, which are system degradation, system measurements and working regimes. The online methodology further consists of an adaptive segmentation method for identification of health stages based on local variation, a variable selection algorithm for selecting related working regime parameters and an Adaptive Kalman Filter (AKF) based online filtering method for model identification and prediction. The methodology is demonstrated and validated using both simulated data and data from real world industrial applications. The case studies show that the proposed approach is able to deliver robust and accurate results with little algorithm tuning needed for different applications, which is ideal for facilitating automated data processing and analytics in online PHM platforms.
To my wife, Shanshan

The only journey is the one with you.

To my parents

It is such a fortune to still be able to pursue childhood dreams after so many years. Thank you for giving them to me and supporting me ever after.
Acknowledgements

Firstly, I would like to express my sincere gratitude to my academic advisor, Professor Jay Lee for his visionary guidance and constant support throughout my studies at University of Cincinnati. His rigorous attitude in academic training and professionalism has affected me deeply and will greatly benefit me for years to come. I am tremendously fortunate to have committee members Dr. David Thompson, Dr. Jay Kim, Dr. Thomas Huston and Dr. Roy Rajkumar who were more than generous with their expertise and precious time. I thank them for giving thoughtful feedback and moving me forward throughout the process with patience and encouragement.

I am greatly thankful to my colleagues at the IMS Center who have helped me through patient mentoring, inspiring discussions and other supports as friends. They are Dr. Masoud Ghaffari, Dr. Linxia Liao, Dr. Tianyi Wang, Dr. Mohamed AbuAli, Dr. Edzel Lapira, Dr. Fangji Wu, Dr. Yan Chen, Dr. David Siegel, Dr. Seunghul Lee, Dr. Xiaoning Jin, Dr. Yixiang Huang, Dr. Mohammad Rezvani, Mr. Hassan Al-Atat, Dr. Xiaoyang Li, Dr. Wenyu Zhao, Mr. Feibai Zhu, Mr. Su Xu, Mr. Hossein Davary, Ms. Wenjing Jin, Dr. Chuan Jiang, Ms. Ann Kao, Mr. Chao Jin, Mr. Zongchang Liu, Mr. Zhe Shi, Mr. Behrad Bagheri, Ms. Christina Lucas, Ms. Xiaorui Tong, Mr. Yuan Di, Mr. Matt Buzza, Mr. Xiaodong Jia, Ms. Laura Pahren, Ms. Ellen Gamel, Ms. Wang Shuai and many others. I am also thankful to Mr. Patrick Brown and Mr. Michael Lyons for all the help they provided with my academic writing and administrative aspects at IMS Center.

Finally, my special thanks go to my family for their unconditional support and encouragement over the years.
Contents

Abstract i

Acknowledgements v

1 Introduction 1
  1.1 Introduction ................................................. 1
  1.2 PHM for Industrial Implementation: Research Gaps and Unmet Needs .................................................. 4
    1.2.1 Complications in Real World Implementation .......... 4
    1.2.2 Unmet Needs in Platform Implementation ............. 7
  1.3 The Scope of the Adaptive Prognostic Research .......... 7
  1.4 Research Objectives ......................................... 11
  1.5 Contributions and Broad Impact ............................ 12

2 State-of-the-art Review of Related PHM Methods and Techniques 17
  2.1 Adaptive PHM Algorithms ...................................... 17
    2.1.1 Adaptive PHM Algorithms for Different Conditions .... 17
    2.1.2 Training-less Adaptive PHM Algorithms ................. 21
  2.2 Prognostic Methods for Systems under Dynamic Working Regimes 23
    2.2.1 Comparison of Data Driven and Model Based Prognostics 23
    2.2.2 Empirical Model based Prognostics ..................... 25
    2.2.3 Prognostics for System under Dynamic Working Regimes 30
      Working Regimes and System Measurement ................. 31
      Working Regimes and Degradation Rate ................... 31
      Prognostic Methods in Literature ....................... 35
  2.3 Cloud Computing Paradigm and Applications ................ 36
    2.3.1 Cloud Computing Paradigm ............................. 36
    2.3.2 Survey of Cloud Based Computing Platforms .......... 38
    2.3.3 Reconfigurable PHM Platforms ....................... 41

3 Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes 43
  3.1 Generalized Degradation Model for Dynamic Regime Systems .. 44
    3.1.1 The State Space Degradation Model and Kalman Filter .. 44
    3.1.2 Discussion of the General Degradation Process ...... 47
List of Figures

1.1 Overall Steps for a Typical PHM Solution .................................. 8
2.1 The General Relationship between Stress and Expected Life .... 32
2.2 The Degradation Process under Changing Stress Levels ....... 34
2.3 The 4-layer Structure for Typical Cloud Platforms .......... 37
2.4 Decoupling Software from Hardware through Virtualization .... 37
3.1 The Overall Steps of the Developed Adaptive Prognostic Methodology ................................................................. 44
3.2 The Overall Relationship among System Health, Measurement (Output) and Working Regimes (Input) from a System Identification Point of View ................................................................. 49
3.3 System Degradation Calculated with Measurement at the Standard Working Regime ................................................................. 50
3.4 Common Types of Stress that Cause System Degradation .... 52
3.5 Simulated Linear Degradation Histories under Different Working Regimes ................................................................. 58
3.6 Simulated Nonlinear Degradation Histories under Different Working Regimes ................................................................. 59
3.7 Simulated Linear Degradation Histories under Dynamic Working Regimes ................................................................. 61
3.8 Comparison of Different Values of Q. (a) Q being too High causes Oscillation (b) Q being too Low causes Bias ................. 65
3.9 Evaluation of AKF Performance on Different $\alpha$ Values .......... 69
3.10 Result of the Myers AKF with N=50 ................................... 70
3.11 Result of the Myers AKF with N=100 ................................. 70
3.12 Result of the Proposed AKF with N=50 ............................... 71
3.13 Result of the Proposed AKF with N=100 ............................... 71
3.14 Simulated Degradation Data for Prediction Model Comparison .... 74
3.15 Distribution of Prediction Error of Prediction Models .......... 75
3.16 Standard Deviation of Prediction Error at Different Noise Levels .... 76
3.17 Mean Squared Error (MSE) of Prediction Error at Different Noise Levels ................................................................. 76
3.18 Adjusted Degradation Rate under Different Stress Levels ...... 79
3.19 An Example of Applying AKF Method to a Poorly Excited Signal 82
3.20 Adaptive Online Segmentation using Median and $Q_n$ Values as boundaries .......................................................... 84
3.21 Adaptive Segmentation Results for Signals with Different Noise Levels ................................................................. 85
3.22 The Result of the Segmentation Method using a Poorly Excited Signal ................................................................. 87
3.23 The Result of the Adaptive Kalman Filter (AKF) using a Poorly Excited Signal with the Segmentation Method ................. 87

4.1 Overall Framework for the Cloud based PHM Platform .............. 99
4.2 Standardized Input and Output for Cloud based Algorithm Modules ................................................................. 101
4.3 An Example of User Interface for Algorithm Configuration ......... 102
4.4 Steps for Implementing the Cloud based PHM Platform .......... 104
4.5 Standardized Data Processing and Feature Extraction Procedure . 106
4.6 Standardized Format for Storing Features in SQL Database ....... 106

5.1 Simulated Degradation Data and AKF Result ....................... 112
5.2 Distribution of the Prediction MSE ..................................... 112
5.3 Detection Rate of the Effect of Working Regimes over Degradation Rate with Different K .................................................. 113
5.4 Prediction MSE with Different K ........................................ 114
5.5 Examples of the Sawing Application .................................... 115
5.6 The Interface of the Web Service for Asset Management and Reporting ................................................................. 117
5.7 Extracted Features by Cut and the Adaptive Segmentation Result. The x axis is time in the unit of number of cuts ............... 119
5.8 System Measurement, Working Regime and PHM Results for a Complete Run-to-failure History of a Saw Blade ................. 123
5.9 Result Comparison of Different Prediction Methods ............... 123
5.10 Data and Results of Failure History 1. From left to right and top to bottom: measurement and identified segments, working regime variables, AKF result, RUL result ........................................... 125
5.11 Data and Results of Failure History 2. From top to bottom: measurement and identified segments, working regime variables, AKF result .................................................. 127

A.1 Convergence Speed of RLSVFF and KF for an Exponential Signal. For RLSVFF, $\alpha = 0.9$, $\beta = 0.7$. For KF, $Q = 1E-5$, $R = 0$ ....... 140
# List of Tables

1.1 Trends in PHM Research Promoted by Industrial Big Data ........ 3

2.1 Model based Prognostics Methods in Literature .................... 28

3.1 Algorithm for Generating a Random Time Series as a Working Regime Variable ........................................ 57
3.2 The Online Segmentation Algorithm ............................... 84

5.1 Comparison of Prediction Errors ................................. 111
5.2 Raw Data Collected From Sawing Machines ...................... 117
5.3 Features Extracted From Sawing Machines in the Format of Fig. 4.6 .............................................................. 118
5.4 Prediction Error Percentage using Different Methods at 70% Total Life ......................................................... 121
5.5 Prediction Error Percentage using Different Methods at 90% Total Life ......................................................... 121
Acronyms

AKF  Adaptive Kalman Filter.

API  Application Programming Interface.

AR  Autoregressive.

ART  Adaptive Resonance Theory.

BOL  Beginning of Life.

CBM  Condition based Maintenance.

CCA  Canonical Correlation Analysis.

CM  Condition Monitoring.

DAQ  Data Acquisition.

EKF  Extended Kalman Filter.

EOL  End of Life.

GPR  Gaussian Process Regression.

HMM  Hidden Markov Model.

i.i.d.  identical and independently distributed.

IIS  Internet Information Services.
IOT  Internet of Things.

IPL  Inverse Power Law.

IT  Information Technology.

KF  Kalman Filter.

LS  Least Squares.

MLE  Maximum Likelihood Estimation.

MSE  Mean Squared Error.

NN  Neural Network.

PCA  Principal Component Analysis.

PF  Particle Filter.

PHM  Prognostics and Health Management.

QFD  Quality Function Deployment.

RLS  Recursive Least Squares.

ROI  Return on Investment.

RUL  Remaining Useful Life.

SIR  Sliced Inverse Regression.

SNR  Signal-to-noise Ratio.

SVM  Support Vector Machine.

TV  Time Varying.
UKF Unscented Kalman Filter.

VFF Variable Forgetting Factor.

VM Virtual Machine.

WLS Weighted Least Squares.
List of Symbols

\( x_i \)  
\( i^{th} \) sample of a single- or multi-variate time series data

\( x(j) \)  
\( j^{th} \) dimension/variable of a multivariate time series data

\( \Delta x_i \)  
Derivative in the discrete form \( \Delta x_i = x_i - x_{i-1} \)

\( \hat{x} \)  
Estimation or \textit{a posteriori} estimation of \( x \)

\( \hat{x}^- \)  
\textit{a priori} estimation of \( x \)

\( t \)  
Time stamp

\( \eta \)  
System degradation

\( \dot{\eta} \)  
System degradation rate \( \dot{\eta} = \frac{\Delta \eta_i}{\Delta t_i} \)

\( y \)  
System measurement or feature

\( y' \)  
Normalized system measurement or feature to the standard working regime

\( \iota \)  
Working regime parameters, such as speed, temperature, load, etc.

\( S_i \)  
Accumulated stress between time \( t_{i-1} \) and \( t_i \) that causes system degradation

\( s_i \)  
Stress level between time \( t_{i-1} \) and \( t_i \), \( s_i = \frac{S_i}{\Delta t_i} \)
Chapter 1

Introduction

1.1 Introduction

Prognostics and Health Management (PHM) focuses on assessing degradation behavior and predicting time to failure of an engineering system using condition monitoring data collected throughout the lifespan of the system. The information of predicted Remaining Useful Life (RUL) and potential failure modes through diagnosis further enables Condition based Maintenance (CBM), reduced operational cost and optimized production. For different applications and customer concern, system failure can be defined with regard of one or several combined aspects including operational safety, reliability, energy efficiency and product quality. The overall tasks of PHM include 1) health assessment for evaluating current health condition of a system, 2) prognostics for RUL prediction and 3) diagnostics for fault identification and root cause analysis. Although the actual technical approach may vary, the objective of PHM methodologies is to transform machine Condition Monitoring (CM) data into useful information that can be used for decision making and/or feedback control [Djurdenovic, Lee, and Ni, 2003]. Generally speaking, such transformation is achieved by comparing machine data collected from different time instances and/or machine units so that patterns, such as similarity and evolution trend,
can be extracted and converted to information on system health and degradation. The development and implementation of traditional PHM strategies can be resource consuming due to the requirement on data availability, expertise on data analytics and domain specific knowledge on system working principles and degradation physics. As a result, the research of PHM initially started with focuses on critical assets whose failure or downtime causes significant amount of monetary cost or safety hazard such as turbine engines [Wang, Yu, Siegel, and Lee, 2008], wind turbines [Lapira et al., 2012] and rotating machinery [Lee et al., 2014]. After over a decade of development, the merits of applying PHM technologies on critical assets have been strongly justified and widely accepted. However, the resource-consuming disadvantage of PHM development still has not been fully overcome which significantly hampers a further dissemination of PHM technologies to a wider range of industrial sectors and applications.

On the other hand, with the recent development of the Internet of Things (IOT), sensor networks and networked machine controllers, data collected from machine fleets have been showing characteristics with higher volume, velocity and variety which is known as the Big Data issue. While Big Data brings challenges to existing PHM strategies particularly with its high diversity and complexity, those characteristics also raise more opportunity for self-learning and adaptive algorithms and start to facilitate an evolution of PHM research to more intelligent, autonomous and cognitive solutions (table 1.1). If properly analyzed, the rich and comprehensive information contained by industrial big data can greatly help industry to further reduce operational cost and increase equipment safety and reliability; otherwise it will only cause increasing burdens on Information Technology (IT) infrastructure and communication networks. To tackle industrial big data, the cloud computing paradigm provides a suitable foundation with its elastic scaling capability of both data storage and
Table 1.1: Trends in PHM Research Promoted by Industrial Big Data

<table>
<thead>
<tr>
<th></th>
<th>Past &amp; Current</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Expensive, scarce</td>
<td>Ubiquitous, abundant</td>
</tr>
<tr>
<td></td>
<td>From controlled experiments</td>
<td>From real world with more complications</td>
</tr>
<tr>
<td>Development &amp;</td>
<td><em>ad hoc</em> for limited machine type and</td>
<td>Industry 4.0: intelligent, autonomous and adaptive</td>
</tr>
<tr>
<td>Implementation</td>
<td>working regime</td>
<td>Centralized, platform based computing for</td>
</tr>
<tr>
<td></td>
<td>On-site, case specific</td>
<td>machine fleets</td>
</tr>
</tbody>
</table>

processing power. A cloud based PHM platform is able to autonomously process and analyze data collected from machine fleets, transform data to health information for better asset management and hence provide a ubiquitous data analytics service for a wide range of industrial applications [Lee, Kao, and Yang, 2014; Yang, Bagheri, Kao, and Lee, 2015]. Such PHM platform is particularly valuable to industry because it offers high Return on Investment (ROI), low investment risk and most importantly, the flexibility to implement and evaluate PHM solutions on different systems and procedures. To facilitate such transition and implement platform-based PHM solutions to large scale industrial applications such as entire factory floors and machine fleets, in addition to necessary platform framework design such as data connectivity, management and visualization functions, the prognostic algorithm also needs to be more robust, generalized and adaptive so that it can be easily applied to different systems without time consuming *ad hoc* user configurations [Sánchez, 2008].
1.2 PHM for Industrial Implementation: Research Gaps and Unmet Needs

1.2.1 Complications in Real World Implementation

As today’s engineering systems are being designed more durable for ambient environments and versatile for a wide range of tasks, machine behavior and degradation pattern is becoming more and more difficult to predict and react upon because existing PHM methods are typically developed for systems under single or limited conditions [Kumar, Tseng, Guo, and Chinnam, 2008]. In order to solve such problem, adaptive on-line analytics and data mining algorithms need to be developed to model the relationships among machine behavior, degradation and working regimes using real-time data from in-field systems. On the other hand, the emerging industrial big data with unprecedented complexity imposes great challenges to the research on autonomous algorithms for health assessment and prognostics. Online data collected from a large amount of in-field machines may consists of many different combinations of working regimes (e.g. different levels of ambient temperature, speed, load, etc.) and types of failure, which is difficult to model comprehensively initially and thus require algorithms to adaptively and autonomously recognize, model and accumulate knowledge of all the regimes and situations observed in real time. For instance, the units in a machine fleet can be utilized for different tasks, at different ambient environment and at different health conditions. Consequently their behavior observed in the condition monitoring data, such as noise level, data variation and degradation rate over time, can be very different and difficult to comprehensively model without prior knowledge of different regimes. In such situation the traditional health assessment model for one or
several pre-defined regimes will not be able to effectively cover all the variations of working regimes and consequently fail to deliver accurate prognostics results.

The complications often observed in real world data can be further summarized into the following categories:

- **Dynamic Working Regimes** It is very common for *in situ* systems to undergo dynamic or multiple working regimes, such as the changes in speed and load for rotating machinery [Heng, Zhang, Tan, and Mathew, 2009] and seasonal change of ambient temperature for outdoor systems such as vehicle fleets [Kurzke, 2003] and wind farms [Lapira et al., 2012]. Currently the common solutions for the multi-regime issue include (see the comprehensive survey in [Siegel, 2013, ch. 3]):

  - **Local Models** Different models are assigned to each possible regime and the most suitable one is selected during the online monitoring [Cholette, Celen, Djurdjanovic, and Rasberry, 2013].

  - **Normalization Based** If the relationship between system measurement and the working regime parameters can be expressed by a function, then regression or data driven (e.g. Neural Networks) methods can be used to model such relationship. The learned model is then used during online monitoring to normalize the measured system signals to a predefined regime level.

However, both of the above mentioned solutions require certain amount of prior knowledge or training data of the multi-regime system. Such requirement is a major hurdle for the desired autonomous platform-based implementation because of the limited data availability in real world applications primarily due to the variations between systems [Reuben and Mba, 2014].
• **System Variations** Variations can be caused by different utilization patterns, ambient environment and maintenance actions applied to systems. Due to such variation, in many cases the data collected from one machine cannot be directly used as training data for other units without careful examination or classification [Lapira et al., 2012]. Moreover, it may cause systems to have different noise and uncertainty levels which also limits the utilization of historical data as training data for other units.

• **Data Availability: Insufficient or Abundant Variables** PHM algorithms typically use controller variables (e.g. speed, temperature, current, etc.) to identify system working regimes. However, the availability of the controller variables is usually decided at the early design phase of the machine during which condition monitoring and prognostics may not be thoroughly considered. Variable availability can also be limited by the type of the controller or network connectivity. As a result, the working regime related variables included in the CM data may be insufficient or abundant (i.e. with unrelated variables). The method therefore needs to first identify working regime parameters that have actual influence on machine performance before further processing.

• **Domain Knowledge/Model Availability** Physical or empirical knowledge based degradation models are widely used in the prognosis and diagnosis [Chiachío et al., 2015; Gašperin, Juričić, Boškoski, and Vižintin, 2011; Orchard and Vachtsevanos, 2009; Zio and Peloni, 2011] which are usually more robust, easier to interpret and less data-dependent comparing to data-driven methods [Spiridonakos and Fassois, 2014]. However, model availability can be an issue when there is little prior knowledge or the system is too complex for a descriptive model to be developed. If such problem cannot be solved directly (e.g. costs too much resource to develop), the alternative approach is to use generalized models such as
1.2.2 Umet Needs in Platform Implementation

Algorithm adaptability is also a key requirement when PHM solutions need to be deployed in cloud platforms for multi-unit applications (e.g. for a fleet of machines or in large factory floors) for autonomous data mining and prognostics. Comparing to traditional ad hoc PHM solutions, cloud based or centralized data analysis requires highly autonomous and adaptive algorithms to handle complex situations and variances among different machines as human configuration, manual algorithm training and parameter tuning toward each machine is often impractical. However, existing research works on cloud based data platforms are for limited or specific types of applications thus lack algorithm adaptability and robustness required for more complex situations. Existing platform development efforts focus mostly on platform framework design such as data connectivity, management and configuration of sensor networks without much improvement of the data analysis functions comparing to traditional ad hoc PHM algorithms [Zhou et al., 2005; Han and Yang, 2006]. As a result, there is an urgent need for developing adaptive prognostic methodologies so that PHM platforms can be more effective when applied to real world engineering systems and more robust when handling real world complications.

1.3 The Scope of the Adaptive Prognostic Research

Generalized and adaptive algorithms are able to handle data collected from different situations automatically without any manual reconfiguration. Although
Figure 1.1: Overall Steps for a Typical PHM Solution
such property is desired for many applications, the development of adaptive algorithms can be difficult and often times involves compromise on algorithm accuracy. For PHM methods specifically, its development often requires domain specific knowledge which can hardly be generalized without any prior information. As a result, in this section a typical PHM solution is separated into connected steps and the possibility of developing adaptive solution for each step is discussed. Based on the discussion, the scope of the proposed research is further clarified.

The development of a PHM solution typically requires two types of knowledge: 1) the domain specific knowledge, which can hardly be generalized, and 2) data analytics such as pattern recognition and time-series prediction, which has the potential of generalization. As a result, as a development strategy, the thesis work will focus on developing adaptive solutions for the data analytics tasks. Figure 1.1 shows a step by step breakdown of a typical PHM solution. The steps can be summarized as follows

- **Data Acquisition and Transmission** As the first step, machine working data is collected from controllers and add-on sensors. Modern machine controllers usually keep records of key operational variables such as speed, position, voltage, environmental conditions and performance variables such as system temperature and power consumption, which all can be used in PHM algorithms for various tasks such as data segmentation by working regime, synchronization and performance assessment. Controller signals are typically generated at a low frequency (i.e. >1 Hz) with the values being averaged within the sampling interval hence are usually not sensitive to incipient faults. As a result, add-on sensors such as accelerometers can be installed so that more detailed machine condition data can be acquired. The selection of sensor type, location and variable type are all specific to the particular system and application.
- **Signal Preprocessing and Feature Extraction** In those steps, the CM data is processed into machine health feature matrix. Before feature extraction, some pre-processing steps may be required such as de-noising and segmentation so that the features can be properly extracted. Depending on different application or data type, different feature extraction techniques can be used [Lee et al., 2014].

- **Health Assessment, Diagnostics and Prognostics using Data Analytics Algorithms** After feature extraction, the original data is transformed into a m by n feature matrix (i.e. feature space) within which each column is one feature variable (i.e. one dimension) and each row is one observation from a certain time point and a machine unit. The tasks of the following data analytics algorithms are to find the most suitable feature variables (i.e. feature selection or dimension reduction), use historical data to train a machine health model and then use test data (i.e. data collected from the machine being currently monitored) to generate system health values with the trained model. The algorithms used in these steps fall into several broad categories including statistics, pattern recognition, similarity measurement and time-series prediction.

To this point it is clear that the PHM steps can be grouped into two parts as shown in figure 1.1. The steps in the left part heavily depend on domain specific knowledge and can vary dramatically with different types of systems while the one on the right side considers the feature matrix as a multi-dimensional data space with less associated physical meanings\(^1\). As a result, the development of adaptive PHM presented in this thesis is focused on the data analytics tasks with the condition that the feature values have already been extracted \(^2\).

\(^1\)An exception is the physical model based PHM methods. See the comparison between data driven and physical model based PHM methods in section 2.2.1

\(^2\)With such condition, the transformation from the original system measurement to health related features is assumed to always exist hence the term *system measurement* and *system feature* will be used interchangeably in the thesis
On the other hand, although the domain specific tasks cannot be generalized, a toolbox approach, which modularizes different signal processing and feature extraction techniques into a collection of tools, can be used to facilitate faster development, configuration and implementation [Djurdjanovic, Lee, and Ni, 2003]. The utilization of the toolbox approach during the implementation stage for a reconfigurable PHM platform will be further discussed in chapter 4.

1.4 Research Objectives

The overall objective of the proposed research is to develop a robust and adaptive algorithm solution that can plug-and-play to a wide range of industrial systems. Such solution can be deployed in cloud computing platforms to provide on-demand and fully automated data processing and promote wider implementation of PHM solutions. More specifically, the developed methodology shall be used

- for a wide range of engineering systems under dynamic working regimes
- with no algorithm training and historical data required because of the machine variation issue
- for online real time health assessment and prediction with little manual configuration so that it can be implemented in automated computing platforms

Most PHM algorithms have been developed based on specific applications and limited working regimes. In reality, many engineered machinery and systems often work under different dynamic working regimes and as a consequence it is always a challenge to implement PHM in such conditions.
Therefore, developing a PHM methodology that can adapt to changing or multiple working regimes is one of the key objectives of this thesis work. Moreover, given that the prior knowledge and/or training data is difficult to prepare or acquire for each machine unit in an automated process, a practical way to solve the machine variation issues is to calculate machine performance and health condition using only the recent data collected from the same machine and streamed to the computing platform in real time [Reuben and Mba, 2014].

To achieve the research objectives, related methods and techniques are first extensively reviewed in order to develop an optimal PHM methodology. Targeting different noise levels, an adaptive segmentation algorithm is proposed to convert feature space observations into discrete health states; a Canonical Correlation Analysis (CCA) based variable selection algorithm is designed to identify most related working regime parameters. For the health assessment and prediction, an state space degradation model is developed based on generalized concepts from accelerated life testing theory and empirical knowledge extracted from degradation patterns of a wide range of industrial systems. Furthermore, an optimized AKF method is developed for online model identification. The robustness, accuracy and reliability of the developed methodology when applied to real world industrial data sets are also analyzed to insure the applicability with regard of platform implementation.

1.5 Contributions and Broad Impact

The thesis work develops an implementation-ready, adaptive and systemically designed health assessment and prognostics methodology that can reliably perform real time data analysis for real world engineering systems that often undergo dynamic working conditions. The methodology is especially optimized
for in-field implementation so that it can plug-and-play for general industrial systems and be commercialization ready. Algorithm adaptability is a critical contributing factor for overall solution robustness and applicability when handling *in situ* machine data. Adaptive algorithms require less effort on algorithm tuning for each application thus lower initial development cost of PHM solutions. Nowadays, the rigidness and inability of handling unprecedented events and other real world complications are some of the major hurdles that prevent current PHM algorithms from being widely implemented in industry. In this context, the developed work is one of the first efforts in the field of PHM research to systematically review these implementation related complications and develop a optimized adaptive methodology to overcome them so that the aforementioned objectives can be achieved.

The contributions and intellectual merits of the research work are further summarized as follows:

- **Development of an Implementation Ready PHM Methodology for Real World Data** The overall methodology is developed targeting complications in real world data that prevent existing PHM solutions from performing effective data analysis. The methodology is adaptive to different applications with little parameter tuning and no algorithm training required hence is suitable for computing platforms to perform unsupervised and autonomous data analysis.

- **Generalized and Comprehensive State Space Model for Degradation Assessment under Dynamic Working Regimes** By investigating typical characteristics and behaviors of condition monitoring data that are commonly seen in industrial cases, a unified mathematical framework is proposed that can be used to model system degradation under dynamic working regimes, different noise level and data type. The degradation
model comprehensively describes the relationships among the three key aspects namely system degradation, system measurement and working regimes. These three aspects and their relationships are the fundamental research issues in the field of PHM study. As a result, the thesis work also provides a complete guideline for future research.

- **Adaptive PHM Algorithm with Online Learning and Autonomous Parameter Tuning** The adaptability of the algorithm is developed with respect of the following two aspects

  - **Adaptive to Dynamic Working Regimes** The developed method requires no prior knowledge or training data of the monitored system and performs normalization on system measurement that is affected by dynamic working regimes. The online normalization is achieved based on a generalized state space degradation model and an online filtering approach for model identification.

  - **Highly Automated for Online Estimation with Little Case-specific Configuration Required** Parameter tuning has always been a major task during algorithm development, training and implementation. Traditionally, parameter tuning has to be done in a manual and time consuming manner. To tackle such issue, various adaptive techniques have been developed in the field of machine learning. For the online estimation applications specifically, due to the fact that the overall variance and noise level cannot be obtained, it has always been an issue on finding the optimal parameter settings such as window length for local/kernel based regression/smoothing methods and covariance metrics for stochastic filtering methods. Targeting online degradation assessment and prediction, an AKF algorithm is developed that is more accurate and robust comparing to existing
AKF methods. The developed AKF method does not require prior-knowledge on system noise level and little application-specific parameter tuning is needed. By itself, the developed AKF can also be used for estimating noise variance metrics in real time for other stochastic filtering methods such as Particle Filters (PFs) and other state-space degradation models in PHM applications.

- **Accurate RUL Prediction under Dynamic Working Regimes** Existing prediction algorithms, especially those based on time-series prediction, require consistent working regimes for both the training and implementation phase in order to perform accurately. Therefore the applicability is very limited in industry where dynamic or multiple working regimes happen frequently. In contrast, the developed working regime based degradation model enables the methodology to perform more accurate and stable prediction when the system measurement (i.e. the time-series data) is compromised by changing working regimes. The online estimation methodology normalizes the system measurement with respect to dynamic working regimes so that a more smooth degradation pattern can be used for prediction. The regime based prognostics further enables the possibility of working regime based control optimization, within which different working regime (e.g. work load) options can be evaluated with regard of system RUL and other concerns such as productivity, reliability and safety.

- **Unified System Framework for Algorithm Implementation in Computing Platforms** Comparing the traditional PHM deployment mechanism, the developed cloud based PHM system can help prompt the implementation of PHM systems in production floors with the advantages including 1) fast and low-cost machine monitoring and PHM service deployment, 2) minimal IT burden for machine users and companies and 3) reliable and
customizable module-based PHM algorithms for in-field machines and real world data.

Moreover, the thesis work will bring about innovation to impact the next generation smart products and systems by using networked sensors and fleet-wide information systems. The potential applications of the developed work, due to its generic nature, cover different industries including the automotive industry, the aerospace industry, the transportation industry, and the semiconductor industry. In addition, the developed methodology can impact how smart analytics can manage the Big Data issue in today’s in-factory and in-field services.
Chapter 2

State-of-the-art Review of Related PHM Methods and Techniques

2.1 Adaptive PHM Algorithms

In general, the adaptability of a PHM solution can be interpreted in two different ways. A solution can be considered adaptive because it is able to generate robust and accurate results by autonomously adjusting its settings toward different conditions such as noise level and other characteristics in the data. On the other hand, one can also define adaptability as the algorithm does not require training data and the manual training phase. According to the discussions in section 1.2, both types of algorithm adaptability are necessary for the PHM solution to effectively handle real world data.

2.1.1 Adaptive PHM Algorithms for Different Conditions

The adaptive PHM algorithms in this category can be further classified into two types, namely the multi-model based and the autonomous parameter tuning based methods.
Multiple Models

The multi-model based adaptive algorithms first build a library of models within which each model can be used to represent a particular condition. Then during implementation, an objective function or measure of fitness is used to determine the most suitable models for the current dataset or machine unit, which will then be used to generate the final result. For instance, [Oh, Kim, and Cho, 2004] proposes an adaptive time-frequency method for diagnosing rotary compressors. With a collection of predefined basis functions, the approach selects the fittest basis function for wavelet packet decomposition according to an entropy-based discriminant distance cost function. The adaptive property makes it possible for the monitoring system to also track the severity of the faults. In [Le Borgne, Santini, and Bontempi, 2007], time-series regression based prediction is used in wireless sensor networks to reduce the amount of data needs to be transferred, with which only the sensor readings that do not agree with the prediction result (e.g. the difference between the two is larger than a given threshold) are transmitted. To improve prediction accuracy, a group of Autoregressive (AR) models with orders from 1 to 5 are used and the racing mechanism with the Hoeffding bound is used to evaluate different models and drop the poor models in real time until only one model is left. In [Reuben and Mba, 2014], the degradation process of a type of rolling element bearings is considered as a combination of three linear processes namely serviceable, slow degradation and fast degradation each of which has a faster degradation rate than the previous. Correspondingly, a switch Kalman filter (SKF) is proposed so that different models are selected and used at different stages. In [Andersson, 1985], an adaptive Recursive Least Squares (RLS) method is proposed using multiple regression models running in parallel each with a different forgetting factor. It is shown that such adaptive forgetting with multiple models (AFMM) is particularly good at tracking parameters with rapid or sudden changes. To
implement AFMM certain level of prior knowledge is still needed to choose the number of models and the forgetting factors as improper settings may worsen the tracking performance [Karasalo and Hu, 2011].

Despite the improvements on accuracy and flexibility introduced by multi-model adaptive approaches, it is not considered as suitable for the research objective because it requires extensive system knowledge or training data to develop the initial model library.

**Parameter Self-tuning**

The autonomous parameter tuning methods on the other hand, utilizes one model but attempts to tune model parameters depending on characteristics observed in the data. Adaptive algorithms with parameter tuning are widely used in online estimation applications for Time Varying (TV) systems whose property changes over time. For instance, the performance of online regression methods such as RLS largely depends on the parameter window length N, which determines how many most recent data samples are used for the regression. One of the common approaches to determine a suitable N uses a Variable Forgetting Factor (VFF) for exponential forgetting. For m data samples $x_i$ where $i \in [1, m]$ that have been collected so far, VFF assigns a weight value to each sample $w_i = \lambda^{m-i}$ where $\lambda$ is the variable forgetting factor. In [Chia-Chang, Hsuan-Yu, and Jyh-Horng, 2005], a fuzzy logic based method is developed to determine the forgetting factor. The method is developed specifically for a DS-CDMA interference suppression receiver to minimize the steady-state error and fasten the convergence speed. In [Shu-hung and So, 2005], the value of the forgetting factor is calculated online using a gradient based formula. The gradient is the change of the MSE over the change of the forgetting factor. Moreover,
arbitrary upper and lower limits of the forgetting factor are also applied to ensure stability of the algorithm. In [Paleologu, Benesty, and Ciochina, 2008], the forgetting factor is calculated each time based on the difference between the a priori and a posteriori error covariance. Theoretically, instead of trying to remove or minimize the noise, the method focuses on reconstructing the add-on noise signal such that the calculated forgetting factor is the optimal. Similar to RLS, the stochastic filtering methods such as Kalman filters and Particle filters also require some of the key parameters to be adaptively tuned to ensure optimal performance. For instance, the noise covariance metrics Q and R in the Kalman filtering approach are key to the overall ability of noise reduction and signal reconstruction. As a result, AKF methods are developed to use the real time a priori and a posteriori error to update the noise metrics [Mehra, 1970; Myers, Tapley, et al., 1976]. For Particle filters such as in [Fox, 2001], the adaptability improvement is focused on choosing the optimal number of particles on-the-fly. The method is based on the KL divergence between the true distribution and the one estimated using the Maximum Likelihood Estimation (MLE). The number of particles is selected so that the error $\epsilon$ is below a given threshold.

Adaptive algorithms with parameter auto-tuning does not require as much algorithm training or prior knowledge and optimize the algorithm parameters in real time using the most recent data samples. As a result, it is a suitable candidate for performing system modeling tasks for TV systems. The general concerns when implementing auto-tuning adaptive algorithms include

- **Initial Conditions and Parameter Values** In an ideal case, the performance of the parameter tuning algorithms shall not depend too much on initial conditions and/or parameter values. In other words, the parameters shall converge to the true value quickly even when the initial setting is inaccurate. If high dependency exists, the algorithm has to rely on some application-specific knowledge to start with a more accurate initial
setting, which in a way reduces the algorithm adaptability.

- **Robustness and Stability to Noise and Outliers** Comparing to algorithms with fixed parameter setting, the parameter tuning methods gain more flexibility with the risk of losing robustness and/or stability. If not properly handled, unexpected data characteristics such as high noise or outliers may cause the estimated new parameter value to diverge and sequentially the whole algorithm to fail. As a result, robustness and stability shall be extensively evaluated during algorithm development.

### 2.1.2 Training-less Adaptive PHM Algorithms

During PHM development and implementation, the traditional training phase requires collection and preparation of historical data as training data, data labeling (e.g. label training data samples as healthy, acceptable and faulty condition or different failure modes) and model learning which are all resource consuming and sometimes tedious. In some other cases where a new engineering system has just been installed, there exists no historical data for training [Coble, Humberstone, and Hines, 2010]. It is a well-known issue for PHM applications and as a result, researchers have started to investigate adaptive PHM solutions that do not require historical data and manual training so that PHM solutions can be quickly deployed to in-field systems. Although actual technical approach varies, training-less adaptive PHM algorithms are typically designed based on the understanding that data collected at the early stage of a system can represent the healthy condition (i.e. baseline) and the more dissimilar the current data to the healthy data, the more degradation has developed. For instance, in [Kumar, Tseng, Guo, and Chinnam, 2008], with the assumption that sensor based historical datasets for several identical assets are available for the entire degradation process, a Hidden Markov Model (HMM) based
adaptive algorithm is developed to determine the health condition of an asset without any predefined health labels. Because the approach requires degradation histories of several assets, the approach is suitable for health assessment of one particular type of machine. In [Coble, Humberstone, and Hines, 2010], targeting the issue of training data availability, the algorithm is first trained using simulated data. Then Principal Component Analysis (PCA) is used to 1) update the model trained from the simulated data using real world data and 2) detect anomaly in the real world data (i.e. fault detection). In [Chen, Kunche, and Pecht, 2013] an incremental learning approach is proposed for fault detection and diagnosis. Instead of using a separated standalone training phase, an online learning approach is used with a repository of degradation models. Using the data collected from engineering systems, new models will be added to the repository when a new failure behavior is observed. Such online learning mechanism is similar to the one introduced by the Adaptive Resonance Theory (ART) [Grossberg, 2003]. Within the framework of ART, a collection of nodes is used to represent different clusters in the data. When a new sample is recorded, depending on the similarity between the new sample and the existing nodes, either one node is updated according to the sample or a new node is created. The sensitivity (i.e. the tendency of creating new nodes) of the algorithm is controlled using a vigilance parameter which shall be optimized toward each application. Another school of training-less algorithms is the online filtering methods such as PF based prognostics in [Orchard and Vachtsevanos, 2009]. Online filtering methods typically do not require and historical data for training. However, expert or empirical knowledge are needed for establishing the model and initial conditions.

In conclusion, adaptive PHM algorithms are more flexible and easy-to-implement thus are more suitable for online applications and TV systems. However, to this date the research efforts on adaptive methods in the field of PHM
mostly target at limited types of application or issues. As a result, it is urgently needed a systematic design and development of a generalized adaptive PHM approach for effective handling of real world data generated by in-field industrial systems.

2.2 Prognostic Methods for Systems under Dynamic Working Regimes

2.2.1 Comparison of Data Driven and Model Based Prognostics

PHM methods can be broadly classified into data-driven and model based approaches. Data-driven methods consider the system as a black box and uses machine learning algorithms such as Neural Networks or statistical models to learn system behavior from the patterns and trends contained in the condition monitoring data [Si, Wang, Hu, and Zhou, 2011]. Comparing to data-driven methods, the model based prognostics incorporates physical or empirical knowledge of the monitored system. Because of the support of such knowledge, the model based methods has less dependency on data availability and are more accurate and stable with noisy or incomplete data [Spiridonakos and Fassois, 2014].

Both data driven and model based methods can be used at most of the steps of a PHM solution (figure 1.1). For example, during feature extraction, both statistics based and physical knowledge based features (e.g. energy at bearing characteristic frequencies) can be used depending on different applications. For health assessment, when training data is available for both healthy and faulty conditions, data driven methods such as logistic regression can be
used; on the other hand, if a well-defined simulation model is available, residuals can be calculated between the model output and the real data as a measure of system degradation.

The physical model based PHM is particularly popular for systems whose models are well established. For instance, battery electric models combined with stochastic filtering methods are widely used in the estimation of State of Charge (SOC), available capacity and internal resistance of Lithium-ion batteries [Rezvanizaniani, Liu, Chen, and Lee, 2014].

Comparing to the model based approach, the data driven approach requires less domain specific knowledge to develop, is more general for different types of engineering systems and thus can be developed and deployed quickly for different applications. However, data driven PHM also suffers from some disadvantages mostly because that its performance is highly limited by data availability and data quality (e.g. noise level/uncertainty, outliers, etc.). The data availability issue is further aggravated for prognostics applications, where for a data driven method, multiple datasets from complete run-to-failure histories are expected so that a representative degradation trend and related variation can be summarized and modeled comprehensively [Wang, Yu, Siegel, and Lee, 2008; Nectoux et al., 2012]. Although multiple failure histories can be generated through simulation [Wang, Yu, Siegel, and Lee, 2008] or accelerated life testing [Nectoux et al., 2012], in general they cost too much resource to be widely available.

The empirical model based methods, which utilizes empirical degradation models generalized from degradation data, achieve a good balance between the physical model based and data driven approaches. Comparing to physical model based PHM, the empirical model based methods does not need the exact working principle of each system thus can be applied in a wider range
of applications. On the other hand, comparing to the data driven methods, the degradation model based method has less dependency on data thus is more robust and accurate for many prognostic cases. As a result, many empirical model based prognostics methods have been developed for a wide range of applications.

In the next subsection, existing empirical degradation models and corresponding model identification techniques are further reviewed in detail.

### 2.2.2 Empirical Model based Prognostics

Model identification methods are used to estimate coefficients and/or states of the established degradation model, either physical or empirical based, using machine CM data. When performed offline with all data available, regression methods such as Least Squares (LS) regression are used for the identification task. However, the thesis work focuses on online identification for which at each time point only the data samples collected at or earlier than the current time are available (i.e. filtering). In general, filtering is more difficult than regression because the overall scale and variation of the data is unknown. Moreover, for TV systems whose property changes over time (e.g. systems with health degradation), it is preferred to use only the recent data samples to perform the estimation so that the property change can be captured. Through literature, the most often used online filtering methods include RLS (see Appendix A for more detail), Kalman Filter (KF) for linear systems and nonlinear filters such as PF, Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF).

The degradation models can be expressed as a function of time [Yang, 2011] or in the form of a state space model. When the former form is used, online regression methods are used to estimate the coefficients in the function. For the latter case, stochastic filtering methods such as KF and PF can be used.
Comparing to the function form, the state space model based health assessment and prognostics is gaining more popularity because in addition to coefficients, the system health can be defined as a state hence estimated directly. Nevertheless, there exists no clear boundary between the two forms as shown in [Gril-lenzoni, 1994], the function form can be written in the state space form as the change of the coefficients can be considered as a random walk process.

**Linear Models**  Here the linear model is defined as that the time function or the state space equations can be expressed in a linear form. In [Reuben and Mba, 2014], the degradation process is considered as a combination of three linear processes namely serviceable, slow degradation and fast degradation each of which has a faster degradation rate than the previous. Correspondingly, a switch Kalman filter (SKF) is proposed so that different models are used at different stages.

In [Lall, Lowe, and Goebel, 2010], KF is used for processing features and performing prognostics for electronics under structural damage, which will cause the resistance components deteriorate. The change of the feature vector under vibration load is nonlinear and has an accelerated degradation toward the End of Life (EOL) hence the exponential curve can be used to approximate such trend. Then Taylor expansion is performed to the 2nd order and the resulting polynomial is used as the state transition matrix.

\[
\Phi(t) = \begin{bmatrix}
1 & T & 0.5T \\
0 & 1 & T \\
0 & 0 & 1 \\
\end{bmatrix}
\] (2.1)

where \( T \) is the time interval. The state vector is \([x \quad \dot{x} \quad \ddot{x}]^T\) where \( x \) is the interconnect resistance. The KF is used for smoothing, identifying state values and prediction of RUL of the component through extrapolation. The result shows
that the predicted RUL is generally larger than the real one, possibly due to the loss of information through the 2nd order Taylor expansion.

In the dissertation work [Yang, 2011], a library of degradation models is proposed to describe the different shapes of the degradation curve. The models are defined as

$$y = ax^2 - (1 + a)x + 1$$  \hspace{1cm} (2.2)

and by varying $a$ within $[-4, 4]$ one gets a collection of degradation curves. Then during online monitoring, Partial Least Squares (PLS) is used to selected the influential regressor variables and the model that gives the minimal MSE is used for prognostics.

**Nonlinear Models** Nonlinear models are usually developed when more comprehensive physical knowledge of the system is available. In [Chiachio et al., 2015] a state equation is developed for estimating micro-crack growth under fatigue cycles for composite materials. The crack propagation model is developed using physics based simulation on crack growth. Using a Bayesian filtering framework, the reliability of the system is estimated using the measured crack length data.

In [Gašperin, Juričić, Boškoski, and Vižintin, 2011] a state space model is used for modeling the damage process dynamics of a single gearbox. Using vibration data collected from the gearbox, the health related features are first extracted through frequency domain analysis. Then the features are used as input for the UKF to estimated the health state of the gearbox system. A similar empirical model for gearbox prognostics is also used in [Orchard and Vachtsevanos, 2009] with PF as the model identification method. Table 2.1 further summarizes the reported applications in literature for model based prognostics.
### Table 2.1: Model based Prognostics Methods in Literature

<table>
<thead>
<tr>
<th>Type</th>
<th>Application</th>
<th>Model</th>
<th>Detail</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>Rolling element bearing</td>
<td>Multiple</td>
<td>Three linear models for different degradation rates are used.</td>
<td>Reuben and Mba, 2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KF</td>
<td></td>
<td>Lall, Lowe, and Goebel, 2010</td>
</tr>
<tr>
<td></td>
<td>Electrical component (resistor)</td>
<td>KF</td>
<td>Second order Taylor expansion of an exponential function</td>
<td>Yang, 2011</td>
</tr>
<tr>
<td></td>
<td>Manufacturing line</td>
<td>Partial</td>
<td>A collection of quadratic functions of different degradation rate is used for different systems.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>LS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonlinear</td>
<td>Crack growth in composite materials</td>
<td>Bayesian filtering</td>
<td>Model is obtained using physics based crack growth simulation then updated with real world measurement.</td>
<td>Chiachío et al., 2015</td>
</tr>
<tr>
<td></td>
<td>Gearbox</td>
<td>UKF</td>
<td>An empirical nonlinear state space model is developed for the damage process.</td>
<td>Gašperin, Juričić, Boškoski, and Vižintin, 2011</td>
</tr>
<tr>
<td></td>
<td>Gearbox</td>
<td>PF</td>
<td>Empirical nonlinear model</td>
<td>Orchard and Vachtsevanos, 2009</td>
</tr>
<tr>
<td></td>
<td>Turban Engine</td>
<td>Multi-model regression</td>
<td>Exponential function for engine degradation over time</td>
<td>Wang, Yu, Siegel, and Lee, 2008</td>
</tr>
<tr>
<td></td>
<td>Lithium-ion battery</td>
<td>PF</td>
<td>Exponential degradation model</td>
<td>An, Choi, and Kim, 2013</td>
</tr>
</tbody>
</table>
Moreover, in [Celaya, Saxena, and Goebel, 2012], the uncertainty during prognostics using KF is extensively discussed. Assuming the states are iid Gaussian random variables, the uncertainty (i.e., variance) will propagate as one attempts to predict future states using the existing state estimations. More specifically, the \( l \) step ahead prediction will be Gaussian distributed \( \hat{x}_p \sim N(\mu_l, \sigma_l^2) \), with

\[
\mu_l = \Phi^l \hat{x}_p + Bu_p
\]

and

\[
\sigma_l^2 = \Phi^{2l} P_p + \sum_{i=1}^{l-1} \Phi^{2i} Q + Q
\]

where \( \hat{x}_p \) and \( P_p \) are the estimated state and state error covariance matrix at the current iteration \( p \) and \( Q \) is the covariance matrix of the process error.

By reviewing the reported model based prognostic methods, it can be seen that although degradation models have been used in a wide range of applications, such as in mechanical, electrical and chemical systems, the overall degradation pattern under constant working regimes, regardless of the type of system, usually follow either linear or exponential functions with respect to time or machine cycle. Even in nonlinear models for specific systems, the general pattern still follows either linear or exponential shape. As a result, when specific degradation knowledge is not available, it is reasonable to use approximation models, especially linear and second order polynomial models as generalized models for health assessment and prognostics. Because availability of the prior knowledge of a given engineering system cannot always be guaranteed, the thesis work utilizes a generalized linear model with adaptive model identification techniques so that it can be quickly applied to different systems without model development and tuning.

Moreover, for the online estimation applications, due to the fact that the overall variance and noise level cannot be obtained with incomplete data, it has
always been an issue on finding the optimal parameter settings such as window length for local/kernel based regression/smoothing methods and covariance metrics for stochastic filtering methods. As a result, an adaptive method for online prognostics needs to be developed to automatically tune the parameter values in real time.

### 2.2.3 Prognostics for System under Dynamic Working Regimes

In industrial applications, it is very common to see engineering systems going through dynamic working regimes. For instance, a machine tool can be used for a set of tasks or steps alternatively with different cutting parameter settings defined by the NC program; outdoor systems such as automobiles and wind turbines go through weather and seasonal effects with changing temperature, humidity and other ambient factors. A working regime is determined by one or several working regime parameters. There are usually two types of parameters: 1) the operational parameters set by operators or machine controllers such as speed and load and 2) the working condition variables such as ambient temperature.

The health assessment and prognostics for systems undergo multiple working regimes is more challenging due to the following two reasons. First, different working regimes can potentially change measured system output. For instance, higher load for a rotating component may cause higher vibration and noise. Second, the working regimes can also be different stress levels applied to the system, which will cause the degradation rate to be inconsistent. The working regime issue needs to be carefully handled to insure the effectiveness of the prognostics solution.
Chapter 2. State-of-the-art Review of Related PHM Methods and Techniques

**Working Regimes and System Measurement**

The influence of working regimes on system measurements/features for most engineering systems can be expressed by a monotonic transformation function. For example, for rolling element bearings, it is shown in [Momono and Noda, 1999] that the vibration energy levels at all frequencies in the inspected spectrum monotonically usually increase when the rotation speed increases. For many other systems, the exponential equation is widely used to model the relationship between system measurements and working regimes (e.g. gas turbine [Kurzke, 2003], battery [Rezvanizaniani, Liu, Chen, and Lee, 2014] and anemometer [Cassity, Aven, and Parker, 2012]), which generally has the form

\[
\frac{y}{y_0} = \left( \frac{\iota}{\iota_0} \right)^\alpha
\]  

(2.5)

where \( y \) and \( \iota \) are system measurement and working regime parameter respectively, \( y_0 \) is the measurement at the standard condition \( \iota_0 \) and \( \alpha \) is a constant typically determined using experimental data.

**Working Regimes and Degradation Rate**

The influence of working regimes on system measurement is instantaneous, i.e. the measured system output will change right after (or with a short delay due to system dynamics) the change of working regime. Such behavior is known as the fast dynamics of a system. Correspondingly, the slow dynamic of the system describes the relationship between the system degradation and the accumulated stress caused by working regimes [Chelidze, 2002]. In general, within a fixed time period, more accumulated stress (e.g. the system is operated at a high speed or temperature) causes higher degradation rate (figure 2.1).
Chapter 2. State-of-the-art Review of Related PHM Methods and Techniques

![Figure 2.1: The General Relationship between Stress and Expected Life](image)

literature, the models that describe such relationship are known as life-stress models.

The Arrhenius Model is one of the most commonly used models to describe material degradation under the influence of thermal stress (i.e. at different temperature) [Kececioglu, 2002].

\[ R(T) = Ae^{-B/T} \]  

(2.6)

where R is the degradation rate, A and B are material related constant values, and T is the temperature in Kelvin.

The Inverse Power Law (IPL) Model is used to model non-thermal stress types such as vibration and voltage.

\[ L(U) = \frac{1}{KU^n} \]  

(2.7)

where L is the life, U is the stress level, K and n are model coefficients that need to be estimated using experiment data. Correspondingly the accelerated
degradation rate with respect to a constant stress level $U_0$ can be expressed as

$$A(U) = \left(\frac{U}{U_0}\right)^n$$  \hspace{1cm} (2.8)

**The Eyring Model**  When both thermal and non-thermal stress types are present, a combination of different stress-life models can be used. Alternatively, the Eyring model also gives a generalized stress-life relationship when multiple stress types are present

$$L(V,U) = \frac{1}{V} e^{A+B/V+CU+DU}$$  \hspace{1cm} (2.9)

where $L$ is the expected life, $V$ is the thermal stress level, $U$ is the non-thermal stress level and $A$, $B$, $C$, $D$ are constants.

**The Accumulative Damage Model**  The stress-life models give the relationship between the overall life and the stress history. Moreover, the accumulative damage model [Zhao and Elsayed, 2005] describes the RUL over time when the stress levels are changed. The two rules the model introduces are

- The degradation rate at any time instance is only determined by the current stress level.

- At the time instance when the stress level is changed, the health condition of the system remains the same while the degradation rate changes according to the new stress level. The health condition can be expressed as the amount of degradation or the probability of failure.

Figure 2.2 further illustrates the degradation process described by the accumulative damage model. The plot on the left shows the degradation process under constant stress levels (i.e. working regimes); the plot on the right shows that when the stress level changes during usage at time $t$, the degradation rate
is changed according to the new stress level but the amount of degradation remains the same.

**Miner’s Rule** is widely used in modeling metal fatigue life under different stress levels [Shimokawa and Tanaka, 1980]. Miner’s rule can be considered as one example of the accumulative damage model. Given $k$ different stress levels $S_1, S_2, \cdots, S_k$ and when only one stress level is applied, the expected life in cycles are $N_1, N_2, \cdots, N_k$. Then when multiple stress levels are applied to a material each with $n_i(S_i)$ cycles, the fatigue failure shall happen when

$$
\sum_{i=1}^{k} \frac{n_i}{N_i} = C \quad (2.10)
$$

where $C$ is a constant between 0.7 and 2.2. When no prior knowledge or data is available, $C$ is usually assumed to be 1.

The Miner’s rule considers the total stress the material can handle is a linear combination of contributions from the applied stress levels. The model does not include the probabilistic nature of failure processes. Nevertheless, the model presents a good approximation of the relationship between stress and fatigue life in many applications.
As a conclusion, various stress-life models have been developed for different type of stress and material. However, most of the models need experiment data before hand for determining the model coefficients. Moreover, the experiment data required by stress-life models usually takes long time to collect because it requires multiple run-to-failure tests, even when accelerated life testing techniques are used. In this thesis work, simplified stress-life models will be used with less coefficients so that it can be estimated using streamed online data.

**Prognostic Methods in Literature**

To tackle the dynamic working regime issue, the common approaches for health assessment and prognostics include 1) separate models for each possible regime and 2) normalize system measurement under dynamic working regimes to a fixed working regime (see the initial discussion in section 1.2). In [Wang, Yu, Siegel, and Lee, 2008], a prognostics method for airplane engines under different working regimes is developed. Using training data that contains run-to-failure history under different working regimes, degradation trends are modeled separately. During testing, for a given ongoing record, the degradation trend observed so far is compared to the modeled degradation trends in training data and the most similar historical trends are used to predict the RUL of the test engine. In [Chen, Kunche, and Pecht, 2013], the author utilizes a similar similarity based prediction method in an online scenario, where new models for new degradation trend can be established using incremental learning with data collected in real time. Comparing to [Wang, Yu, Siegel, and Lee, 2008], [Chen, Kunche, and Pecht, 2013] does not require all training data to be readily available at the beginning with its online learning capability while both efforts require the working regime to be constant for each individual unit. In [Nuhic
et al., 2013], Support Vector Machine (SVM) is used to estimate the RUL of Lithium-ion batteries considering the environmental and load conditions. The input and output vectors of the required SVM learning data set are generated by processing the measured data through different load tests. In [Zhang, Yao, and Jiang, 2014], a degradation assessment and life prediction for electro-hydraulic servo valve (EHSV) are presented. The erosion wear distribution and erosion wear rate of the EHSV under different contaminated oil conditions and working missions are considered. The degradation models are built then according to degradation trends under different erosive wear stages. The performance degradation is assessed using the model and the service time is predicted.

In conclusion, the existing solutions for solving dynamic working regime issues are mostly case specific and require a large amount of training data to develop. As a result, an adaptive and generalized prognostic method for systems under dynamic working regimes that does not require time-consuming algorithm training will greatly improve the applicability and promote wider implementation of PHM algorithms.

2.3 Cloud Computing Paradigm and Applications

2.3.1 Cloud Computing Paradigm

Cloud computing as a new IT paradigm can be considered as an integration of many existing technologies such as computer network, grid computing, distributed computing and virtualization, based on Service-Oriented Architecture (SOA) [Zhang, Cheng, and Boutaba, 2010]. After two years of work and 15 drafts, in 2011 National Institute of Standards and Technology (NIST) published their final definition of cloud computing, which has been widely accepted by industry and academia: [Mell and Grance, 2011]
Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

In addition to this definition, NIST also identifies 5 essential characteristics of cloud computing: on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. The definition also reveals a basic concept in cloud computing: computing as a service, where computing capability is ubiquitous and can be accessed by public on-demand just like other utility services in modern society such as water and electricity.
A cloud platform is usually considered as a 4 layer structure, namely the physical layer, the infrastructure layer, the platform layer and the software layer (figure 2.3). The physical layer consists of a group of physical computer servers. A virtualization technique (figure 2.4) is used to combine all the physical resources and decouple the physical resource from the lower level software such as operating systems. Such separation allows the cloud to treat a group of computer servers as a single pool of hardware resources, and further enables fast and convenient provisioning of virtual machines. The infrastructure layer consists of these virtual machines (instances) created as needed to provide required computing, storage or networking capabilities. The service provided from the infrastructure layer is also referred to as Infrastructure as a Service (IaaS). Some common IaaS providers include Amazon EC2, IBM cloud and GoGrid. The platform layer is based on the virtualized infrastructure and provides a software development environment. Software or applications developed based on the cloud platform are usually more reliable and highly scalable. For instance, Google App Engine is a provider of Platform as a Service (PaaS), the term for the platform layers function, which offers JAVA and Python runtime environment for App developers. Finally, the software layer, contains scalable and on-demand cloud applications, or Software as a Service (SaaS), such as online storage, document editing, social network software, etc. More detailed reviews of cloud services can be found in [Weinhardt et al., 2009; Zhou, Zhang, Zeng, and Qian, 2010].

2.3.2 Survey of Cloud Based Computing Platforms

With its convenient and ubiquitous network accessibility and on-demand scalability, cloud computing is considered to be very suitable for future information systems that host machine data and predictive analytics of machine health.
Many case studies have been reported using cloud computing as the platform for system development in fields including machine health and safety monitoring, asset management, medical and remote patient care, environment monitoring and forecasting, etc.

One benefit of cloud based information systems is that the computational capability (number of CPUs, RAM, storage space, etc.) of the servers or virtual machines can be managed dynamically based on the present demand. In [Sedayao, 2008], a globally distributed web service is migrated into the cloud environment. Cloud computing technology, in this case, enables the web service to automatically initialize the virtual machines hosting the web service in physical servers that are geographically close to the user groups using PlanetLab. Such a mechanism boosts the connection speed for users in different countries and regions. The relatively isolated virtual machines also noticeably increase the robustness of the whole system. Similar utilization of cloud computing can also be found in [Lee, Murray, Hughes, and Joosen, 2010], where a cloud is used to collect data from a network of sensors, and the computational capabilities of the cloud are dynamically managed based on the amount of data flow using Amazon’s CloudWatch service.

Cloud computing also provides a better environment for data processing and analysis services. In [Kurschl and Beer, 2009], a scheme to combine cloud computing with wireless sensor networks using a mechanism called filter-chain is proposed. Besides the boosting of connectivity, the filter-chain mechanism is specially designed for data analysis capabilities in cloud environments. The filters in this case can be considered as cloud based software modules performing data processing and storage tasks, while the chain between filters defines the format of the data flow. A tailings dam monitoring and pre-alarm system is developed by [Sun, Zhang, and Li, 2012] using a cloud computing platform for data storage and failure prognostics. A failure prediction algorithm was
developed based on data collected from the tailing dam to deliver warnings in the cloud server. Cloud computing is also often used as a bridge to gather and combine information from heterogeneous sources. Examples can be found in [McGregor, 2011; Liu, Wang, and Liu, 2012] for medical care service and power grid monitoring respectively. In these systems, information is collected from different sources as well as different subsystems with proprietary protocols. In the cloud, information is organized systematically, then provided to users. Some of the cloud based applications also mention the utilization of a knowledge base (such as a knowledge base of diseases and symptoms) that can learn from collected data using data mining techniques, and be used as reference for future problems.

**Whole System Image Exchange** Standardized algorithms can be used as a repository and for each application, the suitable algorithm can be quickly selected and combined into a workflow to handle data analysis in the cloud platform. Due to the tendency of self-protection of different commercialized cloud platforms and application developers, the standardization across the public cloud may encounter strong resistance. As a consequence [Dudley and Butte, 2010] proposed another mechanism for deploying workflow in the cloud environment, named whole system snapshot exchange (WSSE). WSSE copies the whole system including OS configuration and applications into a digital image and can be shared by researchers easily in the same cloud platform. WSSE can be performed in data level, system level or service level. WSSE is able to reproduce the scientific analysis and workflows without concerns of standardization and workflow construction. However, due to lack of reconfigurability, WSSE is more suitable for scientific researches with objectives clearly defined. An example of WSSE is BIO-LINUX, which provides system images for rapid provisioning of bioinformatics functionalities.
However, aside from data sharing and reporting abilities, the reported cloud based systems and projects still lack systematically organized data analysis capabilities. Cloud based data analytics and machine learning currently are limited in scientific research and very difficult to apply to real industrial cases. One of the reasons is that data processing algorithms deployed in the cloud are usually designed for specific applications, which cannot be shared to solve similar problems in a public cloud. The \textit{ad hoc} development model is effective for relatively simple cases (e.g. limited machine models and/or working conditions) but becomes impractical for many real world situations where working regimes vary. To enable \textit{Analytics as a Service} and effectively tackle industrial big data, a systematically designed, unified framework is needed for an autonomous and adaptive machine health monitoring and prognostics.

\subsection*{2.3.3 Reconfigurable PHM Platforms}

The value of the cloud computing platform is more utilized when it is used to support an adaptive PHM solution so that machine units of different types and applications can be connected and data analyzed.

In literature, the research on automated PHM platform is rather scarce. A reconfigurable PHM platform is proposed in [Liao and Lee, 2010] and the adaptation is realized using a toolbox of PHM algorithms, from which the most suitable algorithm will be automatically picked by the Quality Function Deployment (QFD) algorithm. In [Chen et al., 2012] a generic PHM software structure is presented based on the .NET framework and a collection of algorithm modules. The platform is designed under similar idea that with modularized PHM algorithms, different functions can be quickly selected and combined into a workflow for a specific application. However, in practice it is difficult to develop a fully automated and reliable module selection algorithm for different
applications. Other related work Zhou et al., 2005; Han and Yang, 2006 focuses mostly on the structure and enabling technologies (e.g. WEB, sensor networks, etc.) of the platform rather than the adaptation and automation of the PHM algorithm. Despite those early research efforts, much work is still required for developing a fully adaptive and automated PHM platform.
Chapter 3

Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

This chapter presents the main adaptive prognostic methodology for engineering systems under dynamic working regimes. The methodology assumes no prior knowledge or training data of the system is available and attempts to extract degradation information solely from measured condition monitoring data for each unit.

The approach (figure 3.1) contains a generalized state space model for machine degradation and an adaptive and online methodology for real time degradation assessment and prediction using streamed data. The online methodology further consists of an adaptive segmentation method for identification of health stages based on local variation and noise level observed in the time series data, a Canonical Correlation Analysis (CCA) based variable selection algorithm for selecting related working regime parameters and an Adaptive Kalman Filter (AKF) based online filtering method for model identification and prediction.
3.1 Generalized Degradation Model for Dynamic Regime Systems

3.1.1 The State Space Degradation Model and Kalman Filter

Model based prognostics generally consists of two steps: 1) the health assessment step that estimates the health condition (i.e., state) of the system from noisy measurements through online filtering and 2) the prediction step that extrapolates the health states to future time based on the observed trend of degradation to finally estimate the RUL. The stochastic filters such as KF and PF for linear and nonlinear systems consider the degradation process as non-deterministic and, with measurements/features collected in real time at each time instance, perform model based \textit{a priori} prediction and \textit{a posteriori} update.
using Bayesian theory hence are very suitable for online applications and inherently provide natural ways to handle uncertainties in the model, measurement during health assessment and prediction [Celaya, Saxena, and Goebel, 2012]. Stochastic filtering methods have been widely studied and applied to PHM due to the advantages in uncertainty handling, robustness, high interpretability and low data dependency. For instance, KFs have been applied to prognostics of rolling element bearings [Reuben and Mba, 2014] and electric components [Lall, Lowe, and Goebel, 2010]. For nonlinear cases, PFs have been implemented in crack growth estimation for composite materials Chiachío et al., 2015 and degradation prediction for rotating gears [Orchard and Vachtsevanos, 2009].

Generally the state space model for stochastic filtering has the form:

\[ x_{k+1} = f_t(x_k, w_k) \]  
\[ y_k = g_t(x_k, v_k) \]  

The state vector \( x_k \) consists of the health state of the machine and the coefficients such as degradation rate that also needs to be estimated during the filtering process. The health state can be defined as either a continuous health value (e.g. from 1 to 0 with 1 as healthy and 0 as faulty) or discrete states that represent different conditions of the system such as healthy, incipient and severe fault [Orchard and Vachtsevanos, 2009], where the former is more intuitive and the equations are easier to formulate while the latter may require some additional knowledge to define the relationship between different states and system measurements. \( w_k \) and \( v_k \) are process noise and measurement noise respectively. In the state-space model, the degradation is considered as a Markov process, where the current system state vector \( x_k \) only depends on the previous state \( x_{k-1} \) and the system input. If the process function \( f_t(\cdot) \) and output function
$g_k(\cdot)$ is nonlinear and/or the noise is non-Gaussian, nonlinear filters such as Unscented Kalman Filter (UKF) and PF are used for the online model identification. Otherwise, the linear form of the state-space equation is [Kailath, 1974; Chen, 2003]:

$$
\begin{align*}
    x_{k+1} &= \Phi_k x_k + B_k u_k + w_k \\
    y_k &= H_k x_k + v_k
\end{align*}
$$

where

- $x_k$ : state vector (n by 1)
- $y_k$ : measurement vector (m by 1)
- $\Phi_k$ : state transition matrix (n by n)
- $H_k$ : measurement matrix (m by n)

the process noise $w_k$ and measurement noise $v_k$ are considered i.i.d. Gaussian white noise with zero mean

$$
\begin{align*}
    w &\sim N(0, Q) \\
    v &\sim N(0, R)
\end{align*}
$$

and $u_k$ is the external input vector, which in many applications is constant hence can be omitted.

When the state model defined by eq. 3.3 is linear and the noise is white Gaussian, the KF is the optimal estimator in both maximum likelihood and minimizing MSE. Define the state error covariance matrix as $P_k = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T]$. Given the initial conditions $x_0$, $P_0$ and the noise covariance metrics $Q$ and $R$, the KF process can be expressed as follows
At time \( k \) given the measurement \( y_k \) and \( \hat{x}_{k-1}, u_{k-1}, P_{k-1}, Q_k, R_k \), the prediction step:

\[
\hat{x}^-_k = \Phi_k \hat{x}_{k-1} + B u_{k-1}
\]

\[
P^-_k = \Phi_k P_{k-1} \Phi_k^T + Q_k
\]

The update step:

\[
K_k = \frac{P^-_k H_k^T}{H_k P^-_k H_k^T + R_k}
\]

\[
\hat{x}_k = \hat{x}^-_k + K_k (y_k - H_k \hat{x}^-_k)
\]

\[
P_k = (I - K_k H_k) P^-_k
\]

where \( \hat{\cdot}^- \) and \( \hat{\cdot} \) denote the \textit{a priori} and \textit{a posteriori} estimates respectively. In practice, the initial conditions \( x_0, P_0 \) do not significantly affect the result since KF can converge to the true state quickly. However, the performance of the KF largely depends on the values of \( Q \) and \( R \) while an accurate estimation of \( Q \) and \( R \) beforehand is difficult in practice. As a result, AKF methods are developed to update \( Q \) and \( R \) with the measured data. More discussion on AKFs is given in section 3.3. Next, a generalized degradation model for dynamic-regime systems is proposed.

### 3.1.2 Discussion of the General Degradation Process

The generalized degradation model is primarily designed for implementations within a platform where the traditional \textit{ad hoc} algorithm configuration is impractical due to the large amount of machines being connected and the variations in machine condition, working regimes and failure modes that are common in industrial applications. With the degradation model, system health
condition and degradation rate can be estimated using the system measurements/features as system output and the working regime parameters as input (figure 3.2). From a conceptual point of view, the model describes the relationships among three key aspects of a degradation process which include 1) system degradation, 2) system measurements/features and 3) working regimes. In fact, most of the PHM related research attempts to model or quantify one or several relationships among these three aspects. More specifically, the relationships include

- **System degradation on system measurements** When degradation occurs, system measurements drift away from the original value. This is the basic principle based on which the PHM methodologies are developed. However, there shall be one condition for using such principle, which is that the working regime remains the same when comparing system measurements. If not, the working regime influence shall be also taken into consideration.

- **Working regime on system measurements** On the dynamic working regime condition, system measurement is also affected when working regimes change (section 2.2.3).

- **Working regime on system degradation** On the other hand, the working regime are also stress levels that affect the degradation rate of a system (section 2.2.3).

**The Measurement of System Degradation** In order to develop a generalized model that quantitatively describes the aforementioned relationships, the concept of system degradation in the dissertation work and how it is measured needs to be clarified first. The system health condition can be assessed with the amount of degradation that has occurred so far. In order to measure the
Figure 3.2: The Overall Relationship among System Health, Measurement (Output) and Working Regimes (Input) from a System Identification Point of View

System degradation under the dynamic working regime condition, a standard working regime \( \iota_0 \) needs to be established. Then the amount of degradation can be measured by calculating the difference (e.g. the Euclidean distance) in the system measurement/feature at two time instances at the standard working regime (figure 3.3). Assume there is no degradation at the Beginning of Life (BOL) of a system, the degradation at time \( t \) \( \eta_t \) is

\[
\eta_t = y'_t - y'_0
\]  

(3.9)

where \( y' \) denotes the system measurement measured at the standard working regime. Moreover, if the system measurement is normalized to make \( y'_0 = 0 \), then the system degradation can be directly represented by \( y' \) at any time instance.

Model Assumptions  Several assumptions and conditions are required for developing the degradation model

1. **Direct Estimation** The health condition of the system is directly reflected
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

2. Linear Approximation for Change in Measurement

The change observed in the measured feature values $y$ is caused by a linear combination of system degradation and working regime change:

$$\Delta y = \Delta y_{\text{degradation}} + \Delta y_{\text{regime change}} + \epsilon$$

(3.10)

where $\epsilon$ is Gaussian noise with zero mean. Moreover, within a short period of time, $\Delta y_{\text{regime change}}$ has a linear relationship with the corresponding working regime parameters. Given $n$ working regime parameters $\iota \in \mathbb{R}^n$

$$\Delta y_{\text{regime change}} = \sum_{i=1}^{n} \theta_i \Delta \iota^{(i)} + \beta$$

(3.11)

where $\theta_i$ is the scaling factor (coefficient) of the $i^{th}$ working regime parameter $\iota^{(i)}$, $\beta$ is the constant term (i.e. intersection) in the linear relationship.

3. Linear Approximation for Variable Degradation Rate

The amount of...
degradation that happens within a fixed time period $T$ is linearly proportional to the amount of stress $S$ applied to the system during the same period. Given $m$ stress types $S \in \mathbb{R}^m$

$$\Delta y_{\text{degradation}}|_{t_0-T} = \sum_{i=1}^{m} \rho_i S^{(i)} + b$$ (3.12)

where $b$ is the constant intersection and $S^{(p)}$ is the $p^{th}$ accumulated stress.

4. **Fast Dynamics vs. Slow Dynamics** When the working regime is being changed, the change finishes within a short period of time (i.e. not gradual) such that the system identification algorithm is able to differentiate between the fast dynamics caused by the working regime change and the slow dynamics caused by the system degradation [Chelidze, 2002].

**Discussions of Assumptions**

- Assumption on localized linearization for change in system measurement is essential to the proposed method as it raises the possibility of a generalized degradation model when no explicit model is available 
  \textit{a priori}. At a first glance, specially for the assumption given by (3.11), it seems to be an over simplification as many system models have nonlinear forms. However, by closely examining several engineering systems from a wide range of applications, we find that the linear approximation (i.e. using the first term of the Taylor expansion) is adequate given that the coefficients are dynamically adjusted overtime. For instance, the relationship given in eq. 2.5 can be adequately approximated by localized linearization on the condition that $i$ changes within a relatively limited range without causing highly nonlinear behaviors of the system.

- Assumption 3 is a simplified version of the IPL stress-life model discussed in section 2.2.3 where the power parameter $n$ is set to one. The stress...
value $S$ shall be calculated depending on the type of stress for each working regime parameter. According to the accelerated life testing theory, the common stress types include the constant and variable load \cite{Nelson, 2009} (figure 3.4). For constant and slow variable load conditions, the $p^{th}$ accumulated stress $S^{(p)}$ can be calculated as the integral of the corresponding working regime parameter over the time window ($N$ samples). The discrete form is

$$S^{(p)}_k = \sum_{i=k-n+1}^{k} l^{(p)}_{(i)} \Delta t_i$$ (3.13)

and the stress level

$$s^{(p)}_k = \frac{1}{t_k - t_{k-n+1}} \sum_{i=k-n+1}^{k} l^{(p)}_{(i)} \Delta t_i$$ (3.14)

where $\Delta t_i = t_i - t_{i-1}$ is the time interval between two samples. The stress levels for more complex stress conditions need to be calculated with more advanced techniques. For example, in structural health assessment, the direction of the load changes quickly (i.e. fast variable stress) and causes fatigue in the metal components. For such situation, methods such as the rainflow counting algorithm \cite{Downing and Socie, 1982} can be used. The calculated stress levels $s$ are especially useful for the stage based PHM method described in section 3.5.
• The normal degradation (i.e. gradual and without sudden failures) of a system under a constant working regime is well known to be approximately low-order polynomial or exponential and as a consequence can be modeled using localized linearization.

• For assumption 4, if by nature some working regime parameters change gradually, e.g. ambient temperature, one approach to mitigate such issue is to take averages over relatively long time windows (e.g. by month or season) so that the change caused by working regime can be more prominent and easier to differentiate from caused by degradation.

3.1.3 The Generalized Regime-Adaptive Degradation Model

With the discussed assumptions, the degradation process can be expressed using the state space model as

\[
\begin{align*}
\eta_{k+1} &= \eta_k + \Delta t_k \dot{\eta}_k + w1_k \\
\dot{\eta}_{k+1} &= \dot{\eta}_k + w2_k \\
\end{align*}
\]  

(3.15)

where \( \eta \) denotes the amount of degradation of the system (i.e. health condition), the first order derivative \( \dot{\eta} \) is the degradation rate and \( w1, w2 \) are the noise terms associated with \( \eta \) and \( \dot{\eta} \) respectively. The \( \Delta t \) is the time interval between two measurements. With eq. 3.10 and 3.11, the measured feature value (i.e. system output) \( y \) with one working regime parameter \( \iota \) can then be expressed as

\[
y_k = \eta_k + \theta_k \iota_k + v_k \\
\]  

(3.16)

The model can be expanded to more working regime parameters as needed. The corresponding coefficient \( \theta \) for \( \iota \) is also considered as a state parameter.
whose transition is a random walk process

\[ \theta_{k+1} = \theta_k + w_{3k} \tag{3.17} \]

Eq. 3.15 to 3.17 describe the generalized degradation model. In matrix form, it can be expressed as

\[
\begin{bmatrix}
\eta_{k+1} \\
\dot{\eta}_{k+1} \\
\theta_{k+1}
\end{bmatrix} =
\begin{bmatrix}
1 & \Delta t_k & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\eta_k \\
\dot{\eta}_k \\
\theta_k
\end{bmatrix} +
\begin{bmatrix}
w_{1k} \\
w_{2k} \\
w_{3k}
\end{bmatrix}
\]

\[ y_k = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} \eta_k \\ \dot{\eta}_k \end{bmatrix} + v_k \tag{3.18} \]

The native form of the model is a linear degradation with constant degradation rate \( \dot{\eta} \). However, with proper values of Q and R, the KF process is able to update \( \dot{\eta} \) to track the most recent degradation rate for a time varying system.

The noise covariance metrics, given the noise terms are i.i.d. Gaussian, are

\[
Q = \begin{bmatrix}
E(w_{1}^{2}) & 0 & 0 \\
0 & E(w_{2}^{2}) & 0 \\
0 & 0 & E(w_{3}^{2})
\end{bmatrix} =
\begin{bmatrix}
g_{1} & 0 & 0 \\
0 & g_{2} & 0 \\
0 & 0 & g_{3}
\end{bmatrix}
\]

\[ R = E(v^2) = r \tag{3.19} \]

where \( E(\cdot) \) is the expectation operator.

**Estimation of Degradation Rate** The degradation model in eq. 3.18 considers the degradation rate \( \dot{\eta} \) as constant. However, with KF, the estimated \( \dot{\eta} \) will change in correspondence with the true degradation rate when the stress level
changes. As a result, an extra estimation using Weighted Least Squares (WLS) is used to model the relationship between degradation rate $\dot{\eta}_k$ and the stress level $s_k$. The estimation is presented in section 3.4.

**Estimation of Error Covariance Metrics**  An accurate estimation of $Q$ and $R$ is very difficult [Myers, Tapley, et al., 1976], especially when little prior knowledge or training data is available for each machine unit connected to the computing platform. As a result, an AKF method is developed that is optimized for the proposed degradation model, so that the state values can be accurately estimated without any *ad hoc* configuration.

Besides model identification for estimating the state values, there are two more issues that need to be solved. First, the working regime parameters may stay at the same value for a certain period of time (e.g. a machine is used for one task for several days then switched to a new task) which causes the poor excitation problem. An adaptive segmentation method is developed for solving the poor excitation problem by separating the time series measurement data $y$ into discrete stages. By doing this, measurements that belong to one working regime will be grouped into one segment thus mitigates the poor excitation issue. Second, among all the working regime parameters and accumulated stresses, not all of them have significant influence on the system measurement or degradation. If all variables are included, the model will have too many states to estimate which leads to less accuracy and stability. As a result, a variable selection method is developed so that the length of the state vector can stay minimal.

Note that if the system degradation is nonlinear, the proposed model can only be used for working regime normalization, health assessment and short term prediction while is not suitable for accurate long term prediction since the degradation rate $\dot{\eta}$ may change over time for nonlinear systems. In such
scenario, an extra step is needed to perform long term prediction using the normalized and de-noised health state values generated by the proposed model.

3.2 Simulation of Degradation History

Before introducing the online estimation methodology, a simulation for generating machine degradation data under dynamic working conditions is first introduced. It is used to generate degradation data with different noise level, degradation pattern, sample rate and other system properties so that the performance of the developed methodology can be evaluated under different situations.

Simulation of Working Regime Parameters

Given the time stamps \( t = 1, 2, \ldots, 500 \), the simulator first generates a random time series as a working regime parameter \( l_t \) for dynamic working regime situations. At each time instance, the working regime parameter \( l \) is a random number that follows the uniform distribution between 10 and 60\(^1\). In real world applications, one machine may be used for the same task for several cycles before changing tasks. As a result, the working regime parameter may stay at a fixed value for a certain time period. To simulate such behavior, the algorithm in table 3.1 is used.

When \( a_{\text{regime}}, b_{\text{regime}} \) are large, the working regime parameter may stay unchanged for a long period of time. Under such situation, the online estimation algorithm tends to forget the old information, causes the coefficients to windup and become very sensitive to any random noise. The issue is commonly known

\[^{1}\text{The upper and lower boundaries are set to approximate the temperature (°C) of a battery pack in outdoor settings.}\]
TABLE 3.1: Algorithm for Generating a Random Time Series as a Working Regime Variable

Algorithm starts at $t_0 = 1$

Step 1: at time $t_0$
• Generate a random number $x_1$ that follows the uniform distribution between 10 and 60 as the value of the working regime variable
• Generate a positive random integer $x_2$ that follows the uniform distribution between $a_{\text{regime}}$ and $b_{\text{regime}}$ ($a_{\text{regime}} \leq b_{\text{regime}}$) that determines how many time instances the current working regime value will last
• Set $t_{t_0: (t_0 + x_2 - 1)} = x_1$

Step 2: set new $t_0 = t_0 + x_2$, go to step 1 unless $t_0 > 500$

as the poor excitation issue [Stenlund and Gustafsson, 2002]. Such issue is further discussed and handled in section 3.5 as part of the adaptive prognostic methodology.

Simulation of System Measurement with Degradation

According to previous discussions and review (section 2.2.2), considering the degradation patterns as functions with respect to time, they are usually polynomial or exponential.

The degradation rate is based on the IPL life-stress model

$$R(l) = K \left( \frac{l}{l_0} \right)^n$$

(3.20)

where $l_0$ is the normal condition, $K$ is the slope coefficient and $n$ is the power factor.

The linear degradation form is

$$y = R(l)t + \epsilon = K \left( \frac{l}{l_0} \right)^n t + \epsilon$$

(3.21)

where $\epsilon$ is Gaussian noise of zero mean. Figure 3.5 shows the different degradation histories under constant working regime when $l_0 = 25, K = 100, n =$
The nonlinear form is a combination of a linear and exponential function. It is designed to approximate degradation curves such as for the Lithium-ion batteries in [Design of Electric Drive Vehicle Batteries for Long Life and Low Cost; An, Choi, and Kim, 2013] and electronic components in [Lall, Lowe, and Goebel, 2010]. The degradation under constant working regime $l_0$ is

$$y_0 = f(t) = \frac{1}{2} t + \exp \left( \frac{t}{a} \right) + \epsilon$$ \hspace{1cm} (3.22)$$

where $a = 80$.

The degradation rate for any working regime $l$ can be calculated based on eq. 2.8

$$R(l) = K \left( \frac{l}{l_0} \right)^n \frac{dy}{dt}$$ \hspace{1cm} (3.23)$$

$$R(l) = K \left( \frac{l}{l_0} \right)^n \left[ \frac{1}{2} + \frac{1}{a} \exp \left( \frac{t}{a} \right) \right]$$ \hspace{1cm} (3.24)$$

$$R(l) = \frac{1}{2} r(l) + \frac{r(l)}{a} \exp \left( \frac{t}{a} \right)$$ \hspace{1cm} (3.25)$$
where $r(l) = K \left( \frac{l}{l_0} \right)^n$. The degradation history $y$ can then be calculated by integrating $R(l)$ over $t$

$$y = \int R(l) dt$$  \hspace{1cm} (3.26)

$$y = K \left( \frac{l}{l_0} \right)^n \left[ \frac{1}{2} t + \exp \left( \frac{t}{a} \right) \right]$$  \hspace{1cm} (3.27)

Figure 3.6 shows the different degradation histories under constant working regimes for the nonlinear degradation when $l_0 = 25, K = 100, n = 1, Var(\epsilon) = 0$.

**Simulation of System Measurement with Dynamic Working Regimes**

For the effect of dynamic working regimes upon system measurement, the simulator uses the relationship defined by eq. 2.5 to generate the disturbance caused by changing working regimes.

The final system measurement is a combination of the measurement over degradation and dynamic regimes. The combination can be performed in two ways
• Summation. In this case, the amount of change in system measurement caused by the working regime does not depend on the severity of the degradation.

• Multiplication. In this case, the amount of change in system measurement increases as the system degrades further.

The multiplication scenario causes the system to be highly nonlinear and cannot be handled effectively by the proposed degradation model. For real world systems that follow the multiplication rule, nonlinear models and nonlinear filters shall be used which is further discussed in future work in section 6.2.

For the summation case, the final system measurement can be simulated using the following equations. For the linear form

\[ y = K \left( \frac{l}{l_0} \right)^n t + A \left[ \left( \frac{l}{l_0} \right)^q - 1 \right] + \epsilon \]  (3.28)

The nonlinear form

\[ y = K \left( \frac{l}{l_0} \right)^n \left[ \frac{1}{2} t + \exp \left( \frac{t}{a} \right) \right] + A \left[ \left( \frac{l}{l_0} \right)^q - 1 \right] + \epsilon \]  (3.29)

For example, figure 3.7 shows the degradation history generated by eq. 3.28 with \( A = 5000, q = 0.5, a_{regime} = 10, b_{regime} = 50. \)

The working regime \( l \) is also shifted with \( l = l - l_0 \) so that the new \( l_0 = 0 \).

To keep the scale consistent, an optional normalization step can also be performed

\[ y_k = \frac{y_k - \mu_{y_{1:n}}}{\sigma_{y_{1:n}}} \]  (3.30)

\[ y_k = y_k - y_1 \]  (3.31)
where $\mu, \sigma$ are the variable mean and standard deviation.

$y, t, l$ are the outputs from the simulator, which will be used by online estimation methodology to perform health assessment and prediction. The original number of samples is 500, but it can be down sampled to simulate situations where sample rate is low.

### 3.3 Online Model Identification using the Adaptive Kalman Filter

#### 3.3.1 Review of the Existing Adaptive Kalman Filtering Techniques

The influence of different values of the noise covariance metrics $Q$ and $R$ can be analyzed from the KF process given by eq. 3.4 to 3.8. When $Q \uparrow$ and/or $R \downarrow$, the Kalman gain $K_k \uparrow$ which makes the filter weight more on the measurement error $y_k - H_k x_k$ and less on the state transition model. Vise versa, the estimated
state values will be based more on the state transition matrix and less on the measurements when the Kalman gain $K_k$ is small.

The fundamental idea of the adaptation process is to estimate $Q$ and $R$ during the filtering using a relationship between the immeasurable error statistics (i.e. $Q$ and $R$) and the measurable observers [Bundick, 1988] such as system measurement $y_{1:k}$ and estimated states $\hat{x}_{1:k}$ so that the filter is able to estimate the true values of $Q$ and $R$ and further adjust to changes in system parameters and/or error covariance. With the assumptions such as i.i.d. error terms and constant error covariance within $N$ samples (i.e. steady state), the covariance matrices can be either reconstructed using linear relationships derived from the KF process [Myers, Tapley, et al., 1976; Bundick, 1988; Mehra, 1970] or estimated by minimizing a defined cost function [Mohamed and Schwarz, 1999]. Due to the practical advantages such as less dependency of the initial estimation of $Q$ and $R$ and higher accuracy, the AKF methods have been extensively studied since the 1970s, especially for the applications of target tracking and navigation.

The existing AKF methods can be broadly categorized into several classes, namely the error auto-correlation based, the variance-covariance (V-C) matrix based, the empirical based and the objective function based methods. The autocorrelation based method is first proposed by Mehra in 1970 [Mehra, 1970]. The method is based on the fact that for an optimal KF, the estimation error of measurement $v_i = z_i - H\hat{x}_{i|i-1}$ (i.e. the innovation sequence) shall be white Gaussian noise with the autocorrelation sequence $C_i$ being zero or in practice, within a statistical boundary. Subsequently, if the KF is found to be suboptimal, the noise covariance $Q$ and $R$ can be estimated using the calculated autocorrelation sequence $C_{0:N}$. The estimation requires that KF reaches steady state. In the numerical example discussed in the paper, the estimation is triggered after a batch of 950 points. The optimal $Q$ and $R$ are also approximated recursively and as a consequence the computational requirement is relatively high.
Another AKF method is based on the innovation V-C matrix [Mohamed and Schwarz, 1999] that is calculated through averaging within a window of length $N$

$$\hat{C}_i = \frac{1}{N} \sum_{k=i-N+1}^{i} v_k v_k^T$$

(3.32)

then the estimation of $R$ and $Q$ can be derived using the MLE method

$$\hat{R}_k = \hat{C}_k + H_k P_k H_k^T$$

(3.33)

$$\hat{Q}_k = \frac{1}{N} \sum_{j=k-N+1}^{k} \Delta x_j \Delta x_j^T + P_k - \Phi P_{k-1} \Phi^T$$

(3.34)

where the state correction sequence $\Delta x_k = \hat{x}_k - \tilde{x}_k$. Similar results are also given by the empirical based method [Myers, Tapley, et al., 1976].

Generally speaking, the measurement noise covariance matrix $R$ is easier to estimate than $Q$ because it is more directly associated with the measurement $y_k$. The method developed in [Karasalo and Hu, 2011] assumes $R$ is known and focuses on the estimation of $Q$ with a defined cost function penalizing both the local oscillation caused by high $Q$ and bias caused by low $Q$ (fig. 3.8). A second order polynomial is then fitted to a short window using LS for both $y$ and $\hat{x}$ as the center lines so that the bias and oscillation can be calculated. The cost function is proved to be convex with regard of $\hat{x}$ hence an local optima can be found theoretically. An example is given with an one-dimension random-walk state model with a window size of 3. However, the closed form cost function can be very complex and difficult to derive for multi-dimension state vectors with longer window length. Through numerical simulation, it is also found that the performance of the method is sensitive to the window length for the polynomial fitting.
Discussion on Existing AKF Methods

Most of the existing AKF methods are developed for the application of target tracking, whose data has some different characteristics comparing to the data acquired during a condition monitoring process:

- **Sampling Rate and Noise Level** The data sets based on which the previous AKF methods are developed generally have higher sampling rate and relatively lower noise level within the same number of samples comparing to PHM data sets. As a result, for PHM applications, improper estimation of Q and R can cause KF quickly diverge or compromised by system noise (see more detail in section 3.3.3).

- **Steady State** Steady state in KF is defined as the condition when the state error covariance matrix $P_k$ approaches 0. Due to the high sampling rate, KF can usually reach steady state for target tracking applications before the system changes behavior (e.g. object changes direction or speed). However, for a degradation process with a degradation rate increases exponentially, the KF with a linear model may not have enough samples to reach near steady state.

Due to those aspects, when applying the existing AKF methods to degradation data, it is found that they are not able to give a good estimation of Q that ensures the KF to be both accurate and stable. As a result, a new AKF method is proposed in the next section that is optimized for PHM applications. Comparisons are also presented to demonstrate the advantages of the proposed method.
3.3.2 An Optimized Adaptive Kalman Filter for PHM Applications

Generally speaking, the degradation of engineering systems is a gradual process, i.e. it is not expected to observe any sudden change in the degradation rate $\dot{\eta}_k$. As a result, the guideline of finding the optimal Q for a degradation process is that to find the lowest possible value of Q that does not cause any estimation bias (fig. 3.8 b), such that all random noise is accounted to the measurement noise R. Recall the expression of Q in eq. 3.19:

$$Q = \begin{bmatrix} E(w_1^2) & 0 & 0 \\ 0 & E(w_2^2) & 0 \\ 0 & 0 & E(w_3^2) \end{bmatrix} = \begin{bmatrix} q_1 & 0 & 0 \\ 0 & q_2 & 0 \\ 0 & 0 & q_3 \end{bmatrix}$$

The diagonal elements $q_1, q_2, q_3$ represent the variance of the state noise of system health $\eta_k$, degradation rate $\dot{\eta}_k$ and coefficient for the working regime parameter $\theta_k$. If the degradation process is truly linear, then $q_1 = q_2 = 0$. Otherwise, $q_2$ should be none zero to enable the KF to track the change in degradation rate. For the term $q_3$, we expect the relationship defined by eq. 3.11 is approximately time invariant hence set $q_3 = 0$. In practice, having $q_3$ as zero also improves the stability of the estimation under different noise levels. To this point,
we reduce the estimation task to only focusing on \( q^2 \) with \( Q \) as

\[
Q = \begin{bmatrix}
0 & 0 & 0 \\
0 & q^2 & 0 \\
0 & 0 & 0
\end{bmatrix}
\] (3.35)

For the estimation of \( q^2 \), because the system does not reach steady state, the traditional AKF assumption does not apply and as a result, it cannot be estimated based on the relationships derived from the KF process. Instead, \( q^2 \) can be estimated using the desired change rate of \( \dot{\eta} \) for avoiding local bias since \( q^2 \) is proportional to how much change can happen to \( \dot{\eta} \) at each step. Based on eq. 3.15 and 3.16, one has

\[
y'_{k} = y_{k} - \theta_{k} t_{k} = \eta_{k} + v_{k}
\]

\[
\Delta y'_{k} = \Delta t_{k} \dot{\eta}_{k-1} + w_{1_{k-1}} + \Delta v_{k}
\] (3.36)

where \( y' \) is the normalized measurement and the operator \( \Delta (\cdot) = (\cdot)_{k} - (\cdot)_{k-1} \).

Next define \( \Delta^2 (\cdot) = \Delta [\Delta (\cdot)] \). Because noise terms \( w1 \) and \( v \) have zero mean, one has

\[
q^2 \propto E[\Delta \dot{\eta}_{k}] = E(\Delta^2 y'_{k}) / \Delta t_{k}
\] (3.37)

To this end, reader can see that the value of \( q^2 \) is estimated using the second order derivative of the normalized system measurement. Such idea is potentially applicable to other degradation assessment problems with different models.

In practice, the \( \Delta \) operator is replaced by a local LS estimation with a first order polynomial to improve robustness over noise. Define the slope estimator as

\[
\nabla [(\cdot)_{k}, N] = \Theta_{[1,1]}
\] (3.38)
where $\Theta$ is the coefficient matrix generated by LS regression. $\Theta_{[i,j]}$ denotes the element at $i^{th}$ row and $j^{th}$ column of matrix $\Theta$, and

$$
\Theta = (XX^T)^{-1}XY^T
$$

$$
X = \begin{bmatrix} t_{k-N+1} & t_{k-N+2} & \cdots & t_k \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{2 \times N}
$$

$$
Y = \begin{bmatrix} (\cdot)_{k-N+1} & (\cdot)_{k-N+2} & \cdots & (\cdot)_k \end{bmatrix}_{1 \times N}
$$

Similarly, define $\nabla^2[(\cdot)_k, N] = \nabla \{ \nabla[(\cdot)_k, N], N \}$. Then $q_2$ can be calculated by

$$
q_2(k) = \alpha \nabla^2(y'_{k}, N)
$$

where $\alpha$ is a constant scaling factor and $N$ defines the window size. $\alpha$ and $N$ are both free parameters that need to be predefined and their values affect the sensitivity of the algorithm over different noise levels. Based on experience, $N$ can be determined as $1/10$ to $1/5$ of the number of samples for a complete degradation process. If such estimation is not possible, an initial value between 10 and 20 is applicable for most PHM applications. For the value of $\alpha$, a numerical study is presented. The closed form analysis and asymptotic properties of the proposed method remain open for discussion. For the numerical study and also the performance comparison in the next subsection, data generated by the exponential degradation simulator is used. The performance of the AKF can be affected by the shape parameter $a$ for the exponential component, the noise level $Var(\epsilon)$ and the sample rate.

In order to find the optimal $\alpha$, an objective function needs to be established. Here the objective function proposed in [Karasalo and Hu, 2011] is used that penalizes both local bias and oscillation. The objective function is based on a local second order polynomial LS fitting. Similar to the LS fitting given in eq.
3.9, define

\[ h[(\cdot)_k, N] = \Theta^T X \]
\[ \Theta = (XX^T)^{-1}XY^T \]
\[ X = \begin{bmatrix} t_{k-N+1}^2 & t_{k-N+2}^2 & \cdots & t_k^2 \\ t_{k-N+1} & t_{k-N+2} & \cdots & t_k \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{3 \times N} \]
\[ Y = \begin{bmatrix} (\cdot)_{k-N+1} & (\cdot)_{k-N+2} & \cdots & (\cdot)_k \end{bmatrix}_{1 \times N} \quad (3.41) \]

Then the objective function that needs to be minimized is

\[ J_k = \frac{1}{N} \| h(y'_k, N) - h(x_k, N) \|^2 + \frac{1}{N} \| h(x_k, N) - x_k \|^2 \quad (3.42) \]

where \( \| \cdot \| \) is the Euclidean norm and on the right hand side, the first term is the bias and the second is the oscillation. For the evaluation of \( \alpha \), the values of \( J_k \) are averaged within the last period of the data since that is when the degradation starts to accelerate and may cause large estimation bias.

\[ J = \frac{1}{m} \sum_{i=M-m+1}^{M} J_i \quad (3.43) \]

where \( M \) is the total number of samples and \( m = M/5 \).

Figure 3.9 shows the values of the objective function \( J \) with different values of \( \alpha \). The degradation history is generated with eq. 3.29 with \( a_{\text{regime}} = b_{\text{regime}} = 1, K = 100 \). When \( \alpha \) is small \( J \) is large due to the bias term; as \( \alpha \) increases, \( J \) reaches a minimal value then start to slowly increase due to increasing oscillation. Through evaluating different situations, setting \( \alpha \) at 10 can ensure optimal performance of the proposed AKF method.
3.3.3 Performance Comparison of AKF methods

The performance of the proposed AKF is compared to the existing AKF methods regarding stability and denoising capability. According to the discussions in section 3.3.1, the previous AKF methods are mostly developed for the object tracking applications with the assumption of steady state condition. As a result, it has been found that when applying to the degradation data, e.g. the one given in eq. 3.22, the estimated Q tends to be very high which causes large oscillation and the filter to easily diverge. Another issue is that some of the AKF methods are sensitive to the window size N and require a large N to be stable. Among all methods introduced in section 3.3.1, the one proposed by Myers [Myers, Tapley, et al., 1976] is found to be most stable. The estimation of Q is based on the method given by eq. 3.34 with some optimization for online recursive estimation:

\[
\dot{Q}_k = \dot{Q}_{k-1} + \frac{1}{l_q} \left\{ (q_k - \hat{q}_k)^2 - (q_{k-l_q} - \hat{q}_k)^2 + \frac{1}{l_q} (q_k - q_{k-l_q})^2 + \frac{l_q - 1}{l_q} (\Delta_{k-l_q} - \Delta_k) \right\}
\]

\[
q_k = \tilde{x}_k - \Phi_k \tilde{x}_k - 1
\]

\[
\Delta_k = \Phi_k P_{k-1} \Phi_k^T - P_k
\]

\[
\dot{q}_k = \dot{q}_{k-1} + \frac{1}{l_q} (q_k - q_{k-l_q})
\]  

(3.44)
where \( l_q = N \) is the window length and for a vector \( x \), \( x^2 \triangleq xx^T \).

The result comparison is presented in figures 3.10 to 3.13. The data is down-sampled to 250 points and the results are plotted for both methods with window size 50 (i.e. 1/5 of the total sample size) and 100. The window size 100 is not a suitable choice for the down-sampled data since there are 250 points in total, however, it is found that the Myers method cannot generate a stable estimation smaller lower window size. Several observations can be drawn by comparing the results:

- In general the proposed AKF is more stable than the previous methods for PHM applications, especially when the data has low sample rate and high noise level.
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

**Figure 3.12:** Result of the Proposed AKF with N=50

**Figure 3.13:** Result of the Proposed AKF with N=100
• One of the reasons of such difference is that previous methods estimate the entire matrix of Q which can cause any estimation error quickly propagate and the filter to diverge. It can also be observed that when $q_3$ is not controlled to be 0, the corresponding state $\theta$ has large variation and cannot converge to a stable value. In contrast, the proposed method applies constrains to the covariance matrix Q with the domain knowledge of the degradation model and as a consequence, the adaptive estimation procedure is more stable by only estimating $q_2$.

• The true measurement noise variance generated by the simulator is $Var(\epsilon) = Var(v) = 400$. It can be seen that the proposed method has a more accurate estimation of the measurement noise level.

• Sensitivity on window size: comparing to the Myers method, the proposed method is less sensitive on the window size. In fact, it is able to generate stable results with even smaller window size (e.g. N=20). Such robustness is preferable for automated analysis because it requires less ad hoc configuration.

In conclusion, the proposed AKF has a better performance for typical degradation data collected from engineering systems regarding stability to noise, robustness to initial parameter settings and estimation accuracy. The AKF method estimates the degradation, degradation rate of the system in real time with data collected from multiple working regime conditions and the estimated information can be further used for prediction of RUL.
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

3.4 Prediction and Degradation Rate Modeling

3.4.1 Comparison of Different Models

The most common empirical degradation models used for prediction are linear ($y = ax + b$), quadratic ($y = ax^2 + bx + c$) and exponential ($y = ae^{bx} + c$) models. When sufficient historical data and/or domain knowledge is available, the model that has the best fitting to data can be selected. However, such procedure cannot be performed under the assumptions of this dissertation work since no prior knowledge or data can be guaranteed available for each of the in situ machines. As a result, in order to select the optimal model for such generic purpose, the following aspects are considered:

- **General Fitness** In most engineering systems, the degradation has the tendency to accelerate, especially when approaching the EOL. From this perspective, the quadratic or exponential models should be used.

- **Stability** Data collected from in situ systems may have high noise levels and it is important for the prediction method to be stable with noisy data. For quadratic and exponential functions, the estimation of model coefficients can quickly become unstable with noisy data especially when number of samples is low.

To further compare the different models, a simple numeric example is given to evaluate the prediction accuracy and stability. We construct a exponential degradation curve $y' = \exp(t/3)$ with $t = 1, 2, \cdots, 5$ as the considered time window. The measured degradation $y$ is further added with random noise $y = y' + \epsilon$ where identical and independently distributed (i.i.d.) $\epsilon$ follows normal distribution $\mathcal{N}(0, \sigma)$. The models will be used to predict the value of $y$ at
Figure 3.14: Simulated Degradation Data for Prediction Model Comparison

t = 7 with the true value \( y_t = \exp(7/3) \). Figure 3.14 shows the simulated data when \( \sigma = 1 \).

The model coefficients can be estimated using the LS method given in section A.1. For the exponential model, the form \( y = ae^{bx} \) is used and it can be linearized as \( \log(y) = bx + \log(a) \). The prediction errors are the difference between the predicted value and the true value at time \( t = 7 \). With random noise, the degradation data is generated 500 times so that the statistics of the prediction noise can be calculated. Figure 3.15 shows the boxplot of the prediction error from different models when \( \sigma = 3 \). For the simulated data, it is found that the quadratic model gives the best result regarding the median prediction error. The exponential model is very unstable in this case hence delivers poor results even with the original data being generated with an exponential function. The main reason is that, when the time series is short, random noise has a high chance to trick the nonlinear model estimators to output a very different function to the true one because of the high degree of freedom nonlinear models have. It is true that the generated data is relatively flat and when the curvature increases (e.g. use \( y = \exp(t) \) instead) the accuracy of the exponential model will get better but the instability still exists and can cause unreliable results. On the other hand the results from the linear model, although not as accurate as the
ones from quadratic model, are the most stable ones regarding the distribution of the error. Such pattern be further observed when the standard deviation and MSE of the prediction error is plotted in figure 3.16 and 3.17 respectively with different noise levels. It can be seen that when noise level increases, the stability of nonlinear models drops more quickly than the linear model.

In conclusion, in order to develop a generalized prediction method that is applicable to real world industrial applications, the linear model has a good balance between prediction accuracy and stability since data with high noise level may cause the nonlinear models to give unrealistic prediction results. On the other hand, the accuracy of the linear prediction drops quickly for systems whose degradation rate accelerates rapidly as the degradation propagates. For these systems, only short term predictions can be performed with the linear model. Quadratic or exponential models can be used for system with highly nonlinear degradation patterns but shall be carefully treated with constrains to ensure stability. In this dissertation work, only linear prediction is considered.

### 3.4.2 The Short Term Prediction

Since linear approximation is used in the degradation model, a short term prediction can be performed assuming the degradation rate \( \dot{\eta} \) does not change
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

Figure 3.16: Standard Deviation of Prediction Error at Different Noise Levels

Figure 3.17: MSE of Prediction Error at Different Noise Levels
within a certain time window. The prediction uses the last estimated system states to predict future values through extrapolation. For KFs, the uncertainty can also be calculated as [Celaya, Saxena, and Goebel, 2012]

\[
\sigma^2_l = \Phi^2_k P_k + \sum_{i=1}^{l-1} \Phi^2_i Q + Q
\]  

(3.45)

where \( l \) is the number of steps ahead for prediction and \( \sigma \) is the standard deviation of the prediction uncertainty.

However in practice it is found that even when \( Q \) is set to a very low value, the degradation rate \( \dot{\eta} \) can still be compromised by system noise hence the prediction using the last system state can be unstable and inaccurate. Instead, applying the method to real degradation data shows that the prediction is much more stable through extrapolation using a separate regression performed with the last \( N \) values of system health \( \eta \). Here a WLS with exponential forgetting is used for the regression

\[
\Theta = (X W X^T)^{-1} X W Y^T
\]

(3.46)

\[
Y = \begin{bmatrix}
\eta_{k-N+1} & \eta_{k-N+2} & \cdots & \eta_k
\end{bmatrix}_{1 \times N}
\]

(3.47)

where \( X \) is defined as in eq. 3.39, \( W \) is the diagonal weight matrix with diagonal elements \( w_i = \lambda^{-1} \). \( \lambda \) is the forgetting factor, typically between 0.95 and 0.99. For this case the default value is set as \( \lambda = (2N - 1)/2N \).

Then the health value for time \( t' \) in the future can be calculated as

\[
\eta' = \Theta^T \begin{bmatrix} t' & 1 \end{bmatrix}
\]

(3.48)
For a linear system with given threshold $H$, the RUL can then be calculated as

$$t_{rul} = \frac{H - \Theta_{[2,1]}}{\Theta_{[1,1]}} - t_k$$ (3.49)

### 3.4.3 Variable Degradation Rate under Dynamic Working Regimes

In many engineering systems, the degradation rate is also affected by the different stress levels caused by the dynamic working regimes. As discussed in section 3.1.2, the relationship for a stress pattern with $m$ stress types $s \in \mathbb{R}^m$ can be approximated based on eq. 3.12.

$$\dot{\eta} = \sum_{p=1}^{m} \alpha_p s^{(p)} + \beta$$ (3.50)

where $s^{(p)}$ is the $p^{th}$ stress level calculated using eq. 3.14. The coefficients can be estimated using the same WLS procedure as in eq. 3.46.

$$A = (WX^T)^{-1} XWY^T$$ (3.51)

$$X = \begin{bmatrix} s_{k-N+1} & s_{k-N+2} & \cdots & s_k \\ \vdots & \ddots & \cdots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}_{(m+1) \times N}$$ (3.52)

$$Y = \begin{bmatrix} \dot{\eta}_{k-N+1} & \dot{\eta}_{k-N+2} & \cdots & \dot{\eta}_k \end{bmatrix}_{1 \times N}$$ (3.53)

where $s_i$ is a $m$ by 1 column vector

$$s_i = \begin{bmatrix} s_i^{(1)} & s_i^{(2)} & \cdots & s_i^{(m)} \end{bmatrix}^T$$ (3.54)

and

$$A = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_m & \beta \end{bmatrix}_{(m+1) \times 1}^T$$ (3.55)

With the learned coefficients, the prediction results generated by eq. 3.49
Figure 3.18: Adjusted Degradation Rate under Different Stress Levels

can be further adjusted for different usage scenarios in future. The coefficient $\Theta_{[1,1]}$ calculated in eq. 3.46 is the estimated degradation rate using the estimated system health values $[\eta_{k-N+1} \eta_{k-N+2} \ldots \eta_k]$. The corresponding stress level for such estimation can be calculated using a weighted average

$$\hat{s}^{(p)} = \frac{1}{m} \sum_{i=k-N+1}^{k} w_i s_i^{(p)}$$ (3.56)

where $w_i$ has the same value as in eq. 3.46.

The adjusted degradation rate for a given future stress level $s$ can then be calculated as

$$\hat{\eta}(s) = \Theta_{[1,1]} + A_{[1:m]}^T(s - \hat{s})$$ (3.57)

where

$$s = \begin{bmatrix} s^{(1)} & s^{(2)} & \ldots & s^{(m)} \end{bmatrix}^T_{m \times 1}$$ (3.58)

Then a new extrapolation can be constructed using a new line with slope $\hat{\eta}(s)$.

$$\eta = \hat{\eta}(s)t + b'$$ (3.59)
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

The new line shall also pass the point \( \left( t, \Theta^T \left[ \begin{array}{c} t_k \\ 1 \end{array} \right] \right) \) (figure 3.18), based on which the value of \( b' \) is derived as

\[
b' = \left[ \Theta_{[1,1]} - \hat{\eta}(s) \right] t_k + \Theta_{[2,1]} \tag{3.60}
\]

As a result, the predicted value \( \eta \) at given future time \( t \) with the adjusted extrapolation given future stress level \( s \) is

\[
\eta = \hat{\eta}(s) t + \left[ \Theta_{[1,1]} - \hat{\eta}(s) \right] t_k + \Theta_{[2,1]} \tag{3.61}
\]

The adjusted RUL given \( s \)

\[
\hat{t}_{rul}(s) = \frac{H - \left[ \Theta_{[1,1]} - \hat{\eta}(s) \right] t_k - \Theta_{[2,1]}}{\hat{\eta}(s)} - t_k \tag{3.62}
\]

The adjusted RUL can be further used for mission based control optimization. For instance, a typical control optimization problem is to find the optimal work load so that it is high enough for adequate production but not too high to cause excessive stress on the system and shorten machine life. The regime based variable degradation rate estimation is essential to such optimization process by providing information on the affect of different working loads.
3.5 Adaptive Segmentation Algorithm for Streamed Time-series Data

3.5.1 The Poor Excitation Issue for Online Filtering

For the proposed AKF method and many other online estimation methods, they can only deliver good results when the input variables, in this case the working regimes, constantly have a certain level of variation (i.e. sufficient information). This requirement is commonly known as the persistent excitation assumption, which states that there exist a positive integer $s$ and two positive real numbers $a$ and $b$ such that $||\varphi(t)||^2 < b, \forall t$ and $\sum_{i=t+1:t+s} \varphi(i)\varphi(i)^T > aI, \forall t$. The minimum value of $s$ for which there exists $a>0$ is the order of the persistence excitation [Bittanti, Bolzern, and Campi, 1990]. If this assumption cannot be fulfilled, for instance, one or several input variables stay unchanged for a long period of time, the learning algorithm starts to forget the old information while there is not enough new information to compensate, which eventually leads to parameter windup and high sensitivity to any random noise (i.e. the poor excitation issue) [Kulhavý, 1987].

A common solution for the poor excitation issue is the directional forgetting factor for RLS with variable forgetting factor. The idea is that instead of using one constant forgetting factor for all input variables, a forgetting vector is used so that each input variable has a corresponding forgetting factor which can be determined by its variation. For instance, in [Vahidi, Stefanopoulou, and Peng, 2005], the RLS with multiple forgetting factors (forgetting vector) is implemented for an online estimation of vehicle mass. The different forgetting factors are assigned to each predictor variable to balance the different levels of excitation caused by the different change rates of variables. The method is designed for a particular vehicle application with two variables (mass and grade
of the vehicle) on the assumption that proper forgetting factors for each variable can be properly chosen based on prior knowledge. One potential issue of such method is that the separation of forgetting factors also increases the degree of freedom of the overall problem.

In the field of PHM, the poor excitation issue is common and can be caused by two situations

- A machine is used for a single task for a long time which causes the working regime parameters to have very low variation.

- Some working regimes such as seasonal environment conditions change slowly and can also cause the poor excitation issue.

Figure 3.19 shows an example using a generated degradation history. The simulator in eq. 3.28 is used with $a_{\text{regime}} = 10, b_{\text{regime}} = 50, A = 20000, q = 0.5, \text{Var}(\epsilon) = 3600$. The data is also normalized using eq. 3.30. It can be seen that because the working regime parameter does not change often, the filter has difficulties tracking the true degradation status of the system especially when the working regime changes.
Other Practical Issues  Besides the input variable poor excitation issue, some of the other issues during prognostics include

- Many engineering systems start with a low degradation rate at the early stage of life and the degradation rate accelerates toward the EOL. In some other cases, the Signal-to-noise Ratio (SNR) is so low that the information is totally covered by random noise. For example, if the degradation curve has a long flat segment to start with (e.g. in [Reuben and Mba, 2014]), it is very difficult to perform any prognostics even with the exact degradation model due to the low variation in the system measurement and any fluctuation caused by system noise will destabilize the estimation dramatically.

- **Outliers** are common in real world data due to errors of the Data Acquisition (DAQ) or data transmission system.

Considering those issues, an adaptive segmentation method is developed. Using system measurement data streamed in real time (i.e. time series data), the method separates the time series data into discrete stages based on the local variation. By using the stages instead of the original time series data, isolated outliers can be removed and segments with low variations are grouped together as one data sample so that the aforementioned issues can be mitigated.

### 3.5.2 The Adaptive Segmentation Algorithm

As shown in figure 3.20, the segmentation algorithm is essentially an adaptive thresholding method. Adaptive thresholding algorithms are widely used in many research areas for signal processing and anomaly detection [Feil, Abonyi, Nemeth, and Arva, 2005; Li et al., 2014; Palivonaite, Lukoseviciute, and Ragulskis, 2013]. With Gaussian noise presented in a signal, the method attempts
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

Figure 3.20: Adaptive Online Segmentation using Median and \( Q_n \) Values as boundaries

Table 3.2: The Online Segmentation Algorithm

1. Initialize a new segment \( Seg_i \) with start position \( Ts_i = t_k \), length \( n = 0 \), out-of-boundary count \( p = 0 \)
2. Check if the data sample \( y_k \) exists. If yes, get the data sample \( y_k \) at time \( t_k \), add to \( Seg_i \), segment length \( n = n + 1 \); if no, wait for new data
3. Using all the points in \( Seg_i = [y_i, y_{i+1}, \cdots, y_k] \), calculate the median \( M \) and \( Q_n \) values of the current segment
   \[
   Q_n = d \{ |y_i - y_j|; i < j \}_{(k)}
   \]  
   (3.63)
   The boundary of \( Seg_i \) is \([M - \alpha Q_n, M + \alpha Q_n]\)
4. If \( n \geq N \), go to step 5; else \( k = k + 1 \), go to step 2
5. If \( y_k \) is outside the boundary, \( p = p + 1 \)
6. If \( p > P \), set the end position of the current segment \( Seg_i \) at \( Te_i = t_{k-P-1} \), update \( Seg_i = [y_i, y_{i+1}, \cdots, y_{k-P-1}] \), output \( Seg_i, Ts_i, Te_i \)
7. Set \( k = k - P, i = i + 1 \), go to step 1

to determine an optimal threshold based on noise level and signal variance.

The actual algorithms used for this purpose range from statistical tests to Neural Network based methods. Considering the requirement of computational efficiency for the online solution and properties of condition monitoring data, an adaptive segmentation algorithm for degradation history under dynamic working regimes is developed (table 3.2).

Where \( N = 4 \) is the minimal length threshold, \( P = 3 \) is the out-of-boundary count threshold, \( \alpha = 1.5 \) controls the sensitivity and can be between 1 and 2.
Based on the robust statistics theory, $Q_n$ is an equivalent measure of standard deviation of a random variable while more robust when outliers are present [Rousseeuw and Croux, 1993]. It is shown that $Q_n$ measurement has a 50% breakdown point comparing to 0% of the traditional standard deviation definition. As a result the boundary calculated using $Q_n$ is much more stable when outliers are added to the segment. In eq. 3.63, $d = 2.2219$ is a constant to scale the value of $Q_n$ to standard deviation equivalent, and $\{ \cdot \}_{(k)}$ denotes $k^{th}$ order statistics with

$$k = \left( \frac{n}{2} + 1 \right)$$

The calculation of $Q_n$ is very efficient and thus suitable for online analysis. The proposed segmentation method is proven to be effective for a wide range of applications and for most of the degradation data, the only parameter that needs to be tuned is $\alpha$. 

**Figure 3.21:** Adaptive Segmentation Results for Signals with Different Noise Levels
The segmentation method is performed to the measurement data. After each stage is identified, the corresponding working regime values and stress levels are calculated by averaging related values within the start and end time \((T_{s_i}, T_{e_i})\) of the stage. For segment \(Seg_i\) with \(N\) samples \(y_k, \ldots, y_k\), the time stamp, system measurement, working regimes and stress levels can be calculated as

\[
\begin{align*}
    t'_i &= T_{s_i} \quad (3.65) \\
    y'_i &= \frac{1}{N} \sum_{i=k-N+1}^{k} y_i \quad (3.66) \\
    t'_i &= \frac{1}{N} \sum_{i=k-N+1}^{k} t_i \quad (3.67) \\
    s'_i &= \frac{1}{t_k - t_k-N+1} \sum_{i=k-N+1}^{k} t_i \Delta t_i \quad (3.68) \\
    \end{align*}
\]

Note that for a more robust result, the statistic mean can be replaced by median for representing the segment information.

The AKF can then be performed using the information of the stages instead of the raw time series. Continue with the example given in figure 3.19, the segmentation result and new AKF result with segmented stages as input are shown in figures 3.22 and 3.23. It can be seen that after the segmentation, the AKF is able to deliver a more stable and accurate estimation of the system health.

The advantages of adding the developed segmentation algorithm to the prognostics methodology are further summarized as follows

- **Faster calculation with reduced data samples** the segmentation method clusters adjacent data samples at similar levels into one segment and uses
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

Figure 3.22: The Result of the Segmentation Method using a Poorly Excited Signal

Figure 3.23: The Result of the AKF using a Poorly Excited Signal with the Segmentation Method
the median and $Q_n$ statistics to represent each segment, hence greatly reduces the amount of data needs to be processed by AKF.

- **Better handling of noise and outliers** Isolated outliers within each segment do not affect the statistics of the segment. Moreover, the segmentation method is a native way to handle system noise. As shown in figure 3.21, when SNR is high, more segments are identified with more usable information; on the contrary, when SNR is low, little useful information is available hence fewer segments are identified.

- **Better handling of the poor excitation issue** by reducing the samples that have similar working regime values.

- **Extension of prediction horizon** By aggregating similar data samples into segments, the time interval between each reading is increased which means the n-step-ahead prediction can reach a longer time into the future.

### 3.6 Variable Selection for Working Regime Parameters

Without *ad hoc* algorithm tuning, the working regime variables collected from the real world systems are not guaranteed to be related to the degradation process. Additional unrelated variables increase the dimension of the data and the degree of freedom of the degradation model. Moreover, the additional noise carried by the unrelated variables may reduce the accuracy of the PHM method. As a result, a variable selection method is developed for identifying only the related variables before performing the modeling identification tasks. More specifically, the variable selection method is developed for two purposes
• Variable selection for working regime parameters that have influence on system measurement

• Variable selection for stress types that have influence on degradation rate

For model identification algorithms such as LS regression, they cannot properly handle situations where the dimensionality is high or there are correlated variables in the predictor or response space [Qin, 1998]. To tackle such problem, many modified LS methods are developed. For instance, the Lasso regression adds a penalty term for including more variables in the objective function that needs to be minimized, resulting the coefficients of the least correlated variables to be zero. Some other LS variants add a dimension reduction step before applying the LS regression. Partial Least Square (PLS) handles correlated variables by projecting original variables into orthogonal latent variables using CCA then performing one-dimensional regression on each of the latent variables [Qin, 1998]. Using an approach similar to CCA, Reduced Rank Regression (RRR) is another dimension reduction technique to minimize the sum of squared residuals with reduced rank condition [Izenman, 1975]. In [Huang and Jojic, 2011], a variable sifting method is developed. In order to reduce the number of correlated variables the PCA is first applied to the input space and the first k principal components are used as latent variables. The Lasso regression is then used to learn the model. The number k is determined using cross-validation.

However, for real time processing, there exist two major hurdles of the developed methods

• Most of the variable selection methods have some free parameters that need to be optimized and in many cases cross-validation is needed such as in Lasso regression. The cross-validation process is computational intensive and not suitable for online estimation.
• Some methods use variable mapping techniques such as CCA and PCA to handle correlated variables. However, after being transformed to the latent space, the physical meaning associated to each variable is lost. Moreover, in an online situation the transformation matrix changes with new data reading after each iteration and as a result it is difficult to maintain consistency.

As a result, a simple correlation based variable selection method is used for the adaptive prognostic method. The method removes unrelated variables so that the dimension of the input data can be reduced. However, it does not filter the correlated variables and future work is needed for a more optimized variable selection method for the proposed online estimation methodology.

The basic concept of the correlation based variable selection method is to use hypothesis tests to identify variables that have significant correlation with the output. Given two variables \( X = [x_1, x_2, \ldots, x_n] \) and \( Y = [y_1, y_2, \ldots, y_n] \) the correlation value

\[
\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}
\]

(3.70)

where \( \text{cov}(X,Y) \) is the covariance and \( \sigma \) is the standard deviation. To test the null hypothesis that \( \rho_{X,Y} = 0 \), the variable

\[
t = \rho \sqrt{\frac{n-2}{1-\rho^2}}
\]

(3.71)

follows the student t distribution with degree of freedom of \( n - 2 \). As a result, the p value can be calculated as

\[
p_{X,Y} = 2F(-|t|, n - 2)
\]

(3.72)

where \( F(x, n) \) is the cumulative distribution function of student t distribution. When \( p < 0.1 \), it can be inferred that the two variables have non-zero correlation.
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

with 90% confidence.

3.6.1 Selection of Working Regime Parameters on System Measurement

Given system measurement \( y = [y_1, y_2, \cdots, y_k] \) and multivariate working regime \( \iota = [\iota^{(1)}, \iota^{(2)}, \cdots, \iota^{(n)}] \in \mathbb{R}^n \) where \( \iota^{(n)} = [\iota_1^{(n)}, \iota_2^{(n)}, \cdots, \iota_k^{(n)}] \), first order derivatives are calculated first to reduce the influence caused by degradation

\[
\Delta y_i = y_{i+1} - y_i
\]
\[
\Delta \iota_i^{(j)} = \iota_{i+1}^{(j)} - \iota_i^{(j)}
\]

where \( i = 1, 2, \cdots, k \) and \( j = 1, 2, \cdots, n \). Then the p-values \( p_{\Delta \iota_i^{(j)}, \Delta y} \) can be calculated using eq. 3.72 and the working regime variables with \( p_{\Delta \iota_i^{(j)}, \Delta y} < 0.1 \) are selected.

If the adaptive segmentation method is used, the original variable values shall be replaced with the averaged values within each segment.

3.6.2 Selection of Stress Types on Degradation Rate

The stress types are usually in correspondence with working regime parameters. For instance, the speed working parameter also causes speed related stress and temperature parameter causes thermal stress. As a result, if no additional information is available, the stress \( s \) has the same dimension as \( \iota \). Given the original data or the identified segments with the stress types \( s \in \mathbb{R}^n \) and the estimated degradation rate \( \dot{\eta} \), p-values \( p_{s^{(j)}, \dot{\eta}} \) can be used to select the related stress types.
3.7 The Complete Workflow of the Developed Adaptive Prognostic Methodology

To this point the developed adaptive prognostic methodology is complete. With the data acquisition, transmission, feature extraction and other application specific functions already implemented, the health assessment and prognostics can be performed in real time with the following steps (figure 3.1):

1. At time $t_k$, receive a new reading $y_k, t_k$.
2. Feed the new sample to the adaptive segmentation algorithm given in section 3.5.
3. If new segment is detected, output segment information $t'_i, y'_i, t'_i$ as a new sample and add it to memory; else go back to step 1
4. Select related working regime parameters as in section 3.6
5. If the selection is the same as $t_{k-1}$ go to step 7; else go to step 6 for initialization
6. Initialization
   
   (a) Formulate the degradation model given by eq. 3.18 by adding the selected working regime parameters
   
   (b) Initialization of the AKF: set the initial values: $x_0 = 0, P_0 = Q_0 = 0, R = Var(y_{1:5})$
   
   (c) Start AKF from $t_1$
7. Perform the online model identification with AKF
(a) Online filtering: using the most recent system measurement or feature value \( y_k \) streamed in real time to perform the KF process given by eq. 3.4 to 3.8

(b) Estimate \( R_{k+1} \) using eq. 3.33

(c) Estimate \( Q_{k+1} \) using eq. 3.35 and 3.40

8. Perform the short term prediction given in eq. 3.47

9. If prediction under different stress levels is required, select related stress types based on section 3.6 and calculate the adjusted RUL using the method given in section 3.4.3

10. Repeat from step 1

### 3.8 Limitations and Implementation Considerations

#### 3.8.1 Limitations of the Generalized Degradation Model

**Linear Systems** As discussed in the model assumptions (section 3.1.2), the generalized degradation model is established on the condition that the system behavior, which includes the relationship between the working regime and system measurement and the relationship between the working regime and degradation rate, is approximately linear. Although it is necessary to have such assumption in order to develop a generalized model, it inherently introduces some limitations of the overall methodology. Based on experience and engineering knowledge, the linear approximation assumption is more likely to be valid when

- PHM is performed on a large system within which the major components do not always have close and responsive interaction with each other. For
instance, the health and/or fuel efficiency of a cargo ship is based on the components including engines, generators and pumps whose interactions are indirect, which causes the overall system behavior to be less specific to a particular subsystem. Similar situation can also be found in machine tool applications with major components including tool cutter, spindle, feed axis, tool changer, etc.

- The measured system signals are influenced by multiple subsystems such as power consumption, vibration level and noise level, rather than specific components such as the pressure of a engine and the torque of an induction motor.

- The noise level in system measurement is high or the number of data samples is low which causes the detailed system behavior to be difficult to model.

When the above terms cannot be met and/or when situation allows, the linearity of the system shall always be checked to ensure the methodology can be directly applied. More specifically, the possible nonlinear situations to which the methodology cannot be directly applied include

- Nonlinear relationship between working regime and system measurement for a specific system. For instance, the turbine engine data generated by NASA engine simulator consists of 6 working regimes and a number of sensor measurements. Due to the closely interacted system dynamics of the subsystems of the turbine engine, the relationship between sensor readings and working regime is highly nonlinear and as a result the developed methodology cannot be applied to perform online data normalization. For this case, the normalization can be done manually with some training data so that the algorithm can focus on prediction afterwards [Wang, Yu, Siegel, and Lee, 2008].
Chapter 3. Adaptive Prognostic Methodology for Systems under Dynamic Working Regimes

3.8.2 Considerations during Implementation

Parameter Tuning The entire methodology can be directly applied to most engineering systems with very little parameter tuning required. The two parameters that directly affect the overall performance and may need some tuning for better algorithm performance are

- **The window length** $N$ As discussed in section 3.3.2, the optimal value is $1/10$ to $1/5$ of the approximated total number of samples for the complete life of the system. The default value can be set to 10 when no prior knowledge is available.

- **Boundary width** $\alpha$ in the adaptive segmentation algorithm. The default value is 1.5 and is acceptable for most time series data. Nevertheless, the value can be tuned to allow the methodology to have higher or lower sensitivity to local variation.

Preparation of the Working Regime Parameters Using the proposed degradation model, the system measurement $y$ that is influenced by the changing working regime will be normalized in real time to $y'$ with working regime parameter $\iota = 0$. As a result, in practice if the desired working regime level for
the normalization $\iota_0$ is not zero, the values of the working regime parameters can be shifted with $\iota = \iota - \iota_0$ so that the normalized measurement $y'$ is at the correct level. Moreover because of the local linearization, it is recommended that the standard working regime level $\iota_0$ locates approximately at the center of the possible range of the working regime values so that an optimal result can be achieved.

**Data Normalization**  The performance of the algorithm is tested to be optimal when the variation of the data is normal (e.g. approximately within $1E^3$). When the variation of the data is very high and the number of samples is low, the filter may not have enough samples to converge to the real values. For instance, in the developed AKF method, the noise covariance term $q_3$ for the working regime parameter $\theta$ is set to zero to ensure algorithm stability. Having a non-zero value for $q_3$ can quickly render the filter unstable. However, if the true value of $\theta$ is very large and there are not enough samples, the estimated $\hat{\theta}$ can be far off from the true value. Same situation may happen to the degradation rate $\dot{\eta}$ since the update of $q_2$ does not start until there are at least $N$ samples available.

As a mitigation strategy, if the system measurement or working regime values are high enough to cause the slow convergence, it is recommended to normalize the data before applying the developed methodology. The normalization does not need to be accurate (e.g. to zero mean and unit variance) and only some rough estimation of the possible variance of the data is needed.

**Threshold for RUL Prediction**  In order to obtain RUL, a failure threshold is needed. The threshold has to be set based on domain knowledge or historical data. For instance as shown in the first industrial case study, the threshold is
reached when the system measurement increases to a certain percentage (e.g. 130% or 160%) of the original value.
Chapter 4

System Framework for Algorithm Implementation in the Cloud Computing Platform

Cloud computing has brought about new service models and research opportunities in the manufacturing and service industries with advantages in ubiquitous accessibility, convenient scalability and mobility. With the emerging industrial big data prompted by the advent of the Internet of Things (IOT) and the wide implementation of sensor networks, the cloud computing paradigm can be utilized as a hosting platform for autonomous data mining and cognitive learning algorithms. Facing the complications of real world data, a systematically designed framework for implementing cloud based health prognostics is developed (figure 4.1). The main components of the platform include the cloud server and the local agents connected to each system/machine. The cloud based PHM platform is designed to facilitate faster and wider implementation of PHM technologies.
Figure 4.1: Overall Framework for the Cloud based PHM Platform
4.1 Unified System Framework for Cloud based PHM

Figure 4.1 shows the overall framework of a cloud-based prognostics system for providing a low-cost, easy-to-deploy solution for industrial big data collected in factory floors. The system is built based on the infrastructure provided by the cloud server. The cloud server consists of clusters of Virtual Machines (VMs) where most of the data storage, data processing and analysis will be performed. A monitoring agent, which can be an embedded system or a local server, is assigned to each machine or workshop and is used for the communication between machines and the cloud server. For fast accessibility, users are allowed to use a PC or any smart devices (e.g. smart phones and tablets) to login to the cloud server through user interface and acquire authorized data and machine health information. Moreover, all the data processing and analysis algorithms are located in the cloud server in the form of packaged modules. Such algorithm modules follow a unified programming interface and can be called and combined into workflows in any VM with minimum configuration needed. The algorithm module pool, where the APPs are stored and organized, allows the monitoring system to expand to new cases with minimal effort in system development. Moreover, Application Programming Interfaces (APIs) are open to engineers who can also login to the cloud server as module developers and develop customized modules that perform specialized functions for certain machine, or operating conditions. Firewalls are established between each component to assure data and information security.

- **Cloud Server** The cloud server performs most of the computation work and holds data and user information based on instances of VMs. By provisioning VMs on demand, the cloud server hosts the following functional modules

  - **Data Storage** A cluster of VMs is used specially for data storage and
the data will be transferred to data analyzers only when needed. Such centralized data storage management can help promote the data security and stability. On the other hand centralized database is more preferred by potential data mining algorithms for knowledge-discovery purposes.

- **PHM Algorithm Modules** Beside data storage, PHM algorithms are organized and stored in a repository of algorithm modules. As discussed in section 1.3, a PHM solution includes steps that can hardly be generalized. As a result, a collection of PHM tools such feature extraction, signal analysis for common applications such as rotating machinery can be packaged beforehand into algorithm modules to shorten development time. Algorithms are categorized by usage such as signal de-noising, time domain analysis, frequency domain analysis, feature fusion, data visualization, etc. Under each category the input and output of the algorithms are pre-defined so that they can be combined into different workflows with ease (figure 4.2). Each time a new monitoring service is required, a configuration file will be generated by users based on their particular application and the corresponding algorithm modules will be called from the repository to
construct the new service in a separate instance of the VM.

- **User Interface** is an online graphical software platform provided to end-users so that they can access the machine condition and configure the machine monitoring instances by connecting to the database server. For easy-to-use purposes the user interface can be developed for multiple devices including computer, smart phones and tablets. The User Interface has 2 major functions

  * **Reporting and Visualization** It provides visualization tools that deliver requested machine information to corresponding end users. The visualization tools are reconfigurable so that different types of users can focus on retrieving the particular information they are interested in.

  * **Algorithm Configuration** The user interface also allows users
to configure the data processing strategy by uploading application specific information (e.g. component dimensions) and integrating different algorithm modules into the PHM workflow. For instance, different feature extraction techniques and parameter settings can be selected for different applications (figure 4.3). Users can also quickly test different data processing and analysis methods so that an optimal solution for each case can be further discovered.

- **Local Agent** is an embedded system or a local computer that connects the machine or workshop to the cloud. The agent collects data and information from machines through certain protocols then converts the data to the format required by the cloud server. Possible data includes machine condition data from controller, sensor data (vibration, temperature, noise, etc.) from added sensors for critical components and machine maintenance information that can be input manually. Due to the fact that the sensor data sometimes is too large to transfer, necessary signal preprocessing or feature extraction algorithms can be deployed inside the agent to perform local feature extraction and reduce the data size. The local agent plays an important role on the standardization of the cloud based PHM procedure because it provides the original data for all downstream processes. As a result, in section 4.2 the standardized data acquisition and feature extraction procedure is explained in more detail.

With all the functional modules defined, figure 4.4 shows the steps for implementing a cloud based PHM platform to an industrial setting. The cloud platform handles data formatting, transmission and analysis using standardized procedures. The steps that still require manual configuration are
4.2 Standardized Signal Handling Procedure

One common issue in data analysis for machine condition data collected by the cloud based monitoring platform is that the wide variety of data types and data sources. The cloud based analytics platform shall be able to systematically organize the data from heterogeneous sources into a unified format before feeding into the downstream data analysis algorithms. As a result, a feature extraction scheme is designed to synchronize, segment and transform raw data into feature metrics that are then stored in the cloud database. Within such
scheme, the actual feature extraction methods (e.g. what features should be extracted, number of features and how to segment data) still have to be configured for each specific case but the other modules within the platform can be easily configured accordingly as long as the output feature matrix follows the same standardized format. Moreover, the standardization on the feature extraction procedure and database structure also ensures that the whole platform can be quickly implemented for new manufacturing systems.

A cloud agent, which can be an embedded system or a local server, is assigned to each machine for data acquisition and transmission. For machine condition monitoring, raw data is transformed into machine health related characteristics (features) so that more advanced information such as degradation trend and/or failure modes can be further identified. With the purpose of developing a generalized PHM platform, typical raw data generated by manufacturing systems can be categorized into three groups:

- **Controller Data** collected from machine controller or SCADA systems. The data usually has a low sampling rate (below 10Hz) and contains machine operating parameters (e.g. speed, feed rate, position, etc.) and some performance related key variables such as pressure, temperature, current, etc.

- **Sensor Data** collected from add-on sensors with high sampling rate (>1kHz) or high volume such as vibration, acoustic emission and images. Data from add-on sensors adds additional cost to the monitoring system but can provide much more detailed information of machine health such as early detection and failure modes.

- **Textual Data** such as controller message codes, human generated event logs and notes: such data is usually irregular and sparse. In a cloud based condition monitoring system the textual data can be used for triggers (e.g.
model retrain trigger after maintenance) and validation of health assessment results.

As shown in figure 4.5, because of different sampling rates, the collected data is first synchronized and then segmented using a predefined time unit. For machine tools the time unit can be the time during which one step in the NC program is executed or one complete part is made. The time unit can also take a more generalized form such as by day or month depending on the actual application. Features from all segments are then extracted and combined into a matrix. As shown in figure 4.6, the features can also be categorized into three groups depending on their physical meanings, namely working regimes, performance characteristics and machine health values. The working regime features indicates how the machine is used at each time unit so that the health assessment algorithm is able to differentiate machine performance change caused
by machine degradation and regime change. Such capability enables the data analysis to be adaptive to different machine working regimes. For each identified working regime, the performance characteristics features are used for baseline learning and calculation of machine health values. The textual data, if present, can be used as 1) expert knowledge for assigning physical meaning to each failure case, 2) trigger for baseline retrain after each maintenance event and 3) validation for calculated health values and stages.

The proposed feature extraction scheme transforms various formats in raw data into a unified and information-reach form that can be easily stored in the cloud database and further processed by data analysis algorithms. The standardized feature format is also a necessity for developing a generalized platform and more adaptive machine learning algorithms for health assessment and prognostics of machines under dynamic working regimes.

### 4.3 Advantages of the Cloud based PHM Platform

Through a comprehensive integration of the cloud computing paradigm and adaptive PHM methodologies the cloud based machine monitoring system has the potential to bring machine condition monitoring and prognostics to a new level. In addition to the benefits of convenient connectivity, on-demand computing and minimal management, the system also exceeds other monitoring systems with the following advantages:

- **Adaptive and Generic Machine Health Assessment and Prognosis** The system utilizes the developed adaptive prognostic algorithm for in field systems. In addition, for the application specific steps, a repository of algorithm modules with standardized input and output is developed for common engineering systems such as bearings and gearboxes so that most
suitable modules for signal processing, feature extraction and visualization can be selected and combined into a complete PHM solution.

- **Rapid Deployment** Similar to the Whole System Image Exchange concept, every monitoring instance for one application can be quickly created with one pre-configured VM image that contains the required platform (Operating System, active services and essential software), database functions and combination of algorithm modules. Comparing to traditional PHM deployment where the physical server for data collection and analysis has to be setup for each project, the cloud VM based deployment model greatly reduces the time needed for deployment and configuration.
Chapter 5

Case Studies

5.1 A Numerical Case Study

The robustness of the developed AKF comparing to existing AKF methods has been discussed in section 3.3.3. Before presenting the industrial case studies, a numerical case study using the simulator given in section 3.2 is further presented to quantitatively show the improvement of prediction accuracy using the developed method, which include the generalized degradation model, the AKF online estimation method, the working regime based prediction and the variable selection method.

Due to the fact that using the adaptive segmentation algorithm reduces the number of samples, the following initial parameter values are used for the linear degradation simulator in section 3.2.
\[ t = 1, 21, 41, \ldots, 481 \]

\[ Var(\epsilon) = 10 \]

\[ a_{\text{regime}} = b_{\text{regime}} = 1 \]

\[ K = 1 \]

\[ A = 200 \]

\[ q = 0.5 \]

\[ N = 7 \]

...and the threshold for the p value of determining if the working regime variable has a significant influence on the degradation rate (see section 3.6) is 0.1.

The data is used by the developed method to predict the future value of system measurement \( y_{\text{pred}} \) at time \( t_{\text{pred}} = 531 \) with working regime value \( t_{\text{pred}} = t_0 \). To compare prediction performance, three different methods are used:

- **Method 1**: prediction using the raw measurement data
- **Method 2**: prediction using the AKF processed degradation values \( \eta \) WITHOUT the stress based degradation rate adjustment
- **Method 3**: prediction using the AKF processed degradation values \( \eta \) WITH the stress based degradation rate adjustment

The experiment is repeated for 100 times and the MSE of the three prediction methods is used as the criterion for comparison.

The overall experiment procedure is as follows
1. Generate degradation data using the simulator with random noise and working regime. The outputs are system measurement $y$, time $t$, and one working regime parameter $\iota$.

2. Use the degradation data, perform AKF. The outputs are system degradation $\eta$ and degradation rate $\dot{\eta}$.

3. Using the last N points

   (a) Use system measurement $y$ as inputs, predict the value $\hat{y}_{pred}$ and calculate prediction error using eq. 3.48

   (b) Use system degradation $\eta$ as inputs, predict the value $\hat{y}_{pred}$ and calculate prediction error using eq. 3.48

   (c) For adjusted degradation rate

      i. Calculate the stress levels at each time stamp $s_k = \iota_{k-1}$

      ii. Check for correlation: calculate the p value between the degradation rate and stress level

      iii. If the p value is less than 0.1, calculate the adjusted degradation rate $\hat{\eta}(s_{pred})$ where $s_{pred}$ equals to the last working regime value $\iota$ at $t = 481$. Then get the prediction result using eq. 3.62. Otherwise, the prediction result is the same to the second method.

4. Repeat the above steps for 100 times, collect prediction errors and calculate MSE for each method.

### Table 5.1: Comparison of Prediction Errors

<table>
<thead>
<tr>
<th>Use Measurement $y$ Only</th>
<th>Use Normalized Degradation $\eta$</th>
<th>Use Regime based Prediction</th>
<th>Regime Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>.0528</td>
<td>.0268</td>
<td>.0218</td>
</tr>
<tr>
<td>Error std</td>
<td>.1768</td>
<td>.1638</td>
<td>.1478</td>
</tr>
</tbody>
</table>
Figure 5.1: Simulated Degradation Data and AKF Result

Figure 5.2: Distribution of the Prediction MSE
Figure 5.1 shows the simulated degradation data at one experiment and the corresponding result from AKF. Figure 5.2 and table 5.1 shows the distribution of the prediction MSE after repeated 100 experiments. The boxplot shows that with the working regime based normalization and the noise filtering functions provided by the AKF method, the prediction result can be significantly improved. Moreover it shows that the prediction with adjusted degradation rate over working regimes further improves the prediction result.

The Detection of Variable Degradation Rate over Working Regimes  When $K$ is small, the effect of variable degradation rate caused by dynamic working regimes is weak thus the variable selection method based on the p value is less likely to determine such relationship exists using the noisy data. As a result, the adjusted degradation rate based on working regimes is used more and more often as the value of $K$ increases (figure 5.3). Figure 5.4 shows the prediction MSE with different values of $K$. It can be noticed that the performance of the regime based prediction improves as $K$ gets bigger. Moreover, bigger $K$ with the same $A$ also reduces the working regime influence on system measurement and causes the MSE of the first method to drop.
5.2 Machine Tool Condition Monitoring and Prognostics for Band Saw Machines

The methodology is further illustrated using machine tool condition monitoring in band sawing applications. A manufacturing process usually starts with sawing large pieces of material into designated sizes for the downstream contouring. Bandsaws have been widely used for mostly wood cutting for more than a century but only during the past several decades have the bandsaws been used for advanced metal cutting with expensive bandsaw blades. Due to the up-stream nature of the sawing process, the quality and speed of sawing affect the entire production and any error could be propagated to the following steps and result in bad quality products. As a consequence, the health condition of band sawing machine, especially the saw blade which degrades much quicker than other machine components, play significant role in productivity, product quality and production cost. Similar to other machine tools, a sawing machine can be used with a wide range of cutting parameter settings and material types, which can hardly be modeled comprehensively beforehand hence requires the data analysis solution to be adaptive and flexible for all possible situations.
5.2.1 Cloud Platform Implementation

A number of band saw machines with different sizes and configuration have been setup with accelerometers, microphones, thermal couples and pressure gauges installed near the blade guide. For each machine, the add-on sensors and the PLC controller are connected to an Advantech industrial computer as the cloud agent that enables the machine to communicate with the computing and database server. The Advantech computer performs the following functions:

- **Collect data from add-on sensors** The channels and data types are listed in table 5.2.

- **Collect machine operational parameters** from the PLC. Customized software driver is developed to enable the data collection from the PLC controller.

- **DAQ Triggering** By monitoring the blade position variable in PLC controller, the cloud agent is able to detect when the machine starts to cut a material and finishes cutting. The data acquisition from both add-on sensors and PLC will only be triggered when the machine is working. As a result, the data is naturally separated by cut. When one cut is too long (e.g. more than 30 minutes), the data is further separated into smaller segments each contains two-minute long signals. The feature extraction will
be performed for each data segment. As a result, for this application, one sample of feature values (i.e. one row in figure 4.6) is from one segment of data that contains one cut or two-minute data.

- **Data processing and feature extraction** More detail is presented in section 5.2.2.

- After the feature extraction finishes, **upload the new feature values** to the database server through TCP/IP connection for adaptive prognostics. Each machine is assigned with a unique ID so that the server can identify different machines.

On the other hand, the server side hosts the following functions

- **Database Server** The database server uses Microsoft SQL server as the database engine. The cloud agents and web server (for asset management and reporting) communicate (e.g. upload data, query results, etc.) with the database server using the standard Transact-SQL language.

- **Adaptive Prognostics** The developed adaptive prognostics methodology is deployed in the computing server. The algorithm monitors the feature table in the database and is triggered when new features are uploaded from local machines. The prognostic algorithm writes the prognostic result back to the database so that reporting services can visualize them.

- **Web based Asset Management and Reporting Service** A website is developed using Microsoft ASP.NET framework and deployed to the Internet Information Services (IIS) in the server. The website serves as a secured data visualization and reporting tool for users to view the data and reports. On the website, users can securely log-in to the system, review the real-time status of all connected machines and obtain detailed health history from each particular machine (figure 5.6). The health stages are
TABLE 5.2: Raw Data Collected From Sawing Machines

<table>
<thead>
<tr>
<th>PLC controller data</th>
<th>Add-on sensors</th>
<th>Textual data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saw speed, feed rate, material type, material width, current, cut number count</td>
<td>Vibration, acoustic emission (AE) 10KHz</td>
<td>Blade replacement records, blade failure modes N/A</td>
</tr>
<tr>
<td></td>
<td>Coolant temperature, hydraulic pressure, hydraulic oil temperature</td>
<td></td>
</tr>
</tbody>
</table>

also abstracted into three main stages (good, acceptable, bad) based on the relative feature value increase (24%, 48%, 60% respectively based on previously recorded failure data).

5.2.2 Data Processing and Feature Extraction

Upon the finish of one cut or every two minutes, a set of feature values is extracted from the vibration data and sent to the database server. The detailed list of features that are extracted is in table 5.3. Comparing to other machining processes, the machine tool i.e. the saw blade is an elastic metal band with much less rigidity (figure 5.5) and the study on band saw vibration has been scarce because of the relative short history of band sawing for large scale metal
### Table 5.3: Features Extracted From Sawing Machines in the Format of Fig. 4.6

<table>
<thead>
<tr>
<th>Segment identification</th>
<th>Cut number, time stamp, machine ID</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Working regime</strong></td>
<td>Blade speed, feed rate, material hardness, material width</td>
</tr>
<tr>
<td><strong>Performance characteristics</strong></td>
<td>Time domain features for both vibration and AE signals including RMS, kurtosis, crest factor and peak to peak value Octave frequency band energy for both vibration and AE signals Average coolant and hydraulic temperature, average pressure</td>
</tr>
<tr>
<td><strong>Health condition</strong></td>
<td>Health value, health stage (to be calculated by the analysis algorithm) Failure mode based on maintenance log</td>
</tr>
</tbody>
</table>

cutting. Moreover, the tooth pass frequency is usually low (around 100Hz depending on the blade speed) and is covered with noise. As a result, vibration energies are calculated with a fixed frequency width ($\Delta f = 50Hz$) across the entire frequency range (Nyquist frequency $f_s = 5000Hz$) as vibration and acoustic frequency domain features. Given a segment of measured vibration/acoustic data $x$, the result of the Fourier transform $X$ and corresponding frequency $f$ at time $k$, the features are

$$
\hat{y}^{(i)}_k = \frac{1}{50} \sum_{f=(i-1)\times50+1}^{i\times50} |X| \quad \text{with} \quad i = 1, \ldots, 100
$$

(5.1)

Each feature is normalized into a relative change regarding its initial value, which is calculated by averaging the first $n = 50$ values

$$
\hat{y}^{(i)}_k = \frac{y^{(i)}_k - E(y^{(i)}_{1:n})}{E(y^{(i)}_{1:n})}
$$

(5.2)

The features $y^{(i)}_{1:k}$ are then ranked by the correlation value

$$
\rho_i = \frac{E[(y^{(i)}_{1:k}) - E(y^{(i)}_{1:k})](x - E(x))]}{\sigma_{y^{(i)}_{1:k}} \sigma_x}
$$

(5.3)
where \( x = [1, 2, \cdots, k] \). The selected top 10 features can be different between machines and typically locate at around frequency region of 2000Hz and/or 4000Hz. The final system health measurement \( y_k \) is the average of the top 10 features among all vibration and acoustic signals

\[
y_k = \frac{1}{10} \sum_{j=1}^{10} y_k^{(j)}
\]  \hspace{1cm} (5.4)

To this end, the multi-dimensional features are reduced into a single variable \( y \) as the final system feature. It can be seen that the data processing and feature extraction steps are specific for this application and can hardly be generalized or applied to other machine tool applications. However, as discussed in section 4.1, some of the functions used during the feature extraction process, such as the Fourier Transform and the correlation based feature selection method, can be packaged into function modules and reused by other applications with proper configuration. Next, the adaptive prognostic methodology is applied to the feature \( y \) for health assessment and prediction.
5.2.3 Adaptive Prognostics using Extracted Feature

Figure 5.7 shows the feature $y$ from a complete run-to-failure history of a saw blade. The time is defined as number of cuts with each cut lasting approximately 5 minutes for this machine. The machine is being used for two different tasks alternatively and each task is done repeatedly for half a day. As a result, it can be seen that the system feature is affected by the changing working regime and show a step shaped pattern. From the feature values it can also be noticed that occasionally the feature value drops to a very low level due to error of the DAQ system (i.e. outliers). After applying the segmentation method, the values that are from the same working regime are grouped together so that the poor excitation issue can be mitigated.

Figure 5.8 shows the result of applying AKF to the identified segments. Using the variable selection method proposed in section 3.6, the cutting speed is found to be the related working regime parameter among all working regime parameters in table 5.3.

The window size for AKF is set to $N=8$ due to the low number of samples. The failure is reached when the operator notices crack on the blade or sudden increase of noise level, typically when the normalized measurement reaches 0.6 (i.e. 60% increase from the original vibration level). The blade showed in figure 5.8 is being used for two types of materials with different settings of blade rotation speed (i.e. the regime parameter $\iota$). For plotting purpose, the rotation speed is centered by $\iota_k = \iota_k - \iota_1$ so that the normalized measurement $y'$ is close to the original measurement. It can be clearly observed that the system vibration level is affected by the changing working regime and the proposed AKF method effectively removes the regime caused variations from the measurement so that a smooth degradation curve can be acquired. Because the degradation curve for the saw blades is close to linear, figure 5.8 also shows the
results of the predicted RUL when the blade is approaching EOL. The prediction is carried out 1) using the proposed AKF method with the working regime parameter included in the model, 2) using the proposed AKF method without the working regime parameter and 3) using LS regression directly on measurement using eq. 3.47. For comparison purpose, the failure threshold is set as the value of the last measurement. Since the first two cases provide denoising through filtering and in addition case 1) provides working regime based normalization in real time, the prediction is more accurate and stable. Figure 5.9 shows two more RUL prediction results from two other saw blades. The at a given time point, the relative prediction error is

\[ e = \frac{RUL_{pred} - RUL_{true}}{RUL_{true}} \times 100\% \]  

Table 5.4 and 5.5 shows the error percentage and statistics when prediction is performed at approximately 70% and 90% of the total blade life.
In general, because of the denoising and regime normalization, the prediction with AKF and the working regime adaptive model generates a more stable and accurate result. Those advantages are more pronounced when measurement noise is high. Moreover, in practice the smooth health curves generated by the AKF method also help reduce false alarms especially when the blade is near its EOL.

Moreover, no stress types are found to be significantly related to the degradation rate, which means the data does not carry enough information for variable degradation rate modeling. As a result, no adjusted degradation rate can be calculated using the collected data. In fact, such situation is not uncommon for many industrial applications. In factory floors, the machines are typically being used in a more conserved way to ensure reliability. Under this usage pattern, the system measurement (fast dynamic) is more sensitive to the working regime effects than the degradation rate (slow dynamic). The degradation rates caused by these conservative working regimes all fall in the relative low region with little difference in between hence can be difficult to quantitatively differentiate.

As a conclusion, for the band saw case study, the proposed method is able to process data autonomously collected from different types of saw blades, band saw machine and machine tasks without the need of tuning parameter settings for each machine type. In addition, it works in real time with streamed data and is computationally efficient. As a result, it is very suitable for the implementation within a centralized computing server for providing automated PHM solutions for machine fleets.
Chapter 5. Case Studies

Figure 5.8: System Measurement, Working Regime and PHM Results for a Complete Run-to-failure History of a Saw Blade

Figure 5.9: Result Comparison of Different Prediction Methods
Chapter 5. Case Studies

5.3 Health Prognostics for Electrical Components

To further demonstrate the adaptability of the developed methodology on different industrial systems, another case study of a type of electrical components is presented. For this application, the system measurement \( y \) is the voltage of the system and it is found that the voltage value may drift over time and cause system failure. Due to the low sample rate of the measurement, no feature extraction is needed and the measurement \( y \) can be directly used for the prognostic method. The working regime parameters are the environmental measurements \( \iota(1), \iota(2) \) that change periodically following the seasonal pattern. Due to the low number of working regime parameters, both of them are used in the degradation model without the variable selection process.

Due to confidential agreements, the details of the actual system, DAQ implementation, data and failure conditions cannot be published. For the same reason, all variables are normalized. For the adaptive segmentation method, the boundary width is set to \( \alpha = 1 \) due to the relatively high noise level and for AKF, window length is \( N = 7 \).

**Failure History 1** The original data and results of failure history 1 is showed in figure 5.10. From the system measurement, it can be seen that the degradation process is gradual and meanwhile affected by the environmental working regime variables. As a result, the AKF removes the working regime effects to obtain a smoother degradation curve. The RUL results also show that when the regime based normalization is applied, a more stable prediction result can be obtained.

**Failure History 2** The second degradation history is slow at the beginning and a sudden failure occurs at the end. In general, RUL prediction is based on the
Figure 5.10: Data and Results of Failure History 1. From left to right and top to bottom: measurement and identified segments, working regime variables, AKF result, RUL result.
assumption that only gradual degradation develops without sudden failures. As a result, the RUL prediction for this failure history is not applicable. Still, the adaptive segmentation method is able to extract the information out of the noisy data and the AKF method is able to further normalize the measurement with respect to changing working regimes so that the failure can be detected accurately with no false alarms (figure 5.11). On the contrary, if a threshold is set to the original measurement, an early false alarm will be triggered due to high noise level and the fluctuation caused by working regimes.

For a completely different system comparing to the first case study, the developed adaptive prognostic methodology can still be applied and deliver good results with very little reconfiguration needed. The adaptive algorithm successfully extracts stable degradation patterns from system measurements that are affected by dynamic working regimes as well as noise and outliers. Facing time-series data with high noise level, traditional de-noising techniques (e.g. non-parametric smoothing and parametric regression based methods) usually require a time consuming parameter tuning process such as evaluating different window lengths for achieving the optimal result. Although still requires parameter tuning, the proposed method is less sensitive to different parameter values and has a high tolerance on sub-optimal settings. As a result, the algorithm can be quickly configured and deployed for a wide range of engineering systems.

5.4 Conclusions

The case studies further demonstrate the advantages of the developed regime adaptive prognostic methodology. While traditional ad hoc PHM solutions require case-specific tuning to handle dynamic working regime situations, the
Figure 5.11: Data and Results of Failure History 2. From top to bottom: measurement and identified segments, working regime variables, AKF result.
developed methodology is able to autonomously perform the working regime based normalization and prediction for a wide range of engineering systems. The normalization is based on the model information learned in real time through AKF with online monitoring data, and as a result does not require prior knowledge of the monitored system. The AKF mechanism further ensures the method is able to track the changes in system property over time. Through the normalization and filtering, the variations in the system measurement caused by working regime change and noise are reduced hence a more accurate prediction result can be achieved. It is worth to note that the generalized adaptive prognostics methodology can achieve optimized performance when the relationship between the working regime and system measurement is approximately linear. For highly nonlinear cases, the proposed degradation model shall be replaced with nonlinear models, as discussed in the future works (section 6.2).
Chapter 6

Conclusions and Future Work

6.1 Conclusions

Automated and adaptive data analysis is one of the key aspects for effective handling of the rapidly-growing industrial big data. Existing PHM solutions are usually developed for limited machine types and working regimes. As a result, it is always a challenge applying PHM to real world systems where multiple working regimes and machine models are common. The thesis first discusses the research gaps and practical issues for developing an adaptive PHM methodology and implementing it for real world applications. The main difficulties include the dynamic working regimes, the availability of training data and domain knowledge and data quality issues such as noise level and outliers. Facing these issues, an adaptive prognostic methodology is developed that is comprehensive, flexible and robust on handling data from in-field systems with complex conditions.

To develop the methodology, a generalized, empirical based state-space degradation model for systems under dynamic working regimes is first proposed. Because most engineered systems experience gradual degradation (slow
dynamics) over time while response more quickly to regime change (fast dynamics), the model adopts localized linear approximation to describe the degradation process with the assumption that no prior knowledge or training data is available. To accompany the degradation model, an online model identification algorithm based on AKF is further developed. Benchmarking with other adaptive filtering techniques shows that the proposed AKF method is more robust on noise and accurate on estimating the true degradation process and other system properties. Moreover, an adaptive segmentation algorithm and a variable selection method are also developed as upstream data pre-processing steps to further enhance the ability to handle issues such as poor excitation of the working regime variables, noise and outliers and unrelated variables. The entire methodology is optimized for adaptability, autonomous processing and fast-deployment as it does not require time consuming parameter tuning for different industrial applications.

From the implementation point of view, a complete guideline of deploying the methodology to a computing platform is then presented. Computing platforms such as cloud computing provide centralized data acquisition and analysis services that are suitable for large scale condition monitoring for machine fleets and factory floors. The adaptive prognostic methodology is used as the computing engine in the platform and in order to working with the computing engine, a system framework of a PHM platform is proposed that includes a cloud agent for DAQ, local data processing and transmission, a standardized database structure for data streaming and storage and a reporting service. The framework is designed based on standardized functional modules so that it can be quickly configured for different applications.

Two industrial case studies are then presented to further demonstrate the
advantages of the developed adaptive prognostic methodology and its implementation in computing platforms. The first case study is a machine tool condition monitoring application with bandsaw machines. Similar to other machine tool applications, bandsaws can be used for a wide range of materials with different cutting parameters which creates a typical multiple working regime situation. The combinations of different material types, cutting parameters and ambient environment can be too many to be comprehensively modeled beforehand with experiment data. As a result, an adaptive prognostic method is urgently required. The results show that the developed methodology is able to effectively handle sawing data collected from different machine models with little parameter tuning required. The method is able to normalize the feature values with respect to different cutting conditions and further deliver a more stable and accurate prediction of RUL. The second case study uses data collected from a group of electrical components whose system property is affected by the environmental conditions. The data contains high noise level and frequent outliers. Nevertheless, it is shown that the adaptive methodology can be quickly applied to the new application, handle the data quality issues, remove the environmental effects from system measurement and deliver stable prediction results. The case studies show that the developed methodology is able to deliver robust and accurate results with little tuning needed for different applications, which is ideal for facilitating autonomous and real time data analytics in online PHM platforms.

6.2 Future Work

The research work develops a comprehensive framework for PHM under dynamic working regimes. More specifically, it analysis the functions and interactions among the three key aspects of the degradation process which are system
degradation, working regimes as input and system measurements/features as output. Besides the state space model with linear approximation proposed in the dissertation, more explorations can be done under this theoretical framework.

Future work

- **Nonlinear Degradation Models** The developed degradation model is based on localized linear approximations. The linear model is able to work properly when the system is linear or operates within a limited region. Nevertheless different models can be developed for a wider implementation under dynamic working regimes.

  - For some applications, eq. 3.16 with a nonlinear form as
    \[
    y_k = \left(\frac{t_k}{t_0}\right)^\theta \eta_k + v_k
    \]  
    may deliver better overall filtering performance. Then the state-space model becomes nonlinear and nonlinear filters such as UKF and PF shall be applied.

  - For a specific group of systems for which the physical principle is better understood, the degradation model can be replaced with a more specific one. In those cases the AKF linear model identification method will not work well but the adaptive methods for choosing the optimal parameters such as the forgetting factor can still be referred.

- **Online Working Regime Modeling** For highly nonlinear and dynamic systems such as the system in [Wang, Yu, Siegel, and Lee, 2008], parametric regression based modeling for online working regime normalization cannot work properly. For such case, non-parametric and data driven
methods such as Neural Network (NN) and Gaussian Process Regression (GPR) can potentially be used. Similar to the model identification method developed in this thesis, the non-parametric learning algorithm will also need to differentiate the measurement change caused by degradation and working regime change so that their contributions can be quantitatively modeled.

- **Variable Selection** The variable selection method used in the thesis is simple mostly because it is computationally effective and does not require any parameter tuning comparing to other methods. However, it does not handle highly correlated variables and as a result, improved variable selection methods can be developed based on, for instance, the Sliced Inverse Regression (SIR) algorithm.

- **Long Term Prediction and Prediction Uncertainty** The prediction method used in the thesis is a linear prediction based on the assumption that the degradation rate of the system stays approximately the same within a short period of time. Long term predictions can be developed to compensate such limit, especially when the degradation rate starts to rapidly accelerate toward EOL for some systems. For these cases, the trend of the degradation rate can be analyzed so that the algorithm is able to identify whether the degradation is accelerating. However, as discussed in section 3.4.1, prediction robustness shall be examined with caution when nonlinear prediction is implemented. Moreover, the current prediction uncertainty is handled within the Kalman filter framework (eq. 3.45) which mainly considers the variance of the error terms. For degradation prediction, the uncertainty caused by the degradation model, either linear or nonlinear, can be further evaluated through Monte Carlo simulation.
• **Consistency within Machine Fleets** The advantage of the proposed adaptive prognostic method is that it does not require any historical data for training. On the other hand, such feature may cause inconsistency when the algorithm is applied to multiple units within a machine fleet. For instance, the variable selection algorithm may identify different working regime parameters for different units thus the estimated system degradation after normalization may be at different levels. For the bandsaw case study, it is known beforehand that the saw blades have high risk of failure when the normalized system measurement increases 60% of the original value. However, such knowledge is not always available hence it may be difficult to set a consistent failure threshold for all units. To mitigate such issue, a global observer and optimizer can be developed to compare data among different units, and tune the prognostic algorithm for each unit to ensure the health assessment result is comparable among all units.
Appendix A

Recursive Least Squares Filtering

A.1 Least Squares Regression

For least squares estimation, the linear equation is usually given by:

\[
y(k) = \theta^T(k) \varphi(k) + v(k) \quad (A.1)
\]

where

\[
y(k) : \text{output vector (m by 1)} \\
\varphi(k) : \text{input vector (n by 1)} \\
\theta(k) : \text{coefficient matrix (n by m)} \\
v(k) : \text{measurement noise vector (m by 1), usually considered as } v \sim N(0, \sigma^2)
\]

In PHM research and many other time series analysis, the output is a single variable: \( y \in \mathbb{R}^1 \) i.e. \( m = 1 \).

Depending on the different ways of construction of \( \varphi \), the problem can be further classified into [Lindoff, 1997]

- The FIR model: \( \varphi_k = (u_k, ..., u_{k-n}) \) where \( u_k \) is the system input
• The AR process: \( \varphi_k = (y_k, \ldots, y_{k-m}) \)

• The ARX process: \( \varphi_k = (y_k, \ldots, y_{k-m}, u_{k-1}, \ldots, u_{k-n}) \)

The original LS and the Weighted LS (WLS) was first developed by Gauss and other mathematicians in the 19th century [Nievergelt, 2000]. LS and WLS are offline procedures, which generally means they can only be applied after data collection finishes. Comparing to online and recursive algorithms, the offline procedure requires relatively large amount of storage space and processing power and hence is less preferred in practical applications.

Consider the samples collected from time \( k = 1, 2, \cdots, T \), let

\[ \varphi: \text{n by T multivariable time series, } (\varphi(1), \ldots, \varphi(T)) \]

\[ y: \text{m by T multivariable time series, } (y(1), \ldots, y(T)) \]

The LS regression:

\[ \hat{\theta} = (\varphi\varphi^T)^{-1}\varphi y^T \quad \text{(A.2)} \]

The WLS regression:

\[ \hat{\theta} = (\varphi W \varphi^T)^{-1}\varphi W y^T \quad \text{(A.3)} \]

Where \( \varphi \) is a multivariable vector \( W \) is the diagonal weight matrix with the diagonal entries \( d_t \). The WLS can be used when

• Outliers exist in the input vector. The weights can be calculated based on the deviation of each point from the whole population so that the error caused by outliers can be reduced.

• System behavior changes over time (time-varying (TV) system). In this
case, higher weight is given to more recent observations so that the coefficients can be updated and track system change. One of the commonly used forgetting strategies is the exponential forgetting (EF), given by

$$d_k = \lambda^{N-k}$$  \hspace{1cm} (A.4)

where $N$ is the total number of samples and $\lambda \in (0, 1]$ is the exponential forgetting factor, typically chosen between .95 and .99 [Andersson, 1985]. The selection of the value of $\lambda$ is a compromise between stability and bias. The bigger $\lambda$ is, the regression is more stable to the disturbances caused by noise because more information is being considered; meantime, if the properties of the TV system change rapidly, the estimated coefficients will not be able to catch up in time because the LS is considering too much old information, which causes estimation bias. Vice versa, the bias is small but the estimation is more prone to noise when $\lambda$ is small.

### A.2 Recursive Least Squares Estimation

The RLS is developed to model a TV system in real time and is widely used in adaptive filtering, system identification, control and prediction. Comparing to LS, RLS does not need to store all observations and is more computational efficient as it updates the coefficients recursively using the most recent data. Adaptation of RLS is achieved using the forgetting factor $\lambda$ which functionally is the same to the forgetting factor used in the WLS (see ??). Indeed, although being calculated differently, the results of RLS and WLS are identical when same forgetting factor is used [Lindoff, 1997].
The RLS algorithm [Shu-hung and So, 2005; Paleologu, Benesty, and Ciochina, 2008]:

\[
\hat{\theta}(k) = \hat{\theta}(k - 1) + K(k)e(k) \tag{A.5}
\]

\[
e(k) = y(k) - \hat{\theta}^T(k - 1)\varphi(k) \tag{A.6}
\]

\[
K(k) = \frac{P(k - 1)\varphi(k)}{\lambda(k - 1) + \varphi^T(k)P(k - 1)\varphi(k)} \tag{A.7}
\]

\[
P(k) = \frac{P(k - 1) - K(k)\varphi^T(k)P(k - 1)}{\lambda(k - 1)} \tag{A.8}
\]

Note that if we consider the variation of \(\theta\) is a random walk process, then eq. A.1 can also be written in the state space form [Stenlund and Gustafsson, 2002]

\[
\theta_{k+1} = \theta_k + w_k
\]

\[
y_k = \varphi_k^T\theta_k + v_k \tag{A.9}
\]

which means that RLS can be considered as a special case of KF, with \(\Phi = I\) and \(H_k = \varphi_k^T\). [Guo and Ljung, 1995] discusses in detail the connections and similarities between RLS and KF.

### A.3 Recursive Least Squares with Variable Forgetting Factor (VFF)

The selection of the forgetting factor is a compromise between tracking stability and speed (lower speed causes bias). For TV systems, the system parameters as well as the noise level can vary during the whole lifespan and as a result, the most proper way for such dynamic estimation is to use RLS with VFF. Several VFF RLS methods have been proposed by comparing the MSE of \(a\ priori\) and
Appendix A. Recursive Least Squares Filtering

*a posteriori* estimations. Out of those methods, the one proposed in [Paleologu, Benesty, and Ciochina, 2008] has been shown to have a better tracking stability and speed. Continue with the RLS method given in eq. A.5 to eq. A.8, the proposed method in [Paleologu, Benesty, and Ciochina, 2008] is given as follows:

The *a priori* error $e(k)$ is given by eq. A.6. The *a posteriori* error

$$
\hat{v}(k) = y(k) - \hat{\theta}(k) \ast \varphi(k)
$$

(A.10)

Define

$$
\sigma_e^2 = E(e^T e), \sigma_v^2 = E(v^T v)
$$

(A.11)

And

$$
q(k) = \varphi^T(k) P(k - 1) \varphi(k)
$$

(A.12)

$$
\sigma_q^2 = E(q^T q)
$$

(A.13)

Then the forgetting factor $\lambda(k)$ is given by:

If $\hat{\sigma}_e(k) \leq \gamma \hat{\sigma}_v(k), 1 < \gamma \leq 2$ then $\lambda(n) = \lambda_{max}$. Otherwise,

$$
\lambda(k) = \min \left\{ \frac{\hat{\sigma}_q(k) \hat{\sigma}_v(k)}{\xi + |\hat{\sigma}_e(k) - \hat{\sigma}_v(k)|}, \lambda_{max} \right\}
$$

(A.14)

where $\xi$ is a small constant that prevents the division by zero and $\lambda_{max}$ is very close or equal to one. The estimates $\hat{\sigma}$ are given by

$$
\hat{\sigma}_e^2(k) = \alpha \hat{\sigma}_e^2(k - 1) + (1 - \alpha)e^2(k)
$$

(A.15)

$$
\hat{\sigma}_q^2(k) = \alpha \hat{\sigma}_q^2(k - 1) + (1 - \alpha)q^2(k)
$$

(A.16)

$$
\hat{\sigma}_v^2(k) = \beta \hat{\sigma}_v^2(k - 1) + (1 - \beta)v^2(k)
$$

(A.17)

where $\alpha = 1 - 1/(K_\alpha L), \beta = 1 - 1/(K_\beta L), K_\beta > K_\alpha \geq 2$ and $L$ is an integer
defining the length of the exponential window of the error estimation. The selection of the values of $L, K_\alpha, K_\beta$ depends on the actually application. For example, in [Shu-hung and So, 2005], $\alpha \in [.1, .5]$ and $\beta \in [.9, .99]$ are evaluated.

However, since those methods use MSE as the main criterion, they are unable to differentiate between the error caused by instability (oscillation) and one caused by bias (slow tracking). This problem is more pronounced at the beginning of the online estimation when only a few samples available as the method tends to choose a smaller $\lambda$ due to the high MSE while in fact it is desired to have a large $\lambda$ to achieve faster convergence. As a result, RLS with VFF methods usually have a slower convergence rate comparing to KFs (figure A.1).
Bibliography


Mell, Peter and Tim Grance (2011). “The NIST definition of cloud computing”. In:


Sun, Enji, Xingkai Zhang, and Zhongxue Li (2012). “The internet of things (IOT) and cloud computing (CC) based tailings dam monitoring and pre-alarm system in mines”. In: *Safety science* 50.4, pp. 811–815.


