

Monitoring and Prognostics for Broaching Processes by Integrating Process Knowledge

Wenmeng Tian

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Industrial and Systems Engineering

Jaime A. Camelio, Chair
Ran Jin
Lee J. Wells
William H. Woodall

June 16, 2017
Blacksburg, VA

Keywords: Broaching, Degradation, Health Index, Image Processing, Machine Vision System, Physical Process Models, Profile Monitoring, Prognostics, Remaining Useful Life, Tool Wear

Copyright 2017, Wenmeng Tian

Monitoring and Prognostics for Broaching Processes by Integrating Process Knowledge

Wenmeng Tian

ABSTRACT

With the advancement of sensor technology and data processing capacities, various types of high volume data are available for process monitoring and prognostics in manufacturing systems. In a broaching process, a multi-toothed broaching tool removes material from the workpiece by sequential engagement and disengagement of multiple cutting edges. The quality of the final part, including the geometric integrity and surface finish, is highly dependent upon the broaching tool condition. Though there has been a considerable amount of research on tool condition monitoring and prognostics for various machining processes, the broaching process is unique in the following aspects: 1) a broaching process involves multiple cutting edges, which jointly contribute to the final part quality; 2) the resharpening and any other process adjustments to the tool can only be performed with the whole broaching tool or at least a whole segment of the tool replaced.

The overarching goal of this research is to explore how engineering knowledge can be used to improve process monitoring and prognostics for a complex manufacturing process like broaching. This dissertation addresses the needs for developing new monitoring and prognostics approaches based on various types of data. Specifically, the research effort focuses on 1) the use of in-situ force profile data for real-time process monitoring and fault diagnosis, 2) degradation characterization for broaching processes on an individual component level based on image processing; and 3) system-level degradation modeling and remaining useful life prediction for broaching processes based on multiple images.

Monitoring and Prognostics for Broaching Processes by Integrating Process Knowledge

Wenmeng Tian

GENERAL AUDIENCE ABSTRACT

Big data have been providing both opportunities and challenges for product quality assurance and improvement in modern manufacturing systems. In aerospace industry, broaching processes are one of the most important manufacturing processes as they are used to produce the turbine discs in the jet engine. Nonconforming turbine disc quality, either in terms of compromised surface finish or geometry accuracy, will lead to malfunction or even catastrophic failures in the aircraft engines.

One of the major sources that lead to nonconforming product quality is excessive tool wear accumulation and other abrupt malfunctions of the broaching tools. In broaching processes, multiple cutting edges are sequentially pushed or pulled through the workpiece, and each cutting edge is responsible to shape the workpiece into a specific intermediate shaped contour. Therefore, a broaching process can be regarded as a multistage manufacturing process with variation propagating through the multiple cutting edges.

The overarching goal of this dissertation is to explore how process knowledge can be used to improve process monitoring and prognostics for a complex manufacturing process like broaching. This dissertation focuses on the quality assurance and improvement for broaching processes which includes: 1) timely abrupt process fault detection; 2) tool performance degradation quantification; and 3) remaining tool life prediction, which contributes to both methodological development and practical applications in advanced sensing analytics in manufacturing systems.

TABLE OF CONTENTS

ABSTRACT.....	ii
GENERAL AUDIENCE ABSTRACT.....	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES	viii
LIST OF TABLES	x
1. Introduction	1
1.1 Motivation and Background.....	2
1.2 Significance.....	2
1.3 Research Objectives	3
1.4 Dissertation Outline.....	5
2. Literature Review	6
2.1 Profile Monitoring.....	6
2.2 Multistage Manufacturing Process Monitoring.....	10
2.3 Process Knowledge-based Image Filtering for Tool Wear Characterization.....	12
2.4 System-level Degradation Modeling and Prognostics for Multi-edged Machining Processes	13
3. Experimental Setups for Data Collection and Methodology Validation	17
3.1 Experimental Setup for Real-time Monitoring for Broaching Processes.....	17
3.2 Development of Machine Vision System for Degradation Characterization and Prognostics	18
3.2.1 <i>Experimental Setup</i>	19
3.2.2 <i>Image Acquisition</i>	21

3.2.3	<i>Image Processing</i>	22
3.2.4	<i>Measurement System Analysis</i>	25
4.	Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profile Outputs	27
4.1	Introduction	27
4.2	Multi-stream Extraction from Original Profiles.....	27
4.2.1	<i>Multi-stream Processes</i>	28
4.2.2	<i>Profile Segmentation</i>	28
4.2.3	<i>PLS Modeling</i>	30
4.3	GEMWA Monitoring and Diagnostic Scheme for Global and Local Shift Detection	35
4.3.1	<i>GEWMA Monitoring Scheme for Simultaneous Global and Local Shift Detection</i>	35
4.3.2	<i>The Diagnostic Approach of the Proposed Control Chart</i>	37
4.4	Performance Analysis using Simulation	38
4.5	Case Study.....	46
4.6	Conclusions	48
5.	Process Knowledge Constrained Image Filtering for Tool Wear Characterization ..	49
5.1	Proposed Methodology	49
5.1.1	<i>Notations</i>	49
5.1.2	<i>Image Preprocessing</i>	50
5.1.3	<i>Process Knowledge Constrained Image Filtering</i>	52
5.2	Case Study.....	56

5.2.1	<i>Experimental Setup</i>	56
5.2.2	<i>Image Preprocessing and Process Knowledge Constrained Image Filtering</i> 57	
5.2.3	<i>Performance Comparison</i>	59
5.3	Conclusions	61
6.	Model Integration for System-Level Degradation Modeling in Multi-Edged Machining Processes.....	62
6.1	Acronyms and Abbreviations.....	62
6.2	Notations	62
6.3	Introduction	63
6.4	Methodology Development.....	66
6.4.1	<i>Physical Process Model</i>	66
6.4.2	<i>System-level Degradation Model</i>	69
6.5	Case and Simulation Study	74
6.5.1	<i>Case Study</i>	74
6.5.2	<i>Simulation Study</i>	79
6.6	Conclusions	82
7.	Contributions and Future Research Directions.....	84
7.1	Contributions.....	84
7.1.1	<i>Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profiles</i>	84
7.1.2	<i>Process Knowledge Constrained Image Filtering for Effective Tool Wear Characterization with Noisy Images</i>	84

7.1.3	<i>Model Integration for System-level Degradation Modeling and Prognostics for a Multi-edged Machining Tool</i>	85
7.2	Future Research Directions	86
7.2.1	<i>Process Knowledge Constrained Image Filters for Wear Characterization for General Machining Processes</i>	86
7.2.2	<i>Model Integration for System-level Degradation Modeling for Multi-edged Machining Processes</i>	87
7.2.3	<i>Prognostics-Oriented Inspection Allocation for Multi-edged Machining Tools by Fusing Heterogeneous Data Sources</i>	88
8.	References	89
	Appendix A: Robustness of the proposed control charting system.	109
	Appendix B: Derivation on the posterior probability to determine if region k is noise	111
	Appendix C: Simulation results for various different scenarios in data correlation.....	113

LIST OF FIGURES

Figure 1: Turbine disc (①) is a key component in a jet engine (②) where turbine blades (③) are assembled upon	3
Figure 2: Multistage manufacturing process with non-repeating cyclic profile outputs	9
Figure 3: Cutting force profiles collected from broaching process	10
Figure 4: Experiment setup (adopted from Robertson <i>et al.</i> [31] with permission).....	18
Figure 5: Details of the fixture on the cross-head (adopted from Robertson <i>et al.</i> [31] with permission)	18
Figure 6: Directions of cutting in broaching and relief operations.....	19
Figure 7: Mitutoyo Quick Image microscope and the broach fixture.....	20
Figure 8: Quadrant LED lighting.....	21
Figure 9: Microscope frame (top), block fixtures (left), and adjustment wheel (right)....	22
Figure 10: Image processing procedures	23
Figure 11: Advanced cropping by change-point detection technique	24
Figure 12: Rough-cropped image (left) and advanced-cropped image (right).	24
Figure 13: Cutting edge image after cropping (left) and the final binary image (right)...	25
Figure 14: Profile segmentation.....	29
Figure 15: An illustration of PLS regression model between the profiles at the consecutive stages (redrawn from Höskuldsson [114])	31
Figure 16: The GEWMA chart for shift detection.....	47
Figure 17: The p -values of the Henze-Zirkler's multivariate normality tests	47
Figure 18: Examples of ineffective wear regions	53
Figure 19: An illustration of between-image filtering	55

Figure 20: Sketch of the broach and workpiece.....	57
Figure 21: Images after cropping (a) and thresholding (b)	58
Figure 22: Images after within-image (a) and between-image (b) filtering.....	58
Figure 23: Geometric relationship between the area of wear region and volume loss on the cutting edge	67
Figure 24: Proposed model integration structure	74
Figure 25: Examples of the breakout anomalies in the machined contours (bottom row) and their corresponding healthy contour shapes (top row)	75
Figure 26: Comparison results of absolute value of the mean percentage error.....	78
Figure 27: Comparison of average widths of the 90% CIs	79
Figure 28: Modeling variance over time t_k with the fitted linear relationship (on the left panel) and the residual plot over time t_k (on the right panel).....	80
Figure 29: Comparison results of the mean absolute percentage error.....	81
Figure 30: Comparison of average widths of the 90% CIs	82
Figure 31: Using multiple images to quantify wear on a curved flank face	87
Figure 32: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 2	114
Figure 33: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 3	114
Figure 34: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 4	114

LIST OF TABLES

Table 1: Analysis of Variance (ANOVA) for Gauge R&R Study.....	26
Table 2: Number of PCs extracted from OSFEs.....	40
Table 3: ARL comparison for various types of shifts.....	42
Table 4: Diagnostic accuracy summary when using the GEWMA chart	45
Table 5: List of symbols in notation	50
Table 6: The posterior probability of $R_{e,t}^k$ given the observation of time $t + 1$	55
Table 7: Performance comparison before and after applying proposed image filters	60
Table 8: Simulation scenarios.....	113

1. Introduction

In manufacturing systems, process monitoring and prognostics are most crucial tasks for product quality assurance and system maintenance cost control. During the recent few decades, the advancement of sensing technology and computational power has pushed the frontiers of quality control practices with increasingly high efficiency in data collection, storage, and processing. Various types of data sets, including diverse discrete data, point clouds, signal profiles, and images, become available for system informatics in modern manufacturing environments. These data sets of various types, though complicated in structure and considerable in volume, may contain rich information about the process condition, components' degradation condition, and the final product quality, which poses both opportunities and challenges for quality control practices.

Furthermore, the multiple interdependent components in a manufacturing system add another layer of complication to manufacturing systems analytics. The system components' conditions are inherently dependent with each other due to the physical structure of the system. This leads to high correlation between the sensor data collected from the large number of system components. For example, in a serial multistage manufacturing system, the data collected from each pair of consecutive stages should be highly dependent with each other. This dependency can be used to monitor the process conditions of multistage manufacturing systems.

In addition, various engineering knowledge, such as process kinematics information and physical process models, can be used to guide the statistical modeling for the complex manufacturing systems. Therefore, various types of sensor data collected from multiple components in an interconnected system can be jointly modeled by combining

statistical methods and engineering knowledge for process monitoring and prognostics in manufacturing systems.

1.1 Motivation and Background

This dissertation is motivated by the practical needs in aerospace industry for process monitoring and prognostics, where a broaching process is used to manufacture the complex shaped contours, such as the fir-tree shaped contours in the turbine discs of a jet engine. In a broaching process, the desired contour in a part is shaped through sequential material removal by pushing a series of cutting edges through the workpiece. Each cutting edge is responsible to shape the workpiece into an intermediate geometry, which can be regarded as an incoming product of its down-streaming cutting edges.

For broaching processes, the resharpening and any other adjustments need to be performed with all the cutting edges of the tool (or a tool segment) replaced and thus it is not feasible to make process adjustment and maintenance planning for each individual cutting edge separately. The decisions for process monitoring and prognostics should be made on a system level instead of on an individual component level.

Furthermore, process kinematics knowledge and physical process models will be incorporated in the statistical modeling to enhance both modeling accuracy and engineering interpretability in the proposed research.

1.2 Significance

Turbine discs (as shown in Figure 1) are workpieces with grooves of complex geometry, and they are among the most critical parts in a jet engine. Their geometry integrity and

surface finish play crucial roles in the proper functioning of a jet engine. Broaching processes are most widely used to produce the turbine discs due to their high manufacturing efficiency and low dependency on operator experience. The final part quality of a broached workpiece is highly sensitive to the condition of the multi-edged broaching tool, and thus process condition monitoring for abrupt process change as well as predictive analytics for broaching tool life are of significant concern for quality assurance in a broaching process. However, in current industrial applications, a conservative yet very costly maintenance policy, i.e. to sharpen the broaching tool when a small fixed number of cuts are completed, is often used without taking any consideration of the tool condition.

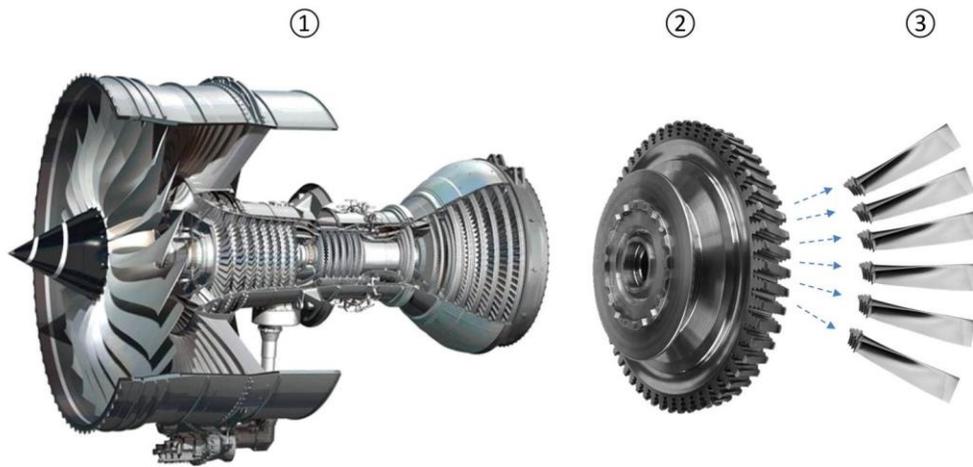


Figure 1: Turbine disc (①) is a key component in a jet engine (②) where turbine blades (③) are assembled upon

1.3 Research Objectives

The overarching goal of the work in this dissertation is to study the statistical monitoring and prognostics by incorporating kinematic information and physical process models for

broaching processes. Though the research in this dissertation was motivated by the broaching process, the general idea of system-level process monitoring and prognostics can be applied in several manufacturing systems or processes with multiple components. Specifically, this dissertation focuses on the following objectives;

Developing a process monitoring methodology to simultaneously detect different types of process shifts in broaching processes: Current profile monitoring approaches mainly focus on repeating cyclic profiles which have an identical distribution over different cycles. Broaching processes, however, generate non-repeating cyclic profiles where the profiles of different cycles do not follow the same baseline distribution. The proposed approach quantifies the relationship between the profiles of consecutive cycles, and simultaneously monitors the standardized one-stage-ahead prediction errors for local shift detection and global mean of the original profiles for global shift detection.

Degradation characterization for single cutting edges based on image processing: Current approaches for image-based degradation characterization are mainly data-driven approaches, i.e. the image processing is purely based on the intensity values of the pixels in the image. A novel approach is proposed to incorporate process knowledge into the data-driven image filtering algorithms by designing process knowledge based constraints.

System-level degradation characterization and remaining useful life prediction for broaching processes: This objective focuses on deriving a general path model for degradation modeling and prognostics for broaching processes considering one single failure mode. The parametric form of the degradation path can be derived from a physical process model.

1.4 Dissertation Outline

This dissertation has eight chapters. Chapter 2 provides a comprehensive literature review for all the following chapters. Chapter 3 describes the experimental setups used in the lab for data collection and methodology validation, including a broaching process setup for process monitoring, and a machine vision system designed for degradation characterization and prognostics [1]. Chapter 4 describes a statistical monitoring and fault diagnosis methodology for multistage manufacturing processes with non-repeating cyclic profiles, which has been published in IISE Transactions [2]. Chapter 5 describes a process knowledge constrained image filtering methodology for tool wear characterization for individual cutting edges, and Chapter 6 introduces a novel methodology for system-level degradation characterization and prognostics by combining the degradation measures of all the cutting edges measured through the integration of a physical process model and statistical degradation models. The general ideas of Chapter 5 and Chapter 6 originated as a conference paper [3] which is currently being extended to two journal papers for submission to the Journal of Manufacturing Science and Engineering and IEEE Transactions on Automation Science and Engineering, respectively. Chapter 7 summarizes the contributions of the work in this dissertation and future research directions, and Chapter 8 lists all the references relevant to the research in the dissertation.

2. Literature Review

This section reviews the relevant literature for the research outlined in this dissertation. Section 2.1 and 2.2 include state-of-the-art methodologies in profile monitoring and multistage process monitoring, which are the relevant literature for multistage process monitoring and diagnostics methods based on non-repeating cyclic profiles in Chapter 4; Section 2.3 summarizes the relevant research in tool wear characterization in machining processes for Chapter 5; Section 2.4 briefly reviews the state-of-the-art degradation modeling and prognostics methodologies as well as the flank wear progression models, which serve as the two crucial pieces of the state-of-the-art approaches for the proposed model integration for system-level prognostics of multi-edged machining processes in Chapter 6.

2.1 Profile Monitoring

Profile monitoring has received considerable attention in the Statistical Process Control (SPC) literature, and it has been used in various applications including calibration process [4], healthcare and public health surveillance [5], and various manufacturing process such as turning [6], stamping [7], and semiconductor manufacturing [8]. In profile monitoring, process quality is characterized by the relationship between a response variable and explanatory variables [9, 10]. Approaches have been developed for monitoring linear profiles, including Kang and Albin [4]; Kim *et al.* [9]; and Zou *et al.* [11]. Also, as a growing number of process variables demonstrate nonlinear relationships, extensive research efforts have been focused on nonlinear profile monitoring. Current nonlinear profile monitoring schemes include parametric (Ding *et al.* [12]; and Jensen and Birch

[13]) and nonparametric approaches (Zou *et al.* [14]; Qiu *et al.* [15]; Paynabar and Jin [16]), most of which have been summarized in Noorossana *et al.* [17].

In recent years, a special class of nonlinear profiles, called cyclic or cycle-based signals, has been studied. Cyclic signals usually refer to signals collected from repeating phenomena or operations in the manufacturing field, such as stamping processes [7, 18-20] and forging processes [21-24]. As the profiles are obtained from repeating operations, they are presumed to follow the same or similar statistical distribution under normal operating conditions. To analyze these profiles, current approaches include signal compression based on wavelet transformation and denoising [18], principal component analysis and clustering methods [21], principal curve method [25], and sparse component analysis method [24]. In terms of SPC, some techniques that have been considered include directionally variant control chart systems for both known and unknown fault detection [22], T^2 control chart based on selected levels of wavelet coefficients [20], process monitoring based on global and local variations in multichannel functional data [23], multichannel profile monitoring and diagnosis based on uncorrelated multilinear principal component analysis [26, 27], and automatic process monitoring technique based on recurrence plot methods [28].

In summary, most approaches in profile monitoring focus mainly on the single-stage processes with profile outputs or multistage processes with repeating cyclic profiles. However, some manufacturing operations consist of multiple stages and the conditions of different stages are characterized as *non-repeating cyclic profiles*, indicating that the signal profiles follow a cyclic pattern while the statistical distributions may vary from cycle to cycle. For example, in a broaching process, the desired contour in a workpiece is

sequentially shaped by multiple teeth. The performance of each tooth or each set of teeth can be reflected by the cutting force collected over time, i.e., a cutting force profile [29, 30]. Each tooth or each set of teeth is considered as a stage in this process. The cutting force profiles of the downstream tooth or set of teeth are largely affected by the amount of material removed by previous teeth and the condition of the currently cutting tooth [31]. As illustrated in Figure 2, when a tooth breaks at the j th stage, the cutting force profile of the j th stage becomes smaller than it should be under normal conditions, due to the changed dimension of the tooth, and the uncompleted material removal left by the j th stage is accomplished by the $(j+1)$ th stage. Thus, the cutting force profile of the $(j+1)$ th stage becomes larger than it is under normal conditions. In Figure 3, four broaching force profiles from different operating conditions are shown. Due to the engagement and disengagement of the multiple teeth on the broach, the profiles collected under normal operating conditions demonstrate cyclic patterns. In addition, given that the material removal rates for different teeth are based on different designed geometries of the teeth, the cutting force profiles at different teeth do not follow the same or similar statistical distributions. Moreover, there are two types of shifts which can both possibly occur in the process, global and local shifts. Global shifts are defined as the process changes which lead to the mean shift at all the stages; local shifts are regarded as the changes which involve the mean shift only at the corresponding stage, such as the shift illustrated in Figure 2. Therefore, the repeating cyclic profile monitoring methods in the literature may not be effective for such a process like broaching.

Even though this work is motivated by broaching process with non-repeating cyclic profiles, it should be noted that non-repeating cyclic profiles are present in various

multistage manufacturing processes, including temperature change over time profiles collected in fruit drying processes [32], CaS level in a SO₂ reduction chemical conversion processes [33], and many other machining operations such as end milling [34].

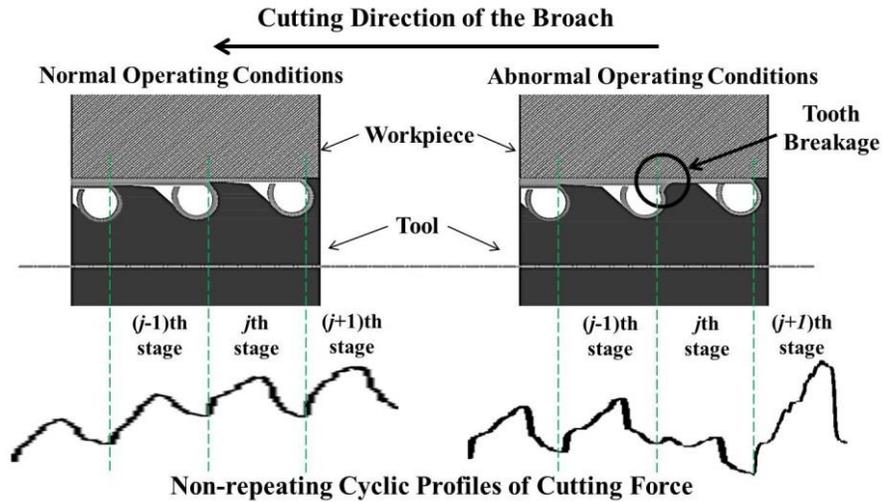


Figure 2: Multistage manufacturing process with non-repeating cyclic profile outputs

In this work, profiles collected from individual stages are called *non-repeating cyclic profiles*. It should be noted that well-studied repeating cyclic operations can be regarded as special cases for the non-repeating cyclic profiles. Additionally, the profiles at all stages are usually continuously collected in temporal order, which results in one profile for each multistage process. The profile containing the information of all the stages in the entire multistage process is called the *original profile* of the process. For example, Figure 3 demonstrates four original profiles collected from the broaching process under different operating conditions.

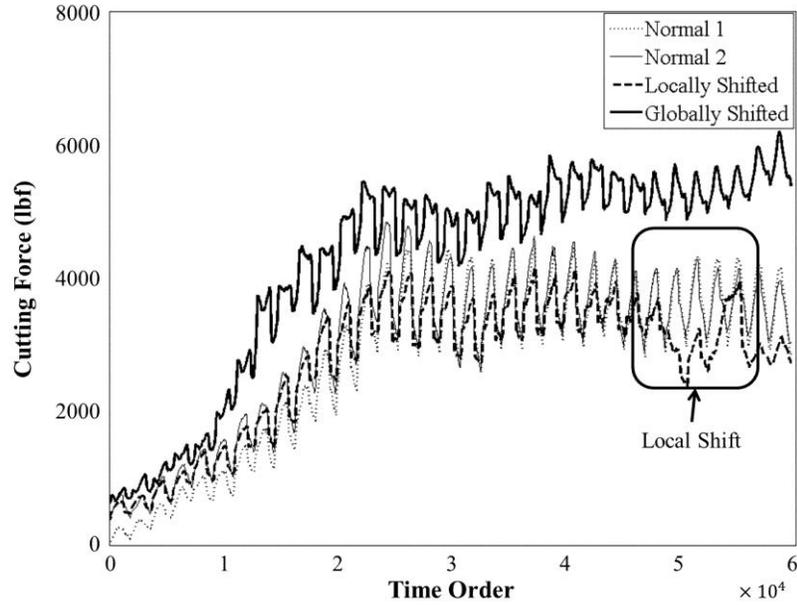


Figure 3: Cutting force profiles collected from broaching process

The objective of the work proposed in Chapter 4 is to detect a shift in multistage manufacturing processes with non-repeating cyclic profiles, and this proposed work focuses on simultaneously detecting two types of process mean shifts, as illustrated in Figure 3, namely the global shifts and the local shifts. The global shifts indicate the process changes resulting in mean shifts occurring at all the stages in the same direction, such as misalignment of the broach tool or large pilot hole in the workpiece. On the other hand, the local shifts are the process changes leading to mean shifts of the profiles at only one or its adjacent stages, which indicates some change of distribution in material removed by those adjacent teeth, such as wear or breakage of a tooth.

2.2 Multistage Manufacturing Process Monitoring

In the monitoring of a non-repeating cyclic profile, a common baseline distribution cannot be identified for profiles at multiple stages. Instead, the correlation between profiles from consecutive stages can be modeled and monitored. In the literature, such

modeling and monitoring methods have been widely used in multistage manufacturing processes, including cause-selecting charts [35, 36] and regression adjustment approaches [37, 38]. Furthermore, Jin and Shi proposed the use of linear state space models to characterize variation propagation between multiple stages [39]. Zantek *et al.* (2006) proposed to use simultaneous CUSUM charts to monitor multiple prediction errors at the same time [40]. A comprehensive review of the approaches and extensions related to multistage process monitoring was given in [41]. Furthermore, Xiang and Tsung proposed an approach to convert the multistage process into a multi-stream process composed of the standardized One-Step ahead Forecast Errors (OSFEs) at all the stages [42]. They adopted the group control charts proposed in [43] to monitor a multi-stream process, which was defined as a process with several streams of outputs. Zou and Tsung developed a directional MEWMA scheme based on generalized likelihood ratio tests for multistage process monitoring [44]. Jin and Liu proposed a control charting system to use regression tree models for serial-parallel multistage manufacturing processes [8]. Zhang *et al.* proposed a Phase I analysis method for multivariate profile data based on functional regression adjustment and functional principal component analysis [45].

All these methods have good capability in identifying potentially shifted stages. However, their methods experienced some limitations to monitor non-repeating cyclic profiles. For example, global shifts cannot be effectively detected by the OSFEs obtained from consecutive stages. The relationship between consecutive stages will not change when all stages experience mean shifts with similar magnitudes in the same direction.

2.3 Process Knowledge-based Image Filtering for Tool Wear Characterization

Extensive research has focused on the use of tool condition monitoring to guarantee product conformity. From these reported works, tool performance degradation is usually characterized by two types of measures, direct and indirect measures. Direct measures focus on quantifying tool wear by directly examining cutting edges, typically through the use of machine vision systems. Highly effective image acquisition and processing methods have been developed to quantify tool wear for various machining processes, as reviewed by Kurada and Bradley [46] and Dutta *et al.* [47]. Indirect measures of tool wear use online process signals; including cutting force, vibration, acoustic emission, and hydraulic pressure as a proxy to monitor and quantify tool condition. These approaches are summarized in Sick [48], Rehorn *et al.* [49], and Siddhpura and Paurobally [50].

As discussed, there is extensive research in directly quantifying tool wear; however, for broaching tool wear characterization, flank wear width, i.e. the maximum width of the wear region on a tooth's relief face, has been mostly used as the wear descriptor for a single cutting edge. The work in broaching tool wear characterization is limited and can be summarized as follows. Mo *et al.* [51] used scanning electron microscopy and chemical composition analysis to measure flank wear on a single cutting edge of a broach. Shi and Gindy [52] developed an on-line indirect tool wear predictive model by using flank wear as the benchmark for a linear broach's wear.

Due to irregular shapes of most wear lands on a broaching tool, flank wear provides very limited tool condition information. Loizou *et al.* [1] proposed to quantify wear for a single cutting edge using the total area covered by the wear region. A Gauge R & R study showed that the proposed approach outperforms the traditional flank wear approach for

wear quantification by comparing the percentage of variance introduced by the measurement system given the same cutting edges to measure.

In addition, current direct wear characterization schemes are mostly data-driven methods, which extract information only based on image intensity. However, there is plenty of process knowledge available in machining processes, such as the location of the possible interaction zone between the workpiece and the cutting edge(s) during the machining, and tool wear progression mechanisms, which could be incorporated in the image filtering procedures to improve wear characterization performance.

The objective of the work in Chapter 5 is to propose novel process knowledge constrained image filtering schemes to identify and eliminate noise in the image in order to obtain the effective wear region(s) on each single cutting edge.

2.4 System-level Degradation Modeling and Prognostics for Multi-edged Machining Processes

2.4.1 Tool Wear Progression Models and Prognostics for Machining Tools

Generally, there are two types of approaches used to model tool wear progression: empirical models and physical process models. Tool wear progression models based on practical experiences and process measurements are categorized as empirical models. These approaches usually select a form of the tool wear progression function and estimate the parameters based on a training data set. Representative research has been reported in [53-55]. One major drawback of the empirical approaches lies in their predefined form of the degradation paths, which cannot be adjusted based on different

types of machining processes or varying machining process parameters, such as material properties, cutting speed, and temperature.

The models that incorporate the physical process mechanisms are classified as the physical process models. These approaches explore the relationship between the tool wear rate and other process parameters, such as cutting speed, temperature, and material properties of cutting tools and workpieces [56-59]. Most of these physical wear progression approaches describe the wear progression in terms of the volume change or the width of flank wear on the cutting edge. For broaching processes, it has been demonstrated that the area of wear region is a robust attainable feature [1] for wear characterization. However, to the best of my knowledge, there are no approaches proposed to model the tool wear progression of the area of the wear region, nor are there any methods on system-level prognostics for multi-edged machining processes.

Currently, most existing studies for Remaining Useful Life (RUL) prediction for machining tools focus on extracting a Health Index (HI) from the process sensor signals to infer the tool wear levels as healthy states, among which Hidden Markovian Models are mostly used [60-63]. Furthermore, various types of other statistical modeling approaches are used to build the statistical relationship between the sensor signals and the tool wear levels, such as Bayesian inferences [64-66] and dynamic Bayesian Network [67], support vector regression [68], neural networks [48, 69-71], and response surface methodology [72, 73]. However, there is very limited research that uses a degradation path model to describe the tool wear progression in terms of the HI extracted from direct image observations from the cutting edges and predict the RULs of the machining tools.

2.4.2 Prognostics Approaches

General prognostics approaches rely on the degradation modeling and RUL prediction based on a predefined threshold value for the corresponding HI. Methodologies for RUL prediction fall into three major categories: data-driven methods, model based methods, and hybrid techniques [74, 75]. Data-driven approaches are the most extensively studied category, and have been comprehensively reviewed in [76, 77]. In general, the degradation of a unit can be characterized by a stochastic process, such as a Wiener process [78-80], a Gamma process [81-83], a Markovian process [84, 85], or a regression based model [86-90] (sometimes also referred to as a general path model [90-92]), and then the constructed degradation models are used to predict future HI values so as to infer when the HI value firstly passes the threshold value.

Model based schemes usually develop the degradation path from first principles, which have been used for applications such as the gear fatigue crack propagation modeling [93, 94] and image contrast degradation modeling [95]. Hybrid methodologies integrate physical process based and data-driven models [96]. Hybrid approaches usually combine the parametric form obtained from the physical process knowledge, and parameter estimation based on data-driven schemes [74]. The approach proposed in this paper falls under the category of hybrid RUL prediction schemes.

In addition, the state-of-the-art RUL prediction approaches have been mostly focused on approaches for units with one single component. There are only very limited efforts proposed for system-level prognostics [74, 97]. Furthermore, most existing research focuses on a unit or system which fails when the HI first exceeds a predefined failure threshold [87, 98], or a random failure threshold which is independent from the corresponding degradation path parameter [99-103]. However, there are scenarios that the

failure threshold distribution is correlated with the corresponding degradation path parameter. The RUL prediction accuracy can be enhanced by taking into consideration the dependency between the random failure threshold and the degradation path parameter, and the proposed methodology uses the correlation structure to update the RUL prediction through combining the historical degradation model parameters and in-situ HI observations.

3. Experimental Setups for Data Collection and Methodology Validation

As the methodologies proposed in this dissertation are motivated by a practiced application in the aerospace industry, experimental setups are built to closely resemble the setups in the industrial settings. To avoid any repeated descriptions on the experimental setups, the major setups built for data collection and methodology validation are briefly introduced in this section, and they will be referred to in the later chapters of the dissertation.

3.1 Experimental Setup for Real-time Monitoring for Broaching Processes

The experimental setup we built is illustrated in Figure 4. It replicates the setup of an industrial broaching process. This setup was modified from a tensile test machine, which could provide a highly controlled displacement of the broach with specified constant cutting speeds.

A 0.75-inch hexagonal broach produced by the duMONT company, LLC, is used in the experiments. The broach is made of high speed steel with 38 teeth and 0.375-inch pitch between consecutive teeth. The 2" × 2" × 1" workpiece is fixed under the cross-head of the tensile test machine as shown in Figure 5. During the cutting process, a real-time cutting force profile can be obtained using a force sensor embedded in the base of the tensile test machine, while the base is moving upwards with a constant speed of three inches per minute. The data collected from this setup are used to validation the effectiveness of the proposed methodology in Section 4.

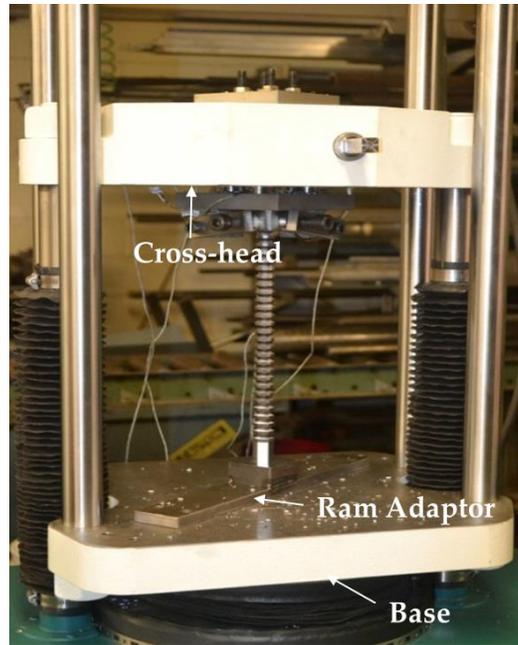


Figure 4: Experiment setup (adopted from Robertson *et al.* [31] with permission)

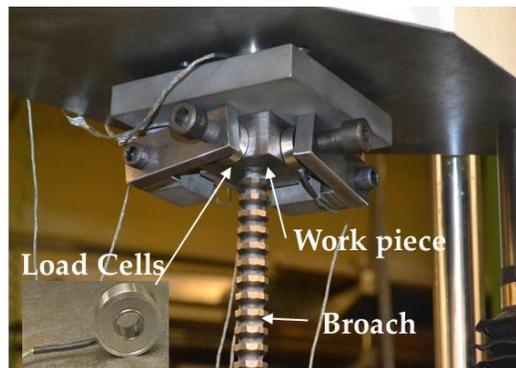


Figure 5: Details of the fixture on the cross-head (adopted from Robertson *et al.* [31] with permission)

3.2 Development of Machine Vision System for Degradation Characterization and Prognostics

In this section, the machine vision system we developed for an automated wear characterization for individual cutting edges is introduced. Image processing procedures are used to identify the wear regions from the noise left by the broaching tool manufacturing processes. Additionally, the area of wear region has been compared with

the traditionally used flank wear width as the degradation measure based on a Gauge R & R study. Most of the contents in this section have been included in [1], which has been completed in collaboration with two other students at Virginia Tech and published in *Journal of Manufacturing Systems*.

It should be noted that this proposed system is designed to quantify tool wear on all kinds of broaches for which texture left by relief operations during broach production is *not parallel* with the cutting direction of broaching. This property applies to a large variety of manufactured broach bars. A couple of examples are shown in Figure 6.

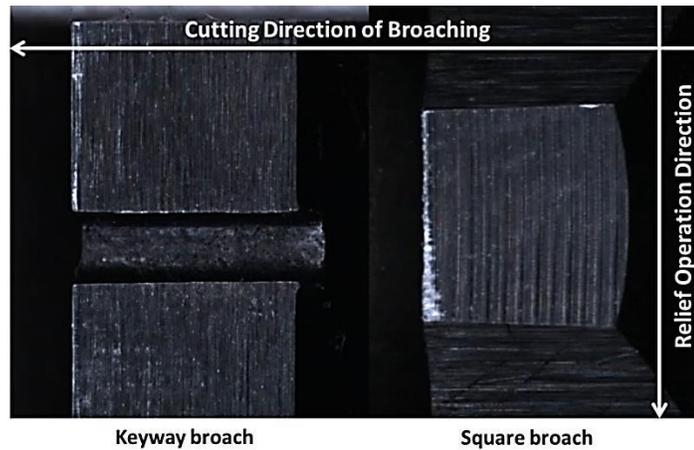


Figure 6: Directions of cutting in broaching and relief operations.

3.2.1 Experimental Setup

The proposed system records image data via a Mitutoyo QI-B3017B Quick Image Telecentric Lens System digital microscope (Figure 7), which has the following properties for wear quantification:

- Large field of view (12.8×9.6mm) and high resolution (1280×960) images – therefore, it can contain the cutting edge in one single image while include highly detailed information in the images;

- Fine scales (0.001mm) in positioning stage that makes the positioning of the part highly repeatable;
- Good measuring repeatability $(5+0.8L)\mu\text{m}$, where L represents the measuring length in mm) that makes it possible to obtain highly repeatable measurements of the same dimension assuming the positioning and operations are identical;
- LED ring light is adjustable by quadrant to enhance contrast of different texture directions (Figure 8).

The analysis used in this section was done operating a hexagonal broach, made of High Speed Steel, size 0.75” with 38 teeth and 0.375” pitch. Tool wear was measured by capturing images of selected cutting edges around one of the 38 teeth on one single broach.

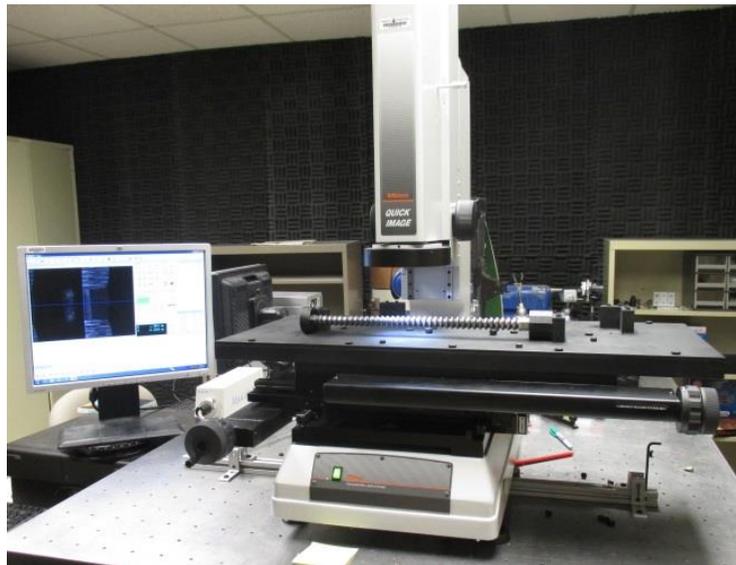


Figure 7: Mitutoyo Quick Image microscope and the broach fixture

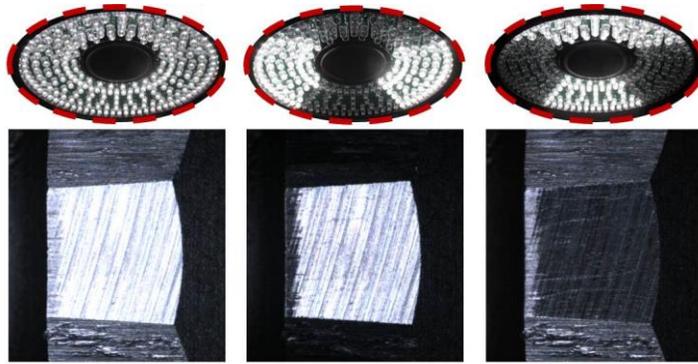


Figure 8: Quadrant LED lighting.

3.2.2 *Image Acquisition*

For image acquisition purposes, the broach is first fixed to blocks on each end so it can be easily rotated when inspecting other edges is necessary. The broach (along with the blocks) is then placed onto the microscope frame for inspection (Figure 9). In this setup, the xy plane coordinates can be changed using the adjustment wheels on the microscope (Figure 7), while the lighting is adjusted in the front and back (LED lights) using the proprietary microscope software interface, QUIPAK software. For improved image contrast, the room lighting was turned off and the microscope focus (z axis) was adjusted based on the software too. To ensure repeatable experimentation, the lighting settings and the position of the broach were recorded digitally, and every time adjusted to the same readings. As described previously, the position of the tool and light settings are critical factors to DIP due to their major effect on the determination of wear [47].

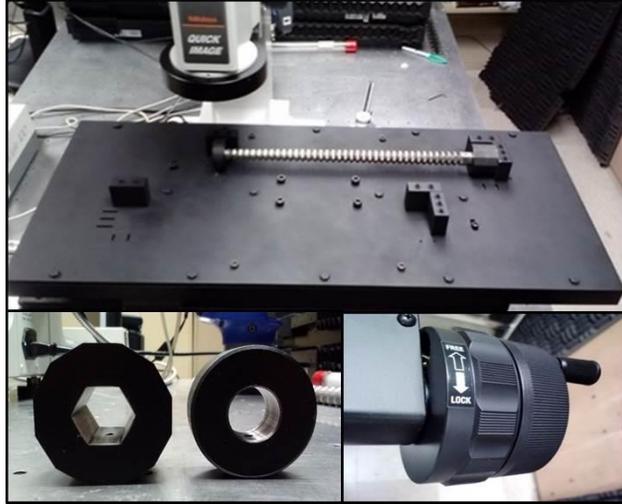


Figure 9: Microscope frame (top), block fixtures (left), and adjustment wheel (right).

3.2.3 Image Processing

Given the images captured from the microscope, various image processing techniques should be applied to extract the feature to describe the wear on the each cutting edge. As shown in Figure 10, the image processing steps involve image cropping, noise reduction, image thresholding, image cleaning, and feature extraction where each method utilizes a common DIP algorithm.

To remove the noise from unrelated areas, the image needs to be cropped to its region of interest (ROI) before any processing. Two cropping operations--rough and advanced cropping--are proposed to determine the ROI of the tool wear. Rough cropping is first applied based on the fixed position and size of each tooth. As teeth sizes change along the length of the broach while the position of broach is highly controlled, these cropping parameters can be determined before the operation and automatically applied to the images.

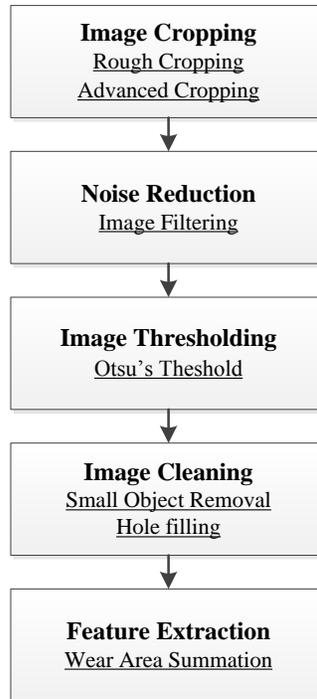


Figure 10: Image processing procedures

Variations in tool striation patterns relative to the cutting direction of broaching cause light to reflect differently, obscuring the delineation between the worn and unworn regions. Therefore, the image processing methodology involves advanced automated cropping by summing the pixel intensity in each row to form a single column of cumulative pixel intensity values. A graph of this column of summed row intensities is shown in Figure 11. The two cropping parameters, circled in red in the figure, are determined according to a variance-based change point detection technique on the summed column vector [104]. These points are used as boundary parameters for the advanced cropping. A similar process is also applied to sum intensity values in each column, forming a single row of summations to determine the left-most and right-most cropping parameters. Figure 12 shows the transition from a rough-cropped to an advanced-cropped image. It should be noted that this process is totally automatic and does not required the individual judgment of the operator.

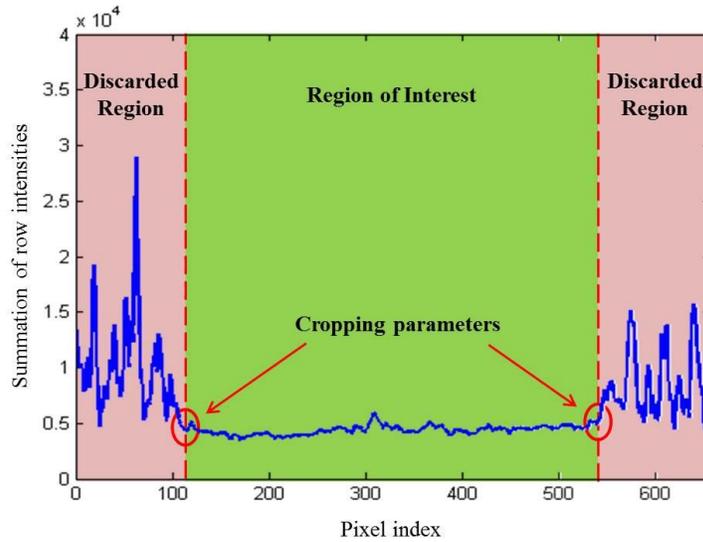


Figure 11: Advanced cropping by change-point detection technique

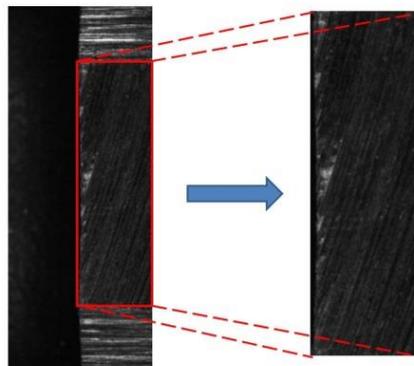


Figure 12: Rough-cropped image (left) and advanced-cropped image (right).

After cropping, the Wiener filtering method, a widely used low-pass adaptive filter based on neighboring pixels' statistics [105] is used to remove noise from the cropped image.

The image is then converted from grayscale to binary by applying a global thresholding method. The selection of these thresholds is informed by using Otsu's method of minimizing intra-class variance of the black and white pixels [106].

Finally, and as a second local filter, small objects—typically manifest as remaining machined chips on the broach—are defined as isolated patches which contain fewer than

8 pixels and are removed from the binary image. Then, any holes in the object are filled to more accurately represent the actual wear on the broach. A hole is defined to be any set of background pixels which cannot be reached by filling the background from the boundary of the image. Figure 13 shows the image after cropping (left) and the final binary image obtained from all the DIP steps (right).

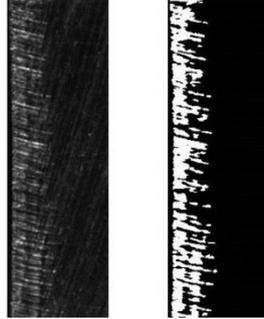


Figure 13: Cutting edge image after cropping (left) and the final binary image (right).

3.2.4 Measurement System Analysis

A widely accepted method to test a measurement system's performance is through a measurement system analysis. Specifically, both industry and academic researchers accept the use of Gauge Repeatability and Reproducibility (R&R) or similar methods to determine measures of the system's precision [46, 107, 108]. The measurement system was tested by focusing on four cutting edges with same design but different levels of tool wear. Two operators were trained and three replications were conducted for each combination of cutting edge and operator. Repeatability and reproducibility were derived by calculating the variances caused by different cutting edges, different operators, and multiple replications for both total area of wear region and flank wear width. All images were processed using the established operational procedures, image acquisition, and processing techniques. The flank wear widths are calculated also based on the images

taken. The summary of the contributed variances based on different features are summarized in Table 1.

Table 1: Analysis of Variance (ANOVA) for Gauge R&R Study

Source	Contribution to Total Variance (%)	
	Wear Area	Flank Wear
Gauge R&R Total	0.0184	21.5211
Repeatability	0.0174	20.6335
Reproducibility	0.0010	0.8876
Operator	0.0010	0.0000
Operator \times Cutting Edge	0.0000	0.8876
Between Cutting Edges	99.9816	78.4789
Total Variance	100	100

It can be noted that the variance component of the measurement system based on flank wear is much larger than the measurement system based on wear area, which essentially means that the resulting wear area measurements are less noisy and thus more distinguishable for differences between different wear levels when compared with the flank wear. This difference occurs because the maximum width of a given portion is more sensitive to noise than the area of that portion. Also, it should be noted that there is one variance component in reproducibility that has zero variance percentage in each column in Table 1. This is because the estimation of those variance components are negative, and in this case the negative variance estimates are set as zero due to it is impossible to have negative variances [109].

In summary, it can be observed that using area of wear region the measurement system contributes only a very small (less than 0.1%) component of the total variation in the measurements, which guarantees a reliable wear characterization based on total area of wear portion. Compared to flank wear width, the measurements of wear area is more robust to operational variation based on the results of the measurement system analysis.

4. Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profile Outputs

4.1 Introduction

In this section, a real-time process monitoring and fault diagnosis for multistage processes with non-repeating cyclic profiles are proposed. A Partial Least Squares (PLS) regression model is used to characterize the relationship between the profiles from consecutive stages and thus standardized OSFEs can be obtained. Then, multiple streams are monitored simultaneously under the assumption that the OSFEs are identically distributed. The multi-stream process is comprised of the following: 1) the streams of the OSFEs which are used to detect local mean shifts in the process; 2) one stream of the global mean of the original profile, which is used for global mean shift detection. Then, a Group Exponentially Weighted Moving Average (GEWMA) control chart is used to monitor the extracted multiple streams simultaneously. In the proposed methodology, when a shift is detected, the potential root cause can be identified. This includes determining whether the shift occurs locally or globally, and locating any locally shifted stage.

4.2 Multi-stream Extraction from Original Profiles

In this section, the configuration of the multi-stream process extracted from the original profiles is introduced. PLS regression models are proposed to model the relationship between the profiles of consecutive stages.

4.2.1 *Multi-stream Processes*

Multiple streams can be extracted from original profiles to isolate the global and local information of different stages in the process. There are quite a few methods including Wavelet Decomposition and Hilbert Huang Transformation that can extract global information from the original profile [110, 111]. However, due to the smoothing concern in most of these transformations, the extracted global trend will always take local shift information from the original profile, which makes it difficult to determine whether the shift is a global change or a local one at a specific location. Therefore, the global mean of the original profile is used to minimize the effect of local change on the global trend information.

Considering that the total number of stages in a process is q , the potential streams available will include one stream of the global mean of the original profile and $(q-1)$ streams obtained from the OSFEs of consecutive stages. Not only can this configuration make full use of the information contained in the original profiles, but it can also ensure that the extracted process streams have explicit engineering explanations. Each process stream has its one-to-one correspondence to one type of shift or one specific location of the local shift. The global stream is responsible for detecting global shifts; while the $(q-1)$ OSFEs are responsible for detecting the local shifts at the corresponding stage.

4.2.2 *Profile Segmentation*

The original profile will be segmented such that each segment represents the output from one stage. The segmentation can provide fault diagnostic guidelines to locate the shifted stage when a shift is detected. It can be observed in Figure 2 that each stage starts with a sharp increase of the force as a new broach tooth engages with the workpiece and

ends with a sharp decrease of the force as a tooth disengages [112, 113]. Thus, one local minimum point indicates the landmark between adjacent stages. Even if there is a local shift, the local minimum point still exists as it represents the engagement of a new tooth to the workpiece. Therefore, the original profiles can be segmented by the valleys in the profiles based on the process knowledge.

Given the sampling rate of the specific sample, the predetermined cutting speed, and the pitch of the broach, the profile valleys can be found easily by searching the minimum point in the potential scope which comes from the tool displacement relative to the workpiece and takes into account the process uncertainty, such as geometry deviation of the tooth, and sensor measurement error in the system. Denote the i th original profile as $C_i(t)$ ($i=1,2,\dots,m$) which can be segmented into q profiles as $X_{i,j}(t)$ ($i=1,2,\dots,m$; $j=1,2,\dots,q$). The procedure is illustrated in Figure 14. After segmentation, each profile only has one maximum value and characterizes the operating condition of its corresponding stage, as shown in Figure 14. Due to the uncertainty in the process, the segmented profiles may not have the same dimension after segmentation. In that case, linear interpolation can be used to adjust the dimensions of different stages to the same value.

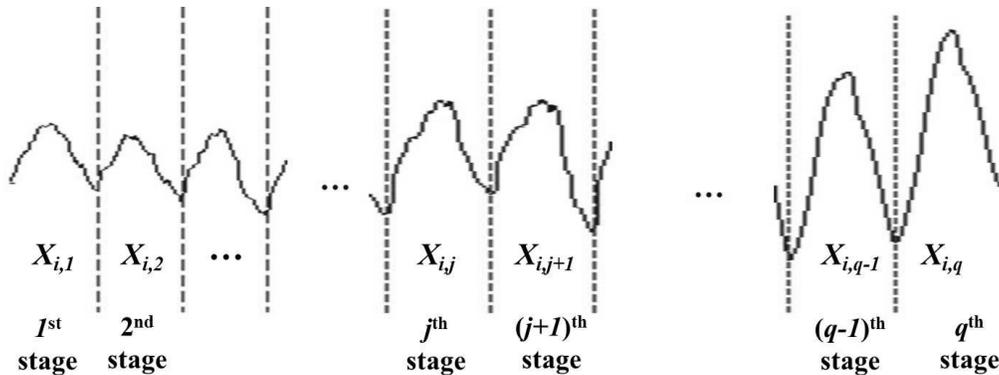


Figure 14: Profile segmentation

4.2.3 PLS Modeling

The objective of PLS modeling is to remove the heterogeneity among the stages to obtain multiple homogeneous process streams under the normal manufacturing conditions. To model the relationships between $\mathbf{X}_{.j}$ and $\mathbf{X}_{.j+1}$ using PLS regression, the underlying assumption is that the major variation pattern in $\mathbf{X}_{.j+1}$ can be described with a small number of Principal Components (PCs) extracted from $\mathbf{X}_{.j}$.

In the PLS modeling approach, the profile of each stage is considered as a multivariate vector. A series of PLS regression models are estimated to describe the correlation between each pair of consecutive profiles. As shown in Figure 15, the covariance between $\mathbf{X}_{.j}$ and $\mathbf{X}_{.j+1}$ are maximized by extracting A_j PCs from these two matrices that could explain a predefined percentage of variance in the response matrix $\mathbf{X}_{.j+1}$. Then the OSFEs of these models are assumed to follow the same distribution under the normal operating conditions. As illustrated in Equation (1), $\boldsymbol{\varepsilon}_{ij+1}$ is the OSFE of the $(j+1)$ th stage in the i th observation ($i=1,2,\dots,m; j=1,2,\dots,q-1$).

$$\mathbf{X}_{ij+1} = f_j(\mathbf{X}_{ij}) + \boldsymbol{\varepsilon}_{ij+1}, \quad (1)$$

where \mathbf{X}_{ij} and \mathbf{X}_{ij+1} denote the profiles collected from the j th and $(j+1)$ th stages in the i th observation. For each pair of consecutive stages $\mathbf{X}_{.j}$ and $\mathbf{X}_{.j+1}$ ($j=1,2,\dots,q-1$), one PLS regression model is estimated based on a fixed set of m samples. The PLS regression models can be obtained using the following procedures [114].

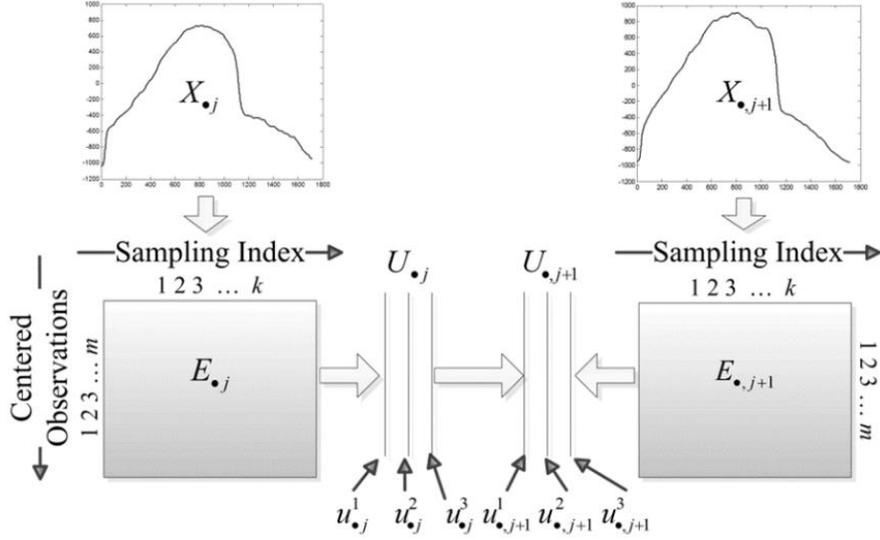


Figure 15: An illustration of PLS regression model between the profiles at the consecutive stages (redrawn from Höskuldsson [114])

Step 1 The estimated mean of $\mathbf{X}_{\bullet,j}$ and $\mathbf{X}_{\bullet,j+1}$ are subtracted from $\mathbf{X}_{\bullet,j}$ and $\mathbf{X}_{\bullet,j+1}$ to obtain $\mathbf{E}_{\bullet,j}$ and $\mathbf{E}_{\bullet,j+1}$, where $\mathbf{X}_{\bullet,j}$ is an $m \times k$ matrix containing m samples of the j th stage. The centering is implemented by

$$\mathbf{E}_{\bullet,j} = \mathbf{X}_{\bullet,j} - \mathbf{1}\boldsymbol{\mu}_j', \quad (2)$$

where $\boldsymbol{\mu}_j$ is a $k \times 1$ vector of the column means for matrix $\mathbf{X}_{\bullet,j}$, and $\mathbf{1}$ is an $m \times 1$ vector of ones.

Step 2 A_j principal components (PCs) are extracted one by one from $\mathbf{E}_{\bullet,j}$ and $\mathbf{E}_{\bullet,j+1}$, respectively, to exceed the total percentage of explained variance in $\mathbf{E}_{\bullet,j+1}$, denoted as p_m . The PCs are extracted based on the following iterations:

Step 2.0 Initialize $\mathbf{E}_{\bullet,j}^0 = \mathbf{E}_{\bullet,j}$, $\mathbf{E}_{\bullet,j+1}^0 = \mathbf{E}_{\bullet,j+1}$, and $\mathbf{u}_{\bullet,j+1}^0 =$ first column of $\mathbf{E}_{\bullet,j+1}^0$;

For $a = 1:A_j$

Extract the PCs $\mathbf{u}_{\bullet,j}^a$ and $\mathbf{u}_{\bullet,j+1}^a$ from $\mathbf{E}_{\bullet,j}^{a-1}$ and $\mathbf{E}_{\bullet,j+1}^{a-1}$ by the following sub-

steps

Step 2.1 Calculate the coefficient vector \mathbf{q}_j^a of the a th PC for $\mathbf{E}_{\cdot,j}^{a-1}$ by

$\mathbf{q}_j^a = \mathbf{E}_{\cdot,j}^{a-1} \mathbf{u}_{j+1}^{a-1} / (\mathbf{u}_{j+1}^{a-1} \mathbf{u}_{j+1}^{a-1})$, where \mathbf{q}_j^a is usually called the weight vector for the a th PC, and then scale \mathbf{q}_j^a to be length of one

Step 2.2 Calculate the a th PC \mathbf{u}_j^a for the j th stage by $\mathbf{u}_j^a = \mathbf{E}_{\cdot,j}^{a-1} \mathbf{q}_j^a$

Step 2.3 Calculate the coefficient vector \mathbf{q}_{j+1}^a of the a th PC for $\mathbf{E}_{\cdot,j+1}^{a-1}$ by

$\mathbf{q}_{j+1}^a = \mathbf{E}_{\cdot,j+1}^{a-1} \mathbf{u}_j^a / (\mathbf{u}_j^a \mathbf{u}_j^a)$ and then scale \mathbf{q}_{j+1}^a to be length of one

Step 2.4 Calculate the a th PC \mathbf{u}_{j+1}^a for the $(j+1)$ th stage by $\mathbf{u}_{j+1}^a =$

$$\mathbf{E}_{\cdot,j+1}^{a-1} \mathbf{q}_{j+1}^a$$

Step 2.5 The loadings vectors of the a th PC for the j th and $(j+1)$ th stage

can be estimated by least squares $\mathbf{f}_j^a = \mathbf{E}_{\cdot,j}^{a-1} \mathbf{u}_j^a / \mathbf{u}_j^{a'} \mathbf{u}_j^a$ and

$$\mathbf{f}_{j+1}^a = \mathbf{E}_{\cdot,j+1}^{a-1} \mathbf{u}_{j+1}^a / \mathbf{u}_{j+1}^{a'} \mathbf{u}_{j+1}^a$$

Step 2.6 Regress \mathbf{u}_{j+1}^a on \mathbf{u}_j^a to estimate the regression coefficients for the

a th PC by $\mathbf{B}_j^a = \mathbf{u}_{j+1}^a \mathbf{u}_j^{a'} / (\mathbf{u}_j^{a'} \mathbf{u}_j^a)$

Step 2.7 Subtract the extracted PCs to obtain new residuals by $\mathbf{E}_{\cdot,j}^a =$

$$\mathbf{E}_{\cdot,j}^{a-1} - \mathbf{u}_{\cdot,j}^a \mathbf{f}_j^{a'}$$
, and $\mathbf{E}_{\cdot,j+1}^a = \mathbf{E}_{\cdot,j+1}^{a-1} - \mathbf{B}_j^a \mathbf{u}_j^a \mathbf{q}_{j+1}^{a'}$

End

The extracted PCs can explain a large proportion of the variance. \mathbf{Q}_j is a $k \times A_j$ transformation matrix containing all the \mathbf{q}_j^a 's ($a = 1, 2, \dots, A_j$), and $\mathbf{U}_{\cdot,j}$ is the extracted principal components containing all the \mathbf{u}_j^a 's ($a = 1, 2, \dots, A_j$).

$$\mathbf{U}_{\cdot,j} = \mathbf{E}_{\cdot,j} \mathbf{Q}_j . \quad (3)$$

After extracting A_j PCs, there is still a small portion of variance that has not been modeled, which is regarded as noise. That component is calculated by

$$\mathbf{E}_{\cdot,j}^{A_j} = \mathbf{E}_{\cdot,j} - \mathbf{U}_{\cdot,j} \mathbf{F}_{\cdot,j}', \quad (4)$$

where $\mathbf{F}_{\cdot,j}$ is usually named as the loading matrices and it contains all the \mathbf{f}_j^a 's ($a = 1, 2, \dots, A_j$).

Step 3 Based on the iterations in Step 2, a multiple linear regression model can be obtained to characterize the correlation between the PCs of the consecutive stages, where \mathbf{B}_j is the $A_j \times A_j$ diagonal coefficient matrix with B_j^a 's at the diagonal, and $\mathbf{H}_{\cdot,j+1}$ is the residual matrix,

$$\mathbf{U}_{\cdot,j+1} = \mathbf{U}_{\cdot,j} \mathbf{B}_j + \mathbf{H}_{\cdot,j+1}. \quad (5)$$

Step 4 Given Equations (2), (3), and (5), Equation (4) can be rewritten and rearranged to obtain Equation (6), where \mathbf{b}_j is the $k \times k$ transformation matrix, \mathbf{L}_j is the intercept term, and $\boldsymbol{\varepsilon}_{\cdot,j+1}$ is the $m \times k$ residual matrix.

$$\mathbf{X}_{\cdot,j+1} = \mathbf{L}_j + \mathbf{X}_{\cdot,j} \mathbf{b}_j + \boldsymbol{\varepsilon}_{\cdot,j+1}, \quad (6)$$

where $\mathbf{b}_j = \mathbf{Q}_j \mathbf{B}_j \mathbf{F}'_{\cdot,j+1}$, $\mathbf{L}_j = \mathbf{1} \boldsymbol{\mu}'_{j+1} - \mathbf{1} \boldsymbol{\mu}'_j \mathbf{Q}_j \mathbf{B}_j \mathbf{F}'_{\cdot,j+1}$, and $\boldsymbol{\varepsilon}_{\cdot,j+1} = \mathbf{E}_{\cdot,j+1}^A - \mathbf{H}_{\cdot,j+1} \mathbf{F}_{\cdot,j+1}'$.

The computing algorithm used for PLS regression modeling is the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm [115]. For more information about PLS regression models and the computing algorithms, see Geladi and Kowalski [116] and Wold *et al.* [117].

The number of stages (q) is determined by how many non-repeating cycles in the process. For example, in a broaching process, the number of stages is defined as the total

number of teeth on the broaching tool. The value of the parameter A_j could be regarded as the tuning parameter in the proposed method. In the literature, there are quite a few methods proposed to select the number of components extracted, among which the methods of minimizing the Predictive Residual Sum of Squares (PRESS) based on cross validation [117, 118] are the most popular. However, in the proposed method, the balance among the accuracies of the models between different stages is more of concern than optimizing the modeling accuracy of every model individually. Therefore, the A_j value for the PLS regression model between the j th and $(j+1)$ th stages has been selected based on the training dataset by searching for the smallest number of components which could exceed a predefined percentage, i.e., 95%, of the total variance explained in the profiles at the $(j+1)$ th stage. Additionally, the ARL performance of the proposed method given different percentages of total variance explained p_m are summarized in the Supplemental Material A, which illustrates high robustness to varying tuning parameters of the proposed control charting system.

The required number of reference samples (m) depends on the dimension of the data and the correlation within the data. For strongly correlated data sets, the number of reference samples m can be smaller than the dimension of the data yet a stable estimation can still be obtained [119].

Based on the reference data set, $(q-1)$ PLS regression models can be estimated. For future observations, $(q-1)$ OSFEs can be calculated by subtracting the predicted profiles from the observed profiles. The OSFEs can be used to characterize the working condition of each stage, which can be monitored simultaneously as $(q-1)$ streams in a control charting scheme.

4.3 GEMWA Monitoring and Diagnostic Scheme for Global and Local Shift Detection

Given multiple streams extracted from original profiles, a Group Exponentially Weighted Moving Average (GEWMA) monitoring scheme is adopted to monitor the different streams simultaneously and detect global and local shifts.

4.3.1 GEWMA Monitoring Scheme for Simultaneous Global and Local Shift Detection

To use the GEWMA chart to monitor a multi-stream process, the streams should follow two assumptions. First, each stream should have the same mean and same variation. Second, the distribution of each single stream should be “approximately normally distributed” [43]. To satisfy the first assumption, the multiple streams extracted in Section 2 can be scaled to have the mean and variation based on their distribution. For the second assumption, the normality assumptions of the multiple streams will be validated later in the case study based on the real data set. Furthermore, the EWMA chart used in the proposed method is usually quite robust to non-normal distributions [120].

The GEWMA monitoring statistics based on the global mean and the OSFEs are calculated to detect the global and local shifts. First, the deviations from nominal $\varepsilon_{i,j}$ ($i = 1, 2, \dots, m$, and $j = 1, 2, \dots, q$) are calculated as

$$\begin{cases} \varepsilon_{ij} = \bar{c}_i - \bar{c}, & \text{if } j = 1 \\ \varepsilon_{ij} = X_{ij} - \widehat{X}_{ij}, & \text{if } j = 2, 3, \dots, q \end{cases} \quad (7)$$

where ε_{i1} denotes the deviation from nominal for the global mean of the original profiles, \bar{c}_i denotes the global mean of the i th original profile, and \bar{c} is the average global mean obtained from the reference samples. Then the GEWMA statistics can be calculated based on the Equations (8) and (9) in which the $\varepsilon_{i,j}$'s of the same stream are

exponentially weighted across the sequential observations, and the multivariate EWMA statistic can be calculated based on the method proposed by Lowry *et al.* [121]. That is

$$\begin{cases} Z_{i1} = \lambda \varepsilon_{i1} + (1 - \lambda)Z_{i-1,1}, \text{ if } j = 1 \\ \mathbf{Z}_{ij} = \lambda \boldsymbol{\varepsilon}_{ij} + (1 - \lambda)\mathbf{Z}_{i-1,j}, \text{ if } j = 2, 3, \dots, q \end{cases}, \quad (8)$$

where $0 < \lambda < 1$ is the smoothing parameter in traditional EWMA charts.

$$T_{ij}^2 = \begin{cases} Z_{i1}^2 / \sigma_{Z1}^2, & \text{if } j = 1 \\ \mathbf{Z}'_{ij} (\boldsymbol{\Sigma}_Z^j)^{-1} \mathbf{Z}_{ij}, & \text{if } j = 2, 3, \dots, q \end{cases}, \quad (9)$$

where

$$\begin{cases} \sigma_{Z1}^2 = \frac{\lambda}{2-\lambda} [1 - (1 - \lambda)^{2i}] \sigma_{\varepsilon 1}^2, \text{ if } j = 1 \\ \boldsymbol{\Sigma}_Z^j = \frac{\lambda}{2-\lambda} [1 - (1 - \lambda)^{2i}] \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^j, \text{ if } j = 2, 3, \dots, q \end{cases}. \quad (10)$$

Here $\sigma_{\varepsilon 1}^2$ denotes the variance of the deviations from the nominal of the global mean, and it is estimated based on the ε_{i1} 's of the historical data; while $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^j$ represents the estimated variance covariance matrix of the OSFEs of the j th stage, and it is estimated using the method based on successive differences, which is recommended by Sullivan and Woodall [122].

As the T_{ij}^2 statistic approximately follows a χ^2 distribution with degree of freedom value p_j and

$$p_j = \begin{cases} 1, \text{ if } j = 1 \\ k, \text{ if } j = 2, 3, \dots, q \end{cases}, \quad (11)$$

the T_{ij}^2 statistics have different mean and variances given various p_j values. To deal with the dimensional inconsistency in all the streams, the adjustment suggested in Xiang and Tsung [42] is performed as

$$\tilde{T}_{ij}^2 = \frac{T_{ij}^2 - p_j}{\sqrt{2p_j}}. \quad (12)$$

Therefore, the \tilde{T}_{ij}^2 statistics of all the streams have the same mean of zero and the same variance of one. The GEMWA statistics of the i th observation can thus be defined as the largest \tilde{T}_{ij}^2 statistics of all the streams, which is

$$MZ_i = \max_{1 \leq j \leq q} (\tilde{T}_{ij}^2), \quad (13)$$

for $i=1,2,\dots,m$. The control limit of the chart is given by

$$h = L \frac{\lambda}{2-\lambda}, \quad (14)$$

where L and h are design parameters of the chart. Given a desired ARL_0 and a predefined λ value, the value of L and h can be obtained by simulation. The control chart signals when an MZ_i value exceeds the control limit h , indicating a shift is detected in the process.

In addition, although the non-repeating cyclic profiles are not assumed to be identically distributed, it is assumed that the monitoring statistics of the OSFEs at all the stages approximately follow the same distribution after the adjustment based on the dimension of the OSFEs. Therefore, after the calculation of Equation (12), the streams derived from the OSFEs can be monitored with one group chart while the global stream can be monitored separately if its dimension varies significantly from the other streams.

4.3.2 The Diagnostic Approach of the Proposed Control Chart

Once a process shift is detected, the fault diagnostic approach is used to determine what type of shift has occurred. If there is a local shift detected, the diagnostic approach will further locate the shifted stage. As illustrated in Equation (13), the maximum of all the \tilde{T}_{ij}^2 statistics is used as the monitoring statistics. Xiang and Tsung [42] suggested that the shifted stage should be $\operatorname{argmax}_{1 \leq j \leq q} (\tilde{T}_{ij}^2)$ for the i th observation. Therefore, the diagnostic operation can be performed by two steps:

Step 1 Determine if there is a global shift occurring by checking if $\tilde{T}_{i1}^2 > h$ is true. If the inequality holds, then a global shift occurs in the process.

Step 2 Determine if there is a local shift by checking if $\tilde{T}_{ij}^2 > h$ ($j = 2, 3, \dots, q$), and if any, identify the shifted stage as $\operatorname{argmax}_{2 \leq j \leq q}(\tilde{T}_{ij}^2)$.

Based on the monitoring scheme and fault diagnostic approach proposed in this section, detection of global and local shifts and fault diagnosis can be achieved simultaneously.

4.4 Performance Analysis using Simulation

The performance of the process monitoring and diagnosis is evaluated using a simulation study. Following similar patterns of the real faults and force signals from the broaching process, original profiles are generated by summing up a signal component of global trend and a component with non-repeating cyclic patterns.

In the simulation, the number of stages in the process is $q=8$. The in-control ARL_0 are adjusted to approximately 370 by varying the control limits and calculate their corresponding run length based on a fixed series of monitoring statistics with a length of 7,000 for each one of the 10,000 replications, and the value of the smoothing constant λ used is 0.2. Here λ is a user specified parameter. If historical data has higher weight, the value of λ should be smaller, and vice versa. The out-of-control signals have three types of shifts:

(1) Global shift: the global component has a mean shift as $\boldsymbol{\mu}_\epsilon = \delta \times \operatorname{diag}(\boldsymbol{\Sigma}_{G0})^{1/2}$, where δ and $\boldsymbol{\Sigma}_{G0}$ are the magnitude of the mean shift and variance covariance matrix of the global trend of the in-control profiles, respectively.

(2) Local wear: the j th stage has a mean shift as $\boldsymbol{\mu}_\varepsilon = \delta \times \text{diag}(\boldsymbol{\Sigma}_\varepsilon^j)^{1/2}$, where δ and $\boldsymbol{\Sigma}_\varepsilon^j$ are the magnitude of the shift and the variance covariance matrix of the added noise, respectively.

(3) Local breakage: the j th stage has a mean shift as $\boldsymbol{\mu}_\varepsilon = -\delta \times \text{diag}(\boldsymbol{\Sigma}_\varepsilon^j)^{1/2}$; while its next adjacent stage has a mean shift as $\boldsymbol{\mu}_\varepsilon = \delta \times \text{diag}(\boldsymbol{\Sigma}_\varepsilon^{j+1})^{1/2}$, where δ and $\boldsymbol{\Sigma}_\varepsilon^j$ represent the magnitude of the shift and the variance covariance matrix of the added noise at the j th stage, respectively.

Such mean shift patterns are generated in the spirit of local wear and breakage from the application case study. The magnitudes of the mean shifts are set as $\delta = 0.1, 0.25, 0.5, 1, 1.5, 2, 2.5, 3$ for all three types of shifts. The ARL is obtained by averaging 10,000 run lengths. The steady-state performances of the charts are compared by first generating ten normal observations in each run and then generating observations with sustained global/local mean shift until the chart signals. Any premature signals during the first ten normal observations are discarded.

To deal with the singularity problem in the OSFEs' variance covariance matrix, PCs are extracted from the OSFEs, and the number of PCs is determined by the percentage of variance explained in the PCs, i.e., any PC which explains more than 0.05% of the total variance in the OSFE will be retained in the monitoring statistics. An excessively large threshold percentage will lead to loss of information in the shifts, while a very small threshold percentage cannot solve the singularity problem. Based on the training data set, the numbers of PCs extracted are listed in Table 2. It is clear that the numbers listed are significantly greater than the dimension of the global stream. Therefore, one one-dimensional EWMA chart is used to monitor the global trend while the GEWMA chart is

used to monitor the streams of the OSFEs. The decision rule of the control chart system is that once either one of the two charts signals, the system signals. To achieve comparable in-control ARL with the benchmark methods, false alarm rates can be applied to the two charts based on their relative costs of false alarms of the two charts and potential loss due to slow detection of different process changes by Bonferroni correction [123]. In this study, the false alarm rate is evenly distributed to the one-dimensional EWMA chart and the GEWMA chart, and thus the control chart system has a combined in-control ARL of 370 approximately.

Table 2: Number of PCs extracted from OSFEs

OSFE	2	3	4	5	6	7	8
Number of PCs	11	11	11	10	11	10	10

The numerical results include an Average Run Length (ARL) performance comparison with two benchmark methods; one is proposed by Zhou *et al.* [20] and the other is proposed by Zou *et al.* [14], with different preset parameter values, respectively. The two benchmark methods are selected because they are effective in detecting mean shifts in complicated profiles. In addition, an EWMA chart based on the features extracted from Zhou’s method is used for the performance comparison. This is because the method proposed in Zhou *et al.* [20] considers a Shewhart-type control chart, while both our proposed method and the method proposed by Zou *et al.* [14] are EWMA-type control charts, which are usually more capable of detecting small shifts. Therefore, the EWMA chart is used to replace the Shewhart chart in Zhou’s method for a fair comparison.

To compare the ARL performances, local wear and local breakage at Stages 2 and 7 are generated in the original profiles, and global shifts are also generated. The simulation

results of the control chart performance are illustrated in Table 2. In this table, GEWMA denotes the proposed approach; Zhou *et al.*'s EWMA denotes the EWMA version of the approach proposed by Zhou *et al.* [20], and the percentage of energy eliminated from the original profiles denoted as Q is directly related to the number of wavelet coefficient levels L involved in the control chart in Zhou *et al.* [20]. In their method, the number of wavelet coefficient levels L used for constructing the monitoring statistics can vary from 1 to 6. Zou *et al.*'s EWMA denotes the method proposed by Zou *et al.* [14], and the parameter c denotes a tuning parameter to determine the bandwidth used to control the error in nonparametric profile representation. In the performance comparison, all the suggested choices of c values are examined.

The best performing ARL values of all the tested methods for each scenario is bolded in Table 3. The proposed GEWMA method outperforms both benchmarking methods in most of the scenarios. In local shift detection, the proposed GEWMA scheme outperforms the two methods in all the tested cases. In global shift detection, our proposed method outperforms the two benchmark methods under most choices of their preset parameter, though when the L value is as small as 1 or 2 the EWMA version of Zhou *et al.*'s method can detect global shifts more quickly. However, when the L value is selected as 1 or 2, the local shift detection is significantly slower than our proposed method, especially when there is a local breakage shift Zhou *et al.*'s EWMA chart can barely detect anything as their out-of-control ARLs are close to the in-control ARL.

Table 3: ARL comparison for various types of shifts

Shift Stage	δ	GEWMA	Zhou <i>et al.</i> 's EWMA					Zou <i>et al.</i> 's EWMA				
			$L=6$	$L=5$	$L=4$	$L=3$	$L=2$	$L=1$	$c=1.0$	$c=1.5$	$c=2.0$	
In-control	-	0	370.6	370.9	370.8	370.3	371.0	370.5	370.3	370.1	370.2	370.0
		0.1	187.7	311.7	324.9	313.7	303.8	340.9	363.7	365.6	364.4	368.0
		0.25	13.9	121.1	182.1	193.9	194.4	284.6	350.2	370.3	359.6	359.6
		0.5	4.0	21.6	49.3	72.8	85.7	189.8	294.8	337.4	323.9	335.2
		1	1.9	5.3	9.6	15.7	21.8	80.3	171.7	256.4	182.2	263.6
		1.5	1.3	3.1	5.0	7.4	10.0	39.7	93.6	156.3	71.6	172.4
		2	1.0	2.3	3.4	4.7	6.2	22.3	55.3	79.6	29.7	97.6
		2.5	1.0	1.9	2.6	3.5	4.5	14.6	35.4	37.5	15.2	50.7
		3	1.0	1.7	2.2	2.9	3.5	10.5	24.3	19.8	9.9	26.3
Local wear		0.1	241.5	255.8	286.3	249.4	377.7	409.9 ¹	362.5	372.4	368.1	360.6
		0.25	19.0	73.0	109.8	82.4	332.0	428.9	352.8	366.3	356.7	357.7
		0.5	4.6	13.9	21.4	18.2	202.2	363.2	280.7	335.6	307.0	335.4
		1	2.1	4.3	5.5	5.3	55.5	161.3	146.2	238.7	145.7	252.0
		1.5	1.5	2.7	3.2	3.1	21.0	69.1	75.4	130.0	54.1	142.6
		2	1.1	2.0	2.4	2.3	11.2	33.4	43.7	58.0	22.3	71.6
		2.5	1.0	1.7	1.9	1.9	7.4	19.6	27.6	27.0	12.3	36.6
		3	1.0	1.4	1.7	1.6	5.5	13.1	19.1	14.5	8.4	19.6
		0.1	50.5	114.6	257.0	330.8	344.1	372.1	359.3	368.4	347.0	375.0
	0.25	5.5	14.2	76.9	161.3	188.1	360.4	364.6	346.5	294.7	357.1	
	0.5	2.4	4.4	15.2	36.5	48.6	351.9	358.3	287.0	201.8	317.8	
	1	1.2	2.0	4.6	7.8	10.3	320.8	361.1	135.8	66.6	209.8	
	1.5	1.0	1.4	2.8	4.2	5.3	289.5	346.9	50.6	23.3	110.5	
	2	1.0	1.1	2.1	3.0	3.6	259.2	341.0	20.5	11.4	52.0	
	2.5	1.0	1.0	1.7	2.4	2.8	231.0	328.6	10.9	7.5	25.5	
	3	1.0	1.0	1.5	2.0	2.3	207.2	310.3	7.4	5.6	14.7	
Local breakage		0.1	106.2	162.2	200.0	162.6	361.3	373.3	360.5	374.7	374.9	369.9
		0.25	8.3	20.9	31.3	24.7	363.0	376.0	360.6	358.9	373.4	365.0
		0.5	3.0	5.3	7.0	6.3	368.9	379.8	359.2	313.9	307.3	344.1
		1	1.6	2.3	2.7	2.6	370.8	377.7	361.7	178.9	132.3	261.4
		1.5	1.0	1.7	1.9	1.8	370.6	371.3	354.8	69.0	36.1	154.0
		2	1.0	1.2	1.5	1.4	361.4	368.6	352.0	23.0	12.9	75.8
		2.5	1.0	1.0	1.1	1.1	367.2	375.8	354.0	10.6	7.9	32.8
		3	1.0	1.0	1.0	1.0	362.7	372.9	357.4	7.0	5.8	15.9
	Global	All	0.1	328.6	348.6	352.4	334.0	324.1	313.2	293.0	365.1	345.3

¹ It is unusual that an out-of-control ARL is larger than the in-control ARL with the same control limit. One reason of this phenomenon may be that this specific shift magnitude at the specific location cancels part of the variability in some of the wavelet coefficients by Harr Transformation.

0.25	167.4	278.5	267.0	222.0	195.8	156.1	130.7	337.4	246.7	318.2
0.5	54.8	140.2	116.8	82.7	63.8	46.6	36.2	225.7	96.2	193.3
1	11.3	30.1	23.1	16.6	13.9	11.1	9.4	63.7	20.1	42.1
1.5	5.7	11.3	9.3	7.4	6.6	5.6	5.0	18.9	8.8	13.7
2	3.9	6.6	5.7	4.8	4.4	3.9	3.5	9.1	5.6	7.4
2.5	3.0	4.7	4.1	3.6	3.3	3.0	2.7	6.1	4.2	5.2
3	2.4	3.7	3.3	2.9	2.7	2.4	2.2	4.6	3.3	4.0

Moreover, the performance of the benchmark methods is very sensitive to the preset tuning parameters. For the method proposed in Zhou *et al.* [20], as the L value gets larger the local shift detection power improves and the global shift detection power deteriorates. This can be explained by the fact that the detection power of either type of shift should be sacrificed for the other type, i.e., the optimal performance on detecting local and global shifts cannot be achieved based on the same choice of the L value. Furthermore, Zhou *et al.*'s method also demonstrates extremely high sensitivity when the shifted location varies. For example, given the preset parameter value $L=3$, satisfactory ARL performance can be achieved when all the tested global shifts and most of the local shifts occur. However, when there is a local breakage at Stage 7, ARL gets extremely large which indicates that the local breakage at Stage 7 is almost undetectable. For the approach proposed in Zou *et al.* [14], the control chart achieves best performance mostly when the value of c is 1.5. However, the ARL performance is also not quite robust when the tuning parameter varies within the suggested range. All these results indicate that the performance of the benchmark methods is not robust enough to the preset parameter when monitoring the non-repeating cyclic profiles. The robustness of the proposed method has been studied, and the ARL performance with respect to different percentage

values of total variance explained p_m in the response matrices has been summarized in Appendix A.

Therefore, the proposed method gives best results for local shift detection and satisfactory performances for global shift detection. This method demonstrates better robustness when shifted location varies as shown in Table 2. The advantage of the proposed method in ARL performances lies in the modeling of consecutive stages, which can reduce the total variance of the profiles by modeling their relationships with their preceding profiles.

To analyze the diagnostic performance, 10,000 simulation replicates are performed for the proposed approach. Table 4 summarizes the diagnostic accuracy. Motivated by the diagnostic performance metric in Zou and Tsung [44], one can compare the probability that the global shift is identified when the control chart signals given a global shift. Furthermore, when there is a local shift, the probabilities that the estimated shifted location $\hat{\zeta}$ falls within a given interval, which centers around the actual shift location ζ , are also summarized. For example, $\Pr(\hat{\zeta} = \zeta)$, $\Pr(|\hat{\zeta} - \zeta| \leq 1)$, and $\Pr(|\hat{\zeta} - \zeta| \leq 2)$ represent the probability of the estimated shifted stage index is exactly the actual shifted stage ζ , the probability of the estimated shifted stage index is included in the interval from $\zeta - 1$ to $\zeta + 1$, and the probability of the estimated shifted stage index is included in the interval from $\zeta - 2$ to $\zeta + 2$, respectively. The three probabilities described above are denoted as P_0 , P_1 , and P_2 .

Table 4: Diagnostic accuracy summary when using the GEWMA chart

Shift type	Shift stage	δ	Proposed Method		
			P ₀	P ₁	P ₂
Local wear	2	0.1	0.5624	0.6500	0.6500
		0.25	0.9637	0.9716	0.9716
		0.5	0.9952	0.9985	0.9985
		1	0.9990	0.9995	0.9995
		1.5	0.9991	0.9998	0.9998
		2	0.9999	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000
		3	1.0000	1.0000	1.0000
	7	0.1	0.3007	0.4600	0.4600
		0.25	0.6485	0.9597	0.9597
		0.5	0.6745	0.9964	0.9964
		1	0.7133	0.9989	0.9989
		1.5	0.7456	0.9996	0.9996
		2	0.7285	0.9998	0.9998
2.5		0.7853	1.0000	1.0000	
3		0.8311	1.0000	1.0000	
Local breakage	2	0.1	0.8681	0.9106	0.9106
		0.25	0.9814	0.9916	0.9916
		0.5	0.9952	0.9993	0.9994
		1	0.9972	0.9998	0.9998
		1.5	0.9999	1.0000	1.0000
		2	1.0000	1.0000	1.0000
		2.5	1.0000	1.0000	1.0000
		3	1.0000	1.0000	1.0000
	7	0.1	0.6852	0.7666	0.7666
		0.25	0.9309	0.9825	0.9825
		0.5	0.9612	0.9986	0.9988
		1	0.9798	0.9994	0.9995
		1.5	0.9827	0.9999	0.9999
		2	0.9969	1.0000	1.0000
2.5		0.9996	1.0000	1.0000	
3		0.9999	1.0000	1.0000	
Global Shift	All	0.1	0.5564	-	-
		0.25	0.7359	-	-
		0.5	0.8854	-	-
		1	0.9327	-	-
		1.5	0.9381	-	-
		2	0.9341	-	-
		2.5	0.9225	-	-
		3	0.9139	-	-

4.5 Case Study

The case study uses data collected from the experimental setup introduced in Section 3.1, which is a hexagonal broaching process. This process considers 33 stages, corresponding to the 33 teeth on the hexagonal broach. To illustrate the monitoring method, 20 force profiles were collected, including 15 normal runs and 5 abnormal runs. The first 10 runs were used as reference data set to estimate the model parameters. The rest of the runs were used for performance evaluation. The five abnormal runs were designed to represent the condition of tool wear on the 30th tooth of the broaching tool to test the monitoring performance. The segmented profiles were processed by linear interpolation, such that all segmented profiles had the same length [124]. The proposed control charting system was implemented to determine if the increasing local wear could be detected. To deal with the dimensional inconsistency, one individual chart was used for the global trend stream while one group chart was used for the streams of the OSFEs.

The Phase II monitoring performance is presented in Figure 16. The MZs of the two charts are plotted against the observational order. On the left panel, the number beside each point indicates the stage index obtained from the fault diagnostic approach. The control limits for the individual chart h_g and the group chart h_l was obtained by using three standard deviations plus the mean of the monitoring statistics obtained from the Phase I data. It is observed that the chart signaled as soon as there was a process change (the sixth observation in the chart), and the diagnostic stages always fell within the ± 1 stage interval of the real shifted stage, while the individual chart doesn't signal.

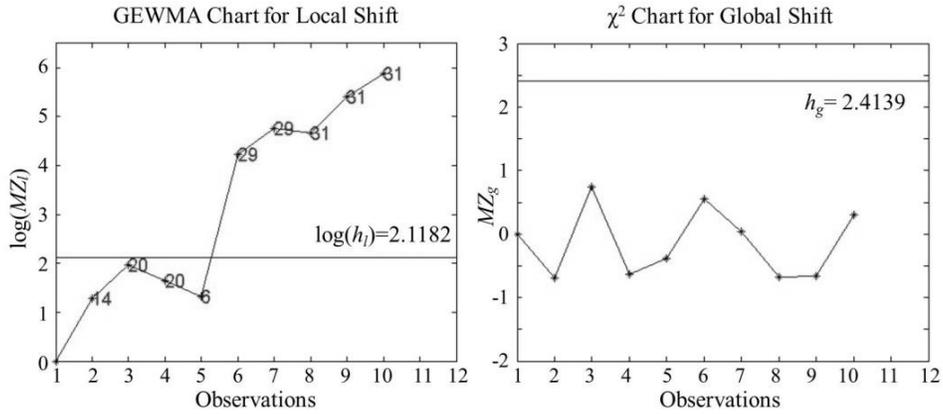


Figure 16: The GEWMA chart for shift detection

Furthermore, Henze-Zirkler's multivariate normality test was performed to the extracted PCs of the OSFEs obtained from the PLS regression model at each stage [125]. In Figure 17, the p -values calculated from all the stages (including the global mean component as the first stage) are plotted. There is not sufficient evidence to reject the null hypothesis that the OSFEs are normally distributed, since almost all the p -values associated to the tests are much larger than the predetermined significance level of 0.05. Therefore, the normality assumption is reasonable based on the data, which is required for the multi-stream process monitoring.

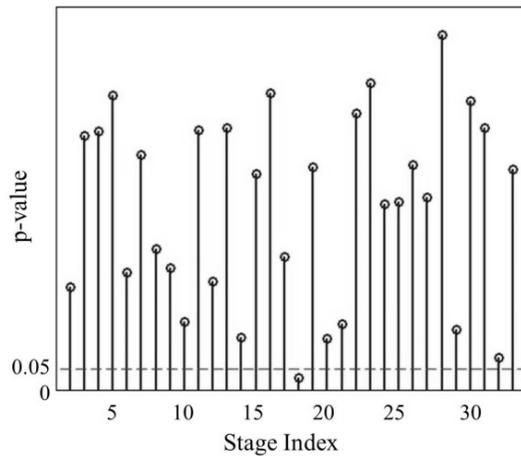


Figure 17: The p -values of the Henze-Zirkler's multivariate normality tests

The case study demonstrates the effectiveness of the proposed GEWMA monitoring and fault diagnostic approach in a real application. The early detection of process shifts can help reduce quality losses, and the precise shifted stage identification can lower the maintenance costs when a change is detected in the process.

4.6 Conclusions

Multistage processes with non-repeating cyclic profile outputs are widely encountered in various industries. However, current profile monitoring methodologies mainly focus on the profiles obtained from single-stage processes or multiple stages with repeating patterns. In this work, a PLS regression method is proposed to model the relationship between profiles from consecutive stages. Thus, the distributional heterogeneity among the multiple stages can be removed and identically distributed OFSEs can be obtained. A Group EWMA control chart is proposed to detect global and local shifts simultaneously. An ARL simulation study shows that the proposed GEWMA chart outperforms previous methods in both detection power and robustness. In addition, the proposed diagnostic approach can accurately identify the shifted stage in the process. A case study involving a push hexagonal broaching process is used to demonstrate the effectiveness of the proposed monitoring and diagnostic methodology. Significant quality improvements are expected in multistage process with non-repeating cyclic output through proactive adjustment using more effective quality control tools.

5. Process Knowledge Constrained Image Filtering for Tool Wear Characterization

In this section, a process knowledge constrained image filtering method for wear characterization is proposed by implementing the three image filtering constraints based on machining process knowledge. It is worth noting that the proposed process knowledge constrained image filtering method can always be incorporated with any intensity-based wear characterization methodology to quantify the tool wear in a more engineering viable way. A case study based on a hexagonal broaching process is described to show the effectiveness of the proposed methodology by comparing the wear characterization results before and after applying the proposed process knowledge constrained image filter right after the intensity-based method.

5.1 Proposed Methodology

Most broach cutting edges are manufactured with relief angles on cutting edge flanks. This angle creates excessive noise in cutting edge images, which are difficult to compensate for, despite efforts to enhance the desirable contrast through various lighting conditions [126]. To reduce this noise, the first task in the proposed method is to identify potential wear regions on the cutting edge at time t by using adaptive image thresholding. Then image filters are designed based on the three constraints based on machining process knowledge.

5.1.1 Notations

Table 5 summarizes the major notations used in the proposed image processing techniques. Generally speaking, the subscripts represent the cutting edge and time indices for corresponding images/features.

Table 5: List of symbols in notation

Notation	Definition
$I_{e,t}$	Original grayscale image of cutting edge e at time t
$RI_{e,t}$	Registered image of cutting edge e at time t
$CI_{e,t}$	Cropped image of cutting edge e at time t
$\alpha_{e,t}^O$	Threshold value obtained from Otsu's method for $CI_{e,t}$
$BI_{e,t}$	Binary image of the e th cutting edge at time t
$s_{e,t}(\alpha)$	Signal of image total intensity when threshold value α varies for $CI_{e,t}$
$\nabla^2 s$	Second-order differencing series of signal s
$\alpha_{e,t}^A$	Threshold value obtained from proposed adaptive thresholding method for $CI_{e,t}$
$K_{e,t}$	Total number of regions in $BI_{e,t}$
$R_{e,t}^k$	Region k in $BI_{e,t}$, $k = 1, 2, \dots, K_{e,t}$
$R_{e,t}^\bullet$	Union of $R_{e,t}^k$'s in $BI_{e,t}$, $k = 1, 2, \dots, K_{e,t}$
δ_1	Maximal allowable slack value of Constraint 1
δ_2	Maximal allowable slack value of Constraint 2
d_1	Deviance from Constraint 1
d_2	Deviance from Constraint 2

5.1.2 Image Preprocessing

In this subsection, the image preprocessing steps to achieve a binary image for wear characterization are proposed. These steps consist of image registration, cropping, and adaptive thresholding.

For image registration, $I_{e,t}$ ($t \geq 1, \forall e$) are registered based on $I_{e,0}$. After registration, all images $RI_{e,t}$ ($\forall e, t$) are cropped based on a cropping parameter that was estimated from $I_{e,0}$. The cropping parameter was determined by the change point method proposed in [126]. After registration and cropping, all region(s) of interest will be consistent for all cropped images $CI_{e,t}$ ($\forall e, t$).

In traditional TCM methods, image filtering schemes are implemented before (or together with) image thresholding to enhance wear region identification [127-129]. In this work, the process knowledge-based image filtering procedures are performed after the image thresholding step to identify any “potential wear regions” in the image.

An adaptive thresholding scheme is proposed under the assumption that actual wear consists of a smaller number of regions with larger sizes compared with noise textures in the image. The threshold value used to segment the image is determined by a change point detection approach based on the total image intensity of the resulting binary image. The threshold for each cropped image $CI_{e,t}$ is determined by the following steps:

Step 1: Calculate the initial threshold value $\alpha_{e,t}^O$ which is specified by Otsu’s method to minimize intra-class variance of segmented regions [130];

Step 2: Range the threshold value α in the neighborhood of the initial value $\alpha_{e,t}^O$ and obtain a signal of total image intensity over the threshold values $s_{e,t}(\alpha)$ ($\alpha \in [\alpha_O - \tau, \alpha_O + \tau]$);

Step 3: Perform second-order differencing to the signal $s_{e,t}(\alpha)$ to obtain the stabilized series $\nabla^2 s$ [131];

Step 4: Perform change point detection algorithm based on the variance in the window of fixed length proposed in [104] to obtain the threshold value $\alpha_{e,t}^A$;

Step 5: Segment the image based on the threshold value $\alpha_{e,t}^A$ to obtain a binary image $BI_{e,t}$.

The change point determined from the signal $\nabla^2 s$ represents the threshold value that produces the most significant increase in magnitude change of the total image intensity. This can be regarded as the turning point to identify irrelevant but massive noise textures.

5.1.3 Process Knowledge Constrained Image Filtering

Intensity-based image filtering methods have been used extensively in the literature to filter images obtained from machine vision systems for the purpose of wear region identification. However, when process knowledge is available, it is beneficial to apply this knowledge to image filtering to enhance the capability of the image processing and feature extraction techniques.

In this work, wear regions, which can affect cutting performance, are defined as *effective wear regions*. The regions that are segmented as wear but don't affect the cutting performance (*ineffective wear regions*) are identified and eliminated by applying the following three constraints to the binary images obtained from intensity-based thresholding.

Constraint 1: The effective wear region will always start accumulating from the cutting edge.

Constraint 2: The effective wear region should be continuous along the cutting direction, but could be discontinuous along the direction perpendicular to the cutting direction.

Constraint 3: Tool wear is an accumulative and irreversible phenomenon that ensures a non-decreasing property in the wear regions of the images. It is highly unlikely that one region of effective tool wear disappears at time $t + 1$ given there is one effective wear region identified at time t at the same location on the cutting edge.

In this subsection, process knowledge constrained image filtering algorithms, including within-image and between-image filters, are proposed to eliminate the noise of the binary images by sequentially applying these three constraints. It should be noted that

process knowledge-based image filtering, specifically the three aforementioned constraints, can be also applied to TCM applications for single-edged machining tools. This will be validated with a case study of a turning process in Section 5.2.2.

5.1.3.1 Within-image Filter

The proposed within-image image filter is designed based on *Constraints 1* and *2*. From the design of the tool, the cutting zone(s) of each cutting edge can be defined in the images obtained from the machine vision system. Figure 18 illustrates two examples of ineffective wear regions that violate *Constraint 1* (labeled as ①) and *Constraint 2* (labeled as ②), respectively, where δ_1 and δ_2 represent the maximal allowable slack values of *Constraints 1* and *2*.

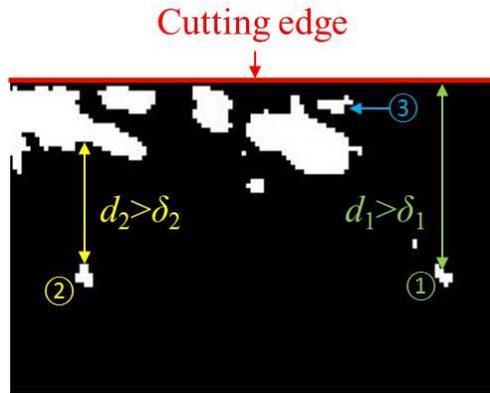


Figure 18: Examples of ineffective wear regions

Without loss of generality, it is assumed that the cutting direction is parallel to the column-direction in the image. If this is not the case, the images can be rotated easily to make the assumption hold. Given the predefined allowable slack value δ_1 for *Constraint 1* and slack value δ_2 for *Constraint 2* obtained from the engineering knowledge of the machining process, the within-image filtering can be performed based on the following steps for each Region $R_{e,t}^k$ in $BI_{e,t}$ ($k = 1, 2, \dots, K_{e,t}$).

Step 1: Find the extrema of $R_{e,t}^k$ by specifying the ranges of indices in rows ($r_1^k \leq r_2^k$) and columns ($c_1^k \leq c_2^k$);

Step 2: Calculate the summation of the columns from c_1 to c_2 in the image by

$$s_c = \sum_{c=c_1^k}^{c=c_2^k} BI_{e,t}(\cdot, c);$$

where $BI_{e,t}(\cdot, c)$ denotes the c^{th} column of binary image $BI_{e,t}$.

Step 3: Determine if $R_{e,t}^k$ is the first region starting from the side of the cutting edge (an example is labeled as ③). If true, proceed to *Step 4*; otherwise proceed to *Step 5*;

Step 4: Filter the image based on *Constraint 1*. Determine if r_1 (equivalent to d_1 in Figure 2) is greater than δ_1 which is specified as the furthest possible point on the cutting edge;

Step 5: Filter the image based on *Constraint 2*. Find the neighboring wear region to $R_{e,t}^k$ in terms of s_c on the side of the cutting edge, and determine if the minimal distance between these two regions (denoted as d_2 in Figure 2) is greater than the predefined δ_2 , which is the maximum allowable distance between two wear regions along the cutting direction.

The five-step algorithm is proposed to be performed iteratively for each wear region in the binary images obtained from Section 5.2.3.1 in an ascending order of the distance from the wear region to the cutting edge. The allowable slack value δ_1 can be determined by possible deformation of the cutting edge, and the slack value δ_2 can be determined by considering the relief angle of the cutting edge.

5.1.3.2 Between-image Filter

The between-image filter is designed based on *Constraint 3*, which applies to tool wear characterization in various types of machining processes. Based on this constraint, the

images can be further filtered by recursively tracking each single wear region along the temporal direction. The mechanism of this filtering algorithm is illustrated in Figure 19 and verified by the posterior probabilistic derivations using Bayes' Theorem (based on an uninformative prior) [132] included in Appendix B, and the results are summarized in Table 6, where

$$\theta = P(R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset \mid R_{e,t}^k \in \text{Information});$$

$$\theta' = P(R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset \mid R_{e,t}^k \in \text{Noise}).$$

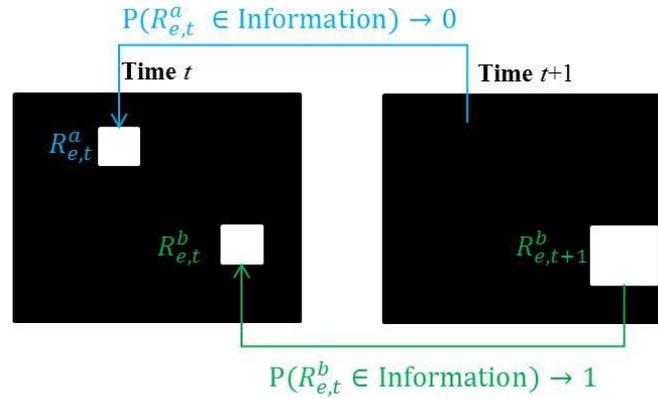


Figure 19: An illustration of between-image filtering

Table 6: The posterior probability of $R_{e,t}^k$ given the observation of time $t + 1$

Event	Posterior Probability
$R_{e,t}^k \in \text{Information} \mid R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset$	$\frac{1}{1 + \frac{\theta'}{1-\theta}} \rightarrow 1$
$R_{e,t}^k \in \text{Noise} \mid R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset$	$\frac{1}{1 + \frac{1-\theta}{\theta'}} \rightarrow 0$
$R_{e,t}^k \in \text{Information} \mid R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset$	$\frac{1}{1 + \frac{1-\theta'}{\theta}} \rightarrow 0$
$R_{e,t}^k \in \text{Noise} \mid R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset$	$\frac{1}{1 + \frac{\theta}{1-\theta'}} \rightarrow 1$

The between-image filtering is performed by recursively tracking $R_{e,t}^k$ ($k = 1, 2, \dots, K_{e,t}$; given $K_{e,t} > 0$) to $R_{e,t+1}^\bullet$ to see if there are any wear region(s) overlapping

with $R_{e,t}^k$ at its neighborhood in $BI_{e,t+1}$. If $R_{e,t}^k$ does not have any overlapping region with $R_{e,t+1}^\bullet$, it is highly probable that $R_{e,t}^k$ is noise in the image $BI_{e,t}$, and thus needs to be eliminated.

Using the between-image filtering, the wear regions from time 1 to t can be automatically updated when the image of time $t + 1$ is obtained. This algorithm is able to remove noise in the binary images for tool wear characterization, especially when images are collected frequently from one cutting edge over time.

After the two filtering procedures, the wear area on each specific cutting edge is extracted and used as the degradation measure for the individual cutting edge. According to [126], wear area is a more robust feature than the traditional flank wear features.

5.2 Case Study

In this section, one cases based on a hexagonal broach is introduced to demonstrate how the proposed methodology are used to characterize tool degradation based on different knowledge on different machining processes.

5.2.1 *Experimental Setup*

The experimental setup used in the case study includes the setup for the broaching tests and a machine vision system for tool wear inspection, which have been both introduced in Chapter 3. There are six samples involved in the dataset which correspond to the six series of cutting edges on one hexagonal broach to shape the six 120°-degree angles of the hexagon (shown in Figure 20).

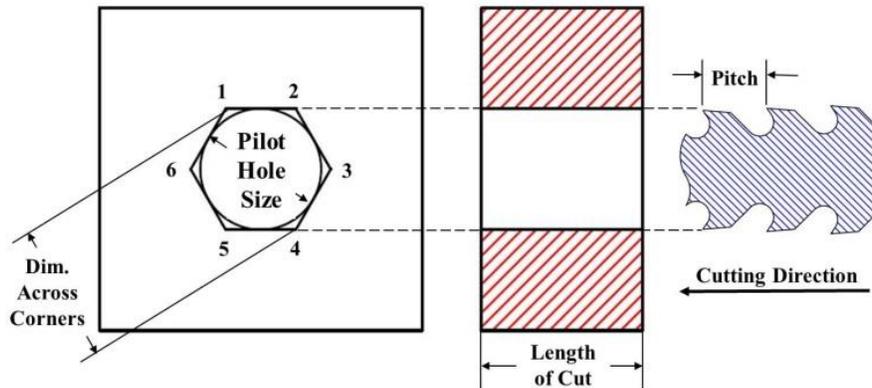


Figure 20: Sketch of the broach and workpiece

The degradation process of the broach has been accelerated by the change of workpiece material. The broach is originally designed to cut carbon steel, while in the accelerated experiments stainless steel was used to accelerate the tool wear progression.

The machine vision system records image data via a Mitutoyo QI-B3017B Quick Image Telecentric Lens System digital microscope (Figure 7), and the images are captured from multiple cutting edges over time by the image acquisition procedure introduced in [126].

5.2.2 Image Preprocessing and Process Knowledge Constrained Image Filtering

The collected images are processed using the proposed preprocessing and process knowledge based filtering methods. In this subsection, examples of the image processing results obtained from each step are illustrated.

5.2.2.1 Image Preprocessing

The collected color images are first converted to gray-scale images. Then image registration and segmentation are performed, as shown in Figure 21(a). Finally, the adaptive thresholding scheme is adopted to obtain the binary images, as shown in Figure 21(b). All the illustrated examples are images taken from one single cutting edge e_0 from consecutive time points t_0 , $t_0 + 1$, and $t_0 + 2$. It can be observed from Figure 21(b) that

the thresholded images contain a significant number of ineffective wear regions, which need to be eliminated.

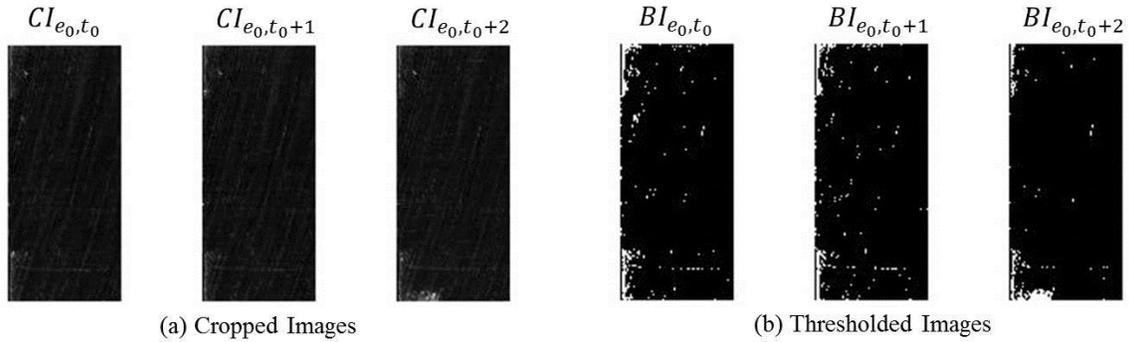


Figure 21: Images after cropping (a) and thresholding (b)

5.2.2.2 Process Knowledge Constrained Image Filtering for Single Cutting Edge

The binary images obtained from the previous section are filtered based on within-image and between image filtering to satisfy the three constraints in tool degradation characterization proposed in Section 5.1.3. Thus the effective wear regions in each image are obtained. Examples of images after the two filtering schemes are shown in Figure 22, where Figure 22(a) shows that a significant amount of within-image noise has been removed from the binary images in Figure 21(b); Figure 22(b) shows that some between-image noise is identified and eliminated from Figure 22(a).

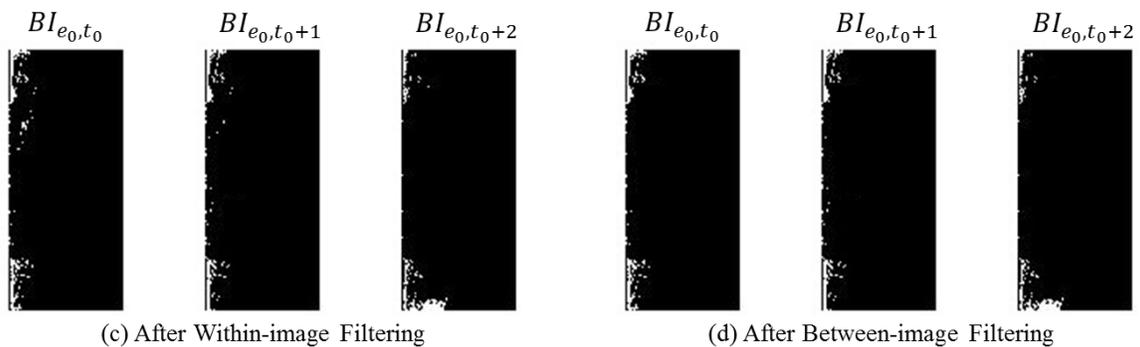


Figure 22: Images after within-image (a) and between-image (b) filtering

5.2.3 Performance Comparison

In this section, three different performance measures are used to evaluate the effectiveness of the proposed process knowledge constrained wear characterization methodology: 1) signal-to-noise ratio of the wear characterization for individual cutting edges; 2) violation of monotonicity; 3) average variance in the wear area values of different samples with the same design when a failure occurs. The performance measures are introduced as follows.

The Signal-to-Noise Ratio (SNR) for the wear characterization based on different image process techniques can be calculated based on the following equation as suggested in [133]. The SNR for sample j is defined as follows,

$$snr^j = \frac{\sum_k \hat{f}_j(t_k)^2}{\sum_k [f_j(t_k) - \hat{f}_j(t_k)]^2}, j = 1, \dots, J$$

where system failures can be identified when a non-conforming part occurs due to excessive tool wear. In addition, $\hat{f}(t_k)$ represents the estimate of $f(t_k)$ at time t_k for unit j . For more information about the model used for $f_j(t_k)$, please refer to Section 6.5.1.

The violation of monotonicity is evaluated using the total weighted sum of violations in monotonicity for each individual cutting edge as follows.

$$M_v = \sum_{j=1}^J \sum_{k=1}^{N_j-1} c_{k,j} \varepsilon_{k,j}$$

where N_j represents the number of observations available until a failure occurs, and $\varepsilon_{k,j}$ is the slack variable that measures the amount of violation in monotonicity which can be calculated as follows,

$$\varepsilon_{k,j} = \max\{0, \tilde{W}_a(t_k) - \tilde{W}_a(t_{k+1})\}$$

where $\tilde{W}_a(t_k)$ represents the standardized area of wear region, and the standardization is performed to scale the ranges of the wear progression paths of different units.

As the degradation progresses, the violation of monotonicity is increasingly crucial to the remaining useful life prediction of the unit. Therefore, geometric series is used as the weights $c_{k,j}$,

$$c_{k,j} = c_{k,1} \times q^{j-1}$$

where q satisfies $c_{k,1} \times q^{j-1} - q + 1 - c_{k,1} = 0$. This definition of the weighting series is suggested in [99] and [101].

The performance measures are summarized in Table 7 for the image processing results before and after applying the proposed process knowledge constrained image filters, where the variance of random failure thresholds for each group of identical units can be calculated, and an average value of all the variances are obtained, and denoted as average VF . The last column in Table 7 summarized the p -values for the pairwise t test for the two compared scenarios to identify if there is a significant difference in performance of the two compared scenarios.

It can be observed that after applying the proposed process knowledge constrained image filters, both average snr and average M_v are improved with average snr increases significantly and the violation of monotonicity M_v decreases significantly. In addition, the average VF is not significantly affected by the proposed image filters as the p -value for the pairwise t test is 0.2649.

Table 7: Performance comparison before and after applying proposed image filters

	Intensity-based Method Only	Proposed Image Filters	p -value
Average snr	13.9665	23.4251	0.0112
Average M_v	0.1525	0.1088	4.0155e-4
Average VF	6.8174	8.0404	0.2649

Therefore, it can be observed from the case study that the proposed process knowledge constrained image filtering method can increase the Signal-to-Noise Ratio for degradation modeling and reduce the violation of monotonicity without significantly affecting the variance of random failure thresholds.

5.3 Conclusions

In this section, a process knowledge constrained image filtering method, including within-image and between-image filtering, is proposed for wear characterization for broaching processes. The constraints proposed can be incorporated with any intensity-based wear characterization methodology to enhance the performance of tool wear characterization a more engineering viable way for various machining processes. A case study based on a broaching process is introduced to show the effectiveness of the proposed methodology by comparing the wear characterization results before and after applying the proposed process knowledge constrained image filters right after the intensity-based method.

6. Model Integration for System-Level Degradation Modeling in Multi-Edged Machining Processes

6.1 Acronyms and Abbreviations

CI	Confidence Interval
CDF	Cumulative Density Function
HI	Health Index
PDF	Probability Density Function
RUL	Remaining Useful Life

6.2 Notations

V	Volume loss at cutting edge
ω	Worn layer width on the relief face
W	Measured flank wear
VB	Actual flank wear
θ	Relief angle on the cutting edge
γ	Rake angle on the cutting edge
m	Mass loss at the cutting edge
L_c	Cumulative cutting thickness
ρ_m	Tool material density
v_c	Cutting speed
t	Cutting time
$\tilde{F}(W_a)$	Integral function of W_a which has a linear degradation path
$L(t_k, v_c)$	System-level HI at time t_k given v_c

L_k	System-level HI at time t_k assuming v_c is constant
S_k	Standard Error of the linear model regressing $\tilde{F}(W_a)$ on ω at time t_k
β	Linear degradation path parameter for system-level prognostics
λ	Random failure threshold for system-level HI
μ_β	Prior mean of linear degradation path parameter β
σ_β^2	Prior variance of linear degradation path parameter β
μ_λ	Prior mean of random failure threshold λ for system-level HI
σ_λ^2	Prior variance of random failure threshold for system-level HI
ρ	Prior correlation coefficient between β and λ
σ^2	Linear model fitting error variance based on system-level HI
\tilde{X}	Posterior parameter of X
$\Phi(\cdot)$	CDF of standard normal distribution
$\phi(\cdot)$	PDF of standard normal distribution
$F_{T L_1, \dots, L_k}(t)$	CDF of RUL conditioning on system-level HI (t_1, \dots, t_k)
$f_{T L_1, \dots, L_k}(t)$	PDF of RUL conditioning on system-level HI (t_1, \dots, t_k)

6.3 Introduction

Remaining useful life (RUL) prediction is a critical task in various complex manufacturing processes. An accurate RUL prediction methodology can be used for condition-based maintenance planning to ensure final product quality and reduce machine downtime due to tool failures. Machining process degradation modeling and RUL prediction has been studied in a large variety of research efforts, which has focused on the study of tool life modeling and in-situ RUL prediction. For tool life modeling, the

Taylor tool life equation is most widely used to empirically model the tool life [134, 135], and Weibull distribution has been used to model the tool life distribution [136]. For in-situ RUL prediction, hidden Markov models have been used for various machining tools [60, 61, 63, 137]. Other approaches include RUL prediction based on engineering knowledge and hazard rate function [138], Bayesian inference methods [64, 67], and various other statistical learning approaches [139-141].

For RUL prediction of machining tools, most of the reported studies focus on individual cutting edges, where a health index (HI) can be constructed based on either features extracted from process sensor signals [48, 142] or direct observations of the wear land on the cutting edge [47, 129]. With rapid development of machine vision systems, image-based tool wear characterization gained increasing popularity as the most direct degradation measure for the tool wear phenomenon [3, 46, 126, 129]. Therefore, this paper focuses on degradation modeling based on direct degradation characterization extracted from high-resolution digital images. Similar to other degradation modeling problems, the HI is characterized by a stochastic model constructed based on a training data set, and the RUL is predicted by inferencing how long it takes for the HI to first pass a predefined threshold value [87, 88, 98, 143, 144].

The large number of cutting edges involved distinguishes broaching process from other machining processes for RUL prediction. In a broaching process, each cutting edge is responsible to shape the workpiece into an intermediate geometry [135, 145]. Although each cutting edge has its own wear characterization as a component-level HI, the downstream cutting edges can, to some extent, compensate the inferior shape integrity finished by their upstreaming cutting edges. The multistage structure makes it difficult to

predict the RUL distribution, and there are no predefined wear thresholds available for each individual cutting edge. The decisions for tool replacement and re-sharpening should be made based on a system-level HI which combines the tool wear measurements of all the cutting edges involved in the process. However, there are only very limited research efforts proposed for system-level prognostics, including a system-level approach that assumes the probability density functions (PDFs) of the RULs of individual components are known [97], and a system-level prognostics scheme given the relationship between the component and system-level degradation [74]. Moreover, there is very limited research reported regarding degradation modeling and RUL prediction for multi-edged machining tools such as broaching tools, and only a few studies on the tool wear characterization for individual cutting edges [51, 113, 126] and multiple cutting edges based on functional descriptors [3] for broaching tools.

In addition to the problems of multiple cutting edges, in practice there can be dependency between the degradation path parameters and the corresponding HI values when a failure occurs. In a general degradation path model, considering the correlation between the failure HI values and the degradation path parameter can enhance the RUL prediction accuracy if the dependency exists. However, to the best of our knowledge, the existing studies usually focus on problems with no correlation existing between the HI failure value and the degradation path parameter(s) [99, 100, 102, 103].

To fill these research gaps, this paper proposes a novel system-level degradation modeling and prognostics approach by integrating the physical model of tool wear progression process and a varying-coefficient model for broaching process prognostics. Instead of being purely data-driven like lots of degradation modeling methods [76, 77],

the proposed methodology makes full use of the physical process knowledge, where a physical process-based flank wear progression model is used to construct the system-level HI for degradation modeling. Additionally, a Bayesian approach framework is used for RUL prediction by incorporating the possible correlation between the random failure threshold and degradation path parameter(s). The rest of this paper is organized as follows. Section II briefly reviews the relevant literature. Section III develops a novel system-level prognostics methodology based on model integration. Section IV demonstrates the effectiveness of the proposed methodology based on a case study and a numerical simulation study. Finally, Section V presents the conclusion and future research directions.

6.4 Methodology Development

This section introduces the proposed model integration methodology. The physical process model is first derived with respect to the area of the wear region on the cutting edge; a function of the wear areas across all the measured cutting edges are then summarized by a system-level HI, which is nested in the varying-coefficient model for degradation modeling and prognostics.

6.4.1 Physical Process Model

(1) Tool Wear Geometric Relationship

Most physical tool wear progression models characterize the wear ratio in terms of the instant volume loss on the cutting edge. However, it has been demonstrated that the area of wear region is a robust attainable feature for broaching processes [3]. Therefore, the

proposed physical process model for tool wear progression is in terms of area of wear region.

For orthogonal machining processes, the geometric relationship between the area of wear region W_a and volume loss V on the cutting edge can be approximated as follows.

$$V = A \cdot \omega = \frac{1}{2} W \cdot VB \cdot \tan\theta \cdot \omega \quad (15)$$

where ω represents the worn layer width on the relief face, W represents the measured flank wear from the cutting edge, and it can be approximated as $W = \frac{W_a}{\omega}$ by assuming that the worn layer width always equals to the designed chip width for each specific cutting edge. $VB = \frac{W}{1 - \tan\theta \tan\gamma}$ represents the real flank wear, where θ represents the relief angle and γ represents the rake angle on the cutting edge, as shown in Figure 23. After simplification, (15) can be simplified as follows. The derivation from (15) to (16) is very similar with the one proposed in [59].

$$V = \frac{W_a^2 \cdot \tan\theta}{2\omega(1 - \tan\theta \tan\gamma)} \quad (16)$$

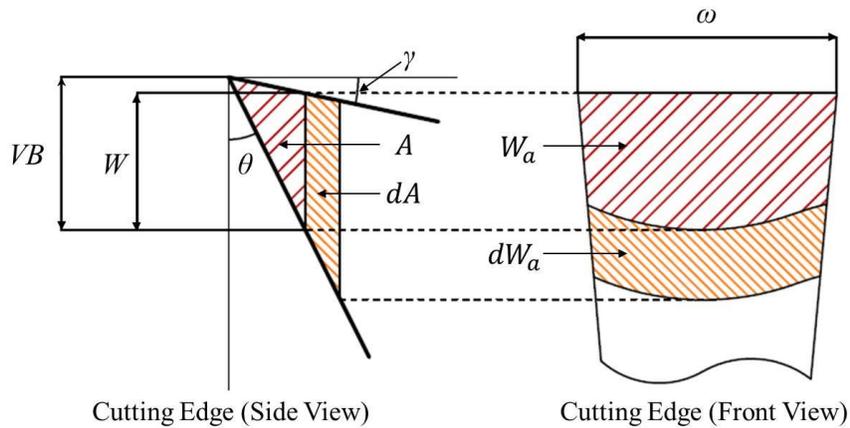


Figure 23: Geometric relationship between the area of wear region and volume loss on the cutting edge

(2) Physical Process Model in Terms of the Area of Wear Region

In this paper, the tool wear progression model proposed in [59] is adopted as the physical process model to describe the wear rate as a nonlinear differential equation as follows.

$$\frac{dm}{dL_c} = \frac{\rho_m}{v_c} \frac{dV}{dt} = C_1 + C_2 \exp\left\{-\frac{Q}{R \cdot \varphi(W)}\right\} \quad (17)$$

On the left side of (17), m represents the mass loss on the cutting edge, L_c is the cumulative cutting thickness, ρ_m and v_c are the density of the tool material and the cutting speed, respectively. On the right side, C_1 , C_2 , Q , and R are constants, and $\varphi(W)$ represents the cutting temperature as a function of flank wear measurement W .

By substituting (16) into (17), a differential equation with respect to W_a can be obtained as follows,

$$d\left\{\frac{W_a^2 \cdot \tan\theta}{2\omega(1 - \tan\theta \tan\gamma)}\right\} = \frac{v_c}{\rho_m} \left\{C_1 + C_2 \exp\left[-\frac{Q}{R \cdot \varphi(W)}\right]\right\} dt$$

$$\frac{1}{F(W_a)} d\left(\frac{W_a^2}{2}\right) = \frac{v_c \omega (1 - \tan\theta \tan\gamma)}{\tan\theta} dt$$

where $F(W_a) = \frac{1}{\rho_m} \left\{C_1 + C_2 \exp\left[-\frac{Q}{R \cdot \varphi(W)}\right]\right\}$. Taking integration on both sides, the following equation can be obtained,

$$\int_{W_a(0)}^{W_a(t_k)} \frac{W_a}{F(W_a)} dW_a = \int_0^{t_k} \frac{v_c \omega (1 - \tan\theta \tan\gamma)}{\tan\theta} dt$$

$$\int_{W_a(0)}^{W_a(t_k)} \frac{W_a}{F(W_a)} dW_a = \frac{v_c (1 - \tan\theta \tan\gamma)}{\tan\theta} t_k \cdot \omega = L(t_k, v_c) \cdot \omega \quad (18)$$

where $L(t_k, v_c) = v_c \frac{1 - \tan\theta \tan\gamma}{\tan\theta} t_k$. Therefore, the integral function of W_a at specific cutting edge is proportional to both the worn layer width ω on the relief face and the cumulative cutting time t_k . (18) illustrates the cumulative tool wear from time 0 to time t_k . Denote the left hand side of the equation as $\Gamma(W_a)$, then (18) can be rewritten as

$$\Gamma(W_a) = L(t_k, v_c) \cdot \omega \quad (19)$$

Therefore, $L(t_k, v_c)$ quantifies the relationship between the measured area of wear regions and the respective cutting edge worn layer width on the relief face of multiple cutting edges, and is a consistent degradation measure across different cutting edges.

It should be noted that $L(t_k, v_c)$ is a proper degradation measure under the assumption that the initial wear on the broaching tool is negligible compared to the cumulative wear progression over machining time. This is a reasonable assumption in broaching processes as the initial wear on a few of the cutting edges are only due to the accidental error in the relief operation in the broach production process.

By incorporating an error term in (19), an estimate of $L(t_k, v_c)$ can be obtained using all the measurements which can be used to summarize the wear measurements from all the cutting edges. This is a robust estimate of $L(t_k, v_c)$ which can combine the degradation measures of multiple cutting edges. The form of $\Gamma(W_a)$ is determined by different cutting tool materials and cutting process parameters. In the cases that there are no closed form for $\Gamma(W_a)$, it can be obtained from numerical integration based on the series of measures of $W_a(t_k)$.

6.4.2 System-level Degradation Model

In this section, the physical process model is nested in a varying-coefficient model for system-level prognostics. First, the system-level HI is estimated based on the areas of wear regions on the cutting edges measured. Second, the form of the degradation path is determined, and the degradation path parameter can be updated by combining historical degradation path parameter distribution and in-situ HI observations. Third, the RUL distribution is determined with a point estimate and an interval estimate as well.

(1) Estimation of the System-level HI

Based on the physical process model obtained in Section 6.4.1, $L(t_k, v_c)$ is used as the system-level HI for degradation modeling. In addition, the parametric form of the degradation path is linear with respect to time t_k and cutting speed v_c , as illustrated in (18). To quantify the randomness in the model, one error term is incorporated into (5), to absorb the uncertainties in the model as follows,

$$\Gamma(W_a) = L(t_k, v_c) \cdot \omega + E(t_k); E(t_k) \sim N(0, S_k^2) \quad (20)$$

$L(t_k, v_c)$ can be estimated by fitting a linear regression model without an intercept based on the vectors of the worn layer width on the relief face $\vec{\omega}_{[m_k \times 1]}(t_k)$, S_k^2 represents the variance of the error term at time t_k , and the measured area of wear region on the cutting edges $\vec{W}_{a[m_k \times 1]}(t_k)$ at time t_k , where m_k is the number of cutting edges observed at time t_k . It is worth noting that neither the measured cutting edges nor the number of the measurements need to be consistent across different evaluation time t_k or different observational units. This provides the proposed methodology significant flexibility which cannot be provided by those existing data-driven composite HI construction and their respective degradation modeling methods [99-101].

(2) Degradation Path Parameter Update

The derivation for the posterior joint distribution of the degradation path parameter and threshold value is introduced in this subsection. It is illustrated in (18) that the system-level HI $L(t_k, v_c)$ is a linear function of the cumulative cutting time t_k , and an error term can be used to absorb the uncertainty in the model.

$$L(t_k, v_c) = \frac{1 - \tan\theta \tan\gamma}{\tan\theta} v_c t_k + \varepsilon_k, \varepsilon_k \sim N(0, \sigma^2) \quad (21)$$

Without losing generality, the cutting speed can be assumed constant. Thus (21) can be rewritten as

$$L(t_k) = L_k = \beta t_k + \varepsilon_k, \varepsilon_k \sim N(0, \sigma^2) \quad (22)$$

In (22), the degradation path parameter β can be obtained by combining a prior parameter distribution obtained from historical data and the in-situ image observations from each tested individual sample. With a Bayesian framework, the RUL distribution can be updated given a failure threshold or threshold distribution [87, 99]. Unlike assuming the independence between the random threshold distribution and the degradation path parameter [99-101, 146], a more generic case is considered where the degradation path parameter and the threshold value are assumed to follow a bivariate normal distribution. This case can be simplified to the independent scenario when the correlation coefficient equals to zero.

The prior joint distribution of the linear degradation path parameter and the failure threshold is assumed to be $\pi \begin{pmatrix} \beta \\ \lambda \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_\beta \\ \mu_\lambda \end{pmatrix}, \begin{bmatrix} \sigma_\beta^2 & \rho\sigma_\beta\sigma_\lambda \\ \rho\sigma_\beta\sigma_\lambda & \sigma_\lambda^2 \end{bmatrix} \right)$. By assuming the system-level HI follows a normal distribution, the conditional density function of k HI observations given β and λ can be expressed as

$$f(L_1, \dots, L_k | \beta, \lambda) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^k \exp \left\{ - \sum_{i=1}^k \frac{(L_i - \beta t_i)^2}{2\sigma^2} \right\}$$

Given the prior distribution of β and λ , the posterior distribution of (β, λ) can be derived as follows.

$$\begin{aligned}
P(\beta, \lambda | L_1, \dots, L_k) &\propto f(L_1, \dots, L_k | \beta) f(\beta | \lambda) f(\lambda) = f(L_1, \dots, L_k | \beta) \pi(\beta, \lambda) \\
&\propto \exp \left\{ - \sum_{i=1}^k \frac{(L_i - \beta t_i)^2}{2\sigma^2} \right\} \exp \left\{ \frac{-1}{2(1-\rho^2)} \left[\frac{(\beta - \mu_\beta)^2}{\sigma_\beta^2} \right. \right. \\
&\quad \left. \left. - \frac{2\rho(\beta - \mu_\beta)(\lambda - \mu_\lambda)}{\sigma_\beta \sigma_\lambda} + \frac{(\lambda - \mu_\lambda)^2}{\sigma_\lambda^2} \right] \right\} \\
&\propto \exp \left\{ - \frac{1}{2} \left[\left(\frac{\sum_{i=1}^k t_i^2}{\sigma^2} + \frac{1}{(1-\rho^2)\sigma_\beta^2} \right) \beta^2 + \left(\frac{1}{(1-\rho^2)\sigma_\lambda^2} \right) \lambda^2 \right. \right. \\
&\quad \left. \left. - 2\beta \left(\frac{\sum_{i=1}^k L_i t_i}{\sigma^2} + \frac{\mu_\beta}{(1-\rho^2)\sigma_\beta^2} - \frac{\rho\mu_\lambda}{(1-\rho^2)\sigma_\beta \sigma_\lambda} \right) \right. \right. \\
&\quad \left. \left. - 2\lambda \left(-\frac{\rho\mu_\beta}{(1-\rho^2)\sigma_\beta \sigma_\lambda} + \frac{\mu_\lambda}{(1-\rho^2)\sigma_\lambda^2} \right) - 2\beta\lambda \left(\frac{\rho}{(1-\rho^2)\sigma_\beta \sigma_\lambda} \right) \right] \right\} \quad (23)
\end{aligned}$$

When comparing (23) with the posterior normal distribution of (β, λ) , the parameters of the distribution are updated as

$$\begin{aligned}
\tilde{\mu}_\beta &= \frac{\sigma_\beta^2 \sum_{i=1}^k L_i t_i + (1-\rho^2)\sigma^2 \mu_\beta}{(1-\rho^2)(\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2)}; & \tilde{\sigma}_\beta^2 &= \frac{\sigma^2}{\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2} \sigma_\beta^2; \\
\tilde{\mu}_\lambda &= \mu_\lambda + \frac{\rho\sigma_\beta \sigma_\lambda (\sum_{i=1}^k L_i t_i - \mu_\beta \sum_{i=1}^k t_i^2)}{\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2}; & \tilde{\sigma}_\lambda^2 &= \frac{(1-\rho^2)\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2}{\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2} \sigma_\lambda^2; \\
\tilde{\rho} &= \frac{\sigma}{\sqrt{(1-\rho^2)\sigma_\beta^2 \sum_{i=1}^k t_i^2 + \sigma^2}} \rho \quad (24)
\end{aligned}$$

With these derived formulas in (24), the joint posterior distribution of (β, λ) can be updated based on the most recent estimated system-level HI.

(3) RUL Prediction

Based on the posterior normal distribution of (β, λ) , at time t_k , the system-level HI can be predicted as $L(t + t_k) = \beta \times (t + t_k) \sim N(\mu_L(t + t_k), \sigma_L^2(t + t_k))$, where $\mu_L(t + t_k) = \tilde{\mu}_\beta \times (t + t_k)$, and $\sigma_L^2(t + t_k) = (t + t_k)^2 \tilde{\sigma}_\beta^2 + \sigma^2$.

Denote the RUL at time t_k as T , given the observations of the system-level HI L_1, \dots, L_k , the conditional cumulative density function (CDF) of T can be expressed as $F_{T|L_1, \dots, L_k}(t) = P(T \leq t | L_1, \dots, L_k) = P(L(t + t_k) \geq \lambda | L_1, \dots, L_k)$. Given the distributions of predicted $L(t + t_k)$ and failure threshold λ , we can have

$$\lambda - L(t + t_k) \sim N(\tilde{\mu}_\lambda - \mu_L(t + t_k), \tilde{\sigma}_\lambda^2 + (t + t_k)^2 \tilde{\sigma}_\beta^2 + \sigma^2 - 2(t + t_k) \tilde{\rho} \tilde{\sigma}_\beta \tilde{\sigma}_\lambda)$$

Therefore, the conditional CDF of T can be obtained as $F_{T|L_1, \dots, L_k}(t) = \Phi(g(t))$,

where $g(t) = \frac{\mu_L(t+t_k) - \tilde{\mu}_\lambda}{\sqrt{\tilde{\sigma}_\lambda^2 + (t+t_k)^2 \tilde{\sigma}_\beta^2 + \sigma^2 - 2(t+t_k) \tilde{\rho} \tilde{\sigma}_\beta \tilde{\sigma}_\lambda}}$, and $\Phi(\cdot)$ denotes the CDF of the

standard normal distribution. In addition, at time t_k it is known that $T \geq 0$, therefore given the system has not failed at time t_k , the conditional CDF of T is

$$F_{T|L_1, \dots, L_k, T \geq 0}(t) = \frac{P\{0 \leq T \leq t | L_1, \dots, L_k\}}{P\{T \geq 0 | L_1, \dots, L_k\}} = \frac{\Phi(g(t)) - \Phi(g(0))}{1 - \Phi(g(0))} \quad (25)$$

The probability density function (PDF) of the RUL distribution can be obtained as follows, where $\phi(\cdot)$ represents the PDF of the standard normal distribution.

$$f_{T|L_1, \dots, L_k, T \geq 0}(t) = \frac{\phi(g(t))g'(t)}{1 - \Phi(g(0))} \quad (26)$$

With the CDF of the predicted RUL distribution in (25), a point estimator can be obtained for RUL by numerically finding the time point t with $F_{T|L_1, \dots, L_k, T \geq 0}(t) = 0.5$, and the confidence interval (CI) of the RUL given a specified confidence level α can be estimated as $[T_L, T_U]$, where T_L and T_U can be found by solving the equations $F_{T|L_1, \dots, L_k, T \geq 0}(t) = \alpha/2$ and $F_{T|L_1, \dots, L_k, T \geq 0}(t) = 1 - \alpha/2$, respectively.

In summary, the proposed methodology combines a physical process model derived based on first principles and a general degradation path model for RUL prediction of multi-edged machining tools. The model integration structure is illustrated in Figure 24.

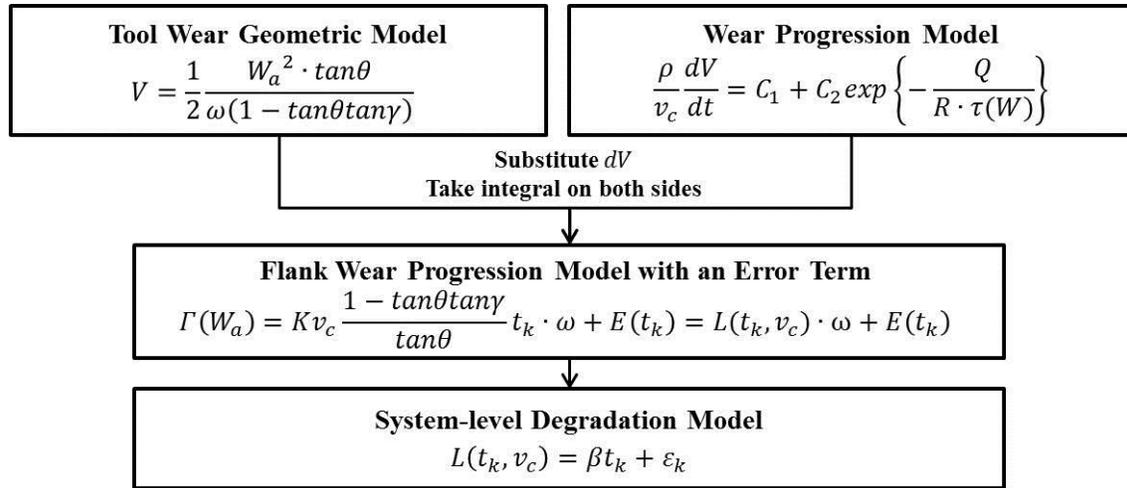


Figure 24: Proposed model integration structure

6.5 Case and Simulation Study

In this section, a case study based on a hexagonal broach is introduced to demonstrate how the proposed system-level prognostics methodology can be used for RUL prediction for a broaching tool. Additionally, a simulation study is conducted by generating data based on the distribution estimated from the case study.

6.5.1 Case Study

(1) Experimental Setup

The experimental setups used in the case study include the broaching test setup and a machine vision system for tool wear measurements. The broaching process is performed on a tensile test machine that can provide the broach with constant speed to travel through the pilot contour. The setup for the broaching process is shown in Figure 4 and closely resembles the operating setup in the industrial application. A 0.75-inch hexagonal broach produced by the duMONT company, LLC, is used in the experiments. The broach is made of high speed steel with 38 teeth and 0.375-inch pitch between consecutive teeth. There are six samples involved in the data set which correspond to the six series of

cutting edges on one hexagonal broach to shape the six 120°-degree angles of the hexagon (shown in Figure 20). The degradation process of the broach has been accelerated by the change of workpiece material. The broach is originally designed to cut carbon steel, while in the accelerated experiments stainless steel was used to accelerate the wear progression.

In the case study, the failure of the broaching process is defined as the occurrence of a severe contour anomaly, a breakout. In the broaching industry, a breakout is usually used to determine the tool replacement and resharpening operation by industrial practitioners. Several examples of the breakout anomalies compared to their corresponding healthy contour shapes are shown in Figure 25.

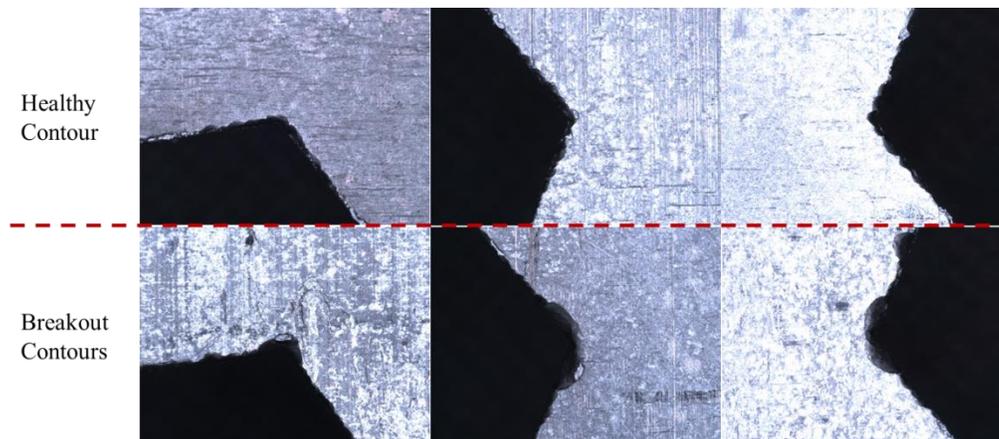


Figure 25: Examples of the breakout anomalies in the machined contours (bottom row) and their corresponding healthy contour shapes (top row)

The case study focuses on one single failure mode, the tool wear progression on the broaching tool. This is one of the major failure modes for the breakout failure to occur. The end of life for each sample is determined by the first time that the maximum area of wear region exceeds a threshold value. The threshold value can be determined by a maximum flank wear criterion, which is usually defined arbitrarily [51, 135, 147]. In this

case study, the threshold value was determined using the maximum area of wear region when the first breakout occurs in the final contours within all the historical observations.

The machine vision system records image data via a Mitutoyo QI-B3017B Quick Image Telecentric Lens System digital microscope (Figure 7), and the images are captured from multiple cutting edges by the image acquisition procedure introduced in [126], and the areas of wear regions are extracted based on the methods introduced in [3].

For the hexagonal broaching process studied in this section, the wear rate is approximately constant, as the temperature is relatively low (under 600°C) [55]. The form of $\Gamma(W_a)$ is also determined by the high speed steel cutting tool material and the relatively low cutting speed. Therefore, for this specific broaching process, the system-level HI can be derived as follows,

$$\Gamma(W_a) = \frac{1}{2}(W_a^2(t_k) - W_a^2(0)) = L(t_k, v_c) \cdot \omega$$

$$W_a^2(t_k) - W_a^2(0) = L'(t_k, v_c) \cdot \omega \quad (27)$$

Based on (27), the system-level HI at time t_k can be estimated using least squares estimation.

(2) RUL Prediction Results

The samples are tested by cross validation. It is observed that one of the samples has exceeded the failure threshold value for the maximum area of wear region, so that specific sample is discarded and five valid samples are used to test the effectiveness of the proposed approach.

In each fold, four of the samples are used as training data to obtain joint prior distribution of the degradation path parameter and threshold value for the proposed HI. It is worth noting that in this paper, broach degradation modeling with single failure mode

is considered. In practice, diagnostics should be performed to the historical observations to make sure that the prior data set only includes this one single failure mode, progression of tool wear. The remaining sample is then used to evaluate the RUL prediction performance as each tested observation is obtained, and the joint posterior distribution of the degradation path parameter and threshold value are updated based on the Bayesian framework derived in (24).

The RUL performance is evaluated based on the absolute percentage error as defined in (28) at different observational time percentages of each individual sample's lifetime, where t_k is current accumulative cutting time; T_{i,t_k} and \hat{T}_{i,t_k} represent the true and estimated RUL of unit i at time t_k , respectively. This performance measure has been used in various RUL prediction research [103].

$$e_{i,t_k} = \left| \frac{\hat{T}_{i,t_k} - T_{i,t_k}}{t_k + T_{i,t_k}} \right| \quad (28)$$

The RUL prediction results are demonstrated in Figure 26, where the means of the absolute percentage error are illustrated. The mean absolute percentage errors are calculated based on different percentage levels of observations used, which mimics the RUL prediction based on in-situ tool wear measurements. The benchmark methods, average and maximum of tool wear, is the most widely used descriptive statistics to summarize the degradation measures of multiple cutting edges. It can be observed from Figure 26 that our proposed method outperforms the average W_a^2 method and the maximum W_a^2 method during most of the lifetime of the tools.

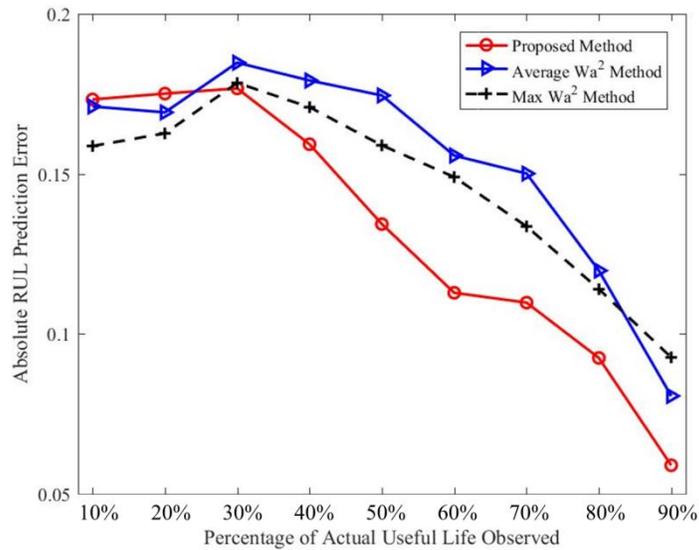


Figure 26: Comparison results of absolute value of the mean percentage error

Additionally, when it comes to the uncertainty of the RUL prediction, the proposed methodology outperforms both benchmark methods. The average widths of the 90% confidence intervals (CIs) at different observational time percentages are summarized in Figure 27. It can be observed that the proposed method yields smaller CI widths compared with the benchmark methods during most of the time in the lifetime of the tools. Specifically, the proposed method's average CI widths are only half of the ones provided by the maximum of the tool wear. Additionally, the average widths of the 90% CIs decrease as a higher observational time percentage of the individual sample's lifetime, which indicates less uncertainty in the RUL prediction as the cutting time approaches to the end of tool life.

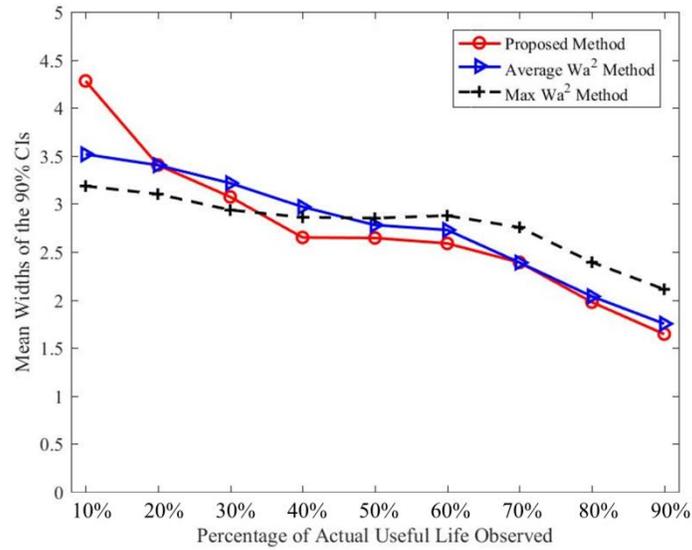


Figure 27: Comparison of average widths of the 90% CIs

6.5.2 Simulation Study

To further verify the performance of the proposed RUL prediction method, simulation studies have been performed based on simulated data which would resemble the properties of the data collected from the case study. The assumptions made in the simulation include *independence* between the system-level HIs over time, *independence* between individual observations of $W_a^2(t_k) - W_a^2(0)$ for unit i at time t_k , and *increasing modeling variance* over time t_k . The increasing variance over time can be justified by the data collected from the case study as shown in Figure 28. In addition, multiple different scenarios in the data correlation structures are used to test the effectiveness of the proposed method, which have been included in Appendix C.

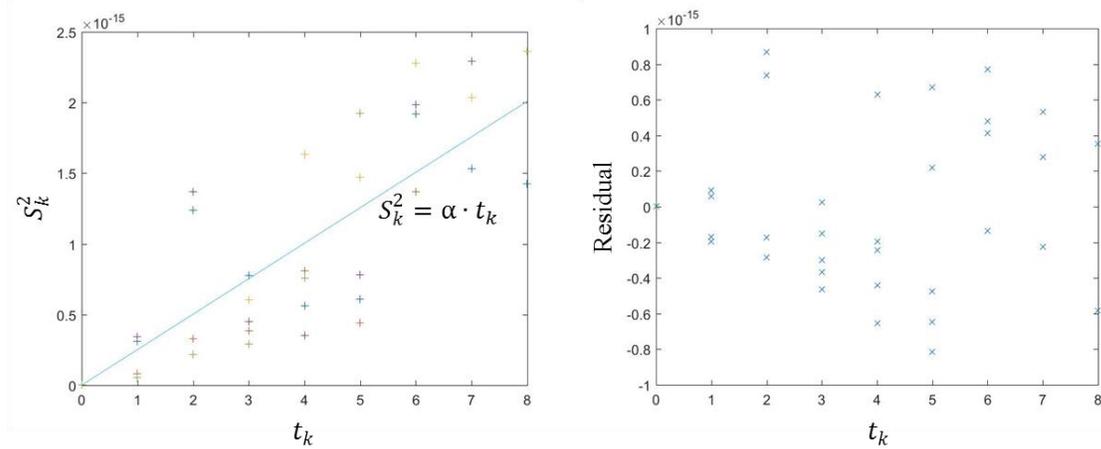


Figure 28: Modeling variance over time t_k with the fitted linear relationship (on the left panel) and the residual plot over time t_k (on the right panel)

(1) Simulation Procedure

The simulation procedure is described as follows:

Step 1: Parameter generation. Degradation path parameter and the threshold value for the system-level HI of each individual are generated from a joint normal distribution which is estimated from the case study data.

Step 2: Observation generation. The individual observations of $W_a^2(t_k) - W_a^2(0)$ can be obtained by the derived relationship in (27) with an error term included in the individual observations at each time t_k .

Step 3: HI estimation. HIs are estimated from all the observations collected at each time t_k , and the proposed system-level HI is obtained from least squares estimation.

Step 4: Iterate between *Step 2* and *3* until the proposed system-level HI exceeds the generated failure threshold value.

Step 5: RUL prediction. The posterior distribution of degradation path parameter and failure threshold value can be updated based on

(24) and the RUL prediction can be obtained based on the CDF in (25), where the prior distribution of the degradation path parameter and the failure threshold value are estimated from a training data set with 20 samples generated from the same procedure of *Steps 1 to 4*.

(2) Simulation Results

Five hundred individual degradation paths are generated and the RULs are predicted at each simulated observational time. Similar to the performance analysis in Section (1), the average RUL performance is evaluated based on the percentage error as defined in (28) at different observational time percentages of each individual sample's lifetime. The results of comparison are summarized in Figure 29.

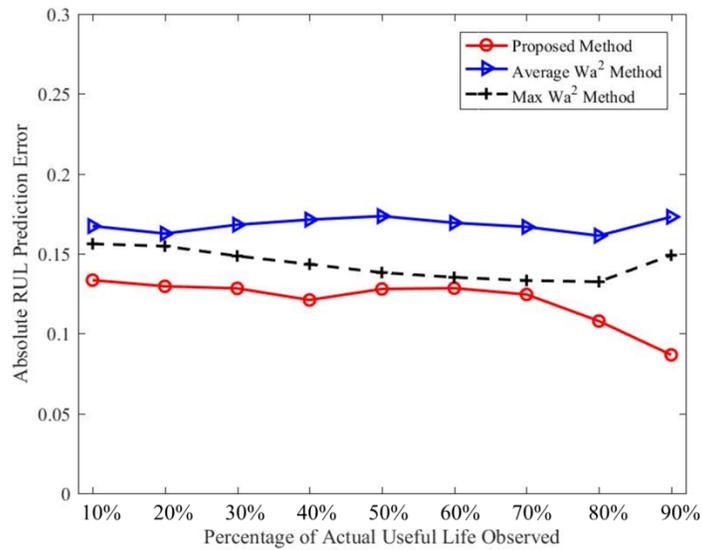


Figure 29: Comparison results of the mean absolute percentage error

Additionally, the average widths of the 90% CIs at different observational time percentages are summarized in Figure 30. It can be observed that the average widths of the 90% CIs decrease as a higher observational time percentage of the individual

sample's lifetime, and the proposed method yields a smaller average widths of the 90% CIs almost all the time across the lifetime of the tools.

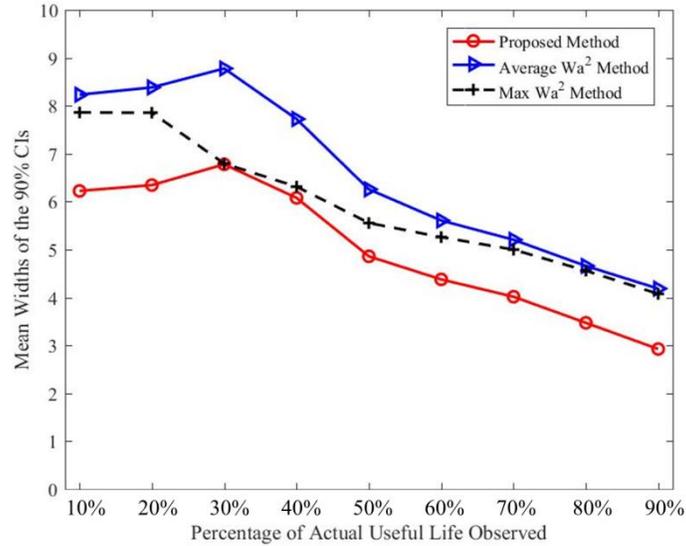


Figure 30: Comparison of average widths of the 90% CIs

6.6 Conclusions

Degradation modeling and RUL prediction for machining processes with multiple cutting edges is a challenging task. The rapid development of machine vision systems makes it possible to use the direct degradation measures, tool wear measurements on the multiple cutting edges, to construct the HI for degradation modeling and RUL prediction for system-level condition-based maintenance planning.

The contributions of this paper are twofold. First, based on a physical process model of tool wear progression, a system-level HI and its corresponding degradation path are derived and expressed as a function of the observable degradation features from the images captured from multiple cutting edges. Second, the Bayesian framework for posterior distribution update is derived to take into consideration the dependency between the degradation path parameter and the random failure threshold value. The effectiveness

of the proposed methodology is verified by both case and simulation study based on a hexagonal broaching process.

The proposed methodology applies to various multi-edged machining processes, including the fir-tree shaped broaching process in aerospace industry. For the broaching process to produce the turbine disc, the form in (18) needs to be adjusted based on the material properties of the workpiece and the broaching tool. Furthermore, the observational unit of the cumulative cutting time t_k for this specific process would be number of turbine discs, and each turbine disc includes a fixed number of cuts by one single broaching tool. With proper adjustment to the observational unit, the proposed method can be applied to the fir-tree shaped broaching process in aerospace industry.

7. Contributions and Future Research Directions

This chapter summarizes the contributions of this dissertation and highlights future research directions for process monitoring and prognostics for broaching processes.

7.1 Contributions

Monitoring and prognostics methodologies proposed in this dissertation provide avenues for incorporating process knowledge into statistical modeling for monitoring and prognostics.

7.1.1 Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profiles

The major contribution of this research is to formulate the multistage process with non-repeating cyclic patterns with multiple streams, each of which will be responsible to detect a specific type and location of possible process shifts. Then the multi-stream process is monitored using a group EWMA control chart to monitor the multiple streams at the same time.

The other important contribution of this research is that the proposed monitoring scheme provides sufficient information for fault diagnosis to identify the type and location of the fault whenever it is detected.

7.1.2 Process Knowledge Constrained Image Filtering for Effective Tool Wear Characterization with Noisy Images

The contribution of this research is to present machine vision based wear characterization from a process knowledge perspective. It is so far the only published research to use process knowledge constrained image filters to incorporate the knowledge of wear mechanism in image processing [3]. In fact, these process knowledge based constraints

can be incorporated with any intensity-based image processing methodology to enhance the wear characterization and remove ineffective wear regions which do not have an effect on the cutting performance and final product quality. With a more effective wear characterization as a direct degradation measure, the degradation modeling and remaining useful life prediction for the cutting tools can also be potentially improved.

7.1.3 Model Integration for System-level Degradation Modeling and Prognostics for a Multi-edged Machining Tool

The major contribution of this research is to derive a physical model driven health index to combine the tool wear conditions of numerous cutting edges. With the physical process knowledge, the degradation of the multi-edged tool can be evaluated by the relationship change between a function of the measured area of wear region and the cutting edge design parameter, which does not require consistency in either the number or the locations of observations across time or different tools.

Another important contribution of this research is that the Bayesian framework for posterior distribution update for remaining useful life prediction is derived to take into consideration the possible dependency between the degradation path parameter(s) and the random failure threshold value. This can be very helpful to deal with prognostics problems with hard failures, which has been studied a lot in the most recent years [148, 149].

7.2 Future Research Directions

The approaches proposed within this dissertation have provided future research directions for incorporating process knowledge into statistical monitoring and prognostics, which are discussed in the following subsections.

7.2.1 *Process Knowledge Constrained Image Filters for Wear Characterization for General Machining Processes*

Process knowledge constrained image filters can be proposed for wear characterization for general machining processes, such as turning and milling process. Due to the differences in the cutting speed of various machining process, additional process constraints can be designed to eliminate the effect of excessive heat accumulation on the surface of the tool, which is not a concern for broaching processes. Additionally, due to difference in the tool production processes and tool geometries, various other geometric descriptors might be more robust to the noise in the images than area of wear region used for broaching processes. For example, for turning processes, due to the curvature in the design of the cutting edges, the volume of loss on the cutting edge can be a more robust attainable degradation measure based on the designed geometry [150, 151].

Furthermore, multiple images can be captured for one machining tool with different orientations and even different lighting conditions for comprehensive feature extraction for tool wear characterization. This would be especially applicable for machining tools with complicated curvatures, such as turning inserts. Unlike broaching tools, the turning inserts usually have a curved nose on the cutting edge [135]. The curved geometry of the cutting edge leads to difficulties to capture the whole region of interest in one image. One possible solution is to capture images from multiple orientations from the same cutting

edge to quantify the nose wear, major and minor flank wear from the single cutting edge. An example of using multiple images to quantify tool wear on a curved cutting edge is illustrated in Figure 31.

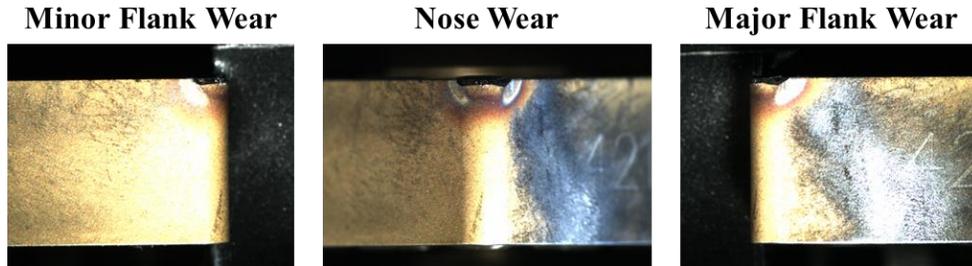


Figure 31: Using multiple images to quantify wear on a curved flank face

7.2.2 Model Integration for System-level Degradation Modeling for Multi-edged Machining Processes

Chapter 6 proposes to integrate tool geometry model and flank wear progression model to derive the form of the degradation path for system-level prognostics. The proposed methodology assumes that the worn layer width on the relief face ω for specific a cutting edge is constant over time, and approximately equals to the designed chip width of the cutting edge. However, this assumption would be possibly violated at the starting phase of the tool degradation. To release this assumption, (18) needs to be modified by replacing ω with a function $\omega(t)$, which will change the results of the integration with respect to time t . The form of the function $\omega(t)$ can be fitted based on historical observations.

There are two other limitations in the proposed method: 1) the initial wear at $t = 0$ is assumed to be negligible comparing to the cumulative wear resulted from the cutting process; 2) the tool wear inspection needs to start at the beginning of the tool life ($t = 0$) as the proposed health index quantifies the accumulative wear regions from time $t = 0$ to

t_k . To release these two limitations, a novel HI with the initial condition taken into consideration is needed.

Additionally, due to the possible spatial correlation in the tool wear progression process in the adjacent cutting edges, the proposed approach in Chapter 6 can be compared with other data driven approaches, such as the spatial-temporal degradation modeling approach proposed by Liu *et al.* [152].

7.2.3 Prognostics-Oriented Inspection Allocation for Multi-edged Machining Tools by Fusing Heterogeneous Data Sources

Due to the flexibility of the proposed method in Chapter 6, prognostics-oriented inspection allocation strategy can be proposed to balance the inspection cost and prediction error for in-situ prognostics for multi-edged machining processes. Furthermore, a data fusion methodology can be applied to combine the image data (direct degradation measure) with the real-time process data (indirect degradation measure), such as cutting force signal, vibration signal, and so forth, to achieve a cost-effective RUL prediction performance with satisfactory prediction accuracy.

8. References

- [1] J. Loizou, W. Tian, J. Robertson, and J. Camelio, "Automated wear characterization for broaching tools based on machine vision systems," *Journal of Manufacturing Systems*, vol. 37, pp. 558-563, 2015.
- [2] W. Tian, R. Jin, T. Huang, and J. A. Camelio, "Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profiles," *IIE Transactions*, vol. 49, pp. 320-331, 2016.
- [3] W. Tian, L. J. Wells, and J. A. Camelio, "Broaching Tool Degradation Characterization Based on Functional Descriptors," in *ASME 2016 11th International Manufacturing Science and Engineering Conference*, 2016, pp. V002T04A030-V002T04A030.
- [4] L. Kang and S. Albin, "On-line monitoring when the process yields a linear," *Journal of quality Technology*, vol. 32, pp. 418-426, 2000.
- [5] W. H. Woodall, M. A. Mohammed, J. M. Lucas, and R. Watkins, "The Use of Control Charts in Health-Care and Public-Health Surveillance," *Journal of Quality Technology*, vol. 38, p. 89, 2006.
- [6] B. M. Colosimo, Q. Semeraro, and M. Pacella, "Statistical Process Control for Geometric Specifications: On the Monitoring of Roundness Profiles," *Journal of Quality Technology*, vol. 40, pp. 1-18, 2008.
- [7] J. Jin and J. Shi, "Automatic feature extraction of waveform signals for in-process diagnostic performance improvement," *Journal of Intelligent Manufacturing*, vol. 12, pp. 257-268, 2001.

- [8] R. Jin and K. Liu, "Multimode variation modeling and process monitoring for serial-parallel multistage manufacturing processes," *IIE Transactions*, vol. 45, pp. 617-629, 2013.
- [9] K. Kim, M. A. Mahmoud, and W. H. Woodall, "On the monitoring of linear profiles," *Journal of Quality Technology*, vol. 35, p. 317, 2003.
- [10] W. H. Woodall, D. J. Spitzner, D. C. Montgomery, and S. Gupta, "Using control charts to monitor process and product quality profiles," *Journal of Quality Technology*, vol. 36, pp. 309-320, 2004.
- [11] C. Zou, Y. Zhang, and Z. Wang, "A control chart based on a change-point model for monitoring linear profiles," *IIE transactions*, vol. 38, pp. 1093-1103, 2006.
- [12] Y. Ding, L. Zeng, and S. Zhou, "Phase I analysis for monitoring nonlinear profiles in manufacturing processes," *Journal of Quality Technology*, vol. 38, p. 199, 2006.
- [13] W. A. Jensen and J. B. Birch, "Profile monitoring via nonlinear mixed models," *Journal of Quality Technology*, vol. 41, p. 18, 2009.
- [14] C. Zou, F. Tsung, and Z. Wang, "Monitoring profiles based on nonparametric regression methods," *Technometrics*, vol. 50, pp. 512-526, 2008.
- [15] P. Qiu, C. Zou, and Z. Wang, "Nonparametric profile monitoring by mixed effects modeling," *Technometrics*, 2012.
- [16] K. Paynabar and J. Jin, "Characterization of non-linear profiles variations using mixed-effect models and wavelets," *IIE Transactions*, vol. 43, pp. 275-290, 2011.
- [17] R. Noorossana, A. Saghaei, and A. Amiri, *Statistical analysis of profile monitoring* vol. 865: John Wiley & Sons, 2011.

- [18] J. Jin and J. Shi, "Feature-preserving data compression of stamping tonnage information using wavelets," *Technometrics*, vol. 41, pp. 327-339, 1999.
- [19] J. Jin and J. Shi, "Diagnostic feature extraction from stamping tonnage signals based on design of experiments," *Journal of manufacturing science and engineering*, vol. 122, pp. 360-369, 2000.
- [20] S. Zhou, B. Sun, and J. Shi, "An SPC monitoring system for cycle-based waveform signals using Haar transform," *IEEE Transactions on Automation Science and Engineering*, vol. 3, pp. 60-72, 2006.
- [21] S. Zhou and J. Jin, "Automatic feature selection for unsupervised clustering of cycle-based signals in manufacturing processes," *IIE Transactions*, vol. 37, pp. 569-584, 2005.
- [22] S. Zhou, N. Jin, and J. Jin, "Cycle-based signal monitoring using a directionally variant multivariate control chart system," *IIE transactions*, vol. 37, pp. 971-982, 2005.
- [23] H. Wang, H. Kababji, and Q. Huang, "Monitoring global and local variations in multichannel functional data for manufacturing processes," *Journal of Manufacturing Systems*, vol. 28, pp. 11-16, 2009.
- [24] Q. Yang and J. Jin, "Separation of individual operation signals from mixed sensor measurements," *IIE Transactions*, vol. 44, pp. 780-792, 2012.
- [25] J. Kim, Q. Huang, J. Shi, and T.-S. Chang, "Online multichannel forging tonnage monitoring and fault pattern discrimination using principal curve," *Journal of manufacturing science and engineering*, vol. 128, pp. 944-950, 2006.

- [26] K. Paynabar, J. Jin, and M. Pacella, "Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis," *IIE Transactions*, vol. 45, pp. 1235-1247, 2013.
- [27] K. Paynabar, C. Zou, and P. Qiu, "A Change-Point Approach for Phase-I Analysis in Multivariate Profile Monitoring and Diagnosis," *Technometrics*, vol. 58, pp. 191-204, 2016.
- [28] C. Zhou, K. Liu, X. Zhang, W. Zhang, and J. Shi, "An Automatic Process Monitoring Method Using Recurrence Plot in Progressive Stamping Processes," *IEEE Transactions on Automation Science and Engineering*, vol. 13, pp. 1102-1111, 2016.
- [29] D. A. Axinte and N. Gindy, "Tool condition monitoring in broaching," *Wear*, vol. 254, pp. 370-382, 2003.
- [30] D. Shi, D. Axinte, and N. Gindy, "Development of an online machining process monitoring system: a case study of the broaching process," *The International Journal of Advanced Manufacturing Technology*, vol. 34, pp. 34-46, 2007.
- [31] J. Robertson, A. C. Rathinam, L. Wells, and J. Camelio, "Statistical monitoring for broaching processes using energy features extracted from cutting force signatures," in *41st North American Manufacturing Research Conference 2013, NAMRC 2013*, 2013.
- [32] J. Ho, S. Chou, K. Chua, A. Mujumdar, and M. Hawlader, "Analytical study of cyclic temperature drying: effect on drying kinetics and product quality," *Journal of Food Engineering*, vol. 51, pp. 65-75, 2002.

- [33] H. Y. Sohn and B.-S. Kim, "A novel cyclic reaction system involving CaS and CaSO₄ for converting sulfur dioxide to elemental sulfur without generating secondary pollutants. 1. Determination of process feasibility," *Industrial & engineering chemistry research*, vol. 41, pp. 3081-3086, 2002.
- [34] J. Sutherland and R. Devor, "An improved method for cutting force and surface error prediction in flexible end milling systems," *Journal of engineering for industry*, vol. 108, pp. 269-279, 1986.
- [35] G. Zhang, "Cause-selecting control charts—a new type of quality control charts," *The QR Journal*, vol. 12, pp. 221-225, 1985.
- [36] G. Zhang, "Cause-Selecting Control Chart and Diagnosis, Theory and Practice, Aarhus School of Business," *Department of Total Quality Management, Aarhus*, 1992.
- [37] D. M. Hawkins, "Multivariate quality control based on regression-adjusted variables," *Technometrics*, vol. 33, pp. 61-75, 1991.
- [38] D. M. Hawkins, "Regression adjustment for variables in multivariate quality control," *Journal of Quality Technology*, vol. 25, pp. 170-170, 1993.
- [39] J. Jin and J. Shi, "State space modeling of sheet metal assembly for dimensional control," *Journal of Manufacturing Science and Engineering*, vol. 121, pp. 756-762, 1999.
- [40] P. F. Zantek, G. P. Wright†, and R. D. Plante, "A self-starting procedure for monitoring process quality in multistage manufacturing systems," *IIE Transactions*, vol. 38, pp. 293-308, 2006.

- [41] F. Tsung, Y. Li, and M. Jin, "Statistical process control for multistage manufacturing and service operations: a review and some extensions," *International Journal of Services Operations and Informatics*, vol. 3, pp. 191-204, 2008.
- [42] L. Xiang and F. Tsung, "Statistical monitoring of multi-stage processes based on engineering models," *IIE transactions*, vol. 40, pp. 957-970, 2008.
- [43] L. Nelson, "Control chart for multiple stream processes," *Journal of Quality Technology*, vol. 18, pp. 255-256, 1986.
- [44] C. Zou and F. Tsung, "Directional MEWMA schemes for multistage process monitoring and diagnosis," *Journal of Quality Technology*, vol. 40, p. 407, 2008.
- [45] J. Zhang, H. Ren, R. Yao, C. Zou, and Z. Wang, "Phase I analysis of multivariate profiles based on regression adjustment," *Computers & Industrial Engineering*, vol. 85, pp. 132-144, 2015.
- [46] S. Kurada and C. Bradley, "A review of machine vision sensors for tool condition monitoring," *Computers in Industry*, vol. 34, pp. 55-72, 1997.
- [47] S. Dutta, S. Pal, S. Mukhopadhyay, and R. Sen, "Application of digital image processing in tool condition monitoring: A review," *CIRP Journal of Manufacturing Science and Technology*, vol. 6, pp. 212-232, 2013.
- [48] B. Sick, "On-line and indirect tool wear monitoring in turning with artificial neural networks: a review of more than a decade of research," *Mechanical Systems and Signal Processing*, vol. 16, pp. 487-546, 2002.

- [49] A. G. Rehorn, J. Jiang, and P. E. Orban, "State-of-the-art methods and results in tool condition monitoring: a review," *The International Journal of Advanced Manufacturing Technology*, vol. 26, pp. 693-710, 2005.
- [50] A. Siddhpura and R. Paurobally, "A review of flank wear prediction methods for tool condition monitoring in a turning process," *The International Journal of Advanced Manufacturing Technology*, vol. 65, pp. 371-393, 2013.
- [51] S. Mo, D. Axinte, T. Hyde, and N. Gindy, "An example of selection of the cutting conditions in broaching of heat-resistant alloys based on cutting forces, surface roughness and tool wear," *Journal of materials processing technology*, vol. 160, pp. 382-389, 2005.
- [52] D. Shi and N. N. Gindy, "Tool wear predictive model based on least squares support vector machines," *Mechanical Systems and Signal Processing*, vol. 21, pp. 1799-1814, 2007.
- [53] L. Filice, F. Micari, L. Settineri, and D. Umbrello, "Wear modelling in mild steel orthogonal cutting when using uncoated carbide tools," *Wear*, vol. 262, pp. 545-554, 2007.
- [54] T. G. Dawson and T. R. Kurfess, "Modeling the progression of flank wear on uncoated and ceramic-coated polycrystalline cubic boron nitride tools in hard turning," *Journal of manufacturing science and engineering*, vol. 128, pp. 104-109, 2006.
- [55] H. Zhao, G. Barber, and Q. Zou, "A study of flank wear in orthogonal cutting with internal cooling," *Wear*, vol. 253, pp. 957-962, 2002.

- [56] E. Usui, T. Shirakashi, and T. Kitagawa, "Analytical prediction of cutting tool wear," *Wear*, vol. 100, pp. 129-151, 1984.
- [57] X. Luo, K. Cheng, R. Holt, and X. Liu, "Modeling flank wear of carbide tool insert in metal cutting," *Wear*, vol. 259, pp. 1235-1240, 2005.
- [58] A. Attanasio, E. Ceretti, S. Rizzuti, D. Umbrello, and F. Micari, "3D finite element analysis of tool wear in machining," *CIRP Annals-Manufacturing Technology*, vol. 57, pp. 61-64, 2008.
- [59] Z. Pálmai, "Proposal for a new theoretical model of the cutting tool's flank wear," *Wear*, vol. 303, pp. 437-445, 2013.
- [60] P. Baruah and R. B. Chinnam*, "HMMs for diagnostics and prognostics in machining processes," *International Journal of Production Research*, vol. 43, pp. 1275-1293, 2005.
- [61] M. Wang and J. Wang, "CHMM for tool condition monitoring and remaining useful life prediction," *The International Journal of Advanced Manufacturing Technology*, vol. 59, pp. 463-471, 2012.
- [62] O. Geramifard, J.-X. Xu, J.-H. Zhou, and X. Li, "A physically segmented hidden Markov model approach for continuous tool condition monitoring: Diagnostics and prognostics," *IEEE Transactions on Industrial Informatics*, vol. 8, pp. 964-973, 2012.
- [63] H. Qiu, H. Liao, and J. Lee, "Degradation assessment for machinery prognostics using hidden Markov models," in *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2005, pp. 531-537.

- [64] J. Karandikar, T. McLeay, S. Turner, and T. Schmitz, "Remaining useful tool life predictions using Bayesian inference," in *ASME 2013 International Manufacturing Science and Engineering Conference collocated with the 41st North American Manufacturing Research Conference*, 2013, pp. V002T02A027-V002T02A027.
- [65] J. M. Karandikar, A. E. Abbas, and T. L. Schmitz, "Tool life prediction using Bayesian updating. Part 1: Milling tool life model using a discrete grid method," *Precision Engineering*, vol. 38, pp. 9-17, 1// 2014.
- [66] J. M. Karandikar, A. E. Abbas, and T. L. Schmitz, "Tool life prediction using Bayesian updating. Part 2: Turning tool life using a Markov Chain Monte Carlo approach," *Precision Engineering*, vol. 38, pp. 18-27, 2014.
- [67] D. Tobon-Mejia, K. Medjaher, and N. Zerhouni, "CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks," *Mechanical Systems and Signal Processing*, vol. 28, pp. 167-182, 2012.
- [68] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechak, "Health assessment and life prediction of cutting tools based on support vector regression," *Journal of Intelligent Manufacturing*, vol. 26, pp. 213-223, 2015.
- [69] C. Drouillet, J. Karandikar, C. Nath, A.-C. Journeaux, M. El Mansori, and T. Kurfess, "Tool life predictions in milling using spindle power with the neural network technique," *Journal of Manufacturing Processes*, vol. 22, pp. 161-168, 2016.
- [70] A. M. Khorasani, M. R. S. Yazdi, and M. S. Safizadeh, "Tool Life Prediction in Face Milling Machiningof 7075 Al by Using Artificial Neural Networks (ANN)

- and Taguchi Design of Experiment (DOE)," *International Journal of Engineering and Technology*, vol. 3, p. 30, 2011.
- [71] U. Natarajan, V. Periasamy, and R. Saravanan, "Application of particle swarm optimisation in artificial neural network for the prediction of tool life," *The International Journal of Advanced Manufacturing Technology*, vol. 31, pp. 871-876, 2007.
- [72] M. Alauddin, M. A. El Baradie, and M. S. J. Hashmi, "Prediction of tool life in end milling by response surface methodology," *Journal of Materials Processing Technology*, vol. 71, pp. 456-465, 1997/11/23 1997.
- [73] M. Alauddin and M. A. El Baradie, "Tool life model for end milling steel (190 BHN)," *Journal of Materials Processing Technology*, vol. 68, pp. 50-59, 1997/06/15 1997.
- [74] H. Khorasgani, G. Biswas, and S. Sankararaman, "Methodologies for system-level remaining useful life prediction," *Reliability Engineering & System Safety*, vol. 154, pp. 8-18, 2016.
- [75] J. Sikorska, M. Hodkiewicz, and L. Ma, "Prognostic modelling options for remaining useful life estimation by industry," *Mechanical Systems and Signal Processing*, vol. 25, pp. 1803-1836, 2011.
- [76] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *European Journal of Operational Research*, vol. 213, pp. 1-14, 2011.

- [77] K. L. Tsui, N. Chen, Q. Zhou, Y. Hai, and W. Wang, "Prognostics and health management: A review on data driven approaches," *Mathematical Problems in Engineering*, vol. 2015, 2015.
- [78] G. Whitmore, "Estimating degradation by a Wiener diffusion process subject to measurement error," *Lifetime data analysis*, vol. 1, pp. 307-319, 1995.
- [79] X. Wang, "Wiener processes with random effects for degradation data," *Journal of Multivariate Analysis*, vol. 101, pp. 340-351, 2010.
- [80] Z.-S. Ye, Y. Wang, K.-L. Tsui, and M. Pecht, "Degradation data analysis using Wiener processes with measurement errors," *IEEE Transactions on Reliability*, vol. 62, pp. 772-780, 2013.
- [81] J. Van Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability Engineering & System Safety*, vol. 94, pp. 2-21, 2009.
- [82] J. Lawless and M. Crowder, "Covariates and random effects in a gamma process model with application to degradation and failure," *Lifetime Data Analysis*, vol. 10, pp. 213-227, 2004.
- [83] C.-C. Tsai, S.-T. Tseng, and N. Balakrishnan, "Optimal burn-in policy for highly reliable products using gamma degradation process," *IEEE Transactions on Reliability*, vol. 60, pp. 234-245, 2011.
- [84] A. Ghasemi, S. Yacout, and M.-S. Ouali, "Parameter estimation methods for condition-based maintenance with indirect observations," *IEEE Transactions on Reliability*, vol. 59, pp. 426-439, 2010.

- [85] D. A. Tobon-Mejia, K. Medjaher, N. Zerhouni, and G. Tripot, "A data-driven failure prognostics method based on mixture of Gaussians hidden Markov models," *IEEE Transactions on reliability*, vol. 61, pp. 491-503, 2012.
- [86] C. J. Lu and W. Q. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, pp. 161-174, 1993.
- [87] N. Gebraeel, "Sensory-updated residual life distributions for components with exponential degradation patterns," *IEEE Transactions on Automation Science and Engineering*, vol. 3, pp. 382-393, 2006.
- [88] N. Chen and K. L. Tsui, "Condition monitoring and remaining useful life prediction using degradation signals: Revisited," *IIE Transactions*, vol. 45, pp. 939-952, 2013.
- [89] J. Son, Y. Zhang, C. Sankavaram, and S. Zhou, "RUL prediction for individual units based on condition monitoring signals with a change point," *IEEE Transactions on Reliability*, vol. 64, pp. 182-196, 2015.
- [90] Z. Xu, Y. Hong, and R. Jin, "Nonlinear general path models for degradation data with dynamic covariates," *Applied Stochastic Models in Business and Industry*, 2015.
- [91] W. Q. Meeker, L. A. Escobar, and C. J. Lu, "Accelerated degradation tests: modeling and analysis," *Technometrics*, vol. 40, pp. 89-99, 1998.
- [92] F. Haghghi, N. Noorae, and N. N. Rad, "On the general degradation path model: Review and simulation," in *Advances in Degradation Modeling*, ed: Springer, 2010, pp. 147-155.

- [93] C. J. Li and H. Lee, "Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics," *Mechanical systems and signal processing*, vol. 19, pp. 836-846, 2005.
- [94] G. Kacprzyński, A. Sarlashkar, M. Roemer, A. Hess, and B. Hardman, "Predicting remaining life by fusing the physics of failure modeling with diagnostics," *JOM Journal of the Minerals, Metals and Materials Society*, vol. 56, pp. 29-35, 2004.
- [95] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE Transactions on Image Processing*, vol. 7, pp. 167-179, 1998.
- [96] R. K. Neerukatti, K. C. Liu, N. Kovvali, and A. Chattopadhyay, "Fatigue life prediction using hybrid prognosis for structural health monitoring," *Journal of Aerospace Information Systems*, vol. 11, pp. 211-232, 2014.
- [97] J. P. P. Gomes, L. R. Rodrigues, R. K. H. Galvão, and T. Yoneyama, "System level RUL estimation for multiple-component systems," in *Proceedings of the 2013 Annual conference of the prognostics and health management society*, 2013, pp. 74-82.
- [98] L. Bian and N. Gebraeel, "Computing and updating the first-passage time distribution for randomly evolving degradation signals," *IIE Transactions*, vol. 44, pp. 974-987, 2012.
- [99] K. Liu, N. Z. Gebraeel, and J. Shi, "A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis,"

- IEEE Transactions on Automation Science and Engineering*, vol. 10, pp. 652-664, 2013.
- [100] K. Liu, A. Chehade, and C. Song, "Optimize the signal quality of the composite health index via data fusion for degradation modeling and prognostic analysis," *IEEE Transactions on Automation Science and Engineering*, 2015.
- [101] K. Liu and S. Huang, "Integration of data fusion methodology and degradation modeling process to improve prognostics," *IEEE Transactions on Automation Science and Engineering*, vol. 13, pp. 344-354, 2016.
- [102] P. Wang and D. W. Coit, "Reliability and degradation modeling with random or uncertain failure threshold," in *Reliability and Maintainability Symposium, 2007. RAMS'07. Annual*, 2007, pp. 392-397.
- [103] H. Yan, K. Liu, X. Zhang, and J. Shi, "Multiple Sensor Data Fusion for Degradation Modeling and Prognostics Under Multiple Operational Conditions," *IEEE Transactions on Reliability*, vol. 65, pp. 1416-1426, 2016.
- [104] M. Lavielle, "Detection of multiple changes in a sequence of dependent variables," *Stochastic Processes and their Applications*, vol. 83, pp. 79-102, 1999.
- [105] J. S. Lim, *Two-dimensional signal and image processing* vol. 1. Englewood Cliffs, NJ: Prentice Hall, 1990.
- [106] C. M. Thompson and L. Shure, *Image processing toolbox: User's guide: The MathWorks*, 1995.
- [107] J. Jurkovic, M. Korosec, and J. Kopac, "New approach in tool wear measuring technique using CCD vision system," *International Journal of Machine Tools and Manufacture*, vol. 45, pp. 1023-1030, 2005.

- [108] S. Kurada and C. Bradley, "A machine vision system for tool wear assessment," *Tribology International*, vol. 30, pp. 295-304, 1997.
- [109] D. C. Montgomery, *Introduction to statistical quality control*, 6th ed.: John Wiley & Sons, 2009.
- [110] C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to wavelets and wavelet transforms: a primer," 1997.
- [111] Z. Wu, N. E. Huang, S. R. Long, and C.-K. Peng, "On the trend, detrending, and variability of nonlinear and nonstationary time series," *Proceedings of the National Academy of Sciences*, vol. 104, pp. 14889-14894, 2007.
- [112] D. A. Axinte, N. Gindy, K. Fox, and I. Unanue, "Process monitoring to assist the workpiece surface quality in machining," *International Journal of Machine Tools and Manufacture*, vol. 44, pp. 1091-1108, 2004.
- [113] F. Klocke, D. Veselovac, S. Gierlings, and L. E. Tamayo, "Development of process monitoring strategies in broaching of nickel-based alloys," *Mechanics & Industry*, vol. 13, pp. 3-9, 2012.
- [114] A. Höskuldsson, "PLS regression methods," *Journal of chemometrics*, vol. 2, pp. 211-228, 1988.
- [115] H. Wold, "Estimation of principal components and related models by iterative least squares. ," *Multivariate Analysis*, vol. 1, pp. 391-420, 1966.
- [116] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Analytica chimica acta*, vol. 185, pp. 1-17, 1986.

- [117] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and intelligent laboratory systems*, vol. 58, pp. 109-130, 2001.
- [118] R. Rosipal and N. Krämer, "Overview and recent advances in partial least squares," in *Subspace, latent structure and feature selection*, ed: Springer, 2006, pp. 34-51.
- [119] H. Chun and S. Keleş, "Sparse partial least squares regression for simultaneous dimension reduction and variable selection," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 72, pp. 3-25, 2010.
- [120] Z. G. Stoumbos and J. H. Sullivan, "Robustness to non-normality of the multivariate EWMA control chart," *Journal of Quality Technology*, vol. 34, p. 260, 2002.
- [121] C. A. Lowry, W. H. Woodall, C. W. Champ, and S. E. Rigdon, "A multivariate exponentially weighted moving average control chart," *Technometrics*, vol. 34, pp. 46-53, 1992.
- [122] J. H. Sullivan and W. H. Woodall, "A comparison of multivariate control charts for individual observations," *Journal of Quality Technology*, vol. 28, pp. 398-408, 1996.
- [123] Z. Wu, Y. C. Lam, S. Zhang, and M. Shamsuzzaman, "Optimization design of control chart systems," *IIE Transactions*, vol. 36, pp. 447-455, 2004.
- [124] P. J. Davis, *Interpolation and approximation*: Courier Corporation, 1975.

- [125] A. Trujillo-Ortiz, R. Hernandez-Walls, K. Barba-Rojo, and L. Cupul-Magana, "hZmvntest: Henze-Zirkler's Multivariate normality test," A *MATLAB file*. URL <http://www.mathworks.com/matlabcentral/fileexchange/loadFile.do>, 2007.
- [126] J. Loizou, W. Tian, J. Robertson, and J. Camelio, "Automated Wear Characterization for Broaching Tools Based on Machine Vision Systems," *Journal of Manufacturing Systems*, vol. 37, pp. 558-563, 2015.
- [127] M. Sortino, "Application of statistical filtering for optical detection of tool wear," *International Journal of Machine Tools and Manufacture*, vol. 43, pp. 493-497, 2003.
- [128] J. Barreiro, M. Castejón, E. Alegre, and L. Hernández, "Use of descriptors based on moments from digital images for tool wear monitoring," *International Journal of Machine Tools and Manufacture*, vol. 48, pp. 1005-1013, 2008.
- [129] M. Castejon, E. Alegre, J. Barreiro, and L. Hernandez, "On-line tool wear monitoring using geometric descriptors from digital images," *International Journal of Machine Tools and Manufacture*, vol. 47, pp. 1847-1853, 2007.
- [130] N. Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, pp. 23-27, 1975.
- [131] G. E. Box, G. M. Jenkins, and G. C. Reinsel, *Time series analysis: forecasting and control* vol. 734: John Wiley & Sons, 2011.
- [132] P. M. Lee, *Bayesian statistics: an introduction*: John Wiley & Sons, 2012.
- [133] R. C. Gonzalez and R. Woods, "Image restoration and reconstruction," *Digital image processing*, Prentice Hall, Upper Saddle River, NJ, 2007, pp. 352-357, 2008.

- [134] F. W. Taylor, "On the art of cutting metals," 1907.
- [135] D. A. Stephenson and J. S. Agapiou, *Metal cutting theory and practice*: CRC press, 2016.
- [136] S. Ramalinganrs, "Tool Life Distributions," *Journal of Engineering for Industry*, vol. 100, p. 193, 1978.
- [137] F. Camci and R. B. Chinnam, "Health-state estimation and prognostics in machining processes," *IEEE Transactions on automation science and engineering*, vol. 7, pp. 581-597, 2010.
- [138] S. E. Chick and M. B. Mendel, "An engineering basis for statistical lifetime models with an application to tribology," *IEEE Transactions on Reliability*, vol. 45, pp. 208-215, 1996.
- [139] H. T. Pham, B.-S. Yang, and T. T. Nguyen, "Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine," *Mechanical Systems and Signal Processing*, vol. 32, pp. 320-330, 2012.
- [140] H.-Z. Huang, H.-K. Wang, Y.-F. Li, L. Zhang, and Z. Liu, "Support vector machine based estimation of remaining useful life: current research status and future trends," *Journal of Mechanical Science and Technology*, vol. 29, pp. 151-163, 2015.
- [141] C. Sun, Z. He, H. Cao, Z. Zhang, X. Chen, and M. J. Zuo, "A non-probabilistic metric derived from condition information for operational reliability assessment of aero-engines," *IEEE Transactions on Reliability*, vol. 64, pp. 167-181, 2015.

- [142] K. Zhu, Y. San Wong, and G. S. Hong, "Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results," *International Journal of Machine Tools and Manufacture*, vol. 49, pp. 537-553, 2009.
- [143] C. J. Lu and W. O. Meeker, "Using degradation measures to estimate a time-to-failure distribution," *Technometrics*, vol. 35, pp. 161-174, 1993.
- [144] R. Zhou, N. Gebraeel, and N. Serban, "Degradation modeling and monitoring of truncated degradation signals," *IIE Transactions*, vol. 44, pp. 793-803, 2012.
- [145] W. Tian, R. Jin, T. Huang, and J. A. Camelio, "Statistical Process Control for Multistage Processes with Non-repeating Cyclic Profiles," *IIE Transactions*, 2016.
- [146] K. Liu, A. Chehade, and C. Song, "Optimize the Signal Quality of the Composite Health Index Via Data Fusion for Degradation Modeling and Prognostic Analysis," *IEEE Transactions on Automation Science and Engineering*, vol. PP pp. 1 - 11, 2015.
- [147] F. Klocke, B. Döbbeler, and M. Seimann, "Dry Broaching Using Carbon Free Steel as Tool Material," *Procedia CIRP*, vol. 46, pp. 496-499, 2016.
- [148] J. Son, Q. Zhou, S. Zhou, X. Mao, and M. Salman, "Evaluation and comparison of mixed effects model based prognosis for hard failure," *IEEE Transactions on Reliability*, vol. 62, pp. 379-394, 2013.
- [149] Q. Zhou, J. Son, S. Zhou, X. Mao, and M. Salman, "Remaining useful life prediction of individual units subject to hard failure," *IIE Transactions*, vol. 46, pp. 1017-1030, 2014.
- [150] M. A. Kuttolamadom, M. L. Mears, and T. R. Kurfess, "On the volumetric assessment of tool wear in machining inserts with complex geometries—part 1:

- need, methodology, and standardization," *Journal of Manufacturing Science and Engineering*, vol. 134, p. 051002, 2012.
- [151] M. A. Kuttolamadom, M. L. Mears, T. R. Kurfess, U. Burger, and A. Bryan, "On the Volumetric Assessment of Tool Wear in Machining Inserts With Complex Geometries—Part II: Experimental Investigation and Validation on Ti-6Al-4V," *Journal of Manufacturing Science and Engineering*, vol. 134, p. 051003, 2012.
- [152] X. Liu, K. Yeo, and J. Kalagnanam, "Statistical Modeling for Spatio-Temporal Degradation Data," *arXiv preprint arXiv:1609.07217*, 2016.

Appendix A: Robustness of the proposed control charting system.

ARL performance given different percentage values of total variance explained in the response matrices is summarized in Table A1.

Table A1. ARL performance of the GEWMA method given different percentages of variance explained in the response matrices p_m

	Shift Stage	δ	GEWMA						
			$p_m=0.85$	$p_m=0.9$	$p_m=0.95$	$p_m=0.96$	$p_m=0.99$		
In-control	-	0	370.27	370.40	370.62	370.21	370.36		
		0.1	196.98	182.38	187.72	202.93	305.42		
		0.25	16.35	15.70	13.86	14.61	66.52		
		0.5	4.42	4.35	4.03	4.12	6.51		
		2	1	2.03	2.01	1.92	1.94	2.57	
			1.5	1.43	1.40	1.29	1.32	1.88	
			2	1.04	1.04	1.01	1.01	1.29	
			2.5	1.00	1.00	1.00	1.00	1.01	
			3	1.00	1.00	1.00	1.00	1.00	
		Local wear		0.1	305.67	298.11	241.53	263.22	368.88
0.25	47.07			44.63	19.03	20.61	201.59		
0.5	7.00			6.95	4.64	4.76	8.68		
7	1			2.72	2.70	2.09	2.13	2.98	
	1.5			1.88	1.88	1.50	1.54	2.00	
	2			1.43	1.43	1.05	1.05	1.66	
	2.5			1.07	1.07	1.00	1.00	1.09	
	3			1.00	1.00	1.00	1.00	1.00	
Local breakage				0.1	62.61	59.66	50.47	56.50	237.65
				0.25	6.23	6.18	5.53	5.71	10.89
		0.5	2.55	2.54	2.37	2.41	3.34		
		2	1	1.32	1.32	1.19	1.23	1.81	
			1.5	1.00	1.00	1.00	1.00	1.03	
			2	1.00	1.00	1.00	1.00	1.00	
			2.5	1.00	1.00	1.00	1.00	1.00	
			3	1.00	1.00	1.00	1.00	1.00	
		7	0.1	294.53	255.91	106.15	122.65	371.79	
			0.25	41.39	26.18	8.30	8.77	27.39	
0.5	6.62		5.52	3.03	3.12	4.67			
1	2.63		2.35	1.61	1.66	2.07			

		1.5	1.85	1.70	1.04	1.04	1.49
		2	1.39	1.18	1.00	1.00	1.01
		2.5	1.04	1.01	1.00	1.00	1.00
		3	1.00	1.00	1.00	1.00	1.00
		0.1	330.49	325.27	328.61	324.38	333.79
		0.25	166.95	162.50	167.44	164.12	168.73
		0.5	48.29	47.74	54.75	48.35	49.46
		1	11.22	11.24	11.33	11.11	11.27
Global	All	1.5	5.67	5.66	5.67	5.66	5.70
		2	3.85	3.81	3.85	3.80	3.85
		2.5	2.93	2.93	2.96	2.93	2.95
		3	2.41	2.41	2.42	2.39	2.43

Appendix B: Derivation on the posterior probability to determine if reigion k is noise

The idea of calculating the posterior probability to determine whether $R_{e,t}^k$ is noise or not is similar to a Naïve Bayes classifier [29]. Given $R_{e,t}^k$, there are two groups of events of interest: Group 1 contains events $A_1 = \{R_{e,t}^k \in \text{Noise}\}$ and $A_2 = \{R_{e,t}^k \in \text{Information}\}$, where A_1 and A_2 is a partition of the full set, i.e. A_1 and A_2 are disjoint and $P(A_1) + P(A_2) = 1$; Group 2 contains events $B_1 = \{R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset\}$ and $B_2 = \{R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset\}$, where A_1 and A_2 is a partition of the full set. From process knowledge, we already know the following probabilities,

$$P(R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset \mid R_{e,t}^k \in \text{Information}) = P(B_1|A_2) = 1 - \theta$$

$$P(R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset \mid R_{e,t}^k \in \text{Information}) = P(B_2|A_2) = \theta$$

$$P(R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset \mid R_{e,t}^k \in \text{Noise}) = P(B_1|A_1) = \theta'$$

$$P(R_{e,t}^k \cap R_{e,t+1}^\bullet = \emptyset \mid R_{e,t}^k \in \text{Noise}) = P(B_2|A_1) = 1 - \theta'$$

where both θ and θ' have very small values and $0 < \theta, \theta' < 1$. As it is not known that if $R_{e,t}^k$ is noise or information, a prior distribution for the two events needs to be specified. Based on the user's preference for the image filtering, the prior distribution could be selected. For example, the more conservative the user is to eliminate possible noise, the smaller prior probability should be given to event A_1 . Provided a neutral perspective of the user, an uninformative prior can be used with $P(A_1) = P(A_2) = 0.5$. Based on Bayes' theorem, the posterior probabilities can be derived as follows,

$$P(R_{e,t}^k \in \text{Information} \mid R_{e,t}^k \cap R_{e,t+1}^\bullet \neq \emptyset) = P(A_2|B_1)$$

$$= \frac{P(B_1|A_2) \times P(A_2)}{P(B_1|A_2) \times P(A_2) + P(B_1|A_1) \times P(A_1)} = \frac{(1 - \theta) \times 0.5}{(1 - \theta) \times 0.5 + \theta' \times 0.5} = \frac{1}{1 + \frac{\theta'}{1 - \theta}}$$

is close to 1 given both θ and θ' have very small values.

The other posterior probabilities can be derived in a similar manner, and the results are summarized in Table 6 in Section 5.

Appendix C: Simulation results for various different scenarios in data correlation

Multiple different scenarios in the data correlation structure are used to test the effectiveness of the proposed method. All the tested scenarios are summarized in Table 8 and the model specifications for data generation include:

- 1) An autoregressive model of order 1 (i.e. AR(1) model) is used to describe the autocorrelation in System HI, which means in (22), the error term $\varepsilon_k = 0.7\varepsilon_{k-1} + \delta_k$, where $\delta_k \sim N(0, \sigma^2)$;
- 2) An exponential kernel is used to describe the spatial covariance function $k(\omega_i, \omega_j) = S_k^2 \exp\left(-\frac{r}{l}\right)$, where l is called the characteristic length scale and is set to 1, and r represents the distance between the two cutting edges of ω_i and ω_j ;
- 3) The increasing variance across t_k in (20) is described with a linear function $S_k^2 = \alpha \cdot t_k$.

Table 8: Simulation scenarios

Scenario #	Autocorrelation in System HI	Spatial Correlation in $W_a^2(t_k) - W_a^2(0)$
1*	N	N
2	N	Y
3	Y	N
4	Y	Y

* Scenario 1 is the scenario shown in Section 6.5.

The simulation results for the eight scenarios are summarized in Figure 32 to Figure 34. It can be observed that our proposed method outperforms its counterparts in both absolute RUL prediction error and mean width of the 90% CIs in most of the scenarios we tested. In addition, our proposed method are always better when we have increasing variance across cutting time t_k .

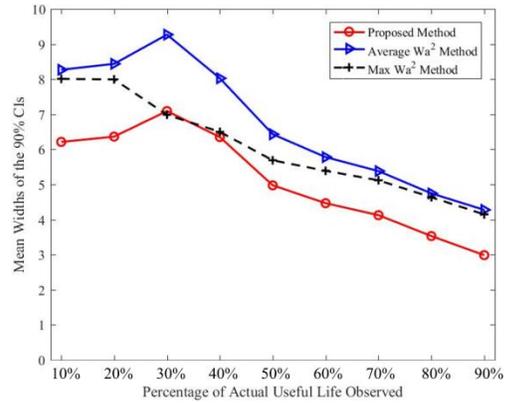
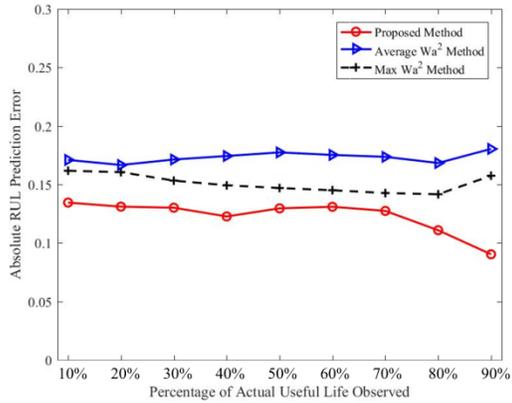


Figure 32: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 2

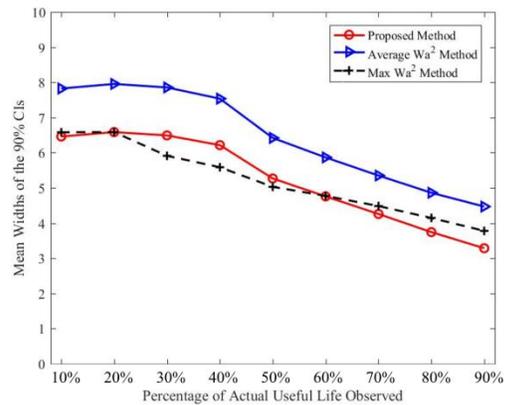
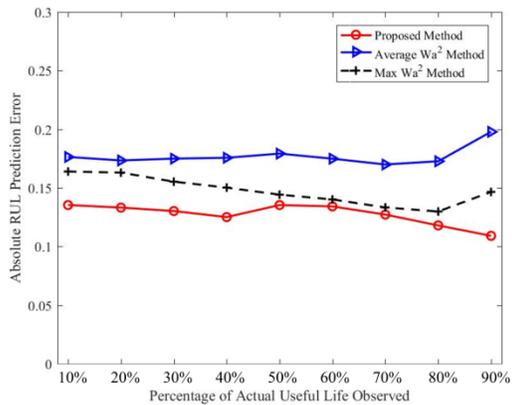


Figure 33: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 3

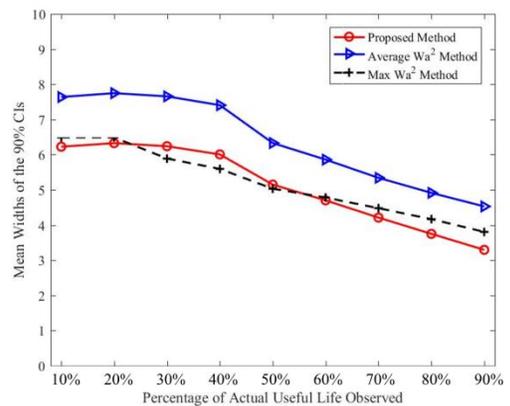
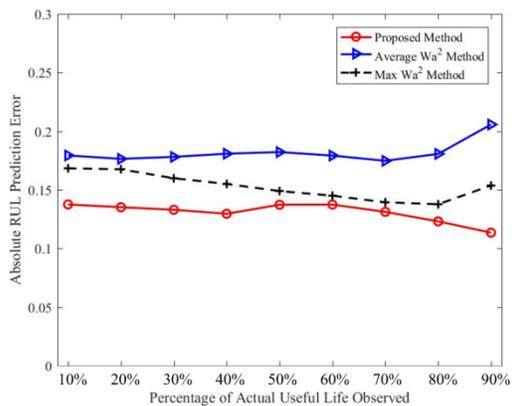


Figure 34: Comparison results of the mean absolute percentage error (left) and average widths of the 90% CIs (right) based on simulated data under Scenario 4