DATA-BASED SPATIAL AND TEMPORAL MODELING FOR SURFACE VARIATION MONITORING IN MANUFACTURING

by

Chenhui Shao

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in The University of Michigan
2016

Doctoral Committee:
  Professor S. Jack Hu, Co-Chair
  Professor Jionghua (Judy) Jin, Co-Chair
  Jeffrey A. Abell, General Motors
  Professor Elijah Kannatey-Asibu Jr.
  Professor Ji Zhu

To my family

Acknowledgments

First and foremost, I would like to express my deepest gratitude and respects to my co-advisors, Professor S. Jack Hu and Professor Judy Jin. This work would not have been possible without their guidance or support. I am very fortunate to have such great advisors, who gave me freedom to explore the sea of knowledge following my own curiosity and meanwhile always provided invaluable and timely pilot. To me, Professor Hu and Professor Jin are more than research advisors—they have given me support, encouragement, and precious career advice, and I have grown both personally and professionally through learning from my role models.

My sincere gratitude also goes to my other committee members, Dr. Abell, Professor Kannatey-Asibu Jr., and Professor Zhu, for their constructive suggestions, insightful vision, and valuable feedbacks. Their knowledge from other points of view have greatly inspired me and broadened my research vision.

I am very grateful to the teachers and professors of the courses that I have taken at the University of Michigan. They have not only taught me knowledge in various fields, but also inspired me and made me determined to become one of them myself.

I would like to thank my colleagues and friends in Professor Hu's lab and Professor Jin's group, for their help and friendship.

None of these would have been possible without the unconditional love, understanding, and support of my family. I want to thank my parents for their constant love and always being supportive no matter what decisions I made. Finally, but most importantly, I would like to acknowledge and thank my beloved wife, Mengjie, for her love, support, and the great happiness she has brought to my life.

Table of Contents

List of Tables

Table

List of Figures

Figure

List of Appendices

Appendix

Abstract

Spatiotemporal processes exist widely in manufacturing, such as tool surface degradation in ultrasonic metal welding and surface shape progression in high-precision machining. High-resolution characterization and monitoring of spatiotemporal processes are crucial for manufacturing process control. The rapid development of 3D sensing technologies has made it possible to generate large volumes of spatiotemporal data for process characterization and monitoring. However, critical challenges exist in effectively acquiring and utilizing such spatiotemporal data in manufacturing, e.g., a high cost in the acquisition of high-resolution spatiotemporal data and a lack of systematic approaches for modeling multi-source data and monitoring spatiotemporal processes.

To address these challenges, this dissertation carries out three research tasks for the development of collecting, modeling and monitoring spatiotemporal data. Specifically,

(1) A novel dynamic sampling design algorithm is developed to efficiently characterize spatiotemporal processes in manufacturing. A state-space model and Kalman filter are used to predictively determine the measurement locations using a criterion considering both the prediction variance and the measurement costs. The determination of measurement locations is formulated as a binary integer programming problem, and genetic algorithm is applied for searching the optimal design. In addition, a new test statistic is proposed to monitor and update the temporal transition

parameters in the spatiotemporal model.

(2) A new surface modeling approach is devised to cost-effectively assess spatial surface variations by integrating an engineering model with multi-task Gaussian process (GP) learning. Surface variation is decomposed into a global trend which is induced by process variables and a zero-mean GP which shares commonality among multiple similar-but-not-identical processes. An iterative algorithm is developed to simultaneously estimate the process-specific parameters and the GP parameters.

(3) A tool condition characterization and monitoring framework is developed for ultrasonic metal welding in lithium-ion battery manufacturing. The geometric progression of the tool surfaces is characterized using high-resolution spatiotemporal data. Classification algorithms are developed with monitoring features extracted from both the space and frequency domains. A novel impression measurement method is designed to effectively measure tool surfaces without the need of disassembling tools for off-line measurement.

Chapter 1
Introduction

1.1 Motivation

Spatiotemporal processes exist widely in a variety of science and engineering areas, such as environmental science (Anderson et al., 2006; Stein et al., 1998; Kyriakidis and Journel, 1999; Kibria et al., 2002), climate prediction and meteorology (Handcock and Wallis, 1994), epidemiology (Waller et al., 1997), agriculture (Mitchell and Gumpertz, 2003), and manufacturing (Shao et al., 2014). In manufacturing processes and systems, it is crucial to characterize and monitor spatiotemporal variations to guarantee satisfactory product quality or tooling conditions. For instance, in ultrasonic welding of lithium-ion batteries for electric vehicles, tool surfaces, including horn and anvil, change progressively as the tool degrades over time. The surface degradation affects the weld quality greatly, and should be closely monitored to ensure satisfactory weld quality (Shao et al., 2014). Another example involves the machining process of automotive engine components. The control of surface shape variation on a machined engine component is of great importance in automotive powertrain manufacturing, because the variation critically influences the downstream sealing and assembly performance, further affecting the overall quality of an engine. The surface variation pattern belongs to spatiotemporal processes, because it is significantly affected by a number of time-variant process variables (Suriano et al., 2015), e.g., cutter degradation, modified product design, and varying process settings.

In the recent years, the rapid development of sensing technologies, including 3D measurement techniques, has created a big data era for manufacturing. Many modern manufacturing plants have become a data-rich environment. For example, in an automotive battery manufacturing plant, production data are continuously collected from various sources, including process parameters, machine identity, tool types, product identity, on-line monitoring functional data, and high-resolution tool surface measurements. Figure 1.1 illustrates the data sources in the ultrasonic metal welding process in lithium-ion battery manufacturing. Product history is recorded for every battery pack, such as the machine identity and quality characteristics of all upstream operations. Within a welding cycle, data are collected from multiple sources. First, process parameters are predetermined, including welding pressure, vibration amplitude, and welding energy. Second, four sensors are used for on-line quality monitoring, including a power sensor, a frequency sensor, a linear variable differential transformer, and a microphone. These sensors generate functional curves in every welding cycle. Third, in order to monitor the tool degradation level, the surfaces of the horn and the anvil are measured every a certain number of welds, and successive measurement of tool surfaces forms spatiotemporal data.

Another example is the machining process for automotive engine components. For the purpose of quality monitoring and control, an engine plant records quality-related data from multiple sources, including (1) product identity, such as product type, station identity, and product history information, (2) machining process variables, e.g., feed rate, depth of cut, cutter type, remaining cutter life, monitoring signals from acoustic emission sensor and force sensor, and (3) surface quality data, which are acquired using one or more 3D

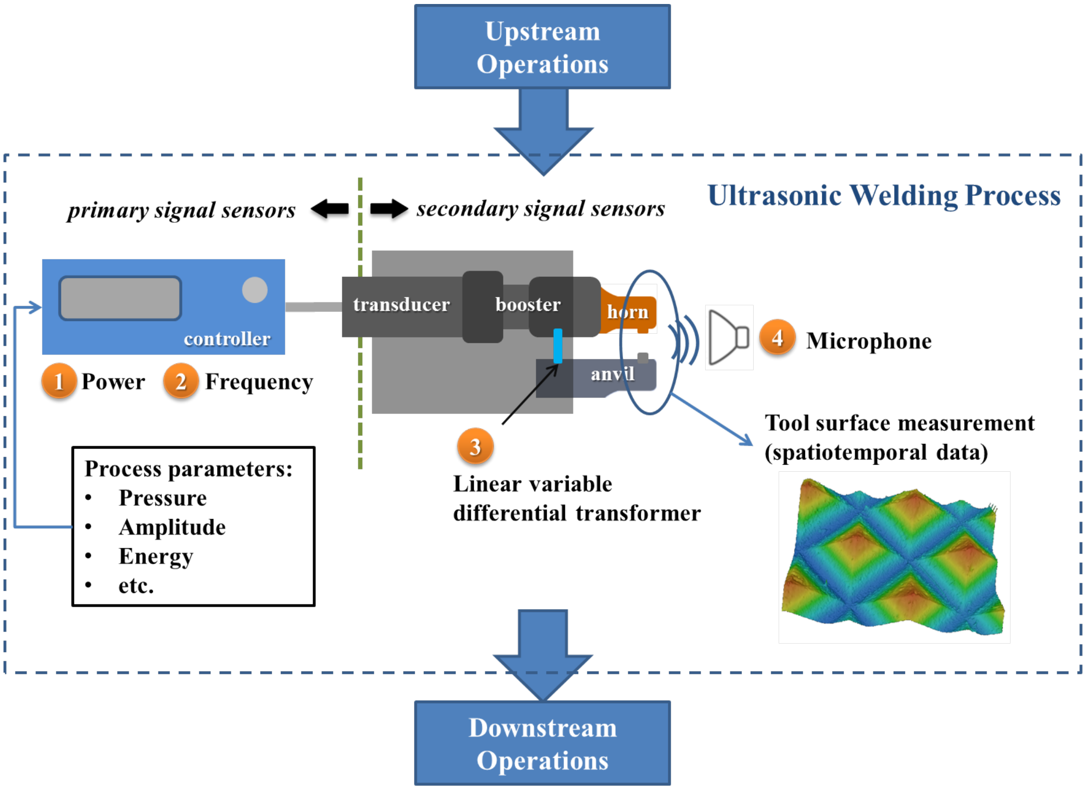


Figure 1.1 Multi-source data in the ultrasonic metal welding of lithium-ion batteries

measurement systems. Examples of 3D measurement systems used in machining processes include coordinate measuring machine, profilometer, and laser holographic interferometer (Suriano et al., 2015).

The availability of big data provides us with new opportunities to understand and monitor manufacturing processes and systems from a much more fine scale, both temporally and spatially. Nevertheless, it is highly challenging to effectively use such data.

First, high-resolution 3D sensing systems are often expensive and time-consuming. For example, a 3D microscope is used to measure tool surfaces in ultrasonic welding of lithium-ion batteries (Shao et al., 2014, 2016). It takes approximately 8 hours to measure a 43 mm × 8 mm surface. Considering that one weld cycle takes less than 1s, the significant downtime caused by the measurement will greatly reduce the production rate (Shao et al., 2013, 2016). Meanwhile, the long measurement time also introduces a significant delay to the tool replacement decision, making real time process adjustment not feasible. The insensitive adjustment actions will further hinder the welding quality, which is very undesirable because electric vehicle manufacturing has critical requirement on the joining quality (Guo et al.). Consequently, the high costs induced by high-resolution 3D measurement systems have limited the potential of using high-resolution spatiotemporal data for process monitoring and control.

Second, there is a lack of modeling approaches for integrating spatiotemporal data with other in-plant process information, such as product design, product history, process settings, on-line monitoring sensor signals, and high-resolution spatiotemporal data. As illustrated by Figure 1.1, modern manufacturing have become a data-rich environment, and data from a variety of sources are readily accessible for on-line process monitoring and control. Nevertheless, the data structure is highly complicated: there are a number of categorical and continuous variables, functional curves, and high-dimensional spatiotemporal data. Existing modeling approaches have seldom considered the cross-correlations among multiple data sources, especially for spatiotemporal data. Therefore, although the big data shows great potentials in improving current process control techniques, nontrivial efforts are necessary for developing new modeling algorithms.

Third, effective and efficient monitoring algorithms are lacking for spatiotemporal processes in manufacturing. Modeling of spatiotemporal processes has received some attention recently (Harvill, 2010), but existing studies have been conducted mostly in the fields other than manufacturing, such as environmental science, climate prediction and meteorology, epidemiology, and agriculture. Modeling and monitoring of such processes in manufacturing are unique, mainly because spatiotemporal processes in manufacturing applications are often high-dimensional and expert process knowledge is generally available to assist the extraction and selection of monitoring features. Hence, new monitoring techniques are highly desirable to characterize and monitor spatiotemporal processes in manufacturing.

1.2 Research Objectives

The objective of this dissertation is to develop methods for efficient data collection, cost-effective spatial variation modeling, and effective monitoring of spatiotemporal processes, with applications to manufacturing processes. Specific research tasks are described as follows.

(1) *Development of a dynamic sampling design approach for spatiotemporal processes*: High-resolution characterization of spatiotemporal processes is crucial for process monitoring and control. One example is the ultrasonic metal welding process. However, the acquisition of high-resolution measurement is often costly and time-consuming. Therefore, a dynamic sampling design algorithm is desirable to adaptively distribute measurement locations, and then spatial interpolation can be conducted to predict the surface height at unmeasured locations. A state-space spatiotemporal model is proposed for the process characterization, and Kalman filter can be applied to derive the prediction variance in a recursive manner. Due to the high-dimensionality of the design space, the genetic algorithm is suggested for solution searching. Moreover, since tool degradation rate is generally time-variant

over one tool life, a monitoring algorithm is necessary to estimate the degradation rate, which may be reflected by model parameter changes.

(2) *Development of a surface shape model to incorporate low-resolution measurement data and process physics*: As discussed in Section 1.1, a new cost-effective surface modeling approach is needed for the monitoring and control of surface manufacturing. Data fusion provides a means to make use of data from multiple sources. In machining processes, the relationship between process variables and the surface quality has been well studied in the literature. Additionally, products machined from the same machining station, or parallel running stations often share many similarities with each other. Hence, there is an opportunity to integrate these two types of information, i.e., process variables and similar-but-not-identical surfaces, to improve the modeling performance for surface variations. A potential method is to combine a process model with a machine learning technique, multi-task learning for Gaussian processes. A model parameter estimation algorithm will be needed to simultaneously estimate parameters in the two parts.

(3) *Design of a tool condition monitoring system for ultrasonic welding of lithium-ion batteries*: It has been reported that tool wear, especially tool surface degradation, is one key factor affecting the welding quality in ultrasonic metal welding for electric vehicle manufacturing. In addition, as a result of lacking accurate tool condition monitoring, the battery manufacturing plant replaces welding tools using a conservative strategy, and therefore some available tool lifetime is being wasted, significantly increasing production costs. Thus, a tool condition monitoring system for ultrasonic metal welding is very important for balancing product quality and

production costs. Using tool surface measurement data from a real-world battery plant, this study will first characterize tool wear progression both qualitatively and quantitatively. Then algorithms for monitoring feature generation and tool condition classification will be developed for spatiotemporal data with expert process knowledge.

1.3 Dissertation Organization

The remainder of this dissertation is organized as follows.

Chapter 2 presents a dynamic sampling design approach for characterizing spatiotemporal processes in manufacturing. A state-space model is utilized to describe a spatiotemporal process, where the residual is modeled as a spatiotemporal noise process. Then a measurement equation is adopted to select a parsimonious subset of the potential measurement locations. Based on Kalman filter, the covariance matrix of the estimation error is evaluated in an iterative fashion. In order to balance between the cost and the precision, a new design criterion is proposed as a combination of measurement costs and weighted average of the prediction uncertainty at all locations. The search process of optimal measurement subset, i.e., sampling design, is formulated as a binary integer programming problem. Genetic algorithm is used to solve for the optimal design. Additionally, a new test statistic is developed to monitor and update the temporal transition parameter. A comparative case study is conducted to show the effectiveness of the proposed method.

In Chapter 3, a novel surface modeling approach is developed to cost-effectively measure the parts for characterizing spatial surface variations by incorporating an engineering model with multi-task learning for Gaussian processes. In this model, part surface variation is decomposed into two levels, namely, a global trend which is mainly induced by process settings, such as material removal rate, and remaining cutter life, and a local variation modeled using a zero-mean Gaussian process. Among multiple similar-but-not-identical processes, e.g., machined parts from the same machining station or parallel stations, local variations share commonalities, and therefore can be estimated using multi-task learning. A new iterative multi-task Gaussian process learning algorithm is devised for the model parameter estimation. Furthermore, some practical suggestions are provided to expedite the iteration convergence. The approach is demonstrated using data collected from a real-world engine plant, and the implementation issues, including the effects of sample size, number of tasks, and the accuracy of the engineering model, are discussed with case studies.

Chapter 4 presents a framework for characterizing and monitoring tool surface degradation in ultrasonic welding of lithium-ion batteries. First, the surface progression of ultrasonic welding tools is characterized using optical images and high-resolution spatiotemporal data. Monitoring features are then extracted from the space and frequency domains of spatiotemporal measurements based on expert knowledge. Classification algorithms are developed based on the extracted monitoring features. In order to reduce production downtime, a novel impression measurement method is designed to effectively measure tool surface geometry. The effectiveness of the framework is demonstrated using the data from an electric vehicle battery manufacturing plant.

In Chapter 5, the contributions of this dissertation are summarized and future research directions are outlined.

Chapter 2

Dynamic Sampling Design for Characterizing Spatiotemporal Processes in Manufacturing

2.1 Introduction

Spatiotemporal processes evolve in both space and time. The characterization and monitoring of such processes are crucial in a variety of scientific and engineering fields, e.g., environmental science (Stein et al., 1998; Kyriakidis and Journel, 1999; Kibria et al., 2002), climate prediction and meteorology (Handcock and Wallis, 1994), epidemiology (Waller et al., 1997), agriculture (Mitchell and Gumpertz, 2003), and manufacturing process control (Shao et al., 2014).

Spatiotemporal processes exist widely in manufacturing. For example, ultrasonic metal welding is adopted for battery joining in electric vehicle manufacturing (Shao et al., 2013). The surface geometry of the welding tools progresses over time (Shao et al., 2014), as illustrated by Figure 2.1; in the machining process of automobile engine components, machined surfaces exhibit varying spatial variation patterns as cutters wear out. Figure 2.2 compares engine head surfaces that are cut using a new and a worn tool. It is indicated that the surface machined with a worn cutter presents significantly larger variation, especially in the edge areas.

Fine-scale characterization of spatiotemporal processes is essential for process monitoring and control in manufacturing. As an example, in ultrasonic metal welding of batteries

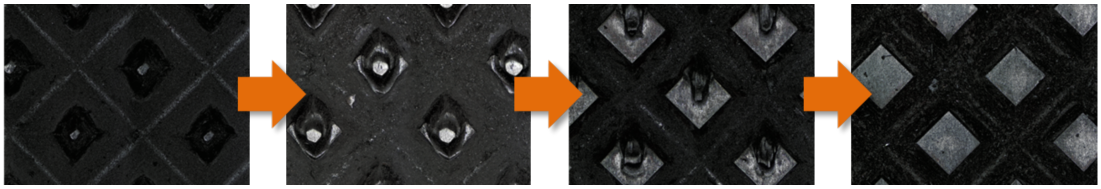
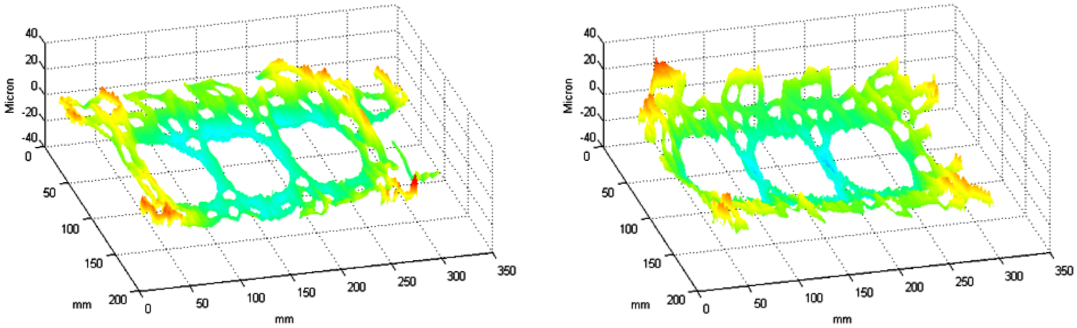


Figure 2.1 Tooling degradation process in ultrasonic metal welding



(a) New tool

(b) Worn tool

Figure 2.2 Engine head surfaces machined using (a) new tool and (b) worn tool

for electric vehicle manufacturing, tool surface degradation is an important root cause accounting for low-quality weld (Shao et al., 2014, 2016). In the example of high-precision machining processes for automotive engine components, surface quality is of great importance for assembly performance and the overall engine quality (Suriano et al., 2015; Nguyen et al., 2013). The spatial variations in both examples fall into the category of spatiotemporal processes. Accurate characterization of such processes is important for both understanding the underlying process mechanism and enabling high-performance monitoring and control of the manufacturing processes.

Critical challenges exist in efficient and effective characterization of spatiotemporal process in manufacturing. First, high-resolution measurement data are expensive and time-consuming to acquire. For instance, a 3D microscope is used by an electric vehicle manufacturer for tool wear monitoring of ultrasonic metal welding. It takes the microscope approximately 8 hours to measure a single tool surface with the dimension of 43 mm ×

8 mm. The low measurement efficiency not only leads to a large amount of production downtime, but also greatly delays the decision-making in tool condition monitoring. The high cost has limited the usage of high-resolution surface measurement systems for on-line process monitoring.

Spatial interpolation is one possible solution to reducing the cost induced by high-resolution measurement systems. In spatial statistics, interpolation refers to the procedure of estimating the value of properties, e.g., surface height, at unsampled/unmeasured locations using measurement data from their vicinities. Examples of spatial interpolation techniques include kriging (Stein, 2012), cokriging (Atkinson et al., 1992), and regression kriging (Odeh et al., 1995). The accuracy and precision of interpolation largely depend on the number of sampling/measurement sites and the distribution of measurement points. Therefore, sampling design is critically needed to achieve satisfactory interpolation results.

Sampling design has received much attention in the area of natural resources and environmental science (McBratney et al., 1981; McBratney and Webster, 1981, 1983a,b; Curran and Williamson, 1986; Curran, 1988; Xiao et al., 2005; Anderson et al., 2006; Zhu and Stein, 2006), and is recently investigated in the area of manufacturing (Jin et al., 2012).

A sampling design problem can be generally formulated as follows. Assume the location set of interest is denoted by $\mathscr{S}$, and the sampled/measured location set is $S_o = \{\mathbf{s}_1, \ldots, \mathbf{s}_n\} \subset \mathscr{S}$. The goal is to adaptively select S_o such that the inference at unobserved locations, denoted as $S_u = \mathscr{S} \setminus S_o$, is optimized. Prediction error variance has been popularly used as a performance measure of one sampling design. If the prediction error variance is denoted as V, the design target is then given by Equation (2.1).

$$\min_{S_o \subset \mathscr{S}} f(V), \tag{2.1}$$

where $f(\cdot)$ is some function of the prediction error variance. Examples of $f(\cdot)$ include minimum, maximum, and average.

In general, sampling design approaches can be divided into three main categories: (1) random sampling, (2) spatial variability based method, and (3) sequential design approaches based on computer experiment methods.

Random sampling is a widely applied sampling design technique. One basic assumption of random sampling is that no spatial correlation of random variables exists among locations to be sampled. Two most popular random sampling designs are simple random sampling (SI) and stratified simple random sampling (STSI) (Brus and De Gruijter, 1997). SI suggests to randomly select a predetermined number of sample locations n_o from the surface of interest $\mathscr{S}$ with equal probabilities of selection and independently from one another. With STSI, the surface of interest $\mathscr{S}$ is first divided into sub-regions, also called strata, $\mathscr{S}_1$, ..., $\mathscr{S}_d$, and then SI is applied to each stratum $\mathscr{S}_i$, $i = 1, \ldots, d$, individually. The sample size of the stratum i, $n_{oi} = |\mathscr{S}_i|$, can be determined such that the probabilities of the locations of being sampled differ among strata. SI is more widely used because it is easier to implement, but this strategy often results in a larger sample size and duplication of information because some measurement locations are too close to each other (Curran, 1988). If an optimal sample size is defined as the number of sample points n_o that minimizes the measurement cost given a pre-specified estimation precision, STSI is able to increase the cost-efficiency by taking the within stratum variances into consideration.

While random sampling is convenient to implement in practice, it does not consider the spatial correlation among sampling locations, possibly leading to non-optimal design and high measurement cost. Geostatistical methods are thus introduced to take the spatial variability into consideration. This type of method is based on the regionalized variable theory which assumes that the closer the sampled locations, the more similar the sampled data. Normally, the spatial variability is decomposed into global and local variability (Anderson et al., 2006). The global spatial variability indicates the spatial variation pattern of a variable. The local variability refers to the variation of the variable within a neighborhood. According to the extent of considering global and local spatial variability, there are two geostatistical sampling techniques: global variability based approaches (Curran and Williamson, 1986; Curran, 1988; Anderson et al., 2006; McBratney and Webster, 1983a,b; Atkinson et al., 1992, 1994; Xiao et al., 2005; Curran and Williamson, 1986; Curran, 1988) and local variability based methods (Anderson et al., 2006).

Sequential design approach is another type of sampling strategy. An exemplary method is reported in Jin et al. (2012), where a systematic measurement strategy is designed for wafer geometric profile estimation based on sequential design methods. The core idea of such methods is to sequentially allocate additional sampling points at the locations where a higher improvement is expected, thus reducing the number of experiments/samples to achieve the optimal solution. Jin et al. (2012) combined the sequential design approach with prior engineering understanding of the wafer cutting process, and developed a sampling method for wafer surface quality monitoring. Once the optimal sampling locations are determined for a representative wafer profile, the strategy will be used for other profiles in the same manufacturing batch.

The sampling methods reviewed so far are applicable for a single spatial process or multiple processes with high similarities, but cannot be directly applied to the sampling design problem of spatiotemporal processes. As a matter of fact, there is relatively limited work on sampling designs for spatiotemporal processes (Fanshawe and Diggle, 2013). For a spatiotemporal process, the spatial structure varies over time, and therefore the optimal design at an earlier time may not always be optimal. In order to cost-effectively characterize a spatiotemporal process, the sampling design is supposed to be determined according to the time-variant spatial variation pattern. Those approaches are generally referred to as "dynamic sampling." Wikle and Royle (1999) is one of the earlies studies on this topic. In this study, a space-time dynamic model is used for characterizing spatiotemporal processes, and Kalman filter is adopted to derive the prediction covariance matrix conditional on the measurements at the current time and all historical data in a recursive manner. Then dynamic sampling design is achieved by minimizing some design criterion based on the prediction error variance. Following this study, dynamic sampling design methods have been developed for ecological monitoring with both Gaussian (Hooten et al., 2009) and non-Gaussian (e.g., Poisson or log-normal) data (Wikle, 2004).

Existing dynamic sampling methods are mostly focused on the applications of environmental monitoring (Wikle and Royle, 1999; Fanshawe and Diggle, 2013) and ecological monitoring (Hooten et al., 2009; Wikle, 2004). However, there is a lack of studies on spatiotemporal processes in manufacturing. Compared with environmental or ecological monitoring, manufacturing processes have some unique characteristics, which make the design problem more complicated. Specifically,

(1) In manufacturing process control, multiple measurement systems are often jointly

used. For instance, in tool condition monitoring for ultrasonic metal welding, a microscope has various lenses with different measurement ranges and resolutions. In surface quality monitoring for high precision machining of automotive engine heads, a coordinate measuring machine and a profilometer are adopted to measure a path following the edges of a rectangle and several rectangular regions, respectively (Suriano et al., 2015). Therefore, the dynamic sampling design should take the measurement capability of all the gages into consideration.

(2) Existing methods have seldom considered the measurement cost during the design process. In ecological and environmental monitoring, the total number of measurement locations or monitoring devices is often determined based on budget constraints (Hooten et al., 2009). However, in manufacturing, it is important to consider the measurement cost to achieve a balance between sampling cost and prediction precision.

(3) The design space in manufacturing is much higher than that in environmental and ecological monitoring. For example, the potential measurement points on a ultrasonic welding anvil can be 9 million. The design space in environmental or ecological monitoring normally consists of several hundred potential locations (Hooten et al., 2009). With such a high-dimensional space, exchange algorithms, which are commonly used in existing studies, may not be effective in searching for the optimal design in manufacturing.

(4) An algorithm for monitoring and updating the model parameters is necessary. Model parameters of spatiotemporal processes in manufacturing are generally time-variant. For example, in a process where tool wear is the major cause for the

spatiotemporal variation, the tool degradation rate is often time-variant. It is critical to monitor and update the model parameters, as inaccurate parameters will likely result in non-optimal designs and inaccurate interpolation.

To address these challenges, this chapter develops a dynamic sampling design approach to characterize spatiotemporal processes in manufacturing. The rest of this chapter is organized as follows. Section 2.2 presents a formulation of the dynamic sampling design problem, a state-space spatiotemporal model, and recursive prediction error estimation based on Kalman filter. Section 2.3 introduces a dynamic sampling design approach with a new design criterion, a genetic algorithm (GA) based design search, and a novel algorithm for estimating and monitoring the temporal transition parameter.

2.2 Problem Formulation

This section formulates the dynamic sampling design problem for spatiotemporal processes using a state-space spatiotemporal model and Kalman filter.

2.2.1 Spatiotemporal Model

Over the recent years, various spatiotemporal models have been proposed, among which a state-space approach has received extensive attention (Xu and Wikle, 2007; Huang and Cressie, 1996; Wikle and Royle, 1999; Wikle, 2004; Hooten et al., 2009). One advantage of this approach is that empirical Bayesian or spatiotemporal Kalman filter can be used to easily implement such models (Wikle and Royle, 1999). We will briefly review the model formulation as follows.

Let $\mathbf{Z}_t = (Z(\mathbf{s}_1,t),\ldots,Z(\mathbf{s}_n,t))^\mathsf{T}$ be an $n \times 1$ vector for an unobservable spatiotemporal process at some fixed set of locations $\mathscr{S} = \{\mathbf{s}_1,\ldots,\mathbf{s}_n\}$. In surface manufacturing, $\mathbf{s}$ is often a two-dimensional geographic coordinate vector, namely, $\mathbf{s} = [x,\ y]$. Additionally, in manufacturing, the location set $\mathscr{S}$ of interest is often a grid, and is constant over time. The spatiotemporal model is given by the following state equation.

$$\mathbf{Z}_t = \boldsymbol{\Phi}_t \mathbf{Z}_{t-1} + \boldsymbol{\eta}_t, \tag{2.2}$$

where $\boldsymbol{\Phi}_t$ is a first-order Markov parameter matrix and $\boldsymbol{\eta}_t$ is a spatiotemporal process with covariance matrix Σ_{η_t}. In spatiotemporal processes where the spatial variation is caused by tool degradation, $\boldsymbol{\Phi}_t$ represents the tool degradation dynamics and can be time-variant.

At a particular time t, a subset of $\mathscr{S}$ is measured, and the measurement data are given by $\mathbf{M}_t = (M(\bar{\mathbf{s}}_1,t),\ldots,M(\bar{\mathbf{s}}_{m_t},t))^\mathsf{T}$, which is an $m_t \times 1$ vector. Supposing the measurements/observations are acquired with errors, a measurement equation can be written as follows.

$$\mathbf{M}_t = \mathbf{D}_t \mathbf{Z}_t + \boldsymbol{\epsilon}_t, \tag{2.3}$$

where $\mathbf{M}_t$ is a vector of m_t measurement data; $\mathbf{D}_t$ is an $m_t \times n$ matrix that maps the true process, $\mathbf{Z}_t$, to the data at observed locations, $\mathbf{M}_t$; and $\boldsymbol{\epsilon}_t$ is a zero-mean measurement error process with covariance matrix $\boldsymbol{\Sigma}_{\varepsilon_t}$. $\mathbf{D}_t$ determines which locations in $\mathbf{Z}_t$ are observed and it is a sparse matrix of 0's and 1's (Wikle et al., 1998). Consequently, the dynamic sampling design is equivalent to the determination of $\mathbf{D}_t$.

In the context of manufacturing processes, we make the following assumptions regarding Models (2.2) and (2.3).

Assumption 2.1 (Model Assumptions for Spatiotemporal Model)

(1) The transition matrix takes the form of

$$\mathbf{\Phi}_t = \phi_t \mathbf{I}, \tag{2.4}$$

where ϕ_t is the surface progression rate and should be monitored and updated, and $\mathbf{I}$ is an $n \times n$ identity matrix. In spatiotemporal processes where tooling degradation is the major factor, ϕ_t can be viewed as the tooling degradation rate.

(2) Σ_{ϵ_t} is a diagonal matrix and is fully determined by the gage repeatability and the design $\mathbf{D}_t$.

(3) Σ_{η_t} is time-invariant, and the parameters can be estimated and updated over time.

2.2.2 Prediction Error Estimation

Given the state-space model (2.2) and the measurement equation (2.3), a state-space Kalman filter can be developed either by Bayesian arguments (Harrison and West, 1999) or projection arguments (Hamilton, 1994). Then we can obtain recursive equations for the prediction error covariance, as shown by Equations (2.5)–(2.8) (Wikle and Royle, 1999).

$$\mathbf{A}_t \equiv \mathrm{Var}(\hat{\mathbf{Z}}_t | \mathbf{M}_t, \ldots, \mathbf{M}_1) \tag{2.5}$$

$$= \mathbf{B}_t - \mathbf{B}_t \mathbf{D}_t^{\mathsf{T}} \left[\mathbf{D}_t \mathbf{B}_t \mathbf{D}_t^{\mathsf{T}} + \Sigma_{\epsilon_t} \right]^{-1} \mathbf{D}_t \mathbf{B}_t \tag{2.6}$$

$$\mathbf{B}_t \equiv \mathrm{Var}(\hat{\mathbf{Z}}_t | \mathbf{M}_{t-1}, \ldots, \mathbf{M}_1) \tag{2.7}$$

$$= \mathbf{\Phi}_t \mathbf{A}_{t-1} \mathbf{\Phi}_t^{\mathsf{T}} + \Sigma_{\eta_t} \tag{2.8}$$

To start the recursion, $\mathbf{A}_0$ needs to be specified and is typically chosen to be the unconditional variance-covariance matrix of the spatiotemporal process (Wikle and Royle, 1999).

Moreover, by Assumption 2.1, Σ_{ϵ_t} is jointly determined by $\mathbf{D}_t$ and the gage repeatability; $\Sigma_{\eta_t} = \Sigma_{\eta}$ is constant and will be updated after each sampling; the calculation of $\mathbf{B}_t$ can be further simplified as the following equation.

$$\mathbf{B}_t = \phi_t^2 \mathbf{A}_{t-1} + \Sigma_{\eta}. \tag{2.9}$$

2.3 Dynamic Sampling Design

This section presents a dynamic sampling design approach for spatiotemporal processes in manufacturing. Particularly, the design problem is formulated in Section 2.3.1; Section 2.3.2 presents a new design criterion which simultaneously considers the prediction precision and the measurement cost; finally, in Section 2.3.3, a binary integer programming problem is formulated and GA is proposed for the solution search.

2.3.1 Measurement System Setting

For the measurement and monitoring of spatial variations in manufacturing, various measurement devices are available, including coordinate measuring machines (CMMs) (Suriano et al., 2015), profilometers (Suriano et al., 2015), 3D microscopes (Shao et al., 2014, 2016), laser holographic interferometers (Suriano et al., 2015), and touching probe sensors (Jin et al., 2012). Simultaneous adoption of multiple gages with different measurement resolutions is a common practice. For instance, in high precision machining of automobile engine components, a CMM and a profilometer are used in combination to monitor the surface quality; in tool condition monitoring of ultrasonic metal welding, different lenses of a 3D microscope may be used together for surface measurement.

In this study, we assume a 3D measurement system with two resolution levels, i.e., Level 1 and Level 2, is applied to characterize a spatiotemporal process. At each sampling time, two levels are used in combination for measurement. We make the following assumptions regarding the measurement from Levels 1 and 2.

Assumption 2.2 (Measurement System Setting)

(1) Level 1 is able to measure the whole surface with one-time scanning but it has a relatively poor precision, which is reflected by a large measurement error variance, or a poor repeatability. We denote the measurement error variance of Level 1 as σ_1^2.

(2) The measurement locations of Level 1 are fixed.

(3) Level 2 has a smaller measurement error variance, denoted as σ_2^2*, but the measurement range is small. We have*

$$\sigma_1^2 > \sigma_2^2. \tag{2.10}$$

(4) The measurement locations of Level 2 are programmable.

(5) In designing the sampling locations for Level 2 measurement, a surface needs to be segmented into a number of grids, and each grid is measured by one single scan. The "Grid Segmentation" process will be explained later. The design problem is then equivalent to selecting a subset of all grids.

(6) The unit measurement costs of Level 1 and Level 2 are assumed to be c_1 *and* c_2*, respective. Level 2 measurement is more expensive to acquire, so we have*

$$c_1 < c_2. \tag{2.11}$$

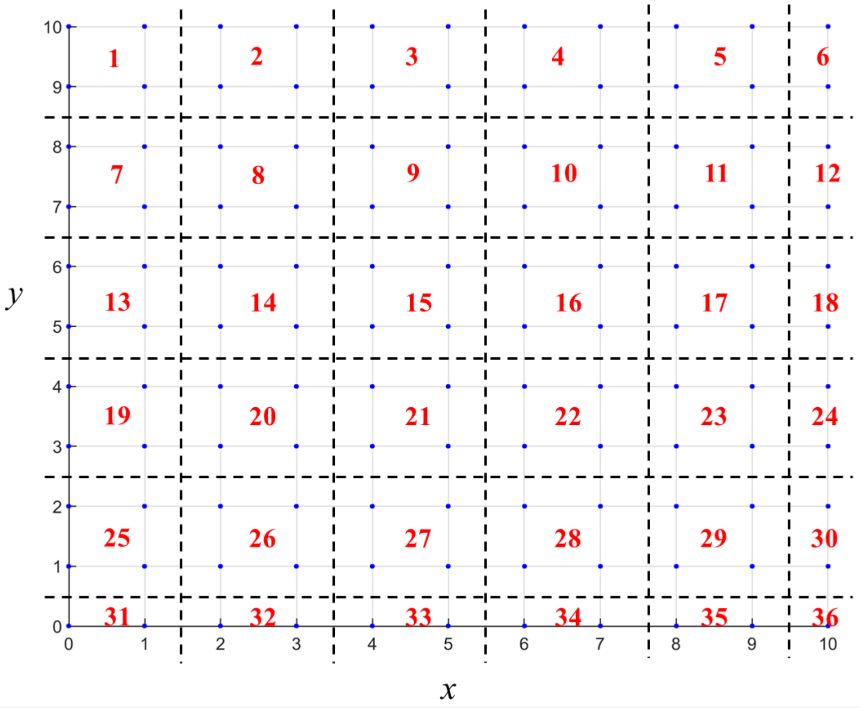


Figure 2.3 Illustration of grid segmentation for Level 2 measurement

Figure 2.3 illustrates the process of grid segmentation. The points in the prediction location set $\mathscr{S}$ are assumed to align with an equally spaced grid, the size of which is $N_x \times N_y$, as shown by the blue dots in Figure 2.3. The measurement range of Level 2 is $n_{d,x} \times n_{d,y}$, indicating that one single scan of Level 2 can measure $n_{d,x} \times n_{d,y}$ points. Then the total number of grids is calculated as

$$N_g = \left\lceil \frac{N_x}{n_{d,x}} \right\rceil \times \left\lceil \frac{N_y}{n_{d,y}} \right\rceil, \tag{2.12}$$

where N_g is the total number of grids, and $\lceil \cdot \rceil$ is a ceiling function.

In the case shown by Figure 2.3, $N_x = N_y = 11, n_{d,x} = n_{d,y} = 2$, so the total number of grids is $N_g = \lceil \frac{11}{2} \rceil \times \lceil \frac{11}{2} \rceil = 36$, and the grids are indexed as $1, 2, \ldots, 36$.

The sampling design is to select a certain number of grids from all N_g grids. The design target is two-fold: (1) to determine the number of measurement grids (n_g) and (2) to determine where to allocate these grids. The number of all possible designs, N_{design} is calculated by the following equation.

$$N_{\text{design}} = \sum_{n_g=0}^{N_g} \binom{N_g}{n_g} = 2^{N_g}. \tag{2.13}$$

For the example in Figure 2.3, the number of all possible designs is $2^{36} \approx 687$ million.

The design matrix, $\mathbf{D}_t$ can be rewritten as

$$\mathbf{D}_t = \begin{bmatrix} \mathbf{D}_1 \\ \\ \mathbf{D}_{2,t} \end{bmatrix}, \tag{2.14}$$

where $\mathbf{D}_1$ is time-invariant, but $\mathbf{D}_{2,t}$ is time-variant and is designed predictively.

The number of measurement points of Level 1 and Level 2 can be calculated as

$$m_1 = \text{trace}\left(\mathbf{D}_1\mathbf{D}_1^\intercal\right), \tag{2.15}$$

$$m_{2,t} = \text{trace}\left(\mathbf{D}_{2,t}\mathbf{D}_{2,t}^\intercal\right), \tag{2.16}$$

where m_1 and $m_{2,t}$ are the number of measurement points of Levels 1 and 2, respectively.

Then the total measurement cost, C_t, is given by the equation below.

$$C_t = c_1 m_1 + c_2 m_{2,t}. \tag{2.17}$$

With the representation of Equation (2.14), the measurement equation, namely, Equation (2.3), is rearranged, as shown by Equation (2.18).

$$\mathbf{M}_t = \begin{bmatrix} \mathbf{M}_{1,t} \\ \\ \mathbf{M}_{2,t} \end{bmatrix}, \tag{2.18}$$

where $\mathbf{M}_{1,t}$ and $\mathbf{M}_{2,t}$ are the measurements of Level 1 and Level 2, respectively. $\mathbf{M}_{1,t}$ and $\mathbf{M}_{2,t}$ can be calculated using

$$\mathbf{M}_{1,t} = \mathbf{D}_1 \mathbf{Z}_t, \tag{2.19}$$

$$\mathbf{M}_{2,t} = \mathbf{D}_{2,t} \mathbf{Z}_t. \tag{2.20}$$

Accordingly, the covariance matrix of the measurement error process, ϵ_t, can be estimated with the following equation.

$$\Sigma_{\epsilon_t} = \begin{bmatrix} \sigma_1^2 \mathbf{I}_{m_1 \times m_1} & \mathbf{0}_{m_1 \times m_{2,t}} \\ \mathbf{0}_{m_{2,t} \times m_1} & \sigma_2^2 \mathbf{I}_{m_{2,t} \times m_{2,t}} \end{bmatrix}, \tag{2.21}$$

where $\mathbf{I}_{m_1 \times m_1}$ and $\mathbf{I}_{m_{2,t} \times m_{2,t}}$ are identity matrices, and $\mathbf{0}_{m_1 \times m_{2,t}}$ and $\mathbf{0}_{m_{2,t} \times m_1}$ are zero matrices.

Note: if some locations are measured by both Level 1 and Level 2, we will only use the measurement from Level 2, because Level 2 is more precise. Meanwhile, the specific rows corresponding to these locations will be removed from $\mathbf{D}_1$, and we denote the modified design matrix as $\mathbf{D}'_{1,t}$. Moreover, the formulation in the equations above will be revised accordingly.

2.3.2 Design Criterion

The design criterion can be constructed using some function of the prediction covariance matrix, $\mathbf{A}_t$, such as average or maximum prediction error variance (Wikle and Royle, 1999). Then the design procedure follows a two-step algorithm.

Algorithm 2.1 (Dynamic Sampling Design Procedure)

Step 1*: Calculate* $\mathbf{B}_t$ *based on* $\mathbf{A}_{t-1}$*;*

Step 2*: Minimize the pre-specified design criterion over all possible designs* $\mathscr{D}$*.*

Average or maximum prediction variance is easy to implement in practice; however, there are two major limitations in using these criteria. First, the measurement cost cannot be incorporated in the design process. Although it may be reasonable in environmental and ecological monitoring to predetermine the total number of observations according to budget constraints, it is not desirable to use such criteria in characterizing spatiotemporal processes in manufacturing. Second, these criteria cannot accommodate the scenario where one has varying precision requirement on a surface. For example, in tool condition monitoring for ultrasonic metal welding, the knurl peaks are of more concern because these areas play a significant role in determining the weld quality.

Here, we propose a new design criterion to simultaneously consider the measurement cost and the "weighted average prediction variance." The weighted average prediction variance is defined as

$$V_t = \mathrm{trace}\left(\mathbf{A}_t\mathbf{W}\right), \tag{2.22}$$

where $\mathbf{A}_t$ is the prediction covariance matrix and $\mathbf{W}$ is an $n \times n$ weight matrix, and it is given by Equation (2.23).

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & w_n \end{bmatrix}, \tag{2.23}$$

where the diagonal entry of $\mathbf{W}$, $\mathbf{W}_{ii}$, is the weight assigned to the ith location in $\mathbf{Z}$. Higher weight corresponds to stricter precision requirement. If the precision requirement is time-variant, one can simply adopt a varying weight matrix, $\mathbf{W}_t$.

The precision measure in Equation (2.23) may be rewritten as Equation (2.24).

$$V_t = \sum_{i=1}^{n} w_i \mathbf{A}_{t(ii)}, \tag{2.24}$$

where $\mathbf{A}_{t(ii)}$ is the ith diagonal entry of $\mathbf{A}_t$.

The design criterion is defined by a loss function, L_t, as shown by Equation (2.25).

$$L_t = C_t + \lambda V_t, \tag{2.25}$$

where λ is a tuning parameter. Larger λ places more emphasis on the prediction precision, and will generally lead to better prediction precision and higher measurement cost. In practice, λ can be determined based on expert knowledge.

2.3.3 GA-Based Binary Integer Programming

The dimension of the design space $\mathscr{D}$ in characterizing a spatiotemporal process in manufacturing could be extremely high. In the example shown by Figure 2.3, there are 687 million possibilities in selecting measurement grids for Level 2. This number will increase exponentially, as shown by Equation (2.13). Exchange algorithms, which have been commonly applied in existing studies on dynamic sampling design, might not be effective in dealing with such a high-dimensional design space. Therefore, a more effective programming technique is needed. In this section, we first formulate the design process as a binary integer programming problem, and then use GA to solve this problem.

The design problem is essentially equivalent to determining whether to conduct measurement at a certain grid, which is a binary decision. Assume there are in total N_g grids in the surface of interest. We define an $N_g \times 1$ design vector d_t, and each element can only take the values of 0 or 1. A 0 value indicates that there is no measurement at the corresponding grid, while 1 suggests that a Level 2 measurement is performed at that grid. Note that the design vector d_t and the Level 2 design matrix, $\mathbf{D}_{2,t}$, are one-to-one mapped. Additionally, given that the Level 1 design matrix, $\mathbf{D}_1$, is fixed, d_t and $\mathbf{D}_t$ are also one-to-one mapped. Hence, the dynamic sampling design problem can be well represented by a binary integer programming problem (Taha, 2014).

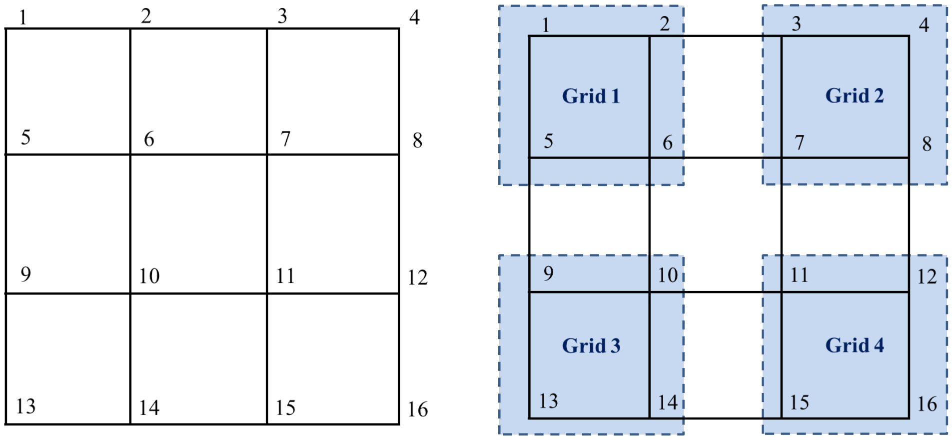
There are various techniques available in the literature to solve integer programming problems (Burer and Letchford, 2012), including exact methods, such as branch-and-bound (Land and Doig, 1960), branch-and-reduce (Ryoo and Sahinidis, 1995, 1996), and heuristic methods, which may be based on tabu search (Exler et al., 2008) and GA (Luo et al., 2007; Young et al., 2007; Schlüter et al., 2009). This study adopts a GA based integer programming approach, which is implemented in Matlab (Attaway, 2013). One practical suggestion in solving this programming problem is that when the number of grids is large, the computation efficiency can be greatly improved by adopting sparse matrix representation (Gilbert et al., 1992), as the design matrix, $\mathbf{D}_t$, is a highly sparse matrix.

2.3.4 An Illustration Example of a Design Process

This section provides an example to illustrate a design process. We consider a surface with the following settings.

$$\begin{aligned} & N_x = N_y = 4;\; n_{d,x} = n_{d,y} = 2; \\ & \sigma_1^2 = 5,\; \sigma_2^2 = 1;\; c_1 = 1,\; c_2 = 2; \\ & \mathbf{W} = \begin{bmatrix} 2 & \mathbf{0}_{16\times 15} \\ \mathbf{0}_{15\times 1} & \mathbf{I}_{15\times 15} \end{bmatrix}. \end{aligned} \tag{2.26}$$

The surface dimension is 4 × 4, and there are 16 prediction locations. We first index these locations using integers 1 to 16, as shown by Figure 2.4a. As the measurement range of Level 2 is $n_{d,x} \times n_{d,y} = 2 \times 2$, we shall segment the surface into 2 × 2 grids, and the segmentation result is shown by Figure 2.4b. After the grid segmentation, there are $N_g = 4$ grids.



(a) Surface with indices

(b) Surface after grid segmentation

Figure 2.4 Illustration of a design process

The weight matrix is a 16 × 16 diagonal matrix. Of the diagonal entries in **W**, only $\mathbf{W}_{11} = 2$, and all other entries are equal to 1, i.e., $\mathbf{W}_{jj} = 1,\ j = 2, \ldots, 16$.

We further suppose that the measurement locations of Level 1 are fixed at Locations 1, 4, 13, and 16. Correspondingly, the Level 1 design matrix is

$$\mathbf{D}_1 = \begin{array}{c} \begin{array}{ccccccc} \text{1st column} & & \text{4th column} & & \text{13th column} & & \text{16th column} \end{array} \\ \begin{pmatrix} 1 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 8} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & \mathbf{0}_{1\times 2} & 1 & \mathbf{0}_{1\times 8} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 8} & 1 & \mathbf{0}_{1\times 2} & 0 \\ 0 & \mathbf{0}_{1\times 2} & 0 & \mathbf{0}_{1\times 8} & 0 & \mathbf{0}_{1\times 2} & 1 \end{pmatrix} \end{array} \tag{2.27}$$

Assume that at a certain time t, Grid 1 is selected for Level 2 measurement. Equivalently, Locations 1, 2, 5, and 6 are measured using Level 2. The design vector for Level 2 is then

$$d_t = \begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\mathsf{T}. \tag{2.28}$$

Correspondingly, the design matrix for Level 2 is given by

$$\mathbf{D}_{2,t} = \begin{array}{c} \begin{array}{cccccc} \text{1st column} & \text{2nd column} & & \text{4th column} & \text{5th column} & \end{array} \\ \begin{pmatrix} 1 & 0 & \mathbf{0}_{1\times 2} & 0 & 0 & \mathbf{0}_{1\times 10} \\ 0 & 1 & \mathbf{0}_{1\times 2} & 0 & 0 & \mathbf{0}_{1\times 10} \\ 0 & 0 & \mathbf{0}_{1\times 2} & 1 & 0 & \mathbf{0}_{1\times 10} \\ 0 & 0 & \mathbf{0}_{1\times 2} & 0 & 1 & \mathbf{0}_{1\times 10} \end{pmatrix} \end{array} \tag{2.29}$$

As shown by Equations (2.27) and (2.29), $\mathbf{D}_1$ and $\mathbf{D}_{2,t}$ are sparse matrices, where there is only one element taking the value of 1 in each row, and the number of rows is equal to the number of measurement points. The numbers of measurement locations for Level 1 and

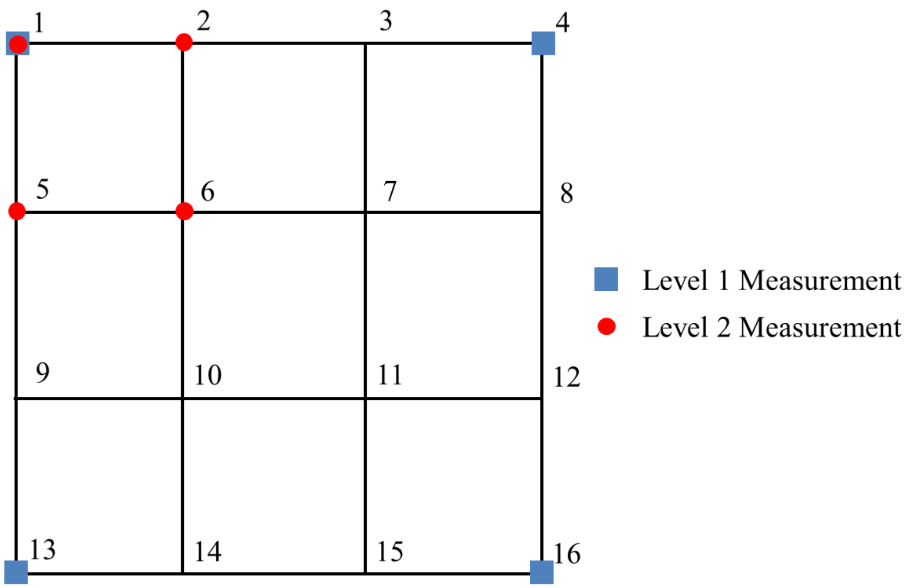


Figure 2.5 Illustration of the measurement allocation

Level 2 can also be calculated using Equations (2.15) and (2.16). In any case, we obtain $m_1 = 4$ and $m_{2,t} = 4$. The measurement cost is then calculated as $C_t = 4 \times 1 + 4 \times 2 = 12$.

This design is further illustrated by Figure 2.5. In Figure 2.5, blue squares indicate Level 1 measurement and red dots represent Level 2 measurement.

Note that Location 1 is measured by both Level 1 and Level 2. Since Level 2 is more precise, we will use the measurement from Level 2 for this point. Correspondingly, we modify $\mathbf{D}_1$ by removing the first row, i.e.,

$$\mathbf{D}'_{1,t} = \begin{array}{c} \begin{array}{cccccc} & \text{4th column} & & \text{13th column} & & \text{16th column} \end{array} \\ \begin{pmatrix} \mathbf{0}_{1\times 3} & 1 & \mathbf{0}_{1\times 8} & 0 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{1\times 3} & 0 & \mathbf{0}_{1\times 8} & 1 & \mathbf{0}_{1\times 2} & 0 \\ \mathbf{0}_{1\times 3} & 0 & \mathbf{0}_{1\times 8} & 0 & \mathbf{0}_{1\times 2} & 1 \end{pmatrix} \end{array} \tag{2.30}$$

Stacking $\mathbf{D}_1'$ and $\mathbf{D}_{2,t}$ vertically, we obtain the overall design matrix:

$$\mathbf{D}_t = \begin{array}{c} \begin{array}{cccccccccc} \substack{\text{1st}\\\text{column}} & \substack{\text{2nd}\\\text{column}} & & \substack{\text{4th}\\\text{column}} & \substack{\text{5th}\\\text{column}} & \substack{\text{6th}\\\text{column}} & & \substack{\text{13th}\\\text{column}} & & \substack{\text{16th}\\\text{column}} \end{array} \\ \left(\begin{array}{cccccccccc} 0 & 0 & 0 & 1 & 0 & 0 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times 6} & 1 & \mathbf{0}_{1\times 2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 2} & 0 \end{array}\right) \end{array} \tag{2.31}$$

The covariance matrix of the measurement error process, ϵ_t, is evaluated as

$$\Sigma_{\epsilon_t} = \begin{bmatrix} 5\mathbf{I}_{3\times 3} & \mathbf{0}_{3\times 4} \\ \mathbf{0}_{4\times 3} & \mathbf{I}_{4\times 4} \end{bmatrix}, \tag{2.32}$$

where $\mathbf{I}_{3\times 3}$ and $\mathbf{I}_{4\times 4}$ are identity matrices, and $\mathbf{0}_{3\times 4}$ and $\mathbf{0}_{4\times 3}$ are zero matrices.

2.3.5 Estimation and Monitoring of the Temporal Transition Parameter

Accurate estimation of the temporal transition parameter, ϕ_t, is crucial in dynamic sampling design, as an inaccurate estimation will lead to incorrect evaluation of $\mathbf{A}_t$ and $\mathbf{B}_t$, and the error will accumulate over time, deteriorating the overall design and prediction performance. Therefore, an algorithm for estimating and monitoring ϕ_t is critically needed.

In the dynamic sampling design literature, Hooten et al. (2009) suggests using a method of moments estimator for ϕ_t using all historical measurement data. Nevertheless, the estimator is developed based on the assumption that the transition parameter is constant, which may not be a valid assumption in manufacturing.

On the monitoring of transition parameters, possible solutions include system identification techniques for state-space models (Juang, 1994; Van Overschee and De Moor, 1994; Ljung, 1998; Pillonetto et al., 2014) and methods for monitoring autoregressive parameters in time series models (Gombay and Serban, 2009; Gombay, 2008; Na et al., 2011; Davis et al., 1995). However, these methods cannot be directly applied in dynamic sampling design either because of their high-complexity or large detection delays.

In fact, the monitoring problem in dynamic sampling design is unique in the sense that for the measurement data, the error variance is known and fully determined by the measurement system's repeatability. This information provides us with an opportunity for monitoring the parameter change. In the rest of this section, we will develop an algorithm to monitor and update the temporal transition parameter, ϕ_t, based on hypothesis testing using the measurement data.

We first define the set of locations measured by Level 1 to be $\mathbf{S}_1$, which is time-invariant by assumption. At time t, we denote the set of locations measured by Level 2 as $\mathbf{S}_{2,t}$. Then, at time $t-1$, the Level 1 and Level 2 measurement locations are $\mathbf{S}_1$ and $\mathbf{S}_{2,t-1}$, respectively. The locations mutually measured by Level 2 at both time $t-1$ and t can be obtained by

$$\mathbf{S}_{2,t(t-1)} = \mathbf{S}_{2,t-1} \cap \mathbf{S}_{2,t}, \tag{2.33}$$

where $\mathbf{S}_{2,(t-1)t}$ represents the set of mutually measured locations by Level 2 at both time $t-1$ and t.

Then we denote the measurements at locations of $\mathbf{S}_{2,(t-1)t}$ at time $t-1$ and t to be $\mathbf{M}^{t-1}_{2,(t-1)t}$ and $\mathbf{M}^{t}_{2,(t-1)t}$, respectively.

The mutual measurements at time $t-1$ and t can be represented as stacked measurement vectors of Level 1 and Level 2, i.e.,

$$\begin{aligned} \mathbf{M}_{(t-1)t}^{t-1} &= \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_{2,(t-1)t}^{t-1} \end{bmatrix}, \\ \mathbf{M}_{(t-1)t}^{t} &= \begin{bmatrix} \mathbf{M}_1 \\ \mathbf{M}_{2,(t-1)t}^{t} \end{bmatrix}, \end{aligned} \tag{2.34}$$

where $\mathbf{M}_{(t-1)t}^{t-1}$ and $\mathbf{M}_{(t-1)t}^{t}$ are the measurement data at mutually measured locations at time $t-1$ and t, including both Level 1 and Level 2 measurement.

Statistics are defined as the average height of the measurements at these mutually locations at time $t-1$ and t, as given by Equation (2.35).

$$\begin{aligned} u_{t-1,(t-1)t} &= \text{avg}\left(\mathbf{M}_{(t-1)t}^{t-1}\right), \\ u_{t,(t-1)t} &= \text{avg}\left(\mathbf{M}_{(t-1)t}^{t}\right), \end{aligned} \tag{2.35}$$

where $\text{avg}(\cdot)$ is an average function.

A one-time temporal transition parameter, which is denoted by h_t, is defined as follows.

$$h_t = \frac{u_{t,(t-1)t}}{u_{t-1,(t-1)t}}. \tag{2.36}$$

When no shift occurs in the temporal transition parameter, ϕ_t, it is estimated using Equation (2.37).

$$\hat{\phi}_t = \frac{1}{t-t_0} \sum_{i=t_0}^{t} h_i = \frac{1}{t_0} \left[(t-t_0)\hat{\phi}_{t-1} + h_t \right], \tag{2.37}$$

where t_0 is the time right after the previous shift occurs. When no shift has occurred, $t_0 = 1$.

When a shift is detected, we will estimate ϕ_t using the one-time temporal transition parameter, h_t, namely,

$$\hat{\phi}_t = h_t. \tag{2.38}$$

The monitoring of ϕ_t is conducted using the following hypothesis test [1].

Test 2.1 (Shift Detection of ϕ_t) *Conclude that H_0: $h_t = \hat{\phi}_{t-1}$ is not supported when*

$$\left| \frac{u_{t,(t-1)t} - \hat{u}_{t,(t-1)t}}{\sqrt{\left(1+\hat{\phi}_{t-1}^2\right)\sigma_{\varepsilon,(t-1)t}^2}} \right| > Z_{\alpha/2}, \tag{2.39}$$

where $Z_{\alpha/2}$ is the standard normal critical value, and

$$\begin{aligned} \hat{u}_{t,(t-1)t} &= \hat{\phi}_{t-1} u_{t-1,(t-1)t}, \\ \sigma_{\varepsilon,(t-1)t}^2 &= \frac{m_1 \sigma_1^2}{\left(m_1 + m_{2,(t-1)t}\right)^2} + \frac{m_{2,(t-1)t}\sigma_2^2}{\left(m_1 + m_{2,(t-1)t}\right)^2}, \\ m_{2,(t-1)t} &= \left|\mathbf{S}_{2,t(t-1)}\right|. \end{aligned} \tag{2.40}$$

The estimation and monitoring procedure is summarized by Figure 2.6.

2.4 Case Study

This section presents a case study to compare the proposed dynamic sampling design approach with the random sampling method. The experiment setting is summarized as follows.

$$\begin{aligned} &N_x = N_y = 16;\ n_{d,x} = n_{d,y} = 2; \\ &\sigma_1^2 = 1,\ \sigma_2^2 = 0.1;\ c_1 = 1,\ c_2 = 1.5; \\ &\lambda = 1.3;\ t = 1, 2, \ldots, 10. \end{aligned} \tag{2.41}$$

Additionally, Level 1 measures 16 points, as given by Equation (2.42). The Level 1

---

[1] Detailed derivation of the hypothesis test is given in Appendix A.

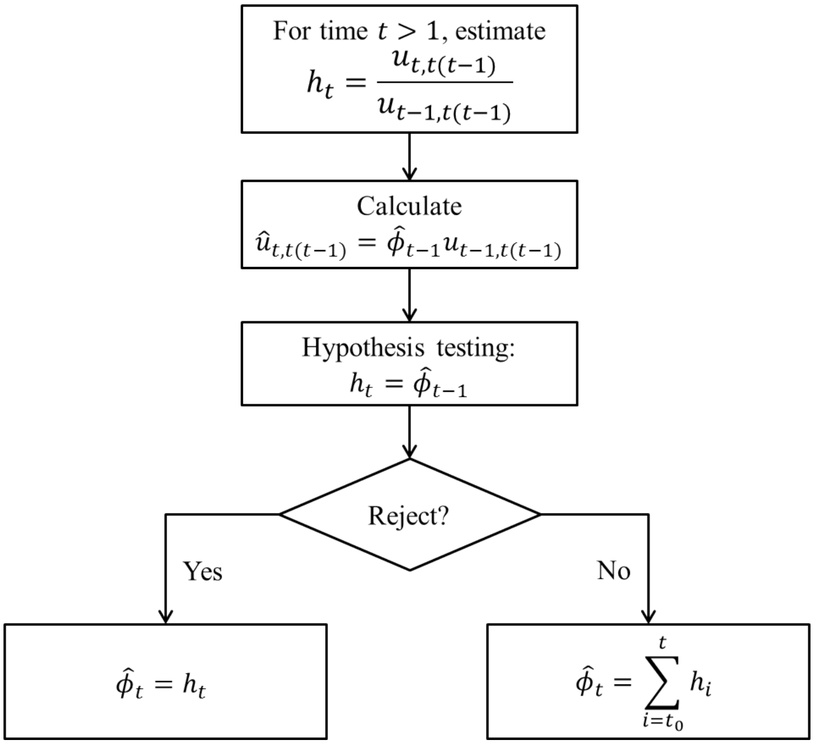


Figure 2.6 Flowchart for the estimation and monitoring of ϕ_t

measurement locations are further illustrated by Figure 2.7.

$$\begin{array}{llll} (1,1), & (1,6), & (1,11), & (1,16) \\ (6,1), & (6,6), & (6,11), & (6,16) \\ (11,1), & (11,6), & (11,11), & (11,16) \\ (1,16), & (16,6), & (16,11), & (16,16) \end{array} \tag{2.42}$$

The weight matrix is given by Equation (2.43).

$$\mathbf{W}_{ii} = \begin{cases} 3 & \text{if } i \in \{61,62,\ldots,70,201,202,\ldots,220\} \\ 1 & \text{otherwise} \end{cases} \tag{2.43}$$

The temporal transition parameter is assumed to be a step function, as given in Equation

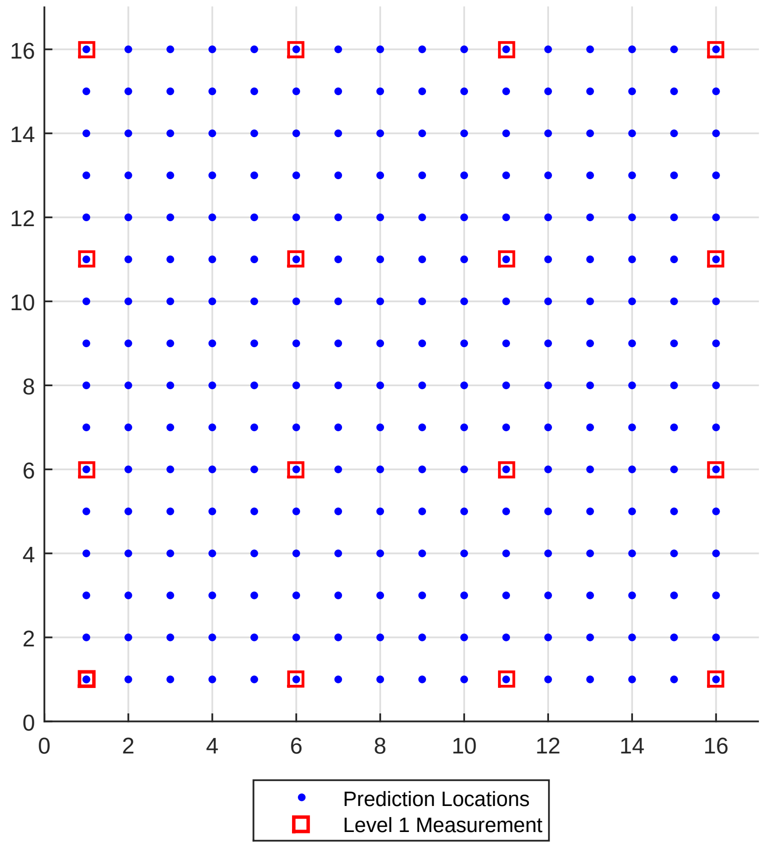


Figure 2.7 Level 1 measurement locations in the case study

(2.44). The surface degradation rate is 0.8 in the first half and becomes 0.95 afterwards.

$$\phi_t = \begin{cases} 0.8 & 1 \leq t \leq 5 \\ 0.95 & 6 \leq t \leq 10 \end{cases} \tag{2.44}$$

The spatiotemporal noise, η_t, is assumed to be constant and known in this case study. In practice, η_t can be estimated by selecting a proper kernel function for the covariance matrix and then estimating the kernel parameters. Normally in spatial statistics, exponential kernels are a common choice. The kernel function could also be selected from some candidate functions by conducing a cross-validation study using historical data.

$\mathbf{A}_0$ needs to be specified to start the Kalman recursion. Here, we assume that the diagonal elements of $\mathbf{A}_0$ are equal to 1, and all other elements are 0.05. In practice, $\mathbf{A}_0$ may

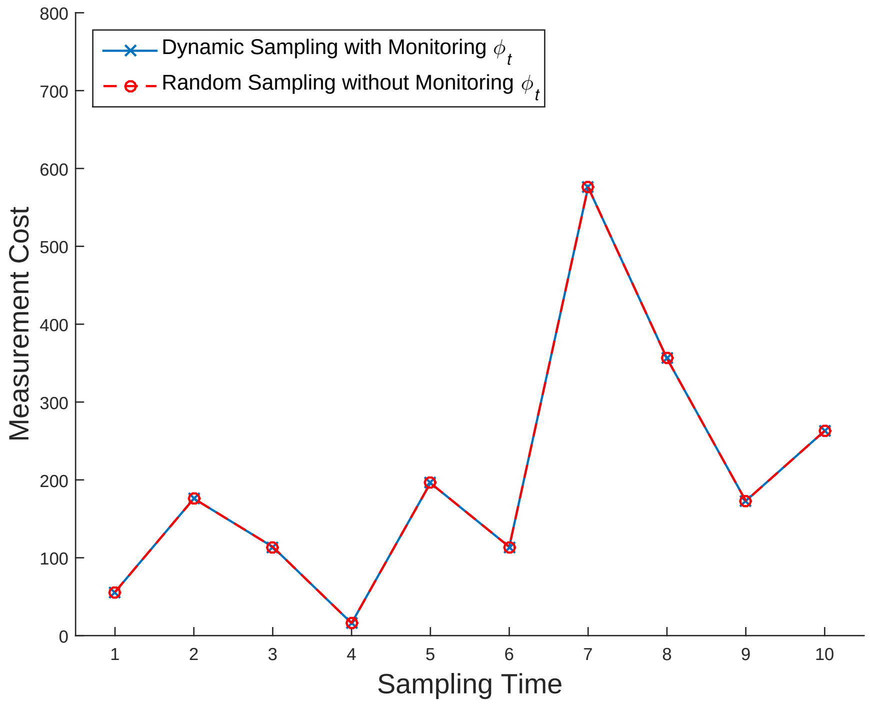


Figure 2.8 Measurement cost comparison of dynamic and random sampling approaches

be specified as the unconditional variance-covariance matrix of the spatiotemporal process (Wikle and Royle, 1999).

The proposed sampling approach is compared with a random sampling strategy. For the convenience of comparison, when conducting random sampling, we predetermine the number of Level 2 measurements as the same with our approach, and randomly allocate these points on the surface. Therefore, the two methods will always have the same measurement cost, and we will only need to compare the prediction precision, V_t. The comparison results are shown by Figures 2.8 and 2.9. Figure 2.8 indicates that two approaches have the same measurement costs for all sampling times. It is suggested by Figure 2.9 that the proposed dynamic sampling approach yields smaller V_t all the time, which demonstrates that the proposed method leads to better prediction performance than random sampling.

Furthermore, we compare the spatial distributions of the prediction variance from two

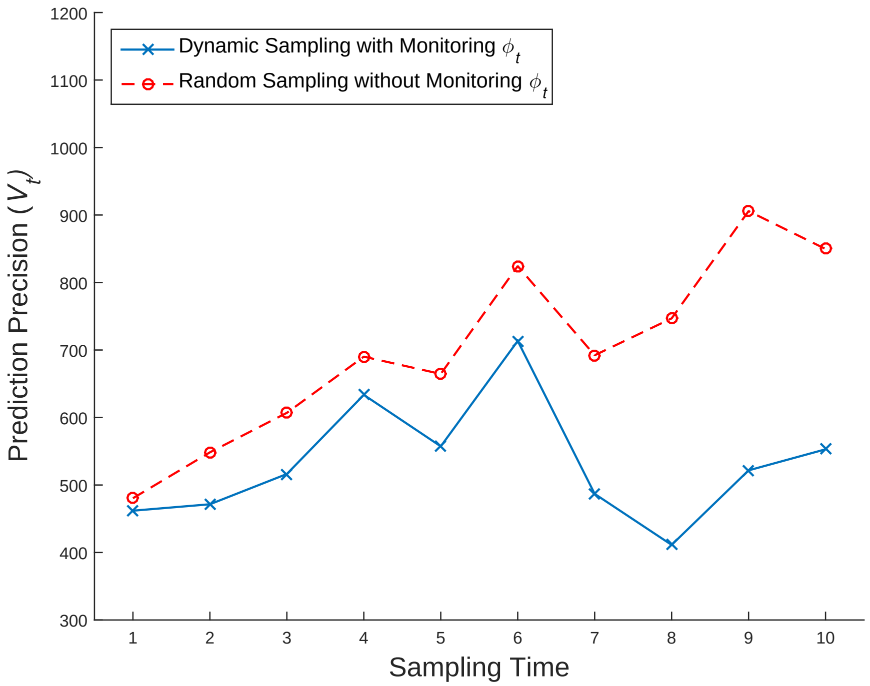


Figure 2.9 Precision comparison of dynamic and random sampling approaches

methods. The weight matrix is plotted as Figure 2.10. In Figure 2.10, each blue dot represents a prediction location, and the location indices are given by the numbers on the left and right sides of the map. It is observed that Locations 61–70 and 201–221 are marked by yellow color, indicating higher weight, and all other locations are shown by blue color. This is consistent with the weight assignment given by Equation (2.43).

The comparison of the overall distribution of Level 2 measurements over all 10 sampling is given by Figure 2.11. As indicated by Figure 2.11a, for the dynamic sampling approach, the highest measurement frequency on the surface are clustered around the locations with higher weights. Note that the rows beginning with Locations 49 and 65 belong to different grids, so the dense measurement locations expand to the rows above and below. On the other hand, rows beginning with Locations 193 and 209 belong to same grids, and therefore these grids connect to become one single band. Particularly, as Locations 201–

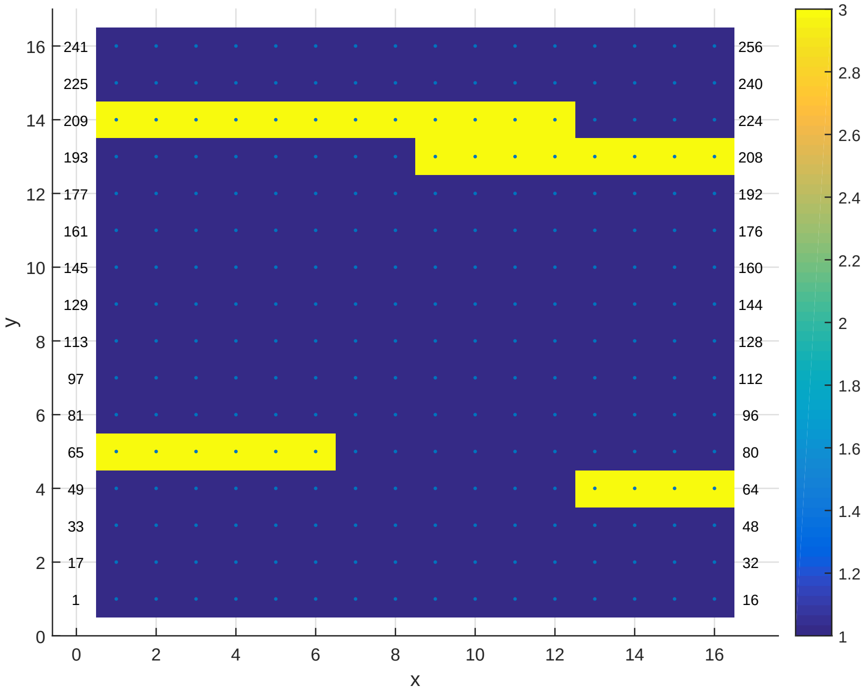


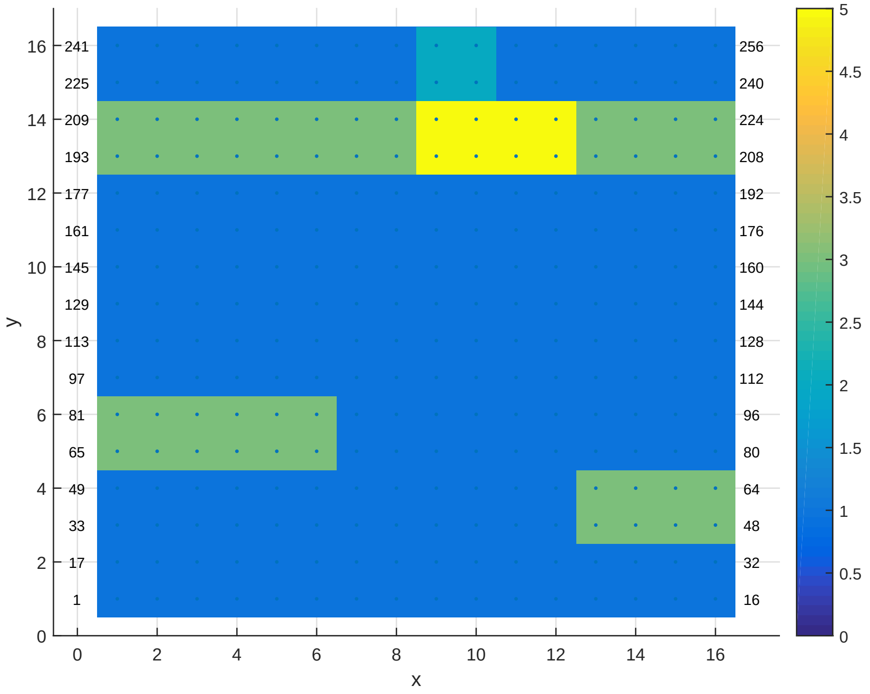
Figure 2.10 Map of the weight matrix

204 and 217–220 are all assigned with high weights, the corresponding two grids have highest measurement frequency, as indicated by yellow color. The overall measurement distribution when using random sampling presents a random pattern, as shown by Figure 2.11b.

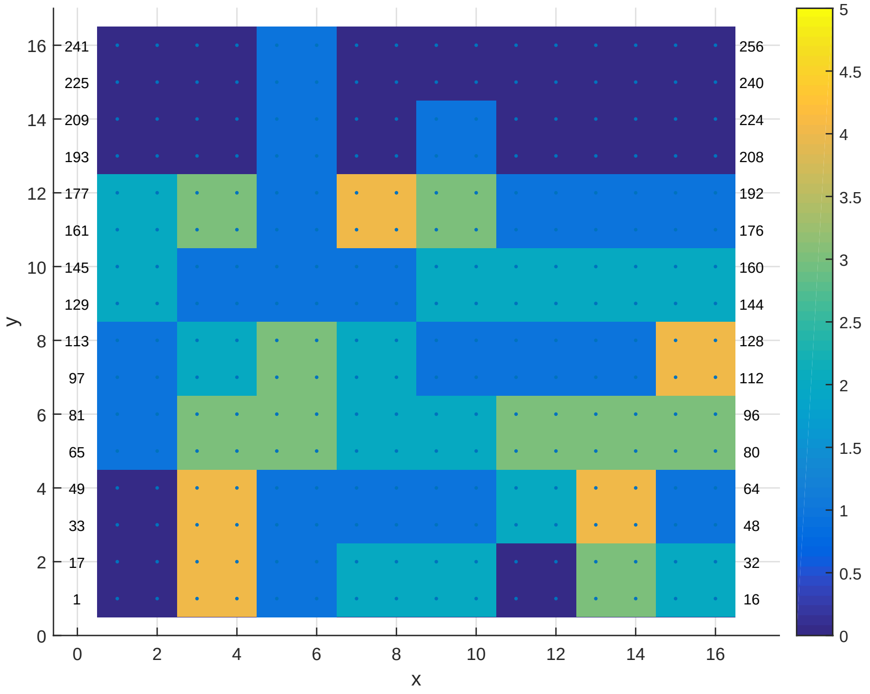
We further compare the prediction variance maps of dynamic and random sampling, and the results are shown by Figure 2.12. From Figure 2.12, we can make the following observations.

(1) In both Figures 2.12a and 2.12b, the locations given by Equation (2.42) have small prediction variance. This is due to the fact that Level 1 measurements are performed at these locations all the time.

(2) In Figure 2.12a, the prediction variance is smaller at locations with high weights. Specifically, Locations at 61–64 and 65–70 along with the locations below and



(a) Dynamic sampling



(b) Random sampling

Figure 2.11 Overall distribution of Level 2 measurements above them have small variance, and this is caused by both the large weights at these locations and that the locations below or above them belong to same Level 2 measurement grids. Since rows beginning from 193 and 209 belong to same Level 2 measurement grids, only one low variability "band" forms at these locations. Moreover, Locations 201–204 and 217–220 have even smaller variability, and it is a result of the fact that all these locations have large weights.

(3) The variance map of random sampling presents a random pattern, as shown by Figure 2.12b.

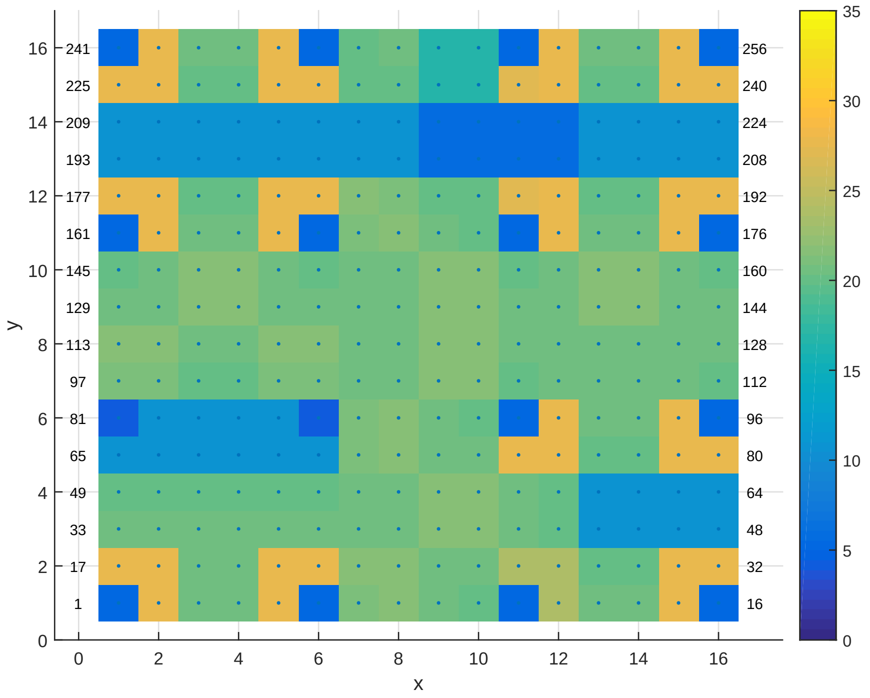
The results presented in this case study demonstrate that the proposed dynamic sampling strategy is able to allocate the measurement efforts according to the weight distribution and the spatiotemporal progression of the surface, thus leading to an improved prediction performance while balancing measurement costs.

2.5 Conclusion

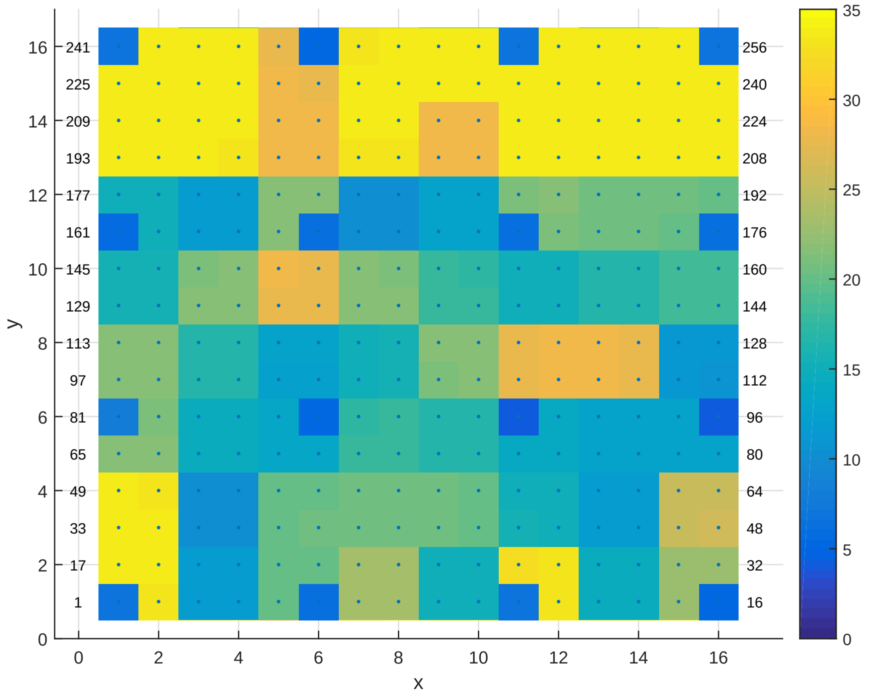
In this chapter, a framework for the dynamic sampling design of spatiotemporal processes in manufacturing is developed to achieve a balance between the prediction performance and the measurement cost. The contributions are summarized as follows.

(1) The design problem is formulated for characterizing spatiotemporal processes in manufacturing, where two precision levels of one measurement system or two gages with varied capabilities are applied in combination.

(2) A new design criterion is proposed to simultaneously consider prediction precision and measurement cost. Then the design search is formulated as a binary integer



(a) Dynamic sampling



(b) Random sampling

Figure 2.12 Comparison of dynamic and random sampling approaches using prediction variance maps programming problem, and GA is suggested to obtain the solution.

(3) A novel algorithm is developed to estimate and monitor the temporal transition parameter based on hypothesis testing using mutually measurement locations at two adjacent sampling times.

Future research may be focused on "spatiotemporal dynamic sampling design," where the design is dynamic in both the temporal and spatial domains. First, the requirement on the prediction precision may be time-variant. For example, in the early stage of tool life, normally less concern exists, but one has higher concern after the tool reaches a certain degradation level. Therefore, the weight matrix becomes time-variant, $\mathbf{W}_t$. Additionally, the covariance structure of the spatiotemporal noise, $\boldsymbol{\eta}_t$, may become more complicated as one tool degrades, e.g., stronger correlations and higher variability. This phenomenon reflects the increased complexity in the tool wear mechanism. Intuitively, one may suggest reducing the sampling time in the early tool stage, but increasing the sampling frequency during the late tool stage. However, in existing literature on dynamic design of spatiotemporal processes, the temporal sampling interval is mostly assumed to be constant. Therefore, a systematic method for spatiotemporal dynamic sampling design is highly desirable to achieve better prediction performance and lower measurement cost.

Chapter 3

Surface Shape Modeling by Integrating an Engineering Model with Multi-Task Gaussian Process Learning

3.1 Introduction

The control of surface shape variation on a machined surface is of great importance in automotive powertrain manufacturing because such variation has a significant influence on the sealing performance and causes distortion during surface assembly, such as those between engine heads and blocks or upper and lower transmission valve bodies. Characterization of the surface shape with high precision has become critical in controlling surface variations and to ensure the functional performance of surface assembly.

Various high-resolution surface measurement systems have been developed, including profilometers and laser holographic interferometers. However, the acquisition of high-resolution measurements using those systems is time-consuming and expensive. Moreover, such measurements may not be robust in response to environmental disturbance. For example, the measurement accuracy based on a laser holographic interferometer system can be affected by the heat dissipated from a machined part, measurement table vibration, and surface contamination, thus resulting in data loss and measurement errors, as shown in Figure 3.1.

Various approaches have been developed to make fine-resolution evaluation of the entire surface shape based on low or multi-resolution surface measurements. These ap-

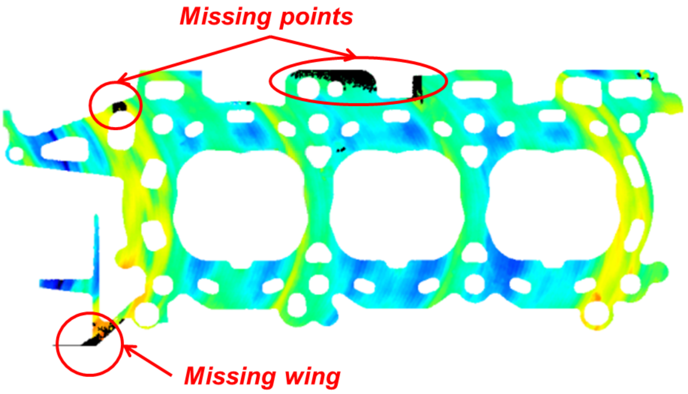
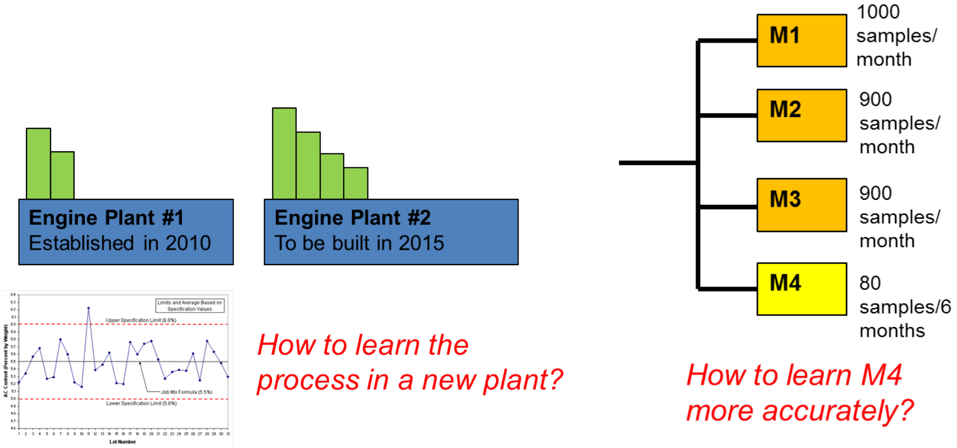


Figure 3.1 Data loss during high-resolution measurement

proaches include interpolating and extrapolating surface shapes from measured points through least squares (Zhu et al., 2004), B-spline methods (Yang and Menq, 1993; Grove et al., 2004; Jung and Kim, 2000) and grid fit through triangulation (D'Errico, 2005). In spatial statistics, the method is to estimate the surface shape by considering spatial correlations among sampled points. Such correlations reflect the spatial similarities between data in the vicinity on the part surface, and has been extensively utilized to interpolate and extrapolate surface data for form error estimation (Yang and Jackman, 2000; Xia et al., 2008) and geographic understanding in remote sensing applications (Curran and Williamson, 1986; Atkinson et al., 1992; Wang et al., 2005; Xiao et al., 2005). One limitation with these data-driven approaches is that the precision of the surface predictions is constrained by the density of measurements. To deal with the limitation, Suriano et al. (2015) proposed an engineering-driven surface model to improve surface prediction without having to increase surface measurements. The prediction improvement was achieved by fusing multi-resolution measurements with cutting force dynamics during surface manufacturing. In summary, state-of-the-art surface modeling techniques mostly focus on the interpolation/extrapolation of the data and process information from *one single manufacturing* *process*. Additionally, most of the data-driven methods lack engineering insights.

This chapter will explore new opportunities in improving surface shape modeling by considering the relationship between multiple manufacturing processes whenever available. The idea is motivated by a common manufacturing scenario when the process of interest lacks sufficient data whereas rich data could be available from other similar-but-not-identical processes. For instance, an automaker plans to build a new engine plant where there is a lack of historical data to establish a baseline for surface quality control. In the meantime, other engine plants have been running for some time and therefore the relevant historical surface data are extensively available. Though the machining processes in the existing plants were developed for different types of engines, the cutting mechanism for the same materials could be very similar and thus those process data can partially contribute to the prediction of the surface shapes in the new plant (Figure 3.2a). Similar scenario applies to a parallel machining station where multiple machines produce the same type of parts in parallel (Figure 3.2b). Surface data from the parallel machines can be potentially utilized to jointly learn the surface shape models for all machines.

These examples motivate us to improve the surface variation prediction by transferring the knowledge learned from data-rich processes to a similar process. Multi-task learning has emerged as a solution to dealing with such a knowledge-transfer problem (Caruana, 1997). It can be particularly useful in the situation when a limited amount of data is available in the target task while data from other related tasks are readily available. It is expected that by learning these tasks simultaneously, superior prediction performance can be achieved over the “no transfer” case (i.e., when each task is learned in isolation). Figure 3.3 summarizes the difference between multi-task learning and traditional single task

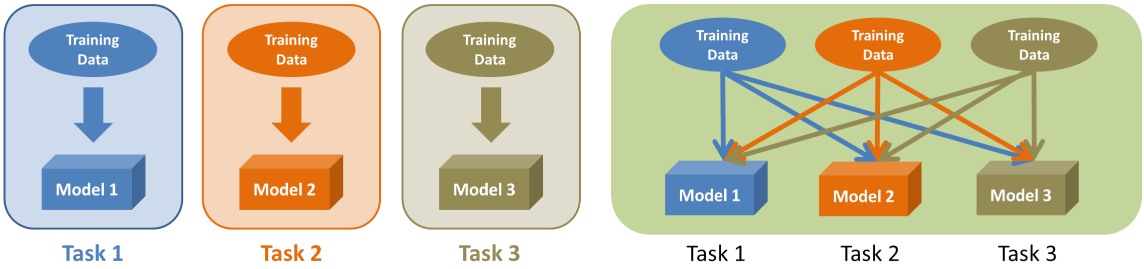


(a) New plant **(b)** Low production station

Figure 3.2 Two instances of the common scenario in manufacturing

learning. Multi-task learning has been employed to Gaussian process prediction, neural network, Dirichlet processes, and support vector machine (Caruana, 1997; Lawrence and Platt, 2004). The key problems of multi-task learning include the modeling of the similarity among multiple processes and the learning algorithms for model parameters. As for Gaussian processes, various similarity models such as block diagonal or not block diagonal covariance matrix and learning algorithms such as expectation-maximization algorithm are proposed (Bonilla et al., 2007; Yu et al., 2005; Schwaighofer et al., 2004). Therefore, it can be anticipated that through multi-task learning, the quality performance of the manufacturing process of interest or the new plant can be predicted by utilizing the observations from different but related manufacturing processes or plants. However, current multi-task learning for Gaussian processes is data-driven and does not incorporate the engineering insight into the model prediction.

This study aims to develop an engineering-guided multi-task learning (EG-MTL) surface model, which fuses engineering knowledge and surface measurement data from a



(a) Single task learning **(b)** Multi-task learning

Figure 3.3 Difference between single task learning and multi-task learning

number of similar-but-not-identical processes (tasks). We first establish a process physics guided spatial model consisting of a global trend that is induced by cutting force dynamics (engineering knowledge) and a local variation term that can be characterized by a zero-mean Gaussian process. The global trend captures the overall shape variation across the surface, and can be explained by the process physics to a great extent (Suriano et al., 2015). An iterative multi-task Gaussian process learning algorithm is then developed to learn the parameters from the global trend and the local variation. The algorithm will be demonstrated based on engine head deck face machining processes.

The remainder of this chapter is organized as follows. First, the surface model is presented in Section 3.2. Section 3.3 presents a case study to compare the proposed modeling approach with existing methods. Then Section 3.4 discusses the selection of hyperparameters, effects of the sample size and the number of tasks, and the effect of the correlation strength between the covariate and the primary variable. Finally, Section 3.5 concludes the chapter.

3.2 Surface Shape Modeling by Integrating Engineering Knowledge with Multi-Task Learning

In this section, the EG-MTL surface model which integrates engineering physics with multi-task learning is first presented, and then an iterative algorithm is developed to estimate the model parameters.

3.2.1 Engineering-Guided Multi-Task Learning Surface Model

Traditional multi-task learning approaches for Gaussian processes are data-driven, and do not incorporate engineering physics into the model. This study develops a new engineering-guided multi-task learning surface model that integrates engineering knowledge with multi-task learning, aiming to improve the surface prediction accuracy based on limited measurement data. Assume there are m similar-but-not-identical surfaces or surface manufacturing processes. For surface l, the surface model takes the form in Equation (3.1).

$$Z_l(\mathbf{s}) = \mathbf{U}_l(\mathbf{s})\boldsymbol{\beta}_l + \eta_l(\mathbf{s}), \tag{3.1}$$

where l is the surface index, and $l = 1, \ldots, m$; $\mathbf{s} = (x, y)$ is the coordinate vector; $Z_l(\mathbf{s})$ is the surface height at location $\mathbf{s}$ in surface l; $\mathbf{U}_l(\mathbf{s})$ is a vector of the location-dependent highly correlated process variables that are identified by expert knowledge; $\eta_l(\mathbf{s})$ is a zero-mean Gaussian process.

In Model (3.1), the surface variation is decomposed into a global trend that is induced by process settings and a local variation part that is modeled by a zero-mean Gaussian process. The global trend is modeled by a linear combination of multiple process variables. When there are both categorical and continuous process variables, a generalized linear model may be used instead (Nelder and Baker, 1972; Fahrmeir et al., 2013). The local variation is modeled as a zero-mean Gaussian process, and will be jointly learned across multiple similar-but-not-identical surfaces.

The rationale behind this decomposition is as follows. First, a linear model is employed to capture the relationship between the process inputs and the surface height in high-precision machining processes. It is reported that the surface height variation in the cut engine deck faces is strongly influenced by the cutting arc length, cutter path, cutting insert engagement, cutting feed rate, and clamping scheme (Suriano et al., 2015). The global trend is determined by machining process conditions. A summary of such factors is given in Table 3.1. More detailed explanation on these factors can be found in Suriano et al. (2015). Second, Gaussian process has been proven to be an effective method to model the local random surface variation that exhibits the spatial distribution similarity (Stein and Corsten, 1991; Atkinson et al., 1992; Suriano et al., 2015).

3.2.2 Iterative Algorithm for Parameter Estimation of the EG-MTL Model

The parameters in Model (3.1) that need to be estimated include: (1) the parameter vector in the global trend, β_l, and (2) the zero-mean Gaussian process parameters, $\eta_l(\mathbf{s})$, where $l = 1, \ldots, m$. The global trend describes the overall spatial variation pattern on a surface, and is mostly determined by the process settings. The local variation part shares similarity across multiple surfaces, and multi-task learning can potentially improve the estimation accuracy. Figure 3.4 illustrates the scheme for learning the parameters. As indicated by

Table 3.1 Summary of factors that impact surface height variation in the high-precision machining process (Suriano et al., 2015)

Engineering factor	Explanation	Example process variables $\mathbf{U}(\mathbf{s})$
Product/surface design	Design patterns of a surface (size, shape, and spatial distribution of holes and slots) or internal structure inside a part can cause cutting force change and local surface deformation	Arc length of a cutter engaged in cutting
Physical attributes of part materials	Due to manufacturing flaws from suppliers, workpiece materials may come with defects and heterogeneous physical attributes causing surface non-uniform deformation	Casting pore size, distribution density, and mass density (that can be acquired by industrial X-ray computed tomography scanning)
Manufacturing process conditions	Manufacturing process and machine tool-related parameters that can change cutting force and torque causing local surface variation	Feed rate, spindle setup tilt, spindle speed, depth of cut, cutter path-related MRR, clamping force
Multistage interdependence	A downstream manufacturing stage could change the surface shapes that have been created in certain upstream stages	Process variables or surfaces manufactured at different stages

Figure 3.4, three information sources are fused in the spatial model, namely, (1) surface measurement from the process of interest, (2) engineering knowledge on the process, and (3) measurement data from other similar-but-non-identical processes. A summary of the difference between the proposed model and existing approaches is given below.

- Traditional multi-task learning approaches only employ information from (1) and (3), and do not incorporate engineering knowledge;
- Simple kriging interpolation uses data from source (1) only;
- Some other kriging approaches, e.g., cokriging and kriging with external drift, jointly

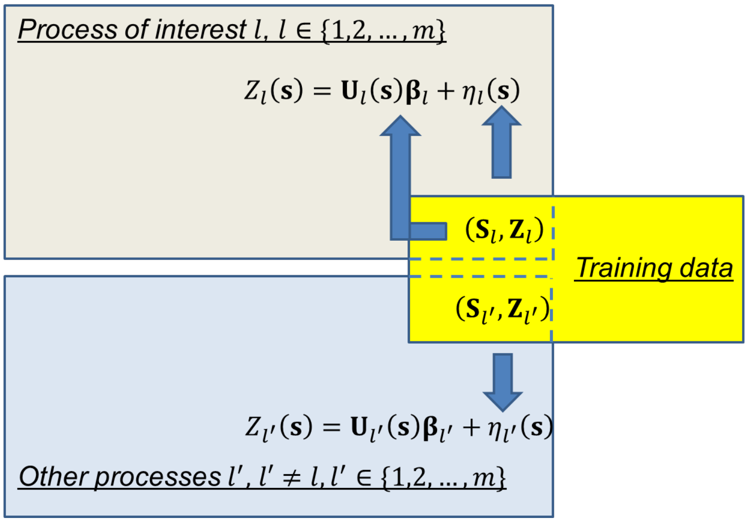


Figure 3.4 Illustration of the learning scheme

utilize information sources (1) and (2), but do not transfer knowledge from other relevant processes.

In the remainder of this section, we first review one data-driven multi-task learning algorithm for zero-mean Gaussian processes, and then present an iterative algorithm for estimating the parameters in Model (3.1).

3.2.2.1 Review of Multi-Task Learning

Several multi-task learning algorithms are available for Gaussian processes in the literature, such as Yu et al. (2005) and Bonilla et al. (2007). The performance of a certain algorithm may depend on a number of factors, e.g., the algorithm assumption, the true data distribution, and the selection of the spatial kernel function. Any multi-task Gaussian process learning algorithm can be incorporated into the proposed model. Here, without losing generality, this study adopts the approach developed by Yu et al. (2005). A brief review of the model is provided below.

The goal of multi-task learning for Gaussian processes is to estimate m related functions η_l, $l = 1, \ldots, m$, based on training data represented by Equation (3.2).

$$\mathbf{D}_l = \left(\mathbf{S}_l, \boldsymbol{\eta}_l\right), \tag{3.2}$$

where $\mathbf{S}_l \in \mathbb{R}^{n_l \times d}$ is a set of the inputs for function l, $\boldsymbol{\eta}_l \in \mathbb{R}^{n_l}$ is a set of the function values at these inputs, and n_l is the size of training data for task l. In surface modeling, the location is generally determined by two coordinates, i.e., x and y, and therefore $d = 2$. It is further assumed that there are in total n distinct data points in $\{\mathbf{D}_l\}$ with $\min\left(\{n_l\}\right) \leq n \leq \sum_{l=1}^{m} n_l$.

The commonality among different tasks is defined via the following inductive model (Yu et al., 2005).

Model 3.1 *Let $\boldsymbol{\eta}^l$ be the values of η_l on a set $\mathbf{S}$, satisfying $\cup \mathbf{S}_l \subset \mathbf{S}$. Given the hyperprior distribution of $\boldsymbol{\mu}_\alpha$ and $\mathbf{C}_\alpha$ described by a Normal-inverse-Wishart distribution as shown by Equation (3.3), the Gaussian process for task l is generated by the following three steps.*

$$p\left(\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha\right) = \mathcal{N}\left(\boldsymbol{\mu}_\alpha \middle| 0, \frac{1}{\pi}\mathbf{C}_\alpha\right) \mathcal{IW}\left(\mathbf{C}_\alpha \middle| \tau, \boldsymbol{\kappa}^{-1}\right). \tag{3.3}$$

1. *$\boldsymbol{\mu}_\alpha$, $\mathbf{C}_\alpha$ are generated once using Equation (3.3);*

2. *For each function η_l,*

$$\boldsymbol{\alpha}^l \sim \mathcal{N}\left(\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha\right). \tag{3.4}$$

3. *Given $\mathbf{s} \in \mathbf{S}_l$,*

$$\eta_l(\mathbf{s}) = \sum_{i=1}^{n} \alpha_i^l \kappa(\mathbf{s}_i, \mathbf{s}) + \varepsilon, \tag{3.5}$$

where $\kappa(\cdot,\cdot)$ *is the base kernel function;* $\mathbf{s} \in \mathbf{S}$*; and* $\varepsilon \sim \mathcal{N}(0,\sigma^2)$.

The estimates of $\Theta = \left\{\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha, \sigma^2\right\}$ can be obtained using an expectation-maximization (EM) algorithm (Yu et al., 2005), and the details of the EM procedure are presented in Appendix B. After obtaining Θ and $\{\hat{\alpha}^l\}$ with the EM algorithm, the function value corresponding to an unknown input $\mathbf{s}_u$ from task l is estimated using the following equation.

$$\hat{\eta}_l(\mathbf{s}_u) = \sum_{i=1}^{n} \hat{\alpha}_i^l \kappa(\mathbf{s}_i, \mathbf{s}_u), \tag{3.6}$$

where the prediction is calculated as a linear combination of $\kappa(\mathbf{s}_i, \mathbf{s}_u)$, and $\mathbf{s}_i \in \mathbf{S}$ includes both the training data from task l and these from other tasks. The kernel function $\kappa(\cdot,\cdot)$ is a measure of the “distance” or similarity between two inputs.

One critical limitation of this approach is that the surface means must be accurately estimated and removed prior to conducting multi-task learning, because it is based on the assumption that the Gaussian processes have zero means. However, estimation of the global means with limited training data is often not sufficiently accurate. The prediction results can be misleading when the zero-mean assumption does not hold. This study utilizes a linear model to characterize the global trend with expert process knowledge. The estimation is expected to be more accurate, and the zero-mean assumption for residual local variation will be valid.

3.2.2.2 Iterative Multi-Task Learning Algorithm

There are two types of model parameters in Model (3.1), i.e., coefficients in the linear model, β_l, and Gaussian process parameters. Simultaneously estimating both types of pa-

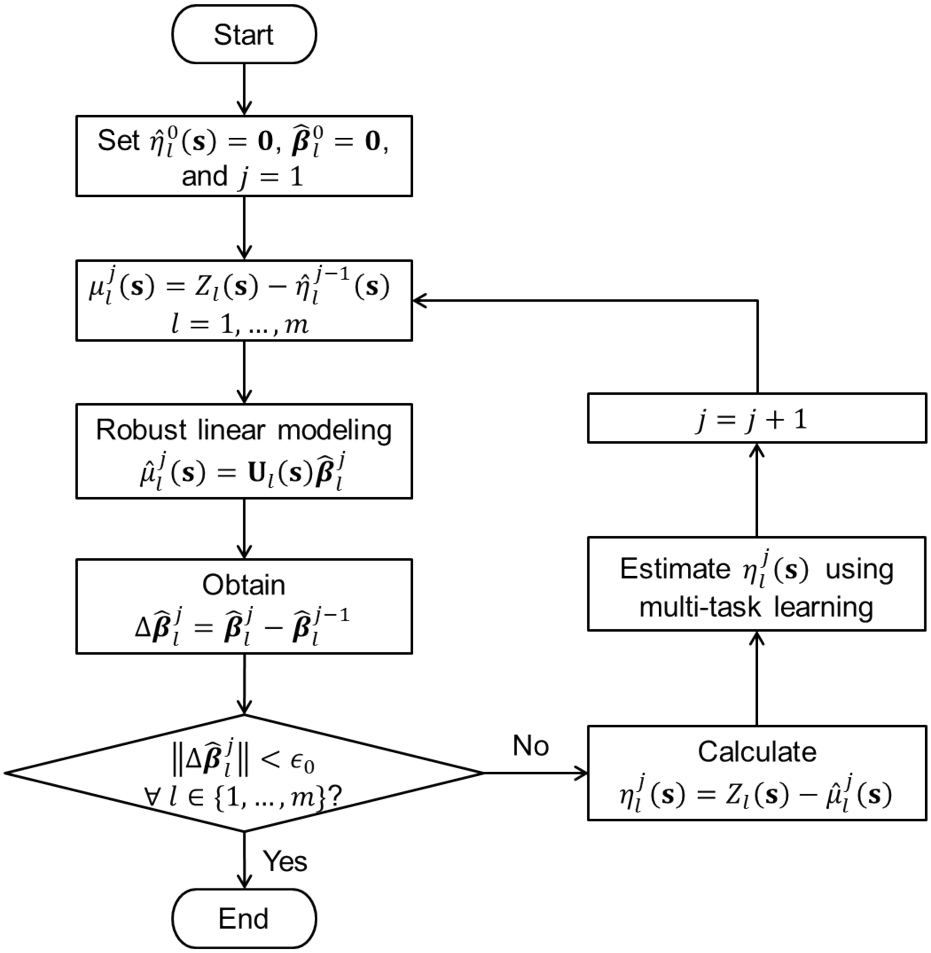


Figure 3.5 Flowchart for the iterative multi-task learning algorithm

rameters is challenging as the changes in one part directly impacts the estimation of the other. To address this issue, an iterative algorithm is developed for the model estimation. The algorithm is illustrated by Figure 3.5. For m similar-but-not-identical surfaces, the global trend and the Gaussian processes are estimated in an iterative manner, and this process is terminated when the coefficients in the global trend converge. In this procedure, the superscript j ($j = 0, 1, \dots$) specifies the iteration index, and the subscript l ($l = 1, \dots, m$) indicates the task/part/surface number. The procedure is further explained as follows.

As an initialization action, the Gaussian process part and the coefficient vector in the global trend are both set to be zero for all surfaces, as shown by Equation (3.7).

$$\begin{aligned} \hat{\eta}_l^0(\mathbf{s}) &= \mathbf{0}, \\ \hat{\beta}_l^0 &= \mathbf{0}, \end{aligned} \tag{3.7}$$

where $l = 1, \ldots, m$.

In the jth iteration, the Gaussian process part estimated from $(j-1)$th iteration, $\hat{\eta}_l^{j-1}(\mathbf{s})$, is first removed from the surface height for all surfaces, as shown by Equation (3.8). Then robust linear regression is applied to estimate the coefficient vector $\hat{\beta}^l$ for $l = 1, \ldots, m$. The model of the global trend for surface l at the jth iteration is given by Equation (3.9).

$$\hat{\mu}_l^j(\mathbf{s}) = Z_l(\mathbf{s}) - \hat{\eta}_l^{j-1}(\mathbf{s}), \tag{3.8}$$

where $Z_l(\mathbf{s})$ is the height of the lth surface; $\hat{\eta}_l^{j-1}(\mathbf{s})$ is the estimated Gaussian process for lth surface at the $(j-1)$th iteration; and $l = 1, \ldots, m$.

$$\hat{\mu}_l^j(\mathbf{s}) = \mathbf{U}_l(\mathbf{s})\hat{\beta}_l^j, \tag{3.9}$$

where $l = 1, \ldots, m$.

There are mainly three categories of robust regression methods for linear models: (1) least squares alternatives (Holland and Welsch, 1977), (2) parametric alternatives (Lange et al., 1989), and (3) unit weights (Wainer and Thissen, 1976). This study adopts the method presented in Holland and Welsch (1977), which applies iteratively reweighted least squares for robust regression. This method has been implemented in some popular computing softwares, such as R and Matlab. Maronna et al. (2006) provides a comprehensive overview of the popular robust regression techniques.

After the robust linear regression, the convergence test is conducted to check whether the coefficients of the linear models converge. The change of the coefficient vector is defined by Equation (3.10). The convergence criterion is given by Equation (3.11). The convergence is considered to be achieved if the changes of the coefficient vectors for all surfaces are within a predetermined threshold.

$$\Delta\hat{\boldsymbol{\beta}}_l^j = \hat{\boldsymbol{\beta}}_l^j - \hat{\boldsymbol{\beta}}_l^{j-1}, \tag{3.10}$$

where $\hat{\boldsymbol{\beta}}_l^j$ and $\hat{\boldsymbol{\beta}}_l^{j-1}$ are the coefficient vectors from iterations j and $j-1$, respectively; and $l = 1,\ldots,m$.

$$\left\|\Delta\hat{\boldsymbol{\beta}}_l^j\right\| < \varepsilon_0, \quad \forall l \in \{l,\ldots,m\}, \tag{3.11}$$

where ε_0 is a predetermined threshold and can be tuned based on applications and the accuracy requirement.

If convergence is achieved at the jth iteration, the iterative process is terminated and the parameter estimation is finished. Meanwhile, the prediction results at the $(j-1)$th iteration are the final results of the algorithm. If convergence is not achieved, multi-task Gaussian process learning (MTGPL) is then conducted to estimate $\eta(\mathbf{s})$'s, and another iteration is performed.

The detailed implementation of multi-task learning for Gaussian processes is illustrated by Figure 3.6. Specifically, the first step is to assign the hyperprior parameters, including

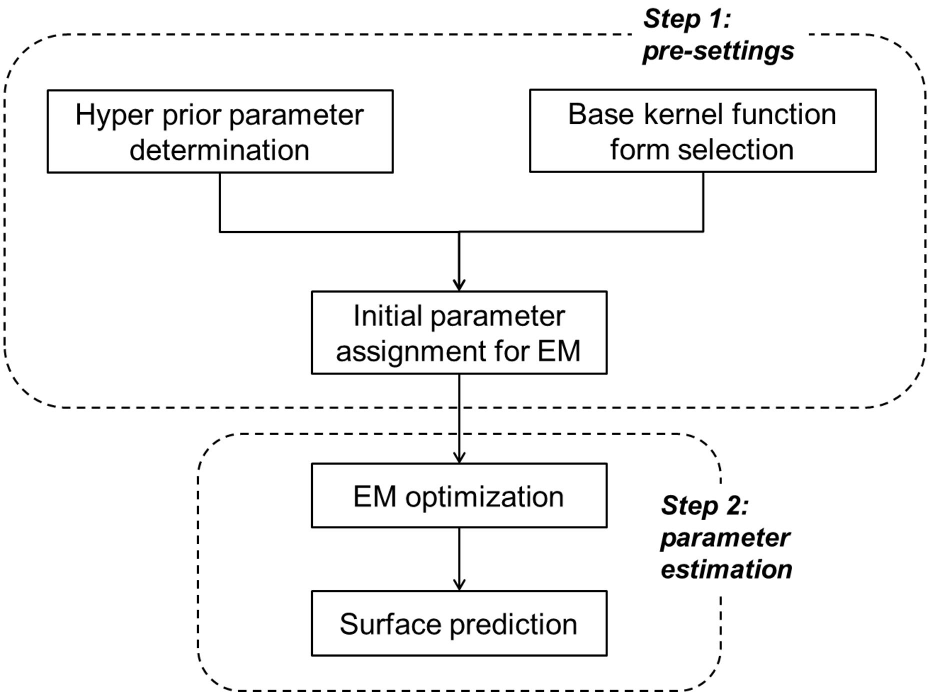


Figure 3.6 Implementation procedure of multi-task learning

τ and π in the Normal-inverse-Wishart distribution as described by Equation (3.3) and to determine the base kernel function form $\kappa(\cdot,\cdot)$. A common practice is to apply cross-validation to select the best settings from a set of candidate parameters and function forms (Yu et al., 2005; Huang et al., 2012). The second step is to estimate the model and perform predictions, and this step follows the procedure reviewed in Section 3.2.2.1.

It should be noted that inappropriate initial settings may lead to undesirable results, e.g., slow convergence or non-convergence and unsatisfactory prediction accuracy. Several practical suggestions are provided as follows to improve the EM efficiency and guarantee the acquisition of the global maximum.

- Random restart (starting with different randomly selected initial parameter values);
- Perform Simulated annealing (Kirkpatrick et al., 1983; Brooks and Morgan, 1995);
- Monitoring the log-likelihood trace, and changing the initial values when (1) fluctu-

ation occurs or (2) convergence is not achieved within a certain number of iterations.

3.3 Case Study

This section compares the proposed engineering-guided multi-task learning surface model (denoted as "EG-MTL model" afterwards) with some representative interpolation methods which have been applied popularly, including the multi-task learning for Gaussian processes (referred to as GPMTL), simple kriging, and kriging with external drift (KED). A summary of these methods is given in Table 3.2.

Table 3.2 Method summary for the case study

Method	Using engineering knowledge?	Transferring knowledge from other processes?
EG-MTL model	Yes	Yes
GPMTL	No	Yes
Simple kriging	No	No
KED	Yes	No

GPMTL applies the algorithm described in Section 3.2.2.1. As this algorithm is applicable to zero-mean Gaussian processes (Yu et al., 2005), the global mean is first estimated using the sample mean from the training data and then multi-task learning is performed. For surface l, $l = 1, \ldots, m$, the global mean is estimated using Equation (3.12).

$$\hat{\mu}_{l,0} = \overline{Z}(\mathbf{s}_{l,o}) = \frac{\sum\limits_{\mathbf{s} \in \mathbf{S}_l} Z(\mathbf{s})}{n_l}, \tag{3.12}$$

where $\mathbf{s}_{l,o}$ represents the observed locations in surface l.

KED is a generalized case of kriging where the overall trend is modeled as a function of some auxiliary predictors (Hudson and Wackernagel, 1994; Bourennane et al., 2000), which are the process variables in the machining example. Linear models may be applied to capture the linear correlations between the target variable and the auxiliary predictors (Bourennane et al., 2000). This interpolation technique incorporates the engineering knowledge but do not employ the similarities between the process of interest and other relevant processes.

A laser holographic interferometer was adopted at an engine manufacturing plant to measure engine block deck surfaces, and the obtained surface data from this plant will be used throughout the case study. Root mean squared error (RMSE) is be chosen as a metric of the prediction performance, and its definition is given by Equation (3.13). Another metric of interest is the improvement percentage of RMSE from the simple kriging method to the EG-MTL model or MTGPL, which is defined by Equation (3.14).

$$\mathrm{RMSE}(l) = \sqrt{\frac{\sum\limits_{\mathbf{s}\in\mathbf{S}} \left(\hat{Z}_l(\mathbf{s}) - Z_l(\mathbf{s})\right)^2}{|\mathcal{G}_l|}}, \tag{3.13}$$

where $l = 1, \ldots, m$ is the task/surface number, $\mathcal{G}_l$ is a set of locations where predictions are made; $|\cdot|$ is an operator to calculate the set size, and therefore $|\mathcal{G}_l|$ is test sample size for surface l.

$$\mathrm{RSME}_{\mathrm{I,kriging}} = \frac{\mathrm{RMSE}_{\mathrm{kriging}} - \mathrm{RMSE}_{\mathrm{engineering}}}{\mathrm{RMSE}_{\mathrm{kriging}}} \times 100\%, \tag{3.14}$$

where $\mathrm{RMSE}_{\mathrm{kriging}}$ and $\mathrm{RMSE}_{\mathrm{engineering}}$ are the RMSEs of simple kriging and the proposed EG-MTL model, respectively.

Table 3.3 Summary of part information

Part number (l)	1	2	3
Machine ID	2N	2N	3N
Training sample size (n_l)	50	50	50
Test sample size ($\lvert\mathscr{G}_l\rvert$)	32563	32790	32673
Correlation strength ($\rho(U,Z)$)	0.84	0.78	0.78

Similarly, a metric is defined to quantify the performance improvement from KED method to the EG-MTL model, as given by Equation (3.15).

$$\mathrm{RSME}_{\mathrm{I,KED}} = \frac{\mathrm{RMSE}_{\mathrm{KED}} - \mathrm{RMSE}_{\mathrm{engineering}}}{\mathrm{RMSE}_{\mathrm{engineering}}} \times 100\%, \tag{3.15}$$

where $\mathrm{RMSE}_{\mathrm{KED}}$ and $\mathrm{RMSE}_{\mathrm{engineering}}$ are the prediction RMSEs of KED and the EG-MTL model, respectively.

The measurement data of three engine head surfaces are utilized for the comparative case study. The product part information is summarized by Table 3.3. The parts were machined at different machines in a machining station: Parts 1 and 2 were machined at Machine 2N, and Part 3 was from Machine 3N. The test data of each part consists of approximately 32600 points, and 50 points are randomly sampled as training data. Figure 3.7 shows an engine head surface example, where red dots represent sampling locations. For each part, a covariate, U, is simulated to have a correlation of around 0.80 with the surface height, Z. Examples of U are listed and explained in Table 3.1.

The prediction results are shown by Figure 3.8. The average RMSEs of the EG-MTL model, GPMTL, simple kriging, and KED over three parts are 3.30, 4.31, 4.71, and 4.14,

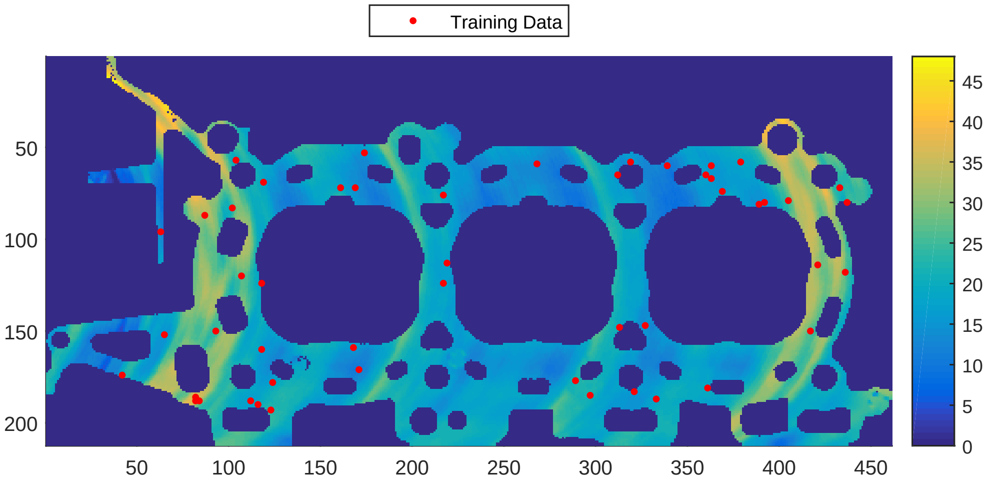


Figure 3.7 An engine head surface example

respectively. Figure 3.8 indicates that for all three parts, the EG-MTL model performs the best, and the simple kriging obtains the worst prediction accuracy. The KED model yields smaller RMSEs than MTGPL and simple kriging, and this implies that the incorporation of engineering knowledge is beneficial. Hence, by incorporating a highly correlated covariate, which can be identified using engineering understanding of the process physics, the EG-MTL model is able to significantly improve the prediction accuracy over existing interpolation techniques. When no such knowledge is available, the data-driven model reviewed in Section 3.2.2.1 is suggested, because GPMTL outperforms simple kriging, as shown by Figure 3.8.

3.4 Discussion

This section discusses the following three issues: (1) hyperparameter selection, (2) effects of the sample size and the number of tasks, and (3) the effect of the correlation strength between the covariate and the surface height.

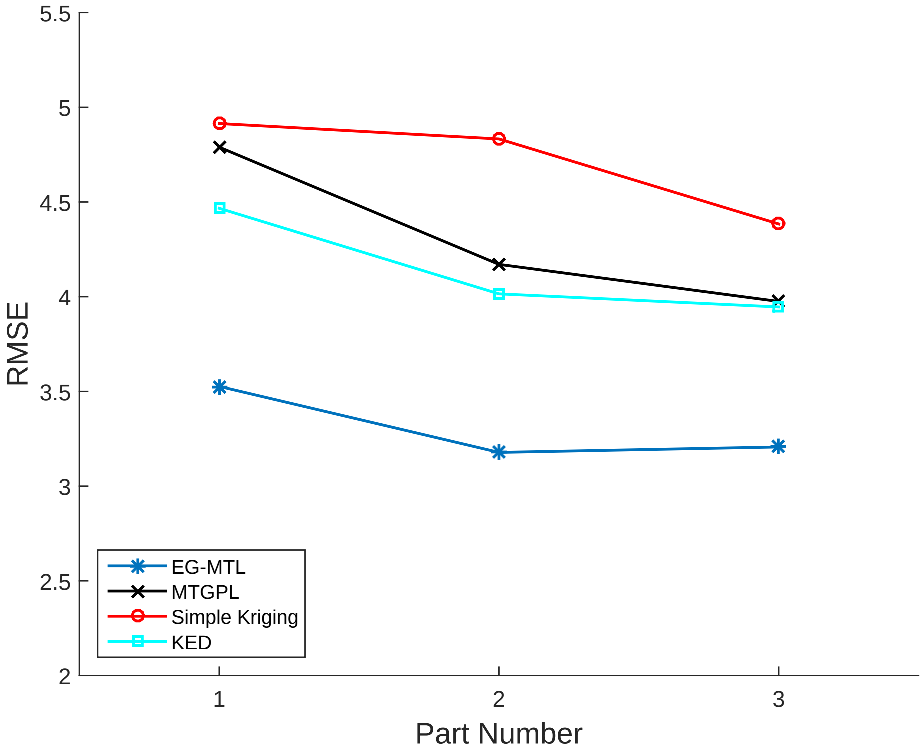


Figure 3.8 RMSE comparison for the EG-MTL model, GPMTL, simple kriging, and KED

3.4.1 Hyperparameter Selection

It was found by a preliminary study that a quadratic exponential kernel function performed well in modeling surface height, and we will use this kernel form in this study. Actually, quadratic exponential functions are commonly used as kernels in spatial statistics (Ripley, 2005). The general form is given by Equation (3.16).

$$\kappa\left(\mathbf{s},\mathbf{s}'\right) = \exp\left(\frac{-\left\|\mathbf{s}-\mathbf{s}'\right\|^2}{\delta^2}\right), \tag{3.16}$$

where δ^2 is an unknown parameter and should be estimated in practice.

Three hyperprior parameters need to be determined prior to the modeling, i.e., τ and π from the Normal-inverse-Wishart distribution, as well as δ^2 from the base kernel function.

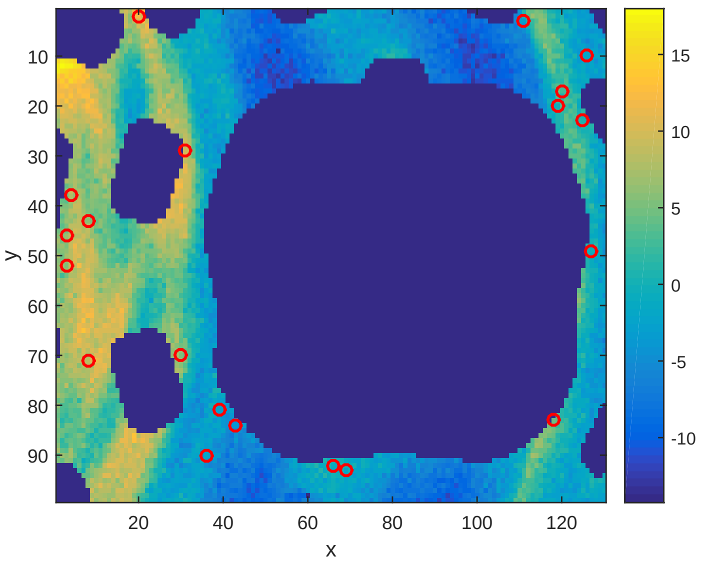


Figure 3.9 Engine surface example for the hyperparameter study

Table 3.4 Experimental setting for the hyperprior parameter study

Number of tasks (m)	5
Training sample size (n_l)	20
Test sample size ($\|\mathscr{G}_l\|$)	approximately 11000
Candidate hyperparameter values	$\{0.0001, 0.001, 0.01, 0.1, 0.5, 1, 6, 11, \ldots, 151\}$
Total number of experimental runs	46656

An experiment is conducted to investigate the effects of the hyperparameters using a partial engine surface. Figure 3.9 displays an example of the surface, where red circles indicate sampling/measurement locations.

A summary of the experimental setting is given as Table 3.4. In this study, the number of tasks and the training sample size are fixed as 5 and 20, respectively; the test sample size is approximately 11000; the candidate parameter values vary from 0.0001 to 151, and all possible combinations are tested.

By testing all possible parameter combinations, it is found that the average RMSE for simple kriging is 6.91, and the lowest prediction error using multi-task learning is 2.87, with corresponding parameters as $\left[\tau, \pi, \delta^2\right] = [0.0001, 1, 111]$. Moreover, Figure 3.10 shows the 1D RMSE trends vs. τ, π, and δ^2. Figure 3.11 displays the RMSE trends when simultaneously changing the pairs of $[\tau, \pi]$, $\left[\tau, \delta^2\right]$, and $\left[\pi, \delta^2\right]$.

Figure 3.10a shows that as the increase of τ, the RMSE first decreases sharply and then increases slowly, meaning that the prediction performance first improves significantly and then degrades. A similar pattern is observed for π, as displayed by Figure 3.10b. From Figure 3.10c, it is noted that when increasing δ^2, the RMSE first decreases, and then increases slowly, meaning that the prediction performance does not change significantly when δ^2 is at a relatively high level, e.g., $\delta^2 > 50$. These patterns can be reconfirmed by the 3D RMSE plots which are shown by Figures 3.11a to 3.11c. It is concluded that

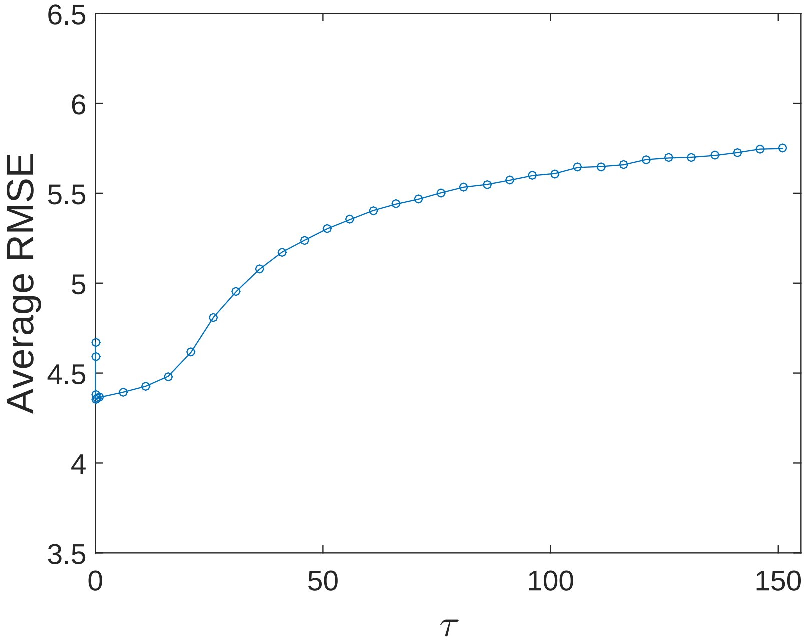
(1) In general, smaller τ and π along with larger δ^2 yield better prediction performance.

(2) In practice, it is suggested to choose τ and π to be around 1, and δ^2 to be between 50 and 100.

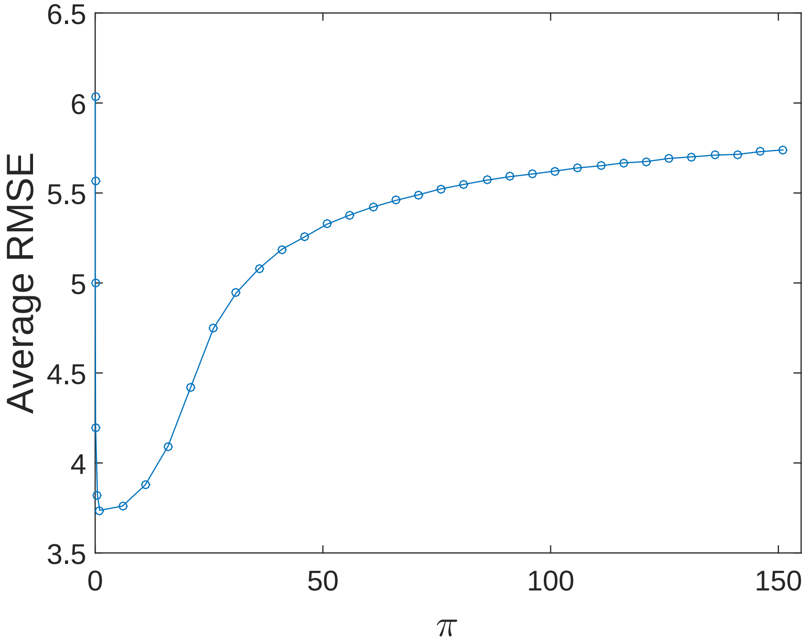
The selection of optimal hyperparameter is mostly determined by the spatial variation pattern. When implementing the proposed modeling approach in practice, a cross-validation study may be conducted using the training data to find appropriate values.

3.4.2 Effects of the Sample Size and the Number of Tasks

The learning performance of multi-task learning can be sensitive to the sample size and the number of tasks; therefore, it is important to analyze the effects of these factors in practice.

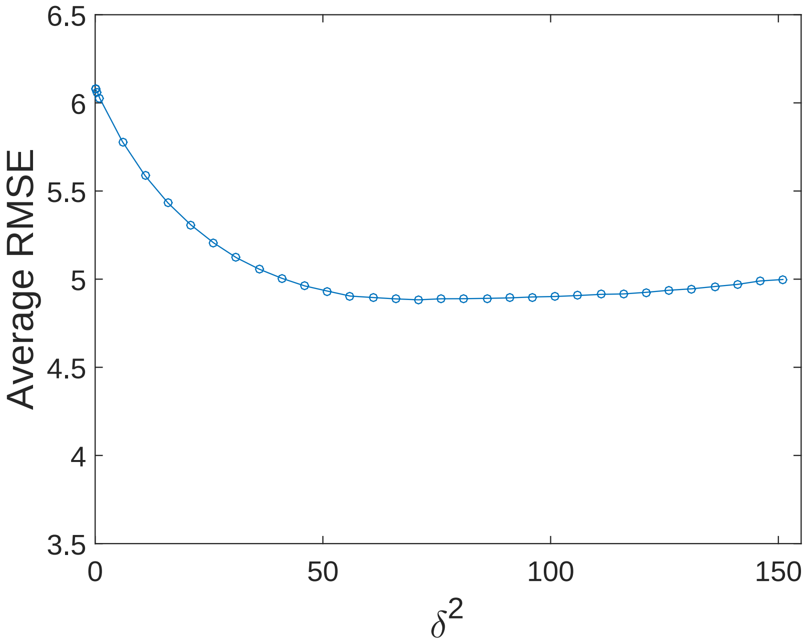


(a) RMSE vs. τ



(b) RMSE vs. π

Figure 3.10 Effects of hyperparameters on RMSE (cont.)



(c) RMSE vs. δ^2

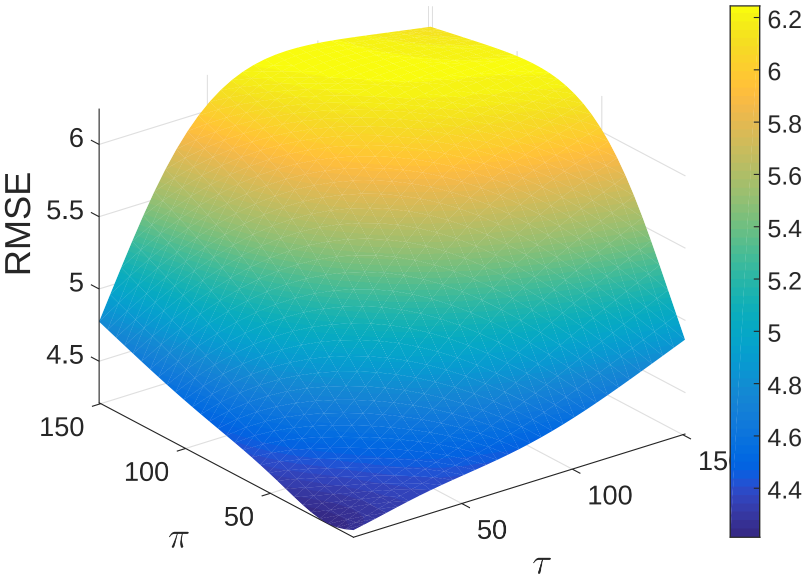
Figure 3.10 Effects of hyperparameters on RMSE

Table 3.5 Summary of experimental settings for the effects of the sample size and the number of tasks

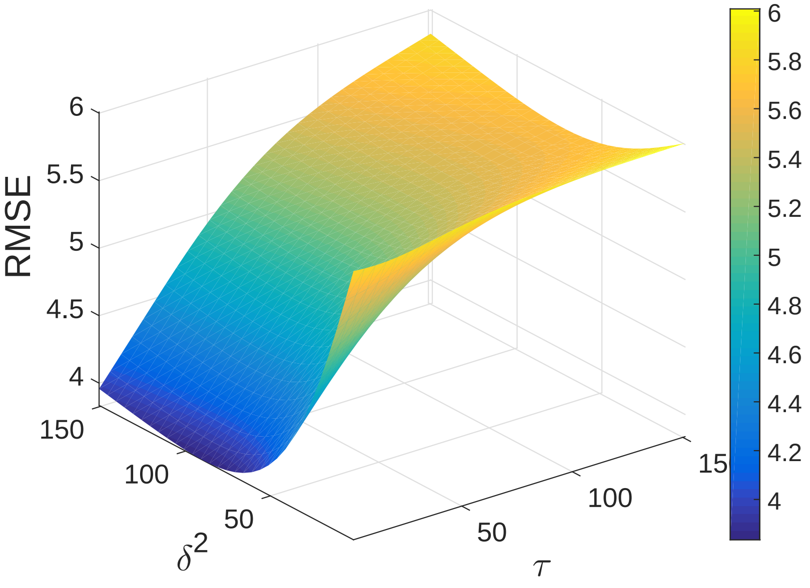
Experiment	1	2	3
Number of tasks (m)	5	$\{3,4,\ldots,20\}$	$\{3,4,\ldots,10\}$
Training sample size (n_l)	$\{10,50,100,\ldots,300\}$	50	$\{10,20,\ldots,150\}$
Test sample size ($\lvert\mathscr{G}_l\rvert$)	≈32600	≈ 32600	≈ 32600
Number of repeated runs	10	10	1

In this section, three experiments are conducted to investigate the effects of the sample size and the number of tasks. Table 3.5 summarizes the experimental settings.

In the first experiment, the number of tasks is fixed at 5, and the sample size varies from 10 to 300. 10 repeated runs are performed for every sample size, and the sampling locations are randomly selected for each run. The experiment results are illustrated using Figure 3.12, where box plots are used to show the range of 10 runs. It is seen that as the

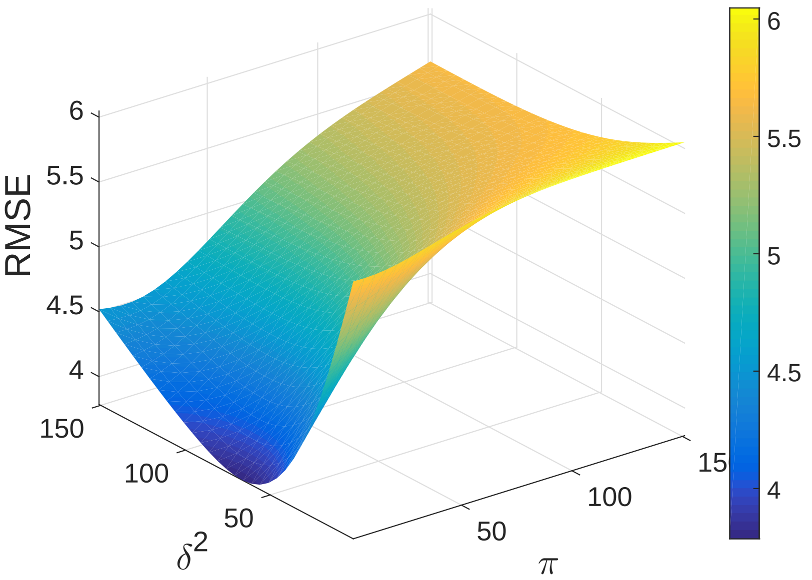


(a) RMSE vs. (τ, π)



(b) RMSE vs. (τ, δ^2)

Figure 3.11 Effects of hyperparameter pairs on RMSE (cont.)



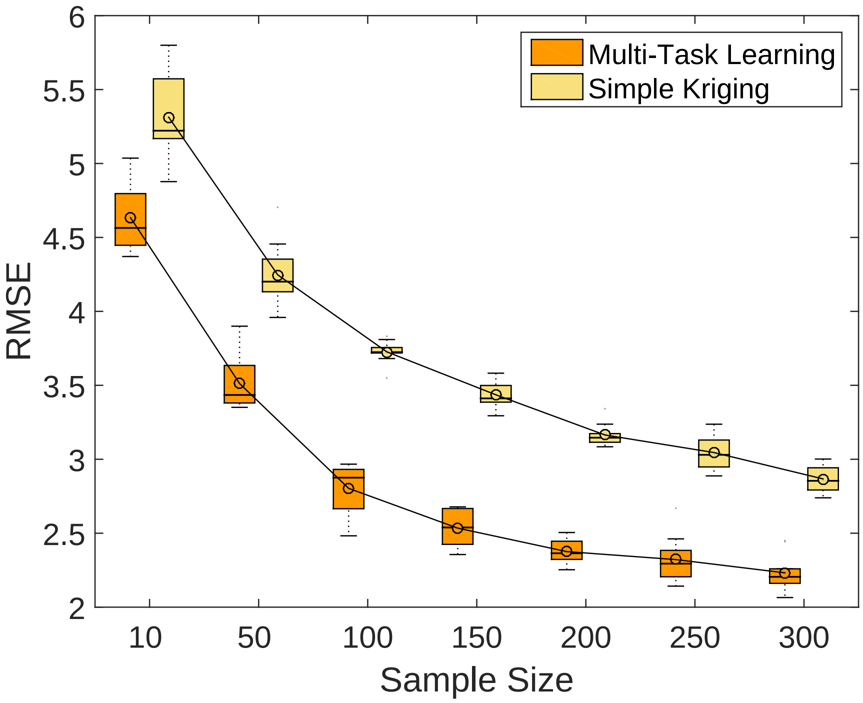
(c) RMSE vs. (π,δ^2)

Figure 3.11 Effects of hyperparameter pairs on RMSE

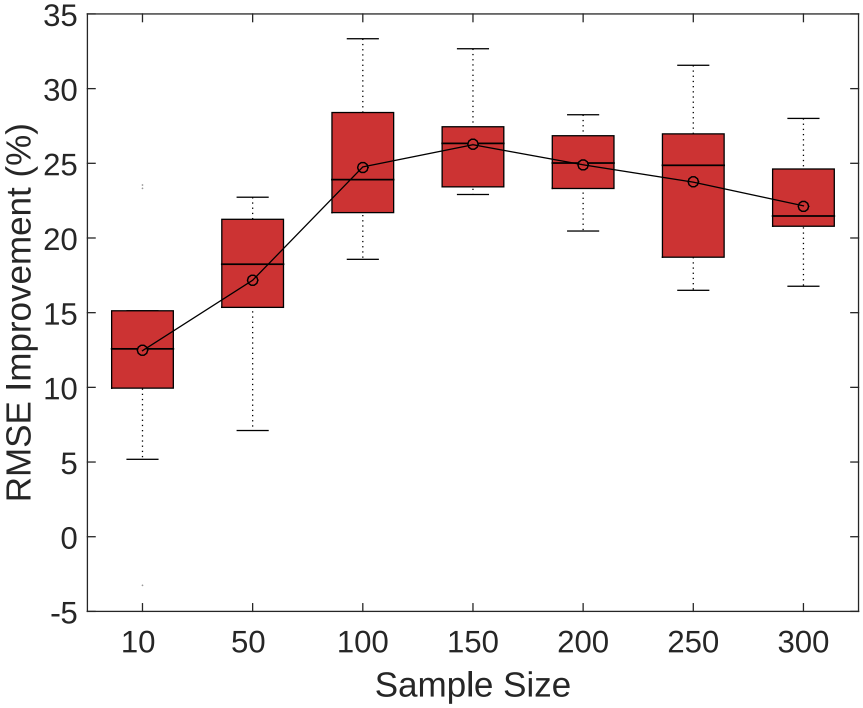
sample size increases, the prediction performance of both multi-task learning and kriging improve, and $\mathrm{RSME_I}$ first increases, then keeps stable, and shows a sign of decreasing in the end.

The second experiment fixes the sample size as 50 for all parts, and varies the number of tasks from 3 to 10. The results are shown by Figure 3.13. As the increase of the number of tasks, $\mathrm{RSME_I}$ first increases quickly, and then keeps stable after the number of tasks reaches 17.

In the third experiment, both the number of tasks and the sample size are changed. The number of tasks takes the values of $\{3,4,\ldots,10\}$, and the sample size varies as $\{10,20,\ldots,150\}$. Figure 3.14a displays the 3D trend of RMSE when varying the sample size and the number of tasks. Clearly, the RMSE decreases with greater sample size and number of tasks. Figure 3.14b shows the 3D plot of $\mathrm{RSME_I}$, and it is indicated that multi-task learning outperforms simple kriging in all cases, and $\mathrm{RSME_I}$ increases as the

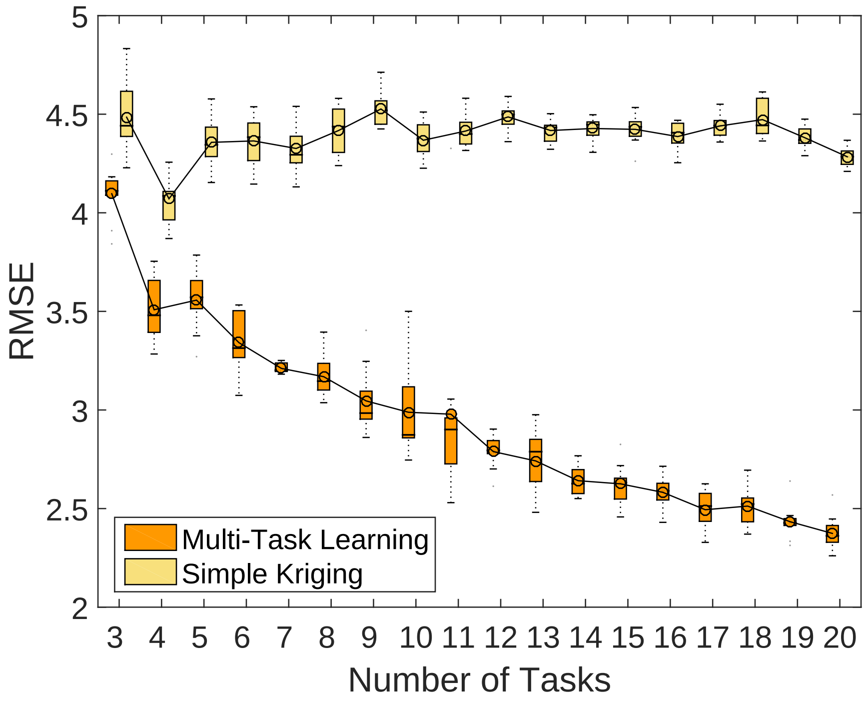


(a) RMSE

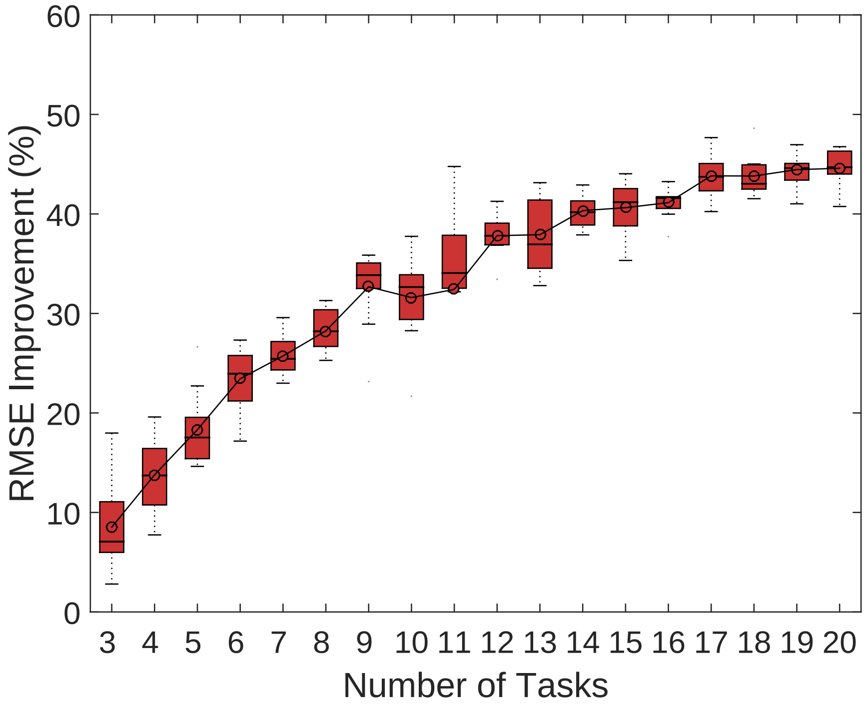


(b) RMSE improvement percent, $RSME_I$

Figure 3.12 Effects of sample size on the prediction performance

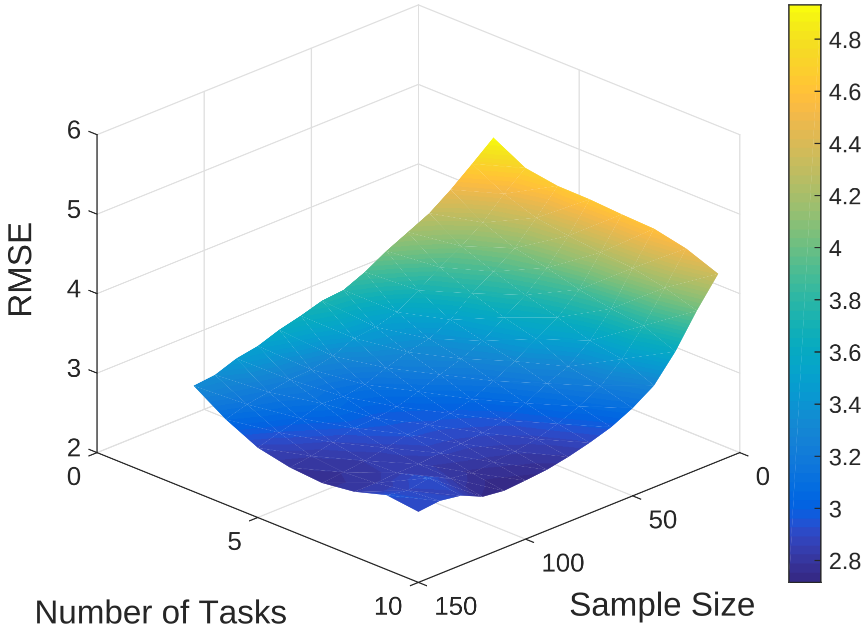


(a) RMSE

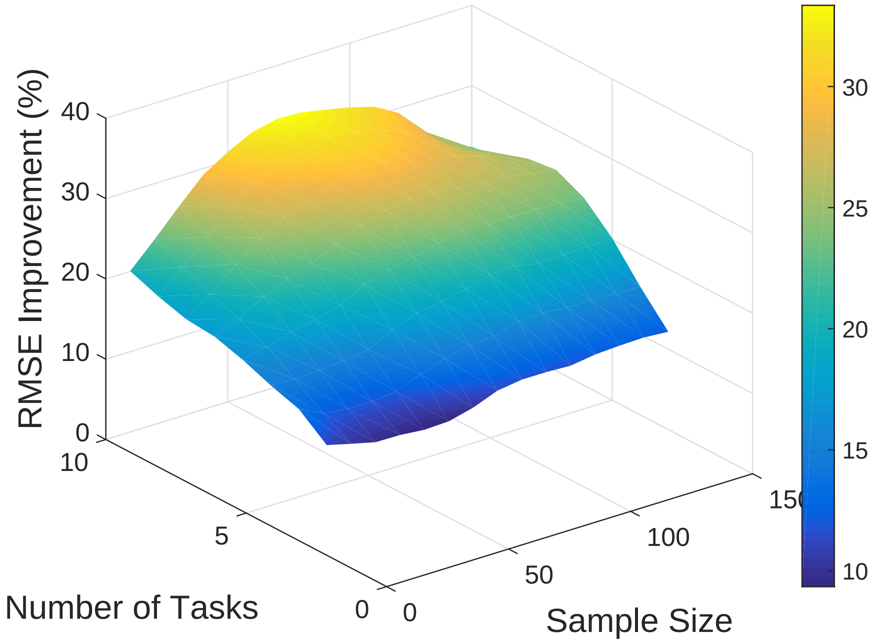


(b) RMSE improvement percent, $\mathrm{RSME_I}$

Figure 3.13 Effects of the number of tasks on the prediction performance



(a) RMSE



(b) RMSE improvement percent, $\mathrm{RSME_I}$

Figure 3.14 Effects of sample size and number of tasks on the prediction performance

increase of sample size and number of tasks.

The above experimental results lead to three major findings:

(1) Multi-task learning is able to improve the prediction accuracy over simple kriging;

(2) For both multi-task learning and kriging, the prediction accuracy improves as the increase of the training sample size;

(3) The prediction performance improves when the number of tasks increases for multi-task learning;

(4) $\mathrm{RSME}_{\mathrm{I}}$ becomes bounded when the sample size and the number of tasks are relatively large.

Intuitively, when the training sample size is sufficient, single-task learning, which is simple kriging in this application, will be able to learn the spatial model very well, and the benefits brought by transferring knowledge thus become limited. Moreover, when the number of tasks increases, duplicate information may exist across tasks, and the improvement becomes less significant.

3.4.3 Effect of the Correlation Strength

To investigate the influence of $\rho(U,Z)$ on the estimation accuracy, we vary the strength of $\rho(U,Z)$, and compare the performance of the EG-MTL model, MTGPL, simple kriging, and the KED method. The comparative results are illustrated using Figures 3.15a and 3.15b, which show the trends of RMSE and $\mathrm{RSME}_{\mathrm{I}}$, respectively. In Figures 3.15a and 3.15b, the results corresponding to each correlation strength are averaged over three parts. The following observations are made from these results.

(1) The performance of KED is strongly influenced by the correlation strength. When the correlation between U and Z is relative weak, e.g., $\rho(U,Z) < 0.75$, its prediction accuracy is worse than simple kriging. When $\rho(U,Z)$ becomes stronger, the prediction accuracy of KED improves, and outperforms simple kriging. Also, the RMSE

improvement from KED to the EG-MTL model reduces as $\rho(U,Z)$ increases, indicating that the benefits introduced by multi-task learning become limited.

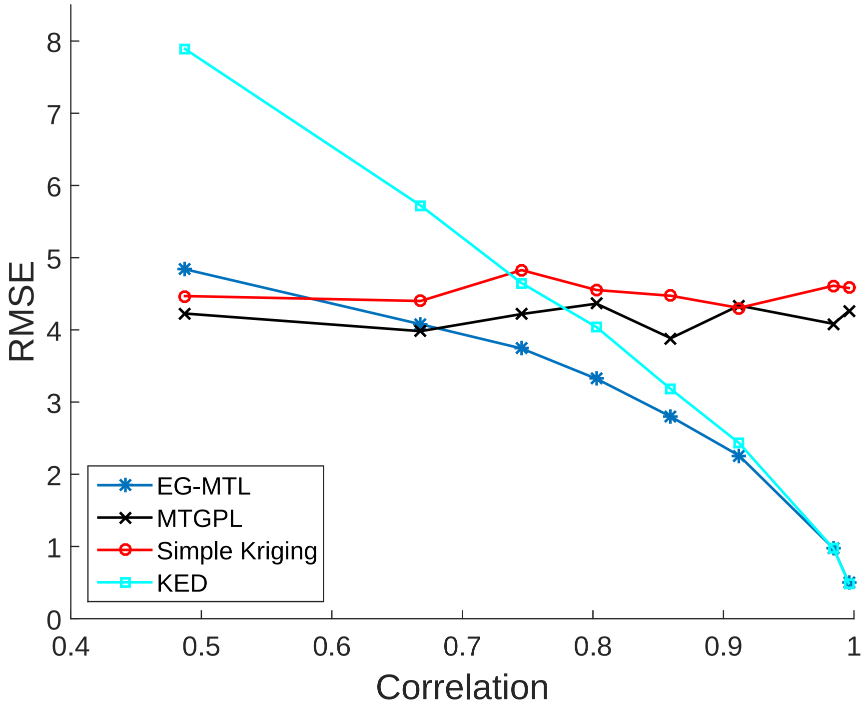
(2) The prediction accuracy using the EG-MTL model is significantly affected by the correlation strength: as $\rho(U,Z)$ increases, RMSE for the EG-MTL model decreases and RSME_I increases. When the correlation is relatively low, e.g., less than 0.5, introduction of such a covariate will even degrade the prediction performance.

(3) GPMTL outperforms simple kriging under all circumstances, indicating that when expert knowledge on the process is not available and thus highly correlated variables cannot be identified, the adoption of multi-task learning is beneficial compared with single task learning.

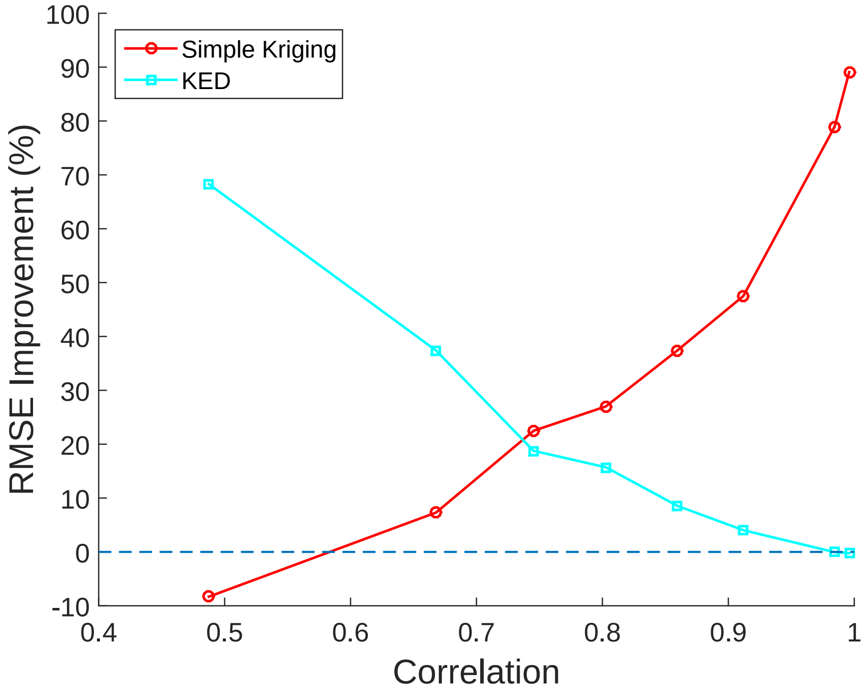
3.5 Conclusion

This chapter develops a novel engineering-guided multi-task learning (EG-MTL) surface model by incorporating an engineering model with multi-task learning for Gaussian processes. Specifically, the following accomplishments are achieved:

(1) *An improved surface shape modeling approach by incorporating engineering insights with multi-task learning for Gaussian processes*: The major uniqueness of the developed model is that information from both the engineering knowledge and spatial data from other similar-but-not-identical processes is fused in one framework. The model is advantageous in predicting surface height at unobserved locations over traditional approaches, such as simple kriging, kriging with external drift, and data-driven multi-task learning for Gaussian processes. The approach



(a) RMSE vs. $\rho(U,Z)$



(b) RMSE improvement percent, $\mathrm{RSME_I}$ vs. $\rho(U,Z)$

Figure 3.15 Effects of $\rho(U,Z)$ on the prediction performance is validated using the data from real-world machining processes in an engine plant.

(2) *An iterative algorithm for estimating model parameters*: A novel algorithm is developed to iteratively estimate the parameters in the engineering model and the Gaussian process. Parameters are updated in each iteration and the procedure is terminated once the convergence criterion is met. Moreover, practical suggestions are provided to expedite the EM optimization process.

(3) *Systematic discussions on the hyperparameter selection and the effects of training data availability*: Response surfaces of the prediction performance are created for the model hyperparameters, the number of tasks, and the training sample size. Additionally, it is revealed that the model performance improves as more engineering insights of the process become available.

The proposed engineering integration model is expected to improve the surface prediction accuracy with limited measurement data, thus alleviating the high costs induced by high-resolution 3D measurement systems and enabling fine-resolution surface characterization on the plant floor.

Future research will be focused on (1) variable selection for the engineering model and (2) task selection for multi-task learning. First, variables reflecting multidisciplinary process information may take a variety of types and formats. These variables may not be equally important for improving the surface shape prediction or they may contain duplicate information. Consequently, a systematic variable selection algorithm is highly desirable to identify a parsimonious subset of process variables for the engineering model. Second, a systematic approach is needed to select tasks/surfaces for multi-task learning. Some algorithms have been reported in the literature to select tasks based on certain task similarity quantification metrics. However, the selection problem is more complicated in the proposed integration model due to the model structure, and nontrivial efforts are expected to devise a systematic task selection algorithm.

Chapter 4

Tool Wear Monitoring for Ultrasonic Metal Welding of Lithium-Ion Batteries

4.1 Introduction

In manufacturing lithium-ion batteries for electrical vehicles such as the Chevrolet Volt, it is critical to create reliable interconnections among battery cells, from module to module, and from module to control unit. Ultrasonic metal welding is advantageous in joining multi-layer dissimilar, thin and conductive materials (Kim et al., 2011; Shao et al., 2013). It is a solid-state joining process which uses ultrasonic vibration to generate oscillating shears between metal sheets clamped under pressure (Lee et al., 2013, 2014). A typical ultrasonic metal welding system is shown in Figure 4.1. The surfaces of the weld tools, i.e., horn and anvil, consist of a large number of pyramid-shape knurls as displayed by Figure 4.2. The horn and anvil wear out quickly in production and are expensive to replace. As a result, monitoring of the horn and anvil wear is critically needed to ensure battery joining quality and reduce production cost.

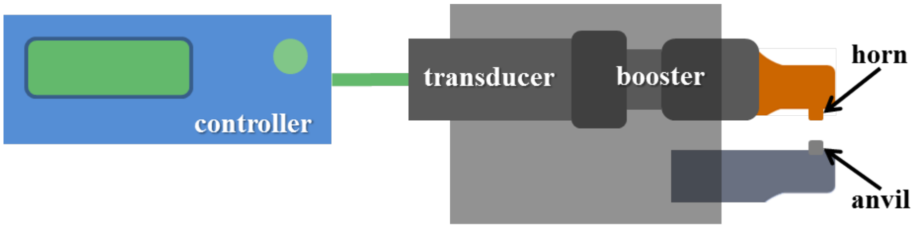
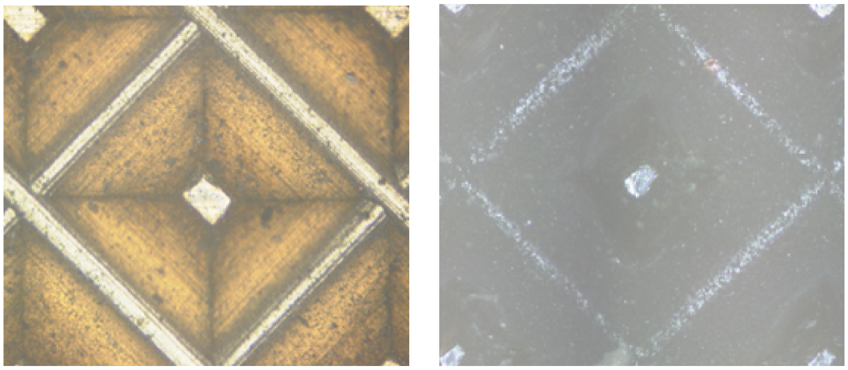


Figure 4.1 A typical ultrasonic metal welding system (Shao et al., 2013)



(a) Horn knurl **(b)** Anvil knurl

Figure 4.2 Pyramid-shape knurls on the horn and anvil (Shao et al., 2014)

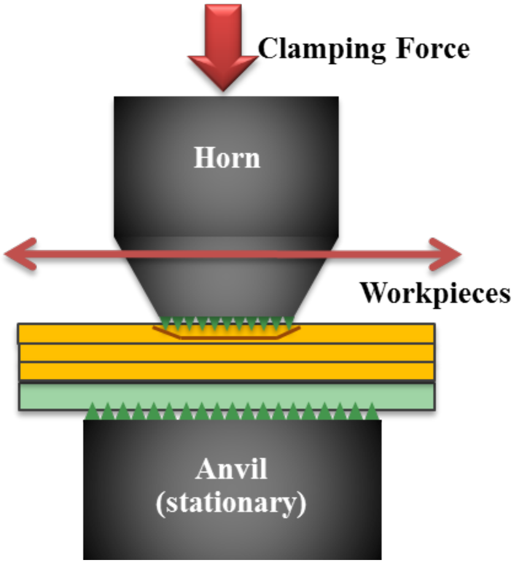


Figure 4.3 Ultrasonic welding mechanism

The ultrasonic metal welding process is illustrated by Figure 4.3. Workpieces are placed between the horn and anvil, and a clamping force is applied to hold the workpieces tightly. During welding, the horn vibrates at a frequency of around 20 kHz, while the anvil is stationary. It is reported in Shao et al. (2014) that relative movements exist between the top metal sheet and the horn, as well as between the bottom sheet and the anvil, and these relative movements are believed to be a major cause of tool wear.

In automotive lithium-ion battery manufacturing, horn and anvil are reported to be a major production cost. Specifically, production costs as a result of tool wear can be divided into two major categories (Shao et al., 2014): (1) costs due to machine down-time as caused by tool wear induced quality problems or time needed for tool replacement; and (2) costs for fabricating, reworking, or refurbishing the replaced tools. Vehicle battery manufacturing has a strict quality requirement for battery tab joining because any low-quality joints may result in the failure of an entire battery pack, causing high production loss (Shao et al., 2013). Consequently, when a TCM system is not available, a conservative tool replacement strategy is generally utilized to ensure satisfactory quality. For example, some battery plant uses the number of welds as a measure of tool wear, and replaces tools once the number of welds reaches a certain limit. While this empirical strategy is straightforward to implement, it may sacrifice some useful tool lives and increase production costs.

TCM has been a popular and important research topic in manufacturing and has received tremendous attention over the past several decades. The majority of the TCM literature has been focused on machining processes (Jantunen, 2002; Cook, 1973; Koren et al., 1991; Abellan-Nebot and Subirón, 2010; Rehorn et al., 2005; Zhou et al., 2011; Ertunc et al., 2001; Dimla, 2000; Kurada and Bradley, 1997a,b; Lanzetta, 2001; Byrne et al., 1995) and forming processes (Kang et al., 1999a,b; Lepadatu et al., 2006; Kong and Nahavandi, 2002). Tool wear mechanism in cutting processes has been investigated in (Cook, 1973; Koren et al., 1991) using physical or empirical models. Tool wear monitoring techniques have been developed in Abellan-Nebot and Subirón (2010); Rehorn et al. (2005); Zhou et al. (2011); Ertunc et al. (2001); Dimla (2000); Kurada and Bradley (1997a,b); Lanzetta (2001). In general, tool wear monitoring techniques can be categorized into direct and indirect methods (Byrne et al., 1995). Direct methods determine tool conditions by measuring tool wear using visual inspection or computer vision. However, direct methods are not attractive economically or technically mainly due to the environmental restrictions on the plant floor (Jantunen, 2002). Hence, indirect methods using on-line signals are often more desirable, and some exemplary scenarios can be found in Jantunen (2002); Koren et al. (1991); Abellan-Nebot and Subirón (2010); Rehorn et al. (2005); Zhou et al. (2011); Ertunc et al. (2001); Dimla (2000). A typical method for developing an indirect monitoring system includes the following key steps (Abellan-Nebot and Subirón, 2010): (i) sensor selection; (ii) signal pre-processing; (iii) feature generation; (iv) feature selection/extraction; (v) monitoring decision and faulty classification using artificial intelligence technique. A comprehensive review on indirect monitoring methods is presented in Abellan-Nebot and Subirón (2010).

Tool wear in forming processes has also been investigated, especially in extrusion and forging processes (Kang et al., 1999a,b; Lepadatu et al., 2006; Kong and Nahavandi, 2002). Archard's wear model is widely applied in studies on extrusion processes (Kang et al., 1999a,b). Statistical process control analysis of the tool wear progression in a metal extrusion process was conducted in Lepadatu et al. (2006). On the TCM of forging processes, an on-line TCM system using artificial neural network was developed to integrate information from multiple sensors (Kong and Nahavandi, 2002).

Despite extensive literature focusing on TCM development for machining and forming processes, limited studies have been conducted on TCM for ultrasonic metal welding. Developing a TCM system for ultrasonic welding is more challenging than machining or forming processes, mainly because: (1) the ultrasonic welding mechanism has not been thoroughly understood, and ultrasonic welding possesses the characteristics of high-frequency (around 20 kHz) and short-duration (each cycle lasts approximately 0.6 s), and (2) the geometry of welding tools is much more complicated (Shao et al., 2014).

To address the above challenges and develop an effective TCM system for ultrasonic welding, this paper (1) characterizes tool wear progression by comparing tool surface measurements in different wear stages; (2) designs an indirect method to efficiently obtain tool surface measurements in plant environment without taking the tools off-line; and (3) develops a tool condition classification algorithm with application-dependent features which are generated from both space and frequency domains.

The remainder of this chapter is organized as follows. Section 4.2 characterizes tool wear progression using changes in the knurl geometry based on high-resolution 3-D measurements. An impression method is presented in Section 4.3 as an indirect tool geometry measurement strategy. Section 4.4 develops a tool condition classification algorithm. Finally, Section 4.5 summarizes the chapter.

4.2 Tool Wear Characterization

This section characterizes the tool wear progression in ultrasonic metal welding based on the comparison of the optical images and height profiles at different wear stages. Representative samples of four different stages were collected from a battery plant and then measured using a 3D microscope. Specifically, Section 4.2.1 compares the optical images; Sections 4.2.2 and 4.2.3 depict the wear progression in the direction perpendicular to vibration and in the vibration direction, respectively. For the sake of simplicity, we define the direction perpendicular to vibration as “horizontal direction” and the vibration direction as “vertical direction.”

4.2.1 Comparison of Optical Images

The optical images of the four stages are shown by Figure 4.4, and the characteristics of each stage are summarized as follows (Shao et al., 2014).

Stage 1: Each knurl possesses a regular pyramid shape.

Stage 2: The peak remains at each knurl, but a small flat top forms, and the knurl shape becomes more complex geometrically than Stage 1: more materials have been removed at the left and right sides than the upper and lower sides. In addition, the colors of peaks become shining.

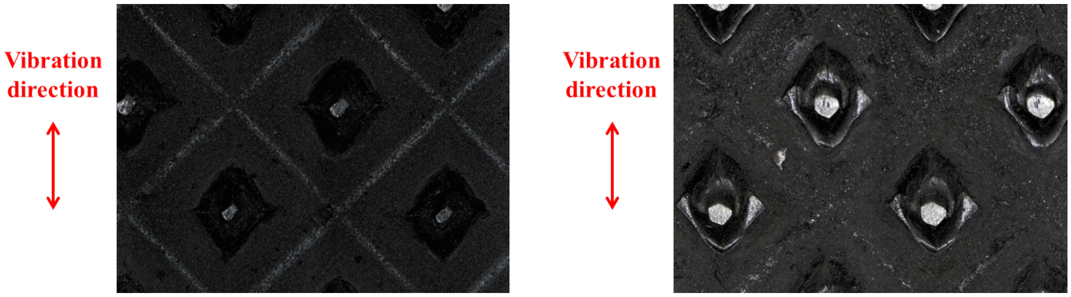
Stage 3: The left and right sides of each knurl are almost flat, and only a small amount of materials remain in the upper side.

Stage 4: All peaks have been removed and the knurls take a frustum shape.

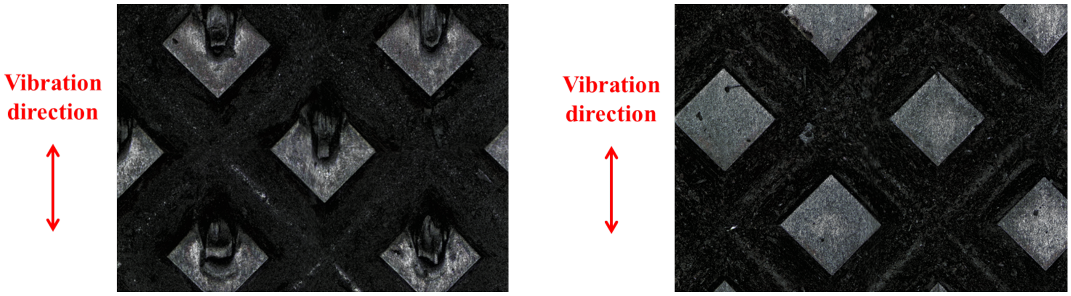
4.2.2 Tool Wear Progression in the Horizontal Direction

In each anvil, typical cross-sectional profiles are extracted in the horizontal direction, and the profiles are shown in Figure 4.5.

In Figure 4.5a, one can see that the cross sections of all knurls of a new tool have triangle shapes and similar peak height. In Figure 4.5b, three types of tool wear patterns can be observed: (1) flank wear, (2) side wear, and (3) breakage. The flank wear is at the upright direction, which occurs with the removal of peaks (height decreases); the side wear represents the wear around the peaks, where two shoulders form at the left and right sides; the breakage happens with a removal of the whole knurl peak and is indicated by a depressed



(a) Stage 1 **(b)** Stage 2



(c) Stage 3 **(d)** Stage 4

Figure 4.4 Optical images of different wear stages (Shao et al., 2014)

shape. In Figure 4.5c, the height reduces significantly compared with Figure 4.5b, and some of the peaks are almost removed. Additionally, the width of shoulders increases. In Figure 4.5d, all peaks have disappeared, and the surface becomes completely flat.

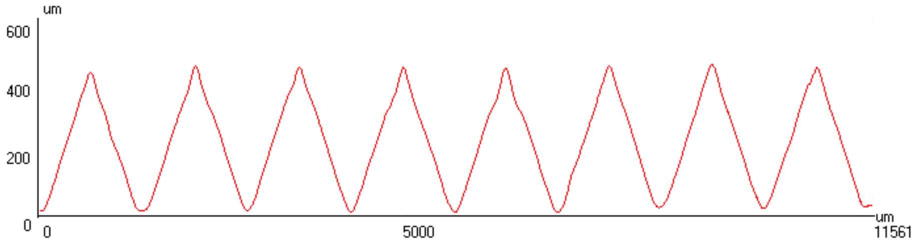
Based on the findings revealed by Figure 4.5, the characterization of each stage in the horizontal direction is summarized by Figure 4.6 and described as follows (Shao et al., 2014).

Horizontal direction

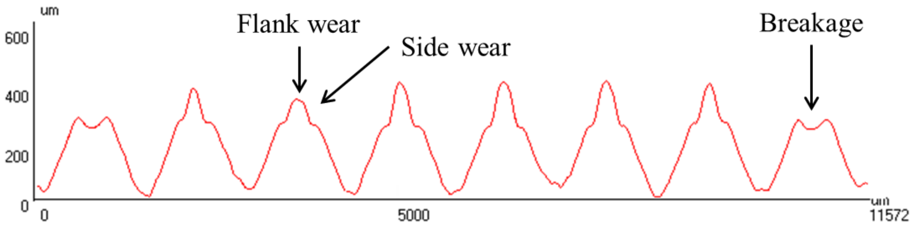
Stage 1: The knurl is new, and it possesses a triangle shape.

Stage 2: Material is removed in both downward and lateral directions, and shoulders appear on the left and right sides.

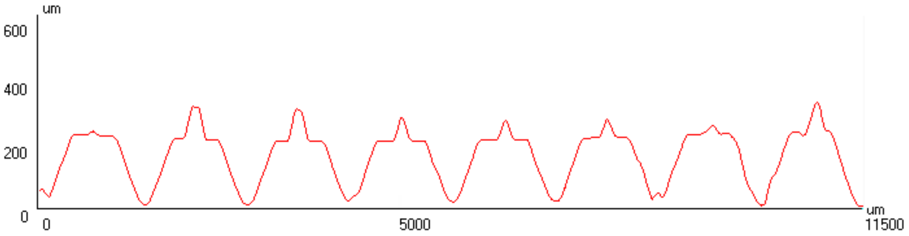
Stage 3: Height decreases significantly, and the width of shoulders increases.



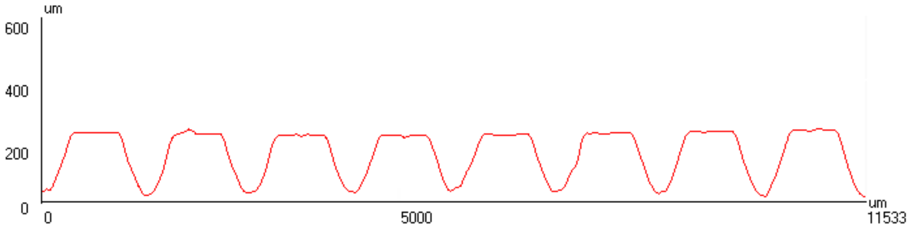
(a) Stage 1



(b) Stage 2



(c) Stage 3



(d) Stage 4

Figure 4.5 Cross-sectional profiles in the horizontal direction (Shao et al., 2014)

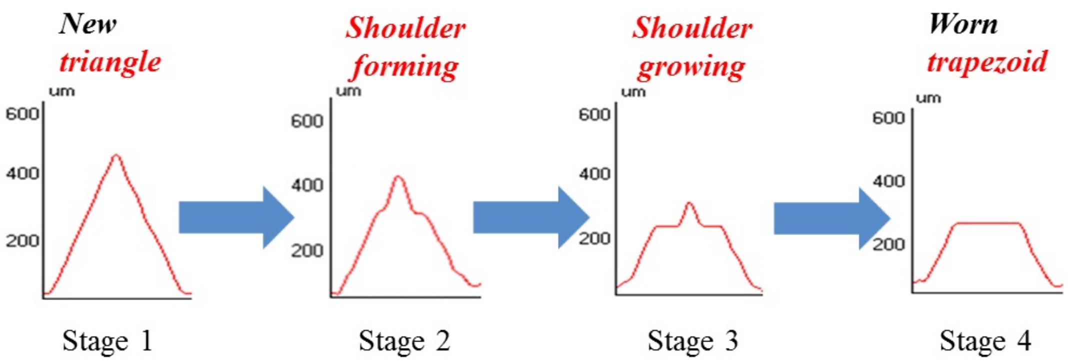


Figure 4.6 Anvil knurl wear progression in the horizontal direction (Shao et al., 2014)

Stage 4: Material is removed until the peak disappears, and a trapezoid shape forms.

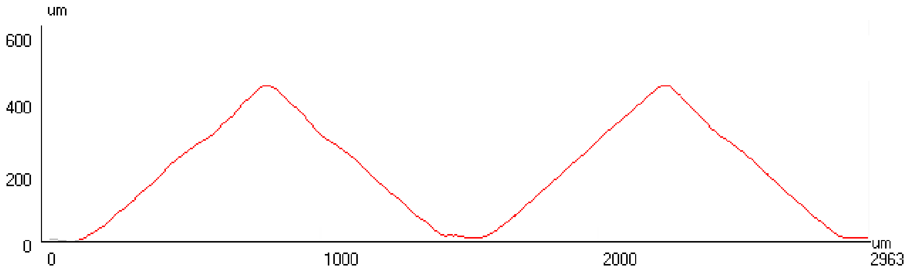
4.2.3 Tool Wear Progression in the Vertical Direction

Similar to Section 4.2.2, the cross-sectional profiles in the vertical direction are also extracted and compared, and the results are shown in Figure 4.7.

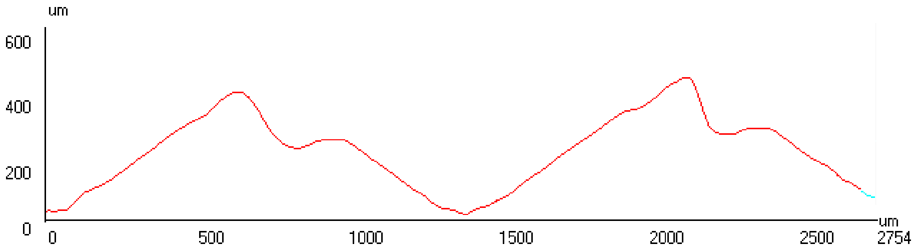
As shown in Figure 4.7a, a new anvil has triangle shapes. In Figure 4.7b, an asymmetric pattern can be seen in two sides of the knurl in the vertical direction: more materials have been removed in the knurl's lower side, and a groove forms. A knurl in this stage has two peaks, i.e., a main peak and a side peak. This asymmetry is further illustrated in Figure 4.8, which was obtained using a high-resolution microscope. Figure 4.8a is an optical image obtained by a microscope, and Figure 4.8b is the corresponding 3D height plot.

In Figure 4.7c, one can see that after more material removal, the side peak in Figure 4.7b disappears and only a main peak remains. Figure 4.7d shows the final wear stage in the vertical direction. The main peak in Figure 4.7c has been completely removed, and a flat surface forms in the end.

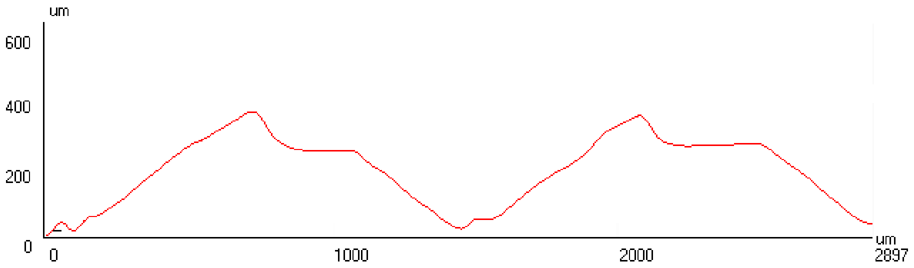
According to the results shown in Figure 4.7, the knurl-level wear progression in the



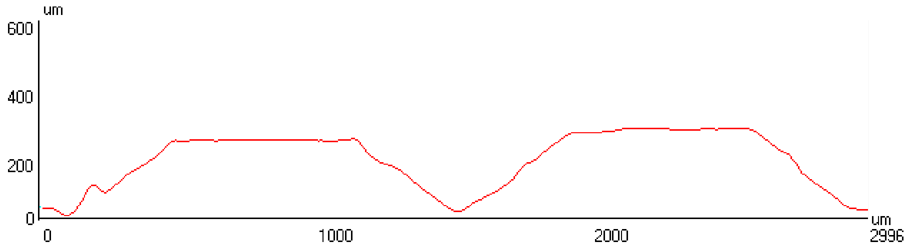
(a) Stage 1



(b) Stage 2

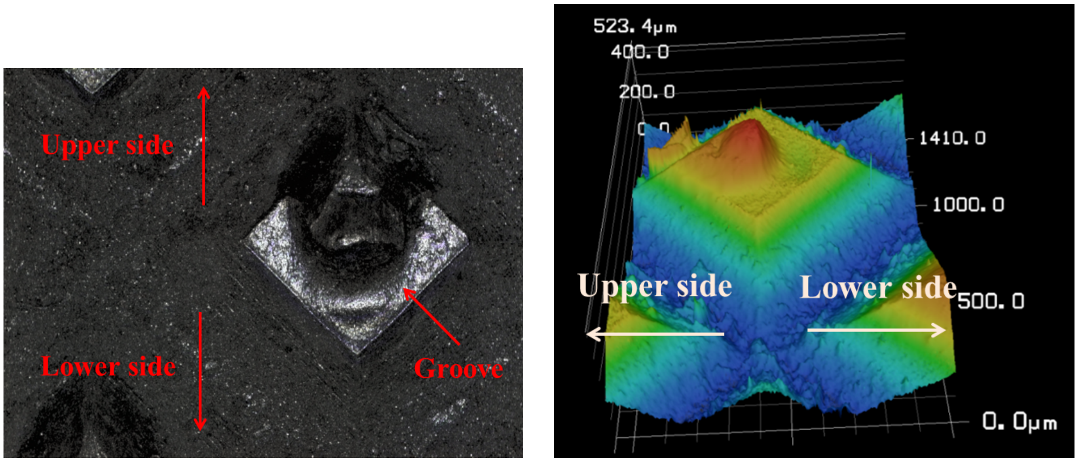


(c) Stage 3



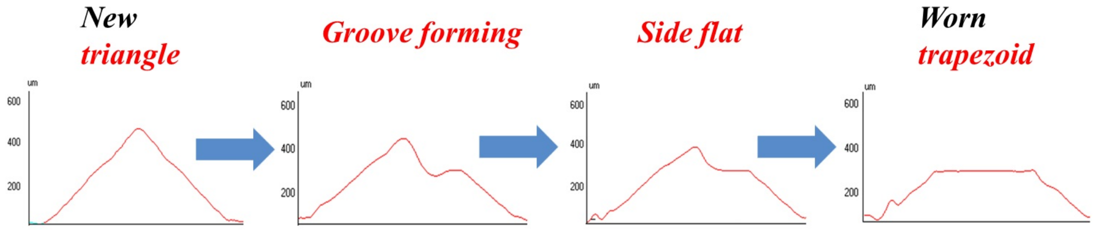
(d) Stage 4

Figure 4.7 Cross-sectional profiles in the vertical direction (Shao et al., 2014)



(a) Optical image **(b)** 3D image

Figure 4.8 Pyramid-shape knurls on the horn and anvil (Shao et al., 2014)



Stage 1 Stage 2 Stage 3 Stage 4

Figure 4.9 Anvil knurl wear progression in the vertical direction (Shao et al., 2014)

vertical direction is illustrated by Figure 4.9. The characteristics of the knurls in each stage are summarized as follows (Shao et al., 2014).

Vertical direction

Stage 1: The knurl possesses a triangle shape.

Stage 2: Materials are mainly removed in the lower side, where a groove and a side peak form. The height of the main peak decreases compared with Stage 1.

Stage 3: More materials are removed until the side peak disappears. The height of the main peak continually decreases.

Stage 4: The main peak is completely removed, and a trapezoid forms.

4.3 Impression Method

In ultrasonic welding, tool surface geometry patterns provide essential tool condition information; hence an efficient tool surface measurement method is necessary. Most 3D measurement systems require the tools to be removed from the welding machine and be placed on a flat and stable fixture. As a result, tool disassembly is necessary in order to directly measure the tool surfaces. However, tool disassembly and assembly for ultrasonic welding machine are complicated and time-consuming, leading to a significant amount of machine down time, which is not desirable in production. As a result, an innovative measurement scheme is crucial in real productions.

To effectively obtain tool surface measurements without introducing significant interruption to production, an "impression method" is developed. Rather than measuring the tool, an impression is made on a weld coupon. The depth of deformation is measured under a microscope and then the inverse of the coupon image is created as a surrogate of the tool image. The process of obtaining the tool image from an "impression" is illustrated in Figure 4.10. When a measurement is needed, one coupon will be generated using pre-determined weld parameters and materials, and then it will be measured using a 3D metrology system. Finally, data processing, i.e., horizontal flip and height inversion, will be applied, and the original tool surface profiles are finally reconstructed.

The weld parameters and the coupon materials need to be carefully selected in order to optimize the quality of the reconstructed tool surfaces. In practice, design of experiment (Wu and Hamada, 2011) can be used to identify the optimal combination of the weld parameters and the coupon materials. In general, soft metals are recommended as they are

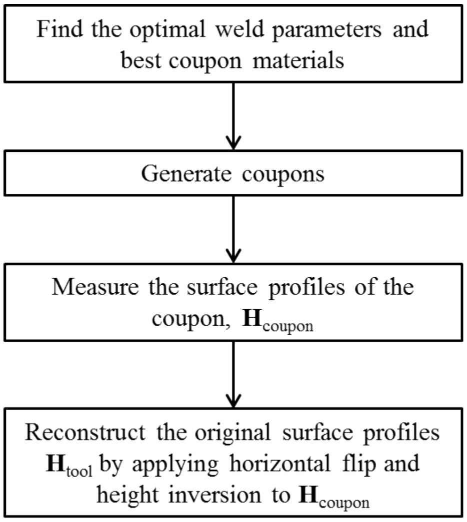
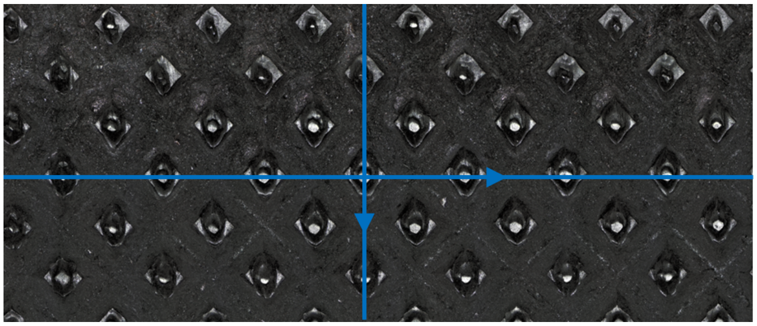


Figure 4.10 Flowchart for impression method

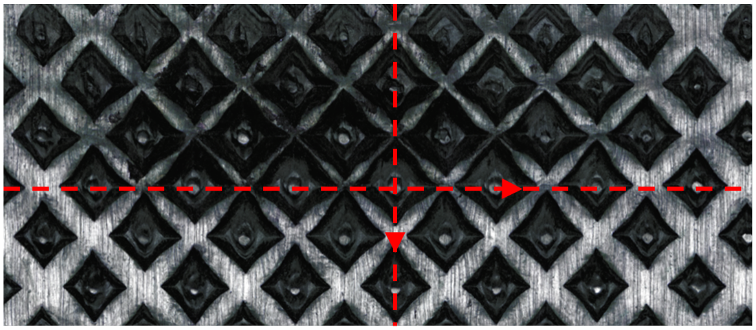
more efficient in capturing tool surface profiles. In this study, four layers of pure aluminum are adopted, and the thickness of each layer is 0.2 mm.

To validate this method, several anvils and corresponding coupons are measured for comparison, and good agreement is achieved between the original tool surfaces and reconstructed ones. For illustrative purposes, only the results from one pair of anvil and coupon are presented, as shown by Figure 4.11.

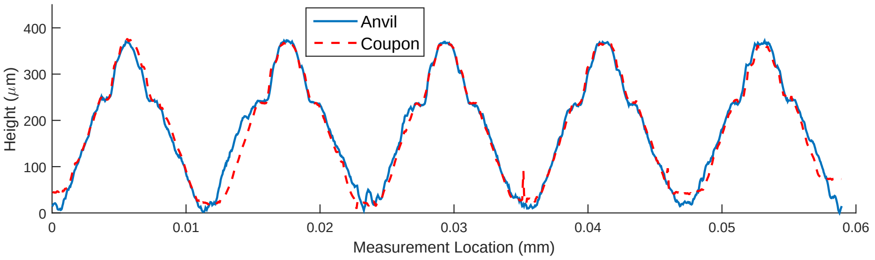
In Figure 4.11, (a) and (b) display the images of anvil and coupon, respectively; (c) and (d) compare the cross-sectional profiles from the anvil and the coupon in the horizontal and vertical directions. The blue solid lines in (a) and red dashed lines in (b) indicate the measurement paths for the cross-sectional profiles. In (c) and (d), the blue solid lines represent the measurement of the anvil, and the red dashed lines correspond to the coupon measurement. It is demonstrated by Figure 4.11 that the coupon impression is able to capture the knurl height profiles effectively. Additionally, the knurl shapes from the impression methods are also very similar to these from the direct measurement. Thus, the designed



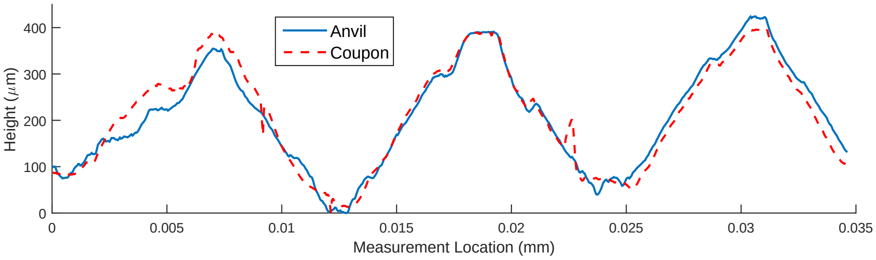
(a) Anvil image



(b) Coupon image



(c) Comparison of horizontal profiles



(d) Comparison of vertical profiles

Figure 4.11 Comparison between measurements of a tool and a coupon

Table 4.1 Comparison results for impression method

	Correlation coefficient	Root mean squared error
Horizontal	0.9864	18.7 μm
Vertical	0.9731	25.9 μm

impression method is effective in constructing the tool surface geometry.

The directly measured profiles and coupon profiles are further quantitatively compared in Table 4.1. The high correlations and low root mean squared errors indicate that the coupon profiles are able to well capture the tool profiles.

4.4 Tool Condition Classification

This section presents a tool condition classification algorithm to identify the state of wear. First, monitoring features are extracted from surface data to characterize tool conditions. Then Fisher's discriminant ratio is used to select features which are closely relevant to tool conditions. Finally, classifiers are developed to classify tool conditions.

4.4.1 Monitoring Feature Generation

In this subsection, several monitoring features are generated from surface data in both the space and frequency domains for tool condition classification. As indicated by Figures 4.6 and 4.9, cross-sectional profiles in the horizontal direction which cut cross through knurl centers can well capture the characteristics of different tool life stages. Accordingly, features are extracted from the cross-sectional profiles to quantify tool wear. The procedure of extracting features is illustrated by Figure 4.12. First, tool surface is reconstructed by applying image processing algorithms, including noise elimination, image rotation, baseline adjustment, horizontal flip, and height inversion, to the coupon data. Then representative cross-sectional profiles are obtained from tool surfaces. Assume the tool surface height is represented by a matrix $\mathbf{H}$, the size of which is $n_r \times n_c$, where n_r and n_c are the numbers of rows and columns, respectively. The extraction of the horizontal profiles can be conducted based on the following algorithm.

Profile Extraction Algorithm

Step 1: Calculate the sum of each row for $\mathbf{H}$. For the ith row, the sum is calculated using Equation (4.1):

$$S_i = \sum_{j=1}^{n_c} \mathbf{H}_{ij}, \tag{4.1}$$

where i is the row index, $i = 1, \ldots, n_r$, and j is the column index.

Step 2: Treat $\{S_i\}$ as a profile, and identify the local maxima. We denote the indices of the local maxima as $m_1, \ldots, m_{n_p}$, where n_p is the number of knurls along the vertical direction.

Step 3: Obtain the horizontal profiles by extracting the row vectors with the indices from Step 2. For the kth index, the corresponding profile is extracted using Equation (4.2).

$$l_k = \mathbf{H}_{m_k \cdot}, \tag{4.2}$$

where l_k is the kth profile, $k = 1, \ldots, n_p$.

Features are extracted from the obtained horizontal profiles in both space and frequency domains. In the space domain, two features are used to describe knurl geometry, i.e., average knurl height variance and average shoulder width. Knurl height variance is defined

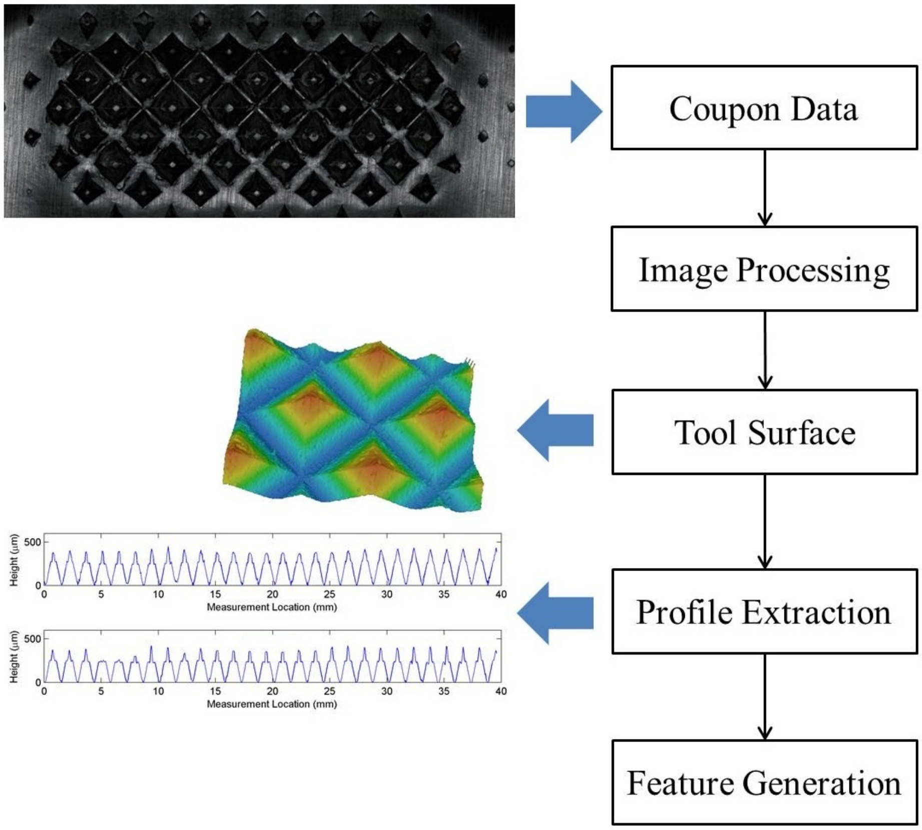


Figure 4.12 Process flowchart for feature extraction

as the variance around knurl peaks. As tool wears out, materials are removed from knurl top, and a smaller height variance will present. Shoulder width is defined as the sum of the left and right shoulders besides knurl peaks, as illustrated by Figure 4.13. As the tool wear becomes more severe, the shoulder width will increase. After calculating knurl-level variance and shoulder width, tool level features are obtained by averaging them over all knurls.

Frequency-domain features are the amplitudes corresponding to the dominant frequencies after applying fast Fourier transform (FFT) to the cross-sectional profiles. These features are able to capture the overall periodic pattern which is closely related to the tool wear level. Figure 4.14 shows (a) original profiles in four wear stages and (b) FFT profiles.

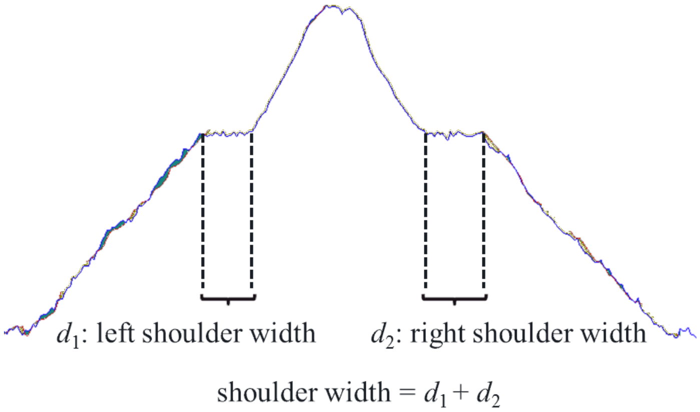
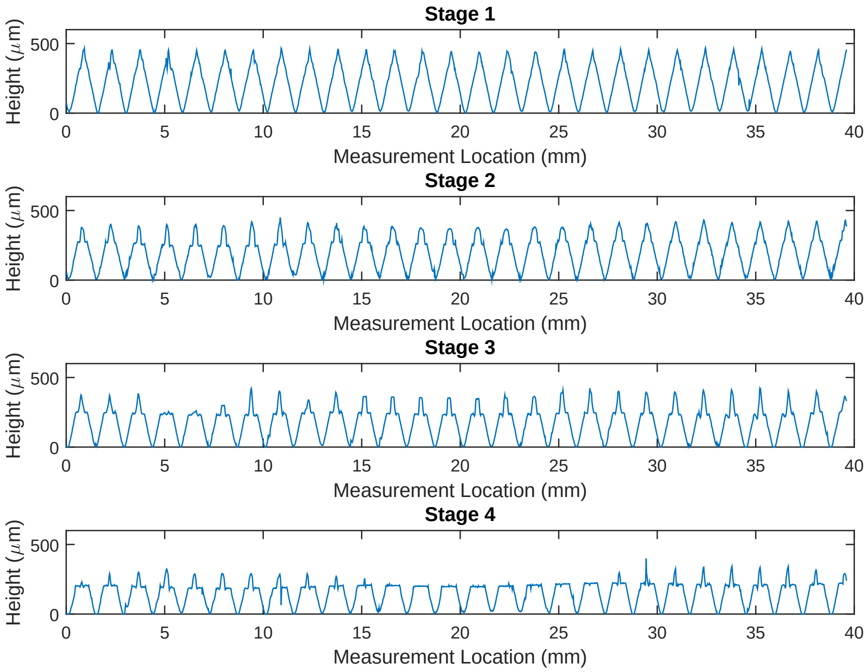


Figure 4.13 Shoulder width calculation

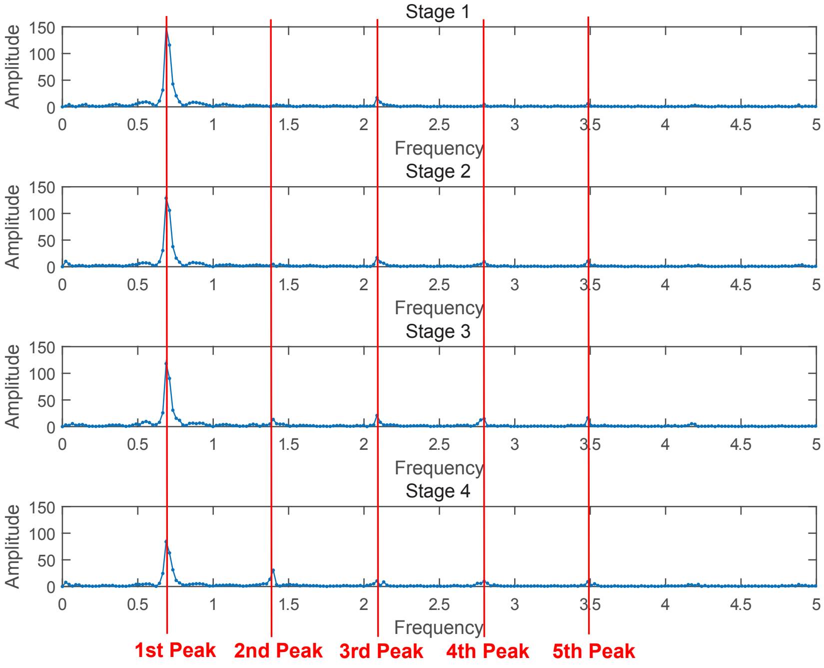
From (b), it is shown that the amplitude of the 1st dominant frequency, 0.7040, which is corresponding to the diameter of a knurl, 1.4 mm, has a decreasing trend as tool wear gets more severe; however, the amplitude of the 2nd dominant frequency, 1.43, corresponding to half period, 0.7 mm, has an increasing trend. Thus, the amplitudes of 1st to 5th dominant frequencies are considered as candidate features.

A summary of candidate features is given by Table 4.2. Two features are extracted from the space-domain to depict the average knurl geometry and five features are generated from the frequency-domain to acquire periodic patterns of cross-sectional profiles.

Figure 4.15 depicts the trend of the extracted features over the number of welds. In each subplot, the horizontal axis is the number of welds, and the vertical axis is the feature value. It is seen that, Features 1, 3, and 4 have decreasing trends; Features 2, 5, and 6 have increasing trends; while Feature 7 first increases and then decreases as the number of welds increases.



(a) Profiles in the space domain



(b) Frequency-domain features

Figure 4.14 Frequency features for different stages of wear

Table 4.2 Candidate feature list

Feature No.	Feature Name	Category
Feature 1	Average knurl-level variance	Space-domain
Feature 2	Average shoulder width	
Feature 3	1st peak amplitude	Freq.-domain
Feature 4	2nd peak amplitude	
Feature 5	3rd peak amplitude	
Feature 6	4th peak amplitude	
Feature 7	5th peak amplitude	

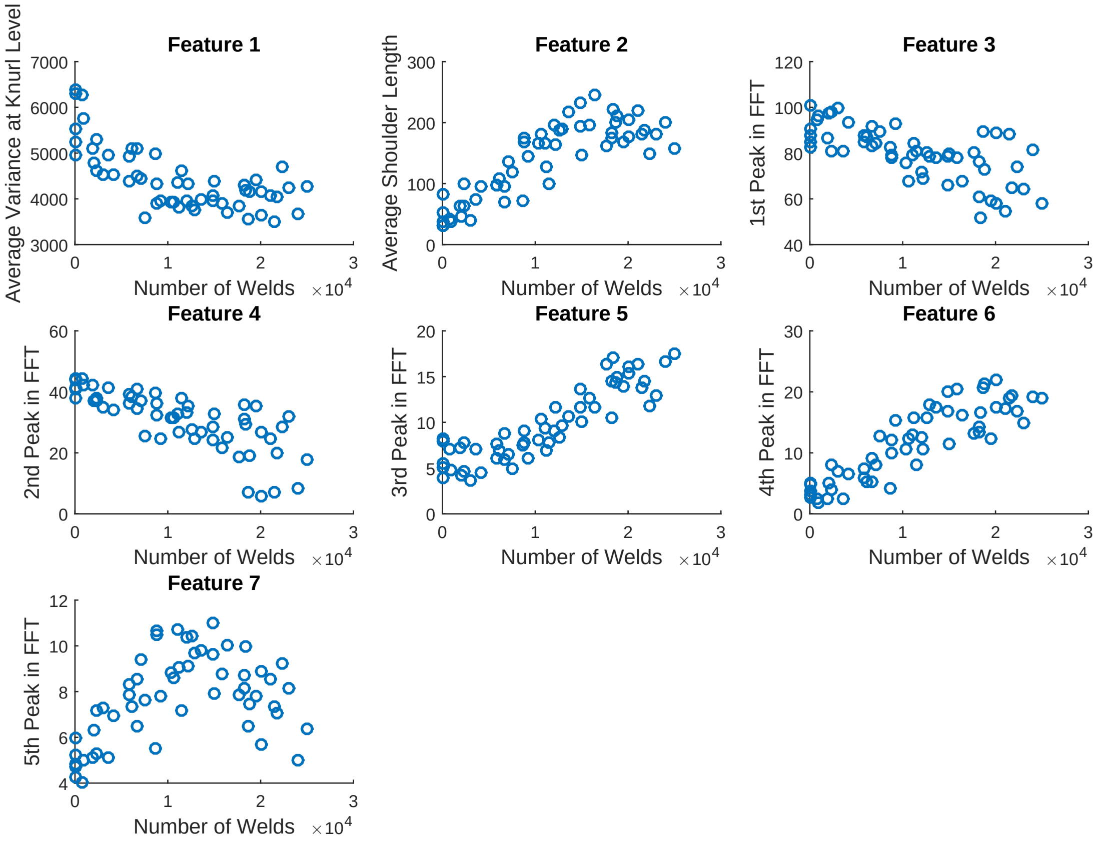


Figure 4.15 Feature trend vs. the number of welds

4.4.2 Feature Selection

In the development of monitoring algorithms, feature selection is an essential step for achieving best monitoring performance, as not all features have good separability between different classes (Shao et al., 2013). Fisher's discriminant ratio is applied to perform feature screening in a computationally simple and fast manner. Fisher's discriminant ratio was first presented in Fisher (1936), and it is a separability measure for feature selection (Duda et al., 2012). A larger Fisher's ratio indicates more significant difference between two classes. Fisher's ratio is defined as

$$J = \frac{(\mu_1 - \mu_2)^2}{s_1^2 + s_2^2}, \tag{4.3}$$

where μ_1 and μ_2 are the means of new and worn classes, s_1^2 and s_2^2 are respective variances.

56 coupons have been collected from a battery manufacturing plant, and expert knowledge is used to classify them as new (Class 1) and slightly worn (Class 2). Fisher's ratios are then calculated for all candidate features, the results of which are shown by Figure 4.16. Features 2, 5, and 6 have higher ratios, indicating that they can provide better separability between new and slightly worn coupons. As a result, Features 2, 5, and 6 (average shoulder width, 3rd and 4th peak amplitudes in the frequency-domain) are chosen for monitoring tool conditions. Scatter plots of these features are shown by Figure 4.17, which indicates good class separability.

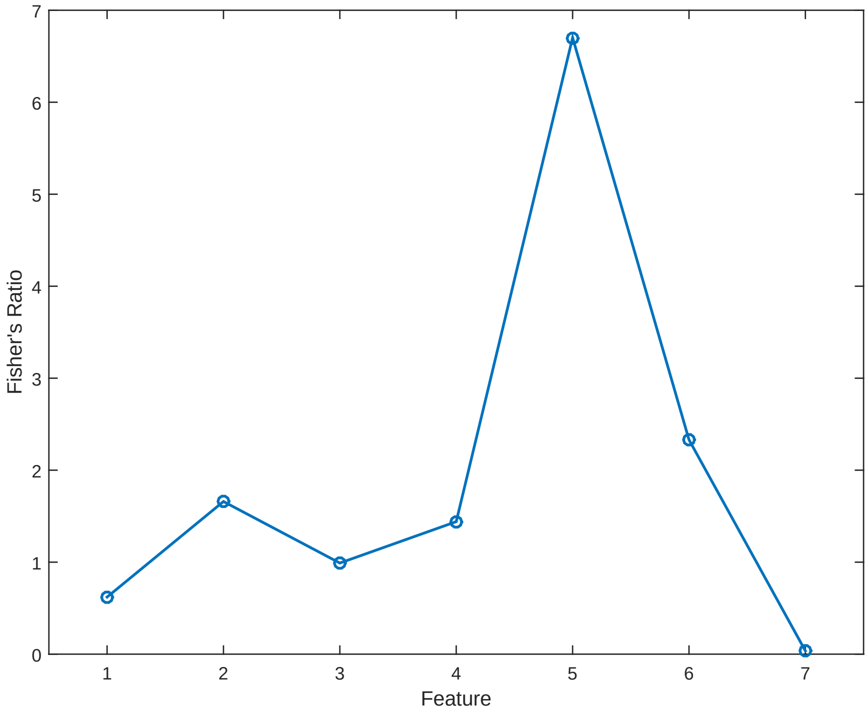


Figure 4.16 Fisher's ratio for all features

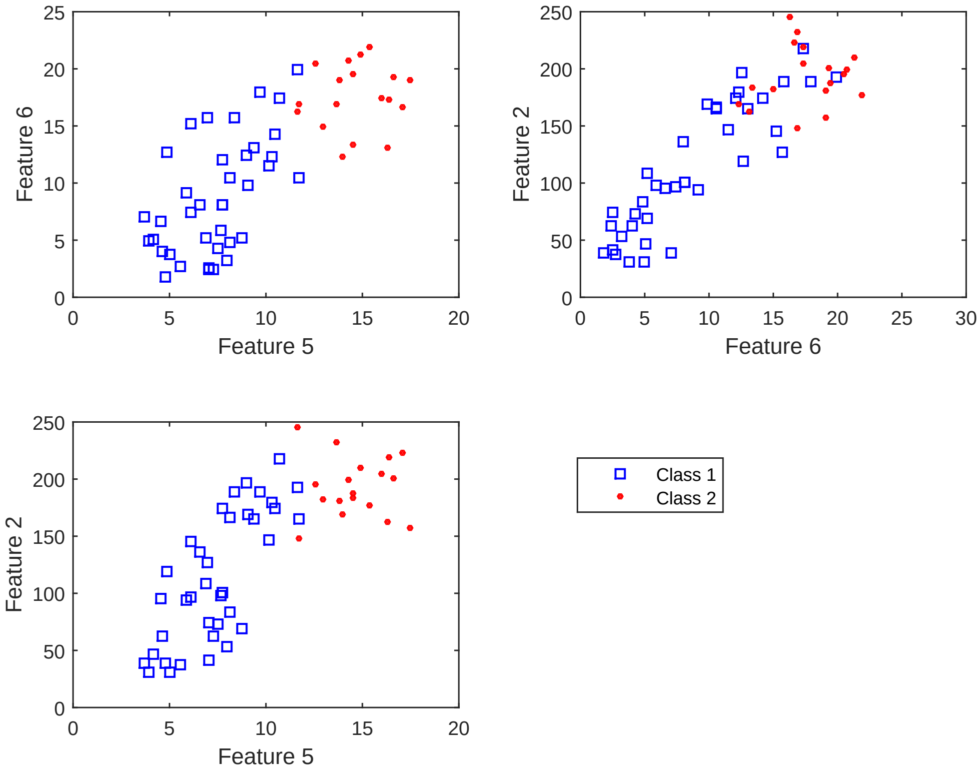


Figure 4.17 Scatter plots of selected features

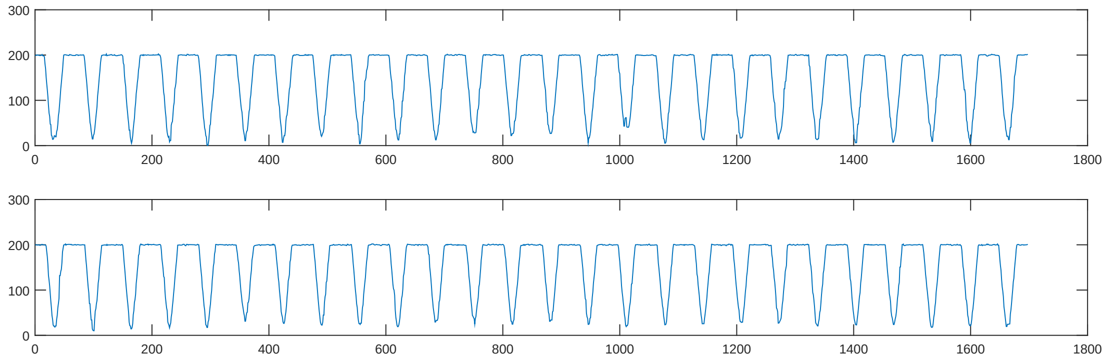


Figure 4.18 Simulated profiles for worn tools

4.4.3 Classification

In this subsection, classifiers are designed for tool condition classification, and leave-one-out cross-validation (LOOCV) is applied to evaluate the performance. Candidate classifiers include linear classifier (Fisher, 1936), quadratic classifier (Zhang, 1997), and support vector machine (SVM) (Suykens and Vandewalle, 1999). When the data sample size is limited, cross-validation is commonly used for evaluating and comparing learning algorithms (Shao et al., 2013). LOOCV proposes to partition the data into two sets: one set only includes one observation and is used for model validation, and the remaining observations are used for model training. By repeating this partition for all observations, LOOCV is able to predict and compare the performance of different learning algorithms.

In order to ensure satisfactory welding quality, the battery plant applies a conservative tool replacement strategy, i.e., the tools are always replaced before the degradation can affect the weld quality. Therefore, it is very difficult to collect coupons that reflect truly worn tools' surface profiles from production. To obstacle this issue, 8 coupon surfaces are simulated by truncating the knurl peaks from new or slightly worn surfaces, and examples of simulated profiles are shown by Figure 4.18.

Table 4.3 Misclassification rates of different classifiers

	Linear	Quadratic	SVM
Training error rate	15.63% (10/64)	1.56% (1/64)	15.63% (10/64)
CV error rate	15.63% (10/64)	3.13% (2/64)	18.75% (12/64)

In this research, 64 coupons, including 38 new coupons (Class 1), 18 slightly worn coupons (Class 2), and 8 simulated completely worn coupons (Class 3) are used for classifier training and LOOCV. The comparison of linear classifier, quadratic classifier, and SVM is shown by Tables 4.3 and 4.4.

From Tables 4.3 and 4.4, it is noticed that the quadratic classifier achieves the best performance, and the corresponding cross-validation misclassification rate is 3.13%. Additionally, all coupons in Class 3 are classified correctly, and no Class 1 or Class 2 coupons are misclassified as Class 3, implying that the quadratic classifier is able to accurately distinguish Class 3 from the other two classes. On the other hand, the linear classifier and SVM have significantly larger misclassification rates, and neither of them is able to completely distinguish Class 3 from the other two classes. Consequently, for this application, the quadratic classifier is able to achieve satisfactory classification performance.

Note to Practitioners: When applying this tool wear monitoring framework on the plant floor, it is recommended to adopt a more efficient and cost-effective measurement system to measure the coupons. As shown by the previous results in this section, cross-sectional profiles along the horizontal direction contain rich information on the tool degradation level. Therefore, one potential measurement system may be a line scanner with sufficient resolu-

Table 4.4 Confusion matrices for cross-validation

(a) Linear Classifier

		Predicted class		
		Class 1	Class 2	Class 3
True class	Class 1	76.32% (29/38)	2.63% (1/38)	21.05% (8/38)
	Class 2	5.56% (1/18)	94.44% (17/18)	0
	Class 3	0	0	100% (8/8)

(b) Quadratic Classifier

		Predicted class		
		Class 1	Class 2	Class 3
True class	Class 1	97.37% (37/38)	2.63% (1/38)	0
	Class 2	5.56% (1/18)	94.44% (17/18)	0
	Class 3	0	0	100% (8/8)

(c) SVM

		Predicted class		
		Class 1	Class 2	Class 3
True class	Class 1	68.42% (26/38)	5.26% (2/38)	26.32% (10/38)
	Class 2	0	100% (18/18)	0
	Class 3	25.00% (2/8)	0	75.00% (6/8)

tions, the measurement of which is generally completed within minutes. In this way, the tool condition monitoring can be conducted in a more timely manner.

4.5 Conclusion

Tool wear characterization and monitoring for ultrasonic welding of lithium-ion batteries have been investigated in this paper. By comparing tool surface measurements at different tool life stages, tool wear is characterized by four stages using changes in the knurl geometry. A novel impression method is then developed to efficiently and accurately obtain tool surface profiles without introducing significant interruption to plant production. Finally, an effective monitoring algorithm is developed using a quadratic classifier and features that are extracted from space and frequency domains of cross-sectional profiles on tool surfaces.

This study enhances our understanding of the tool wear mechanism in ultrasonic metal welding, and the algorithm can be utilized to accurately identify tool conditions, leading to decreased production costs while ensuring good joining quality in battery manufacturing.

Chapter 5

Conclusion and Future Work

5.1 Conclusion and Contributions

Characterization and monitoring of spatiotemporal processes have become increasingly important for quality evaluation and improvement in manufacturing. This dissertation develops novel methods for efficient sampling, cost-effective surface variation modeling, and effective process monitoring. The major achievements and contributions of this dissertation are summarized as follows.

(1) *Dynamic sampling design for characterizing spatiotemporal processes in manufacturing.* In this study, we first formulate the dynamic sampling design problem for spatiotemporal tooling degradation processes. Then a new design criterion is defined by simultaneously considering the prediction precision and the measurement costs. A weight matrix is introduced to account for the spatially varied precision requirement. The design search process is formulated as a binary integer programming problem, and the genetic algorithm is adopted for solution search. Moreover, an innovative algorithm is devised to monitor and update the temporal transition parameter in the spatiotemporal model. A comparative case study is conducted to demonstrated the effectiveness of the developed framework.

(2) *Cost-effective surface variation modeling by integrating an engineering model with*

multi-task learning. The proposed model decomposes spatial variation into a global trend and a local variation part. The global trend is induced by process settings, and it is modeled using an engineering model. The local variation part is modeled as a zero mean Gaussian process, which shares commonalities among multiple similar-but-non-identical surfaces, and it is estimated using multi-task learning. An iterative algorithm is developed to estimate the model parameters in the global trend and the local variation part. A case study using engine deck surface data from an automobile plant has demonstrated that the developed approach outperforms those existing methods, including simple kriging, kriging with external drift, and data-driven Gaussian process multi-task learning. Additionally, the effects of number of tasks, sample size, and correlation strength between process variables and surface height have been discussed.

(3) *Tool wear monitoring of ultrasonic metal welding in lithium-ion battery manufacturing.* A framework for tool condition monitoring of ultrasonic metal welding is developed for the first time. Tool wear progression is first characterized using optical images and surface height measurement of welding tools in various wear stages. A new impression method is then developed to efficiently obtain tool surface profiles. Based on expert knowledge, monitoring features are generated from both the spatial and temporal domains of high-resolution spatial data. Classification algorithms are designed for tool condition determination using the extracted monitoring features. Measurement data from a real-world electric vehicle battery manufacturing plant is used to validate the developed algorithms.

5.2 Suggested Future Work

Future research efforts are proposed in the following several directions.

(1) *Spatial and temporal dynamic sampling design for spatiotemporal processes*: Traditional dynamic sampling design methods for spatiotemporal processes often assume the temporal sampling interval to be constant. For instance, in environmental and ecological monitoring, measurement is conducted daily, monthly, or annually; in tool condition monitoring of ultrasonic metal welding, tool surfaces are measured every a certain number of welds. Nevertheless, constant sampling interval may not be optimal. Further improvement can be achieved by adaptively varying the sampling interval according to the spatiotemporal progression pattern.

(2) *Data fusion for spatiotemporal modeling in manufacturing*: The availability of multi-source data, including process variables, functional curves, and spatiotemporal data, provides a good opportunity for further improving process monitoring and control. However, modeling the cross-correlations among different data sources is very complicated, and there is a lack of systematic modeling methods. It is very promising to integrate process monitoring and control in one framework based on data fusion. For example, for the ultrasonic welding process in Figure 1.1, the welding quality monitoring algorithm should be adaptive to the tool degradation; tool condition monitoring can be developed by incorporating tool surface measurement and the monitoring signals; process parameters, such as pressure, amplitude, and energy, are supposed to be adjusted based on tool degradation. Through this integration, quality monitoring performance will be improved; cost-effective tool

condition monitoring can be enabled; the welding quality will be robust to tool degradation; and tool life may be extended to reduce production costs.

(3) *Variable and task selection for fusion model with multi-task learning*: Variables reflecting multidisciplinary process information may take various types and formats. These variables may not be equally important or they may contain duplicate information. Thus, a systematic variable selection algorithm is necessary to select a parsimonious subset of all candidate variables for the engineering model. Moreover, task selection is one important topic in multi-task learning. As shown by Figure 3.13, adding more tasks is not always able to improve the prediction performance, but the computational load will be significantly increased. Therefore, a task selection approach is desirable for the model developed in Chapter 3.

(4) *Multi-level tool wear model development for ultrasonic metal welding*: In Chapter 4, a tool condition monitoring system is developed to classify a tool into three categories: new, early worn, and worn. However, this system is not able to predict remaining tool life or estimate when to conduct tool replacement. Hence, a continuous tool life model is highly desirable for quantification of the tool wear level and prediction of remaining tool life. A knurl-based spatiotemporal model may be developed using knurl-level features, e.g., average knurl height, knurl volume, and shoulder width. The model can predict the progression of knurl features, and further remaining tool life. Additionally, this model will be able to reveal abnormal degradation progression patterns, assisting process diagnosis and control. For instance, tool misalignment is a common issue in ultrasonic welding and may greatly reduce tool life. The misalignment often yields nonuniform degradation patterns,

and it may be effectively detected using the spatiotemporal model at an early stage. Then tool adjustment can be conducted accordingly.

Appendices

Appendix A

Hypothesis Testing for Monitoring the Temporal Transition Parameter

$u_{t-1,(t-1)t}$ and $u_{t,(t-1)t}$ in Equation (2.35) can be rewritten as Equations (A.1) and (A.2).

$$u_{t-1,(t-1)t} = \frac{1}{m_1 + m_{2,(t-1)t}} \left(\text{sum}\left(\mathbf{M}_{1,t-1}\right) + \text{sum}\left(\mathbf{M}_{2,(t-1)t}^{t-1}\right) \right), \tag{A.1}$$

$$u_{t,(t-1)t} = \frac{1}{m_1 + m_{2,(t-1)t}} \left(\text{sum}\left(\mathbf{M}_{1,t}\right) + \text{sum}\left(\mathbf{M}_{2,(t-1)t}^{t}\right) \right), \tag{A.2}$$

where $\text{sum}(\cdot)$ is a summation function; m_1 is the number of Level 1 measurement points, and it is time-invariant; and $m_{2,(t-1)t}$ is the number of measurement locations which are measured by Level 2 at both time $t-1$ and t.

It is known that the measurement error variances of Level 1 and Level 2 are σ_1^2 and σ_2^2, respectively, according to Assumption 2.2. Therefore, we can assume the Level 1 measurement points are independent normal random variables with mean as the true height and variance as the measurement error variance, σ_1^2. Similarly, Level 2 measurement points are also independent normal random variables with mean as the true height and variance σ_2^2.

Both $u_{t-1,(t-1)t}$ and $u_{t,(t-1)t}$ are normal random variables. Their variances are equal and determined by the repeatability of Level 1 and Level 2 measurements. We denote their variances as

$$\sigma_{\varepsilon,(t-1)t}^2 \equiv \text{Var}\left(u_{t-1,(t-1)t}\right) = \text{Var}\left(u_{t,(t-1)t}\right). \tag{A.3}$$

$\sigma^2_{\varepsilon,(t-1)t}$ can by calculated using the following equation.

$$\begin{aligned}\sigma^2_{\varepsilon,(t-1)t} =& \frac{1}{\left(m_1+m_{2,(t-1)t}\right)^2}\left(m_1\sigma_1^2+m_{2,(t-1)t}\sigma_2^2\right) \\ =& \frac{m_1\sigma_1^2}{\left(m_1+m_{2,(t-1)t}\right)^2}+\frac{m_{2,(t-1)t}\sigma_2^2}{\left(m_1+m_{2,(t-1)t}\right)^2}.\end{aligned} \tag{A.4}$$

Define $\hat{u}_{t,(t-1)t}$ as the estimation of $u_{t,(t-1)t}$ using $u_{t-1,(t-1)t}$ and $\hat{\phi}_{t-1}$, as shown by Equation (A.5).

$$\hat{u}_{t,(t-1)t} = \hat{\phi}_{t-1}u_{t-1,(t-1)t}. \tag{A.5}$$

A statistic is further defined as the difference between $u_{t,(t-1)t}$ and $\hat{u}_{t,(t-1)t}$:

$$\Delta u_{(t-1)t} = u_{t,(t-1)t} - \hat{u}_{t,(t-1)t}. \tag{A.6}$$

Under H_0: $\hat{\phi}_{t-1} = h_t$ in Test 2.1,

$$\begin{aligned}\mathrm{E}\left(\Delta u_{(t-1)t}\right) &= \mathrm{E}\left(u_{t,(t-1)t} - \hat{\phi}_{t-1}u_{t-1,(t-1)t}\right) \\ &= \mathrm{E}\left(h_t u_{t-1,(t-1)t} - \hat{\phi}_{t-1}u_{t-1,(t-1)t}\right) \\ &= \left(h_t - \hat{\phi}_{t-1}\right)\mathrm{E}\left(u_{t-1,(t-1)t}\right) \\ &= 0,\end{aligned} \tag{A.7}$$

$$\begin{aligned}\mathrm{Var}\left(\Delta u_{(t-1)t}\right) &= \mathrm{Var}\left(u_{t,(t-1)t} - \hat{\phi}_{t-1}u_{t-1,(t-1)t}\right) \\ &= \mathrm{Var}\left(u_{t,(t-1)t}\right) + \hat{\phi}^2_{t-1}\mathrm{Var}\left(u_{t-1,(t-1)t}\right) \\ &= \left(1+\hat{\phi}^2_{t-1}\right)\sigma^2_{\varepsilon,(t-1)t}.\end{aligned} \tag{A.8}$$

Consequently,

$$\Delta u_{(t-1)t} \sim \mathscr{N}\left(0,\ \left(1+\hat{\phi}^2_{t-1}\right)\sigma^2_{\varepsilon,(t-1)t}\right). \tag{A.9}$$

Consequently, we can test H_0 using two-tailed hypothesis test for normal distributions and reject H_0 if the following condition holds.

$$\left| \frac{\Delta u_{(t-1)t}}{\sqrt{\left(1+\hat{\phi}_{t-1}^{2}\right)\sigma_{\varepsilon,(t-1)t}^{2}}} \right| > Z_{\alpha/2}. \tag{A.10}$$

□

Appendix B

Multi-Task Gaussian Process Learning Algorithm

The estimates of $\boldsymbol{\Theta} = \left\{\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha, \sigma^2\right\}$ can be obtained using the following EM algorithm (Yu et al., 2005). Detailed derivation of the algorithm is available in Yu et al. (2005).

- *E-step*: Estimate the expectation and covariance of $\boldsymbol{\alpha}$, $l = 1, \ldots, m$, given the current $\boldsymbol{\Theta}$.

  $$\hat{\boldsymbol{\alpha}}^l = \left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^\intercal \boldsymbol{\kappa}_l + \mathbf{C}_\alpha^{-1}\right)^{-1} \left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^\intercal \eta_l + \mathbf{C}_\alpha^{-1}\boldsymbol{\mu}_\alpha\right), \tag{B.1}$$

  $$\mathbf{C}_{\alpha^l} = \left(\frac{1}{\sigma^2}\boldsymbol{\kappa}_l^\intercal \boldsymbol{\kappa}_l + \mathbf{C}_\alpha^{-1}\right)^{-1}, \tag{B.2}$$

  where $\boldsymbol{\kappa}_l \in \mathbb{R}^{n_l \times n}$ is the base kernel $\kappa(\cdot,\cdot)$ evaluated between $\mathbf{S}_l$ and $\mathbf{S}$.

- *M-step*: Optimize $\boldsymbol{\Theta} = \left\{\boldsymbol{\mu}_\alpha, \mathbf{C}_\alpha, \sigma^2\right\}$.

  $$\boldsymbol{\mu}_\alpha = \frac{1}{\pi + m}\sum_{l=1}^{m} \hat{\boldsymbol{\alpha}}^l, \tag{B.3}$$

  $$\mathbf{C}_\alpha = \frac{1}{\tau + m}\left\{\pi \boldsymbol{\mu}_\alpha \boldsymbol{\mu}_\alpha^\intercal + \tau \boldsymbol{\kappa}^{-1} + \sum_{l=1}^{m} \mathbf{C}_{\alpha^l} + \sum_{l=1}^{m}\left[\hat{\boldsymbol{\alpha}}^l - \boldsymbol{\mu}_\alpha\right]\left[\hat{\boldsymbol{\alpha}}^l - \boldsymbol{\mu}_\alpha\right]^\intercal\right\}, \tag{B.4}$$

  $$\sigma^2 = \frac{1}{\sum_{l=1}^{m} n_l}\sum_{l=1}^{m} \|\boldsymbol{\eta}_l - \boldsymbol{\kappa}\hat{\boldsymbol{\alpha}}_l\|^2 + \mathrm{tr}\left[\boldsymbol{\kappa}_l \mathbf{C}_{\alpha^l} \boldsymbol{\kappa}_l^\intercal\right], \tag{B.5}$$

  where $\mathrm{tr}(\cdot)$ is the trace operator.

Bibliography

Abellan-Nebot, J. V. and Subirón, F. R. A review of machining monitoring systems based on artificial intelligence process models. *The International Journal of Advanced Manufacturing Technology*, 47(1-4):237–257, 2010.

Anderson, A., Wang, G., and Gertner, G. Local variability based sampling for mapping a soil erosion cover factor by co-simulation with landsat tm images. *International Journal of Remote Sensing*, 27(12):2423–2447, 2006.

Atkinson, P., Webster, R., and Curran, P. Cokriging with ground-based radiometry. *Remote Sensing of Environment*, 41(1):45–60, 1992.

Atkinson, P., Webster, R., and Curran, P. Cokriging with airborne mss imagery. *Remote Sensing of Environment*, 50(3):335–345, 1994.

Attaway, S. *Matlab: a practical introduction to programming and problem solving*. Butterworth-Heinemann, 2013.

Bonilla, E. V., Chai, K. M., and Williams, C. Multi-task gaussian process prediction. In *Advances in neural information processing systems*, pages 153–160, 2007.

Bourennane, H., King, D., and Couturier, A. Comparison of kriging with external drift and simple linear regression for predicting soil horizon thickness with different sample densities. *Geoderma*, 97(3):255–271, 2000.

Brooks, S. P. and Morgan, B. J. Optimization using simulated annealing. *The Statistician*, pages 241–257, 1995.

Brus, D. and De Gruijter, J. Random sampling or geostatistical modelling? choosing between design-based and model-based sampling strategies for soil (with discussion). *Geoderma*, 80(1):1–44, 1997.

Burer, S. and Letchford, A. N. Non-convex mixed-integer nonlinear programming: a survey. *Surveys in Operations Research and Management Science*, 17(2):97–106, 2012.

Byrne, G., Dornfeld, D., Inasaki, I., Ketteler, G., König, W., and Teti, R. Tool condition monitoring (tcm)–the status of research and industrial application. *CIRP Annals-Manufacturing Technology*, 44(2):541–567, 1995.

Caruana, R. Multitask learning. *Machine Learning*, 28(1):41–75, 1997.

Cook, N. H. Tool wear and tool life. *Journal of Manufacturing Science and Engineering*, 95(4):931–938, 1973.

Curran, P. and Williamson, H. Sample size for ground and remotely sensed data. *Remote sensing of environment*, 20(1):31–41, 1986.

Curran, P. J. The semivariogram in remote sensing: an introduction. *Remote Sensing of Environment*, 24(3):493–507, 1988.

Davis, R. A., Huang, D., and Yao, Y.-C. Testing for a change in the parameter values and order of an autoregressive model. *The Annals of Statistics*, pages 282–304, 1995.

D'Errico, J. R. Surface fitting using gridfit, 2005. URL `http://www.mathworks.com/matlabcentral/fileexchange/8998-surface-fitting-using-gridfit`.

Dimla, D. E. Sensor signals for tool-wear monitoring in metal cutting operationsa review of methods. *International Journal of Machine Tools and Manufacture*, 40(8):1073–1098, 2000.

Duda, R. O., Hart, P. E., and Stork, D. G. *Pattern classification*. John Wiley & Sons, 2012.

Ertunc, H. M., Loparo, K. A., and Ocak, H. Tool wear condition monitoring in drilling operations using hidden markov models (hmms). *International Journal of Machine Tools and Manufacture*, 41(9):1363–1384, 2001.

Exler, O., Antelo, L. T., Egea, J. A., Alonso, A. A., and Banga, J. R. A tabu search-based algorithm for mixed-integer nonlinear problems and its application to integrated process and control system design. *Computers & Chemical Engineering*, 32(8):1877–1891, 2008.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. Generalized linear models. In *Regression*, pages 269–324. Springer, 2013.

Fanshawe, T. R. and Diggle, P. J. Adaptive sampling design for spatio-temporal prediction. *Spatio-Temporal Design: Advances in Efficient Data Acquisition*, pages 249–268, 2013.

Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2):179–188, 1936.

Gilbert, J. R., Moler, C., and Schreiber, R. Sparse matrices in matlab: design and implementation. *SIAM Journal on Matrix Analysis and Applications*, 13(1):333–356, 1992.

Gombay, E. Change detection in autoregressive time series. *Journal of Multivariate Analysis*, 99(3):451–464, 2008.

Gombay, E. and Serban, D. Monitoring parameter change in time series models. *Journal of Multivariate Analysis*, 100(4):715–725, 2009.

Grove, D. M., Woods, D. C., and Lewis, S. M. Multifactor b-spline mixed models in designed experiments for the engine mapping problem. *Journal of Quality Technology*, 36 (4):380, 2004.

Guo, W., Shao, C., Kim, T. H., Jin, J. J., Hu, S. J., Wang, H., and Spicer, J. P. Integrating univariate control charts and mahalanobis distance for process monitoring with near zero type ii error. *Journal of Manufacturing Systems*, accepted.

Hamilton, J. D. *Time series analysis*, volume 2. Princeton university press Princeton, 1994.

Handcock, M. S. and Wallis, J. R. An approach to statistical spatial-temporal modeling of meteorological fields. *Journal of the American Statistical Association*, 89(426):368–378, 1994.

Harrison, J. and West, M. *Bayesian Forecasting & Dynamic Models*. Springer, 1999.

Harvill, J. L. Spatio-temporal processes. *Wiley interdisciplinary reviews: computational statistics*, 2(3):375–382, 2010.

Holland, P. W. and Welsch, R. E. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods*, 6(9):813–827, 1977.

Hooten, M. B., Wikle, C. K., Sheriff, S. L., and Rushin, J. W. Optimal spatio-temporal hybrid sampling designs for ecological monitoring. *Journal of Vegetation Science*, 20 (4):639–649, 2009.

Huang, H.-C. and Cressie, N. Spatio-temporal prediction of snow water equivalent using the kalman filter. *Computational Statistics & Data Analysis*, 22(2):159–175, 1996.

Huang, S., Li, J., Chen, K., Wu, T., Ye, J., Wu, X., and Yao, L. A transfer learning approach for network modeling. *IIE Transactions*, 44(11):915–931, 2012.

Hudson, G. and Wackernagel, H. Mapping temperature using kriging with external drift: theory and an example from scotland. *International journal of Climatology*, 14(1):77–91, 1994.

Jantunen, E. A summary of methods applied to tool condition monitoring in drilling. *International Journal of Machine Tools and Manufacture*, 42(9):997–1010, 2002.

Jin, R., Chang, C.-J., and Shi, J. Sequential measurement strategy for wafer geometric profile estimation. *IIE Transactions*, 44(1):1–12, 2012.

Juang, J.-N. Applied system identification. 1994.

Jung, H. and Kim, K. A new parameterisation method for nurbs surface interpolation. *The International Journal of Advanced Manufacturing Technology*, 16(11):784–790, 2000.

Kang, J., Park, I., Jae, J., and Kang, S. A study on a die wear model considering thermal softening:(i) construction of the wear model. *Journal of Materials Processing Technology*, 96(1):53–58, 1999a.

Kang, J., Park, I., Jae, J., and Kang, S. A study on die wear model considering thermal softening (ii): application of the suggested wear model. *Journal of Materials Processing Technology*, 94(2):183–188, 1999b.

Kibria, B. G., Sun, L., Zidek, J. V., and Le, N. D. Bayesian spatial prediction of random space-time fields with application to mapping $pm_{2.5}$ exposure. *Journal of the American Statistical Association*, 97(457):112–124, 2002.

Kim, T., Yum, J., Hu, S., Spicer, J., and Abell, J. Process robustness of single lap ultrasonic welding of thin, dissimilar materials. *CIRP Annals-Manufacturing Technology*, 60(1): 17–20, 2011.

Kirkpatrick, S., Vecchi, M., et al. Optimization by simmulated annealing. *Science*, 220 (4598):671–680, 1983.

Kong, L. X. and Nahavandi, S. On-line tool condition monitoring and control system in forging processes. *Journal of Materials Processing Technology*, 125:464–470, 2002.

Koren, Y., Ko, T.-R., Ulsoy, A. G., and Danai, K. Flank wear estimation under varying cutting conditions. *Journal of dynamic systems, measurement, and control*, 113(2): 300–307, 1991.

Kurada, S. and Bradley, C. A machine vision system for tool wear assessment. *Tribology International*, 30(4):295–304, 1997a.

Kurada, S. and Bradley, C. A review of machine vision sensors for tool condition monitoring. *Computers in Industry*, 34(1):55–72, 1997b.

Kyriakidis, P. C. and Journel, A. G. Geostatistical space–time models: a review. *Mathematical geology*, 31(6):651–684, 1999.

Land, A. H. and Doig, A. G. An automatic method of solving discrete programming problems. *Econometrica: Journal of the Econometric Society*, pages 497–520, 1960.

Lange, K. L., Little, R. J., and Taylor, J. M. Robust statistical modeling using the t distribution. *Journal of the American Statistical Association*, 84(408):881–896, 1989.

Lanzetta, M. A new flexible high-resolution vision sensor for tool condition monitoring. *Journal of Materials Processing Technology*, 119(1):73–82, 2001.

Lawrence, N. D. and Platt, J. C. Learning to learn with the informative vector machine. In *Proceedings of the 21st international conference on Machine learning*, page 65. ACM, 2004.

Lee, S. S., Kim, T. H., Hu, S. J., Cai, W. W., Abell, J. A., and Li, J. Characterization of joint quality in ultrasonic welding of battery tabs. *Journal of Manufacturing Science and Engineering*, 135(2):021004, 2013.

Lee, S. S., Shao, C., Kim, T. H., Hu, S. J., Kannatey-Asibu, E., Cai, W. W., Spicer, J. P., and Abell, J. A. Characterization of ultrasonic metal welding by correlating online sensor signals with weld attributes. *Journal of Manufacturing Science and Engineering*, 136 (5):051019, 2014.

Lepadatu, D., Hambli, R., Kobi, A., and Barreau, A. Statistical investigation of die wear in metal extrusion processes. *The International Journal of Advanced Manufacturing Technology*, 28(3-4):272–278, 2006.

Ljung, L. *System identification*. Springer, 1998.

Luo, Y., Yuan, X., and Liu, Y. An improved pso algorithm for solving non-convex nlp/minlp problems with equality constraints. *Computers & chemical engineering*, 31(3):153–162, 2007.

Maronna, R., Martin, D., and Yohai, V. *Robust statistics*. John Wiley & Sons, Chichester. ISBN, 2006.

McBratney, A. and Webster, R. The design of optimal sampling schemes for local estimation and mapping of regionalized variables–ii: program and examples. *Computers & Geosciences*, 7(4):335–365, 1981.

McBratney, A. and Webster, R. Optimal interpolation and isarithmic mapping of soil properties. *Journal of Soil Science*, 34(1):137–162, 1983a.

McBratney, A. and Webster, R. How many observations are needed for regional estimation of soil properties? *Soil Science*, 135(3):177–183, 1983b.

McBratney, A., Webster, R., and Burgess, T. The design of optimal sampling schemes for local estimation and mapping of of regionalized variables–i: Theory and method. *Computers & Geosciences*, 7(4):331–334, 1981.

Mitchell, M. W. and Gumpertz, M. L. Spatio-temporal prediction inside a free-air co_2 enrichment system. *Journal of agricultural, biological, and environmental statistics*, 8(3): 310–327, 2003.

Na, O., Lee, Y., and Lee, S. Monitoring parameter change in time series models. *Statistical Methods & Applications*, 20(2):171–199, 2011.

Nelder, J. A. and Baker, R. Generalized linear models. *Encyclopedia of Statistical Sciences*, 1972.

Nguyen, H. T., Wang, H., and Hu, S. J. Characterization of cutting force induced surface shape variation in face milling using high-definition metrology. *Journal of Manufacturing Science and Engineering*, 135(4):041014, 2013.

Odeh, I. O., McBratney, A., and Chittleborough, D. Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. *Geoderma*, 67(3):215–226, 1995.

Pillonetto, G., Dinuzzo, F., Chen, T., De Nicolao, G., and Ljung, L. Kernel methods in system identification, machine learning and function estimation: A survey. *Automatica*, 50(3):657–682, 2014.

Rehorn, A. G., Jiang, J., and Orban, P. E. State-of-the-art methods and results in tool condition monitoring: a review. *The International Journal of Advanced Manufacturing Technology*, 26(7-8):693–710, 2005.

Ripley, B. D. *Spatial statistics*, volume 575. John Wiley & Sons, 2005.

Ryoo, H. S. and Sahinidis, N. V. Global optimization of nonconvex nlps and minlps with applications in process design. *Computers & Chemical Engineering*, 19(5):551–566, 1995.

Ryoo, H. S. and Sahinidis, N. V. A branch-and-reduce approach to global optimization. *Journal of Global Optimization*, 8(2):107–138, 1996.

Schlüter, M., Egea, J. A., and Banga, J. R. Extended ant colony optimization for non-convex mixed integer nonlinear programming. *Computers & Operations Research*, 36 (7):2217–2229, 2009.

Schwaighofer, A., Tresp, V., and Yu, K. Learning gaussian process kernels via hierarchical bayes. In *Advances in neural information processing systems*, pages 1209–1216, 2004.

Shao, C., Paynabar, K., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., Wang, H., and Abell, J. A. Feature selection for manufacturing process monitoring using cross-validation. *Journal of Manufacturing Systems*, 32(4):550–555, 2013.

Shao, C., Guo, W., Kim, T. H., Jin, J. J., Hu, S. J., Spicer, J. P., and Abell, J. A. Characterization and monitoring of tool wear in ultrasonic metal welding. In *9th International Workshop on Microfactories*, pages 161–169, 2014.

Shao, C., Kim, T. H., Hu, S. J., Jin, J. J., Abell, J. A., and Spicer, J. P. Tool wear monitoring for ultrasonic metal welding of lithium-ion batteries. *Journal of Manufacturing Science and Engineering*, 138(5):051005, 2016.

Stein, A. and Corsten, L. Universal kriging and cokriging as a regression procedure. *Biometrics*, pages 575–587, 1991.

Stein, A., Van Groenigen, J., Jeger, M., and Hoosbeek, M. Space-time statistics for environmental and agricultural related phenomena. *Environmental and Ecological Statistics*, 5(2):155–172, 1998.

Stein, M. L. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 2012.

Suriano, S., Wang, H., Shao, C., Hu, S. J., and Praveen, S. Progressive measurement and monitoring for multi-resolution data in surface manufacturing considering spatial and cross correlations. *IIE Transactions*, in press(ja):00–00, 2015. doi: 10.1080/0740817X.2014.998389. URL `http://dx.doi.org/10.1080/0740817X.2014.998389`.

Suykens, J. A. and Vandewalle, J. Least squares support vector machine classifiers. *Neural Processing Letters*, 9(3):293–300, 1999.

Taha, H. A. *Integer programming: theory, applications, and computations*. Academic Press, 2014.

Van Overschee, P. and De Moor, B. N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.

Wainer, H. and Thissen, D. Three steps towards robust regression. *Psychometrika*, 41(1): 9–34, 1976.

Waller, L. A., Carlin, B. P., Xia, H., and Gelfand, A. E. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical association*, 92(438):607–617, 1997.

Wang, G., Gertner, G., and Anderson, A. Sampling design and uncertainty based on spatial variability of spectral variables for mapping vegetation cover. *International Journal of Remote Sensing*, 26(15):3255–3274, 2005.

Wikle, C. K. Dynamic design of ecological monitoring networks for non-gaussian spatio-temporal data. *US Geological Survey*, 2004.

Wikle, C. K. and Royle, J. A. Space–time dynamic design of environmental monitoring networks. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 489–507, 1999.

Wikle, C. K., Berliner, L. M., and Cressie, N. Hierarchical bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154, 1998.

Wu, C. J. and Hamada, M. S. *Experiments: planning, analysis, and optimization*. John Wiley & Sons, 2011.

Xia, H., Ding, Y., and Wang, J. Gaussian process method for form error assessment using coordinate measurements. *IIE Transactions*, 40(10):931–946, 2008.

Xiao, X., Gertner, G., Wang, G., and Anderson, A. B. Optimal sampling scheme for estimation landscape mapping of vegetation cover. *Landscape Ecology*, 20(4):375–387, 2005.

Xu, K. and Wikle, C. K. Estimation of parameterized spatio-temporal dynamic models. *Journal of Statistical Planning and Inference*, 137(2):567–588, 2007.

Yang, B.-D. and Menq, C.-H. Compensation for form error of end-milled sculptured surfaces using discrete measurement data. *International Journal of Machine Tools and Manufacture*, 33(5):725–740, 1993.

Yang, T.-H. and Jackman, J. Form error estimation using spatial statistics. *Journal of Manufacturing Science and Engineering*, 122(1):262–272, 2000.

Young, C.-T., Zheng, Y., Yeh, C.-W., and Jang, S.-S. Information-guided genetic algorithm approach to the solution of minlp problems. *Industrial & engineering chemistry research*, 46(5):1527–1537, 2007.

Yu, K., Tresp, V., and Schwaighofer, A. Learning gaussian processes from multiple tasks. In *Proceedings of the 22nd international conference on machine learning*, pages 1012–1019. ACM, 2005.

Zhang, M. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proceedings of the National Academy of Sciences*, 94(2):565–568, 1997.

Zhou, J.-H., Pang, C. K., Zhong, Z.-W., and Lewis, F. L. Tool wear monitoring using acoustic emissions by dominant-feature identification. *Instrumentation and Measurement, IEEE Transactions on*, 60(2):547–559, 2011.

Zhu, X., Ding, H., and Wang, M. Y. Form error evaluation: An iterative reweighted least squares algorithm. *Journal of Manufacturing Science and Engineering*, 126(3):535–541, 2004.

Zhu, Z. and Stein, M. L. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(1):24–44, 2006.