Heterogeneous Sensor Data based Online Quality Assurance for Advanced Manufacturing using Spatiotemporal Modeling

Jia Liu

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

Doctor of Philosophy
In
Industrial and Systems Engineering

Zhenyu (James) Kong, Chair
Jaime A. Camelio
Ran Jin
Christopher B. Williams

07/10/2017
Blacksburg, VA

Keywords: Dirichlet process, recurrent hierarchical Dirichlet process, online process monitoring, spatial Dirichlet process, statistical control scheme, wafer profiles modeling, augmented point pattern, augmented spatiotemporal log Gaussian Cox process, porosity prediction

Heterogeneous Sensor Data based Online Quality Assurance for Advanced Manufacturing using Spatiotemporal Modeling

Jia Liu

ABSTRACT

Online quality assurance is crucial for elevating product quality and boosting process productivity in advanced manufacturing. However, the inherent complexity of advanced manufacturing, including nonlinear process dynamics, multiple process attributes, and low signal/noise ratio, poses severe challenges for both maintaining stable process operations and establishing efficacious online quality assurance schemes.

To address these challenges, four different advanced manufacturing processes, namely, fused filament fabrication (FFF), binder jetting, chemical mechanical planarization (CMP), and the slicing process in wafer production, are investigated in this dissertation for applications of online quality assurance, with utilization of various sensors, such as thermocouples, infrared temperature sensors, accelerometers, etc. The overarching goal of this dissertation is to develop innovative integrated methodologies tailored for these individual manufacturing processes but addressing their common challenges to achieve satisfying performance in online quality assurance based on heterogeneous sensor data. Specifically, three new methodologies are created and validated using actual sensor data, namely,

(1) Real-time process monitoring methods using Dirichlet process (DP) mixture model for timely detection of process changes and identification of different process states for FFF and CMP. The proposed methodology is capable of tackling non-Gaussian data from heterogeneous sensors in these advanced manufacturing processes for successful online quality assurance.

(2) Spatial Dirichlet process (SDP) for modeling complex multimodal wafer thickness profiles and exploring their clustering effects. The SDP-based statistical control scheme can effectively detect out-of-control wafers and achieve wafer thickness quality assurance for the slicing process with high accuracy.

(3) Augmented spatiotemporal log Gaussian Cox process (AST-LGCP) quantifying the spatiotemporal evolution of porosity in binder jetting parts, capable of predicting high-risk areas on consecutive layers. This work fills the long-standing research gap of lacking rigorous layer-wise porosity quantification for parts made by additive manufacturing (AM), and provides the basis for facilitating corrective actions for product quality improvements in a prognostic way.

These developed methodologies surmount some common challenges of advanced manufacturing which paralyze traditional methods in online quality assurance, and embody key components for implementing effective online quality assurance with various sensor data. There is a promising potential to extend them to other manufacturing processes in the future.

Heterogeneous Sensor Data based Online Quality Assurance for Advanced Manufacturing using Spatiotemporal Modeling

Jia Liu

GENERAL AUDIENCE ABSTRACT

This dissertation work develops novel online quality assurance methodologies for advanced manufacturing using various sensor data. Four advanced manufacturing processes, including fused filament fabrication, binder jetting, chemical mechanical planarization, and wafer slicing process, are investigated in this research. The developed methodologies address some common challenges in the aforementioned processes, such as nonlinear process dynamics and high variety in sensor data dimensions, which have severely hindered the effectiveness of traditional online quality assurance methods. Consequently, the proposed research accomplishes satisfying performance in defect detection and quality prediction for the advanced manufacturing processes.

In this dissertation, the research methodologies are constructed in both space and time domains based on different types of sensor data. Sensor data representation and integration for a variety of data formats (e.g., online data stream, profile data, image data) with the dimensionality covering a wide range (from ~10^0 to ~10^5) are researched to extract effective features that are sensitive to manufacturing process defects; the devised methods, based on the extracted features, utilize spatiotemporal analysis to realize timely detection and accurate prediction of process defects. These integrated methodologies have a promising potential to be extended to other advanced manufacturing processes for efficacious process monitoring and quality assurance.

The accomplished work in this dissertation is an effective effort towards sustainable operations of advanced manufacturing. The achieved performance not only enables improvement in defect

detection and quality prediction, but also lays the foundation for future implementation of corrective actions that can automatically mitigate the process defects.

DEDICATION

To all my family members for their unconditional love and support over these years.

ACKNOWLEDGMENTS

Over the past four years, Dr. James Kong has been a dedicated advisor and inspirational mentor, who wholeheartedly guided me through this challenging but rewarding journey. Without his supervision and guidance, my work could not have been accomplished with high quality. I would also like to thank my Ph.D. advisory committee, Dr. Jaime Camelio, Dr. Ran Jin, and Dr. Chris Williams, for their valuable suggestions and constant help with my research.

Throughout my graduate study at Virginia Tech, I have received help and encouragement from many other individuals. Dr. Prahalad Rao not only created innovative testbeds for my research, but also offered plenty of advice to me. David Roberson aided me in conducting experiments and proofread this dissertation meticulously. Yun Bai contributed invaluable process knowledge in our collaboration. My lab members, Dr. Kaveh Bastani, Chenang Liu and Babak Barazandeh, generously shared many inspiring insights in both research and life.

Finally, I am grateful for the support from the Grado Department of Industrial and Systems Engineering at Virginia Tech. It has been a delightful experience for me to work with every person here.

Table of Contents

1 Introduction 1

1.1 Motivation 1

1.2 Research gaps 2

1.3 Research objectives 3

1.4 Dissertation organization 4

2 Research Methodologies and Literature Review 5

2.1 Overall research methodologies 5

2.1.1 Task 1: Real-time process monitoring based on Dirichlet process mixture model 8

2.1.2 Task 2: Spatial Dirichlet process modeling based statistical control scheme for quality assurance with wafer thickness profiles 9

2.1.3 Task 3: Layer-wise porosity modeling and prediction for additive manufacturing (AM) using spatial point process 11

2.2 Literature review 12

2.2.1 Statistical process control (SPC) for non-Gaussian sensor data 12

2.2.2 Gaussian process (GP) and its applications in modeling wafer profiles 15

2.2.3 Porosity detection and quantification in products made by additive manufacturing (AM) 18

2.2.4 Point process analysis and its applications 19

3 Real-time Process Monitoring based on Dirichlet Process Mixture Model 21

3.1 Introduction 21

3.1.1 Fused filament fabrication (FFF) process 21

3.1.2 Chemical mechanical planarization (CMP) process 23

3.1.3 Real-time process monitoring methods based on DP mixture model 24

3.2 Research methodology 25

3.2.1 Dirichlet process (DP) mixture model 25

3.2.2 Multi-sensor real-time detection of process anomalies in FFF using Dirichlet process (DP) mixture model and evidence theory (ET) 27

3.2.3 Dirichlet process (DP)-based SPC for monitoring non-Gaussian data signals 31

3.2.4 Recurrent hierarchical Dirichlet process (RHDP) for evolutionary clustering of process states 33

3.3 Simulation studies 39

3.3.1 Application of DP-based SPC for process monitoring 39

3.3.2 RHDP clustering analysis for simulated data in sequential epochs 45

3.4 Application results 47

3.4.1 Application of integrated DP mixture model and ET approach in FFF process 47

3.4.2 Application of DP-based SPC and RHDP clustering in CMP process ... 52
3.5 Summary ... 58
4 Spatial Dirichlet Process Modeling based Statistical Control Scheme for Geometric Quality Assurance with Wafer Thickness Profiles ... 60
4.1 Introduction ... 60
4.2 Research methodology ... 63
4.2.1 Spatial mean fitting by Zernike polynomial regression ... 65
4.2.2 Spatial randomness modeling by spatial Dirichlet process (SDP) ... 66
4.2.3 Spatial Dirichlet process (SDP) based statistical control scheme ... 70
4.3 Simulations for spatial Dirichlet process modeling and statistical control ... 71
4.3.1 Complex spatial data modeling with spatial Dirichlet process ... 72
4.3.2 SDP-based statistical control scheme for complex spatial data ... 75
4.4 Application of spatial Dirichlet process approaches for wafer thickness profiles ... 80
4.4.1 Measurement sampling with profile data reduction ... 81
4.4.2 Measurement sampling with maximin Latin hypercube designs (LHD) ... 85
4.5 Summary ... 87
5 Spatiotemporal Modeling and Prediction of Layer-wise Porosity in Additive Manufacturing ... 89
5.1 Introduction ... 89
5.2 Research methodology ... 92
5.2.1 Introduction to spatiotemporal log Gaussian Cox process (ST-LGCP) modeling ... 94
5.2.2 Data representation by augmented point pattern ... 95
5.2.3 Augmented spatiotemporal log Gaussian Cox process (AST-LGCP) ... 96
5.2.4 Discretization of augmented spatiotemporal log Gaussian Cox process (AST-LGCP) ... 97
5.2.5 Parameter estimation in AST-LGCP ... 101
5.2.6 Porosity prediction for future layers in AST-LGCP ... 102
5.3 Application of AST-LGCP for prediction of porosity ... 103
5.3.1 Spatiotemporal analysis for numerically simulated porosity data ... 104
5.3.2 Application of AST-LGCP for quality assurance in binder jetting ... 107
5.4 Summary ... 112
6 Conclusions and Future Work ... 114
References ... 118
Appendix ... 127

List of Figures

Figure 2-1 The overall framework of proposed methodologies for online quality assurance. 5

Figure 2-2 For the investigated four advanced manufacturing processes, the dimensions of sensor data surges exponentially with the increasing complexity of the measurements, and their corresponding methodologies evolve from temporal analysis, to spatial analysis, then to spatiotemporal analysis. 7

Figure 2-3 Real-time process monitoring methodology based on DP mixture model. 9

Figure 2-4 Wafer profile modeling and geometric quality assurance are achieved by the proposed spatial Dirichlet process (SDP) and the SDP-based statistical control scheme. 10

Figure 2-5 Overall methodology of augmented spatiotemporal log Gaussian Cox process for layer-wise porosity modeling and prediction. 12

Figure 3-1 (a) Schematic of the FFF process. (b) Schematic of the FFF setup instrumented with multiple *in situ* sensors used in this work for measuring process conditions in real-time. 22

Figure 3-2 (a) (b) Buehler Automet® 250 experimental CMP setup with the integrated wireless sensor, (c) (d) near-specular CMP finished copper wafers. 24

Figure 3-3 The summary of the Dirichlet process (DP) mixture model and evidence theory (ET) approach developed in this work for real-time monitoring of process states in FFF. 28

Figure 3-4 Histogram with fitted probability density functions of IR sensor data at normal, abnormal, and failure process states. The signal patterns have a non-Gaussian probability density function, which is particularly pronounced for the normal state. 29

Figure 3-5 Overall methodology of RHDP for different process states clustering. 34

Figure 3-6 Possible evolutions of data distribution in a physical process. 35

Figure 3-7 True pdf, Gaussian components, and approximated distribution by DP mixture model for data generated from (a) $\mathbb{N}(\mu,1)$ and (b) χ_1^2. 40

Figure 3-8 Case N1 results - OOC ARL_1 values of different control charts when ARL_0 is fixed at 500, and actual IC data generated from (a) $\mathbb{N}(\mu,1)$ and (b) χ_1^2. Scale on the y-axis is in natural logarithm. 40

Figure 3-9 True pdf, Gaussian components and fitted distribution by DP model for data generated from a bimodal distribution consisting of $\mathbb{N}(10,1)$ and χ_1^2. 42

Figure 3-10 Case N2 results - OOC ARL_1 values of three SPC methods when ARL_0 is 500, and actual IC data are generated from a bimodal distribution consisting of $\mathbb{N}(10,1)$ and χ_1^2. Scale on the y-axis is in natural logarithm. 42

Figure 3-11 (a) The Rössler attractor delineated in (18), (b) a sample 1000 data points from the Rössler attractor contaminated with white noise $\mathbb{N}(\mathbf{0}, \boldsymbol{I}_3)$, (c) a new sample from the approximated distribution of the Rössler attractor. 43

Figure 3-12 Case N3 results - OOC ARL_1 values of three SPC methods when IC ARL_0 is 500; scale on the y-axis is in natural logarithm. (a) ARL_1 results for mean shifts, (b) ARL_1 detection for data variation σ^2. 44

Figure 3-13 (a) Generated three-part data from different Gaussian mixture distributions, (b) average log-likelihood values of data in time epochs. Three different shades indicate data from distributions D_1, D_2 and D_3, and the white areas between different parts of data are transition periods. 46

Figure 3-14 Comparison of F-scores for various sensor combinations using DP-ET approach. The error bars are one standard deviation (σ) long. 50

Figure 3-15 Comparison of F-scores for the three pre-defined process states in FFF using high-dimensional DP mixture modeling and our proposed DP-ET approach. The error bars are one standard deviation (σ) long. 51

Figure 3-16 Comparison of F-scores for the three pre-defined process states in FFF with four widely-used classification approaches. The error bars are one standard deviation (σ) long. 52

Figure 3-17 (a) Representative vibration signal patterns obtained under changing load conditions, (b) comparison of ARL_1 in changing load conditions. 54

Figure 3-18 (a) Representative vibration signal patterns obtained for pad wear experiments, (b) comparison of ARL_1 for pad wear. 55

Figure 3-19 (a) Representative vibration signal patterns obtained for pad degradation experiments, (b) comparison of ARL_1 for pad degradation. 56

Figure 3-20 Vibration data time series for the multiple process states (Case E3), including normal condition (C_1), high load (C_2) and low slurry (C_3). 57

Figure 4-1 An example of the thickness profile of a wafer in a slicing process; (b) The thickness profiles of twenty wafers in the slicing process show clear clustering phenomenon. 62

Figure 4-2 Overall methodology of the proposed SDP approaches under the spatial linear model structure. 64

Figure 4-3 The simulated two spatial random effects and the spatial observations generated from the two spatial random effects by adding pure errors. 73

Figure 4-4 Two designs of measured locations in the simulation: (a) 50 measured locations and (b) 100 measured locations. 73

Figure 4-5 The estimates of the number of spatial random effects, spatial variance σ^2, decay parameter ϕ and pure error variance τ^2 from 100 replicates for Case A1 (a-d), Case A2 (e-h), Case A3 (i-l), and Case A4 (m-p). The red vertical lines indicate the true values. 74

Figure 4-6 Out-of-control data detection efficiency by the AGP-based statistical control scheme and the proposed SDP-based statistical control scheme in terms of type II error for Case B1(a-d) and Case B2 (e-h). 77

Figure 4-7 Out-of-control data detection efficiency by the AGP-based statistical control scheme and the proposed SDP-based statistical control scheme in terms of type II error for Case B1(a-b) and Case B2 (c-d) under the scenario of no clustering phenomenon existing among spatial data. 79

Figure 4-8 (a) The spatial mean of 88 wafers obtained by Zernike polynomial regression; (b) the profile residuals of 20 wafers including spatial random effects (spatial deviation) and pure error (non-spatial deviation). 80

Figure 4-9 Top view and left-side view of eight wafer profile residuals. The data in the first row (a-d) belong to one cluster and both top view and left-side view show the homogeneity; the data in the second row (e-h) are all different. 81

Figure 4-10 The percentage of total variance explained by the variation among groups for each wafer. 74 wafers have more than 70% total variance explained by the data variation among groups. 82

Figure 4-11 Six clusters of profile residuals are identified by the proposed SDP modeling. Different clusters have different numbers of wafers. Within each cluster, the profile residuals are similar, centering on the spatial random effect with some pure errors. 84

Figure 4-12 The proposed SDP-based statistical control scheme correctly detects all the out-of-control (OOC) wafers with unusually high negative log likelihood values in one fold of cross-validation. 84

Figure 5-1 (a) A copper product fabricated by binder jetting. (b) A CT scan image with pores on the stem of the product. (c) Sequential zoomed images of pores on the stem. It is noticed that the pores on these sequential layers are dependent in space-time. 91

Figure 5-2 Overall methodology of AST-LGCP for layer-wise porosity modeling and prediction. 92

Figure 5-3 (a) A CT scan image from a metal part manufactured by binder jetting. Pores are shown as dark spots in this RGB image. (b) The visualization of augmented point pattern for this CT scan. Pores are represented by a set of spatial pairs with location and morphology. The pore in the circle occupies five pixels, and hence its information about size and form is captured. 95

Figure 5-4 (a) A CT scan image from a metal part manufactured by binder jetting. (b) The visualization of augmented point pattern for this CT scan. A pore in the blue circle has five pixels, and hence its information about size and form is captured by such data representation. (c) After the discretization, the pore in the blue circle occupies two cells ($D_{12,9}$ and $D_{12,10}$), therefore, is represented by the discrete format of augmented point pattern ($\tilde{Z}_{(12,9)} = 1$ and $\tilde{Z}_{(12,10)} = 4$). 98

Figure 5-5 One set of simulated data is shown as sequential images of pores with spatial and temporal correlations. The aim of this simulation is to predict the high-risk areas with occurrence of pores on the fifth layer based on the previous four layers. 104

Figure 5-6 For the set of simulated data in Figure 5-5, (a) spatial covariance of the pores along the distance, (b) temporal correlation of the pores among sequential layers, and (c) the predicted intensity function in the region of interest for the fifth layer. 105

Figure 5-7 (a) The simulated pores on the fifth layer; (b) overly segmented high-risk areas identified by the empirical approach (benchmark method); (c) small high-risk areas identified by ST-LGCP; and (d) relatively large and continuous high-risk areas identified by the proposed AST-LGCP. 106

Figure 5-8 (a) The print head of the ExOne R2 printer; (b) the sketch of binder jetting. 108

Figure 5-9 Sequential CT scan images on the stem region of the copper part made in binder jetting are selected for layer-wise porosity modeling and prediction. 109

Figure 5-10 The number of pores within the selected region on each layer is extracted from the CT scans of the copper product. The range of layers (layer 10 - layer 40) are used for layer-wise porosity prediction. 109

Figure 5-11 (a) The pores on the layer 22; (b) overly segmented high-risk areas identified by the empirical approach (benchmark method); (c) high-risk areas with low local intensity identified by ST-LGCP; and (d) high-risk areas with high local intensity identified by the proposed AST-LGCP. 110

Figure 5-12 The average F-score results for predicting total thirty layers (layer 11 - layer 40) with different numbers of previous layers by AST-LGCP. 111

Figure 5-13 The layer-wise predicted number of pores on each scanned layer in the specified region on the stem of the copper product by the empirical approach, ST-LGCP, and AST-LGCP. 112

List of Tables

Table 3-1 F-score results for data series with three distributions – comparison of RHDP clustering vs. mean shift (The values in the parenthesis are the standard deviation). 47

Table 3-2 Comparison of F-scores for the three pre-defined process states in FFF using our proposed DP-ET approach for various sensor combinations. The numbers in the parentheses indicate standard deviation over four cross-validation samples. Typically, the DP-ET approach using all three sensors, i.e., extruder vibration (V_E), table (or bed) vibration (V_T), and IR meltpool temperature (T_{IR}), demonstrates high precision and accuracy (97%) classification of the failure state. 49

Table 3-3 Comparison of F-scores for the three pre-defined process states in FFF using high-dimensional DP mixture modeling and our proposed DP-ET approach. The numbers in the parentheses indicate standard deviation over four cross-validation samples. Typically, DP-ET outperforms the traditional DP approach. 50

Table 3-4 Comparison of F-scores for the three pre-defined process states in FFF with four widely-used classification approaches. The numbers in the parentheses indicate standard deviation over four cross-validation samples. Our proposed DP-ET approach outperforms these popular methods in classifying the respective process states. 51

Table 3-5 Comparison of ARL_1 values for two predefined process anomalies with traditional SPC and DP-based SPC. The units are in milliseconds (ms). 56

Table 3-6 Clustering results for multiple process states in CMP experiment – comparison of RHDP clustering vs. mean shift (the values in the parenthesis are the standard deviation) 58

Table 4-1 The estimates of the parameters in the underlying model for the abovementioned four cases by using spatial Dirichlet process modeling. Mode and range are used to summarize the estimation. 74

Table 4-2 The coefficients of truncated Zernike polynomial basis for estimating the spatial mean of 88 wafers from the slicing process. 81

Table 4-3 The parameter estimation by spatial Dirichlet process model with spatial profile residuals at 30 measured locations. Mode and range are used to summarize the estimation. 83

Table 4-4 Type II error of testing the out-of-control (OOC) wafers - comparison of the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme with sampling from one-dimensional profiles. 85

Table 4-5 Type II error of testing the out-of-control (OOC) wafers - comparison of the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme with sampling from maximin Latin hypercube designs. 86

Table 5-1 Nomenclature and notations used in this work. 93

Table 5-2 Parameter estimates (variance σ^2, spatial scale parameter ϕ, temporal scale parameter θ in (42)) using the proposed AST-LGCP. The numbers in parenthesis are the standard deviation of the estimates (4000 iterations in MALA). 105

Table 5-3 F-score results for five sets of simulated data – Comparison of the empirical approach, ST-LGCP, and AST-LGCP (The values in the parenthesis are the standard deviation for five replications). 107

Table 5-4 F-score results for predicting total thirty layers (layer 11 - layer 40) – Comparison of the empirical approach, ST-LGCP, and AST-LGCP (The values in the parenthesis are the standard deviation). 111

1 Introduction

1.1 Motivation

Advanced manufacturing is transforming the manufacturing industries with insertion of various innovative technologies to fulfill the ever-growing demand for high quality products in every aspect of the modern world. For instance, one popular advanced manufacturing technology, additive manufacturing, has started to change the landscape of the manufacturing and service industries by producing highly complex and personalized products, far beyond the capability of traditional subtractive manufacturing.

However, the inherent complexity of advanced manufacturing also creates severe challenges for continuous process reliability and product consistency. For example, fused filament fabrication (FFF), a popular additive manufacturing process involving thermodynamics, can produce highly complex plastic parts by depositing melted plastic filaments in a layer-wise manner. If the process deviates from the normal process state, however, the manufactured part will suffer from various defects (such as overfilling, voids, air gaps, cracks, etc.). Chemical mechanical planarization (CMP), a back-end-of-line (BEOL) process in semiconductor manufacturing involving both corrosion and erosion, can polish a copper wafer to a near-optical (arithmetic average roughness, Sa ~ 5 nm) quality surface finish. If the process conditions are not optimized, the polished copper wafer would end up with a burnt surface full of scratches.

Therefore, online quality assurance is desired to improve the product quality and enhance the productivity of advanced manufacturing. Currently, the rapid development of sensor technology paves the way for online quality assurance by enabling measurements on different aspects of an

advanced manufacturing process, e.g., thermocouples and infrared sensors measure the temperature, and accelerometers detect the vibration, in FFF.

Four different advanced manufacturing processes, namely, fused filament fabrication (FFF), binder jetting, chemical mechanical planarization (CMP), and the slicing process in wafer production, are investigated in this dissertation to develop integrated online quality assurance methodologies by using heterogeneous sensor data. These methodologies will be effective in helping to maintain stable operation and improve product quality for advanced manufacturing.

1.2 Research gaps

As imperative as it is, online quality assurance for advanced manufacturing faces grave challenges to be effectively deployed into various processes due to their inherent complexity, such as nonlinear process dynamics, multiple process attributes, high noise in sensor signals, etc. Furthermore, there are some vital research gaps in the area of online quality assurance for advanced manufacturing:

1) The lack of effective online process monitoring schemes for advanced manufacturing with non-Gaussian signal data;
2) The inefficiency in sensor data usage for quality assurance due to the mismatch between the stagnant methods and the expanding capacity of data collection;
3) The absence of rigorous methods to quantify product quality in an online manner for parts produced by advanced manufacturing.

These common challenges and research gaps for online quality assurance are embodied in the four investigated manufacturing processes in this dissertation, and including the following.

(i) The sensor signals, with evident asymmetry and non-normality generated from FFF and CMP, prohibit traditional statistical process control (SPC) charts from efficacious real-time process monitoring.

(ii) The high-dimensional wafer profile data, with strong spatial non-normality, are oversimplified for quality assurance in current industrial standards and existing research, leading to inefficiency in detection of out-of-control wafers.

(iii) The insufficient quantification of porosity in additive manufactured parts, either by a porosity value or by simple visualization, hampers prognostic and corrective actions to improve the part quality.

In this dissertation, these aforementioned research gaps and challenges will be properly addressed by developing integrated online quality assurance methodologies for these manufacturing processes.

1.3 Research objectives

The overarching goal of this dissertation is to develop innovative methodologies tailored for aforementioned individual manufacturing processes but addressing their common challenges to achieve satisfying performance in online quality assurance based on heterogeneous sensor data. In pursuit of this goal, the objectives of this dissertation are concretized into three elemental tasks in different advanced manufacturing processes:

1) To design real-time process monitoring methodologies for FFF and CMP, surmounting challenges caused by non-Gaussian signal data and exceling at various scenarios of process change detection (Task 1);

2) To devise an effective statistical control scheme for complex spatial profile data, capable of discovering subtle spatial features in high-dimensional wafer thickness profiles and detecting aberrant wafers with high power of test for quality assurance (Task 2);
3) To develop a novel statistical modeling methodology to quantify the spatiotemporal evolution of porosity in AM parts and achieve accurate prediction of porosity in a layer-wise manner (Task 3).

The developed methods from these three tasks constitute integrated online quality assurance methodologies, contributing to product quality improvement and productivity increase for advanced manufacturing.

1.4 Dissertation organization

The rest of the dissertation is organized as follows: in Chapter 2 the integrated methodologies are outlined, followed by presenting related research and literature; real-time process monitoring methods based on Dirichlet process mixture model are developed in Chapter 3, adaptable to various scenarios of process change detection in FFF and CMP; a statistical control scheme based on spatial Dirichlet process modeling is proposed in Chapter 4 for strict quality control in a wafer slicing process with high-dimensional non-Gaussian wafer thickness profiles; augmented spatiotemporal log Gaussian Cox process is detailed in Chapter 5, constituting a rigorous layer-wise porosity quantification and prediction for online quality assurance in binder jetting; and lastly, the dissertation research is summarized in Chapter 6, and the future work built on this dissertation is envisioned.

2 Research Methodologies and Literature Review

2.1 Overall research methodologies

In this dissertation, the integrated online quality assurance methodologies are developed to improve the product quality and boost the productivity in advanced manufacturing. This work is summarized in a research framework ("Processes – Sensors – Data – Methods") as shown in Figure 2-1.

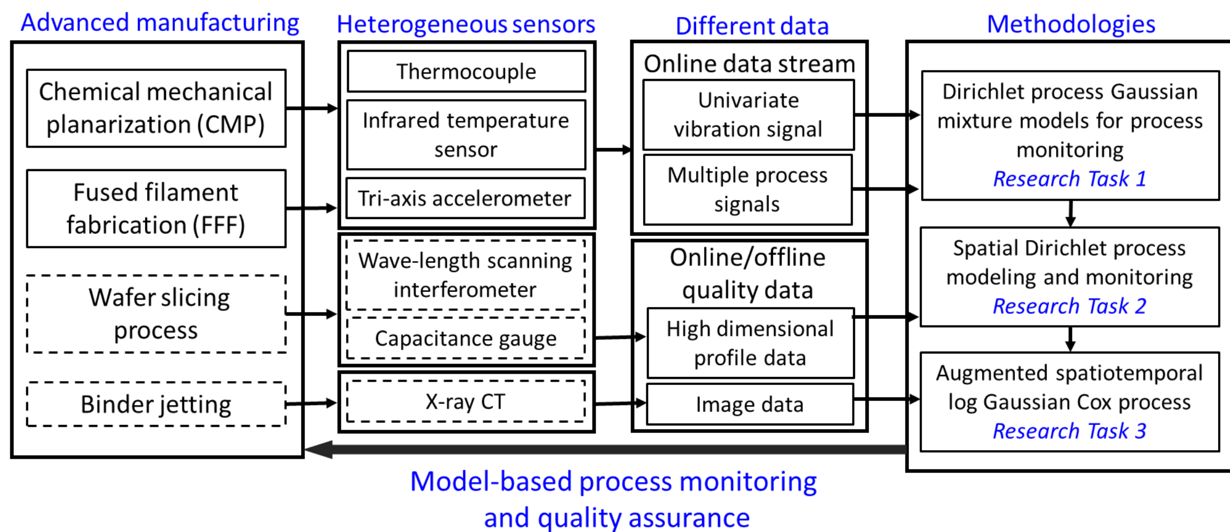

Figure 2-1 The overall framework of proposed methodologies for online quality assurance.

In Figure 2-1, four advanced manufacturing processes are inspected via different sensors and devices. Among them, chemical mechanical planarization (CMP) is a vital back-end-of-line (BEOL) process in semiconductor manufacturing, which aims to polish the blanket copper wafers to near-optical (arithmetic average roughness, Sa ~ 5 nm) quality surface finish; fused filament fabrication (FFF) is an additive manufacturing (AM) process in which an object is manufactured by depositing progressive layers of extruded molten material through a nozzle in a controlled manner [1-4]. Since they involve movements and heat transfer, vibration sensors (tri-axis accelerometers) and temperature sensors (thermocouples and infrared temperature sensors) are

installed on these two processes to acquire online data streams for process monitoring. The dimension of these data (i.e., the number of sensor channels) are not high (up to 13) due to the limitation of sensor installation, but these sensors can provide valuable data in the temporal domain.

Wafer slicing refers to the cutting of a silicon ingot into slices by using multiple-wire saws in wafer production with the aim to produce repeatable in-control slices in terms of geometric quality (e.g., thickness, taper, and bow) and surface quality (e.g., surface finish, surface damage). The geometric quality data can be measured by non-contact measuring methods (such as wave-length scanning interferometer and capacitance gauge), which provide a great amount of profile data (many thousands of measurements on a single wafer), viz., high-dimensional spatial data.

Binder jetting is one of the powder-based AM processes, which forms objects by using a liquid binding agent to selectively adhere powder layers. X-ray computed tomography (CT) is widely used to examine the internal structure of AM products in a layer-wise manner. CT-scanned images in the case study of a copper part fabricated by binder jetting have more than 100*100 pixels, which is very high in dimensionality ($10^4 \sim 10^5$). Moreover, its CT-scanned image data also possess spatial and temporal correlations due to the layer-wise printing.

With the increasing complexity and dimensionality of sensor data, the methodologies evolve from temporal analysis, to spatial analysis, then to spatiotemporal analysis (Figure 2-2). The first proposed methodology is real-time process monitoring for FFF and CMP based on Dirichlet process (DP) mixture model, which tackles non-Gaussian and noisy signals from heterogeneous sensors. By extending DP to spatial analysis, the second methodology (spatial Dirichlet process (SDP) modeling and statistical control scheme) is designed to model high-dimensional non-Gaussian wafer profiles and identify the aberrant wafers with high accuracy. Moreover, further

incorporating temporal correlation among data, the third methodology (augmented spatiotemporal log Gaussian Cox process (AST-LGCP)) is developed to investigate the spatiotemporal evolution of porosity in AM parts and predict the pore-prone areas in a layer-wise manner to achieve quality assurance during printing.

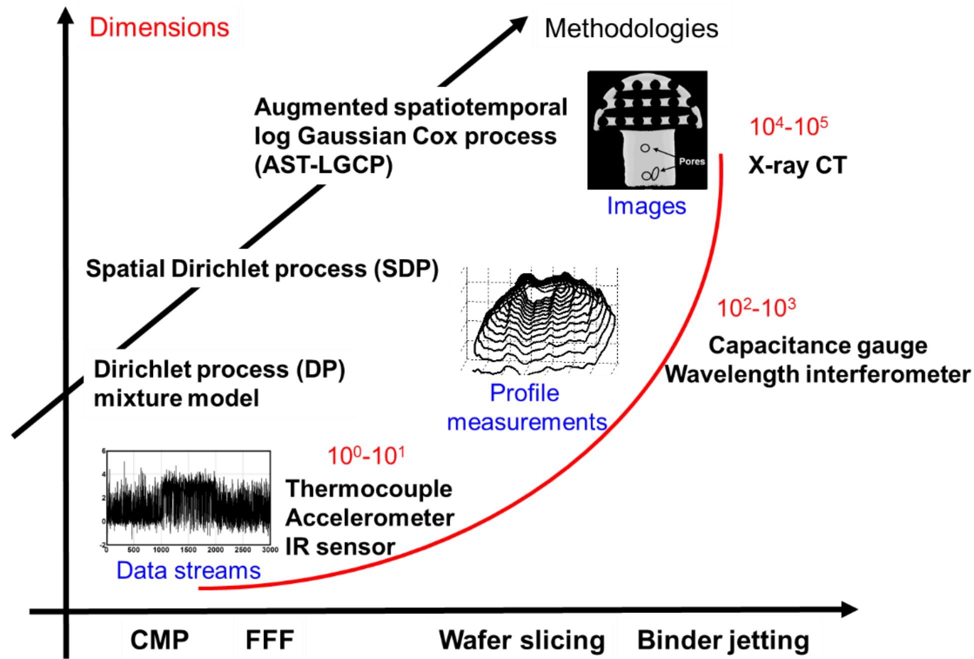

Figure 2-2 For the investigated four advanced manufacturing processes, the dimensions of sensor data surges exponentially with the increasing complexity of the measurements, and their corresponding methodologies evolve from temporal analysis, to spatial analysis, then to spatiotemporal analysis.

The developed integrated methodologies not only achieve superior performance in timely fault detection and accurate quality assurance for these aforementioned individual processes, but also conquer some common challenges of advanced manufacturing for efficacious online quality assurance. They have an encouraging potential to be extended to other advanced manufacturing processes in the future.

2.1.1 Task 1: Real-time process monitoring based on Dirichlet process mixture model

Due to the nonlinear dynamics of advanced manufacturing [5], the data distributions of sensor signals usually violate the Gaussian or symmetrical assumptions required by traditional statistical process control (SPC) methods. In order to accommodate the non-normality within the signal data, real-time process monitoring methodologies are proposed based on Dirichlet process (DP) mixture model.

The essential concept of these process monitoring methodologies is to represent a non-Gaussian probability distribution of signal data as a mixture of multiple Gaussian components. This implication can be stated mathematically as follows,

$$p(x) = \sum_{j=1}^{k} \pi_j \mathbb{N}(x|\theta_j), \tag{1}$$

where x represents a time series collected by sensors from the process, $p(x)$ is its data distribution, k is the number of Gaussian components $\mathbb{N}(\cdot)$ in the mixture, each of which is modeled with weight π_j and parameters θ_j (mean μ_j and variance ${\sigma_j}^2$). Dirichlet process (DP) mixture model is applied, which is a data-driven nonparametric Bayesian approach, to approximate a non-Gaussian distribution without any a *priori* knowledge of k [6].

Benefited from accurate approximation of data distribution by using DP mixture model, the real-time process monitoring concretizes into three different methods as shown in Figure 2-3. If the process states cannot be observed and labeled, (for instance, the polishing surface is impossible to be seen in CMP since it faces down against the polishing pad during the operation), recurrent hierarchical Dirichlet process (RHDP) clustering is developed to discover the transitions among different process states; if only two process states (e.g., in-control and out-of-control) exist, DP-based SPC can quickly detect deviations from the in-control process state; if more process states exist (for instance in FFF, three process states are observed: Normal, Abnormal, and Failure), the

DP-based evidence theoretic method (DP-ET) can achieve real-time process monitoring by accurately classifying the current signal data to the right process state.

These three methods for different monitoring scenarios compose of the proposed real-time process monitoring methodology based on Dirichlet process mixture model, which is detailed in Chapter 3.

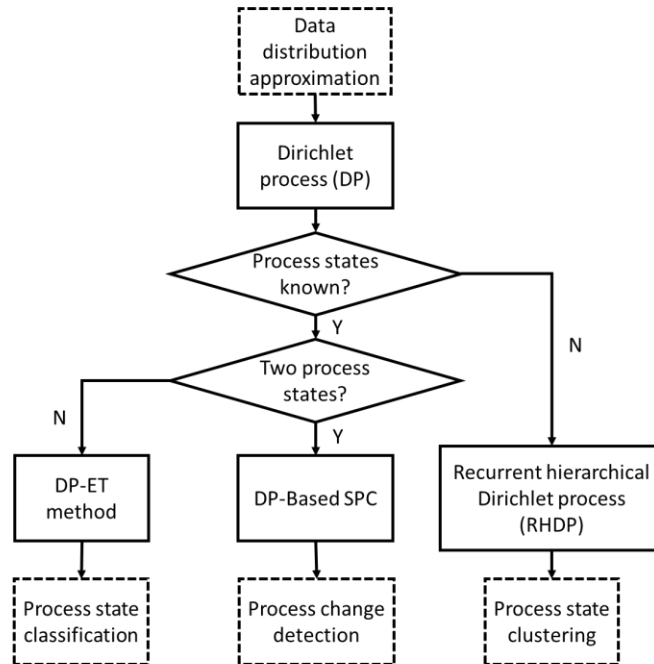

Figure 2-3 Real-time process monitoring methodology based on DP mixture model.

2.1.2 Task 2: Spatial Dirichlet process modeling based statistical control scheme for quality assurance with wafer thickness profiles

In order to achieve satisfying quality assurance for wafer thickness profiles from a slicing process and preventing defective wafers from being passed along to downstream production, a statistical control scheme based on spatial Dirichlet process is proposed.

The fundamental element, spatial Dirichlet process (SDP), extends the superior performance of DP mixture model for modeling multimodal data into spatial analysis, aiming for accurate spatial non-Gaussian data modeling by incorporating the clustering phenomenon among wafer profiles,

which is a research gap in the current literature. It decomposes the spatial profile data, implementing both deterministic modeling and stochastic modeling on them. The deterministic modeling is to estimate the spatial mean of the profile data, whereas the stochastic modeling is to model their randomness. The randomness can further be divided into spatial random effects (the spatial deviation) and pure error (the non-spatial deviation). It is the spatial variation that includes the non-normality within the profile data. Such profile decomposition as illustrated in Figure 2-4 systematically preserves sufficient information from the profiles. SDP further approximates the non-normality in the spatial variation by a mixture of spatial random effects. Consequently, the wafer profiles with the same spatial random effects can be clustered together due to their resemblance.

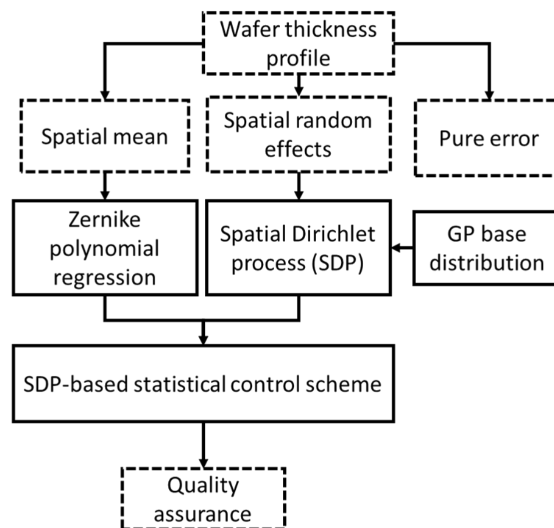

Figure 2-4 Wafer profile modeling and geometric quality assurance are achieved by the proposed spatial Dirichlet process (SDP) and the SDP-based statistical control scheme.

Built on the SDP modeling, the statistical control scheme based on SDP is created for more accurate out-of-control wafer detection than the current methods by adopting the learned clustering information existing in wafer thickness data. Specifically, it calculates the probability of new wafers belonging to the existing clusters of the in-control wafers. The wafers having small

probabilities within the existing clusters of the in-control wafers are deemed as out-of-control. With the presence of multiple clusters in the in-control wafers, the SDP-based statistical control scheme is expected to have high sensitivity and test power in detecting out-of-control wafers.

2.1.3 Task 3: Layer-wise porosity modeling and prediction for additive manufacturing (AM) using spatial point process

Instead of limiting the porosity quantification in AM parts to a single value or mere porosity visualization, a rigorous statistical method for layer-wise porosity modeling and prediction is proposed based on spatial point process analysis by using CT scanned images. The CT scanned images of porosity in AM parts possess valuable information of pores, such as their distribution characteristics (e.g., numbers and locations) and morphological features (e.g., sizes and forms). Moreover, due to the layer-wise production of AM parts, there are potential correlations among pores on each layer and across different layers. Therefore, a spatial point process analysis, specifically augmented spatiotemporal log Gaussian Cox process (AST-LGCP), is devised to extract both distribution characteristics and morphological features of pores from CT scan images of a binder jetting part, and explore the correlations among pores within each layer and across layers, for predicting pore-prone areas.

In AST-LGCP, augmented point pattern is designed to represent the pores on each layer using the morphological features and the distribution characteristics of pores from the CT scanned images for analysis. AST-LGCP, a flexible hierarchical spatial point process model, describes the augmented point pattern with a non-parametric intensity function in the interest of effective quantification of complex point patterns. By tracking the intensity function in space-time, AST-LGCP can model the spatiotemporal evolution of porosity and predict susceptible areas on future layers based on previous layers, facilitating corrective actions for quality improvements in a

proactive way. It can be generalized into different processes for layer-wise porosity modeling and prediction due to its data-driven nature. The overall framework of the proposed methodology is summarized in Figure 2-5:

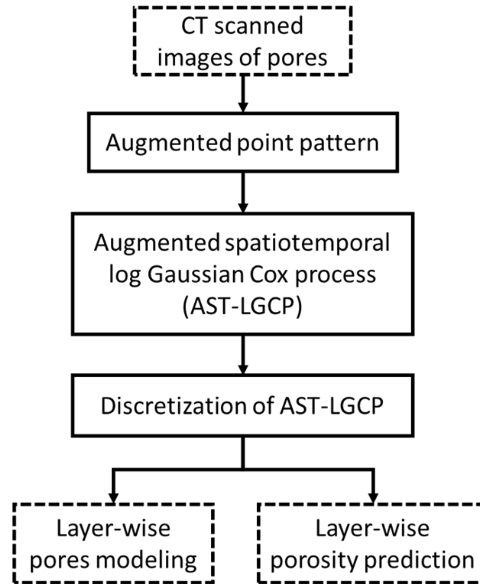

Figure 2-5 Overall methodology of augmented spatiotemporal log Gaussian Cox process for layer-wise porosity modeling and prediction.

2.2 Literature review

2.2.1 Statistical process control (SPC) for non-Gaussian sensor data

Traditional parametric SPC charts, such as Shewhart Xbar and R, cumulative sum (CUSUM), and exponentially weighted moving average (EWMA), have been widely used in various scenarios ranging from manufacturing to service industries for process improvement [7]. Despite the underlying normality and independence assumptions (NID) the effectiveness of Shewhart control charts have been attested; they are particularly useful for situations where sub-grouped measurements can be made and the process shifts are significant (> 1 standard deviation) [7]. CUSUM and EWMA control charts can be applied for both sub-grouped and individual measurements, and are particularly suited for detecting small drifts. However, the latter (EWMA) are not directionally invariant, i.e., the control chart has a certain inertia effect in reacting to process

drifts [7].

To overcome these restrictive assumptions with traditional parametric control charts, researchers devised nonparametric SPC charts, which are also called distribution-free SPC charts. Chakroborti *et al*. [8] provided a comprehensive review of nonparametric SPC charts. Although a specific type of distribution does not restrain nonparametric charts, nonetheless, most are based on data ranking methods, which entail that the data is implicitly assumed to be symmetric about the median.

To overcome this drawback, Qiu and Li [9] proposed a categorization-based nonparametric SPC chart for univariate data sequences. Their method relaxes the data symmetry assumption and is shown to be effective for non-Gaussian data. However, relying on a *priori* categorization of data for analysis results in information loss. Particularly, the selection of the number of groups for categorization, which is a heuristic parameter, is critical to the performance of control chart.

In another article, Qiu and Li [10] devised nonparametric SPC charts leveraging Gaussian transformations, i.e., transforming data belonging to an unknown distribution to approximately Gaussian. However, the normality of transformed data cannot be universally guaranteed for cases where the data is patently multimodal and complex, such as in the CMP vibration signals used in this work.

To overcome these challenges, researchers have explored wavelet and neural network-based SPC. These techniques can accommodate complex process dynamics, and have also been applied in CMP process [11]. Wavelet analysis has been successfully implemented in modeling and monitoring of functional data in advanced manufacturing [12]. For instance, Ganesan *et al*. [13] developed the wavelet-based SPC approach for real-time identification of delamination defects in CMP process. Guo *et al.* [14] presented an approach that uses wavelet coefficients in an SPC setting for detecting process drifts. Their method involves multi-scale decomposition of a signal

using a predetermined Harr wavelet basis function. Subsequently, they tracked the wavelet coefficients at a predetermined optimal (wavelet) level using CUSUM and EWMA control charts. Jeong *et al.* [15] described a similar wavelet-SPC procedure using the Symlet-8 wavelet basis function for functional data analysis of radio antenna reception patterns. Their approach uses a customized control chart with control limits derived from a statistic resembling the multivariate Hotelling's T^2 [7].

Pugh [16] showed that feed-forward neural networks (NN) has significantly lower type I and type II errors compared to traditional Shewhart X-bar and R charts, and therefore could be valuable for process monitoring applications. Subsequently, several researchers [17] have developed methods that employ neural networks (NN) for process monitoring applications. As an example of NN-based process monitoring, Rao *et al.* [18, 19] integrated a feedback-delay embedded recurrent neural network (RNN) with Bayesian particle filtering (PF) for real-time detection of mean shift in ultraprecision diamond turning process. The evolving surface morphology of diamond turned workpieces is predicted in real-time from *in situ* heterogeneous sensor data using PF-updated RNN weights. The network weights are subsequently monitored in an SPC setting using mean shift clustering [20].

Although these wavelet and NN-based SPC methods are applicable to complex signals without being constrained by the underlying assumptions of data distribution, they are nonetheless computationally demanding and engender a large number of variables that have to be tracked simultaneously. Moreover, these approaches require a predetermined model or basis function, such as the structure of the NN, and the basis and scaling function for wavelet decomposition. Therefore, decision uncertainty due to model selection remains a contentious challenge.

In contrast, SPC methods with Gaussian mixture modeling (GMM) overcome these

aforementioned data distribution and model selection limitations. In this context, Choi *et al.* [21] and Thissen *et al.* [22] proposed PCA-based monitoring techniques, where GMM-derived models constructed via EM algorithms are used to approximate the data pattern. Similarly, Chen *et al.* [23] utilized infinite Gaussian mixture models (IGMM) to construct the control chart. DP mixture model based methods [24, 25] were applied to identify process anomalies in fused filament fabrication (FFF) - an additive manufacturing process. Moreover, statistical process control methods dealing with high-dimensional massive sensor data for online process monitoring are also emerging [26, 27].

2.2.2 Gaussian process (GP) and its applications in modeling wafer profiles

A stochastic spatial model called Gaussian process (GP) is widely used in spatial statistics for spatial data analysis due to its easy formulation and affordable computation [28-30]. A Gaussian process represents a collection of random variables, which have a joint Gaussian distribution and can be completely specified by its mean and covariance functions. The covariance function encodes the spatial correlation among the data based on the distance (i.e., observations at closer locations are likely to be more correlated) with a few parameters. Therefore, it is very computationally efficient to apply Gaussian process to model spatial data. Different extension methods of Gaussian process are capable of solving even more challenging spatial data [30-35].

In the literature about wafer applications, research based on Gaussian process (GP) is prevalent in modeling the high-dimensional data in wafer profiles [36-42]. Interpolation for wafer profiles by these GP-based modeling methods can obtain the predicted thickness values at unmeasured locations, which are useful to evaluate the overall wafer quality. For instance, Jin *et al.* [36] proposed to use a Gaussian process model with a minimum number of thickness measurements to depict the entire geometric profile by sequentially measuring the wafer thickness profile. It reduces

the time and cost of evaluating the geometric quality of a whole wafer. Zhao *et al.* [38] devised a partial derivative equation (PDE) constrained Gaussian process model, which integrates physical knowledge of the slicing process into the modeling and prediction of a wafer thickness profile. Bao *et al.* [39] developed a hierarchical model to characterize wafer thickness profile by decomposing it into macro-scale and micro-scale components. The macro-scale component is modeled by a third-order polynomial function, while the micro-scale component is modeled as a first-order intrinsic Gaussian Markov random field (IGMRF) with the spatial correlation between the neighboring sites. Furthermore, researchers also take into account the circular geometry of the wafer during the modeling and interpolation. Pistone and Vicario [41] formulated a spatial interpolation approach for calculating values on unmeasured locations in circular grids for a wafer. Padonou and Roustant [40] proposed a polar Gaussian process for spatial interpolation and prediction of values in circular domains by representing spatial covariance with functions of angles and radius in polar coordinates.

However, when a large amount of profile data is easily available by non-contact measuring method, such as a wavelength scanning interferometer, or a capacitance gauge, the significance of interpolation at unmeasured locations by aforementioned wafer profile modeling methods is reduced. Moreover, these methods are not extended to statistical control to detect out-of-control wafers, since they are developed for modeling an individual wafer and lack the quantification of the variations among different wafers.

Plumlee *et al.* [42] developed a statistical model for multiple produced wafer surfaces. It uses an additive linear model with cubic-spline basis functions for the global mean trend and a Gaussian process for the stochastic error. However, it does not include the statistical rules to determine whether the fitted model is in control or not, therefore, still not suitable for statistical control.

An additive Gaussian process (AGP) model proposed by Zhang *et al.* [37] is the most recently published work regarding spatial analysis of wafer profile data. Two independent Gaussian processes with different covariance structures are used to model thickness profiles for multiple wafers and quantify variations among wafers: the first Gaussian process is used to approximate the mean geometric profile of multiple wafers, while the second is used to quantify the spatially correlated deviations from the mean profile. It not only outperforms abovementioned spatial models for wafer profile data, but also overcomes the limitations of quality monitoring: it establishes a new statistical control scheme based on the developed AGP model to detect the out-of-control wafers with high sensitivity.

From the aforementioned literature review [37, 41, 42], one evident but remaining untapped characteristic of the wafer profile data among the aforementioned wafer models is the clustering phenomenon existed in wafer data (i.e., after grouping the wafers according to their homogeneous profiles (refer to Figure 4-1(b)), the variation within each cluster is much smaller than the variation among different clusters). Even with AGP model, the state-of-the-art in GP-based spatial modeling, without considering the clustering phenomenon in wafer data, the AGP-based statistical control scheme is prone to miss detection because it cannot distinguish different spatial data but still with the same spatial correlation structure [30, 37].

Such data underutilization can be addressed by the proposed spatial Dirichlet process (SDP) modeling and SDP-based statistical control scheme, extension of Dirichlet process based SPC [43] into spatial analysis. Since the proposed SDP modeling not only estimates the spatial correlation structure of the spatial data but also captures the clustering phenomenon among them, the proposed SDP-based statistical control scheme can effectively detect the out-of-control wafers even when

they still have the same spatial correlation structure with the in-control wafers, therefore, having higher power of test than the state of the art AGP-based statistical control scheme.

2.2.3 Porosity detection and quantification in products made by additive manufacturing (AM)

Porosity is a frequent and prominent defect in powder-based AM processes, namely, directed energy deposition (DED), laser and electron beam powder bed fusion (LPBF and EPBF), and binder jetting (BJ)[44]. Porosity has a negative influence on the mechanical performance of AM parts, leading to premature failure, and remains an impediment to wider application of AM processes [45, 46].

Porosity occurs typically due to: (a) the part design, (b) material properties, (c) machine-related errors, and (d) processing and environmental conditions [44]. For example, in powder bed fusion AM process, incomplete melting of particles due to insufficient energy leads to acicular pores at the meso-scale (10 μm to 100 μm); while excessive energy leads to material vaporization and hence micro-level gas pores (< 10 μm) [47-49]. In ceramic or metal parts fabricated with binder jetting, pores are mainly generated due to low dosing of the powder bed, i.e., insufficient amount of powder is raked across the powder bed [50-52]. Since the nature of porosity is contingent on its root cause, thorough quantification of its attributes (type, size, form, and location) is a prerequisite for process improvement.

As a common practice, percentage porosity i.e., the percentage of pores relative to the bulk volume of the part, is used to quantify porosity in AM parts [45]. The porosity value is obtained by several well-established methods, such as Archimedes method [53, 54], ultrasonic testing [45, 55], microscopic image analysis methods [56, 57], and X-ray computed tomography (CT) [58-60]. While Archimedes' method delivers the most reliable and accurate result of porosity based on measured density, it provides no spatial and morphological information of porosity. In contrast,

microscopic analysis and X-ray computed tomography (CT) methods capture the layer cross-section information. These techniques, apart from providing a volume measurement of porosity, can be used to analyze the morphological features and occurrence of pores [58-60].

The CT method to assess porosity in AM parts provides non-destructive measurement of the internal morphology [58, 61, 62]. The application of CT in AM can be traced back to the early nineties, when it was mainly used for reverse engineering purposes, and has gradually evolved since then as an established method for AM part qualification [63]. For instance, Taud *et al*. [64] measured overall percentage porosity in AM parts from CT scans by calculating the ratio of the number of voxels representing pores to the number voxels representing the bulk material. Siddique *et al*. [58] applied dimensional measurements of individual pores from CT scans in their characterization of fatigue performance. Likewise, Tammas-Williams *et al.* [61] identified areas of a part frequently afflicted with porosity from CT scans by visualizing the pores in a reconstructed 3D space. However, investigation of the distribution and prorogation of pores within a layer and across different layers from CT scan images, and the subsequent layer-wise prediction of areas where pores are liable to occur, remain an open research problem.

2.2.4 Point process analysis and its applications

As a component of spatial statistics, point process analysis aims to investigate the distribution-related characteristics and mutual dependence of events occurring within a defined region of interest [28, 65]. More specifically, point processes use dimensionless points to capture the number and the locations of events within a region of interest through a stochastic intensity function [28, 65].

Examples of point processes are common in the natural sciences, such as ecology and epidemiology [66, 67]. For instance, the spreading of certain tree species (as events) in a forest (as

the region of interest) [66]. The use of point process analysis has recently grained traction in manufacturing. For instance, Zhou *et al.* [68] used point process analysis to explain the clustering phenomena of particles in the manufacture of nanocomposites. Likewise, Dong *et al.* [69] invoked point process modeling to quantify the mixing of nanoparticles within composite materials. However, the application of point process analysis for modeling porosity in AM remains unexplored.

Among point processes, spatiotemporal log Gaussian Cox process (ST-LGCP) is a hierarchical modeling approach [70, 71]. ST-LGCP models capture complex spatial heterogeneity and temporal correlation among events through a nonparametric intensity function that encapsulates a random process in space-time. ST-LGCP is the latest embodiment in spatiotemporal point process analysis, and is further improved in this paper to investigate the spatiotemporal evolution of porosity in AM parts.

A salient aspect of ST-LGCP is that it reduces events within the region of interest to dimensionless points by disregarding the morphological aspects (shape and size) of the events. Accordingly, ST-LGCP introduces a bias which is magnified in the modeling of porosity in AM, because, pores in AM parts have different sizes and shapes contingent on the causal process phenomena. For example, in powder bed fusion, gas pores which are largely spherical occur at the micro-scale (< 10 μm), while elongated acicular pores occur at the meso-scale (10 μm to 100 μm) [72]. Disregarding the effect of pore morphology (size and shape) is therefore not physically tenable and will inevitably lead to inaccurate prediction of porosity [69]. The augmented ST-LGCP (AST-LGCP) framework proposed in this work overcomes this inherent morphology-related limitation of ST-LGCP by formulating the spatiotemporal distribution of pores through an augmented point pattern which includes the morphological features (size and form) of pores.

3 Real-time Process Monitoring based on Dirichlet Process Mixture Model

In advanced manufacturing, the data signals are usually generated by complex nonlinear dynamics, such as the vibration sensor data acquired from the FFF and CMP processes, the data distribution may not be Gaussian. This poses a significant challenge for process modeling and monitoring with traditional methods based on the presumption of normality and symmetry of data. Pertinently, since a non-Gaussian distribution can be modeled by using a mixture of Gaussian distributions, such impediments could be overcome.

Dirichlet process (DP) mixture model, which is a data-driven nonparametric Bayesian approach, approximates an empirical (arbitrary) data distribution via a mixture of finite Gaussian distributions without a *priori* knowledge of the number of mixture components. Such distinctive feature can be used in process monitoring in advanced manufacturing systems, where the signal data from sensors usually have arbitrary distributions.

The goal of this chapter is to develop different real-time monitoring methodologies for different advanced manufacturing processes based on Dirichlet process mixture model. These processes are fused filament fabrication (FFF) and chemical mechanical planarization (CMP). Due to their difference in process and obtained sensor data, the real-time monitoring methods will be different depending on the production of the product and the monitoring intentions.

3.1 Introduction

3.1.1 Fused filament fabrication (FFF) process

Fused filament fabrication (FFF) is an additive manufacturing (AM) process in which an object is manufactured by depositing progressive layers of extruded molten material [1-4]. In FFF, typically

a thermoplastic material is heated past its glass transition temperature and extruded through a nozzle in a controlled manner. Thermoplastic polymers, such as ABS and PLA, given their relatively low glass transition temperatures (approximately 105 °C and 65 °C, respectively), are the materials of choice in FFF. Although thermoplastic materials are ubiquitous in FFF, the process has evolved to accommodate novel materials and application areas, e.g., mortar, clay, and lunar regolith [73, 74]. In AM processes, such as FFF, major challenges with respect to process reliability and consistency remain to be resolved [2, 75-77]. In the absence of real-time process monitoring, quality control in FFF is largely limited to offline techniques [78, 79], leading to high scrap rates [3]. Therefore, there is a need to develop effective process control and monitoring techniques in FFF [80]. Equipped an FFF machine (MakerBot Replicator 2X) with multiple *in situ* sensors, such as accelerometers (vibration sensors) and temperature sensors. The schematic diagram of our sensor-instrumented FFF setup is shown in Figure 3-1. This setup allows real-time monitoring and detection of FFF process drifts from ideal conditions, so that quick corrective action can be taken and product defects can be reduced. Combined the sensor data using an integrated Bayesian Dirichlet process (DP) mixture model and evidence theoretic (ET) approach to FFF detect process drifts. By analyzing the *in situ* sensor data, process defects can be identified at an early stage.

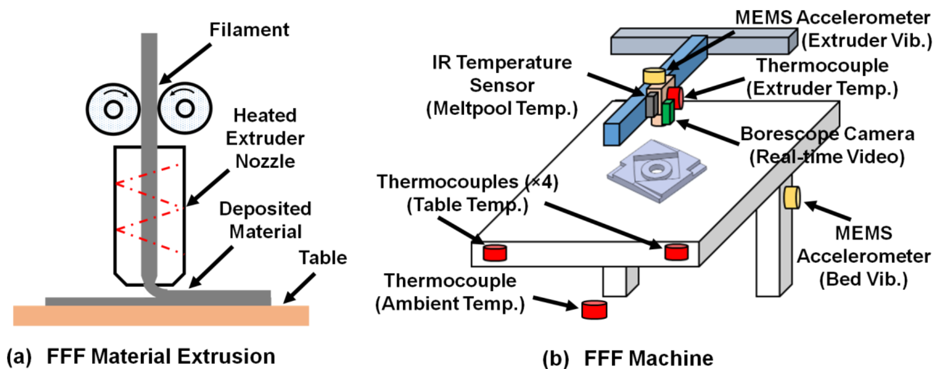

Figure 3-1 (a) Schematic of the FFF process. (b) Schematic of the FFF setup instrumented with multiple *in situ* sensors used in this work for measuring process conditions in real-time.

3.1.2 Chemical mechanical planarization (CMP) process

CMP is a vital back-end-of-line (BEOL) process in semiconductor manufacturing. Semiconductor wafer defects resulting from CMP process drifts can lead to high yield losses [81]. It is therefore desirable to ensure defect-free operation in CMP by employing real-time *in situ* sensor-based process monitoring approaches [82]. Various sensors, such as acoustic emission, force, and vibration sensors, have been applied for CMP process monitoring [82-90]. Miniature wireless MEMS devices are particularly attractive for *in situ* monitoring applications due to their weight, and energy efficiency. MEMS vibration sensors have been successfully used hitherto for model-based monitoring, material removal rate estimation, and endpoint detection in CMP process [87, 88].

We use a Buehler Automet 250 benchtop CMP apparatus for our experiments. Further details of the setup and experimental outcomes are available in our recent publication (Ref. [82]). A tri-axis MEMS vibration sensor (ADXL 335) manufactured by Analog Devices Inc. is mounted on the apparatus to collect sensor data. The sensor signals are sampled at 670 Hz and transmitted wirelessly to a desktop computer with a matching wireless receiver unit. The CMP setup and wireless sensor network are shown in Figure 3-2 (a) and (b). Blanket copper wafer disks of Φ1.625 inch (40.625 mm) are polished in KOH-based alkaline colloidal silica slurry medium, which has a constant flow rate of 20ml/min. Near-optical (arithmetic average roughness, Sa ~ 5 nm) quality surface finish blanket copper wafers are obtained by polishing with a *priori* identified optimal processing conditions (Figure 3-2(c) and (d)) [91].

However, sensor signals acquired from CMP process are complex; they may violate normality and linearity conditions. Consequently, traditional SPC and mean shift clustering approaches may not lend towards detection of CMP process anomalies.

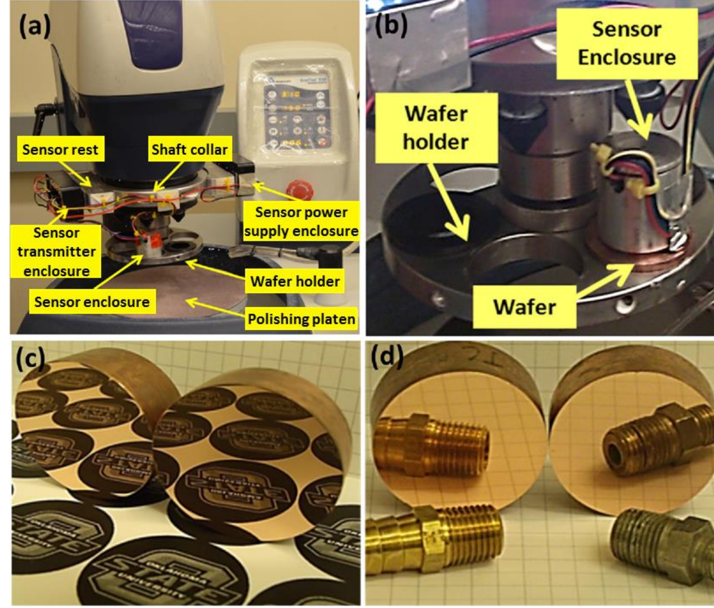

Figure 3-2 (a) (b) Buehler Automet® 250 experimental CMP setup with the integrated wireless sensor, (c) (d) near-specular CMP finished copper wafers.

3.1.3 Real-time process monitoring methods based on DP mixture model

The essential concept of monitoring methods based on DP mixture model is to represent a non-Gaussian probability distribution as a mixture of multiple Gaussian distributions. This implication can be stated mathematically as follows,

$$p(x) = \sum_{j=1}^{k} \pi_j \mathbb{N}(x|\theta_j), \tag{2}$$

where x represents a time series collected by sensors from the process, $p(x)$ is its data distribution, k is the number of Gaussian components $\mathbb{N}(\cdot)$ in the mixture, each of which is modeled with weight π_j and parameters θ_j (mean μ_j and variance ${\sigma_j}^2$). In reality k may be unknown. We apply the Dirichlet process (DP) mixture model, which is a data-driven nonparametric Bayesian approach to approximate a non-Gaussian distribution without any a *priori* knowledge of k [6].

The process monitoring methods for FFF and CMP are built on the fact that DP mixture model enables accurate modeling of data distributions in different process states, and then developed according to their own characteristics.

3.2 Research methodology

3.2.1 Dirichlet process (DP) mixture model

DP mixture model is a data-driven nonparametric Bayesian approach to approximate an empirical (arbitrary) data distribution via a mixture of finite Gaussian distributions without a *priori* knowledge of the number of mixture components [6, 92]. DP mixture model is advantageous over the parametric estimation methods, such as EM algorithm [93], due to its nonparametric characteristics in estimating data distributions. It can estimate complex data distribution with a few simple probability density function. It has been applied to broad areas, such as bioinformatics, healthcare, document clustering and image processing [94-97].

In Dirichlet process, the limit for the number of clusters k goes to infinity [92]. In other words, when $k \rightarrow \infty$, the conditional prior distribution for the component indicators reaches its limit as follows,

$$p(c_i = j | \boldsymbol{c}_{-i}, \alpha) \propto \begin{cases} \dfrac{n_{-i,j}}{N-1+\alpha} & \text{if exsisting component } j \text{ is chosen,} \\ \dfrac{\alpha}{N-1+\alpha} & \text{if a new component is created,} \end{cases} \quad (3)$$

where $\boldsymbol{c} = (c_1, \dots, c_N)$ are indicators of data points for components, α is the concentration parameter, n_j is the number of data points in Gaussian component j, N is the number of data points, $N = \sum_{j=1}^{k} n_j$. The subscript $-i$ indicates all indices except i, and similarly, $n_{-i,j}$ indicates the number of observations in component j for all data points except point i.

For each component indicator c_i drawn conditioned on all other component indicators from the multinomial distribution, there is a corresponding component parameter θ_i drawn from a base distribution G_0. This result signifies a Dirichlet process (DP) mixture model, which can be used to model a set of observations $(x_1, \dots, x_i, \dots, x_N)$, with latent variables of $\boldsymbol{\theta} = (\theta_1, \dots, \theta_i, \dots, \theta_N)$ as follows,

$$G \sim DP(\alpha, G_0), \tag{4}$$

$$\theta_i \sim G,$$

$$x_i \sim \mathbb{N}(.\,|\theta_i),$$

where $DP(\alpha, G_0)$ is the Dirichlet process (DP) with base distribution G_0 and concentration parameter α; G is a random discrete distribution drawn from $DP(\alpha, G_0)$; each θ_i is drawn from the discrete distribution G; and each data point x_i (which may include statistical features, e.g., mean, variation, etc., from the sensor data) is drawn from a normal distribution with parameter θ_i. Because the empirical distribution G is discrete, the same values can be assigned to multiple θ_i. Data points which have the same latent value belong to the same component [6, 92]. Furthermore, on integrating out G, the following conditional distribution for θ_i is obtained [98],

$$\theta_i|\boldsymbol{\theta}_{-i}, \mathrm{G}_0, \alpha \sim \frac{\alpha}{N-1+\alpha}\mathrm{G}_0 + \sum_{j=1 \cup j \neq i}^{N-1} \frac{1}{N-1+\alpha}\delta(\theta_j), \tag{5}$$

where $\delta(\theta_j)$ is the Dirac delta function peaked on θ_j. It is noted that DP has the clustering tendency within its components, i.e., first, a few number of distribution components could sufficiently represent all the existing data; second, the component with more data have higher probability to generate new data.

Subsequently, combining the prior distribution for θ_i of data i in (5) and the likelihood function in (4) results in the following posterior distribution for Gaussian components parameters,

$p(\theta_i = j | \boldsymbol{\theta}_{-i}, x_i) \sim$

$$\begin{cases} \frac{n_{-i,j}}{N-1+\alpha} \mathbb{N}(x_i|\theta_j) & \text{if existing component } j \text{ is chosen (a),} \\ \frac{\alpha q}{N-1+\alpha} H(\theta|x_i) & \text{if a new component is created (b),} \end{cases} \quad (6)$$

where $q = \int G_0(\theta)\mathbb{N}(x_i|\theta)d(\theta)$, $H(\theta|x_i) = \frac{G_0(\theta)\mathbb{N}(x_i|\theta)}{\int G_0(\theta)\mathbb{N}(x_i|\theta)d(\theta)}$. Equation (6(a)) shows the probabilities of θ_i having the same value with existing Gaussian component parameter θ_j, and (6(b)) is the posterior probability of θ_i choosing a new value which is randomly generated from $H(\theta|x_i)$.

There is no direct way to compute the posterior distribution of θ under a Dirichlet process prior. Approximate inference methods are required for DP mixtures and Markov chain Monte Carlo (MCMC) sampling methods have become the methodology of choice [99-103]. For instance, Gibbs sampling algorithms is used to get posterior distribution of θ in DP mixture model when conjugate priors are used [104]. It will provide accurate and flexible data distribution modeling. The joint posterior distribution for the parameters of all existing components $\{\theta_l,\ l = 1, 2, \ldots\}$ can be obtained. If the base distribution G_0 is chosen to be conjugated with data distributions, a closed form of posterior distribution could be obtained. If the base distribution G_0 and data distributions are chosen as Gaussian distribution, the posterior distribution is mixture of Gaussian distributions, known as Gaussian mixture model [6, 105, 106].

3.2.2 Multi-sensor real-time detection of process anomalies in FFF using Dirichlet process (DP) mixture model and evidence theory (ET)

A novel data-driven approach based on DP mixture model is developed for online monitoring in FFF. Specifically, a non-parametric Bayesian Dirichlet process (DP) mixture model is used to classify the process into three process states, namely, normal operation, abnormal operation, and

build failure [106]. These DP models will be invoked in an evidence theoretic (ET) framework for real-time detection of process drifts using online sensor data [107].

There are two steps in our proposed method for AM online process monitoring:

1) In the first step, complex non-linear and non-Gaussian signals from individual sensors under different process states are represented using DP mixture model (Figure 3-3 (left)) [106]. During the offline training phase, DP mixture model is used to estimate the underlying probability density function of heterogeneous sensor data;
2) In the second step, Dempster-Shafer evidence theory (ET) [107] is used to integrate the signal patterns from seven different sensor signals and classify the process state based on the DP mixture distributions. Incoming sensor data (Figure 3-3 (right)) can be classified in real-time by using this two-stage algorithm.

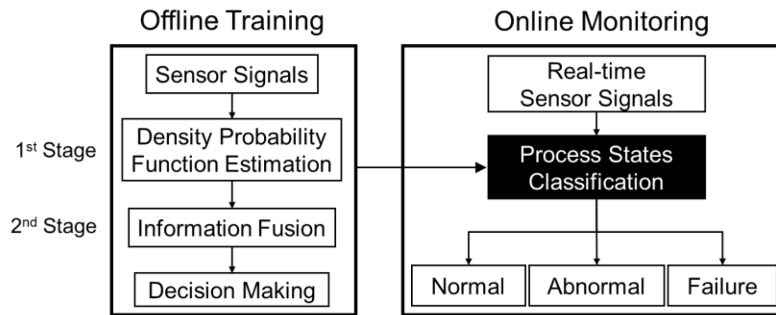

Figure 3-3 The summary of the Dirichlet process (DP) mixture model and evidence theory (ET) approach developed in this work for real-time monitoring of process states in FFF.

Step 1 - Dirichlet Process Mixture Model for Non-Gaussian Data

We observed that the FFF sensor signal patterns acquired from our experiments were distinctly non-Gaussian owing to the complex dynamics of the process (see Figure 3-4). Nonetheless, Gaussian mixture models can be used to model complex nonlinear probability distributions by convex combination of Gaussian distributions [108]. Figure 3-4 shows the probability density

function of IR temperature data for different process states; the distribution of values is evidently complex and non-Gaussian. For example, in the normal process state, the distribution of meltpool temperature can be modeled as a mixture of two Gaussian distributions.

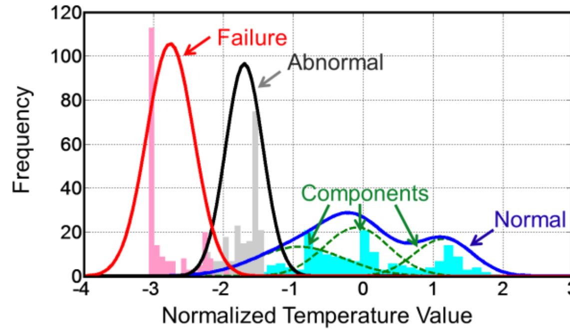

Figure 3-4 Histogram with fitted probability density functions of IR sensor data at normal, abnormal, and failure process states. The signal patterns have a non-Gaussian probability density function, which is particularly pronounced for the normal state.

Therefore, all the distribution of sensor data could be represented by (2). Our approach using Dirichlet process (DP) modeling relaxes the need to determine the value of k beforehand, by adaptively modeling the signal distribution using an infinite DP mixture model and determining k in a data driven manner. Thus, using the DP modeling approach, even complex sensor signal patterns having an arbitrary probability distribution, can be approximated in the form of (2). For further mathematical details of DP mixture model, are discussed in depth in our recent research (see Refs. [109, 110]).

Step 2 - Dempster-Shafer Evidence Theory

In the previous step, by using DP mixture model, we can model the signal patterns from different sensors separately. The aim of this stage is to combine this information in order to provide a cohesive decision regarding the process state. Dempster-Shafer evidence theory (ET) [107] takes into consideration the conflicts among the sources of evidence to handle uncertainty during

information fusion, and has been widely used for multi-sensor condition monitoring purposes [111-113].

For instance, considering the case where we have N sensors and M process states (corresponding to different build conditions in AM processes), for a new sensor signal, we want to determine which state it belongs to by using the information collected from the N sensors. For each sensor and each state, the distribution of sensor signal can be estimated using Eq. (1). Thus, each of these N sensors will have a probability p associated to a state. We denote the probability of the process state j from sensor k as p_{jk}. We wish to evaluate the overall probability of the process being in state j by fusing information from multiple sensors. We achieve this using Dempster-Shafer evidence theory (ET). These sensor-state probabilities can be arranged in the form of a matrix $\boldsymbol{P}$ as follows:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1\mathrm{N}} \\ p_{21} & p_{22} & \cdots & p_{2\mathrm{N}} \\ \vdots & \vdots & \ddots & \vdots \\ p_{\mathrm{M}1} & p_{\mathrm{M}2} & \cdots & p_{\mathrm{MN}} \end{bmatrix}. \tag{7}$$

The values in column k ($[p_{1k} \quad \cdots \quad p_{\mathrm{M}k}]^{\mathrm{T}}$, k =1, 2,…, N) indicate how the kth sensor supports each of the M process states, hence, $\sum_{j=1}^{\mathrm{M}} p_{jk} = 1$ (k =1, 2,…, N). We now describe the procedure to populate the matrix $\boldsymbol{P}$ from the previous DP mixture modeling stage in the following manner. First, we obtain the likelihood estimate $\mathbb{f}(\boldsymbol{y})$ that the process is in a particular state given an incoming data point $\boldsymbol{y}$ by a type of sensor from the LHS of Eq. (1). Next, these likelihood estimates are normalized for each of the N sensors, so that the probabilities over all states sum to one. The normalized values for different states from the distribution $\mathbb{f}$ for sensor k are essentially the sensor-state probabilities in the column k of the matrix P. With the information in Eq. (1) available, if the types of process states are assumed to be independent, the probability mass function for process states j can be computed as follows using the evidence theory [114]:

$$m_j = \prod_{k=1}^{n} p_{jk} / (1 - \sum_{j=1}^{\mathrm{M}} \prod_{k=1}^{n} p_{jk}), \tag{8}$$

where m_j represents the fused probability mass function for process states j (j =1, 2, …, M), we term it state probability. The denominator $1 - \sum_{j=1}^{\mathrm{M}} \prod_{k=1}^{n} p_{jk}$ in Eqn. (8) is the normalization factor; $\sum_{j=1}^{\mathrm{M}} \prod_{k=1}^{n} p_{jk}$ is in effect the probability mass associated with conflicting information among the n sources. From Eqn. (8), we can obtain the state probability estimate (m_j) for process state j for different sensor combinations. We select the combination with the highest state probability m_j. Thus, a fused decision regarding the process states can be made in real-time.

3.2.3 Dirichlet process (DP)-based SPC for monitoring non-Gaussian data signals

A control chart is a visual tool that is used for monitoring whether a process or system at a given time is under the influence of common cause (chance) variation or special cause (assignable) variation [7, 115]. The limits of the control charts represent thresholds that are obtained when a system operates wholly under common cause variation (in-control (IC) condition). In IC condition, the monitoring statistic falls within the control limits threshold. If special causes take effect, the control chart should presumably signal a change in terms of the monitoring statistic drifting outside the control limits (out-of-control (OOC) condition). Thus, the control chart is effectively a two-state or binary classifier as it signals only IC or OOC process states. The control chart does not identify, explicitly, the type of anomaly/special cause.

DP-based SPC for detecting process anomalies is introduced by using Dirichlet process mixture model which approximates an empirical data distribution as a mixture of Gaussian components. In DP-based SPC, the likelihood values of new data are calculated under IC data distribution acquired by DP mixture model. Process changes are detected once the likelihood values drop, indicating such data are not likely generated under IC condition. An effective way to detect the OOC

operation is by monitoring average log-likelihood value in a subgroup of incoming data under IC data distribution as in (9),

$$\frac{1}{w}\log\left[L\left(x_1,\dots,x_i,\dots,x_w\middle|\theta_1,\theta_2,\dots,\theta_j\right)\right]=\frac{1}{w}\sum_{i=1}^{w}\log\left[\sum_{k=1}^{j}\pi_k\mathbb{N}(x_i|\theta_k)\right], \tag{9}$$

where x_i is the incoming data, j is the number of components for mixture distribution for IC condition, w is the subgroup size of testing data. The larger the value of w, the more reliable the detection of OOC operation, but a longer delay is caused to detect process changes. Based on empirical results, we choose w as the minimal number of observations to achieve the average run length (ARL_0) for type I error of likelihood values below a certain value, e.g., 0.05, in order to balance fast detection and detecting accuracy.

By the central limit theorem, the average log-likelihood values of incoming data are approximately normally distributed. Therefore, the problem of monitoring original complex non-Gaussian data reduces to a scheme of monitoring normally distributed average log-likelihood values. For simplicity, we construct the DP-based SPC by closely emulating the framework of the CUSUM chart with the average log-likelihood values as the monitoring target. Therefore, representing the average log-likelihood value in time epoch t as y_t, we have monitoring statistics for DP-based SPC:

$$\begin{cases} C_t^+ = max[0, y_t-(\mu_0+K)+C_{t-1}^+], \\ C_t^- = max[0, (\mu_0-K)-y_t+C_{t-1}^-], \end{cases} \tag{10}$$

with the control limits (threshold) for the chart set at

$$\mathrm{H} = L\sigma, \tag{11}$$

where σ and μ_0 are the standard deviation and mean of the sequential data y_t under IC condition; the parameters K and L are adjusted for a given average run length criteria (ARL_0) [7]; the average log-likelihood value y_t is obtained from (9). The cumulative sums C_t^+ and C_t^- are tracked over

time, if these quantities are greater than H, then OOC status is signaled. The cumulative sums C_t^+ and C_t^- are never negative.

3.2.4 Recurrent hierarchical Dirichlet process (RHDP) for evolutionary clustering of process states

Towards identifying different types of process anomalies from a continuous sensor data stream, two extensions to DP models are forwarded to accommodate the following aims:

1) Estimating distribution characteristics within a contiguous time epoch given the autocorrelation in the data.
2) Tracking the evolution of the data distribution between time epochs.

The first aim is realized using the recurrent Dirichlet process (RDP) method proposed by Ahmed and Xing [116]. The second involves using the hierarchical Dirichlet process (HDP) developed by Teh *et al*. [117]. The integration of these two entities is accomplished in this work, and is termed recurrent hierarchical Dirichlet process (RHDP). Specifically, the RDP and HDP parts resolve the following questions:

- RDP – What are the characteristics of the data distribution at the current time epoch, given the knowledge of the distribution characteristics at the previous time epochs?
- HDP – What category is the process state (anomaly/fault) at the current time epoch, given the distribution characteristics estimated using RDP?

In other words, RDP determines characteristics of data distribution at the current time epoch by including information from previous time epochs, while HDP enables data to be classified by allowing distributions falling under the same DP model to be clustered together. The overall framework of the proposed methodology is summarized in Figure 3-5.

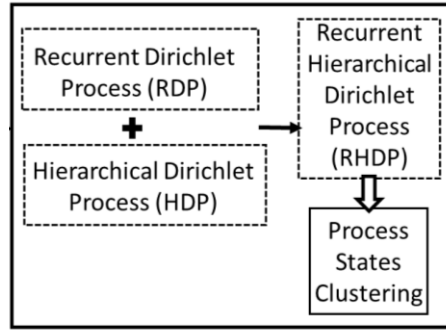

Figure 3-5 Overall methodology of RHDP for different process states clustering.

For complex manufacturing processes, each process state manifests in unique signal distributions. A control chart cannot classify the differences in process states because the control limits are estimated based on the so-called, in-control state alone. In order to identify the specific process anomalies (drifts), we herewith propose recurrent hierarchical Dirichlet process (RHDP) clustering.

Recurrent hierarchical Dirichlet process (RHDP) model could not only estimate the data distribution at each time epoch in a sequential manner, but also consider the evolution of distributions between time epochs. For instance, during the physical process, as depicted in Figure 3-6, the following four possibilities exist:

1) the signals dynamics remain stationary (no change);
2) new components may emerge;
3) the parameters of mixture components may change over time; and
4) an existing component may die.

Therefore, to classify the process, it is essential to track the evolution of mixture components between time epochs.

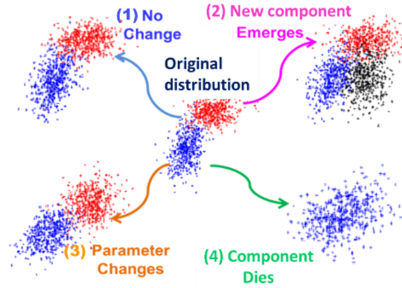

Figure 3-6 Possible evolutions of data distribution in a physical process.

Recurrent Dirichlet process (RDP) model is proposed by Ahmed and Xing [116] to overcome the challenge in DP modeling. In other words, the autocorrelation of the data is not accounted in the basic DP modeling. The recurrent Dirichlet process (RDP) model divides time series data into contiguous sequential epochs (windows); data points within the same epoch are assumed to be exchangeable, while the temporal order is maintained across epochs. Thus, the autocorrelation in consecutive epochs is accounted in the RDP model. In the implementation of RDP, the incoming sensor data is divided using a sliding window technique; the data inside a sliding window is an epoch.

The concept of hierarchical Dirichlet process (HDP), which is essentially a multiple-level Dirichlet process, as proposed by Teh *et al*. [117]. Unlike with DP, in a two-level HDP model, the parent Dirichlet process G_0 is a random variable distributed with concentration parameter γ and base distribution H. The so-called child Dirichlet processes $G_j's$ have concentration parameter α and base distribution G_0. Since G_0 is discrete, the child Dirichlet processes $G_j's$ share atoms (mixture components) with each other. Data distributions parametrized by $G_j's$ with the same atoms, will have the similar Gaussian components, and therefore could be clustered together [117].

Given a temporal dataset, RHDP model could be used to monitor the distribution evolution among multiple sequential epochs, subsequently the time epochs with similar distribution characteristics can be grouped/clustered together [117, 118]. The RHDP model is formulated as,

$$\begin{gathered} G_0|\gamma \sim DP(\gamma, H), \\ G^t|\alpha \sim DP(\alpha, G_0), \\ \theta_i^t|G^t \sim G^t, \\ x_i^t|\theta_i^t \sim \mathbb{N}(\cdot\,|\theta_i^t), \end{gathered} \tag{12}$$

where x_i^t for i =1,…, N^t are observations in time epoch t; $\mathbb{N}(\cdot\,|\theta_i^t)$ denotes the Gaussian component parameterized by θ_i^t, which is sampled from the child Dirichlet process G^t.

Data in different epochs are modeled by using Gaussian mixture distributions with parameters $\{\theta_1^t \dots, \theta_{N^t}^t\}$ sampled from G^t. If the process is stationary, the parameters of the mixture distribution would remain constant. However, if there is a change in the underlying process, entailing a change in the data distribution, the current data distribution will not suit the new data, i.e., the existing parameters drawn from G^t will not appropriately model the new data. Accordingly, new samples for G_0 will be drawn from the base function H of parent Dirichlet process.

We can estimate marginal distributions of mixture component at two levels of DP by integrating out G_0 and G^t. The conditional distribution for θ_i^t can be calculated by integrating out G^t as follows,

$$\theta_i^t|\boldsymbol{\theta}_{\cdot}^{t-1}, \theta_1^t, \dots, \theta_{i-1}^t, G_0, \alpha \sim \frac{1}{N^{t-1} + i - 1 + \alpha}\left[\sum_{j \in J^{t-1} \cup J^t} \left(n_{\cdot j}^{t-1} + n_{-i,j}^t\right) \delta\left(\phi_j^t\right) + \alpha G_0\right], \tag{13}$$

where ϕ_j^t represents the distribution parameter of mixture component j at time epoch t. Notice that in (13) G_0 is not fixed, but distributed as Dirichlet process.

The subsequent step is to integrate out G_0 to get the conditional distribution for ϕ_j^t. Since G_0 is distributed as Dirichlet process, it can be integrated out as follows,

$$\phi_j^t|\boldsymbol{\phi}_{\cdot}^{t-1}, \phi_1^t, \dots, \phi_{j-1}^t, \mathrm{H}, \gamma \sim \frac{1}{M^{t-1}+j-1+\gamma}\left[\sum_{l\in L^{t-1}\cup L^t}\left(m_{\cdot l}^{t-1}+m_{-j,l}^t\right)\delta(\tau_l)+\gamma H\right], \tag{14}$$

where τ_l denotes a value drawn from base distribution H, M^{t-1} is the number of all Gaussian components in epoch $t-1$, $m_{\cdot l}^{t-1}$ is the number of Gaussian components associated with τ_l at time epoch $t-1$, $m_{-j,l}^t$ is the number of the Gaussian components except components j associated with τ_l at time epoch t, and L^t denotes the collection of samples drawn from H at epoch t.

Subsequently, we obtain the posterior probability distributions for both the component values of Dirichlet process G_0 in (15), and its child Dirichlet process G^t in (16),

$$p\left(\phi_j^t = l\middle|\boldsymbol{\phi}_{\cdot}^{t-1}, \phi_1^t, \dots, \phi_{j-1}^t, \{c(x_i^t)=j\}\right) \sim \begin{cases}\left(m_{\cdot l}^{t-1}+m_{-j,l}^t\right)F\left(\phi_j^t\middle|\tau_l\right) & \text{if component } l \text{ is chosen,}\\ \gamma s T\left(\tau\middle|\phi_j^t\right) & \text{if a new component is created,}\end{cases} \tag{15}$$

where $s=\int H(\tau)F\left(\phi_j^t\middle|\tau\right)d(\tau)$, $T(\tau|\phi_j^t)=\frac{H(\tau)F\left(\phi_j^t\middle|\tau\right)}{\int H(\tau)F\left(\phi_j^t\middle|\tau\right)d(\tau)}$; and $F\left(\phi_j^t\middle|\tau_l\right)$ is the probability of ϕ_j^t getting the value of τ_l, which can be represented by likelihood of all data belonging to component j in the mixture distribution at epoch t (i.e., all data with indicator $c(x_i^t)=j$).

$$p(\theta_i^t = j|\boldsymbol{\theta}_{\cdot}^{t-1}, \theta_1^t, \dots, \theta_{i-1}^t, x_i^t) \sim \begin{cases}(n_{\cdot j}^{t-1}+n_{-i,j}^t)\mathbb{N}\left(x_i^t\middle|\theta_j^t\right) & \text{if component } j \text{ is chosen,}\\ \alpha q R(\theta|x_i^t) & \text{if a new component is created,}\end{cases} \tag{16}$$

where $q=\int G_0(\theta)\mathbb{N}(x_i^t|\theta)d(\theta)$, $R(\theta|x_i^t)=\frac{G_0(\theta)\mathbb{N}\left(x_i^t\middle|\theta\right)}{\int G_0(\theta)\mathbb{N}\left(x_i^t\middle|\theta\right)d(\theta)}$.

In (16), x_i^t represents data observation i during time epoch t; θ_i^t represents the distribution parameter for data x_i^t at time epoch t; ϕ_j^t represents the jth atom value of child Dirichlet process G^t(i.e., mixture component j at time epoch t). If the base distributions H is Gaussian, i.e., it is conjugate with distribution of observations, then the integrals in (15) and (16) have analytical solutions.

RHDP could attain unsupervised clustering of process states by monitoring the change of mixture components among time epochs, i.e., the evolution of sequential data distributions. RHDP clustering includes the following two major steps:

1) RDP modeling is used for sequential process data, which are segregated into sliding windows. Gibbs sampling is adopted to update the data distribution, and Pearson's Chi-square goodness of fit (GoF) test is used to evaluate the accuracy of distribution modeling.
2) Cluster data of which the mixture distributions are from the same realizations in HDP. The average log-likelihood value of current data under previous distribution is continuously calculated and monitored as follows,

$$\frac{1}{w}\log\left[L\left(x_1^t,\dots,x_i^t,\dots,x_w^t\middle|\theta_1^{t-1},\theta_2^{t-1},\dots,\theta_j^{t-1}\right)\right]=\frac{1}{w}\sum_{i=1}^{w}\log\left[\sum_{k=1}^{j}\pi_k\mathbb{N}(x_i^t|\theta_k^{t-1})\right], \qquad (17)$$

where x_i^t is the incoming data in epoch t, j is the number of components in epoch $t-1$. If the average log-likelihood values calculated as (17) remain stable and without significant drop, it indicates that the data in these consecutive windows have the similar distribution, therefore, could be clustering as one process state. This is computationally amenable than tracking the change of mixture components.

In this way, by tracking the evolution of mixture distributions at consecutive time epochs using RHDP model, process drifts in complex manufacturing processes, such as semiconductor CMP,

can be monitored, and different process states (e.g., different anomalies) can be identified. We demonstrate this assertion herewith using a numerical example.

3.3 Simulation studies

3.3.1 Application of DP-based SPC for process monitoring

In this section, we show that DP-based SPC can capture changes in the data despite the underlying distribution being asymmetric and multimodal. We compare the results with two conventional control charts, namely, exponentially weighted moving average (EWMA) and cumulative sum (CUSUM) [7]. The traditional control charts monitor the raw data values, while the DP-based SPC uses the average log-likelihood values in (9) within the CUSUM framework in (10). For comparison purposes, we use the following two average run length criteria (ARL) as widely used for performance evaluation of control charts: ARL_0 and ARL_1 [7].

We now test the hypothesis that DP-based SPC has superior ability (i.e., smaller ARL_1) in capturing the changes in incoming data compared to EWMA and CUSUM given identical ARL_0. The following three scenarios are investigated:

Case N1: Detecting mean shifts in univariate, unimodal Gaussian and non-Gaussian distributions.

Case N2: Detecting mean shifts in univariate, multimodal non-Gaussian distributions.

Case N3: Detecting shifts in a multivariate, nonlinear, quasi-periodic data from the Rössler chaotic attractor [119].

3.3.1.1 Case N1: DP-based SPC for data from univariate, unimodal Gaussian and non-Gaussian distributions

The aim of this study is to ascertain the ARL_1 performance of DP-based SPC towards detecting a shift in mean (location parameter) of a distribution. Furthermore, we contrast the ARL_1 performance of DP-based SPC with that of EWMA and CUSUM control charts.

This study is conducted with data generated from two basic univariate distributions: the Gaussian distribution with mean μ and variance 1, $\mathbb{N}(\mu,1)$; and the Chi-squared distribution with one degree of freedom, χ_1^2. The mean of these distributions will be shifted from the in-control state of zero mean, and the ARL_1 will be evaluated for CUSUM, EWMA, and DP-based SPC.

We note that the latter distribution (χ_1^2) is inherently asymmetric (right skewed), and is theoretically equivalent to a F distribution, $F(1,\infty)$. Both the Gaussian and Chi-squared distributions are approximated by mixtures of Gaussian components using DP mixture model, as exemplified in Figure 3-7.

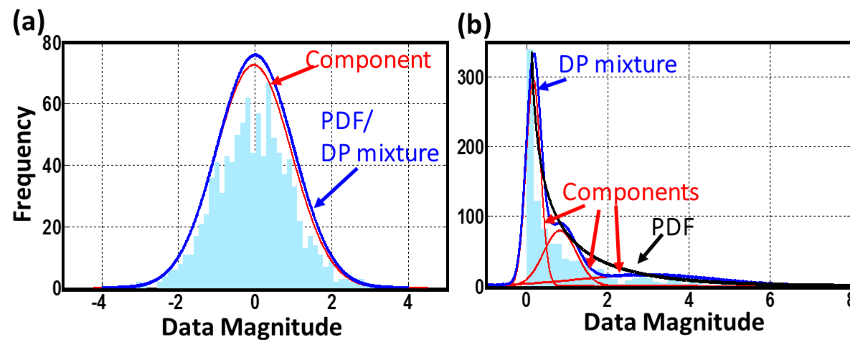

Figure 3-7 True pdf, Gaussian components, and approximated distribution by DP mixture model for data generated from (a) $\mathbb{N}(\mu,1)$ and (b) χ_1^2.

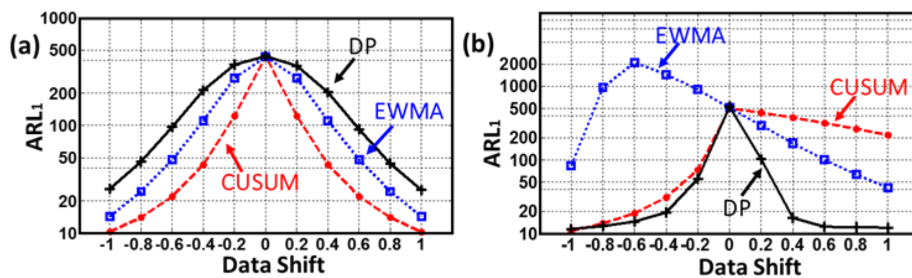

Figure 3-8 Case N1 results - OOC ARL_1 values of different control charts when ARL_0 is fixed at 500, and actual IC data generated from (a) $\mathbb{N}(\mu,1)$ and (b) χ_1^2. Scale on the y-axis is in natural logarithm.

The out-of-control (OOC) data are obtained by mean shift, ranging from -1.0 to 1.0 with a step of 0.2. The control limit is acquired by adjusting parameter L in (11) to obtain average ARL_0 in 5,000

repetitions at 500 under IC condition; the ARL_1 values are reported based on 10,000 replications. The average ARL_1 results for the EWMA, CUSUM, and DP-based SPC are reported in Figure 3-8. The following observations can be tendered from Figure 3-8:

i. Figure 3-8 (a): When the normality assumption is not violated, as the case with the Gaussian distribution $\mathbb{N}(\mu,1)$, CUSUM and EWMA perform better (lower ARL_1) than DP-based SPC.
ii. Figure 3-8 (b): If the data is patently non-Gaussian, i.e., the normality condition as in the case of χ_1^2 is violated, then the ARL_1 of DP-based SPC is smaller than EWMA and CUSUM control charts.

3.3.1.2 Case N2: DP-based SPC for data from univariate, multimodal, non-Gaussian distributions

In this case study, the data are obtained from an underlying bimodal distribution consisting of $\mathbb{N}(10,1)$ and χ_1^2. As evident in Figure 3-9, DP mixture model closely approximates the data distribution, which corroborates our assertion that DP mixture model can capture complex distributions.

As in the previous case (Case N1) OOC data are obtained by shifting the mean of the data in the range of -1.0 to 1.0 with a step of 0.2. Once again, the control limit is acquired by adjusting parameter L in (11) to obtain average ARL_0 of 5,000 repetitions at 500 under IC condition, and ARL_1 results from a 10,000-replication study are reported (Figure 3-10).

It can be inferred from Figure 3-10 that under a multimodal distribution, and when the data is patently non-Gaussian and asymmetric, the performance of EWMA and CUSUM is considerably inferior to DP-based SPC; the ARL_1 of DP-based SPC is smaller than EWMA and CUSUM. Indeed, the performance of the DP-based SPC is almost identical to Figure 3-8 (b), thus further affirming that the DP-based SPC is not influenced by symmetry and modes of the underlying data.

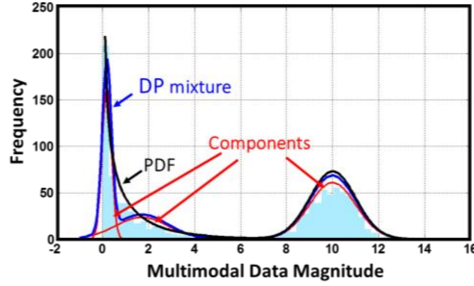

Figure 3-9 True pdf, Gaussian components and fitted distribution by DP model for data generated from a bimodal distribution consisting of $\mathbb{N}(10,1)$ and χ_1^2.

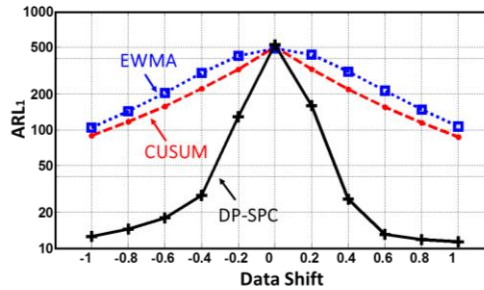

Figure 3-10 Case N2 results - OOC ARL_1 values of three SPC methods when ARL_0 is 500, and actual IC data are generated from a bimodal distribution consisting of $\mathbb{N}(10,1)$ and χ_1^2. Scale on the y-axis is in natural logarithm.

3.3.1.3 Case N3: DP-based SPC for multivariate, nonlinear, quasi-periodic data

Real-world signals customarily portray strong nonlinearity and high dimensionality; such behavior has been observed in several practical instances in manufacturing processes, including CMP [19, 120, 121]. In this case study, we show that DP mixture model can accommodate multidimensional data depicting nonlinear quasi-periodic dynamics [121].

The three-dimensional Rössler system, as delineated in (18), is used in this case study [119]; it consists of three coupled ordinary differential equations to define a continuous-time dynamical system, which exhibits chaotic nonlinear behavior predicated by the choice of three parameters, namely, *a, b* , and *c* in (18) [119].

$$\frac{dx}{dt} = -y(t) - z(t),$$
$$\frac{dy}{dt} = x(t) + a \cdot y(t), \tag{18}$$
$$\frac{dz}{dt} = b + z(t) \cdot [x(t) - c].$$

We fix the parameters as follows: a = 0.2, b = 0.2, c = 5. The Rössler system depicts prominent chaotic dynamics under these conditions; the dynamics of the Rössler system has been extensively investigated by Crutchfield *et al.* [119]. The Rössler attractor state-phase diagram obtained as a result of (18) are shown in Figure 3-11 (a). We note that, in this simulation, data generated from the Rössler system of (18) are purposely contaminated with Gaussian white noise $\mathbb{N}(\mathbf{0}, \sigma^2\boldsymbol{I_3})$, where $\boldsymbol{I_3}$ is the identity matrix of order 3; the effect of variance σ^2 on ARL_1 of DP-based SPC is tested in this case study. Shown in Figure 3-11 (b) is a sample 1,000 data points from the contaminated Rössler attractor. Next, DP mixture model is used to approximate the data distribution of Rössler contractor using a mixture of multivariate Gaussian distributions. 1,000 new data points are generated from the DP approximated distribution of the contaminated Rössler attractor as shown in Figure 3-11(c). It is apparent from Figure 3-11(c) that the data generated from the DP approximated distribution resembles closely the data sampled from the contaminated Rössler attractor Figure 3-11 (b). The Chi-square goodness of fit (GoF) test attests that there is no significant difference between the actual and DP approximated data in Figure 3-11 (b) and (c), respectively.

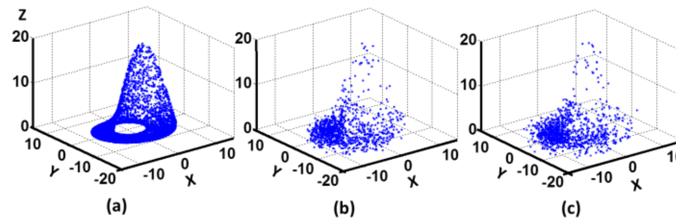

Figure 3-11 (a) The Rössler attractor delineated in (18), (b) a sample 1000 data points from the Rössler attractor contaminated with white noise $\mathbb{N}(\mathbf{0}, \boldsymbol{I}_3)$, (c) a new sample from the approximated distribution of the Rössler attractor.

In order to detect the effects of mean and variance shifts, OOC data are generated as follows (the IC state is the data obtained from (18) with white noise $\mathbb{N}(\mathbf{0},\boldsymbol{I_3})$:

- for the mean shift case, OOC data are obtained by translating the original data from (18) in all directions $(x(t), y(t), z(t))$ in the range of 0.5 to 2.5 (step size 0.5);
- for variance shifts, the OOC data are obtained by contaminating original data with different levels of Gaussian noise $\mathbb{N}(\mathbf{0}, \sigma^2\boldsymbol{I_3})$ with variance (σ^2) ranging from 1.5 to 4 (step size 0.5).

The ARL_0 of the multivariate extension of the EWMA (MEWMA), Hotelling's T^2 multivariate control chart, and DP-based SPC is maintained at 500 to obtain the control limit, and ARL_1 is assessed [7]. We use the Hotelling's T^2 instead of the CUSUM, because the Hotelling's T^2 is easier to implement than to extend the CUSUM to the multivariate case, and it is also considered one of the standard multivariate control charts [122].

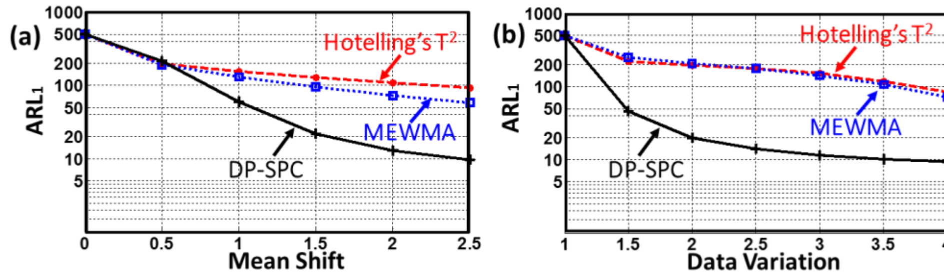

Figure 3-12 Case N3 results - OOC ARL_1 values of three SPC methods when IC ARL_0 is 500; scale on the y-axis is in natural logarithm. (a) ARL_1 results for mean shifts, (b) ARL_1 detection for data variation σ^2.

As in previous cases (Case N1 and N2), the ARL_1 results from 10,000 replications are reported for the three control charts. The ARL_1 results of the DP-based SPC are compared with Hotelling's T^2 and MEWMA in Figure 3-12. It can be inferred, based on the evidence presented in Figure 3-12, that the DP-based SPC has significantly smaller ARL_1, i.e., DP-based SPC is able to detect data shifts and variability earlier than either of the conventional control charts compared (Hotelling's T^2 and MEWMA) for multivariate, nonlinear, quasi-periodic data.

Note on Computational Time: For univariate data (Case N1 and Case N2), the computational time of DP-based SPC is about 0.02ms per data point, and that of EWMA and CUSUM is about 0.005ms per data point; for the complex multivariate data (Case N3), the computational time of DP-based SPC is about 0.2ms per data point, and that of MEWMA and Hotelling's T^2 is about 0.02ms per data point (with Intel® Core™ i7-4770 CPU@ 3.40GHz). Although DP-based SPC is slower than traditional SPC charts, it is fast enough (~50KHz for one-dimensional data, and ~5KHz for three-dimensional multivariate data) to handle many manufacturing processes (e.g., in CMP, the sampling frequency of vibration sensors is ~670Hz), and it is superior in monitoring complex signal data.

3.3.2 RHDP clustering analysis for simulated data in sequential epochs

The aim of this case study is to demonstrate the ability of the RHDP clustering to group non-Gaussian, nonstationary sequentially acquired time series data. We show that by using numerically generated data the unsupervised clustering technique of RHDP identifies specific process states contingent on their data distributions.

As noted in the preceding section, we continuously monitor the average log-likelihood value of new data as in (17). For a stationary process, the data distribution does not change over time, therefore average log-likelihood values remain stable. If the average log-likelihood values were to change dramatically, it indicates that the current data is not generated from the existing distribution but from a new one. Therefore, all the time epochs preceding the change of log-likelihood values are grouped into the same cluster, given their distribution similarity. Additionally, we note that a transition period between two process states is inevitable, because RHDP model splits the data into time epochs (windows), consequently, some windows will contain data from two temporally adjacent process states.

In this study, we define following three mixture distributions from which the data is sequentially generated (Figure 3-13 (a)):

- D1: $x_t \sim 0.5\mathbb{N}(0,0.2) + 0.5\mathbb{N}(1.5,1)$
- D2: $x_t \sim 0.2\mathbb{N}(0,0.5) + 0.8\mathbb{N}(3,0.5)$
- D3: $x_t \sim 0.5\mathbb{N}(0,0.5) + 0.5\mathbb{N}(2,0.7)$

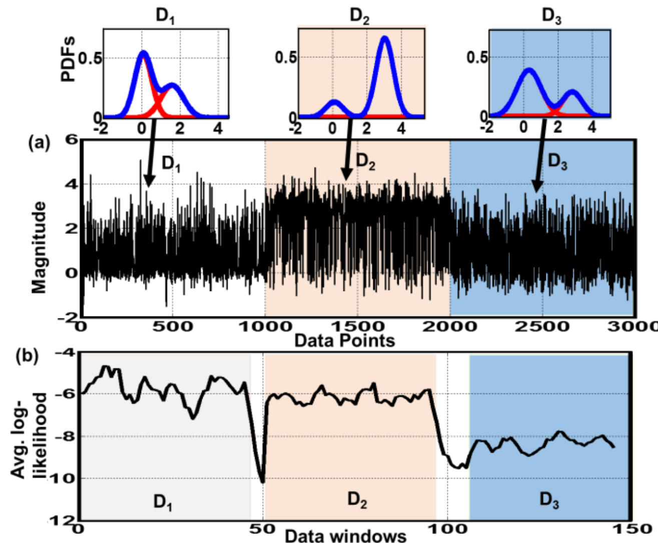

Figure 3-13 (a) Generated three-part data from different Gaussian mixture distributions, (b) average log-likelihood values of data in time epochs. Three different shades indicate data from distributions D1, D2 and D3, and the white areas between different parts of data are transition periods.

Referring to Figure 3-13 (a), the data naturally clusters into three parts, as shaded by different colors. The corresponding average log-likelihood values as estimated using RHDP are shown in Figure 3-13 (b). The unshaded parts indicate the transition periods.

We report results from a ten-replication study, and compare the clustering results from RHDP with mean shift method [20], a frequently used unsupervised clustering method; mean shift uses the raw data as opposed to utilizing average log-likelihood values by RHDP. Since the labels of sequential data are known, in order to evaluate the effectiveness of RHDP in terms of percentage of correctly clustering data, we use the F-score (precision and sensitivity) as the evaluation metric

[123]. The higher the F-score, the more accurate the model is. The clustering results are presented in Table 3-1, and it is evident that RHDP clustering has both higher precision and sensitivity compared to mean shift, consequently, the F-score for RHDP clustering is significantly higher than mean shift (98% vs 85%). This is because RHDP utilizes all the characteristics of data distribution to compute average log-likelihood values (see (17)), while mean shift only uses the average values of data.

Table 3-1 F-score results for data series with three distributions – comparison of RHDP clustering vs. mean shift (The values in the parenthesis are the standard deviation).

	RHDP Clustering			Mean Shift		
	D_1	D_2	D_3	D_1	D_2	D_3
Precision	0.9872 (0.0117)	0.9784 (0.0150)	0.9826 (0.0097)	1 (0)	0.6573 (0.2644)	0.9825 (0.0304)
Sensitivity	0.9913 (0.0194)	1 (0)	1 (0)	0.9263 (0.0798)	0.9889 (0.0248)	0.6852 (0.3060)
F-score	**0.9892**	**0.9891**	**0.9912**	0.9617	0.7897	0.8074
Average F-score	**0.9898**			0.8592		

<u>**Note on Computational Time**</u>: Due to continuous updates to the distribution estimates on sequential data, the computational time of RHDP clustering is about 1.3ms per data point (200-point window with 10-point overlap is used in this simulation), and that of mean shift is about 0.03ms per data point (with Intel® Core™ i7-4770 CPU@ 3.40GHz). Still the computational time of RHDP clustering is fast enough (with sampling frequency ~700Hz) for our application in CMP.

3.4 Application results

3.4.1 Application of integrated DP mixture model and ET approach in FFF process

We now apply the integrated Bayesian Dirichlet process (DP) mixture model and evidence theoretic (ET) approach (DP-ET) for classifying FFF data contingent on the three process states demarcated at the beginning of Sec. 3.2.2. The F-score metric (range 0~1) is used to quantify the

accuracy of this integrated DP-ET approach in comparison to other signal analysis approaches. The F-score can be interpreted as a combination of the precision and sensitivity [123]. Precision (or positive predictive value (PPV)) measures the proportion of true positives among test outcome positives, while sensitivity (or true positive rate (TPR)) measures the proportion of actual positives which are correctly identified as such. An algorithm with high F-score tends to make correct and reliable classification decisions.

The balanced F-score is the harmonic mean of precision and sensitivity:

$$\text{F-score} = 2 * \frac{\text{PPV} * \text{TPR}}{\text{PPV} + \text{TPR}}, \quad (19)$$

Before we proceed to apply this approach to our experimentally acquired FFF data, we note the following:

i. the sensors measuring the table (bed) and extruder temperature are static because they are not responsive to process dynamics, in that the machine controller will try to maintain the setting regardless of the build status; and
ii. the ambient temperature is an independent variable; the door can be opened or closed at any time.

Therefore, it is not tenable to assume that the table, extruder, and ambient temperature sensors are capable of capturing the process dynamics. We will therefore not include these sensors in our analysis. This leaves us with a total of seven sensor signals to consider, i.e., vibration in the three Cartesian axes acquired from each of the accelerometers mounted on the table (V_T) and extruder (V_E), and the meltpool temperature (T_{IR}) measured using the non-contact IR temperature sensor. However, because we have a total of seven sensor signals to study, there would be 2^7-1= 127 combinations. Considering that testing all combinations is unreasonable, we cull the number of sensor combinations using the following physical basis.

In the experimental setup, consider the accelerometer mounted on the extruder. This accelerometer has three signal channels to detect vibration of the extruder in X, Y, and Z axes. Thus, we can construct a three-dimensional mixture model for this accelerometer, since it measures the dynamics of one machine element, namely, the extruder. From a physical perspective, this implicitly assumes that the three channels of the accelerometer are statistically correlated or dependent, which is reasonable given that machine tool elements have coupled dynamics. Thus, there will be three mixture models for sensors (extruder vibration, table (or bed) vibration, meltpool) in our setup. And the number of their combinations will be reduce to 2^3-1= 7. Dempster-Shafer evidence theory can fuse any combination from all the sensors to get the best model with maximum accuracy. The F-scores for all these seven combinations are presented in Table 3-2 for the three different process states.

Table 3-2 Comparison of F-scores for the three pre-defined process states in FFF using our proposed DP-ET approach for various sensor combinations. The numbers in the parentheses indicate standard deviation over four cross-validation samples. Typically, the DP-ET approach using all three sensors, i.e., extruder vibration (V_E), table (or bed) vibration (V_T), and IR meltpool temperature (T_{IR}), demonstrates high precision and accuracy (97%) classification of the failure state.

Process State	V_E	V_T	T_{IR}	$V_E + V_T$	$V_E + T_{IR}$	$V_T + T_{IR}$	$V_E + V_T + T_{IR}$
Normal	0.500 (0.066)	0.586 (0.108)	0.741 (0.327)	0.538 (0.117)	0.738 (0.067)	**0.806** (0.113)	0.777 (0.075)
Abnormal	0.343 (0.079)	0.378 (0.117)	0.529 (0.411)	0.339 (0.124)	0.538 (0.213)	0.654 (0.241)	**0.775** (0.062)
Failure	0.492 (0.138)	0.668 (0.116)	0.911 (0.156)	0.639 (0.128)	0.876 (0.124)	0.901 (0.124)	**0.974** (0.030)
Average F-score	0.445	0.544	0.727	0.505	0.717	0.787	**0.842**

The following inferences can be drawn from these results (see Figure 3-14):

1) When the vibration sensors are a taken one at a time or a combination of vibration sensors is used without the IR sensor, the average F-scores are generally the lowest. This means using the vibration sensors alone would not be desirable.

2) The IR temperature sensor gives the most information about the process among all the sensors, because melting temperature is a direct indicator of filament extrusion and layer binding. However, it may not reliable to use the IR temperature alone for classification purposes due to large variance in the result (probably due to ambient condition fluctuations).
3) The combination of all sensors generally gives high average F-scores and low variance for all process states. Especially in abnormal and failure status, it has the highest average F-scores and lowest variance among all the combinations.

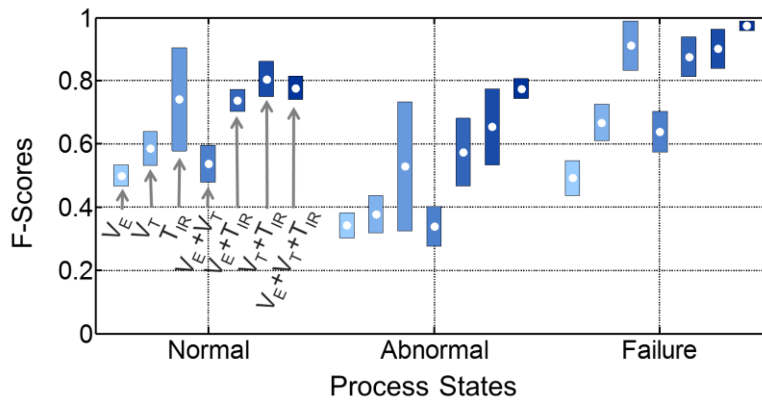

Figure 3-14 Comparison of F-scores for various sensor combinations using DP-ET approach. The error bars are one standard deviation (σ) long.

To illustrate the superiority of Dempster-Shafer evidence theory in accurate decision making, the DP-ET approach is compared with the high-dimensional DP mixture model in terms of classification accuracy. For this purpose, a seven-dimension DP mixture model is built with the same sensor information and compared with the best results from Table 3-2. The F-scores for the two methods are presented in Table 3-3 for the three different process states. The DP-ET approach gives the highest average F-scores and lowest variance for all process states (see Figure 3-15). This shows that the developed DP-ET approach is superior to the more complex high-dimensional DP mixture model.

Table 3-3 Comparison of F-scores for the three pre-defined process states in FFF using high-dimensional DP mixture modeling and our proposed DP-ET approach. The numbers in the parentheses indicate standard deviation over four

cross-validation samples. Typically, DP-ET outperforms the traditional DP approach.

Process State	High-dimensional DP	DP-ET
Normal	0.645 (0.096)	**0.777** (0.075)
Abnormal	0.623 (0.181)	**0.775** (0.062)
Failure	0.927 (0.069)	**0.974** (0.030)
Average F-score	0.732	**0.842**

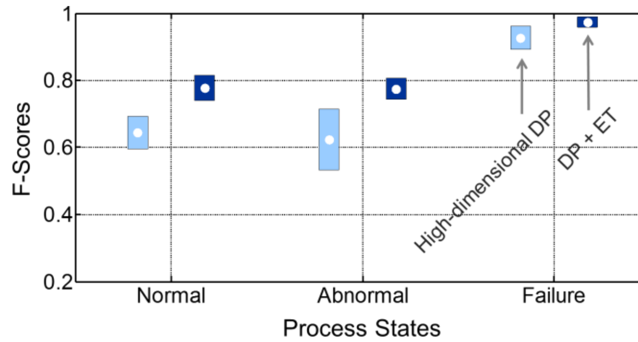

Figure 3-15 Comparison of F-scores for the three pre-defined process states in FFF using high-dimensional DP mixture modeling and our proposed DP-ET approach. The error bars are one standard deviation (σ) long.

The poor performance of high-dimensional DP may be caused by sparseness of training data. In contrast, the DP-ET approach fuses decisions from three lower dimensional DP mixture models (the maximum dimension is three) achieving the best performance since the training data may be adequate in lower dimension space.

Four different conventional classification algorithms are used to compare with our DP-ET approach. These are Quadratic Discriminant Analysis (QDA), Naïve Bayesian Classifier (NBC), Support Vector Machine (SVM), and Probabilistic Neural Networks (PNN) [124-128]. The F-scores for all of these approaches are presented in Table 3-4 for the three different process states.

Table 3-4 Comparison of F-scores for the three pre-defined process states in FFF with four widely-used classification approaches. The numbers in the parentheses indicate standard deviation over four cross-validation samples. Our proposed DP-ET approach outperforms these popular methods in classifying the respective process states.

Process States	QDA	NBC	SVM	PNN	DP-ET
Normal	0.713 (0.073)	0.725 (0.173)	0.728 (0.018)	0.622 (0.108)	**0.777** (0.075)
Abnormal	0.644 (0.130)	0.652 (0.206)	0.550 (0.259)	0.388 (0.086)	**0.775** (0.062)
Failure	0.919 (0.095)	0.886 (0.090)	0.680 (0.181)	0.689 (0.091)	**0.974** (0.030)
Average F-score	0.759	0.754	0.653	0.566	**0.842**

The following inferences can be drawn from these results (see Figure 3-16):

1) The DP-ET approach gives the highest average F-scores and generally the lowest variance for all process states. This implies that the DP-ET approach can detect the onset of FFF process anomalies more reliably and consistently than the other classification approaches studied.

2) PNN generally gives the lowest average F-scores for all process states. QDA, NBC and SVM give higher F-scores than PNN, but they show large variance in F-scores. This is probably because these methods are prone to be affected by the amount of training data. For instance, variance of F-scores becomes larger in the abnormal process state (where the training data is small) than the other two process states.

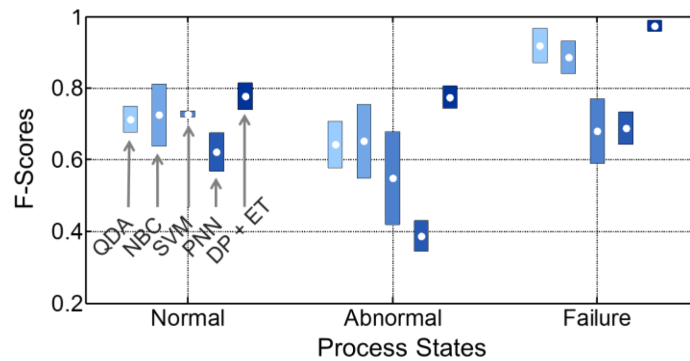

Figure 3-16 Comparison of F-scores for the three pre-defined process states in FFF with four widely-used classification approaches. The error bars are one standard deviation (σ) long.

3.4.2 Application of DP-based SPC and RHDP clustering in CMP process

In our experimental tests, certain CMP process parameters are deliberately changed to induce precisely controlled defects on the semiconductor wafer (e.g., scratches on the wafer). The following practical case studies are illustrated in this section.

1) Case E1 – changes in polishing load or downforce;
2) Case E2 – wear of the polishing pad; and
3) Case E3 – sequential changes in processing conditions.

The first two of the above cases are instances where DP-based SPC will be applied for detecting

process anomalies; the last case, Case E3, involves identification of specific anomalies using RHDP clustering.

Case E1 – Capturing changes in CMP polishing load (downforce) with data slightly violating normal assumption

The polishing load is one of the most significant factors in CMP and determines not only physical aspects, such as the nature of tribological contact, but also key process output variables, namely, material removal rate, within wafer non-uniformity, surface quality, etc. [81].

In this experiment, a change in polishing load (downforce) is monitored based on acquired vibration sensor data. As depicted in Figure 3-17 (a), after a low load (5 lb.) is active for the first half time, the load is suddenly increased to a high load (8 lb.) condition. All other factors, namely, head speed and base speed are maintained constant at 60 RPM and 150 RPM, respectively. We acquire 4000 data points in total, amounting to about 6 seconds of polishing, during which the change of load occurs approximately midway. A visible prominent shift in signal mean, as well as variation, is evident; the signal mean and variation increase with an increase in downforce.

CUSUM, EWMA, and DP-based SPC are applied to the same time series data, allowing us to compare their ARL_1 results. The control limits are adjusted *a priori* to maintain identical type I error probabilities (α-error) at 5%, this translates to an ARL_0 of 200. The results from a ten-replication study are presented in Figure 3-17 (b). Moreover, it is observed that the CMP vibration data departs from Gaussian distribution as indicated by the Anderson-Darling goodness-of-fit test. Depending on the severity (p-value) of the non-Gaussian nature of the data distribution, the DP-based SPC charts are faster (low ARL_1) compared to the CUSUM and EWMA charts. For instance, referring to Table 3-5, the DP-based SPC detects the change in polishing load within ~ 21ms (14 data points) on average, whereas CUSUM and EWMA require ~ 27ms (18 data points).

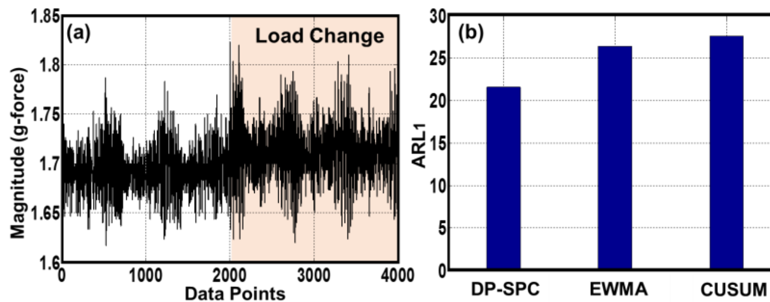

Figure 3-17 (a) Representative vibration signal patterns obtained under changing load conditions, (b) comparison of ARL_1 in changing load conditions.

Case E2 – Capturing wear of CMP polishing pad with data severely violating normal assumption

Degradation of the polishing pad is caused by wear overtime, selection of sub-optimal process conditions, or improper post-process handling [82]. For instance, inadequate post-process cleaning allows the residual slurry to dry and coagulate on the pad. Also, some portions of the polishing pad may be sheared away during polishing, thus exposing the underlying hard layer. Such polishing pads are *glazed*, i.e., the fibers of the polishing pad become entangled and lose the ability to retain slurry abrasives [82]. Polishing with a glazed pad leads to deep scratches and non-uniform wafer morphology [82].

The DP-based SPC aims to detect a degraded pad condition. It is constructed by training the in-control (IC) mixture distribution with operational data using good pads, then applied to monitor CMP runs. This is akin to building a Phase I control chart based on an *a priori* in-control process state [7]. The degraded pad is treated as the shifted process state.

In the case study, in order to verify the efficiency of DP-based SPC in detecting a degraded pad, we combine data from two experiments. The first half of the data (2000 data points, ~ 3 sec) is obtained from an experiment where a new pad is used, while the second half is gathered from an experiment conducted with a glazed pad. DP-based SPC is compared with CUSUM and EWMA in terms of detection of the pad condition change.

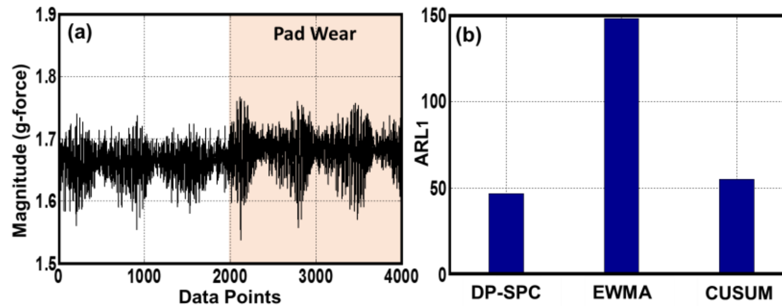

Figure 3-18 (a) Representative vibration signal patterns obtained for pad wear experiments, (b) comparison of ARL_1 for pad wear.

Figure 3-18 (a) depicts the vibration time series data gathered under the following CMP conditions: 8 lb. contact load, 150 RPM base speed and 60 RPM head speed. We discern from Figure 3-18 (a) that, not only does the mean of the vibration signal change, but also the variance of the signal slightly increases.

Moreover, compared to the previous in this instance, the vibration signals are found to depart more severely from Gaussian behavior. Therefore, DP-based SPC significantly outperforms the other two methods; it detects the pad wear earlier than CUSUM, and twice as quicker than EWMA control charts. Referring to Table 3-5, the DP-based SPC detects the change in pad wear within ~ 47ms (31 data points) on average, whereas CUSUM requires ~ 56ms (37 data points), and EWMA over 140ms (99 data points).

In addition, another polishing experiment is conducted to show the effectiveness of DP-based SPC in detecting signal change with a mildly-used pad (neither brand new nor glazed). From Figure 3-19 (a), it is noticed that there is a slight mean shift of vibration signals after switching to the mildly-used pad. Comparing with the results in Fig. 13(b), the time of detecting pad degradation by DP-based SPC increases to ~ 65ms (43 data points) on average, and CUSUM increases to ~ 71ms (48 data points) on average. Yet DP-based SPC still outperforms EWMA and CUSUM.

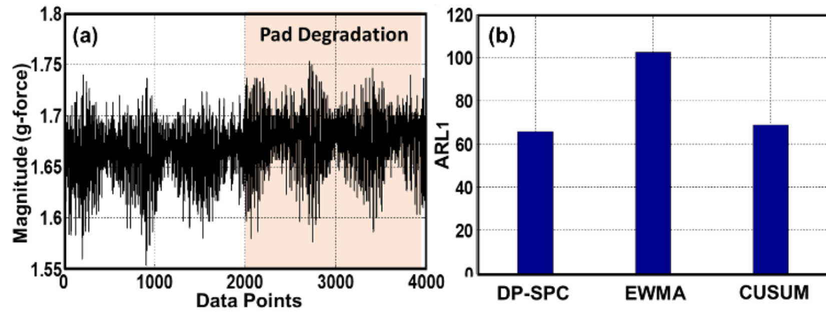

Figure 3-19 (a) Representative vibration signal patterns obtained for pad degradation experiments, (b) comparison of ARL_1 for pad degradation.

The ARL_1 results from the foregoing cases are summarized in Table 3-5. The following inferences can be obtained:

- If the in-control data slightly deviate from Gaussian distributed as in Case E1, then DP-based SPC detects the process anomalies nearly as quickly as EWMA and CUSUM.
- If the normality or symmetry conditions for the data distributions are violated severely as in Case E2, then DP-based SPC significantly outperforms CUSUM and EWMA.

These results agree closely with the implications from the numerical studies. It is noticed that the relative performance of DP-based SPC against traditional SPC drops with experimental data: while numerical case studies are generated from highly non-Gaussian and nonlinear systems (χ_1^2, bimodal data, or Rössler attractor), the experimental data manifest a modicum of similarity to Gaussian data [18, 82].

Table 3-5 Comparison of ARL_1 values for two predefined process anomalies with traditional SPC and DP-based SPC. The units are in milliseconds (ms).

Case		DP-SPC	EWMA	CUSUM
Case E1:	Load change	**21**	26	28
Case E2:	Pad wear	**47**	140	56
	Pad Degradation	**65**	102	69

Case E3 – Identifying multiple sequentially occurring CMP process anomalies using RHDP clustering

After having demonstrated the utility of DP-based SPC for detecting process anomalies in the last two cases, we now apply the RHDP unsupervised evolutionary clustering approach explained in Sec. 3.2.4 for identifying multiple sequentially occurring faults in CMP. This is important from an application standpoint for CMP process, since if the process is out of control, it is valuable to know what type of anomaly is prevalent so that the appropriate corrective action can be taken.

In this study, three different kinds of CMP process operating conditions are sequentially activated during a single experimental run, these are:

- The normal condition (C_1) occurs under nominally optimal process conditions, viz., 5 lb. polishing load, 150 RPM base speed, and 60 RPM head speed.
- Condition C_2 occurs after 3 seconds of operation (~2000 data points), the polishing load is increased to 8 lb. (the other settings are maintained at constant) for 3 seconds.
- Condition C_3 is when the slurry feed is low for 3 seconds (2000 data points) while the polishing load is kept at 8 lb.

Vibration signal patterns acquired for this experiment are presented in Figure 3-20.

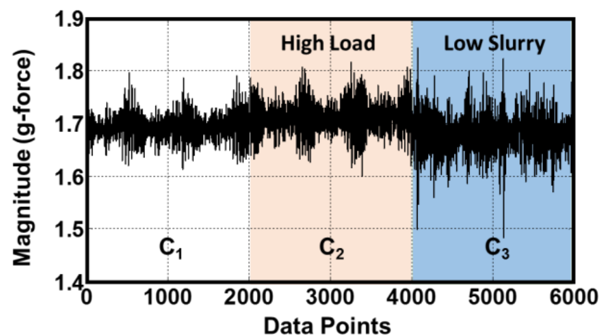

Figure 3-20 Vibration data time series for the multiple process states (Case E3), including normal condition (C_1), high load (C_2) and low slurry (C_3).

Comparison between RHDP clustering and mean shift method is also based on F-score (precision

and sensitivity) borrowed from classification. The results for Case E3, presented in Table 3-6, indicate that despite the continual change in CMP operating conditions, RHDP clustering identifies the different process states with higher precision and sensitivity compared to the conventional mean shift clustering method. For our three process states, RHDP clustering achieves an average F-score of 0.7923, which is about 10% higher than mean shift method. Moreover, the sampling frequency of RHDP clustering is ~700Hz, faster than the sampling frequency of vibration sensors (~670Hz) in CMP experiments.

Table 3-6 Clustering results for multiple process states in CMP experiment – comparison of RHDP clustering vs. mean shift (the values in the parenthesis are the standard deviation)

	RHDP CLUSTERING			MEAN SHIFT		
	C_1	C_2	C_3	C_1	C_2	C_3
PRECISION	0.8834 (0.0746)	0.7699 (0.1942)	0.8181 (0.1696)	0.9765 (0.0132)	0.7121 (0.1639)	0.7075 (0.1888)
SENSITIVITY	0.7750 (0.1554)	0.7822 (0.1911)	0.7367 (0.2557)	0.7401 (0.0548)	0.5644 (0.1679)	0.5422 (0.2253)
F-SCORE	0.8256	**0.7760**	**0.7752**	**0.8419**	0.6297	0.6139
AVERAGE F-SCORE		**0.7923**			0.6952	

3.5 Summary

In this chapter, Dirichlet process based real-time process monitoring methods are developed for FFF and CMP process.

Real-time process monitoring is proposed by combining Dirichlet Process (DP) mixture model with evidence theory to classify process states in FFF using heterogeneous sensor signals. We demonstrate that this approach can identify FFF process failures (e.g., nozzle clog) with high accuracy and reliability (average F-score ~ 85%). Consequently, this DP-based evidence theoretic framework can be valuable for real-time detection of AM process anomalies, thereby advancing closed loop process control in the future.

DP-based SPC and RHDP clustering are tested on an experimental chemical mechanical planarization (CMP) setup with obtained wireless vibration signals. They outperform traditional methods under conditions where the sensor signal patterns are nonlinear and non-Gaussian. Practical outcomes from this research are as follows: DP-based SPC detects the onset of CMP process anomalies, such as changes in pad wear, within 50 milliseconds of their inception. In contrast, the traditional methods, such as exponentially weighted moving average (EWMA) control chart has a delay of over 140 milliseconds. RHDP clustering model classifies with about 80% fidelity (F-score) multiple, sequential process drifts; traditional mean shift clustering accounts for F-score under 70%.

Consequently, this work addresses one of the significant challenges for process monitoring in advanced manufacturing applications. As part of our future research, we aim to improve the DP based modeling in the following manner:

- To increase the accuracy of distribution approximation by using extracted features instead of raw data in DP model;
- To improve the computational tractability of RHDP clustering model for high dimensional data by incorporating dimension reduction techniques.

4 Spatial Dirichlet Process Modeling based Statistical Control Scheme for Geometric Quality Assurance with Wafer Thickness Profiles

The objective of this chapter is to develop a statistical control scheme based on spatial modeling for wafer thickness profiles data in an industrial wafer slicing process. The current industrial practice uses the summary quality features for quantifying wafer thickness, such as total thickness variation, bow, and warp, which, however, underuse the abundant profile measurements and may lead to inconsistent product quality assurance. The existing spatial modeling and statistical quality control schemes for wafer thickness profiles do not take into account the clustering phenomenon among the wafers, therefore are insensitive to out-of-control wafers. In this chapter, the proposed spatial Dirichlet process (SDP) modeling utilizes the spatial clustering phenomenon existing in wafer thickness profile data and consequently the developed SDP-based statistical control scheme can detect the aberrant spatial data with higher power of test than the state of the art method. A key advantage of the proposed SDP modeling stems from its adaptive means in determining the clusters in the spatial data by using Dirichlet process modeling with a nonparametric and data-driven nature. Its effectiveness is validated by numerical simulation studies, and an actual experimental application with wafer thickness profiles acquired from an industrial wafer slicing process. The proposed SDP-based statistical control scheme detects the out-of-control wafers with much smaller type II error than the benchmark method in the comparison study.

4.1 Introduction

As raw materials for manufacturing integrated circuits, high-quality semiconductor wafer products are the prerequisite for the advancement of electronic products and Internet-based technologies. The production of wafers includes a series of sequential processes, such as ingot growth, grinding, slicing, lapping, etching, polishing, and cleaning. Each of them has a significant impact on the final product quality [129].

Slicing is one of the most critical operations in the wafer production. It refers to the cutting of the silicon ingot into slices by using multiple-wire saws with the aim to produce repeatable in-control slices in terms of geometry quality (e.g., thickness, taper, and bow) and surface quality (e.g., surface finish, surface damage) [130]. Failing to meet this aim either requires extra processing steps or rejects wafers, which directly impacts production cost and wafer yield [129]. Moreover, damages occurring during this early stage can carry over to produce defective finished devices in the following processes.

Quality features according to industrial standards [131, 132], such as total thickness variation (TTV), bow, warp, etc. [129] are usually used to summarize the wafer geometric quality. However, they only provide limited information about the geometric quality due to their utilization of a few measurements. Consequently, their contributions to rejecting out-of-control wafers and providing production insight are compromised.

With rapid development of sensor technology [133, 134], the availability of a great amount of wafer profile measurements (more than thousands on a single wafer) provides the opportunity to complement the limited geometric quality summaries by using spatial analysis for wafer profile data. The measurements are usually obtained by non-contact measuring methods (such as a wavelength scanning interferometer, or a capacitance gauge). For instance, the measurements of a real wafer thickness profile (provided by our industrial collaborator) from a slicing process are

illustrated in Figure 4-1 (a). There are about 3800 thickness measurements on a single wafer, as indicated by small black dots. Such profiles provide plenty of information about the geometric quality of the wafer. Each measurement on the thickness profile of a wafer includes both the thickness value and the location (i.e., the coordinates). Therefore, the profile measurements are eligible for spatial statistical analysis [28, 29, 65].

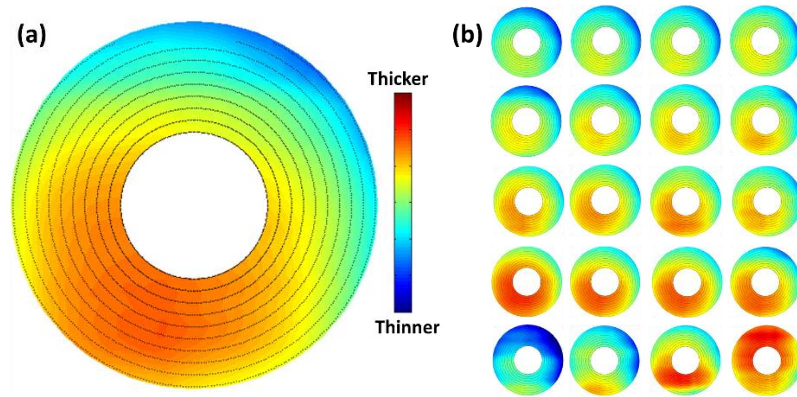

Figure 4-1 An example of the thickness profile of a wafer in a slicing process; (b) The thickness profiles of twenty wafers in the slicing process show clear clustering phenomenon.

The obtained wafer profile data also have their own complexity as shown in Figure 4-1. First, a large number of the measurements on a single wafer profile provide very high-dimensional data (more than thousands). Such data impose a great challenge for modeling and computation (e.g., implementing matrix inverse operation). Second, strong spatial correlation exists among the measurements at different measured locations, causing data redundancy. Yet it also provides opportunities to reduce data dimension. Third, spatial deviation caused by non-randomness error exists among the wafer profiles. Failing to incorporate it in statistical modeling would cause inaccurate description of the wafer profile data. Last, wafer profile data show clustering phenomenon. Such phenomenon was often neglected in previous research of wafer spatial modeling.

Specifically, clustering phenomenon among wafer profile data was reported by Pistone and Vicario [41]. The clustering phenomenon is indeed noticed from the real wafer thickness profiles in Figure 4-1 (b), by which the wafers can be grouped into different clusters. For instance, among the twenty wafers from the slicing process in Figure 4-1(b), the wafers in the first four rows can form four different clusters with the homogeneous thickness profiles (having similar thickness values at every location) in each cluster, while the wafers at the fifth row are different from all above four clusters. Admittedly, other clustering results are also possible if choosing different criteria for the homogeneity among the wafer thickness profiles. The major observation obtained here is that wafers can be clustered according to their homogeneity in thickness profiles (refer to Figure 4-1(b)).

To address the challenges of utilizing wafer profile data for quality assurance, a statistical control scheme based on spatial Dirichlet process (SDP) modeling is developed in this chapter to flexibly model the wafer thickness profiles by capturing their spatial correlation and spatial clustering phenomenon existing in wafer thickness profiles and thus effectively detect the out-of-control wafers. The proposed SDP modeling can adaptively cluster the wafer thickness profile in a data driven. Thus, no prior assumption of cluster number and structures is needed.

4.2 Research methodology

Inspired by Dirichlet process mixture model and spatial statistics, a statistical modeling by spatial Dirichlet process (SDP) is proposed, and further, an SDP-based statistical control scheme is devised. The proposed SDP modeling incorporates the deterministic modeling and stochastic modeling of the spatial data, where deterministic modeling is to estimate the spatial mean of the spatial data and stochastic modeling is to model randomness in the spatial data. The randomness in the spatial data can further be divided into spatial random effects (spatial deviation) and pure

error (non-spatial deviation). Therefore, the formulated SDP approaches can be illustrated under the framework of the spatial linear model [29] as shown in Figure 4-2.

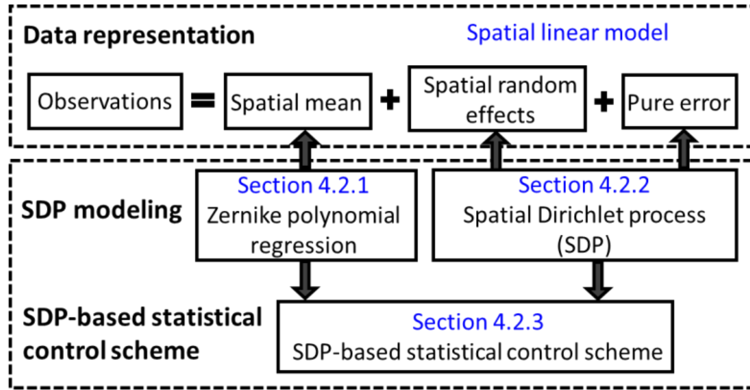

Figure 4-2 Overall methodology of the proposed SDP approaches under the spatial linear model structure.

In the proposed SDP modeling, measured locations are denoted by $\boldsymbol{s} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n)$. The observed values $\boldsymbol{Y} = (Y(\boldsymbol{s}_1), Y(\boldsymbol{s}_2), \ldots, Y(\boldsymbol{s}_n))$ at the locations $\boldsymbol{s}$ are modeled in (20) under the framework of spatial linear model [28, 65]:

$$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{G} + \boldsymbol{\epsilon}, \quad (20)$$

where $\boldsymbol{\mu}$ is the spatial mean at the locations $\boldsymbol{s}$, $\boldsymbol{G}$ is the spatial random process inducing the spatial deviation to the observations at the locations $\boldsymbol{s}$, and $\boldsymbol{\epsilon}$ is the pure error at the locations $\boldsymbol{s}$ distributed as $\mathbb{N}(\boldsymbol{0}, \tau^2\boldsymbol{I})$ (also called nugget effects to model the residuals and additional variability from replications) [28, 135]. The proposed SDP modeling utilizes the spatial correlation among measured locations to reduce the high dimensionality and estimates the spatial deviation by including spatial random effects, which capture the clustering phenomenon among the wafer profiles.

In the proposed SDP modeling, the spatial mean $\boldsymbol{\mu}$ are obtained by Zernike polynomial regression which performs well for spatial data with circular patterns [136], such as wafer profiles shown in Figure 4-1. The spatial random process $\boldsymbol{G}$ is modeled by a spatial Dirichlet process (SDP), which

not only guarantees the similar spatial correlation among the observations, but also adaptively approximates the spatial random effects for different wafer clusters. It is noticed that only the spatial random effects contribute to the wafer clustering phenomenon, since spatial mean is the same for all the wafers and pure error is independent among all the wafers. Therefore, the proposed SDP-base statistical control scheme targets on detecting out-of-control spatial data generated from aberrant spatial random effects.

4.2.1 Spatial mean fitting by Zernike polynomial regression

The spatial mean of the wafer profile data is obtained by Zernike polynomial regression [40]. Zernike polynomials form a complete, orthogonal basis over a unit disk, which is popular for spatial data within a circular area [136]. Zernike polynomials basis consists of even and odd functions, which are defined at a location $\tilde{\boldsymbol{s}} = (\rho, \varphi)$ as

$$\begin{cases} Z_n^m(\rho,\varphi) = R_n^m(\rho)\cos(m\varphi), & even\ function, \\ Z_n^{-m}(\rho,\varphi) = R_n^m(\rho)\sin(m\varphi), & odd\ function, \end{cases} \tag{21}$$

where m and n are nonnegative integers with $n \geq m$, φ is the angle, ρ is the radial distance $0 \leq \rho \leq 1$, and R_n^m are the radial polynomials defined in (22) [136].

$$R_n^m(\rho) = \begin{cases} \displaystyle\sum_{k=0}^{\frac{n-m}{2}} \frac{(-1)^k (n-k)!\,\rho^{n-2k}}{k!\left(\frac{n+m}{2}-k\right)!\left(\frac{n-m}{2}-k\right)!}, & n-m\ is\ even, \\ 0, & n-m\ is\ odd. \end{cases} \tag{22}$$

Therefore, the spatial mean value $\mu(\tilde{\boldsymbol{s}})$ at location $\tilde{\boldsymbol{s}} = (\rho, \varphi)$ could be written in terms of Zernike polynomial basis as

$$\mu(\tilde{\boldsymbol{s}}) = \sum_{i=-m}^{m} \sum_{j=0}^{n} \alpha_{i,j} Z_j^i(\rho,\varphi), \tag{23}$$

where $\alpha_{i,j}$ is the coefficient for Zernike polynomial basis.

4.2.2 Spatial randomness modeling by spatial Dirichlet process (SDP)

In the spatial linear model, the spatial random effects, generated from a spatial random process, are the source of spatial variation. Unlike the pure error, which generates non-spatial variation, the spatial random effects can be attributed to certain causes, such as different machines or operational conditions in manufacturing. Therefore, estimating the spatial random effects among different clusters of spatial data not only differentiates product quality, but also provides insight for operational conditions and production diagnosis. In the chapter, the spatial random process which generates spatial random effects is modeled by spatial Dirichlet process (SDP) with Gaussian process as its base distribution. This spatial random process is able to adaptively estimate the spatial random effects and cluster the observations according to the homogeneity of their spatial random effects in a data-driven way.

The advantage of our proposed SDP modeling is to find the clustered data without a prior of the number of clusters and the form of the spatial random effect for each cluster. By setting a random process (here is SDP) as the prior of the spatial random effects, the proposed SDP modeling obtains the posterior of the spatial random effects, which is also a random process. The posterior provides the number of clusters within the spatial data and the shared spatial random effect for each cluster. There is no parametric form for the spatial random effects either.

4.2.2.1 Spatial Dirichlet process

Dirichlet process (DP) is a stochastic process in Bayesian nonparametric modeling [6, 101, 137]. It is usually used as a prior for a mixture model, since DP does not require any *a priori* knowledge of the number of mixing components (k). The mixture model can be learned by a data-driven nonparametric Bayesian approach. DP has the form of $DP(v, G_0)$, with a precision parameter v

and a base distribution G_0. The base distribution G_0 is the mean of realizations from DP, and the precision parameter v control the spread of components in a realization.

In the proposed spatial Dirichlet process (SDP) modeling, a Gaussian process is adopted as the base distribution. A Gaussian process is a stochastic process such that any finite subcollection of random variables has a multivariate Gaussian distribution. For instance, a realization $\boldsymbol{\theta}$ from a Gaussian process GP_0 with mean $\mathbf{0}$ and covariance function $\boldsymbol{K}(\cdot,\cdot)$ is distributed as

$$\boldsymbol{\theta} \sim N\big(\mathbf{0}, \boldsymbol{K}(\cdot,\cdot)\big). \tag{24}$$

The covariance function $\boldsymbol{K}(\cdot,\cdot)$ is usually represented by a kernel function based on the assumption of a shorter distance resulting in a higher correlation. For instance, the exponential kernel function can be used in the covariance function as $\boldsymbol{K}(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma^2 \exp(-\phi|\boldsymbol{s}_1 - \boldsymbol{s}_2|)$, where σ^2 is the variance, ϕ is the decay parameter in the covariance function.

By adopting Gaussian process GP_0 as the base distribution in the Dirichlet process prior, the mixture model is extended into spatial analysis, since the Gaussian process GP_0 ensures the spatial correlations among data. A realization $\boldsymbol{G}$ from $SDP(v, GP_0)$ is discrete and admits the representation as $\boldsymbol{G} = \sum_{l=1}^{\infty} \omega_l \delta_{\boldsymbol{\theta}_l}$, where $\omega_1 = z_1$, $\omega_l = z_l \prod_{r=1}^{l}(1 - z_r)$, $l = 2, 3, \ldots$, with $\{z_r$, $r = 1, 2, \ldots\}$ *i.i.d.* from $Beta(1, v)$, and $\delta_{\boldsymbol{\theta}_l}$ denotes a point mass at $\boldsymbol{\theta}_l$, with $\{\boldsymbol{\theta}_l,\ l = 1, 2, \ldots\}$ *i.i.d.* from GP_0 [138].

In a mixture model for spatial data $\mathbb{F}(\cdot) = \sum_{l=1}^{L} \pi_l F(\cdot\,|\boldsymbol{\theta}_l)$, the number L of its unique components, the parameters $\boldsymbol{\theta}_l$ $(l = 1, \ldots L)$ and mixture weight $\pi_l (l = 1, \ldots L)$ for each unique component can all be obtained by applying SDP prior and observed spatial data in a Bayesian estimation [6, 105, 106].

Assuming the format of the components $F(\cdot\,|\,\cdot)$ are known, and the observation $\boldsymbol{Y}_p$ is generated from its respective component with parameter $\boldsymbol{\theta}_{(p)}$: $\boldsymbol{Y}_p \sim F\big(\cdot\,|\boldsymbol{\theta}_{(p)}\big)$, $\boldsymbol{\theta}_{(p)} \in \{\boldsymbol{\theta}_l,\ l =$

$1, 2, \ldots, L\}$, $p = 1, \ldots, P, L \ll P$. When a spatial Dirichlet process $SDP(v, GP_0)$ is used as the prior for the mixture model, the conditional distribution of $\boldsymbol{\theta}_{(p)}$ is

$$\boldsymbol{\theta}_{(p)} | \boldsymbol{\theta}_{(1)}, \ldots, \boldsymbol{\theta}_{(p-1)}, GP_0, v \sim \frac{v}{p-1+v} GP_0 + \sum_{l=1}^{L} \sum_{j=1}^{p-1} \frac{\delta_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}_{(j)})}{p-1+v}, \tag{25}$$

where $\delta_{\boldsymbol{\theta}_l}$ is Dirac delta function with point mass at $\boldsymbol{\theta}_l$. It is noted in (25) that SDP has the clustering functionality, i.e., the spatial data $\boldsymbol{Y}_p$ (or their respective component parameters $\boldsymbol{\theta}_{(p)}$, $p = 1, \ldots, P$) could be grouped into a few clusters and sufficiently be represented by a few unique component parameters ($\boldsymbol{\theta}_l, l = 1, \ldots L, L \ll P$).

The posterior conditional distribution on $\boldsymbol{\theta}_{(p)}$ can be calculated using Bayes' law from (26) given the observations $\boldsymbol{Y}_p$ ($p = 1, \ldots, P$) are exchangeable.

$$\boldsymbol{\theta}_{(p)} | \boldsymbol{\Theta}_{(-p)}, \boldsymbol{Y}_p, GP_0, v \sim \frac{v \cdot g(\boldsymbol{\theta}|\boldsymbol{Y}_p)}{p-1+v} \int F(\boldsymbol{Y}_p|\boldsymbol{\theta}) GP_0(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \sum_{l=1}^{L} F(\boldsymbol{Y}_p|\boldsymbol{\theta}_l) \sum_{j \neq p}^{P} \frac{\delta_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}_{(j)})}{P-1+v}, \tag{26}$$

where $\boldsymbol{\Theta}_{(-p)}$ includes the parameters for P observations except the one for the pth observation, and $g(\boldsymbol{\theta}|\boldsymbol{Y}_p) = \frac{F(\boldsymbol{Y}_p|\boldsymbol{\theta}) GP_0(\boldsymbol{\theta})}{\int F(\boldsymbol{Y}_p|\boldsymbol{\theta}) GP_0(\boldsymbol{\theta}) d\boldsymbol{\theta}}$.

By Gibbs sampling from (26), the joint posterior distribution for the parameters of all existing unique components $\{\boldsymbol{\theta}_l,\ l = 1, 2, \ldots, L\}$ can be obtained, and the mixture weight $\{\pi_l, l = 1, \ldots L\}$ for each component can be estimated by the number of observations in that component [6, 102, 138]. The implementation of Gibbs sampling is introduced in Appendix.

4.2.2.2 Spatial Dirichlet process (SDP) modeling

By using the spatial Dirichlet process as prior, the spatial linear model in (20) can be formulated as a spatial mixture model. In a random field $\mathbf{Y}_D \equiv \{\boldsymbol{Y}(\boldsymbol{s}): \boldsymbol{s} \in D\}$, $D \in \mathbb{R}^d$, let $\boldsymbol{s} = (\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_n)$ denote n specific distinct locations in D, observations $\boldsymbol{Y}_p = (Y_p(\boldsymbol{s}_1), Y_p(\boldsymbol{s}_2), \ldots, Y_p(\boldsymbol{s}_n))$, ($p = 1, 2, \ldots P$) are obtained at locations $\boldsymbol{s}$:

$$\boldsymbol{Y}_p|\boldsymbol{G},\tau^2 \sim \sum_{l=1}^{L} \omega_l\, \mathbb{N}(\boldsymbol{\mu}+\boldsymbol{\theta}_l, \tau^2\boldsymbol{I}_n),$$

$$\boldsymbol{G} = \sum_{l=1}^{L} \omega_l\, \boldsymbol{\theta}_l, \qquad (27)$$

$$\boldsymbol{G} \sim SDP(v, GP_0).$$

where $\boldsymbol{\mu}$ is the spatial mean of the observations, $\boldsymbol{\theta}_l$ is the spatial random effect for cluster l and the parameter for component l in the mixture model, ω_l is mixture weight for component l, $\boldsymbol{G}$ is a realization from the SDP prior with precision parameter v and base distribution GP_0, GP_0 is a Gaussian process with mean $\boldsymbol{0}_n$ and covariance matrix $\boldsymbol{K}(\cdot,\cdot\,|\sigma^2,\phi) \coloneqq \sigma^2\boldsymbol{H}_n(\phi)$ parameterized by the variance σ^2 and the decay parameter ϕ in correlation matrix $\boldsymbol{H}_n(\phi)$. It is noticed that spatial Dirichlet process modeling can adaptively cluster the spatial data into L cluster with respective spatial random effects.

Analyzed by the spatial linear model in (20) the spatial mean $\boldsymbol{\mu}$ of the observations is obtained by Zernike polynomial regression. Since the spatial mean $\boldsymbol{\mu}$ will be the same for all the observation and it has no effect on the clustering of observations, it can be subtracted from the observed data during the analysis for spatial random effects. Without loss of generality, we set spatial mean $\boldsymbol{\mu} = \boldsymbol{0}$, then the observations can be updated in the format of a mixture model:

$$\boldsymbol{Y}_p|\boldsymbol{G},\tau^2 \sim \sum_{l=1}^{L} \omega_l\, \mathbb{N}(\boldsymbol{\theta}_l, \tau^2\boldsymbol{I}_n),$$

$$\boldsymbol{G} = \sum_{l=1}^{L} \omega_l\, \boldsymbol{\theta}_l\,, \qquad (28)$$

$$\boldsymbol{G} \sim SDP(v, GP_0).$$

Under the Bayesian nonparametric scheme, after applying appropriate priors on parameters and hyperparameters as in (29), the proposed SDP modeling could be fitted by the MCMC algorithms [6, 102, 138]. The implementation of MCMC algorithm is introduced in Appendix.

$$\tau^2 \sim InvGamma(\alpha_\tau, \beta_\tau),$$
$$v \sim Gamma(\alpha_v, \beta_v),$$
$$\sigma^2 \sim InvGamma(\alpha_\sigma, \beta_\sigma), \quad (29)$$
$$\phi \sim U(0, \beta_\phi) \coloneqq [\phi].$$

where α_τ, β_τ, α_v, β_v, α_σ, β_σ, β_ϕ are relevant hyperparameters selected by some knowledge and trial-and-error [105].

4.2.3 Spatial Dirichlet process (SDP) based statistical control scheme

By clustering the spatial observations and estimating the spatial random effects with spatial Dirichlet process (SDP), the proposed SDP modeling not only estimates the spatial correlation structure but also models the clustering phenomenon among the spatial data. Since the spatial random effects possess information to differentiate the product quality and possibly the operational conditions, the change of spatial random effects in the spatial data could potentially indicate a change in the product quality and the operational conditions.

An SDP-based statistical control scheme is proposed here to detect the variations in product quality according to the changes in spatial random effects among spatial data. In Phase I, in-control spatial data are collected and divided as the training dataset, the validation dataset, and the testing dataset. In order to establish the statistical monitoring scheme, the training dataset from in-control spatial data are used to acquire the mixture model (as the spatial linear model in (20) and the mixture model in (28)) for the in-control conditions. The spatial random effects among the training set are identified by the proposed SDP modeling. There might be one or more spatial random effects among the training dataset, depending on their operational conditions.

The negative log-likelihood value under the mixture model is used as monitoring statistic as in (30) in order to simplify monitoring original complex spatial data into monitoring a nearly

normally-distributed statistic value. Due to the approximate normality of negative log-likelihood values, the control limits can be set by calculating statistic values of the validation dataset for a given type I error (e.g., $\alpha = 0.01$) [7]. The testing dataset is used to verify the effectiveness of the trained SDP-based statistical control scheme.

$$\mathcal{S} = -\log[L(\boldsymbol{Y}|\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_l, \dots, \boldsymbol{\theta}_L)] = -log \sum_{l=1}^{L} \omega_l \mathbb{N}(\boldsymbol{Y}|\boldsymbol{\theta}_l). \quad (30)$$

In Phase II, the proposed SDP-based statistical control scheme is used to check the conforming of the new spatial data by calculating their negative log-likelihood values under the trained in-control mixture model. Product quality changes are detected once the monitoring statistic (negative log-likelihood values) drifts outside the control limits (out-of-control).

By adaptively clustering the spatial observations based on their homogeneity, the proposed SDP modeling provides a flexible description about the spatial observations. Moreover, the proposed SDP-based statistical control scheme is sensitive to the change of spatial data but robust to the minor variations among spatial data within each cluster.

4.3 Simulations for spatial Dirichlet process modeling and statistical control

In this section, the first simulation is implemented to verify the effectiveness of the proposed spatial Dirichlet process in modeling spatial data, especially when multiple clusters exist in the spatial data. Later, the second simulation is employed to validate the detection efficiency of the proposed SDP-based statistical control scheme about the out-of-control spatial data in terms of Missed Detection (type II error). The additive Gaussian process (AGP)-based statistical control scheme, the state-of-the-art method regarding spatial analysis with wafer profile data [37], is used as a benchmark for a good faith comparison with our proposed method. As detailed in the literature review in the Section 2.2.2, the AGP model not only outperforms other Gaussian process based

methods in terms of modeling multiple thickness profiles and quantifying variations among wafers, but also overcomes the limitations of quality monitoring and establishes a new control scheme for detecting the out-of-control wafers with high sensitivity. The AGP model employs both modeling and statistical control scheme for wafer profiles data, it is the most relevant work to our research in this area of spatial modeling for wafer profile data.

4.3.1 Complex spatial data modeling with spatial Dirichlet process

The aim of the first simulation is to verify that the proposed spatial Dirichlet process (SDP) modeling is sufficient to summarize complex spatial data. Its sufficiency is validated by checking whether the posterior distributions of the parameters estimated by the proposed SDP modeling cover the true values in the underlying model, which generated the spatial data.

In this simulation, the underlying model to generate the simulated spatial observations is a zero-mean Gaussian process with variance $\sigma^2 = 2$, decay parameter $\phi = 0.1$. Spatial random effects are drawn from this underlying model and the spatial observations are obtained by adding pure error with variance $\tau^2 = 0.1$ to the spatial random effects. The spatial mean is assumed to be zero for the sake of the simplicity of the simulation, since the spatial random effects are the major contributor to the clustering phenomenon among the spatial observations. Each spatial random effect has impact on one cluster of observations. These spatial random effects incur different spatial variations for different data clusters, but they maintain the same spatial correlation structure from the underlying Gaussian process model.

For the simulated spatial data, two conditions are considered: two clusters of spatial observations and five clusters of spatial observations. For instance, the training data for the condition with two clusters of spatial observations are shown in Figure 4-3, where two different spatial random effects are generated from the underlying model and two clusters of observations are obtained by adding

pure error on each spatial random effect. These obtained spatial observations satisfy our assumption for the proposed SDP modeling: spatial data with clustering phenomenon.

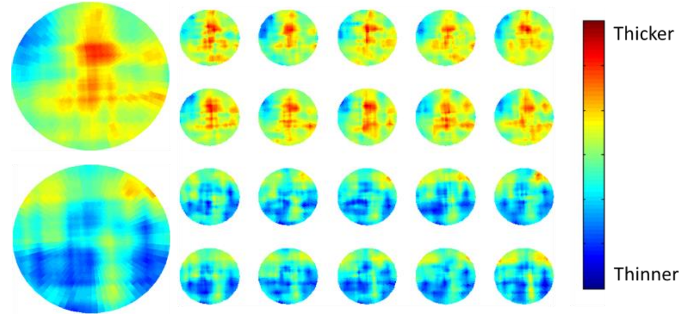

Figure 4-3 The simulated two spatial random effects and the spatial observations generated from the two spatial random effects by adding pure errors.

In this simulation, the training locations are selected by using the maximin Latin hypercube designs (LHD) [139]. As in Figure 4-4 two designs are chosen in the simulation: 50 measured locations and 100 measured locations.

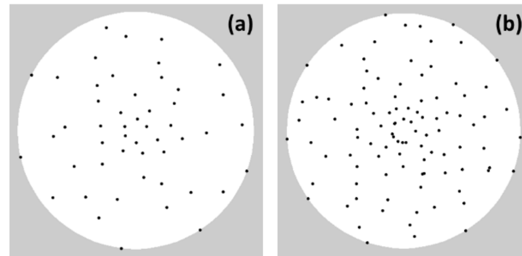

Figure 4-4 Two designs of measured locations in the simulation: (a) 50 measured locations and (b) 100 measured locations.

Therefore, total four cases are investigated in the simulation:

Case A1: two clusters of spatial observations with 50 measured locations;

Case A2: two clusters of spatial observations with 100 measured locations;

Case A3: five clusters of spatial observations with 50 measured locations;

Case A4: five clusters of spatial observations with 100 measured locations.

In each case, we simulate ten observations for every spatial random effect by adding pure errors to form the training dataset, which is used to estimate the parameters of the underlying model by using the proposed SDP modeling. Since the posterior distributions of the parameters estimated by the proposed SDP modeling do not have close form, MCMC algorithms are used to obtain their posterior distributions as introduced in Appendix. The MCMC is implemented for 10000 iterations with first 1000 iterations as the burning period. Furthermore, 100 replicates are implemented for each case. The estimates of the number of spatial random effects, spatial variance σ^2, decay parameter ϕ and pure error variance τ^2 for the 100 replicates are shown in Figure 4-5 and summarized in Table 4-1.

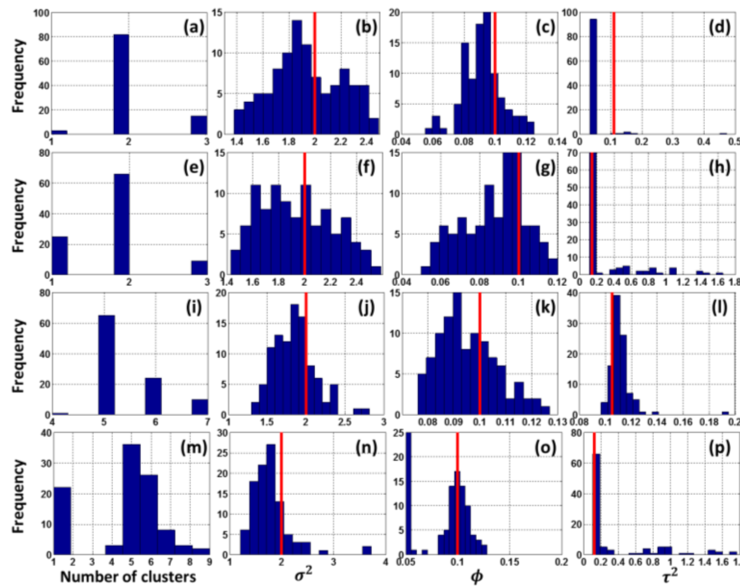

Figure 4-5 The estimates of the number of spatial random effects, spatial variance σ^2, decay parameter ϕ and pure error variance τ^2 from 100 replicates for Case A1 (a-d), Case A2 (e-h), Case A3 (i-l), and Case A4 (m-p). The red vertical lines indicate the true values.

Table 4-1 The estimates of the parameters in the underlying model for the abovementioned four cases by using spatial Dirichlet process modeling. Mode and range are used to summarize the estimation.

Case		$\sigma^2 = 2$	$\phi = 0.1$	$\tau^2 = 0.1$
A1	Mode	1.855	0.095	0.050
	Range	1.112	0.070	0.439

A2	Mode	2.040	0.095	0.152
	Range	1.161	0.070	1.545
A3	Mode	1.900	0.093	0.109
	Range	1.519	0.051	0.098
A4	Mode	1.830	0.05	0.145
	Range	2.518	0.078	1.666

By comparing the results in the aforementioned four cases, it can be inferred from Figure 4-5 and Table 4-1 that:

(1) The proposed SDP modeling is able to identify the right number of clusters in the spatial data most of the time, especially when spatial data have a few clusters and a small number of measured locations. For instance, in Cases A1, A2 and A3, the proposed SDP modeling identifies the right number of clusters in the spatial data at least 65 times out of 100 replicates. Not identifying the exact number of clusters is reasonable due to its data-driven nature (e.g., in cases when the two clusters of generated spatial observation are indeed similar and therefore clustered as one group, or when the observations generated from one spatial random effect look different after adding pure error and therefore are separated into two groups). In Case A4, the proposed SDP modeling identifies the right number of clusters only 37 times out of 100 replicates. The reason is that the proposed SDP modeling becomes a bit volatile due to its sensitivity to the choice of the hyperparameters when the number of measured locations increases.

(2) The proposed SDP modeling is sufficient to estimate the parameters of the underlying model which generates the clustered spatial data. It is noticed in Figure 4-5 all the true values of the three parameters (variance σ^2, decay parameter ϕ, and pure error variance τ^2) are within the posterior distributions in all the four cases, except the pure error variance τ^2 in Case A1.

4.3.2 SDP-based statistical control scheme for complex spatial data

The aim of the second simulation is to test the effectiveness of the proposed SDP-based statistical control scheme in detecting out-of-control spatial data. It is compared with a benchmark method - additive Gaussian process (AGP) based statistical control scheme - in terms of Missed Detection (type II error). The AGP method can model multiple thickness profiles and quantify variations among wafers. Therefore, the established new AGP-based statistical control scheme overcomes the limitations of quality monitoring and achieves prominent detection of changes in spatial data with different spatial correlation structures [37].

Under the assumption of this chapter, the spatial data are clustered and the spatial random effects are the contributors to the differences among the clustered spatial data. Without loss of generality, the spatial mean is assumed as zero, and the spatial random effects are drawn from a zero-mean Gaussian process with variance $\sigma^2 = 2$, decay parameter $\phi = 0.1$. The in-control data are generated by adding pure error with variance $\tau^2 = 0.1$ to the generated spatial random effects.

The in-control statistical control model is trained by 100 in-control training data, then the control limits are obtained by calculating the monitoring statistic of 10000 in-control validation data as in (9) and achieving type I error as 0.01 under the trained statistical control model. In order to verify the effectiveness of detecting out-of-control data, type II error of the proposed SDP-based statistical control scheme on the 10000 out-of-control data is obtained and compared with the AGP-based statistical control scheme. 100 replicates are implemented to reduce the bias caused by the generation of the spatial random effects for the in-control statistical model.

Four different types of out-of-control spatial data are used for the proposed SDP-based statistical control scheme: (1) out-of-control spatial data come from a new spatial random effect; (2) out-of-control spatial data come from a new Gaussian process with increased variance parameter σ^2; (3)

out-of-control spatial data come from a new Gaussian process with increased decay parameter ϕ; and (4) out-of-control products have increased variance τ^2 of pure error.

The aim of the first type of out-of-control data is to investigate the efficiency of detecting out-of-control spatial data with the increasing number of clusters among in-control data. We set the in-control model to include from one up to five clusters of spatial data. For the other types of out-of-control data, the aim is to investigate the efficiencies of detecting out-of-control spatial data with the changes of parameters, therefore the in-control model is fixed to have two clusters of spatial data.

Two cases are investigated in the simulation:

Case B1: spatial observations with 50 measured locations;

Case B2: spatial observations with 100 measured locations.

The comparisons of type II error between the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme under the two cases with four types of out-of-control data are shown in Figure 4-6. It is noted from Figure 4-6 that the relative performances of the AGP-based statistical control scheme and the proposed SDP-based statistical control scheme are not quite affected by the number of measured locations as in Case B1 and Case B2.

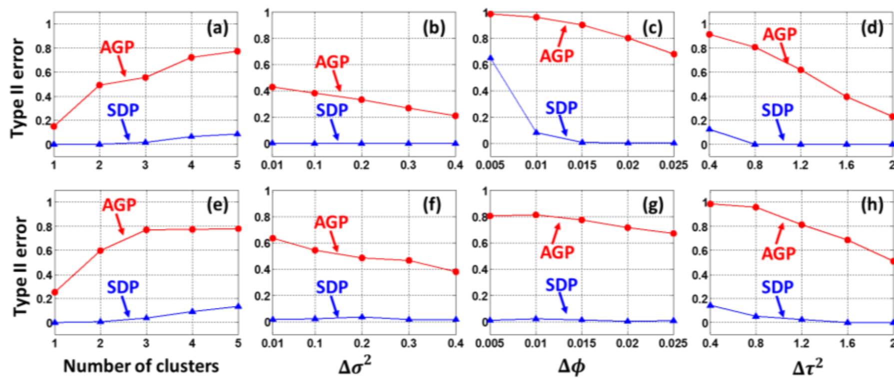

Figure 4-6 Out-of-control data detection efficiency by the AGP-based statistical control scheme and the proposed SDP-based statistical control scheme in terms of type II error for Case B1(a-d) and Case B2 (e-h).

For the out-of-control spatial data come from a new spatial random effect, increasing the number of clusters in the in-control data incurs miss detection for both statistical control schemes based on AGP and the proposed SDP. However, since the proposed SDP modeling can identify the clusters in the in-control data and label the spatial data not in the existing clusters, the type II error for the proposed SDP-based statistical control scheme is much lower than the AGP-based statistical control scheme.

For the out-of-control spatial data come from a new Gaussian process with increased parameters (either the variance σ^2 or the decay parameter ϕ), a large increase of the parameter results in an obvious change in the out-of-control data, therefore reduces miss detection for both statistical control schemes based on AGP and the proposed SDP. The proposed SDP-based statistical control scheme has consistently low type II error, since it is based on checking whether spatial data are within the existing clusters or not, very sensitive to the out-of-control data by modeling clusters in the in-control data.

For out-of-control products having increased variance τ^2 in pure error, the detection efficiency of both methods increases with the increased variance of the pure error on the observations. Since the proposed SDP modeling takes into account the pure error term in the modeling, the proposed SDP-based statistical control scheme has lower type II error than the AGP-based statistical control scheme.

Therefore, it can be inferred from Figure 4-6 that the proposed SDP-based statistical control scheme has high sensitivity to out-of-control spatial data, since it models the in-control spatial data by taking into account the clustering phenomenon among the spatial data. It outperforms the AGP-based statistical control scheme when the in-control spatial data do have clustering phenomenon.

When no clustering phenomenon exists among spatial data, the proposed SDP modeling has similar performance to the GP-based modeling because the base distribution of SDP is a Gaussian process. This scenario is under Case B1 and Case B2 with two types of simulated out-of-control spatial data: (1) the out-of-control spatial data come from a new Gaussian process with increased variance parameter σ^2; (2) the out-of-control spatial data come from a new Gaussian process with increased decay parameter ϕ. The comparisons of type II error between the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme under the two cases with two types of out-of-control data are shown in Figure 4-7.

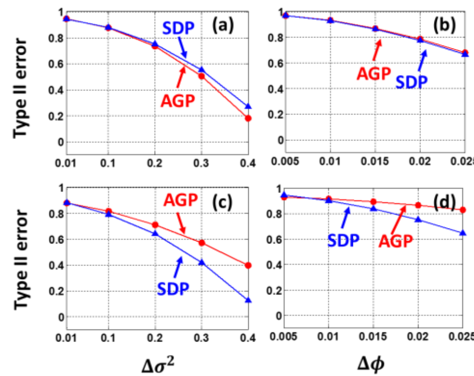

Figure 4-7 Out-of-control data detection efficiency by the AGP-based statistical control scheme and the proposed SDP-based statistical control scheme in terms of type II error for Case B1(a-b) and Case B2 (c-d) under the scenario of no clustering phenomenon existing among spatial data.

It is noted from Figure 4-7that if there is no clustering phenomenon among the in-control spatial data, the proposed SDP-based statistical control scheme would have the similar type II error rate in detecting the out-of-control spatial data as the AGP-based statistical control scheme. Consequently, from the above simulations with two scenarios of spatial data (with and without clustering phenomenon), it is evident that the proposed SDP-based statistical control scheme indeed has superior performance in detecting out-of-control spatial data over the benchmarked AGP-based statistical control scheme when spatial data has clustering structures.

4.4 Application of spatial Dirichlet process approaches for wafer thickness profiles

Eighty-eight wafers are provided by our industrial collaborator. These wafers are obtained from a slicing process, which is the early stage of wafer manufacturing, by cutting an ingot with multiple-wire saws [130]. Poor quality wafers should be identified and discarded at the early stage in the production to avoid unnecessary costs, since their rough surfaces and non-uniform thickness impact the geometric quality of final products. The motivation of applying the proposed SDP-based statistical control scheme on these wafers is to model their thickness profiles shown in Figure 4-1 by taking into account their homogeneity (refer to Figure 4-1(b)) and consequently achieve effective out-of-control wafer detection. By taking the advantage of clustering phenomenon among wafers, the proposed SDP-based statistical control scheme has high sensitivity to detect out-of-control wafers.

In the proposed SDP modeling, the wafer thickness profiles are decomposed into the spatial mean, the spatial random effects, and the pure error as in (20). The spatial mean of the 88 wafers shown in Figure 4-8 (a) is obtained by Zernike polynomial regression with truncated basis. The estimated coefficients of truncated Zernike polynomial basis are shown in Table 4-2. The profile residuals shown in Figure 4-8 (b) are obtained by subtracting the spatial mean from the measurements and analyzed by spatial Dirichlet process to find the spatial random effects for different clusters.

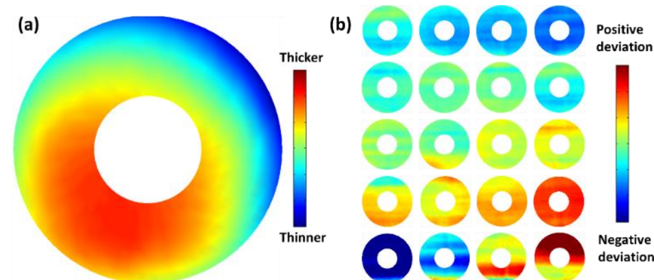

Figure 4-8 (a) The spatial mean of 88 wafers obtained by Zernike polynomial regression; (b) the profile residuals of 20 wafers including spatial random effects (spatial deviation) and pure error (non-spatial deviation).

Table 4-2 The coefficients of truncated Zernike polynomial basis for estimating the spatial mean of 88 wafers from the slicing process.

Basis indices	i	0	1	1	2	4
	j	0	-1	1	0	0
Coefficients	α_{ij}	209.82	-5.95	-4.51	-5.04	-1.16

4.4.1 Measurement sampling with profile data reduction

It is noted from Figure 4-8 (b) that for each wafer the values on the profile residual at the same horizontal level are quite similar. If we observe the wafer profile residuals from the left-side view, they can be reduced to spatial data only along one axis as shown in Figure 4-9. Such data reduction can be justified by process knowledge that the wafers are sliced in the vertical direction by the multiple-wire saws [130].

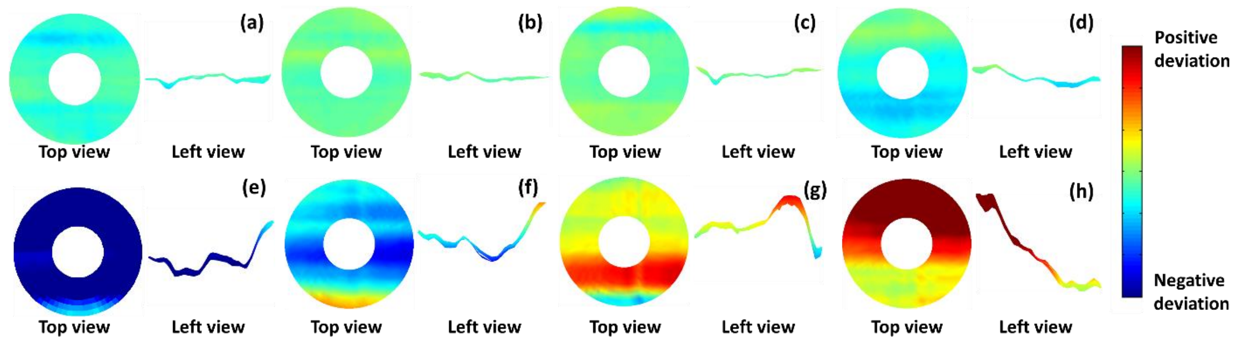

Figure 4-9 Top view and left-side view of eight wafer profile residuals. The data in the first row (a-d) belong to one cluster and both top view and left-side view show the homogeneity; the data in the second row (e-h) are all different.

Furthermore, the data reduction from wafer profile residuals to one-dimensional profiles is verified by investigating the proportion of total variance of the wafer profile residuals explained by the one-dimensional profiles. The wafer profile residuals are stratified into different groups along the horizontal levels. As observed in Figure 4-9, the data within each horizontal group are speculated to have small variation, whereas the data variation among groups would have significant contribution to the total variance of the wafer profile residuals. The percentage of total variance explained by the variation among groups for each wafer is calculated and shown in Figure 4-10.

Among the 88 wafers, there are 74 wafers, where more than 70% of the total variance is explained by the variation among the stratified horizontal groups. It means the one-dimensional profiles preserve most information about the wafer profile residuals, and validates the proposed data reduction. In addition, for the several wafers, where less than 70% of the total variance is explained by the variation among the horizontal groups, their spatial residuals are often tilted along the horizontal level, causing large data variation within each horizontal group.

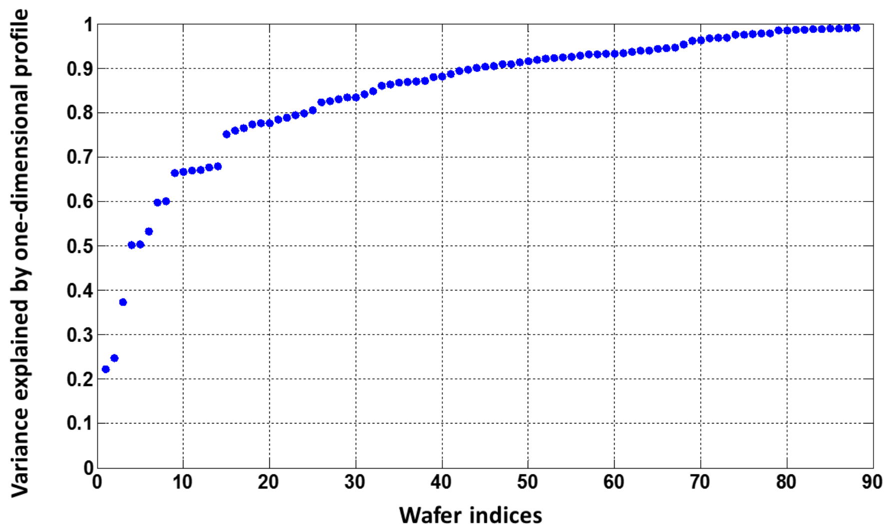

Figure 4-10 The percentage of total variance explained by the variation among groups for each wafer. 74 wafers have more than 70% total variance explained by the data variation among groups.

Therefore, in the spatial analysis of wafers from a slicing process, the computation can be greatly simplified with little compromise to the clustering information in the spatial data by taking the cutting direction into account. As shown in Figure 4-9, the wafers in the first row (Figure 4-9 (a-d)) belong to one cluster, and their spatial profiles in the left-side view are similar (close to zero with minor variation); the wafers in the second row (Figure 4-9 (e-h)) are quite different, and so are their spatial profiles in the left-side view.

The available wafers are classified into two groups as in-control (IC) and out-of-control (OOC) based on the variation of the profile residuals. Among 88 wafers, total 70 wafers are classified as

in-control since they have relatively flat profile residuals as shown in Figure 4-9 (a-d), and total 18 wafers are classified as out-of-control since their profile residuals are highly fluctuating as shown in Figure 4-9 (e-h). By using such classification, the objective of this case study is to verify the efficiency of detecting out-of-control wafers (highly fluctuating thickness) by the proposed SDP-based statistical control scheme. Thirty measured points are selected by maximin LHD [139] along the spatial profiles as shown in Figure 4-9. Ten-fold cross validation is implemented to obtain the proposed SDP-based statistical control model with 70 in-control wafers. The control limit is set with type I error ($\alpha = 0$) by calculating the monitoring statistic as in (30) for the testing dataset in cross validation. The proposed SDP-based statistical control scheme is evaluated by using18 out-of-control wafers.

The parameter estimation by the proposed SDP modeling is listed in Table 4-3. Furthermore, the proposed SDP modeling identifies six clusters among the in-control wafers based on the homogeneity of the profile residuals as shown in Figure 4-11. These clusters are formed in a data-driven way with different numbers of wafers, where some clusters have positive deviations (thicker) and others have negative deviations (thinner), and the profile residuals in each cluster look similar. Based on the obtained SDP model, if new profile residuals do not belong to the identified in-control clusters, they probably are out-of-control. The proposed SDP-based statistical control scheme detects the out-of-control wafers by evaluating the likelihood value of the profiles residuals belonging to the in-control clusters as in (30).

Table 4-3 The parameter estimation by spatial Dirichlet process model with spatial profile residuals at 30 measured locations. Mode and range are used to summarize the estimation.

	σ^2	ϕ	τ^2
Mode	0.579	0.050	0.152
Range	0.350	0.002	0.741

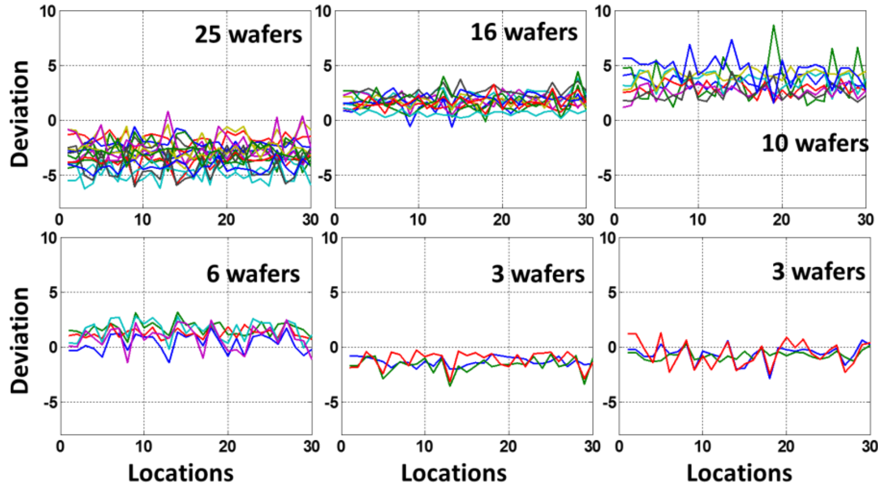

Figure 4-11 Six clusters of profile residuals are identified by the proposed SDP modeling. Different clusters have different numbers of wafers. Within each cluster, the profile residuals are similar, centering on the spatial random effect with some pure errors.

The effectiveness of the proposed SDP-based statistical control scheme in detecting the out-of-control wafers is shown in Figure 4-12 and Table 4-4. The monitoring statistic used in the proposed SDP-based statistical control scheme is the negative log-likelihood value $\mathcal{S}$ as shown in (30). The result of one cross validation is shown in Figure 4-12, where all 18 out-of-control wafers are detected with evidently larger monitoring statistic values than the control limit.

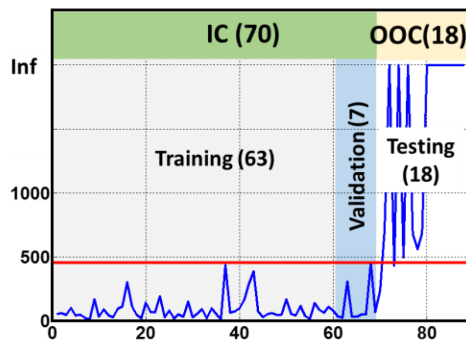

Figure 4-12 The proposed SDP-based statistical control scheme correctly detects all the out-of-control (OOC) wafers with unusually high negative log likelihood values in one fold of cross-validation.

The type II error of the ten-fold cross validation is summarized in Table 4-4 for the comparison between the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme, with the same type I error (α=0 due to limited testing dataset). The AGP-based statistical control scheme detects the out-of-control wafers with high sensitivity [37]. It is noted that by considering the clustering phenomenon existing among the wafer profiles, the average type II error for the proposed SDP-based statistical control scheme is significantly lower than AGP-based statistical control scheme (0.039 vs 0.122) with the wafer thickness profiles from the slicing process.

Table 4-4 Type II error of testing the out-of-control (OOC) wafers - comparison of the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme with sampling from one-dimensional profiles.

Methods	Type II error of ten-fold cross-validation										Avg.	Std.
AGP (Benchmark)	0.056	0	0	0	0.056	0	0.389	0.278	0.444	0	0.122	0.177
SDP (Proposed)	0.056	0	0	0	0	0	0.222	0	0.111	0	0.039	0.074

4.4.2 Measurement sampling with maximin Latin hypercube designs (LHD)

Another sampling strategy, maximin LHD [139] without considering the vertical slicing direction, is applied on the profile residuals in Figure 4-8. Fifty measured points are selected from the total 3800 measured points for the proposed SDP-based statistical control scheme, as shown in Figure 4-4. In contrast, fifty sampled points from the total 3800 measured points varies among wafers for the benchmark AGP-based statistical control scheme. Such sampling is to cater for the characteristic of AGP: it is capable of formulating spatial statistical control scheme with varying samples from different wafers.

By using this sampling strategy, total eight clusters are found among the in-control wafer by the proposed SDP modeling. The numbers of wafers in these clusters are 12, 11, 10, 8, 8, 6, 4 and 4 respectively. The proposed SDP-based statistical control scheme achieves superior detection of

out-of-control wafers by considering the clustering phenomenon existing among the wafer profiles (Table 4-5). Its average type II error is significantly lower than AGP-based statistical control scheme (0.062 vs 0.350). Comparing to the type II error results in Table 4-4, it is noticeable that the average type II error of the SDP-based statistical control scheme is consistent for these two sampling strategies. However, AGP-based statistical control scheme has much larger type II error with the maximin LHD sampling and varying samples among wafers. Despite the fact that the AGP-based statistical control scheme can detect the defective wafers with different measured points, missing the alignment among the measured points in the training data could lead to insufficient number of samples at each measured points and the increase of variation in the in-control model, consequently, resulting in the failure of detecting the out-of-control wafers with high power of test.

Table 4-5 Type II error of testing the out-of-control (OOC) wafers - comparison of the proposed SDP-based statistical control scheme and the AGP-based statistical control scheme with sampling from maximin Latin hypercube designs.

Methods	Type II error of ten-fold cross-validation										Avg.	Std.
AGP (Benchmark)	0.333	0.278	0.222	0.333	0.167	0.611	0.333	0.556	0.389	0.278	0.350	0.139
SDP (Proposed)	0.056	0.056	0.111	0.056	0.056	0.056	0.056	0.056	0.056	0.056	0.062	0.018

In this case study, total 88 wafers from a slicing process show clear clustering phenomenon and sufficiently verify the effectiveness of proposed SDP approaches. It can be inferred from this case study with wafer thickness profiles:

(1) The proposed SDP modeling can cluster the spatial profile data in a nonparametric data-driven way in terms of homogeneity of the spatial random effects.

(2) The proposed SDP-based statistical control scheme takes advantage of spatial data clustering and possesses high test power and sensitivity in detecting out-of-control spatial data.

(3) By incorporating process characteristics such as the cutting direction in a wafer slicing process, the proposed SDP modeling can simplify computation with little compromise to the spatial data clustering.

4.5 Summary

In this chapter, the spatial Dirichlet process (SDP) modeling and the SDP-based statistical control scheme are proposed to achieve flexible spatial data modeling and effective out-of-control profiles detection when clustering phenomenon exists in the spatial data. Their effectiveness is verified by numerical simulations and an experimental case study about wafer thickness profiles from a slicing process. From the numerical simulations, it is evident that the proposed SDP modeling is adequate to model complex spatial data with multiple clusters; and the proposed SDP-based statistical control scheme can achieve aberrant spatial data detection with high sensitivity and test power by taking advantage of clustering phenomenon within the spatial data. From the experimental case study, it is shown that the proposed SDP-based statistical control scheme is sensitive to the out-of-control spatial data and robust to the in-control spatial data. The practical outcome of this case study is that the proposed SDP-based statistical control scheme detects out-of-control wafers with higher power of test (average type II error of 0.039 from the 10-fold cross validation) than the benchmark method (average type II error of 0.122 from the 10-fold cross validation). Consequently, this work improves the product modeling and quality assurance in wafer profile data applications.

Clustering phenomenon among product quality data possesses additional information about the products and the processes. By incorporating such clustering phenomenon into modeling, the proposed SDP modeling and the formulated SDP-based statistical control scheme attain flexible modeling, accurate anomaly detection, and even process diagnosis. For instance, an out-of-control

wafer may indicate the malfunction of the wire saw (e.g., exhaustion of diamond splinters, partial breakage of the wire saw, etc.). These approaches can be easily generalized to different processes so long as their spatial data have clustering phenomenon.

As part of our future research, we aim to improve the proposed SDP modeling in the following manners:

- To integrate the estimation of the spatial mean by Zernike polynomial regression into the MCMC algorithm;
- To incorporate the anisotropy among spatial data into the spatial Dirichlet process modeling.

5 Spatiotemporal Modeling and Prediction of Layer-wise Porosity in Additive Manufacturing

The objective of this chapter is to model and predict the layer-wise evolution of porosity in parts made using additive manufacturing (AM) processes. This is an important research area because porosity has a direct impact on the functional performance of AM parts, such as fatigue life and strength. To realize this objective, an augmented spatiotemporal log Gaussian Cox process (AST-LGCP) model is proposed. The AST-LGCP approach quantifies the spatial distribution of pores within each layer of the AM part, and tracks their temporal evolution across layers. The spatiotemporal modeling of porosity in AM parts leads to deeper understanding of where (at what location), when (at what layer), and to what extent or severity (size and number) pores are formed in a part. This is a significant improvement over the current approach used to quantify porosity, namely, the percentage porosity relative to the bulk part volume. Unmasking the spatiotemporal behavior of pores is the first-step towards initiating remedial corrective actions, e.g., by changing the process parameters or part design. Accordingly, this work is a consequential step towards the future goal of online quality assurance of AM parts. In this paper, the AST-LGCP approach is used to predict the spatiotemporal behavior of porosity for metal parts made using a binder jetting AM process (ExOne R2). Based on non-destructive X-Ray computed tomography (CT) scans of the part, the approach predicts the areas with high risk of porosity with statistical fidelity ~85% (F-score). This approach is extensible to other AM processes such as powder bed fusion.

5.1 Introduction

The potential of additive manufacturing (AM) to transcend design and material constraints associated with traditional subtractive and formative manufacturing processes has been

conclusively demonstrated in the preceding decade [80]. For instance, functional metal AM parts made from Inconel and Titanium are entering service in aerospace and biomedical industries [80]. Despite its transformative potential, the wider exploitation of AM is encumbered due to poor process repeatability and frequent occurrence of defects. For instance, porosity is known to significantly curtail the part mechanical strength and fatigue life [72, 140]. Currently, a common quantifier for porosity in AM parts is the percentage porosity relative to the bulk volume of the part. Percentage porosity does not reveal descriptive information of pores, for example, whether there is an underlying pattern to which pores form and propagate [141].

A salient aspect of AM is that the raw material is deposited and formed simultaneously layer-by-layer. Hence, not only does the integrity of certain areas impacts that of adjacent areas within the same layer, but also the integrity of the previous layer influences subsequent layers [44, 140]. This within-layer and across-layers dependence motivates the investigation of the spatiotemporal aspects of defect formation in AM. Specifically, this work focuses on understanding the spatiotemporal behavior of porosity in a specific type of AM process called binder jetting (also called 3D Printing, 3DP).

The scientific rationale for this work is that porosity attributes such as the number, location, size, form, and type of pores are intimately connected to distinctive process phenomena [44]. By tracking and quantifying these porosity attributes a deeper understanding of the causal process phenomena is obtained. This understanding is the first-step for initiating offline or online corrective actions to improve product quality. In other words, the process parameters and design features can optimized based on quantifying the spatiotemporal dynamics of porosity formation [48, 49]. Despite significant recent developments in using X-ray computed tomography (CT) data for visualizing the internal structure of AM parts, quantitative methods for systematically

analyzing the CT scan images towards quantifying the layer-wise evolution of porosity are still not reported in the literature. This work satisfies the extant research gap [58, 61].

The objective of this work is to quantify and predict the spatiotemporal evolution of porosity in AM parts using an augmented spatiotemporal log Gaussian Cox process (AST-LGCP) model. The approach uses CT data to isolate where (at what location), when (at what layer), and to what extent or severity (size and number) pores are formed. AST-LGCP is subsequently applied to understand the spatiotemporal behavior of pores for parts made using binder jetting AM process.

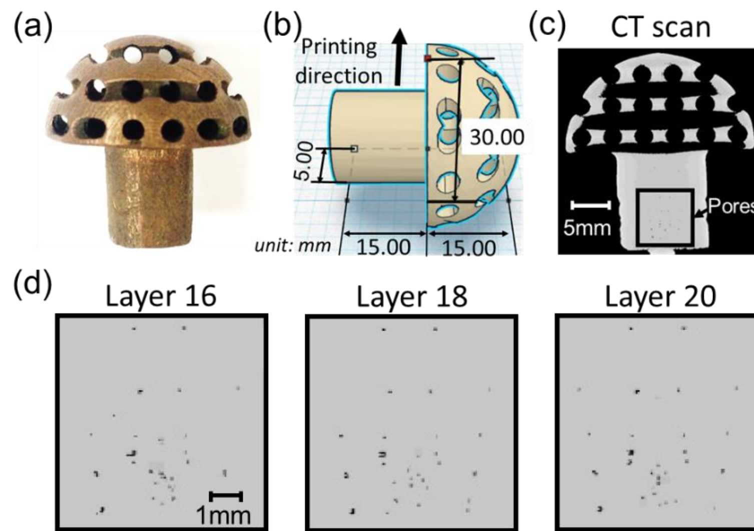

Figure 5-1 (a) A copper product fabricated by binder jetting. (b) A CT scan image with pores on the stem of the product. (c) Sequential zoomed images of pores on the stem. It is noticed that the pores on these sequential layers are dependent in space-time.

The central hypothesis is that the occurrence of pores is not independent, but that there is a spatial correlation in the distribution of pores on each layer and a temporal correlation in the distribution of pores from layer-to-layer (as shown in Figure 5-1). By modeling this spatiotemporal aspect of porosity, the occurrence of pores can be predicted with certain statistical fidelity. For instance, Figure 5-1 (a) shows a copper part made using binder jetting. The part design and the printing direction are indicated in Figure 5-1(b). The CT scan of the part, and zoomed in portions within its cylindrical stem region are shown in Figure 5-1(c) and Figure 5-1(d) respectively. The pore

patterns for three consecutive layers show that pores tend to cascade across layers and occur repeatedly in certain regions.

The rest of the chapter is structured as follows: the proposed AST-LGCP is detailed in Sec. 5.2; further corroboration of AST-LGCP with a numerical case study and application to a copper part fabricated using binder jetting is presented in Sec. 5.3; and conclusions and avenues for further research are summarized in Sec. 5.4.

5.2 Research methodology

The framework of the overall research methodology is summarized in Figure 5-2.

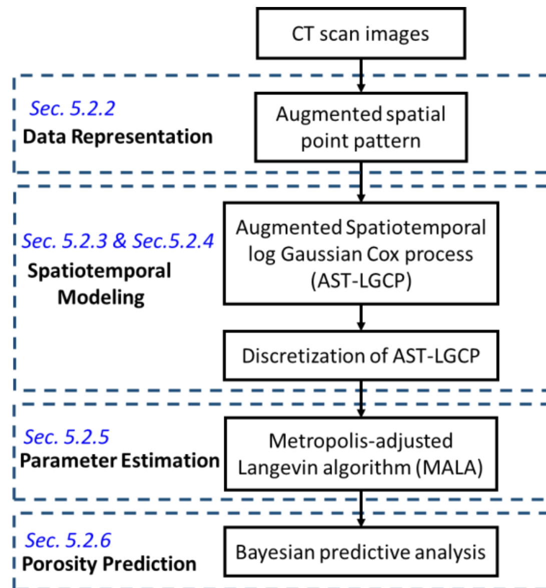

Figure 5-2 Overall methodology of AST-LGCP for layer-wise porosity modeling and prediction.

The proposed AST-LGCP methodology for layer-wise porosity modeling and prediction includes four steps:

Step 1 (Sec. 5.2.2): Pores on each CT scan image (i.e., a layer) are represented by augmented point patterns.

Step 2 (Sec. 5.2.3 and Sec. 5.2.4): AST-LGCP is formulated for spatiotemporal analysis.

Step 3 (Sec. 5.2.5): Parameters for AST-LGCP are estimated using the Metropolis-Adjusted Langevin algorithm (MALA).

Step 4 (Sec. 5.2.6): The porosity-prone areas on next layers are predicted using AST-LGCP.

Key notations that appear in the following sections are summarized in Table 5-1.

Table 5-1 Nomenclature and notations used in this work.

$\boldsymbol{W}$	The region of interest, $\boldsymbol{W} \subseteq \mathbb{R}^d$
$\boldsymbol{u}$	Spatial location of a pore, $\boldsymbol{u} \in \boldsymbol{W}$
t	Layer number of the sequential images of an AM part, $t \in [1, \mathrm{T}]$
n_t	Number of the pores in an augmented point pattern on layer t
$\boldsymbol{Z}_t$	Augmented point pattern in the region of interest $\boldsymbol{W}$ on layer t with n_t pores at locations $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{n_t}$
$\boldsymbol{Y}_t$	A realization from the Gaussian process in AST-LGCP on layer t
$\mathbf{X}$	Covariate of the realization in the Gaussian process
$\boldsymbol{\beta}$	Parameters for the covariates in the Gaussian process
$C_{\boldsymbol{Y}}$	Covariance function in the Gaussian process
σ^2	Variance parameter in the covariance matrix for the Gaussian process
ϕ	Spatial scale parameter in the covariance matrix for the Gaussian process
θ	Temporal scale parameter in the covariance matrix for the Gaussian process
$\boldsymbol{\Lambda}_t$	Intensity function for the augmented point patterns in AST-LGCP on layer t
$\bar{\mathrm{f}}_t$	Average severity parameter representing the average size of the points within the region of interest $\boldsymbol{W}$ on layer t
$\mathrm{D}_{i,j}$	The (*i, j*) cell in the discretized grid over the region of interest $\boldsymbol{W}$
M	The number of grid tracks (row or column) in the discretized region of interest
$\widetilde{\boldsymbol{Z}}_t$	Discretized point pattern on layer t, restructured as a $\mathrm{M} \times \mathrm{M}$ vector, with elements $\tilde{Z}_{(i,j),t}$ in cell $\mathrm{D}_{i,j}$
$\widetilde{\boldsymbol{Y}}_t$	Discretized random realization from the Gaussian process in AST-LGCP on layer t, restructured as a $\mathrm{M} \times \mathrm{M}$ vector, with elements $\tilde{Y}_{(i,j),t}$ in cell $\mathrm{D}_{i,j}$

5.2.1 Introduction to spatiotemporal log Gaussian Cox process (ST-LGCP) modeling

ST-LGCP is defined as a hierarchical model, the first-level of which is a Gaussian process (GP) that accommodates a nonparametric intensity function $\boldsymbol{\Lambda}_t$, where t is an AM part layer and $\boldsymbol{Y}_t$ is a realization from the GP [70, 71].

$$\boldsymbol{\Lambda}_t = \exp(\boldsymbol{Y}_t), \tag{31}$$

$$\boldsymbol{Y}_t \sim \mathrm{GP}(\mathbf{X}'\boldsymbol{\beta}, C_{\boldsymbol{Y}}), \tag{32}$$

where mean $\mathrm{E}(\boldsymbol{Y}_t) = \mathbf{X}'\boldsymbol{\beta}$, $\mathbf{X}$ is the covariate of the realization $\boldsymbol{Y}_t$ $(t = 1, \ldots, \mathrm{T})$, $\boldsymbol{\beta}$ is the parameters for the covariates. The covariance function $C_{\boldsymbol{Y}}$ is typically represented as a distance-based kernel function with the assumption that a shorter distance results in higher correlation. The radial basis function is a popular choice due to its compact form since it involves only three parameters, namely, variance (σ^2), spatial scale parameter (ϕ) and temporal scale parameter (θ). Accordingly, a separable spatiotemporal covariance function can be written as [28],

$$\begin{aligned} C_{\boldsymbol{Y}}\big((\boldsymbol{u}, t), (\boldsymbol{s}, v)\big) &= \mathrm{cov}[Y_t(\boldsymbol{u}), Y_v(\boldsymbol{s})] \\ &= \sigma^2 \exp\left(-\frac{\|\boldsymbol{u} - \boldsymbol{s}\|}{\phi}\right) \exp\left(-\frac{|t - v|}{\theta}\right), \end{aligned} \tag{33}$$

where $\boldsymbol{u}$ and $\boldsymbol{s}$ are two locations within the region of interest $\boldsymbol{W}$ on the layer t and the layer v of CT scan images respectively, $(\boldsymbol{u}, t) \in \boldsymbol{W} \times [1, \mathrm{T}]$, $(\boldsymbol{s}, v) \in \boldsymbol{W} \times [1, \mathrm{T}]$.

In the second-level of ST-LGCP, a spatiotemporal point process $\boldsymbol{Z}_t$ is used to model the point pattern data on layer t. Conditional on the intensity function $\boldsymbol{\Lambda}_t$ from the first level formulated by (31), the spatiotemporal point process $\boldsymbol{Z}_t$ for the layer-wise porosity modeling is an inhomogeneous Poison process on the layer t

$$\boldsymbol{Z}_t \sim \mathrm{Possion}[\boldsymbol{\Lambda}_t], \tag{34}$$

with the expected number of pores on the layer t image as

$$\mathrm{E}[n_t|\boldsymbol{\Lambda}_t] = \int_{\boldsymbol{u}\in\boldsymbol{W}} \Lambda_t(\boldsymbol{u})d\boldsymbol{u}. \tag{35}$$

5.2.2 Data representation by augmented point pattern

Before proceeding to the details of AST-LGCP, the augmented point pattern and average severity parameter are defined.

Definition 1 (Augmented point pattern): An augmented point pattern is a set of spatial pairs $\{(\boldsymbol{u}_i, \boldsymbol{r}_i) : i = 1, \cdots, n\}$ depicting the point i in a region of interest $\boldsymbol{W}$ with its centroid coordinate $\boldsymbol{u}_i$ and its morphology $\boldsymbol{r}_i$.

The morphology $\boldsymbol{r}_i$ could take different formats to describe the morphological features of the points, such as the size and the form. In this work, $\boldsymbol{r}_i$ is defined as a matrix representing the circumscribed rectangle of a pore, with element equal to 1 if the corresponding pixel belongs to the pore.

Figure 5-3 illustrates the use of augmented point pattern for a CT scan from an AM part. While the CT scan is a RGB image (Figure 5-3(a)), the augmented point pattern in Figure 5-3(b) represents a set of spatial pairs. To illustrate, a pore demarcated in Figure 5-3(b) is represented by a spatial pair $(u, \boldsymbol{r})$ in the augmented point pattern, where location u is the centroid coordinate of the pore $\boldsymbol{u} = (29, 36)$, and morphology of the pore is translated as a matrix $\boldsymbol{r}$.

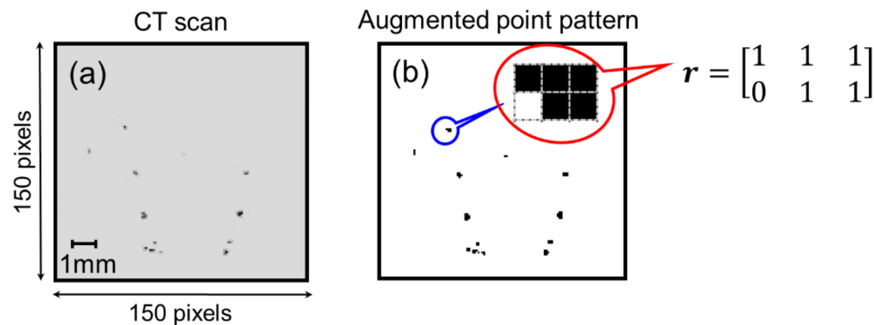

Figure 5-3 (a) A CT scan image from a metal part manufactured by binder jetting. Pores are shown as dark spots in this RGB image. (b) The visualization of augmented point pattern for this CT scan. Pores are represented by a set of spatial pairs with location and morphology. The pore in the circle occupies five pixels, and hence its information about size and form is captured.

Since the augmented point pattern has the morphological features (e.g., size and form) of pores, an average severity parameter is defined to quantify the detrimental impact of pores.

Definition 2 (Average severity parameter): An average severity parameter $\bar{\mathrm{f}}$ is defined as the average size of all the pores within a region of interest $\boldsymbol{W} \subset \mathbb{R}^d$ on a layer. It is calculated from an augmented point pattern $\{(\boldsymbol{u}_i, \boldsymbol{r}_i) : i = 1, \cdots, n\}$ as

$$\bar{\mathrm{f}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{e}_{\mathrm{C}}{}^{\mathrm{T}} \boldsymbol{r}_i \boldsymbol{e}_{\mathrm{R}}, \tag{36}$$

where $\boldsymbol{e}_{\mathrm{C}}$ and $\boldsymbol{e}_{\mathrm{R}}$ are all-ones column vectors with lengths equal to the column number and row number of morphology $\boldsymbol{r}_i$ respectively. According to this definition, larger pores on a layer lead to higher average severity parameter, which is in alignment with practical observations [142].

5.2.3 Augmented spatiotemporal log Gaussian Cox process (AST-LGCP)

AST-LGCP is defined as a hierarchical model; the first-level is to model the intensity function, which depicts the distribution-related characteristics in the augmented point pattern; the second-level is to model the augmented point pattern of pores.

In the first-level of AST-LGCP, a Gaussian process (GP) is used to accommodate the complex form of a intensity function $\boldsymbol{\Lambda}_t$ for the augmented point pattern on the layer t in a nonparametric manner with its realization $\boldsymbol{Y}_t$ [70, 71]. The intensity function $\boldsymbol{\Lambda}_t$ can be calculated via a realization $\boldsymbol{Y}_t$ on the layer t as,

$$\boldsymbol{\Lambda}_t = \frac{\exp(\boldsymbol{Y}_t)}{\bar{\mathrm{f}}_t}, \tag{37}$$

where $\bar{\mathrm{f}}_t$ is the average severity parameter for the layer t (see Definition 2), which denotes the average size of the pores on this layer.

In the second-level of AST-LGCP, a spatiotemporal point process $\boldsymbol{Z}_t$ on the layer t is used to model the augmented point pattern data. Conditional on the intensity function $\boldsymbol{\Lambda}_t$ from the first-level formulated by (37), the spatiotemporal point process $\boldsymbol{Z}_t$ for the layer-wise porosity modeling is an inhomogeneous Poison process on the layer t,

$$\boldsymbol{Z}_t \sim \text{Possion}[\boldsymbol{\Lambda}_t], \tag{38}$$

with the expected number of pores on the layer t image as,

$$\text{E}[n_t|\boldsymbol{\Lambda}_t] = \int_{\boldsymbol{u}\in\boldsymbol{W}} \Lambda_t(\boldsymbol{u})d\boldsymbol{u} = \frac{\int_{\boldsymbol{u}\in\boldsymbol{W}} \exp\big(Y_t(\boldsymbol{u})\big)d\boldsymbol{u}}{\bar{\text{f}}_t}. \tag{39}$$

5.2.4 Discretization of augmented spatiotemporal log Gaussian Cox process (AST-LGCP)

The proposed AST-LGCP approach uses a Gaussian process to model the complex intensity function of augmented point patterns in a nonparametric manner. Despite its flexibility, the Gaussian process poses a computational challenge in spatiotemporal analysis. Since the dimension of its realization $\boldsymbol{Y}_t$ as shown in Eqn. (32) for layer t $(t = 1, \cdots, \text{T})$ depends on the number of pores on the particular layer, the computational complexity in tracking the pores across layers becomes untenable. To overcome this difficulty, the CT scan images are discretized into a grid [143, 144]. Discretizing a region of interest (e.g., a unit square) on a grid with $\text{M} \times \text{M}$ cells as shown in Figure 5-4, where M is the number of horizontal (vertical) cells in the grid. The cell on the row i and column j is represented as $\text{D}_{i,j} = \left[\frac{i-1}{\text{M}}, \frac{i}{\text{M}}\right] \times \left[\frac{j-1}{\text{M}}, \frac{j}{\text{M}}\right]$, $(i = 1,2, \dots, \text{M}, j = 1,2, \dots, \text{M})$, with the centroid $\widetilde{\boldsymbol{u}}_{i,j} = (\frac{2i-1}{\text{M}}, \frac{2j-1}{\text{M}})$. With discretization, the observed point pattern $\boldsymbol{Z}_t$ within the region of interest on layer t is translated into a $\text{M} \times \text{M}$ matrix with elements $\widetilde{Z}_{(i,j),t}$ $(i = 1,2, \dots, \text{M}, j = 1,2, \dots \text{M})$ equaling the number of pixels from a certain pore in cell $\text{D}_{i,j}$, which is reshaped as a vector $\widetilde{\boldsymbol{Z}}_t$. In the same vein, the realization of Gaussian process $\boldsymbol{Y}_t$ is discretized to a $\text{M} \times \text{M}$ matrix with elements $\widetilde{Y}_{(i,j),t}$ $(i = 1,2, \dots, \text{M}, j = 1,2, \dots, \text{M})$, and then reshaped into a vector of a

multivariate Gaussian random variable $\widetilde{\boldsymbol{Y}}_t$. The discretization facilitates the subsequent computation by making both $\widetilde{\boldsymbol{Z}}_t$ and $\widetilde{\boldsymbol{Y}}_t$ on all the layers $(t = 1, \cdots, \mathrm{T})$ have the same dimension $(\mathrm{M} \times \mathrm{M})$. The choice of discretization is contingent on the smoothness of the realizations of the Gaussian field.

Due to discretization, a pore can occupy more than one cell depending on its size and form, in such a case it will be represented by the numbers of pixels in these cells together. That is, if a pore occupies two neighboring cells $\mathrm{D}_{i,j}$ and $\mathrm{D}_{i,j+1}$ on layer t, then $\tilde{Z}_{(i,j),t}$ and $\tilde{Z}_{(i,j+1),t}$ are used together to represent this pore. For instance, the augmented point pattern in Figure 5-4(b) is discretized with a 38×38 grid with each cell including 16 pixels (Figure 5-4(c)), the pore $\{\boldsymbol{u} = (29, 36), \boldsymbol{r} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}\}$ occupies two cells $\mathrm{D}_{12,9}$ and $\mathrm{D}_{12,10}$, and is represented by the number of pixels of the pore in these two cells together (one pixel and four pixels in these two cells, respectively, in this example), thus, assigning $\tilde{Z}_{(12,9)} = 1$ and $\tilde{Z}_{(12,10)} = 4$.

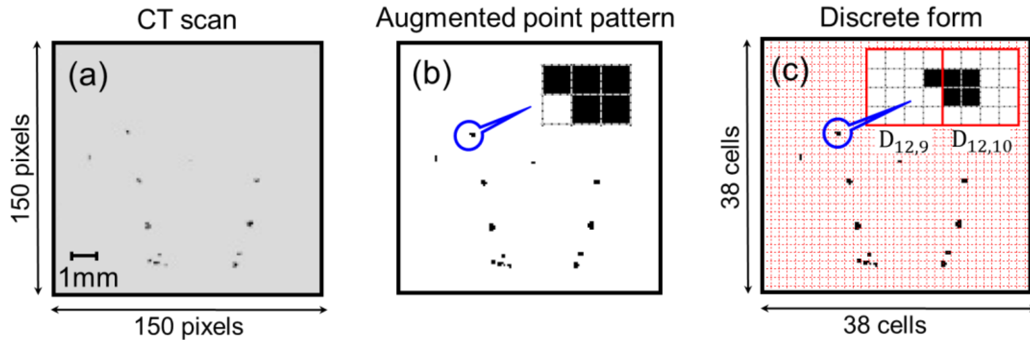

Figure 5-4 (a) A CT scan image from a metal part manufactured by binder jetting. (b) The visualization of augmented point pattern for this CT scan. A pore in the blue circle has five pixels, and hence its information about size and form is captured by such data representation. (c) After the discretization, the pore in the blue circle occupies two cells ($D_{12,9}$ and $D_{12,10}$), therefore, is represented by the discrete format of augmented point pattern ($\tilde{Z}_{(12,9)} = 1$ and $\tilde{Z}_{(12,10)} = 4$).

Moreover, since $\tilde{Z}_{(i,j),t}$ and $\tilde{Z}_{(i,j+1),t}$ represents a pore, the spatial correlation between $\tilde{Y}_{(i,j),t}$ and $\tilde{Y}_{(i,j+1),t}$ in the latent Gaussian distribution should be 1. If the exponential kernel is chosen to

calculate the correlation among different cells of pores, then the correlation matrix of the Gaussian distribution would also be an exponential function, except some entries with value 1, corresponding to these cells occupied by certain pores. Such a correlation matrix is an effective representation for the discrete augmented point pattern. However, it increases the computational cost.

Calculating the covariance between the two nonadjacent cells with the furthest distance of a pore from an augmented point pattern:

$$\begin{aligned}\operatorname{cov}\left[\tilde{Y}_{(i-a,j),t}, \tilde{Y}_{(i,j+b),t}\right] &= \sigma^2 \exp\left(-\frac{\left\|\tilde{\boldsymbol{u}}_{i-a,j} - \tilde{\boldsymbol{u}}_{i,j+b}\right\|}{\phi}\right) \\ &= \sigma^2 \exp\left(-\frac{(a^2+b^2)^{0.5}}{\phi \mathrm{M}}\right) \\ &= \sigma^2 \exp\left(-\frac{d}{\phi \mathrm{M}}\right) \approx \sigma^2,\end{aligned} \tag{40}$$

where M is the number of horizontal (vertical) cells in the grid, d is the maximum distance of the two cell (a and b are distances along two dimensions).

In the discretized AST-LGCP for pore modeling, the covariance matrix of the multivariate Gaussian random variable $\tilde{\boldsymbol{Y}}_t$ can be approximated by a distance-based kernel function as in (40), if the following three assumptions are satisfied:

1) The pores are not overly big, i.e., d is small;
2) the grid is reasonably fine, i.e., M is large; and
3) the spatial correlation is not negligible, i.e., ϕ is large.

The justification of the above statement is as follows. In the case study with a metal part fabricated by using binder jetting, a fine grid with $\mathrm{M} = 50$ is used to discretize the CT scan image. The pores observed typically span less than 10 cells ($d < 10$), and the spatial scale parameter ϕ is also

estimated around 10 ($\phi \approx 10$). Consequently, $\exp\left(-\frac{d}{\phi \mathrm{M}}\right) \approx 1$. Therefore, based on the CT scan images for the AM part, the choice of these parameters with a simple parametric exponential kernel as in Eqn. (40) is found to be adequate for the covariance matrix in the discretized format of AST-LGCP.

This leads to the following discretized format of AST-LGCP model. The Gaussian process at the first-level is approximated by multivariate Gaussian distribution $\mathbb{N}(\cdot)$ on the computational grid:

$$\widetilde{\boldsymbol{Y}}_t \sim \mathbb{N}(\widetilde{\mathbf{X}}'\boldsymbol{\beta}, C_{\widetilde{\boldsymbol{Y}}}), \tag{41}$$

where $\widetilde{\mathbf{X}}$ is the covariates of $\widetilde{\boldsymbol{Y}}_t$ on the computational grid on layer t $(t = 1, \dots, \mathrm{T})$, and a separable spatiotemporal covariance function is defined as

$$\begin{aligned} C_{\widetilde{\boldsymbol{Y}}}\left((\widetilde{\boldsymbol{u}}_{i,j}, t), (\widetilde{\boldsymbol{u}}_{i',j'}, t')\right) &= \operatorname{cov}\left[\widetilde{Y}_{(i,j),t}, \widetilde{Y}_{(i',j'),t'}\right] \\ &= \sigma^2 \exp\left(-\frac{\left\|\widetilde{\boldsymbol{u}}_{i,j} - \widetilde{\boldsymbol{u}}_{i',j'}\right\|}{\phi}\right) \exp\left(-\frac{|t - t'|}{\theta}\right), \end{aligned} \tag{42}$$

where (i, j) and (i', j') are the indices of cells occupied by pores $(i, i' = 1,2, \dots, \mathrm{M}, j, j' = 1,2, \dots, \mathrm{M})$ on the layer t and the layer t' respectively $(t, t' \in [1, \mathrm{T}])$. Denote $\eta = \{\sigma^2, \phi, \theta\}$ for notation simplicity in the MCMC algorithm in Sec. 5.2.5.

A spatiotemporal point process $\widetilde{\boldsymbol{Z}}_t$ for layer t is at the second-level with the intensity function

$$\widetilde{\boldsymbol{\Lambda}}_t = \frac{\exp(\widetilde{\boldsymbol{Y}}_t)}{\bar{\mathrm{f}}_t}, \tag{43}$$

where $\bar{\mathrm{f}}_t$ is the average severity from (36) on layer t, which is estimated as the ratio between the number of pores ($\hat{n}_t$) and the number of cells ($\widehat{N}_t$) occupied by pores on layer t: $\bar{\mathrm{f}}_t \approx \frac{\hat{n}_t}{\widehat{N}_t}$.

Conditional on the intensity function $\widetilde{\boldsymbol{\Lambda}}_t$, the spatiotemporal point process $\widetilde{\boldsymbol{Z}}_t$ is expressed as an inhomogeneous Poisson point process as follows,

$$\widetilde{\boldsymbol{Z}}_t \sim \text{Possion}[\widetilde{\boldsymbol{\Lambda}}_t], \tag{44}$$

and the expected number of pores in region $\boldsymbol{W}$ is calculated as

$$\mathrm{E}[n_t|\widetilde{\boldsymbol{\Lambda}}_t] \cong \sum_{i=1}^{\mathrm{M}}\sum_{j=1}^{\mathrm{M}} \tilde{\Lambda}_{(i,j),t} * \mathrm{C_A}. \tag{45}$$

where $\tilde{\Lambda}_{(i,j),t}$ is the Poisson rate in the cell $\mathrm{D}_{i,j}$ of the computational grid at layer t, and $\mathrm{C_A}$ is the cell area.

5.2.5 Parameter estimation in AST-LGCP

Bayesian estimation is used to obtain the posterior distribution of the parameters from the prior belief functions of parameters and the observed augmented point patterns. This leads to a way for estimating parameters in AST-LGCP, which includes the parameters $(\eta, \boldsymbol{\beta})$ of the Gaussian process and the random process $\widetilde{\boldsymbol{Y}}_t$ (used for predicting the intensity function $\widetilde{\boldsymbol{\Lambda}}_t$ in the region of interest) in (41)-(43) [143, 144].

Through space-time discretization, the likelihood function of the augmented point patterns (to layer t) $\pi(\widetilde{\boldsymbol{Z}}_1, \dots, \widetilde{\boldsymbol{Z}}_t | \eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t)$ and the priors $\pi(\eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t)$ are obtained from finite dimensional distributions over the region of interest, enabling parameter estimation for Gaussian the process $(\eta, \boldsymbol{\beta})$ and prediction for the random process $\widetilde{\boldsymbol{Y}}_t$ in a Bayesian framework.

$$\begin{aligned}\pi(\eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t | \widetilde{\boldsymbol{Z}}_1, \dots, \widetilde{\boldsymbol{Z}}_t) &\propto \pi(\widetilde{\boldsymbol{Z}}_1, \dots, \widetilde{\boldsymbol{Z}}_t | \eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t)\pi(\eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t) \\ &= \pi(\widetilde{\boldsymbol{Z}}_1, \dots, \widetilde{\boldsymbol{Z}}_t | \eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t)\pi(\eta)\pi(\boldsymbol{\beta})\pi(\widetilde{\boldsymbol{Y}}_t).\end{aligned} \tag{46}$$

Accordingly, $\pi(\widetilde{\boldsymbol{Y}}_t)$ is the corresponding finite dimensional Gaussian distribution in the grid on layer t, and the likelihood function $\pi(\widetilde{\boldsymbol{Z}}_1, \dots, \widetilde{\boldsymbol{Z}}_t | \eta, \boldsymbol{\beta}, \widetilde{\boldsymbol{Y}}_t)$ is a joint distribution of multiple Poisson distributions in the cells up to layer t. $\pi(\eta)$ and $\pi(\boldsymbol{\beta})$ are Gaussian distributions with mean and variance set as the estimated values from minimum contrast parameter estimation [144].

Consequently, the posterior distribution $\pi(\widetilde{\boldsymbol{Y}}_t|\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_t)$ can be obtained by marginalizing η amd $\boldsymbol{\beta}$ in Eqn. (46) [144].

A Metropolis-adjusted Langevin algorithm (MALA), a Markov chain Monte Carlo approach, is used to obtain sample-based estimate for the posterior distributions where the target distribution $\pi(\eta,\boldsymbol{\beta},\widetilde{\boldsymbol{Y}}_t|\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_t)$ is approximated by sequential samples $\left\{\eta^{(j)},\boldsymbol{\beta}^{(j)},\widetilde{\boldsymbol{Y}}_t^{(j)}\right\}_{j=1}^{N}$from a Markov chain whose stationary distribution is the target [145, 146]. The design of the proposed density q herein is a mix of random walk and Langevin kernels. It is used to exploit the gradient information on the target to help guide movements towards areas of higher posterior probability [147]. The samples drawn from the proposal density q are accepted in a probabilistic way. The samples are drawn from the proposal density q and are accepted in a probabilistic way. For instance, in the *j*th step of the algorithm, a candidate $\{\eta^*,\boldsymbol{\beta}^*,\widetilde{\boldsymbol{Y}}^*\}$ is drawn from the proposal density $q(\eta^*,\beta^*,\widetilde{\boldsymbol{Y}}^*|\eta^{(j-1)},\beta^{(j-1)},\widetilde{\boldsymbol{Y}}^{(j-1)})$ and accepting it as the *j*th sample, i.e., setting $\{\eta^{(j)},\beta^{(j)},\widetilde{\boldsymbol{Y}}^{(j)}\}$=$\{\eta^*,\beta^*,\widetilde{\boldsymbol{Y}}^*\}$, with probability

$$\min\left\{1,\frac{\pi(\eta^*,\boldsymbol{\beta}^*,\widetilde{\boldsymbol{Y}}_t^*|\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_t)}{\pi\left(\eta^{(j-1)},\boldsymbol{\beta}^{(j-1)},\widetilde{\boldsymbol{Y}}_t^{(j-1)}\middle|\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_t\right)}\frac{q\left(\eta^{(j-1)},\boldsymbol{\beta}^{(j-1)},\widetilde{\boldsymbol{Y}}_t^{(j-1)}\middle|\eta^*,\boldsymbol{\beta}^*,\widetilde{\boldsymbol{Y}}_t^*\right)}{q\left(\eta^*,\boldsymbol{\beta}^*,\widetilde{\boldsymbol{Y}}_t^*\middle|\eta^{(j-1)},\boldsymbol{\beta}^{(j-1)},\widetilde{\boldsymbol{Y}}_t^{(j-1)}\right)}\right\}. \tag{47}$$

5.2.6 Porosity prediction for future layers in AST-LGCP

Since the observations are time-dependent, a further step prediction of the intensity function and then the porosity-prone areas become possible, and subsequently contributes to AM online quality assurance. Represented in the grid, the task here is to predict the multivariate Gaussian distribution of $\widetilde{\boldsymbol{Y}}_{T+1}$ given the observed data $\{\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_T\}$ up to layer T as in $\pi(\widetilde{\boldsymbol{Y}}_{T+1}|\widetilde{\boldsymbol{Z}}_1,\dots,\widetilde{\boldsymbol{Z}}_T)$, and consequently estimate the intensity function of observed augmented point patterns $\widetilde{\boldsymbol{\Lambda}}_{T+1}$ with the mean of $\widetilde{\boldsymbol{Y}}_{T+1}$ with [143, 144]. A practical approach is to adopt recent observed data

$\{\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\}$, (L is a user-defined lag parameter), which have a bigger impact on the future observations, to predict the distribution $\pi(\widetilde{\boldsymbol{Y}}_{\mathrm{T+1}}|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}})$ [70]. The conditional independence properties of the model imply $\pi\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T+1}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right)$ follows a multivariate normal distribution with mean

$$\mathrm{E}\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T+1}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right) = \xi \mathrm{E}\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right) + (1-\xi)\widetilde{\mathbf{X}}'\boldsymbol{\beta}, \tag{48}$$

and variance,

$$\mathrm{var}\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T+1}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right) = \xi^2 \mathrm{var}\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right) + (1-\xi)^2 C_{\widetilde{\mathbf{Y}}}, \tag{49}$$

where $\xi = \exp(-\theta)$ represents the temporal evolution of porosity across layers; $\pi\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right)$, viz., the distribution of $\widetilde{\boldsymbol{Y}}_{\mathrm{T}}$, can be estimated from Eqn. (46) by MALA in Sec.5.2.5.

Finally, the predicted intensity function in the region of interest is calculated from (48) as follows,

$$\widetilde{\boldsymbol{\Lambda}}_{\mathrm{T+1}} = \exp\left(\mathrm{E}\left(\widetilde{\boldsymbol{Y}}_{\mathrm{T+1}}\middle|\widetilde{\boldsymbol{Z}}_{\mathrm{T-L}}, \dots, \widetilde{\boldsymbol{Z}}_{\mathrm{T}}\right)\right). \tag{50}$$

The predicted intensity function indicates the potential for pores to occur within the region of interest based on the information from previous layers. High intensity at certain locations suggests high probability of the occurrence of pores at those locations. It is utilized to characterize the spatial distribution of pores on the future layers and predict the high-risk areas prone to the occurrence of pores.

5.3 Application of AST-LGCP for prediction of porosity

In this section, the proposed AST-LGCP is first illustrated with numerically generated sequences of pores, and subsequently, applied to a metal part fabricated using binder jetting AM process. The aim to predict the porosity distribution in a subsequent layer given the porosity of preceding layers. In addition, the spatiotemporal log Gaussian Cox process (ST-LGCP) [70] is also proposed for this new application of layer-wise porosity prediction in AM part.

The prediction results are compared with a benchmark method in terms of the statistical fidelity (F-score). The benchmark method is an empirical approach previously applied by Tammas-Williams *et al*. [61], that directly uses all the pores on previous layer images for prediction and defines the areas with pores on existing layers as the high-risk areas on the next layer.

5.3.1 Spatiotemporal analysis for numerically simulated porosity data

Five sets of data with spatiotemporally correlated pores are numerically generated from the spatiotemporal model using Eq. (37) - (38), where the spatial scale parameter(ϕ) and temporal scale parameter (θ) determine the spatiotemporal correlation among the simulated pores, and the variance (σ^2) regulates the dispersion of the pores within the region of interest and across different layers. One such set of the sequential images generated numerically is shown in Figure 5-5. These images are generated by setting $\sigma^2 = 4$, $\phi = 10$, and $\theta = 10$. The aim of this study is to predict the high-risk areas with occurrence of pores on the fifth layer based on the previous four layers. The reason for choosing the previous four layers is due to the gradual decrease in the temporal correlation between layers. The temporal correlation between layer 1 and the fifth layer in this study is decreased to ~0.6.

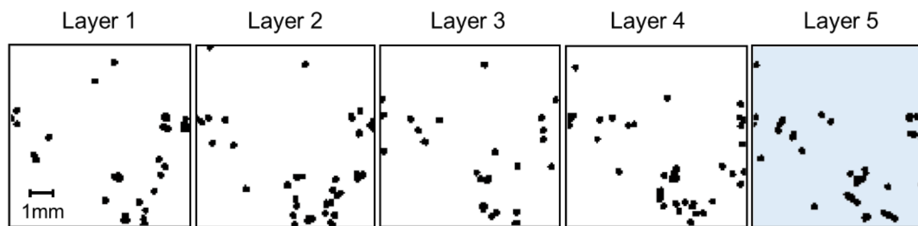

Figure 5-5 One set of simulated data is shown as sequential images of pores with spatial and temporal correlations. The aim of this simulation is to predict the high-risk areas with occurrence of pores on the fifth layer based on the previous four layers.

AST-LGCP is used to model the obtained augmented point patterns from the simulated data in Figure 5-5. Using Metropolis-adjusted Langevin algorithm (MALA), 5000 iterations of sampling are implemented for AST-LGCP with the first 1000 iterations as burn-in period. The model

converges swiftly after the burn-in period. The parameter estimates (variance σ^2, spatial scale parameter ϕ, temporal scale parameter θ) by using AST-LGCP are summarized in Table 5-2, from which the parameters estimated by AST-LGCP are within 8% of their true values.

Table 5-2 Parameter estimates (variance σ^2, spatial scale parameter ϕ, temporal scale parameter θ in (42)) using the proposed AST-LGCP. The numbers in parenthesis are the standard deviation of the estimates (4000 iterations in MALA).

Parameter notations in the spatiotemporal model	σ^2	ϕ	θ
True values of parameters for simulated data (Figure 5-5)	4	10	10
Parameter estimates using AST- LGCP	3.750 (0.730)	9.246 (0.669)	10.781 (2.897)
Relative error between the estimates and the true values	6.25%	7.54%	7.81%

With a separable covariance structure in AST-LGCP, the spatial covariance among the pores within a layer are shown in Figure 5-6 (a), whereas the temporal correlation among the pores across layers are shown in Figure 5-6 (b). It is noted that both spatial correlation and temporal correlation decreases with the increasing distance within a layer and with progressive layers. Moreover, the predicted intensity function in the region of interest for the fifth layer using AST-LGCP is shown in Figure 5-6 (c), which indicates areas prone to porosity and their corresponding severity.

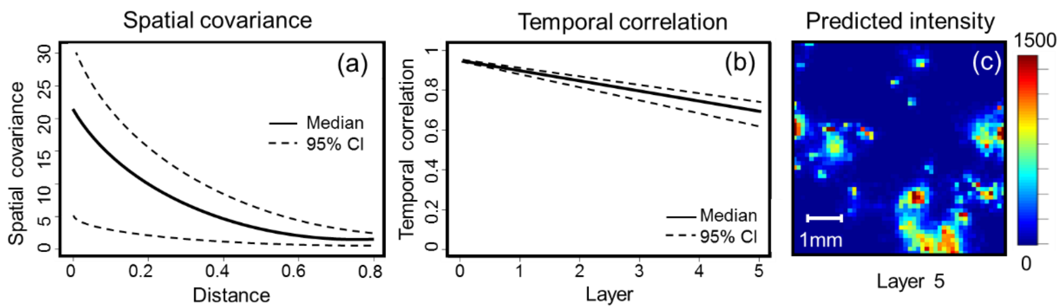

Figure 5-6 For the set of simulated data in Figure 5-5, (a) spatial covariance of the pores along the distance, (b) temporal correlation of the pores among sequential layers, and (c) the predicted intensity function in the region of interest for the fifth layer.

The statistical accuracy of AST-LGCP and ST-LGCP in predicting the porosity-prone areas in the future layers is compared with the empirical approach used by Tammas-Williams *et al*. [61]. The

identified high-risk areas are juxtaposed in Figure 5-7. The empirical approach overestimates the occurrence of pores by identifying overly segmented high-risk areas (Figure 5-7(b)); ST-LGCP underestimates the high-risk areas (Figure 5-7 (c)); AST-LGCP marks several relatively large and contiguous areas (Figure 5-7 (d)).

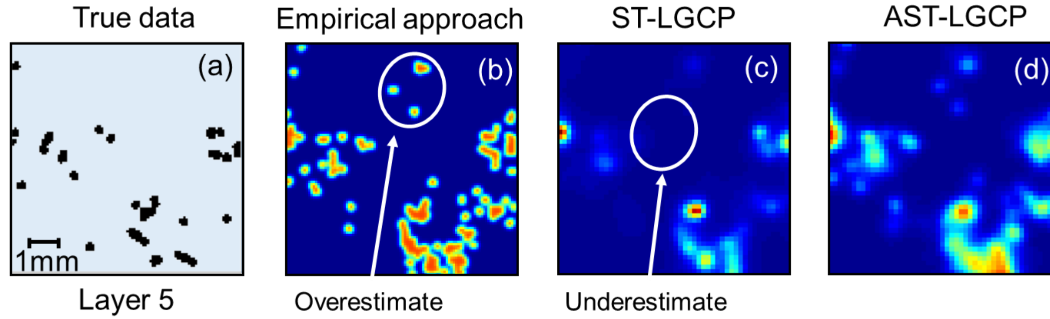

Figure 5-7 (a) The simulated pores on the fifth layer; (b) overly segmented high-risk areas identified by the empirical approach (benchmark method); (c) small high-risk areas identified by ST-LGCP; and (d) relatively large and continuous high-risk areas identified by the proposed AST-LGCP.

A statistical measure, F-score (a combination of precision and sensitivity), is used to quantify the accuracy of the three methods in predicting the porosity-prone areas on the next layer [123]. In this context, precision represents the percentage of real pores among the predicted ones, and sensitivity indicates the percentage of real pores correctly identified. In the implementation of F-score calculation for AST-LGCP, three major steps are taken:

1) Normalize the predicted intensity in the region of interest $\widetilde{\boldsymbol{\Lambda}}_{\mathrm{T}+1}$ in Eqn. (50) into cell-wise probability of the occurrence of pores;
2) Threshold the cell-wise probability by setting the probability to zero if it is smaller than the threshold for *inactive cell* or keeping the probability value if it is larger than the threshold for *active cell*;
3) Consider the region within *active cells* as the prediction of pores, and the region within *inactive cells* as the prediction of normal condition. With these F-score can be calculated as,

$$\text{Precision} = \frac{\text{areas of } \mathit{active\ cells} \text{ with pores}}{\text{total areas of } \mathit{active\ cells}},$$

$$\text{Sensitivity} = \frac{\text{number of pores within } \mathit{active\ cells}}{\text{total number of pores}}, \quad (51)$$

$$\text{F} - \text{score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}.$$

The empirical approach and ST-LGCP adopt the same steps to calculate the F-score for an adequate comparison. Since the empirical approach does not generate an intensity function, it assigns probability one to the cells with pores from previous layers, and probability zero to the cells without pores from previous layers in step 1).

The F-score results for the total five sets of simulated sequential images are summarized in Table 5-3. From this, it is evident that AST-LGCP has the highest F-score in predicting high-risk areas on the next layer. It has high precision and sensitivity, since it has robust performance in tracking the pores on the next layers by incorporating the pore size into modeling. In contrast, ST-LGCP has low sensitivity because of its tendency to underestimate the porosity-prone areas since it disregards the pore morphology information. The empirical approach has very low precision, because it ignores the spatiotemporal correlation among pores and generally overestimates the high-risk areas.

Table 5-3 F-score results for five sets of simulated data – Comparison of the empirical approach, ST-LGCP, and AST-LGCP (The values in the parenthesis are the standard deviation for five replications).

	Empirical approach	ST-LGCP	AST-LGCP
Precision	0.6866 (0.1470)	0.7898 (0.0846)	**0.8089 (0.1061)**
Sensitivity	0.9589 (0.0368)	0.7744 (0.1428)	**0.9779 (0.0186)**
F-score	0.7910 (0.0835)	0.7627 (0.0420)	**0.8884 (0.0460)**

5.3.2 Application of AST-LGCP for quality assurance in binder jetting

In this case study, the effectiveness of AST-LGCP in predicting high-risk areas prone to pores on next layer is verified with a copper part (Figure 5-1(a)) made on the authors' binder jetting machine (ExOne R2).

Binder jetting: an additive manufacturing process in which a liquid bonding agent is selectively deposited to join powder materials [148]. As shown in Figure 5-8, the jetted binder droplets interact with the powder particles to form primitives that stitch together to form a cross-sectional layer. Once a layer is deposited, a new layer of powder is recoated on top of the previous layer, which is then printed and stitched to the previous layer by the liquid binder. The layer-by-layer process is repeated to create the complete green part, which will be placed in a sintering furnace to vaporize the binder and bind the powder particles together to obtain the final product. The part may also be subjected to a secondary infiltration process with a lower melting alloy to minimize porosity. For example, copper parts are infiltrated with brass after sintering. In this work, the secondary infiltration process is not conducted to maintain a single-phase material.

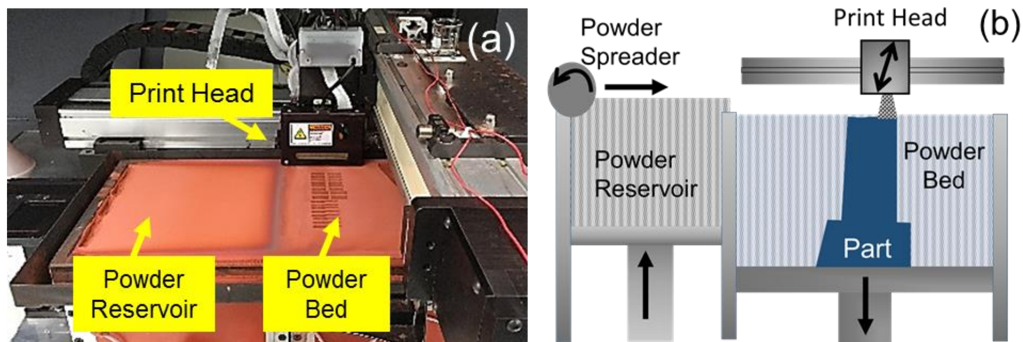

Figure 5-8 (a) The print head of the ExOne R2 printer; (b) the sketch of binder jetting.

The copper part with intricate features made using binder jetting was shown in Figure 5-1 (a). CT scanning is used to visualize the internal morphology on the stem region of the part as in Figure 5-1 (c). CT scans from layers 16 through 21 are shown in Figure 5-9. The denser material is rendered with bright color, while hollow features and pores are displayed with dark color.

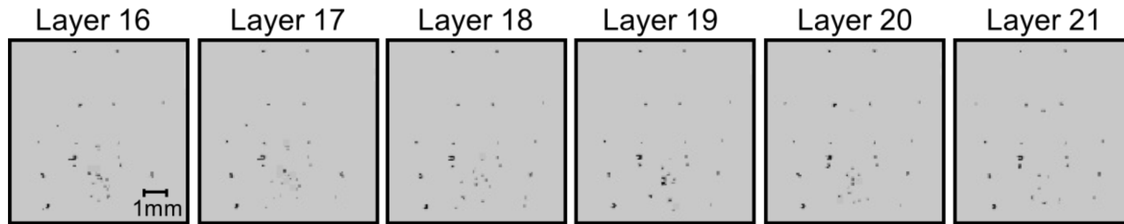

Figure 5-9 Sequential CT scan images on the stem region of the copper part made in binder jetting are selected for layer-wise porosity modeling and prediction.

The number of pores within the selected region on each layer is extracted from the CT scans and plotted in Figure 5-10. From Figure 5-10, a definite trend is evident in the range of layers (layer 10 – layer 40). The temporal correlation is estimated to be at least six layers. Hence, ignoring the temporal correlation by assuming that the pores between layers are independent is not physically tenable.

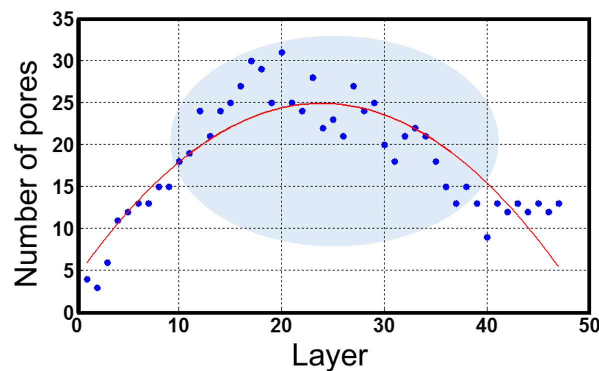

Figure 5-10 The number of pores within the selected region on each layer is extracted from the CT scans of the copper product. The range of layers (layer 10 - layer 40) are used for layer-wise porosity prediction.

In this case study, AST-LGCP is used to predict the high-risk areas prone to pores on the next layer based on the historical data of previous layers. The pores on the thirty images shaded in Figure 5-10 (layer 11 - layer 40) are predicted based on information from the previous six layers. In this demonstration, AST-LGCP is applied on the obtained augmented point patterns from the images (layer 16 - layer 21) in Figure 5-9 for spatiotemporal analysis of layer-wise pore evolution. The predicted high-risk areas on layer 22 by AST-LGCP, ST-LGCP and the empirical approach

are shown in Figure 5-11 for comparison. The empirical approach identifies the segmented porosity-prone areas (Figure 5-11(b)) by directly using all the pores from previous layers for prediction; ST-LGCP finds relatively large areas with low local intensity (Figure 5-11(c)) in that it reduces pores into dimensionless points; hence it lends to underestimate the severity of the pores; In contrast, AST-LGCP marks relatively large and continuous areas with high local intensity (Figure 5-11(d)) by accounting for the size of the pores and also the decrease in temporal correlation among previous images.

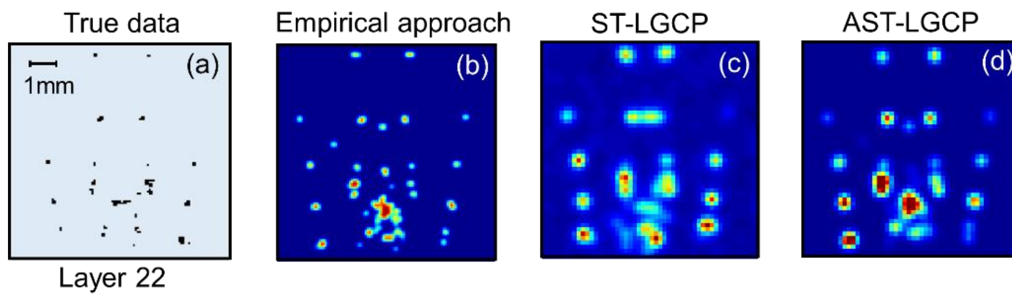

Figure 5-11 (a) The pores on the layer 22; (b) overly segmented high-risk areas identified by the empirical approach (benchmark method); (c) high-risk areas with low local intensity identified by ST-LGCP; and (d) high-risk areas with high local intensity identified by the proposed AST-LGCP.

The F-score results for predicting the porosity for total thirty layers (layer 11 - layer 40) are summarized in Table 5-4. AST-LGCP has ~ 5% higher F-score in predicting high-risk areas on the next layer than the two methods. ST-LGCP also has high precision and sensitivity. As discussed in Sec.5.3.1, the performance of ST-LGCP would improve with small pores. In contrast, the empirical approach has the inferior sensitivity because the identified high-risk is too segmented to achieve efficient prediction for the small pores. In a practical context, lower porosity leads to higher mechanical strength and conductivity of AM parts [149, 150]. With accurate layer-wise porosity prediction by the proposed AST-LGCP, corrective action could be initialized for improving product quality of AM parts.

Table 5-4 F-score results for predicting total thirty layers (layer 11 - layer 40) – Comparison of the empirical approach, ST-LGCP, and AST-LGCP (The values in the parenthesis are the standard deviation).

	Empirical approach	ST-LGCP	AST-LGCP
Precision	0.7698 (0.1441)	0.7227(0.1226)	**0.7871 (0.1051)**
Sensitivity	0.8451 (0.1190)	0.9428 (0.0582)	**0.9526 (0.0560)**
F-score	0.7978 (0.1085)	0.8121 (0.0896)	**0.8564 (0.0638)**

In addition, the effects of different number of previous layers in predicting the porosity on the next layer by AST-LGCP are investigated. The prediction accuracy reaches the peak average F-score value with six previous layers for prediction as shown in Figure 5-12.

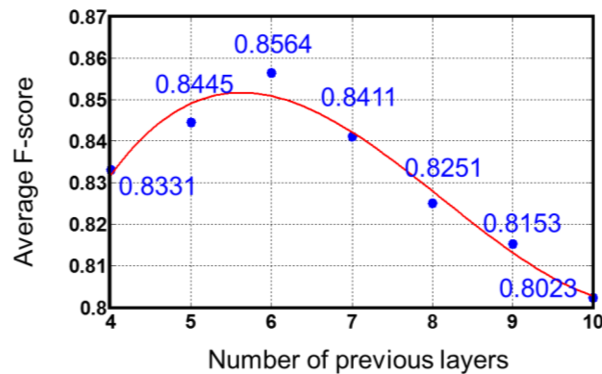

Figure 5-12 The average F-score results for predicting total thirty layers (layer 11 - layer 40) with different numbers of previous layers by AST-LGCP.

Furthermore, apart from the prediction for pore-prone areas, the number of pores on next layers is also predicted. While the empirical approach uses the moving average of numbers on previous six layers; ST-LGCP and AST-LGCP calculate the predicted number of pores on next layer through Eqn. (35) and (39), respectively. The predicted trend for layer 11- 40 by these three methods is juxtaposed in Figure 5-13.

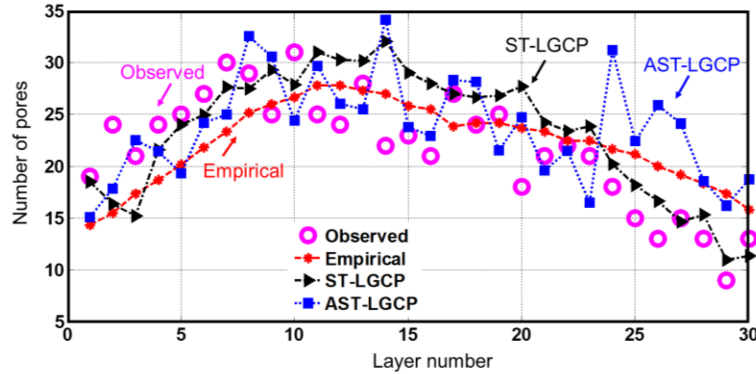

Figure 5-13 The layer-wise predicted number of pores on each scanned layer in the specified region on the stem of the copper product by the empirical approach, ST-LGCP, and AST-LGCP.

From these results it is evident the proposed AST-LGCP outperforms existing empirical and spatiotemporal modeling approach in the predicting the spatiotemporal trends in pore formation for AM parts.

5.4 Summary

This research proposed an augmented spatiotemporal log Gaussian Cox process (AST-LGCP) to understand and predict the spatiotemporal behavior of pore formation in additive manufacturing. Specifically, AST-LGCP first represents the pores observed from CT scan images of an AM product with augmented point patterns, including the information about the number, location, size and form of the pores. It further utilizes the spatial and temporal correlations among the pores on different layers, and predicts the areas susceptible to pores on next layers. The AST-LGCP approach is applied to parts made using binder jetting; it predicts the areas susceptible to porosity with statistical fidelity approaching ~85% (F-score).

This is a fundamentally new analytical direction to understand and quantify porosity in AM parts. The practical outcome is that porosity-prone areas of a part can be ascertained, and the appropriate design or process parameters can be modified in advance. Given its data-driven nature, the

proposed approach can be readily generalized to different powder-based AM processes. Two main focuses of the forthcoming research are:

- To verify the repeatability of AST-LGCP in porosity prediction by using parts from different AM processes, and improve its prediction for multiple subsequent layers ahead;
- To diagnose the root causes of the porosity in AM parts, and optimize part design or process parameters accordingly for quality improvement.

6 Conclusions and Future Work

In this dissertation, integrated methodologies for online quality assurance are developed and deployed for four complex advanced manufacturing processes, namely, fused filament fabrication (FFF), chemical mechanical planarization (CMP), wafer slicing, and binder jetting to improve product quality and enhance productivity. These integrated methodologies successfully utilize a great variety of measurement data (e.g., online signal streams, profile data, and images), and achieve effective online quality assurance for the individual advanced manufacturing processes. The common challenges in the above manufacturing processes, such as nonlinear process dynamics, multiple process attributes, and non-Gaussian sensor signals are addressed successfully by the developed methodologies, which are more adaptable and robust than traditional methods, and are promising to be generalized to other complex advanced manufacturing processes.

Moreover, this dissertation bridges the existing gaps in online quality assurance for advanced manufacturing, namely, (1) overcomes restrictive Gaussian or symmetry assumptions in statistical process control by approximating the non-Gaussian sensor signal with Dirichlet process mixture model; (2) discovers subtle spatial features in wafer thickness data for comprehensive modeling and strict quality assurance; and (3) formulates efficient quality quantification and prediction for porosity. Therefore, the research in the dissertation constitutes a step further along the lines of online quality assurance for advanced manufacturing.

Specifically, three innovative online quality assurance methodologies utilizing spatiotemporal analysis with heterogeneous sensor data are developed and implemented, and they achieve superior performance in timely fault detection and accurate quality assurance for these advanced manufacturing processes.

Real-time process monitoring based on Dirichlet process (DP) mixture model constitutes an effective online quality assurance methodology for advanced manufacturing with strong non-Gaussian sensor signals. It eliminates the restrictive assumption of normality or symmetry central to traditional statistical process control methods, and achieves timely process monitoring for FFF and CMP by three specific methods with different prerequisites. The DP evidence theoretic method applied in FFF attains accurate real-time process state classification and detection of process anomalies. This method identifies the FFF process failures (e.g., nozzle clog) with high accuracy and reliability (average F-score ~ 85%). DP-based SPC and RHDP clustering are employed in CMP for detecting process changes and clustering process states by adequately handling the non-Gaussian and high-noise vibration signals. DP-based SPC detects the onset of CMP process anomalies in only the half of the time used by traditional methods, such as the exponentially weighted moving average (EWMA) control chart. The RHDP clustering model identifies multiple, sequential process drifts with high accuracy (F-score ~ 80%).

Spatial Dirichlet process (SDP) modeling innovatively tackles the challenging non-Gaussian spatial profile data of wafer thickness by exploring the clustering phenomenon within wafers. With accurate modeling of in-control profiles, the proposed SDP-based statistical control scheme achieves effective out-of-control profile detection for quality assurance. This work emulates existing statistical control schemes for spatial profile data. When applied to the wafer thickness profile data from a slicing process, the SDP-based statistical control scheme detects out-of-control wafers with a higher power of test (average type II error of 0.039) than the benchmark method (average type II error of 0.122), and is therefore capable of preventing defective wafers being passed along to downstream production.

Layer-wise porosity modeling and prediction establish a new systematic methodology to investigate the spatiotemporal evolutions of pores in additive manufacturing (AM) from CT scans and predict pore-prone areas on consecutive layers, for online quality improvement. This research has a consequential impact on quality assurance in additive manufacturing, because porosity directly impacts the functional performance of AM parts and early corrective action is desired to improve product quality. The proposed augmented spatiotemporal log Gaussian Cox process (AST-LGCP) incorporates both point patterns (numbers and locations) and morphological features (sizes and forms) of pores into a spatiotemporal analysis, provides an insightful view of pores within a spatiotemporal structure, and accurately predicts the pore-prone areas on subsequent layers of the part with statistical fidelity (F-score ~ 85%). This work advances the porosity quantification and online quality improvements in AM.

In the future, two research areas are of great interest built on this dissertation:

1) Video-based online quality assurance

   Video has become a prevalent data format in various areas with the easy accessibility of cameras for primary data acquisition. It preserves complete information in space-time with ample details. Therefore, adoption of video in online quality assurance would enhance the accuracy and responsiveness for process monitoring and product quality improvement. Severe challenges, however, can be foreseen regarding data analytics: first, difficulties in tracking targets among a complex background; second, complication in coping with massive high-dimensional data in real-time; and third, challenges in balancing computational efficiency and analytical accuracy. Some components from this dissertation are potentially helpful to overcome these challenges: for instance, data representation with augmented point pattern for investigating event occurrence instead of using raw video data,

data compression by clustering images in the video before analysis, and spatiotemporal analysis for the video by exploring spatial and temporal correlations.

2) Voxel-based porosity quantification for AM parts

Pores essentially are 3D objects scattered in AM parts. Describing pores with voxels from CT scans and extending the proposed pixel-based porosity quantification to voxel-based porosity quantification are a natural and necessary step in future work.

Pores in voxels preserve even more morphological features of pores, such as shapes, volumes, etc., which provide comprehensive basis for analyzing the impact of part design and process conditions. Exploring spatial correlation among these pores in 3D space provides more accurate quantification of dependence among the pores than on 2D layer. Therefore, voxel-based porosity quantification could be used for process diagnosis with distribution characteristics and pore morphologies, and quality improvement with feedback to the part design and the process parameters.

References

[1] I. Gibson, D. W. Rosen, and B. Stucker, *Additive manufacturing technologies: rapid prototyping to direct digital manufacturing*: Springer, 2010.

[2] D. Bourell, J. Beaman, M. Leu, and D. Rosen, "A brief history of additive manufacturing and the 2009 roadmap for additive manufacturing: looking back and looking ahead," in *US-Turkey Workshop on Rapid Technologies*, 2009.

[3] J.-P. Kruth, M. Leu, and T. Nakagawa, "Progress in additive manufacturing and rapid prototyping," *CIRP Annals-Manufacturing Technology,* vol. 47, pp. 525-540, 1998.

[4] G. N. Levy, R. Schindel, and J.-P. Kruth, "Rapid manufacturing and rapid tooling with layer manufacturing (LM) technologies, state of the art and future perspectives," *CIRP Annals-Manufacturing Technology,* vol. 52, pp. 589-609, 2003.

[5] C. Cheng, A. Sa-Ngasoongsong, O. Beyca, T. Le, H. Yang, Z. Kong*, et al.*, "Time series forecasting for nonlinear and non-stationary processes: a review and comparative study," *IIE Transactions,* vol. 47, pp. 1053-1071, 2015.

[6] C. E. Rasmussen, "The infinite Gaussian mixture model," *Advances in Neural Information Processing Systems,* vol. 12, pp. 554-560, 2000.

[7] D. C. Montgomery, *Introduction to Statistical Quality Control*, 6 ed. New York, NY: John Wiley & Sons, 2008.

[8] S. Chakraborti, P. Van der Laan, and S. Bakir, "Nonparametric control charts: an overview and some results," *Journal of Quality Technology,* vol. 33, 2001.

[9] P. Qiu and Z. Li, "On nonparametric statistical process control of univariate processes," *Technometrics,* vol. 53, 2011.

[10] P. Qiu and Z. Li, "Distribution-free monitoring of univariate processes," *Statistics & Probability Letters,* vol. 81, pp. 1833-1840, 2011.

[11] R. Ganesan, T. K. Das, and V. Venkataraman, "Wavelet-based multiscale statistical process monitoring: A literature review," *IIE Transactions,* vol. 36, pp. 787-806, 2004.

[12] J.-C. Lu, S.-L. Jeng, and K. Wang, "A review of statistical methods for quality improvement and control in nanotechnology," *Journal of Quality Technology,* vol. 41, pp. 148 - 164, 2009.

[13] R. Ganesan, T. K. Das, A. K. Sikder, and A. Kumar, "Wavelet-based identification of delamination defect in CMP (Cu-low k) using nonstationary acoustic emission signal," *IEEE Transactions on Semiconductor Manufacturing,* vol. 16, pp. 677-685, 2003.

[14] H. Guo, K. Paynabar, and J. Jin, "Multiscale monitoring of autocorrelated processes using wavelets analysis," *IIE Transactions,* vol. 44, pp. 312-326, 2012.

[15] M. K. Jeong, J.-C. Lu, and N. Wang, "Wavelet-based SPC procedure for complicated functional data," *International Journal of Production Research,* vol. 44, pp. 729-744, 2006.

[16] G. A. Pugh, "A comparison of neural networks to SPC charts," *Computers & Industrial Engineering,* vol. 21, pp. 253-255, 1991.

[17] F. Zorriassatine and J. Tannock, "A review of neural networks for statistical process control," *Journal of Intelligent Manufacturing,* vol. 9, pp. 209-224, 1998.

[18] P. K. Rao, "Sensor-based monitoring and inspection of surface morphology variations in ultraprecision manufacturing processes," PhD Dissertation, Industrial Engineering and Management, Oklahoma State University, Stillwater, OK, 2013.

[19] P. Rao, S. Bukkapatnam, O. Beyca, Z. J. Kong, and R. Komanduri, "Real-Time Identification of Incipient Surface Morphology Variations in Ultraprecision Machining Process," *Journal of Manufacturing Science and Engineering,* vol. 136, pp. 021008-1 - 021008-11, 2014.

[20] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 603-619, 2002.

[21] S. W. Choi, J. H. Park, and I.-B. Lee, "Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis," *Computers & Chemical Engineering,* vol. 28, pp. 1377-1387, 2004.

[22] U. Thissen, H. Swierenga, A. De Weijer, R. Wehrens, W. Melssen, and L. Buydens, "Multivariate statistical process control using mixture modelling," *Journal of Chemometrics,* vol. 19, pp. 23-31, 2005.

[23] T. Chen, J. Morris, and E. Martin, "Probability density estimation via an infinite Gaussian mixture model: application to statistical process monitoring," *Journal of the Royal Statistical Society: Series C (Applied Statistics),* vol. 55, pp. 699-715, 2006.

[24] P. K. Rao, J. P. Liu, D. Roberson, and Z. J. Kong, "Sensor-Based Online Process Fault Detection in Additive Manufacturing," in *ASME 2015 International Manufacturing Science and Engineering Conference*, 2015, p. 13.

[25] P. K. Rao, J. P. Liu, D. Roberson, Z. J. Kong, and C. Williams, "Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors," *Journal of Manufacturing Science and Engineering,* vol. 137, p. 061007, 2015.

[26] K. Bastani, P. K. Rao, and Z. Kong, "An online sparse estimation-based classification approach for real-time monitoring in advanced manufacturing processes from heterogeneous sensor data," *IIE Transactions,* vol. 48, pp. 579-598, 2016.

[27] S. Huang, Z. Kong, and W. Huang, "High-dimensional process monitoring and change point detection using embedding distributions in reproducing kernel Hilbert space," *IIE Transactions,* vol. 46, pp. 999-1016, 2014.

[28] N. Cressie and C. K. Wikle, *Statistics for spatio-temporal data*. Hoboken, NJ: John Wiley & Sons, 2011.

[29] P. Diggle and P. J. Ribeiro, *Model-based geostatistics*. New York, NY: Springer, 2007.

[30] C. E. Rasmussen and C. K. I. Williams, *Gaussian processes for machine learning*. Cambridge, MA: The MIT Press, 2005.

[31] J. Quiñonero-Candela and C. E. Rasmussen, "A unifying view of sparse approximate Gaussian process regression," *The Journal of Machine Learning Research,* vol. 6, pp. 1939-1959, 2005.

[32] R. B. Gramacy and H. K. Lee, "Bayesian treed Gaussian process models with an application to computer modeling," *Journal of the American Statistical Association,* vol. 103, pp. 1119-1130, 2008.

[33] E. Snelson and Z. Ghahramani, "Local and global sparse Gaussian process approximations," in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 524-531.

[34] C. Walder, K. I. Kim, and B. Schölkopf, "Sparse multiscale Gaussian process regression," in *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1112-1119.

[35] C. Park, J. Z. Huang, and Y. Ding, "Domain decomposition approach for fast Gaussian process regression of large spatial data sets," *Journal of Machine Learning Research,* vol. 12, pp. 1697-1728, 2011.

[36] R. Jin, C.-J. Chang, and J. Shi, "Sequential measurement strategy for wafer geometric profile estimation," *IIE Transactions,* vol. 44, pp. 1-12, 2012.

[37] L. Zhang, K. Wang, and N. Chen, "Monitoring wafer's geometric quality using an additive Gaussian process model," *IIE Transactions,* vol. 48, pp. 1-15, 2016.

[38] H. Zhao, R. Jin, S. Wu, and J. Shi, "PDE-constrained Gaussian process model on material removal rate of wire saw slicing process," *Journal of Manufacturing Science and Engineering,* vol. 133, p. 021012, 2011.

[39] L. Bao, K. Wang, and R. Jin, "A hierarchical model for characterising spatial wafer variations," *International Journal of Production Research,* vol. 52, pp. 1827-1842, 2014.

[40] E. Padonou and O. Roustant, "Polar Gaussian Processes for Predicting on Circular Domains," *HAL, Hal-01119942,* 2015.

[41] G. Pistone and G. Vicario, "Kriging prediction from a circular grid: application to wafer diffusion," *Applied Stochastic Models in Business and Industry,* vol. 29, pp. 350-361, 2013.

[42] M. Plumlee, R. Jin, V. Roshan Joseph, and J. Shi, "Gaussian process modeling for engineered surfaces with applications to Si wafer production," *Stat,* vol. 2, pp. 159-170, 2013.

[43] J. P. Liu, O. F. Beyca, P. K. Rao, Z. J. Kong, and S. T. Bukkapatnam, "Dirichlet Process Gaussian Mixture Models for Real-Time Monitoring and Their Application to Chemical Mechanical Planarization," *IEEE Transactions on Automation Science and Engineering,* vol. 14, pp. 208-221, 2017.

[44] W. J. Sames, F. List, S. Pannala, R. R. Dehoff, and S. S. Babu, "The metallurgy and processing science of metal additive manufacturing," *International Materials Reviews,* vol. 61, pp. 315-360, 2016.

[45] J. A. Slotwinski, E. J. Garboczi, and K. M. Hebenstreit, "Porosity Measurements and Analysis for Metal Additive Manufacturing Process Control," *Journal of Research of the National Institute of Standards and Technology,* vol. 119, pp. 494-528, 2014.

[46] S. Leuders, M. Thöne, A. Riemer, T. Niendorf, T. Tröster, H. Richard*, et al.*, "On the mechanical behaviour of titanium alloy TiAl6V4 manufactured by selective laser melting: Fatigue resistance and crack growth performance," *International Journal of Fatigue,* vol. 48, pp. 300-307, 2013.

[47] H. Gong, K. Rafi, T. Starr, and B. Stucker, "The Effects of Processing Parameters on Defect Regularity in Ti-6Al-4V Parts Fabricated By Selective Laser Melting and Electron Beam Melting," presented at the 24th Annual International Solid Freeform Fabrication Symposium—An Additive Manufacturing Conference, Austin, TX, 2013.

[48] A. Spierings and G. Levy, "Comparison of density of stainless steel 316L parts produced with selective laser melting using different powder grades," in *Proceedings of the Annual International Solid Freeform Fabrication Symposium*, Austin, TX, 2009, pp. 342-353.

[49] I. Yadroitsev and I. Smurov, "Selective laser melting technology: from the single laser melted track stability to 3D parts of complex shape," *Physics Procedia,* vol. 5, pp. 551-560, 2010.

[50] C. B. Williams, J. K. Cochran, and D. W. Rosen, "Additive manufacturing of metallic cellular materials via three-dimensional printing," *The International Journal of Advanced Manufacturing Technology,* vol. 53, pp. 231-239, 2011.

[51] Y. Bai and C. B. Williams, "An exploration of binder jetting of copper," *Rapid Prototyping Journal,* vol. 21, pp. 177-185, 2015.

[52] S. Gaytan, M. Cadena, H. Karim, D. Delfin, Y. Lin, D. Espalin*, et al.*, "Fabrication of barium titanate by binder jetting additive manufacturing technology," *Ceramics International,* vol. 41, pp. 6610-6619, 2015.

[53] A. Spierings, M. Schneider, and R. Eggenberger, "Comparison of density measurement techniques for additive manufactured metallic parts," *Rapid Prototyping Journal,* vol. 17, pp. 380-386, 2011.

[54] D. Bourell, B. Stucker, A. Spierings, N. Herres, and G. Levy, "Influence of the particle size distribution on surface quality and mechanical properties in AM steel parts," *Rapid Prototyping Journal,* vol. 17, pp. 195-202, 2011.

[55] N. Karthik, H. Gu, D. Pal, T. Starr, and B. Stucker, "High Frequency Ultrasonic Non Destructive Evaluation of Additively Manufactured Components," in *Proceeding of the Twenty Forth Annual International Solid Freeform Fabrication Symposium*, 2013, pp. 311-325.

[56] K. Monroy, J. Delgado, and J. Ciurana, "Study of the pore formation on CoCrMo alloys by selective laser melting manufacturing process," *Procedia Engineering,* vol. 63, pp. 361-369, 2013.

[57] H. Gu, H. Gong, D. Pal, K. Rafi, T. Starr, and B. Stucker, "Influences of energy density on porosity and microstructure of selective laser melted 17-4PH stainless steel," in *2013 Solid Freeform Fabrication Symposium*, 2013, p. 474.

[58] S. Siddique, M. Imran, M. Rauer, M. Kaloudis, E. Wycisk, C. Emmelmann*, et al.*, "Computed tomography for characterization of fatigue performance of selective laser melted parts," *Materials & Design,* vol. 83, pp. 661-669, 2015.

[59] F. Léonard, S. Tammas-Wiliams, P. B. Prangnell, I. Todd, and P. J. Withers, "Assessment by X-ray CT of the effects of geometry and build direction on defects in titanium ALM parts," in *Conference on Industrial Computed Tomography (ICT)*, 2012, pp. 85-93.

[60] E. Girardin, C. Renghini, J. Dyson, V. Calbucci, F. Moroncini, and G. Albertini, "Characterization of porosity in a laser sintered MMCp using X-ray synchrotron phase contrast microtomography," *Materials Sciences and Applications,* vol. 2, p. 1322, 2011.

[61] S. Tammas-Williams, H. Zhao, F. Léonard, F. Derguti, I. Todd, and P. Prangnell, "XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by Selective Electron Beam Melting," *Materials Characterization,* vol. 102, pp. 47-61, 2015.

[62] B. Sharrat, "Non-Destructive Techniques and Technologies for Qualification of Additive Manufactured Parts and Processes: A Literature Review," Defence Research and Development Canada, 2015.

[63] A. Thompson, I. Maskery, and R. K. Leach, "X-ray computed tomography for additive manufacturing: a review," *Measurement Science and Technology,* vol. 27, p. 072001, 2016.

[64] H. Taud, R. Martinez-Angeles, J. Parrot, and L. Hernandez-Escobedo, "Porosity estimation method by X-ray computed tomography," *Journal of petroleum science and engineering,* vol. 47, pp. 209-217, 2005.

[65] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical modeling and analysis for spatial data*. Boca Raton: CRC Press, 2014.

[66] D. Stoyan and A. Penttinen, "Recent applications of point process methods in forestry statistics," *Statistical Science,* pp. 61-78, 2000.

[67] Y. Ogata, "Space-time point-process models for earthquake occurrences," *Annals of the Institute of Statistical Mathematics,* vol. 50, pp. 379-402, 1998.

[68] Q. Zhou, J. Zhou, M. De Cicco, S. Zhou, and X. Li, "Detecting 3D spatial clustering of particles in nanocomposites based on cross-sectional images," *Technometrics,* vol. 56, pp. 212-224, 2014.

[69] L. Dong, X. Li, Y. Qian, D. Yu, H. Zhang, Z. Zhang*, et al.*, "Quantifying Nanoparticle Mixing State to Account for Both Location and Size Effects," *Technometrics,* vol. 0, pp. 1-13, 2017.

[70] A. Brix and P. J. Diggle, "Spatiotemporal prediction for log-Gaussian Cox processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 63, pp. 823-841, 2001.

[71] J. Møller, A. R. Syversveen, and R. P. Waagepetersen, "Log Gaussian Cox processes," *Scandinavian Journal of Statistics,* vol. 25, pp. 451-482, 1998.

[72] M. Grasso and B. M. Colosimo, "Process defects and in situ monitoring methods in metal powder bed fusion: a review," *Measurement Science and Technology,* vol. 28, p. 044005, 2017.

[73] B. Khoshnevis, D. Hwang, K.-T. Yao, and Z. Yeh, "Mega-scale fabrication by contour crafting," *International Journal of Industrial and Systems Engineering,* vol. 1, pp. 301-320, 2006.

[74] B. Khoshnevis, M. P. Bodiford, K. H. Burks, E. Ethridge, D. Tucker, W. Kim*, et al.*, "Lunar contour crafting–A novel technique for ISRU based habitat development," in *American Institute of Aeronautics and Astronautics Conference, Reno, January*, 2005.

[75] K. P. Cooper, "Layered manufacturing: challenges and opportunities," in *Materials Research Society Symposia*, 2003, pp. 23-34.

[76] J. Scott, N. Gupta, C. L. Weber, S. Newsome, T. Wohlers, and T. Caffrey, "Additive Manufacturing: Status and Opportunities," *Science and Technology Policy Institute,* pp. 1-29, 2012.

[77] (2013) Additive Manufacturing: Opportunities and Constraints. *Royal Academy of Engineering*.

[78] P. K. Rao, Z. Kong, C. E. Duty, R. J. Smith, V. Kunc, and L. J. Love, "Assessment of dimensional integrity and spatial defect localization in additive manufacturing using spectral graph theory," *Journal of Manufacturing Science and Engineering,* vol. 138, p. 051007, 2016.

[79] M. S. Tootooni, A. Dsouza, R. Donovan, P. K. Rao, Z. J. Kong, and P. Borgesen, "Classifying the Dimensional Variation in Additive Manufactured Parts from Laser-Scanned 3D Point Cloud Data using Machine Learning Approaches," *Journal of Manufacturing Science and Engineering,* 2017.

[80] N. Guo and M. C. Leu, "Additive manufacturing: technology, applications and research needs," *Frontiers of Mechanical Engineering,* vol. 8, pp. 215-243, 2013.

[81] J. M. Steigerwald, S. P. Murarka, and R. J. Gutmann, *Chemical mechanical planarization of microelectronic materials*. Weinheim, Germany: Wiley-VCH, 2008.

[82] P. K. Rao, M. B. Bhushan, S. T. Bukkapatnam, Z. Kong, S. Byalal, O. F. Beyca*, et al.*, "Process-Machine Interaction (PMI) Modeling and Monitoring of Chemical Mechanical Planarization (CMP) Process Using Wireless Vibration Sensors," *IEEE Transactions on Semiconductor Manufacturing,* vol. 27, pp. 1-15, 2014.

[83] Z. Kong, A. Oztekin, O. F. Beyca, U. Phatak, S. T. S. Bukkapatnam, and R. Komanduri, "Process performance prediction for chemical mechanical planarization (CMP) by integration of nonlinear bayesian analysis and statistical modeling," *IEEE Transactions on Semiconductor Manufacturing,* vol. 23, pp. 316 - 327, 2010.

[84] H. Jeong, H. Kim, S. Lee, and D. Dornfeld, "Multi-sensor monitoring system in chemical mechanical planarization (CMP) for correlations with process issues," *CIRP Annals-Manufacturing Technology,* vol. 55, pp. 325-328, 2006.

[85] A. Sikder, F. Giglio, J. Wood, A. Kumar, and M. Anthony, "Optimization of tribological properties of silicon dioxide during the chemical mechanical planarization process," *Journal of Electronic Materials,* vol. 30, pp. 1520-1526, 2001.

[86] J. Tang, D. Dornfeld, S. K. Pangrle, and A. Dangca, "In-process detection of microscratching during CMP using acoustic emission sensing technology," *Journal of Electronic Materials,* vol. 27, pp. 1099-1103, 1998.

[87] Z. Kong, O. Beyca, S. Bukkapatnam, and R. Komanduri, "Nonlinear Sequential Bayesian Analysis-Based Decision Making for End-Point Detection of Chemical Mechanical Planarization (CMP) Processes," *IEEE Transactions on Semiconductor Manufacturing,* vol. 24, pp. 523-532, 2011.

[88] U. Phatak, S. Bukkapatnam, Z. Kong, and R. Komanduri, "Sensor-based modeling of slurry chemistry effects on the material removal rate (MRR) in copper-CMP process," *International Journal of Machine Tools and Manufacture,* vol. 49, pp. 171-181, 2009.

[89] Z. Wang, S. T. Bukkapatnam, S. R. Kumara, Z. Kong, and Z. Katz, "Change detection in precision manufacturing processes under transient conditions," *CIRP Annals-Manufacturing Technology,* vol. 63, pp. 449-452, 2014.

[90] M. S. Tootooni, P. K. Rao, C.-A. Chou, and Z. Kong, "A Spectral Graph Theoretic Approach for Monitoring Multivariate Time Series Data From Complex Dynamical Processes," *IEEE Transactions on Automation Science and Engineering,* 2016.

[91] P. K. Rao, O. F. Beyca, Z. Kong, S. T. Bukkapatnam, K. E. Case, and R. Komanduri, "A graph-theoretic approach for quantification of surface morphology variation and its application to chemical mechanical planarization process," *IIE Transactions,* vol. 47, pp. 1088-1111, 2015.

[92] M. D. Escobar and M. West, "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association,* vol. 90, 1995.

[93] T. K. Moon, "The expectation-maximization algorithm," *Signal processing magazine, IEEE,* vol. 13, pp. 47-60, 1996.

[94] M. Medvedovic and S. Sivaganesan, "Bayesian infinite mixture model based clustering of gene expression profiles," *Bioinformatics,* vol. 18, pp. 1194-1206, 2002.

[95] A. R. Ferreira da Silva, "A Dirichlet process mixture model for brain MRI tissue classification," *Medical Image Analysis,* vol. 11, pp. 169-182, 2007.

[96] J. Zhang, Z. Ghahramani, and Y. Yang, "A probabilistic model for online document clustering with application to novelty detection," in *Advances in Neural Information Processing Systems*, 2004, pp. 1617-1624.

[97] A. Torralba, A. S. Willsky, E. B. Sudderth, and W. T. Freeman, "Describing visual scenes using transformed dirichlet processes," in *Advances in neural information processing systems*, 2005, pp. 1297-1304.

[98] D. Blackwell and J. B. MacQueen, "Ferguson distributions via Pólya urn schemes," *The Annals of Statistics,* vol. 1, pp. 353-355, 1973.

[99] S. N. MacEachern and P. Müller, "Estimating mixture of Dirichlet process models," *Journal of Computational and Graphical Statistics,* vol. 7, pp. 223-238, 1998.

[100] S. N. MacEachern, "Estimating normal means with a conjugate style Dirichlet process prior," *Communications in Statistics-Simulation and Computation,* vol. 23, pp. 727-741, 1994.

[101] M. D. Escobar and M. West, "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association,* vol. 90, pp. 577-588, 1995.

[102] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of computational and graphical statistics,* vol. 9, pp. 249-265, 2000.

[103] H. Ishwaran and L. F. James, "Gibbs sampling methods for stick-breaking priors," *Journal of the American Statistical Association,* vol. 96, 2001.

[104] A. Ranganathan, "The Dirichlet process mixture (DPM) model," Citeseer, 2006.

[105] D. Görür and C. E. Rasmussen, "Dirichlet process gaussian mixture models: Choice of the base distribution," *Journal of Computer Science and Technology,* vol. 25, pp. 653-664, 2010.

[106] M. D. Escobar, "Estimating normal means with a Dirichlet process prior," *Journal of the American Statistical Association,* vol. 89, pp. 268-277, 1994.

[107] G. Shafer, *A mathematical theory of evidence* vol. 1: Princeton university press Princeton, 1976.

[108] T. S. Ferguson, "Bayesian density estimation by mixtures of normal distributions," *Recent advances in statistics,* vol. 24, pp. 287-302, 1983.

[109] O. F. Beyca, "Sensor-based Real-time Process Monitoring for Ultra-Precision Manufacturing Processes with nonlinearity and nonstationarity," PhD Dissertation, Industrial Engineering and Management, Oklahoma State University, 2013.

[110] O. Beyca, P. Rao, Z. J. Kong, and S. Bukkapatnam, "Heterogeneous Sensor Data Fusion Approach for Real-time Monitoring in Ultraprecision Machining (UPM) Process using Non-parametric Bayesian Clustering and Evidence Theory," *IEEE Transactions on Automation Science and Engineering,* vol. 13, pp. 1033-1044, 2016.

[111] O. Basir and X. Yuan, "Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory," *Information Fusion,* vol. 8, pp. 379-386, 2007.

[112] B.-S. Yang and K. J. Kim, "Application of Dempster–Shafer theory in fault diagnosis of induction motors using vibration and current signals," *Mechanical Systems and Signal Processing,* vol. 20, pp. 403-420, 2006.

[113] C. R. Parikh, M. J. Pont, and N. Barrie Jones, "Application of Dempster–Shafer theory in condition monitoring applications: a case study," *Pattern Recognition Letters,* vol. 22, pp. 777-785, 2001.

[114] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics,* vol. 5, pp. 115-133, 1943.

[115] P. Qiu, *Introduction to statistical process control*. Boca Raton, FL: CRC Press, 2013.

[116] A. Ahmed and E. P. Xing, "Dynamic non-parametric mixture models and the recurrent chinese restaurant process," in *Proceedings of the 2008 SIAM International Conference on Data Mining*, Atlanta, GA, 2008, pp. 219-230.

[117] Y. W. Teh, M. I. Jordan, M. J. Beal, and D. M. Blei, "Hierarchical dirichlet processes," *Journal of the American statistical Association,* vol. 101, pp. 1566-1581, 2006.

[118] A. Rodriguez, D. B. Dunson, and A. E. Gelfand, "The nested Dirichlet process," *Journal of the American statistical Association,* vol. 103, pp. 1131-1154, 2008.

[119] J. Crutchfield, D. Farmer, N. Packard, R. Shaw, G. Jones, and R. Donnelly, "Power spectral analysis of a dynamical system," *Physics Letters A,* vol. 76, pp. 1-4, 1980.

[120] S. Bukkapatnam, P. Rao, and R. Komanduri, "Experimental dynamics characterization and monitoring of MRR in oxide chemical mechanical planarization (CMP) process," *International Journal of Machine Tools and Manufacture,* vol. 48, pp. 1375-1386, 2008.

[121] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, 2 ed. vol. 7. Cambridge, UK, New York: Cambridge University Press, 2004.

[122] C. A. Lowry and D. C. Montgomery, "A review of multivariate control charts," *IIE Transactions,* vol. 27, pp. 800-810, 1995.

[123] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies,* vol. 2, pp. 37-63, 2011.

[124] G. H. John and P. Langley, "Estimating continuous distributions in Bayesian classifiers," in *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*, 1995, pp. 338-345.

[125] P. A. Lachenbruch, *Discriminant analysis*: Wiley Online Library, 1975.

[126] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning,* vol. 20, pp. 273-297, 1995.

[127] D. F. Specht, "Probabilistic neural networks," *Neural networks,* vol. 3, pp. 109-118, 1990.

[128] M. T. Hagan, H. B. Demuth, and M. Beale, *Neural Network Design*. Boston, MA.: PWS Publishing, 1997.

[129] Y. Nishi and R. Doering, *Handbook of semiconductor manufacturing technology*. Boca Raton, FL: CRC Press, 2000.

[130] C.-W. Chang, C.-R. Wu, C.-T. Lin, and H.-C. Chen, "Evaluating and controlling silicon wafer slicing quality using fuzzy analytical hierarchy and sensitivity analysis," *The International Journal of Advanced Manufacturing Technology,* vol. 36, pp. 322-333, 2008.

[131] SEMI, "SEMI MF1530-0707: Test Method for Measuring Flatness, Thickness, and Total Thickness Variation on Silicon Wafers by Automated Noncontact Scanning," in *SEMI International Standards*, ed, 2012.

[132] SEMI, "SEMI M49-1016: Guide for Specifying Geometry Measurement Systems for Silicon Wafers for the 130 nm to 16 nm Technology Generations," in *SEMI International Standards*, ed, 2016.

[133] J. Schmit, K. Creath, and J. Wyant, "Surface profilers, multiple wavelength, and white light interferometry," in *Optical Shop Testing*, ed Hoboken, NJ: John Wiley & Sons, 2007, pp. 667-755.

[134] B. Sopori, C. Auriemma, C. Li, and J. Madjdpour, "A Rapid, Non-Contact Method for Measurement of Si Wafer Thickness: Principles and Preliminary Results," in *13th Workshop on Crystalline Silicon Solar Cell Materials and Processes*, 2003, p. 225.

[135] W. Huang, Z. Kong, D. Ceglarek, and E. Brahmst, "The analysis of feature-based measurement error in coordinate metrology," *IIE Transactions,* vol. 36, pp. 237-251, 2004.

[136] R. Navarro and J. Arines Piferrer, "Complete modal representation with discrete zernike polynomials-critical sampling in non redundant grids," in *Numerical Simulation of Physical and Engineering Processes*, ed: InTech, 2011, pp. 221-238.

[137] P. Müller and R. Mitra, "Bayesian nonparametric inference-why and how," *Bayesian Analysis,* vol. 8, 2013.

[138] A. E. Gelfand, A. Kottas, and S. N. MacEachern, "Bayesian nonparametric spatial modeling with Dirichlet process mixing," *Journal of the American Statistical Association,* vol. 100, pp. 1021-1035, 2005.

[139] M. D. McKay, R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics,* vol. 42, pp. 55-61, 2000.

[140] W. E. Frazier, "Metal additive manufacturing: A review," *Journal of Materials Engineering and Performance,* vol. 23, pp. 1917-1928, 2014.

[141] D. Bourell, J. P. Kruth, M. Leu, G. Levy, D. Rosen, A. M. Beese*, et al.*, "Materials for additive manufacturing," *CIRP Annals-Manufacturing Technology,* 2017.

[142] N. T. Aboulkhair, N. M. Everitt, I. Ashcroft, and C. Tuck, "Reducing porosity in AlSi10Mg parts processed by selective laser melting," *Additive Manufacturing,* vol. 1, pp. 77-86, 2014.

[143] B. M. Taylor, T. M. Davies, B. S. Rowlingson, and P. J. Diggle, "lgcp: An R Package for Inference with Spatio-Temporal Log-Gaussian Cox Processes," *arXiv preprint arXiv:1110.6054,* 2011.

[144] B. Taylor, T. Davies, B. Rowlingson, and P. Diggle, "Bayesian inference and data augmentation schemes for spatial, spatiotemporal and multivariate Log-Gaussian Cox processes in R," *Journal of Statistical Software,* vol. 63, pp. 1-48, 2015.

[145] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *Journal of Chemical Physics,* vol. 21, pp. 1087-1092, 1953.

[146] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika,* vol. 57, pp. 97-109, 1970.

[147] G. O. Roberts and R. L. Tweedie, "Exponential convergence of Langevin distributions and their discrete approximations," *Bernoulli,* pp. 341-363, 1996.

[148] ISO/ASTM, "ISO/ASTM 52900:2015 Additive manufacturing-General principles-Terminology," ed, 2015, pp. 1-19.

[149] M. P. Meeder, "Modeling the Thermal and Electrical Properties of Different Density Sintered Binder Jetted Copper for Verification and Revision of The Wiedemann-Franz Law," Virginia Tech, 2016.

[150] J. Koh and A. Fortini, "Prediction of thermal conductivity and electrical resistivity of porous metallic materials," *International Journal of Heat and Mass Transfer,* vol. 16, pp. 2013-2022, 1973.

Appendix

The MCMC procedures for the proposed SDP modeling include the updating of the spatial random effects, the parameters of base distribution – Gaussian process (variance parameter σ^2 and decay parameter ϕ), the precision parameter v, and the variance parameter τ^2 in the pure error by iterating the following four steps [65, 138]. The priors for these parameters are defined in (29).

Step 1: Updating the spatial random effects requires the full conditional of spatial Dirichlet process (SDP) as in (26),

$$\boldsymbol{\theta}_{(p)}|\boldsymbol{\Theta}_{(-p)}, \boldsymbol{Y}_p, GP_0, v \sim \frac{v \cdot g(\boldsymbol{\theta}|\boldsymbol{Y}_p)}{p-1+v}\int F(\boldsymbol{Y}_p|\boldsymbol{\theta})GP_0(\boldsymbol{\theta})\, d\boldsymbol{\theta} + \sum_{l=1}^{L} F(\boldsymbol{Y}_p|\boldsymbol{\theta}_l)\sum_{j\neq p}^{P}\frac{\delta_{\boldsymbol{\theta}_l}(\boldsymbol{\theta}_{(j)})}{P-1+v},$$

where $\boldsymbol{\theta}_{(p)}$ is the respective spatial random effect for observation $\boldsymbol{Y}_p, p = 1, \dots, P$; $\boldsymbol{\Theta}_{(-p)}$ includes the spatial random effects for P observations except for $\boldsymbol{Y}_p$; the unique spatial random effects for all the observations are included in the set $\boldsymbol{\Theta}_L = \{\boldsymbol{\theta}_l\}, l = 1, \dots L$, and $\boldsymbol{\theta}_{(p)} \in \boldsymbol{\Theta}_L$, $p = 1, \dots, P, L \ll P$; the observation $\boldsymbol{Y}_p \sim F(\cdot\,|\boldsymbol{\theta}_{(p)}),\ p = 1, \dots P$, $g(\boldsymbol{\theta}|\boldsymbol{Y}_p) = \frac{F(\boldsymbol{Y}_p|\boldsymbol{\theta})GP_0(\boldsymbol{\theta})}{\int F(\boldsymbol{Y}_p|\boldsymbol{\theta})GP_0(\boldsymbol{\theta})d\boldsymbol{\theta}}$.

Step 2: Updating the parameters of the base distribution – Gaussian process (variance parameter σ^2 and decay parameter ϕ) by the following full conditionals,

$$\sigma^2|\boldsymbol{\Theta}_L, L, \phi \sim InvGamma(\tilde{\alpha}_\sigma, \tilde{\beta}_\sigma),$$

where $\tilde{\alpha}_\sigma = \alpha_\sigma + 0.5 \cdot n \cdot L$, $\tilde{\beta}_\sigma = \beta_\sigma + 0.5\sum_{l=1}^{L}\boldsymbol{\theta}_l{}^T\boldsymbol{H}_n^{-1}(\phi)\boldsymbol{\theta}_l$, n is the number of the measured locations for spatial data.

$$\phi|\boldsymbol{\Theta}_L, L, \sigma^2 \sim [\phi]\left(\det(\boldsymbol{H}_n(\phi))\right)^{-\frac{L}{2}}\exp\left(-\sum_{l=1}^{L}\frac{\boldsymbol{\theta}_l{}^T\boldsymbol{H}_n^{-1}(\phi)\boldsymbol{\theta}_l}{2\sigma^2}\right).$$

Step 3: Updating the precision parameter v by the following full conditional,

$$v|\boldsymbol{\Theta}, \mathbb{Y}, \eta \sim \lambda \cdot Gamma(\alpha_v + L, \beta_v - \log(\eta)) + (1-\lambda)\cdot Gamma(\alpha_v + L - 1, \beta_v - \log(\eta)),$$

where $\lambda = \frac{\alpha_v + L - 1}{\alpha_v + L - 1 + P(\beta_v - \log(\eta))}$, $\eta | v, \mathbb{Y} \sim Beta(v+1, P)$, and observation set $\mathbb{Y} = \{\boldsymbol{Y}_p\}$, $p = 1, \ldots P$.

Step 4: Updating the variance parameter τ^2 in pure error by the following full conditional,

$$\tau^2 | \boldsymbol{\Theta}, \mathbb{Y} \sim InvGamma(\tilde{\alpha}_\tau, \tilde{\beta}_\tau),$$

where $\tilde{\alpha}_\tau = \alpha_\tau + 0.5 \cdot n \cdot P$, $\tilde{\beta}_\tau = \beta_\tau + 0.5 \sum_{p=1}^{P} (\boldsymbol{Y}_p - \boldsymbol{\theta}_{(p)})^T (\boldsymbol{Y}_p - \boldsymbol{\theta}_{(p)})$.