IISE Transactions

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uiie21

In-process quality improvement: Concepts, methodologies, and applications

Jianjun Shi

To cite this article: Jianjun Shi (2022): In-process quality improvement: Concepts, methodologies, and applications, IISE Transactions, DOI: 10.1080/24725854.2022.2059725

To link to this article: https://doi.org/10.1080/24725854.2022.2059725

Published online: 23 May 2022.

IISE TRANSACTIONS
https://doi.org/10.1080/24725854.2022.2059725

In-process quality improvement: Concepts, methodologies, and applications

Jianjun Shi

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT
This article presents the concepts, methodologies, and applications of In-Process Quality Improvement (IPQI) in complex manufacturing systems. As opposed to traditional quality control concepts that emphasize process change detection, acceptance sampling, and offline designed experiments, IPQI focuses on integrating data science and system theory, taking full advantage of in-process sensing data to achieve process monitoring, diagnosis, and control. The implementation of IPQI leads to root cause diagnosis (in addition to change detection), automatic compensation (in addition to off-line adjustment), and defect prevention (in addition to defect inspection). The methodologies of IPQI have been developed and implemented in various manufacturing processes. This paper provides a brief historical review of the IPQI, summarizes the developments and applications of IPQI methodologies, and discusses some challenges and opportunities in the current data-rich manufacturing systems. Future research directions are discussed at the end of the article with a special focus on leveraging emerging machine learning tools to address quality improvements in data-rich advanced manufacturing systems.

ARTICLE HISTORY
Received 8 January 2022
Accepted 19 March 2022

KEYWORDS
Data fusion; engineering-driven statistics; in-process quality improvement; Industry 4.0; smart manufacturing

1. Introduction

About 25 years ago, while in the author's early career, he saw a significant opportunity to combine his training in systems and control theories with the quality improvement demands in the emerging, data-rich manufacturing systems. He coined the term, "In-Process Quality Improvement", or IPQI, to describe this new branch of quality science research and dedicated his career to developing it into a full-blown research field. This article aims to provide a historical review of IPQI: discussing the concepts, methodologies, and applications of IPQI; and exploring IPQI opportunities for future research and applications. Due to the rich literature in the past two and a half decades, this article cannot cover all the achievements related to IPQI methodologies and applications. Instead, this article focuses on selected IPQI topics and uses them to illustrate the development and implementation of IPQI concepts in selected applications with which the author is more familiar.

- *How does the concept of IPQI fit into the general framework of quality engineering?*

Before the introduction of the IPQI, the quality improvement methodologies had four major components: design of experiments (DOE) for response surface modeling and robust parameter design, statistical process control (SPC), acceptance sampling, and quality management (Figure 1).

Robust parameter design via DOE response surface modeling emphasizes how to design and optimize products or processes that will be robust to disturbances in a manufacturing system; *SPC* mainly focuses on monitoring and change detections for manufacturing processes; *acceptance sampling* makes the lot sentencing decisions; and *quality management* emphasizes policies and procedures of quality control at the organization level. Although those four methodologies played important roles in quality control and improvement, there were inherent limitations. For example, the robust parameter design assumes that all input variables and the distributions of anticipated disturbances are known in advance to achieve a robust design of the product/process. However, these assumptions may not always be satisfied in practice, and some unanticipated disturbances (e.g., random machine degradations/failures with unknown root causes and associated severities) may occur during the real-time production. Hence, off-line DOE methods cannot adequately address all unanticipated disturbances occurring during the production period. SPC can detect process changes, but relies on laborious manual efforts to find root causes of those changes, which can be time-consuming and heavily relies on operators' experiences. Acceptance sampling evaluates the lot quality, but does not improve its quality. Quality management defines the procedures and policies on quality in an organization, but quality management is not designed to address the root causes of failures in machines or production systems. Moreover, one fundamental component is missing in the existing quality control frameworks: how to make use of multiple *in-situ* sensing signals, integrated with other process and product data and engineering knowledge, to achieve *in-process* quality improvements.

CONTACT Jianjun Shi Jianjun.shi@isye.gatech.edu

Figure 1. Conventional quality control methods vs. IPQI.

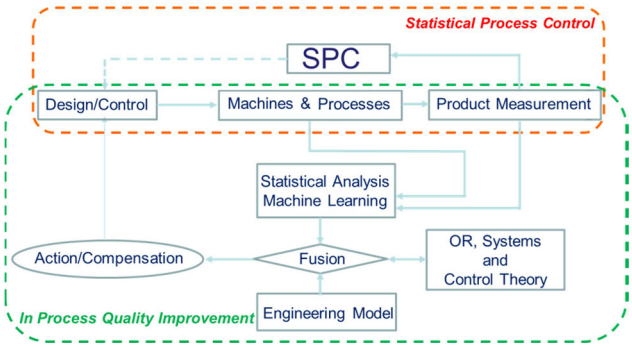


Figure 2. Comparison of SPC vs. IPQI.

- *What is the innovation of the IPQI, and what are its unique characteristics?*

IPQI refers to a set of methodologies of engineering-driven data fusion for process monitoring, root cause diagnosis, and feedback and feed-forward control. The data concerned in IPQI embodies those methodologies throughout the life cycle of a process and product, ranging from product design, process design, *in-situ* sensors, product quality measurement, and maintenance information, among others. Data fusion is achieved by developing advanced statistical and machine learning methods guided by engineering knowledge. This resulting method or decision-making is further enhanced by optimization methods and control theories (Figure 2). By implementing IPQI, one expects to achieve root cause diagnosis (in addition to change detection), online automatic control (in addition to off-line adjustment), and defect prevention (in addition to defect inspection) in manufacturing systems.

- *How does IPQI enhance conventional machine automation with a focus on quality improvements?*

Machine automation is essential for manufacturing systems. High-precision robots, machine tools, and equipment are well developed by using conventional system and control theory. As

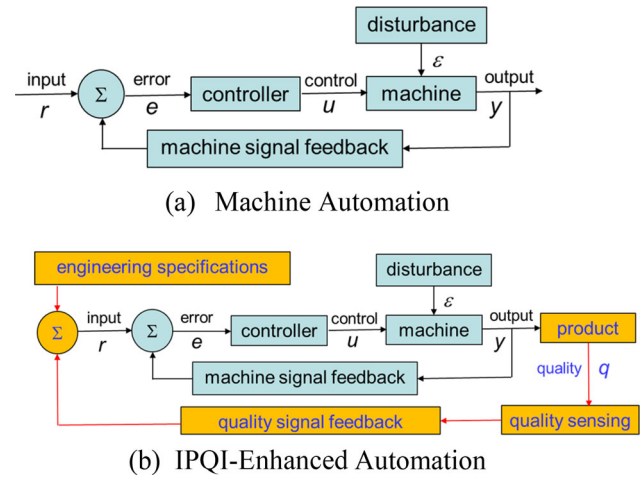


Figure 3. Machine automation vs. IPQI-enhanced automation.

shown in Figure 3(a), the typical control objective of "machine automation" is to adjust the system's inputs to achieve high precision of the outputs, typically reaching and maintaining the setting points of position, speed, temperature, force, torque, etc. Using machine automation methods, the best settings of a machine or production line are needed *a priori* for delivering the best products by a manufacturing system.

Although machine automation provides the potential capability to produce a high-quality product, it does not completely close the whole loop when *in-situ* product quality measurements are available. The *IPQI-enhanced automation* closes this loop by fully utilizing the *in-situ* quality sensing signals with feedback or feed-forward control to determine the (best) machine setting points as illustrated in Figure 3(b). This outer loop provides quality measurement feedback to the machine inputs/controls to achieve engineering specifications defined by the product design.

A systematic comparison between traditional machine automation and IPQI-enhanced automation is provided in Table 1. The IPQI-enhanced automation has several unique characteristics in terms of measurement signals, sampling frequencies, system dynamics, which raise challenges in system modeling and control algorithm development. Depending on the combinations of those characteristics, some unique algorithms are worthy of investigation.

- *What is the evolution of the IPQI and its applications?*

Since the introduction of the IPQI (Shi, 1996), a tremendous amount of effort has been made by the author himself and others interested in IPQI to develop and enrich IPQI methodologies and applications. The evolution of IPQI can be observed from different aspects related to types of data

Table 1. Comparison between machine automation and IPQI-enhanced automation.

	Machine automation	*IPQI-enhanced automation*
Feedback variables	machine operation status	product quality status
Output data types	homogeneous	heterogeneous (digital, text, image)
Model	differential or difference equations	Various depending on quality data type/format, automation capabilities
Data sampling freq.	uniform high freq. sampling	mixed sampling rate, low frequency
Control algorithms	PID, adaptive, robust, fuzzy, etc.	depending on the model & objective
System characteristics	dynamic system	dynamic or static, or mixed system

addressed, system control and statistical methods, and applications. First, the types of data addressed in IPQI have evolved from real-time univariate data to multivariate data, time series, functional/waveform signals, high-dimensional streaming signals, high-resolution images, and high-resolution video signals. Second, regarding system control and statistical methods, IPQI methodologies span the fields of univariate and multivariate statistical methods (e.g., principal component analysis, factor analysis, singular value decomposition, partial least squares regression, etc.), wavelet analysis, causal discovery and modeling, tensor-based modeling and analysis, and more recently, various emerging machine learning methods. Third, applications of IPQI have been implemented in different manufacturing systems, such as automotive assembly, machining, forming, rolling, semiconductor manufacturing, nanomanufacturing, and aerospace manufacturing processes.

2. IPQI methodologies for assembly, machining, and forming

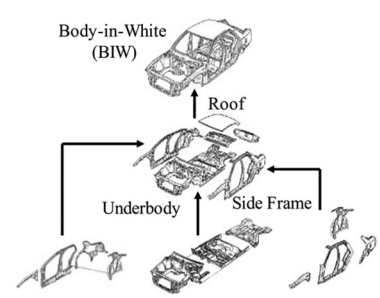
In the 1990s, *in-situ* sensing technology were introduced and implemented in some commercial production systems and provided in-process measurements of quality characteristics on 100% of the products. At the same time, the real-time sensing of process variables and machine conditions (e.g., temperature, force, torque, vibration, etc.) was widely adopted with advanced data acquisition systems. The data format is typically characterized as vectors (e.g., dimensional measurements of stamped parts, assemblies, or machined parts) and functional waveform signals (e.g., forming force, pressure, welding current, etc.). The conventional SPC methods are ineffective in dealing with those types of data generated from 100% measurement of products and functional waveform signals. Those new challenges and opportunities motivated the development of IPQI methodologies and applications. A key characteristic of the IPQI methodology development is the fusion of advanced statistics, control theory, and engineering domain knowledge. This section will discuss the IPQI methodologies development and applications, with the focus on root cause diagnosis for assembly, waveform signature analysis for forming, Stream of Variation (SoV) theory for multistage manufacturing, and causation-based process control.

2.1. Root cause diagnosis for automotive body assembly

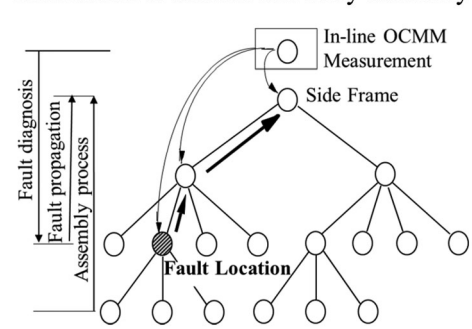
In the early 1990s, an in-line Optical Coordinate Measurement Machine (OCMM) was installed in the assembly line to measure every car assembly, which provided about 130 critical dimensions for a final Body-In-White (BIW). The conventional SPC control charts cannot handle those massive data due to unavoidable false alarms or false negatives (i.e., missed defects). Furthermore, even if SPC control charts correctly detect the changes of the dimensions in a BIW, the root cause identification of those changes remains a significant challenge, since a BIW assembly typically has 150 stamped parts assembled in hundreds of assembly stations with thousands of assembly fixtures/locators (Ceglarek *et al.*, 1994). The challenges, i.e., how to fully utilize the in-line OCMM sensing data and quickly identify root causes for dimensional variation reduction, triggered the initial research of IPQI.

Based on the OCMM measurement, the first effort constitutes the identification of the assembly station that contributes to the final dimensional variation in the BIW. In order to achieve this goal, a hierarchical model of the BIW assembly process was developed to represent the stamped parts, assembly stations, subassemblies, and assembly sequence based on the engineering design, as shown in Figure 4 (Ceglarek *et al.*, 1994). Then, in-line OCMM data was analyzed by focusing on sensing data with large variation, clustering those data, and mapping each data cluster into the hierarchical model to find the candidate parts and candidate stations of root causes. A set of decision rules were developed to find the station that constitutes the root cause of the dimensional variation (Ceglarek *et al.*, 1994; Ceglarek and Shi, 1995).

After a root cause station is identified, the next question is which fixture locator failed, leading to the large dimensional variation. Principal Component Analysis (PCA) (Wold *et al.*, 1987) was utilized to analyze the parts dimensional measurements to address this problem. It has been proven that the first principal component has a bijective (one-to-one) correspondence with the failure vector obtained from the fixture design if a single locator fails (Ceglarek and Shi, 1996). Further study was conducted to consider multiple faults (Apley and Shi, 1998) and with measurement noise and model uncertainties (Ceglarek and Shi, 1999).



(a) Illustration of automotive body assembly process



(b) Hierarchical structure representation of an assembly process

Figure 4. Automotive body assembly process and hierarchical structure representation (reproduced from Shi (2006)).

To further improve the IPQI methodology for assembly, a series of topics have been investigated, including optimal sensor layout for fixture diagnosis (Khan *et al.*, 1999), diagnosis with multiple assembly stations (Shiu *et al.*, 1996), and root cause diagnosis for compliant parts assembly (Rong *et al.*, 2000, 2001).

2.2. Waveform analysis for cyclic signals for stamping or forming

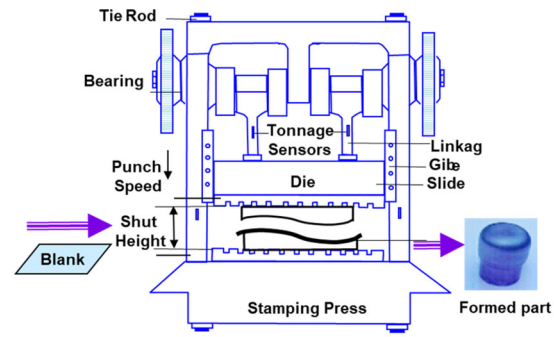
The waveform analysis for cyclic signals was motivated by stamping or forming applications. A typical quality assurance strategy in a stamping plant is to take a sample of parts and inspect each one for its Key Product Characteristics (KPCs). If a quality concern arises, production engineers will try to identify the root causes and take appropriate actions to resolve the quality issue. However, this practice has two limitations: (i) the inspected samples only constitute a small subset of the fabricated products, and therefore, cannot guarantee a 100% quality satisfaction; (ii) it is difficult to find root causes from more than 40 process variables that may lead to a defective part.

To address these limitations, IPQI concepts were used in the stamping process control. In a stamping press, there are many sensors to measure the stamping press and the die in each stamping cycle, including tonnage sensors, shut heights, nitrogen cushion pressure, vibration sensors, etc. These sensors provide cyclic signals corresponding to each stamped part. Furthermore, each cycle of the waveform signals can be divided into multiple segments, and each segment corresponds to specific mechanic interactions among the press, the die, and the part (Figure 5). Thus, these in-press/in-die sensors provide rich information for the IPQI methodology development for stamping.

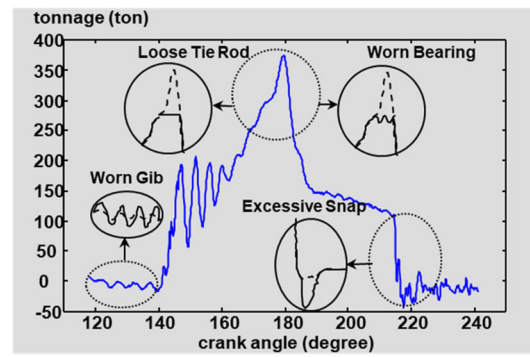
Koh *et al.* (1995) studied the tonnage signal analysis with press maintenance information. Jin and Shi (1999a) further identified the time and frequency characteristics in the segmented tonnage signals and used wavelet analysis to extract features from those signals for process monitoring and diagnosis. With the development of the models between tonnage signal and product quality (Jin and Shi, 2000; Ding *et al.*, 2006, Kim *et al.*, 2006; Kim *et al.*, 2007), the tonnage signal can be used to directly monitor the quality of the stamped parts. In essence, the IPQI study for waveform signal sets a basis for later research and the expansion of the IPQI research for automotive engine manufacturing (Paynabar and Jin, 2011), nanomanufacturing (Yue *et al.*, 2016), and semiconductor manufacturing (Zhang *et al.*, 2018).

2.3. Stream of Variation theory for multistage manufacturing processes

A multistage system refers to a system consisting of multiple units, stations, or operations to finish the final product or service. The multistage system is very common in modern manufacturing processes. In most cases, the final product quality of a Multistage Manufacturing Process (MMP) is determined by complex interactions among multiple stages



(a) Stamping press and *in situ* sensors



(b) Tonnage signal, segmentation, and associated faults

Figure 5. Segmentation of a tonnage signal (reproduced from Jin and Shi (1999a)).

– the quality characteristics of one stage are not only influenced by the local variations at that stage, but also by the propagated variations from upstream stages. Multistage systems present significant challenges, yet also opportunities for quality engineering research.

The concept of Stream of Variation has been proposed to describe the complex production stream and the data stream involved in modeling and analysis of variation and its propagation in an MMP (Figure 6). An interpretation of the SoV reflects the complex data relationships in an MMP (Figure 7). As shown in Figure 7, the X-axis represents the manufacturing stages; the Y-axis represents the time (or product job number over time); the Z-axis represents the quality attributes. Thus, $M_{x,y,z}$ represents the z-th quality attributes in the stage x at time y. In an MMP, there are three types of correlations among those data streams: (i) the quality attributes are auto-correlated in terms of the stages along the production line along the X-axis; (ii) the quality attributes are cross-correlated among them within the same stage along the Z-axis; and (iii) each quality attribute is also auto-correlated in terms of time due to the degradation or wear of production tooling over time along the Y-axis.

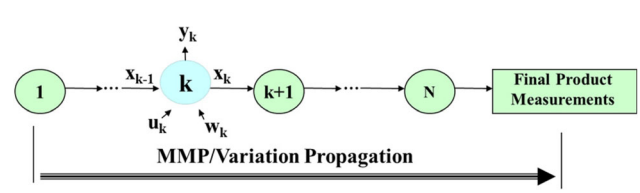


Figure 6. Variation propagation and notations in the SoV modeling (reproduced from Shi (2006)).

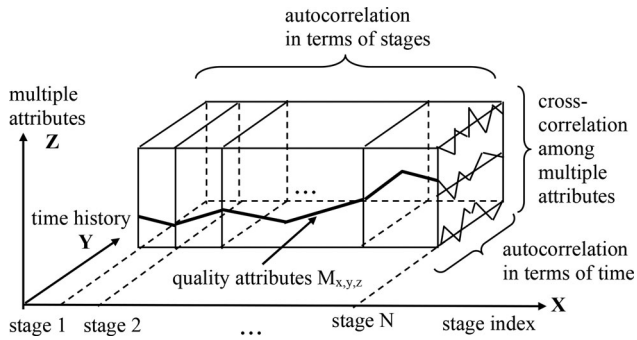


Figure 7. A 3D diagram of the MMP illustrating the complex data relationships (revised from (Ding *et al.*, 2002).

These three types of correlations, observed in the data stream, introduce significant challenges in variation modeling, analysis, and control. The SoV methodology (Shi, 2006) was developed to investigate the variations of these data streams.

The foundation of the SoV methodology is a mathematical model that links the key product quality characteristics with key control characteristics (e.g., fixture error, machine error, etc.). This model has a state space representation (Jin and Shi, 1999b) that describes the deviation and its propagation in an N-station process (as shown in Figures 6 and 7), i.e.

$$\boldsymbol{x}_k = \boldsymbol{A}_{k-1}\boldsymbol{x}_{k-1} + \boldsymbol{B}_k\boldsymbol{u}_k + \boldsymbol{w}_k, \qquad k \in \{1,\ 2,\ \ldots,\ N\}, \tag{1}$$

$$\boldsymbol{y}_k = \boldsymbol{C}_k\boldsymbol{x}_k + \boldsymbol{v}_k, \qquad k \in \{1,\ 2,\ \ldots,\ N\}, \tag{2}$$

where k is the stage index and $k \in \{1,\ 2,\ \ldots,\ N\}$. $\boldsymbol{x}_k$ is the state vector representing the key quality characteristics of the product (or intermediate workpiece) after stage k. $\boldsymbol{u}_k$ is the control vector representing the tooling errors (e.g., small random deviation within tolerance when no faults occur, or large deviation when failures occur on the tooling) at stage k. $\boldsymbol{y}_k$ is the measurement vector representing product quality measurements at stage k. $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are modeling error and sensing error, respectively. The coefficient matrices $\boldsymbol{A}_k$, $\boldsymbol{B}_k$, and $\boldsymbol{C}_k$ are determined by product and process design information: $\boldsymbol{A}_k$ represents the impact of deviation transition from stage $k-1$ to stage k, $\boldsymbol{B}_k$ represents the impact of the local tooling deviation on the product quality at stage k, and $\boldsymbol{C}_k$ is the measurement matrix, which can be obtained from the defined quality features of the product at stage k. Evidently, the SoV theory assumed linearity in its models, i.e., Equations (1) and (2). The suitability of the linearity was confirmed by Ren *et al.* (2006) through their study quantifying the impact of nonlinearity in typical MMPs.

With the mathematical models (1) and (2), variation reduction can be achieved in both design and manufacturing stages through rigorous mathematical analysis for decision-making. However, significant challenges exist in both model development and utilization to realize the benefits of the analytical capability of this model. These challenges were addressed in the SoV methodological research (Shi, 2006). In detail, the SoV methodology addressed the following important questions for the variation reduction and IPQI in an MMP:

- *How to model and integrate the product and process design information for variation reduction?*
  In Shi (2006), two basic methods were investigated: the physics-based modeling method and the data-driven modeling method. In the former method, the kinematics relationship between Key Control Characteristic (KCC) and KPC is identified through a detailed physical analysis of the manufacturing process (Ding *et al.*, 2002); whereas in the latter method, the model is achieved through a statistical estimation procedure based on historical measurement data. The details of SoV modeling are discussed in Chapters 6, 7, and 8 of Shi (2006).
- *How to systematically identify the root causes of variability in terms of which manufacturing station and what faults in the station introduce the variability?*
  During continuous production, a product variation may occur at any stage of an MMP, due to the tooling wearing out, tooling breakage, and incoming part variation. The SoV book (Shi 2006) presents a systematic approach for root cause identification. In this approach, a new concept of "statistical methods driven by engineering models" is proposed to integrate the product and process design knowledge with statistical analysis. The variation models (1) and (2), developed from the design information, are used to link the product variation (quality characteristics) with the tooling variation (or potential failures) (Ding *et al.*, 2002). The product features are measured during the production, and the data are used to conduct statistical analysis – based on the model (1) – to identify root causes. To this end, advanced statistics and estimation theories are utilized. The SoV book (Shi, 2006) presented two types of diagnostic techniques for root cause identification for the MMPs: a variation pattern matching method for pre-defined faulty patterns in Chapter 10 and an estimation-based diagnosis for detecting changes in monitoring statistics in Chapter 11.
- *How to distribute measurement sensors for effective process control in an MMP by determining when, where, and what to measure regarding the final and intermediate workpieces?*
  One of the major tasks in variation reduction is to design gauging strategies to measure product features in an MMP. Most of the existing industrial practices focus on product conformity inspection (i.e., product-oriented measurements), which is effective for detecting product imperfection, but may not be effective in identifying root causes of product variation. The SoV book proposed a "process-oriented" measurement concept with a distributed sensing strategy (Ding *et al.*, 2003; Liu *et al.*, 2005; Ding and Apley, 2007). In this strategy, selected key control characteristics, as well as selected key product characteristics, will be measured in the selected stages in order to simultaneously detect product defects and identify their root causes. Chapter 12 of Shi (2006) discusses a wide range of issues and methods on the optimal

sensor placement and distribution in an MMP. One fundamental metric used in guiding sensor placement is the so-called diagnosability (Ding *et al.*, 2002; Zhou *et al.*, 2003; Apley and Ding, 2005), namely the capability to identify potential root causes of the process variation for a given measurement strategy. The issue of diagnosability is comprehensively analyzed in Chapter 9 of Shi (2006).

- *How to conduct design evaluation and tolerance synthesis to ensure product quality for an MMP?*
  Variation analysis and design evaluation are conducted in the product and process design stage to identify critical components, features, and manufacturing operations (Kim and Ding, 2004; Ding et al., 2005; Kim and Ding, 2005). With the SoV model defined in (1), the following three tasks can be performed: (i) tolerance analysis by allocating the intermediate part tolerance ($\boldsymbol{x}_0$) and tooling tolerance ($\boldsymbol{u}_k$) and then predicting the final product natural tolerance ($\boldsymbol{x}_N$) by solving the difference equations; (ii) tolerance synthesis by fixing the final product specifications on tolerance ($\boldsymbol{x}_N$) and then assigning the natural tolerance for individual parts ($\boldsymbol{x}_0$) and tooling components ($\boldsymbol{u}_k$) by minimizing the specified cost objective functions; and (iii) sensitivity analysis by identifying the critical parts ($\boldsymbol{x}_k$) or tooling components ($\boldsymbol{u}_k$) that have a significant impact on the final product variation by evaluating the defined sensitivity indices. Details of these topics are discussed in Chapters 13, 14, and 15 in Shi (2006).
- *How to integrate product quality and production tooling reliability for an effective system configuration and tooling design, and maintenance decision-making?*
  There is a complex, intriguing relationship between product quality and tooling reliability. A degraded (or failed) production tooling will lead to a larger product variability or number of defects; meanwhile, the variability of product quality features will impact the degradation rate or failure rate of production tooling. For an MMP, these interactions are more complex as variations propagate from one stage to the next stage (Chen *et al.*, 2006). Thus, a "chain effect" between the product quality (Q) and tooling reliability (R) can be observed and is thus denoted as the "QR Chain" effect. The modeling of the QR Chain is an integrated effort of the SoV model and Semi-Markov process models. Chapter 16 of Shi (2006) discusses the modeling of the QR Chain effects for MMPs. Chapter 17 of Shi (2006) investigates the applications of the QR Chain effect in reliability and maintenance decisions.

The SoV theory for MMP is fully explained in Shi (2006); a summary of the SoV concepts can be found in Shi (2014), and Shi and Zhou (2009) provide a survey of emerging methodologies for tackling various challenges in multi-stage systems.

2.4. Causation-based quality control

Most of the existing multivariate quality control research and methods focus on *correlation* or *association* among variables, which concerns how to predict some system features reliably and accurately from other features of a system. However, for effective process control, there is a need to identify the *cause–effect* relationships (also called causal relationships) among variables that cannot be inferred solely based on the correlation or association. This idea leads to the concept of "causation-based quality control" (Li and Shi, 2007; Li *et al.* 2008; Liu and Shi, 2013, Liu *et al.* 2013), which is built upon observational data, causal modeling, causal inference, and decision-making. It should be pointed out that the causation in the "causation-based quality control" is inherent cause–effect relationships among variables in a system. The findings of causal relationships can be based on the physical laws or physical understanding of a system, or by using data–driven causal discovery methods from observationary data, or combining both methods as "engineering-driven causal discovery from observational data".

Observational data is referred to as the sensing data obtained from a manufacturing system during the production, as well as other data generated in the product and process design. In general, they are passively observed, as opposed to experimental data in which one or more variables are manipulated (often randomly) and the effects on other variables are measured. Observational data is more readily available than experimental data, especially in complex manufacturing systems where the excessive number of variables and practical constraints prohibit the execution of designed experiments. As observational data becomes increasingly available, the opportunities for successful causal discovery increase.

Causal discovery from observational (uncontrolled non-experimental) data is challenging and of interest to many researchers and practitioners. The challenges in causal discovery involve an intricate interplay between the assumptions on the data generating process, patterns of associations in the data, and aspects of causal processes that are consistent with the constraints and can explain the patterns in the data. Although various research efforts have been made to develop generic causal modeling algorithms, most of the implementations and applications focus on problems in social and medical science. However, the literature on causal modeling based on manufacturing data is sparse, especially for process control problems.

In a real manufacturing system, the causal relationships are complicated, nonlinear, and dynamic, which generates considerable difficulties for causal modeling of an underlying system. It is almost impossible to develop a universal causal modeling method without knowledge of the underlying manufacturing system. Therefore, the research emphasis is placed on developing an engineering-driven causal modeling approach that integrates the generic statistical causal discovery algorithm with manufacturing domain knowledge. In a manufacturing system, the information flow is determined by the nature of each physical activity and the topology of the physical system. The information related to the key process and product features are evolving in the system following engineering principles. Some engineering domain

knowledge exists from product and process design, which helps identify the key variables and potential causal relationships. Meanwhile, the data captured by the in-process sensor records the process changes and interrelationships among variables in practice. By integrating those two sets of information (information flow defined by production system layout and engineering principles and data), a causal model can be discovered from the observational data and further used to develop effective process control strategies. In the past decade, concepts of causation-based quality control have been proposed. Several topics have been investigated, including but not limited to the following:

- *Causation-based quality control for rolling processes* (Li *et al.*, 2008): In this study, an integrated approach is proposed to discover the causal model represented by a causal Bayesian network (or causal network for short). In the developed causal discovery approach, engineering domain knowledge is embedded in the statistical causal discovery algorithms in various critical stages of the modeling process. With the integrated approach, an effective and efficient causal model is obtained (Figure 8). The approach is demonstrated with a rolling process control problem, where the real production data are collected for causal discovery. In the rolling process, the product quality is measured by the number of surface defects, and the process variation is measured by 22 variables, collected from two major manufacturing stages, continuous casting (pre-rolling) and rolling. With the causal network representation, causal relationships among variables can be identified both qualitatively and quantitatively. The results can further facilitate diagnosis, prediction, and the development of control strategies.
- *Optimal sensor allocation by integrating causal models and optimization algorithms* (Li and Jin, 2010; Liu and Shi, 2013): Optimal sensor allocation for system anomaly detection is an important research topic in quality engineering. An optimal sensor allocation method is developed in a distributed sensing network to timely detect anomalies in an underlying physical system. This method involves two steps: first, a Bayesian Network (BN) is built to represent causal relationships among physical variables in the system; second, an integrated algorithm by combining the BN and a set-covering algorithm is developed to determine which physical variables should be sensed, in order to minimize the total sensing cost as well as satisfying a prescribed detectability requirement (Li and Jin, 2010). Further studies were conducted to develop better sensor allocation strategies based on the causal relationships to achieve optimal performance with the minimal sensor cost (Liu and Shi, 2013). Case studies were performed on a hot forming process and a large-scale cap alignment process, showing that the developed algorithm satisfies both cost and detectability requirements.
- *Causation-based T^2 decomposition for multivariate process monitoring and diagnosis* (Li *et al.*, 2008): Multivariate process monitoring and diagnosis is an important and challenging issue. The widely adopted Hotelling T^2 control chart can effectively detect a change in a system, but cannot diagnose the root causes of the change. The MTY approach (Mason *et al.*, 1995) makes efforts to improve the diagnosability by decomposing the T^2 statistic. However, this approach is computationally intensive and has a limited capability in root cause diagnosis for high-dimensional variables. The developed causation-based T^2 decomposition method (Li *et al.*, 2008) integrates the causal relationships revealed by a BN with the traditional MTY approach. Theoretical analysis and simulation studies demonstrate that the proposed method substantially reduces the computational complexity and enhances the diagnosability compared with the MTY approach.

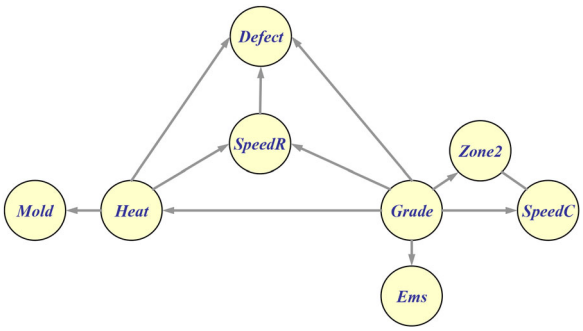


Figure 8. An abstracted causal model developed from rolling process data (Li *et al.*, (2008)).

3. IPQI methodologies for high-dimensional streaming data

In the past decade, high-resolution images and video signals have been increasingly introduced to generate multichannel, High-Dimensional (HD) streaming data in a production system. As shown in Figure 9, those HD streaming data can be obtained in each manufacturing stage to measure both process variables and product quality characteristics. Those HD streaming data contain rich information about production system conditions and quality responses. Furthermore, there are inherent interrelationships among those HD streaming data, due to interactions within a manufacturing stage and between manufacturing stages due to the quality/variation propagation. New IPQI methodologies need to be developed to address the challenges and availabilities of HD streaming data in more sophisticated multistage manufacturing systems. Those challenges and opportunities lead to new IPQI research by using advanced algorithms, such as tensor-based analysis and

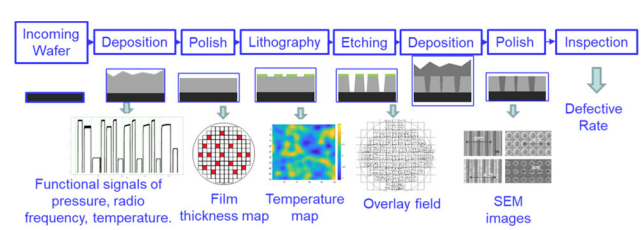


Figure 9. Heterogeneous inputs and outputs with HD streaming data in a semiconductor MMP.

modeling, engineering-driven machine learning, advanced optimization algorithms, and computation capabilities.

3.1. HD streaming data monitoring and diagnosis

HD streaming data refers to multiple functional waveform signals or high-resolution video signals that are used to monitor manufacturing process variables or product quality variables. With the wide adoption of *in-situ* sensors, HD streaming data is prevalent in many advanced manufacturing processes, such as semiconductors, hot rolling, nanomanufacturing, 3D printing, etc. There are a set of challenges in monitoring and diagnosis based on HD streaming data, including unknown patterns of anomalies, random time and location of occurrence of anomalies in HD streaming data. Furthermore, predominantly in-control samples, but a small number of out-of-control samples (or anomalies) in a production system, are available. Labeling anomalies in HD streaming data is either too costly or even infeasible, due to the lack of a complete set of anomalies that can cover all possible types of manufacturing defects. Thus, supervised learning is infeasible or too costly in developing effective monitoring and diagnosis algorithm for HD streaming data. A set of algorithms have been developed in the past decade to address those challenges. We discuss two categories of methodologies in the remainder of this subsection.

3.1.1. Adaptive sampling-based statistical decision-making for change detection

In the literature, most of the existing methods focus on process monitoring for lower-dimensional data or complete measurement data. To extend the monitoring framework to HD data with incomplete measurement, Liu *et al.* (2015) developed an adaptive sampling algorithm with top-r statistical decision–making to automatically update CUSUM statistics in HD streaming data to detect and localize the anomalies. The method assumes that there are a limited number of sensors (denoted by s) to monitor multivariate HD stream data (denoted as dimension p, $s \ll p$), and the sensor can be reallocated to monitor a new data stream at each sampling point. A CUSUM statistic is updated with real sensing data if the data stream is measured with a sensor; or is updated with a predetermined off-set value (Δ) if the data stream is not measured with a sensor. Based on the CUSUM statistics, two decisions will be made: (i) sensors will always be reallocated to measure the data streams that have top-s largest CUSUM statistics, and (ii) the system is out of control if the summation of the top-r largest CUSUM statistics is larger than a decision threshold. The values of off-set value Δ, r, and the decision threshold are design parameters of the algorithm.

The algorithm of an adaptive sampling algorithm with top-r statistics decision-making (Liu *et al.*, 2015) has desirable properties, including (i) it guarantees to cover all HD data streams if there is no anomaly occurring, and (ii) it detects the anomaly quickly with probability 1 if anomalies occur. Furthermore, a top-r decision-making strategy has a unique interpretation and excellent intuition as it always focuses on a few (e.g., top-r) statistics that behave the worst in the system. Thus, it provides effective capabilities for system-level anomaly detection.

Since the algorithm was developed, various efforts have been made to generalize the concept to other HD streaming data monitoring and diagnosis problems. Zhang and Mei (2020) extended this framework into a Bayesian decision framework. Nabhan *et al.* (2021) extended the above method by considering the cross-correlation of variables, which significantly improves the anomaly detection performance when the data is cross-correlated.

3.1.2. Decomposition-based methods for anomaly detection and the extensions

For more efficient anomaly detection, Yan *et al.* (2017) first proposed the Smooth Sparse Decomposition (SSD) method to estimate the smooth background data, detect sparse anomalies, and filter noises automatically and simultaneously. In the SSD method, the problem formulation can be represented as

$$\text{argmin}_{\boldsymbol{\theta},\boldsymbol{\theta}_a} \|\boldsymbol{e}\|_2^2 + \lambda \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{R} \boldsymbol{\theta} + \gamma \|\boldsymbol{\theta}_a\|_1, \ \ s.t. \ \ y = \boldsymbol{B}\boldsymbol{\theta} + \boldsymbol{B}_a \boldsymbol{\theta}_a + \boldsymbol{e}, \tag{3}$$

where $\boldsymbol{B}\boldsymbol{\theta}$, $\boldsymbol{B}_a\boldsymbol{\theta}_a$, and $\boldsymbol{e}$ indicate the background, anomalies, and error, respectively. $\boldsymbol{B}$ and $\boldsymbol{B}_a$ are the smooth bases for the background (mean) and anomalies, which need to be specified by the engineers, $\boldsymbol{R}$ is the roughness matrix and λ, γ are tuning parameters. Finally, proximal gradient methods are developed to optimize (3) efficiently.

Several extensions on the decomposition methods have been developed. The first direction is to extend SSD to higher-order tensors, or spatio-temporal data, with larger dimensionality. In the original SSD work, the authors also proposed to extend the method by defining the basis $\boldsymbol{B}$ and $\boldsymbol{B}_a$ as the tensor product of the basis in each dimension $\boldsymbol{B} = \boldsymbol{B}_2 \otimes \boldsymbol{B}_1$ and $\boldsymbol{B_a} = \boldsymbol{B_{a2}} \otimes \boldsymbol{B_{a1}}$, which is illustrated in Figure 10. To deal with HD, nonstationary data with spatio-temporal correlation structure, Yan *et al.* (2018) extended the SSD to the ST-SSD algorithm that addresses HD streaming video signals. Shen *et al.* (2022) further extended the SSD to smooth and sparse tensor decomposition to address the challenges of incomplete tensor data, as well as learning the basis automatically from the data.

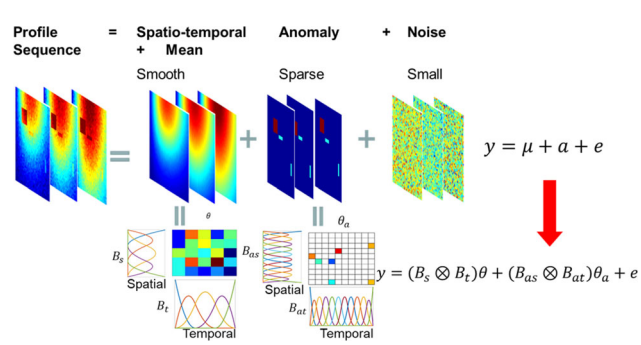


Figure 10. Illustration of SSD for image/video signals (Yan *et al.*, 2017).

Another line of research focuses on extending the SSD framework to other data structures, such as non-smooth background or temporally clustered anomalies. For example, Yue *et al.* (2017) extended the decomposition idea to the waveform signals with wavelet basis to penalize mixed-effects decomposition and applied it to multichannel profile change detection with random effects in nanomanufacturing. Yan *et al.* (2021) further extended this method by incorporating the temporal consistency of anomaly structures in hotspot detection in the additive manufacturing. Mou *et al.* (2021) extended the SSD to Additive Tensor Decomposition (ATD) by incorporating structural data information such as low rank and smoothness along specific modes or their combinations to decompose tensor data.

Finally, some other works improved the efficiency by adaptively observing incomplete high-resolution video signals. For example, Yan *et al.* (2020) further extended the SSD to the AKM^{2}D algorithm by integrating adaptive sampling, which combines the exploration and exploitation of HD data streams. Guo *et al.* (2020) extended the SSD to a Bayesian SSD method that adopts an adaptive sampling framework to decide the best sampling locations in a dynamic setting.

3.2. Tensor decomposition-based modeling and analysis for multi-channel HD data

In many engineering applications, both the system inputs (or process variables) and outputs (product quality variables) are multi-channel signals measured by multiple sensors, image signals, or video signals. Those waveform signals or image signals can be represented in tensor format, as shown in Figure 11. Early works focus on various PCA methods to reduce the dimensionality of the profile data or multi-channel profile (Zhang *et al.*, 2018). However, these methods require vectorizing the original tensor data into vectors, which leads to a loss of the important structural information contained in the original tensor.

The major idea of tensor-based process monitoring and quality modeling is dimensionality reduction. This can be achieved by utilizing the tensor structures with the help of various tensor decomposition methods, such as Tucker and Candecomp/Parafac (CP) decompositions.

For the process monitoring, Yan *et al.* (2014) developed a tensor-based process monitoring technique for flame monitoring in a steel manufacturing process. The main idea is to utilize Tucker and CP decompositions to reduce the dimensionality of the original tensor data and build monitoring statistics on the core tensors with the reduced dimensionality and the residual tensor. There are also other tensor-based process monitoring methods developed, including UMPCA (Paynabar *et al.*, 2013) and multichannel PCA (Paynabar *et al.*, 2016).

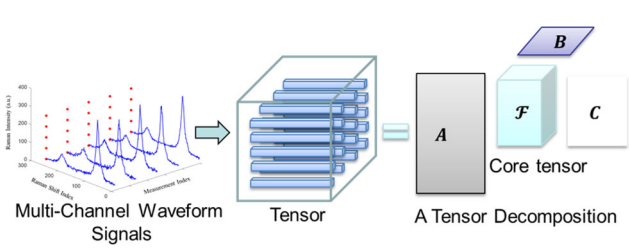


Figure 11. Multichannel Raman spectra, a tensor; A type of tensor decomposition.

For the quality monitoring, Yan et al. (2010) developed a tensor-on-scalar regression model of the output tensor variable with respect to the input variables with Tucker decomposition to model the influence of the process variables on the shape of the product. Gahrooei *et al.* (2021) studied a more general case on multiple tensor-on-tensor regression (MTOT), where both input and output variables are tensors; and Tucker decomposition (Kolda and Bader, 2009) is then applied on the tensor coefficient to reduce the parameter dimensionality as

$$\mathcal{Y}_i = \sum_{j=1}^{p} \mathcal{X}_{ji} * \mathcal{B}_j + \mathcal{E}_i, \;\; i \in \{1, \ldots, M\}, \tag{4}$$

where $\mathcal{Y}_i$ is the response tensor $i \in \{1, \ldots, M\}$, $\mathcal{X}_{ji}$ is the input tensor $i \in \{1, \ldots, M\}$, $j \in \{1, \ldots, p\}$, $\mathcal{B}_j$ is the model parameter to be estimated, and $\mathcal{E}_i$ is an error tensor. p is the number of inputs.

The proposed MTOT method addresses the problem of modeling the relationship between the system output (a scalar, curve, image, or structured point cloud) and heterogeneous system input variables. Wang *et al.* (2021) extended the MTOT model to address missing data/signal problems by integrating the concept of tensor completion. Miao *et al.* (2022) extended the MTOT regression by considering the interaction effects of input tensors in the model.

Yue *et al.* (2020) proposed a Tensor Mixed-Effects (TME) model to analyze massive HD Raman spectroscopy data with complex structures. The proposed TME model can (i) separate fixed effects and random effects in a tensor domain; (ii) explore correlations along different dimensions; and (iii) realize the efficient parameter estimation by a proposed iterative double Flip-Flop algorithm. Yue *et al.* (2020) also investigated properties of the TME model, the existence and identifiability of parameter estimation, and applied it to the in-line Raman signal monitoring of a nanomanufacturing process. Gao *et al.* (2020) proposed an optimal integration of tensor decomposition and ensemble learning, where tensor decomposition with regularization can select the critical features and enhance the performance of ensemble learning. This method has been applied to *in-situ* quality evaluation of friction stir blind riveting.

3.3. IPQI for other heterogeneous data types such as distributed data, categorical data, and event data

Heterogeneous HD data (such as sparse inputs, mixed profile outputs, event data, categorical data, and point cloud data) can be generated in some MMPs, which raises new challenges for IPQI. With the advancements of Machine Learning (ML) algorithms, the IPQI has entered a new era. Here are a few examples of how ML algorithms were developed to solve IPQI problems, which involves complex and heterogeneous data

types (such as distributed data, categorical data, and event data), that conventional methodologies cannot solve.

The first direction is to model the heterogeneous data distributed in different stages of a production line in MMPs. MMPs are equipped with complex sensing systems, which generate data with several unique characteristics: the output quality measurements from each stage are of different types, the comprehensive set of inputs (or process variables) have distinct degrees of influence over the process, and the relationship between inputs and outputs is sometimes ambiguous, and multiple types of faults repetitively occur during the process operation. To address those challenges, Wang and Shi (2021) proposed a holistic modeling approach for MMPs, aiming at understanding how intermediate quality measurements of mixed profile/image outputs relate to sparse effective inputs across the entire MMP.

The second direction involves modeling categorical variables, denoting the process configurations and product customizations. These categorical variables lead to a flexible relationship between input process variables and output quality measurements, because there are many potential configurations of the manufacturing process. It causes significant challenges for data-driven process modeling and root cause diagnosis. To combine both categorical quality variables and the continuous quality variables, Deng and Jin (2015) proposed the Quantitative and Qualitative (QQ) model to combine both types of quality responses in a manufacturing process for joint modeling. Later, Sun *et al.* (2017) extended this into functional quantitative and qualitative quality response variables. Miao *et al.* (2022) proposed a data-driven additive model to address the effects of different categorical variables on the relationship between process variables and quality measurements. The estimation algorithm automatically identifies variables that have a significant impact on the product quality, aggregates the levels of each categorical variable based on *a priori* knowledge of level similarity, and provides an accurate model that describes the relationship between process variables and quality measurements.

Event data is another common data type in manufacturing systems including maintenance events and failure events, etc. The first line of research focuses on modeling soft failure, which is typically modeled by degradation signals. Techniques such as Bayesian analysis (Gebraeel *et al.*, 2005), data-fusion-driven prognostics (Liu *et al.* 2013; Liu and Huang, 2014; Yan *et al.*, 2016), and deep learning prognostic analysis methods have been developed (Fink *et al.*, 2020; Kim and Liu, 2020; Wang *et al.* 2021). The second line of research focuses on modeling hard failure, where time-to-event types of models and point process-based models are integrated. Rather than modeling each individual event, recent research has been developed on monitoring and modeling multiple event sequences (Deep *et al.*, 2021a; Deep *et al.* 2021b; Jahani *et al.* 2021). Another direction is to identify the *correct* event type from other information, including the maintenance report and real-time signals. For example, the word embedding model is developed to cluster the maintenance reports for different event data automatically (Bhardwaj *et al.*, 2021). Furthermore, a retrospective analysis method (Wang *et al.* 2021) is proposed to identify multiple events from the multichannel functional signals by updating the event signature and sequences.

4. IPQI-enhanced automation

As discussed in the introduction, the IPQI demands further development of machine automation methods with a focus on quality improvements by closing the loop of in-process measurements of quality data, which leads to IPQI enhanced automation. Depending on system dynamics, characteristics, measurement information, and control actuator capabilities, different modeling and control algorithms should be investigated to develop the IPQI-enhanced automation. Here we discuss four topics related to IPQI-enhanced automation.

4.1. DOE-based automatic process control

Design of Experiments has been widely used in the product and process design stage. It is often used to conduct efficient experiments to build a process/product model, based on which the Robust Parameter Design (RPD) can be performed off-line to effectively minimize the system performance variation to disturbances that may occur during system operations. RPD assumes that distributions of noise factors (or disturbances) are known in the design stage, thus the control factors can be pre-set to the optimal values that make the system performance robust to the changes of noise factors.

In many systems, *in-situ* sensors are used to measure some noise factors during system operations. Thus, an accurate measurement of noise variables can be obtained in real-time. At the same time, some control factors can be adjusted online during the system operation, which provides capabilities to feedback/feed-forward in-line sensing signals to improve the control performance. Thus, the DOE-based Automatic Process Control (APC) (Jin and Ding, 2004; Zhong *et al.*, 2009) was developed to fully take advantage of in-line sensing, real-time control, and DOE modeling of the system. The framework of DOE-based APC is illustrated in Figure 12.

In a DOE-based APC (Figure 12), the process variables are classified as follows: noise factors as in-line measurable factors ($\boldsymbol{e}$) and non-measurable factors ($\boldsymbol{n}$), control factors as off-line setting factors ($\boldsymbol{X}$) and in-line adjustable factors ($\boldsymbol{U}$). A DOE modeling procedure (Wu and Hamada, 2011) is adapted to identify the relationships among those

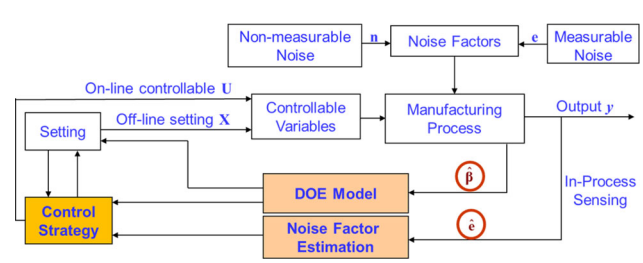


Figure 12. Illustration of DOE-based APC.

variables ($\boldsymbol{\beta}_i$ and $\boldsymbol{B}_i$, $i = 1, 2, 3, 4$) in the model below:

$$y = \beta_0 + \boldsymbol{\beta}_1^{\mathrm{T}}\boldsymbol{X} + \boldsymbol{\beta}_2^{\mathrm{T}}\boldsymbol{U} + \boldsymbol{\beta}_3^{\mathrm{T}}\boldsymbol{e} + \boldsymbol{\beta}_4^{\mathrm{T}}\boldsymbol{n} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}_1\boldsymbol{e} + \boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}_2\boldsymbol{e} + \boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}_3\boldsymbol{n} + \boldsymbol{U}^{\mathrm{T}}\boldsymbol{B}_4\boldsymbol{n} + \epsilon. \tag{5}$$

Based on the model, an optimal control index is defined as

$$J_{APC}\left(\boldsymbol{X}, \boldsymbol{U}|\widehat{\boldsymbol{e}}, \widehat{\boldsymbol{\beta}}\right) = E_{\boldsymbol{e}, \boldsymbol{n}, \boldsymbol{\beta}, \epsilon}\left[c(y - t)^2|\widehat{\boldsymbol{e}}, \widehat{\boldsymbol{\beta}}\right]. \tag{6}$$

By solving the optimization problem, a DOE-based APC strategy can be obtained for both off-line setting control variables ($\boldsymbol{X}$) and in-line control variables ($\boldsymbol{U}$). If model parameter estimation errors are considered in the control law calculation, a cautious control strategy (Shi and Apley, 1998) can be obtained to improve the robustness to modeling errors.

The DOE-based APC strategy has broad applications. Any system that uses DOE to do modeling and robust parameter design has the potential to apply the DOE-based APC if (i) there are *in-situ* sensing to measure noise variables, and (ii) there are online control/adjustment capabilities for control variables. Furthermore, SPC methods can be used to monitor process changes and the adequacy of the DOE model. The monitoring results can be integrated into supervisory strategies to achieve supervised DOE-based APC.

4.2. SoV model-based dimensional variation reduction and control

In an MMP, there are opportunities to make in-line adjustments of tooling settings at stage k to impact the output quality of this stage, as well as the output quality of all downstream stages. In the SoV model (1), such a control strategy is reflected as the adjustment of $\boldsymbol{u}_k$ to impact the quality $\boldsymbol{y}_j$ ($j = k+1, \ldots, N$). Thus, an SoV model-based control strategy can be formulated as

$$J = \min_{\boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_N} \widehat{\boldsymbol{y}}_{N/k}\boldsymbol{Q}\widehat{\boldsymbol{y}}_{N/k}^T \tag{7}$$

where $\widehat{\boldsymbol{y}}_{N/k}$ is the prediction made at stage k for the final product quality at stage N as a function of $\{\boldsymbol{u}_k, \boldsymbol{u}_{k+1}, \ldots, \boldsymbol{u}_N\}$ and $\boldsymbol{Q}$ is a weight matrix.

The key ideas of the SoV model-based control can be summarized as follows:

1. At stage k, all quality features ($\boldsymbol{x}_k$) up to stage k can be obtained via *in-situ* sensing and estimation.
2. A prediction of the final product quality at stage N is made by using the SoV model (1) at stage k, assume $\boldsymbol{u}_j \equiv 0$, $j \in \{k+1, k+2, \ldots, N\}$.
3. An adjustment of tooling locator ($\boldsymbol{u}_k$) at stage k is performed to minimize the variance of the predicted quality at stage N.
4. At next stage $k+1$, we repeat the steps 1 to 3 until the product arrives at the final stage N.

The SoV model-based control has been implemented for dimensional variation reductions in the automotive assembly process (Izquierdo *et al.*, 2007). The method was further improved with the Bayesian estimation to adaptively estimate unknown parameters during production, which serves as inputs to obtain the optimal control strategy via dynamic programming (Chaipradabgiat *et al.*, 2009). To address model uncertainties, SoV-based cautious control was investigated for MMPs (Zhong *et al.*, 2010).

4.3. Tensor-based feedback control

Based on the tensor-on-tensor regression method, Zhong *et al.* (2022) proposed an image-based feedback control method where the system outputs are image/video signals with temporal and spatial correlation. Zhong *et al.* (2022) developed a novel tensor-based process control approach by incorporating the tensor time series and regression techniques, which models the image output as a tensor $\mathcal{Y}_t$ at time t:

$$\mathcal{Y}_t = \Sigma_{j=1}^{p}\mathcal{Y}_{t-j} * \mathcal{A}_j + \Sigma_{n=1}^{l}\mathcal{X}_{t-n} * \mathcal{B}_n + E_t, \tag{8}$$

where $\mathcal{X}_t$ is the control signal and E_t is the temporally correlated error at time t. $\mathcal{A}_j$, $j \in \{1, \ldots, p\}$ and $\mathcal{B}_n$, $n \in \{1, \ldots, l\}$ are coefficient tensors. Using Tensor Basis representation, we have:

$$\mathcal{B}_n = \mathcal{C}_{B_n} \times_1 U_{B_n 1} \ldots \times_l U_{B_n l} \times_{l+1} V_{B_n 1} \times_{l+2} \ldots \times_{l+d} V_{B_n d}, \tag{9}$$

$$\mathcal{A}_j = \mathcal{C}_j \times_1 U_{j1} \times_2 \ldots \times_d U_{jd} \times_{d+1} V_{j1} \times_{d+2} \ldots \times_{2d} V_{jd}, \tag{10}$$

where U and V are orthogonal basis matrices and $\mathcal{C}$ is the core tensor. The unknown parameters to be estimated constitute $\mathcal{C}$, U, V to get $\mathcal{A}$ and $\mathcal{B}$. More discussions on the notations and modeling algorithems can be found in Zhong *et al.* (2022). By using tensor representation, the number of parameters to be estimated is reduced from 4.3×10^9 unknown parameters to 3153 unknown parameters, if we assume that a 256×256 image is used as the system output.

Based on the process model (8), a control law $\mathcal{X}_t$ is obtained by minimizing an objective function;

$$\min_{\mathcal{X}_t} E\left(\widehat{\mathcal{Y}}_{t+1}(\mathcal{X}_t) - T\right)^2, \tag{11}$$

where $\widehat{\mathcal{Y}}_{t+1}(\mathcal{X}_t)$ is the one-step-ahead prediction made at time t, and T is the control target.

It should be pointed out that the image-based feedback control is still in its infant stage. There are strong assumptions to obtain model (8): the low-rank property of control signals and response signals. There are a few topics worthy of further investigation, including (i) how to relax low-rank assumptions; (ii) how to consider model parameter and modeling uncertainties; (iii) how to adopt model predictive control concepts, and using a sliding prediction horizon (Shi and Apley, 1998), instead of one-step-ahead output prediction, as the objective function; and (iv) how to address the heterogeneous outputs with a mixture of images, functional curves, etc.

4.4. Automatic shape control for composite fuselage assembly

Another important IPQI issue in the aerospace industry is the automatic shape control for composite fuselage assembly.

Due to the dimensional variation among fuselages, there could be a mismatch, or gap, between two fuselages to be joined (Gates, 2007; Sloan, 2020), as shown in Figure 13. Thus, two fuselages need shape control methods to adjust their dimensions to match each other for assembly. In order to realize automatic shape control for the composite fuselage, efforts have been made to develop the foundation of fuselage control concerning (i) the modeling of the fuselage control systems; and (ii) the development of optimal control algorithms.

4.4.1. Modeling of the fuselage assembly system

The shape control for fuselage assembly requires modeling the relationships between the fuselage shape deviation ($\boldsymbol{Y}_k$) and the actuator inputs ($\boldsymbol{F}_k$), which can be represented as

$$\boldsymbol{Y}_k = f(\boldsymbol{F}_k) + \boldsymbol{\varepsilon}$$

In order to find the relationship, various efforts have been made with different assumptions and conditions. Yue *et al.* (2018) developed a surrogate model considering uncertainties (e.g., actuators' uncertainty, part uncertainty, and unquantified uncertainty), and derived the best linear unbiased predictor for this model. Zhang and Shi (2016a, 2016b) proposed the SoV modeling for compliant composite parts, in which the part manufacturing error, fixture position error, and relocation-induced error are integrated within a state space model. Wen *et al.* (2018, 2019) developed a finite element simulation platform to exactly mimic the fabrication process of composite fuselages, which considers the detailed material property, ply design, fixture structure, and actuators installation. The simulation platform was calibrated with real fuselage experimental data by using the sparse learning method (Wang *et al.*, 2020), and then it was used to conduct feasibility analysis, virtual assembly, and stress analysis for composite fuselage assembly.

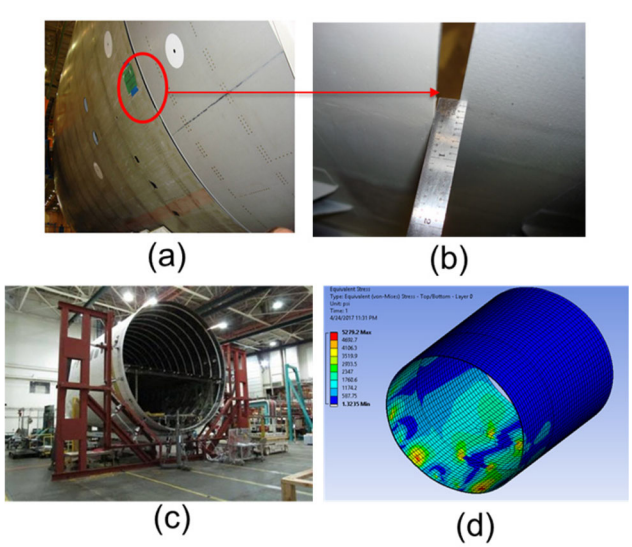


Figure 13. Illustration of Shape control of composite fuselage: (a-b) dimensional gap between two fuselage sections (Gates, 2007; Sloan, 2020); (c) shape control system with actuators, and (d) stress analysis for shape control (Wen *et al.*, 2018).

4.4.2. Optimal control of the fuselage assembly

Based on the model of the shape control system, an optimal control strategy is obtained by minimizing the gap between the deviations of a real fuselage from its target shape, e.g.,

$$\boldsymbol{F}^* = \underset{\boldsymbol{F}}{\operatorname{argmin}}\, \boldsymbol{J} = (\boldsymbol{Y}_c + \boldsymbol{Y}(\boldsymbol{F}) - \boldsymbol{Y}^*)^T \boldsymbol{W} (\boldsymbol{Y}_c + \boldsymbol{Y}(\boldsymbol{F}) - \boldsymbol{Y}^*),$$

where $\boldsymbol{Y}_c$ is the initial dimension, $\boldsymbol{Y}(\boldsymbol{F})$ is the magnitude of the shape change due to actuator force $\boldsymbol{F}$, and $\boldsymbol{Y}^*$ is the target shape. To obtain the control law, initial efforts were made to develop an accurate predictive model for the shape control of a fuselage (Yue *et al.*, 2018). However, each fuselage may have its own unique dimension errors in practice. Thus, it is desirable to place the actuators at the best positions on the edge of a fuselage, which will lead to the best shape control precision while minimizing the adjustment forces. This was a challenging problem until Du *et al.* (2019) applied the sparse learning method to find the optimal locations for actuators and the optimal force applied by each actuator. In the sparse learning method, the actuator location selection problem was formulated into a convex optimization problem, which minimizes the loss function $L(\boldsymbol{F}) = (\boldsymbol{\Psi} + \boldsymbol{U}\boldsymbol{F})^T \boldsymbol{B} (\boldsymbol{\Psi} + \boldsymbol{U}\boldsymbol{F}) + \lambda \|\boldsymbol{F}\|_1$. In this function, $\boldsymbol{\Psi}$ represents the initial shape distortions, $\boldsymbol{U}$ represents the displacement matrix, $\boldsymbol{F}$ is the force vector associated with the actuators. $\boldsymbol{B}$ is a diagonal weighted matrix indicating the importance of different measurement points. By regularizing the L_1 norm of $\boldsymbol{F}$, a limited number of actuators' locations will be selected to conduct the shape control.

Considering the modeling error in parameter estimations, a cautious control method (Zhong *et al.*, 2021) is developed to improve the robustness of the control results. Recently, a sparse sensor placement-based adaptive control strategy (Mou *et al.* 2021) has been developed to actively compensate the control performance based on the feedback information. To this end, a minimal subset of *in-situ* dimensional measurements is obtained to efficiently reconstruct the entire response surface based on the concept of compressive sensing and this information is applied to the perturbed system caused by process and product variability for feedback control.

Even though this subsection discusses the dimensional shape control for fuselage assembly, the concepts and procedures are applicable to other assembly processes where there are larger inherent dimensional variations on each part/subassembly, and meanwhile, there are high precision requirements on the assembled products.

5. The future of IPQI methodologies and applications

We are currently in the era of Industry 4.0. The rapid advancements in sensor technologies, communication networks, and computing power have resulted in temporally and spatially dense data-rich environments in modern manufacturing systems. In addition to the volume, the data have an increased complexity such as streaming data, spatial-temporal structures, or network relationships; please refer to a

recent book dedicated to the same challenges in the nanomaterial and nanomanufacturing (Park and Ding, 2021). Furthermore, the Industrial Internet of Things (IIoT) provides a great capability for data harvesting, which leads to deeply intertwined physical and software components (Nguyen *et al.*, 2021). The IPQI for such a sophisticated cyber-physical system will demand interdisciplinary efforts to capture and characterize many challenges associated with collecting, managing, analyzing, and visualizing massive amounts of data. In detail, these challenges/opportunities can be put into four categories:

- *Data harvesting and curation.* With the arrival of tremendous amounts of data in contemporary smart manufacturing systems, organizing, labeling, and curating the data for effective learning and knowledge extraction remains challenging.
- *Application of emerging ML and Artificial Intelligence (AI) techniques to IPQI.* Many recently developed ML/AI techniques have demonstrated significant modeling and predictive power for complex engineering problems such as image analysis, natural language processing, etc. Applying these techniques to IPQI presents a significant opportunity in achieving the next quantum leap in IPQI.
- *Application of IPQI strategies to emerging manufacturing processes.* In addition to traditional metal-based manufacturing processes, IPQI can be applied to other manufacturing processes such as additive manufacturing, nano/bio manufacturing processes, and process industry and create significant quality improvements.
- *Cybersecurity issues in smart manufacturing.* The wide application of information technology in manufacturing systems creates great opportunities for smart process control and operations. However, those opportunities are accompanied by more potential security vulnerabilities. Hence, cybersecurity issues need to be addressed to confidently apply and implement IPQI strategy in manufacturing, which heavily depends on information technologies.

The following list of topics consists of timely topics of IPQI worthy of further investigation according to the author's point of view. However, it is not intended to be a comprehensive list of all future IPQI research topics; rather, it is intended to serve as inspiring examples and guiding directions that could be extended to more open issues and applications in the years to come.

5.1. Data harvesting and curation

(1) *Automatic data pre-processing, synchronization, and alignment*: In an MMP, sensing data are continuously collected over time by each process sensor. In order to conduct meaningful data analytics, the data need to be processed to make the sensing data aligned with individual product and tasks. (i) In a discrete part manufacturing process, efforts are needed to connect those continuous sensing data from all MMP with *each* product and further align them correctly. The abnormal parts or signals need to be identified and cleaned in this effort. More complexity may occur if the MMP has a mix of serial-parallel configurations. To this end, Miao *et al.* (2022) have made some initial efforts for hot rolling processes. More efforts are needed to make such automatic process-to-product data connections and alignments. (ii) In a continuous manufacturing process, a series of chemical/physical actions occur one after another. Each action may involve different process variables with different relationships. It is essential to automatically detect the change points of such a process action and automatically identify the clusters of variables associated with each change/action. Zhang *et al.* (2021) investigated dynamic clustering of functional data assuming that functional data has the same mathematical basis in each segment. Qian *et al.* (2017) and Qian *et al.* (2019) investigated, both retrospectively and prospectively, change detection for delineating the growth stages in nanomaterial production. Overall, manually aligning, segmenting, and labeling the data could be very time-consuming in practice. There is a strong demand to develop more generic, robust, and scalable algorithms to automate this process.

(2) *Holistic retrospective data analytics*: In practice, the massive data generated from a production system are not re-visited again, which leads to a significant loss of information and knowledge. As an example, one day of production in a semiconductor plant will generate 2TB of data. Less than 20% of those data are processed for quality control and system improvement, and more than 80% of data has less usage or has not been processed. Thus, holistic *retrospective* data analytic methodologies are needed to systematically identify special events that occurred in specific production periods, and further identify when they occurred and which variables/machines/products are involved in those special events. This retrospective data analysis will provide opportunities to find the relevant datasets, develop models, and further develop decision-making algorithms that are capable and efficient for the IPQI in future production systems. Wang and Shi (2021) explored this idea in a hot rolling process, but more investigations on this topic are needed.

(3) *Synthetic defect data generation*: Nowadays, companies attempt to train automatic inspection or diagnostic systems based on ML techniques. However, one main challenge of building such an ML-based system is the lack of defect samples in the training dataset because: (i) the in-control manufacturing process is data-rich but defect-rare, which creates a highly imbalanced training dataset. Using those imbalanced datasets will lead to very low or even no detection in conventional classification methods (Byon *et al.*, 2010); (ii) the lack of defect annotations due to the expensive and time-consuming defect labeling process, which requires domain expertise and needs to be conducted by experienced engineers. To circumvent the challenge of the

scarcity of defect data, developing effective synthetic defect data generation is critical. Synthetic data generation has been studied in autonomous vehicles and robotic navigation (Tsirikoglou *et al.*, 2017) and healthcare (Walonoski *et al.*, 2018), where data is hard to collect, but a large amount of data is needed. We also note that one branch of methods in handling imbalanced datasets is based on synthetic data generation (Chawla *et al.*, 2002; Pourhabib *et al.*, 2015). However, the explorations in the advanced manufacturing field are limited (Bo *et al.*, 2020; Niu *et al.*, 2020; Li *et al.* 2021), where the integration of ML techniques with physical knowledge of the manufacturing process plays an important role. More discussions on synthetic data generation can be found in Libes *et al.* (2017).

5.2. Application of emerging ML/AI techniques to IPQI

(1) *Deep learning-enabled IPQI*: Deep learning has achieved great success in many different domains, such as computer vision and natural language processing. Many ongoing efforts apply deep learning methods to the monitoring and modeling of manufacturing systems. Related to process monitoring, given no labeled data is provided, initial efforts such as autoencoders (Alfeo *et al.*, 2020), variational autoencoders (Sergin and Yan, 2021), and generative adversarial learning (Yan *et al.* 2019; Kusiak, 2020) have been developed for process monitoring of images or profile data in a manufacturing system. Related to failure mode classification and quality prediction, recurrent neural networks (Mozaffar *et al.*, 2018; Tian *et al.*, 2021; Wang *et al.* 2021) and convolutional neural networks (Lee *et al.,* 2017; Guo *et al.*, 2022) have been developed. Bayesian neural networks have also been developed to quantify the uncertainty in a manufacturing system (Ferreira *et al.*, 2019). Recently, there is also some research focusing on utilizing deep learning for joint modeling of multiple manufacturing stages transitions (Yan *et al.* 2021), with the ability to identify the most important input variables in each stage. However, the adaptation of deep learning methods to heterogeneous data, unstructured point cloud data, and HD streaming data in manufacturing systems with moderate sample sizes remains an open challenge and is worthy of further investigation. Please refer to Wang *et al.* (2018) for a more detailed review of deep learning methods in manufacturing systems.

(2) *Reinforcement learning-enabled data-driven automation control for IPQI*: The ubiquitous data and information flows provide us with new opportunities for the control and automation of manufacturing systems. As mentioned before, DOE-based APC and surrogate model-based control are two examples that leverage data-driven methods to estimate the model for improving control and automation. It can be foreseen that manufacturing automation will be dramatically enhanced at the device level, machine level, and system levels via interconnected cyber physical systems by making full use of the data resources and ML techniques such as Reinforcement Learning (RL). Recently, RL has achieved great success in the ML literature. For example, AlphaGo and AlphaZero have been developed to master chess, shogi, the game of Go, which can realize self-play without human interventions and can beat human experts (Silver *et al.* 2017; Silver *et al.,* 2018). However, such a RL strategy requires a large number of simulated games, and the amount of data required to learn an action policy may be infeasible for manufacturing systems. Furthermore, these RLs are typically applied to simulated games with little randomness or noise. Some initial efforts on applying RL to manufacturing systems have emerged for maintenance optimization (Xanthopoulos *et al.* 2017; Liu *et al.*, 2020; Yang *et al.*, 2021), task scheduling (Park *et al.*, 2019; Dong *et al.*, 2020), and adaptive control (Dornheim *et al.*, 2020; Kuhnle *et al.*, 2021). More research is needed to apply RL methods for quality improvement in manufacturing systems. At this stage, it remains unclear whether an AI-based system (or an "AlphaZero" for manufacturing) can achieve self-optimization and automatic IPQI of manufacturing systems or even outperforms human manufacturing experts under some scenarios.

(3) *Transfer learning and domain adaptation learning-enabled IPQI*: One specific challenge about the manufacturing system is that each manufacturing process is unique due to different physical processes, different machines, and different process settings. Unlike computer vision or natural language processing, there is no "gold standard" dataset that can be used for all manufacturing systems. Even within the same manufacturing process, such as additive manufacturing, each part design is unique and leads to completely different process variable observations. Being able to transfer the knowledge from one manufacturing process (i.e., source domain) to the other (i.e., target domain) is essential to obtain an accurate model for a new manufacturing process even with limited data. To address this challenge, transfer learning and domain adaptation learning techniques (Azamfar, Li, and Lee, 2020) are developed for quality modeling (Cheng *et al.*, 2017; Sabbaghi and Huang, 2018; Li *et al.* 2021) and statistical process control (Tsung *et al.*, 2018). More efforts are needed to identify which knowledge, at what level and accuracy, in what format, to conduct transfer learning for a given class of manufacturing systems.

(4) *Weakly supervised learning*: A broad field to deal with limitations in the quantity and quality of labeled samples is known as weakly-supervised learning. Weakly-supervised learning assumes that for an unlabeled or partially labeled sample Y, there may be domain or within-data knowledge that can provide some weak labelling information of Y. Weakly-supervised learning

can be divided into three major sub-fields: (i) Incomplete-supervision: only a subset of samples is labeled, (ii) inexact-supervision: only coarse labels are available, and (iii) inaccurate-supervision: given labels are not always ground truth (Zhou, 2018; Nodet *et al.*, 2020). Incomplete supervision is often expected in the quality control of an industrial process, where anomaly labels are scarce by nature, or comprehensive quality testing to produce enough accurate labels is costly and possibly infeasible. One example is the laser welding application, in which error events can be distinguished from regular frames in terms of their spatial appearance and temporal dynamics, whose class labels are only available for the complete sequence, and the quality of the produced part is determined to be either error-free or faulty (Jager *et al.*, 2008). Two major sub-fields that branch from incomplete supervision are active learning (Settles, 2009) and semi-supervised learning (Chapelle *et al.*, 2006; Zhu, 2008; Zhou and Li, 2010).

Active learning assumes that we can interactively query an 'oracle', such as a human expert, to get labels for selected unlabeled samples (Settles, 2009). In the statistical literature, it is sometimes also called sequential experimental design. Conventional ML models acquire a training dataset first before training, whereas the active learning approach queries the data interactively to maximize the data efficiency and ultimately reduces the number of samples needed for labeling. Readers are referred to Shim *et al.* (2021) for initial use of the active learning approach in manufacturing systems.

Semi-Supervised Learning (SSL) assumes no 'oracle' intervention; instead, it aims to integrate labeled and unlabeled samples to train a model to predict or classify Y (e.g., product quality varables) using X (e.g., process variables). There are different types of SSL algorithms such as cluster-based methods (Chapelle, Weston, and Schölkopf 2003) and graph-based methods (Blum and Chawla 2001). Readers are referred to Okaro *et al.* (2019) and Kang *et al.* (2016) for some applications of SSL in manufacturing.

(5) *Distributed computing and federated learning*: Industry 4.0 emphasizes interconnected systems, cyber-physical systems, and IIoT, which generates tremendous data to represent system operation conditions and performances. Overall, the amount of data collected in the entire manufacturing system is too large to be analyzed by existing cloud data storage and cloud computing, due to the communication loss. In order to effectively take advantage of cloud data storage and cloud computing, several fundamental problems need to be addressed, including how to extract valuable and relevant information quickly, how to decide what data should be calculated locally and what should be sent to the server. For IPQI, those questions are more relevant, as it typically requires real-time decision-making. A system failure analysis should be well conducted to provide a basis for those decisions. Another related challenge is the data fusion for time-variant sampling data: the *in-situ* data may have different sampling frequencies, including high frequency, such as vibration, intermediate frequency, such as pressure or temperature, and low frequency such as the quality measurement for discrete parts. There are some initial efforts in this direction: On the software side, the usage of distributed computing platforms such as Apache Hadoop and Spark has enabled large-scale distributed computing. On the algorithm side, methods such as federated learning have enabled efficient model inference without aggregating the raw data across different machines while obtaining the near-optimal solution in a reasonable amount of computational time. Interested readers are referred to O'Donovan *et al.* (2015) and Kusiak (2017) for a review of big data challenges in manufacturing systems and federated learning.

5.3. IPQI application to emerging manufacturing processes

3D printing, or Additive Manufacturing (AM), has been increasingly adopted in the defense, aerospace, and medical industries. Various *in-situ* sensing systems have been installed to monitor AM processes in real-time to prevent poor dimensional tolerances, surface roughness, and materials and structural defects (Tapia and Elwany, 2014; Rao *et al.*, 2015). An important direction is how to achieve *in-situ* product evaluation and quantification on the material or functional characteristics, which provides confidence on the part and saves time from the post-fabrication evaluations (Mani *et al.*, 2017). To this end, IPQI is a very suitable framework to conduct data fusion for real-time monitoring and anomaly detection to enable *in-situ* product evaluation and quantification. To achieve this goal, a model needs to be developed to link the raw material properties, product design, *in-situ* sensing signals during printing, and post-printing part testing and quantifications. Uncertainty quantifications and analysis should be conducted to consider the impacts of data acquisition, signal-to-noise ratio, and tolerance of the engineering specifications (Colosimo *et al.*, 2018).

5.4. Cybersecurity issues in IPQI

In the era of Industry 4.0, the increased software-defined automation, control, and monitoring of manufacturing assets across connected networks also increase the risk of cyber-attacks. The cyber-threats may compromise the integrity of manufacturing assets (manufacturing systems and processes, machine tools, fabricated parts), reduce manufacturing productivity, and increase costs. Some cyber-threats, including integrity attacks, are only partially observable in cyberspace alone, and therefore need to be detected and diagnosed through the inter-dependency analysis of both cyber and physical signals. Thus, there is a significant opportunity to jointly analyze physical and cyber signals to advance the trustworthiness and resilience of manufacturing system operations. By utilizing the inherent strength of IPQI

methodologies such as the monitoring and control of the complex system via *in-situ* sensing data, there is a significant opportunity to extend IPQI to perform Cyber-threat Detection and Diagnosis (CDD) in manufacturing systems. The IPQI-CDD framework should monitor various cyber and physical signals and perform cyber-threat detection and root cause diagnosis through advanced cyber-physical data fusion and taint analysis. The goal is to enable the prevention and mitigation of potential harms at an early stage via proactive and predictive countermeasures and system design. This effort needs to integrate and analyze the *in-situ* process and quality signals and the signals from cyber networks of manufacturing systems to detect and diagnose cyber-threats. New data analytics methodologies are to be developed to integrate cyber and physical signals to gain the fundamental understanding of cyber-threat detection and diagnosis in manufacturing systems, and to further expand to other cyber-physical systems. In line with the IPQI principles, engineering domain knowledge needs to be integrated to analyze and synthetically generate attacks on CPS and develop a comprehensive and holistic CDD framework. The underlying data assumptions of IPQI methodologies need to be revisited in the face of adversarial, stealthy attacks executed by intelligent attackers of a CPS, and novel methodologies need to be developed accordingly. Readers are referred to Elhabashy *et al.* (2019) for more discussions on this topic.

6. Summary

This article provides a historical review of the evolution of concepts, methodologies, and applications of the IPQI research and applications. The IPQI methodologies have been evolving and call upon significant new developments due to (i) the advancements of advanced data analytics and ML/AI techniques and unprecedented computation capabilities; (ii) the availability of tremendous sensing signals, data acquisition, and networking capabilities; and (iii) the requirements of high precision, performance, productivity, flexibility, agility, and low cost in manufacturing systems. The author's sincere belief is that IPQI will advance alongside Industry 4.0, Cyber-Physical Systems, Smart Manufacturing, and in general, the advancements of modern manufacturing systems.

ORCID

Jianjun Shi http://orcid.org/0000-0002-3774-9176

References

Alfeo, A.L., Cimino, M.G., Manco, G., Ritacco, E. and Vaglini, G. (2020) Using an autoencoder in the design of an anomaly detector for smart manufacturing. *Pattern Recognition Letters*, **136**, 272–278.

Apley, D.W. and. Ding, Y. (2005) A characterization of diagnosability conditions for variance components analysis in assembly operations. *IEEE Transactions on Automation Science and Engineering*, **2**(2), 101–110.

Apley, D.W. and Shi, J. (1998) Diagnosis of multiple fixture faults in panel assembly. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **120**, 793–801.

Azamfar, M., Li, X. and Lee, J. (2020) Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, **33**(3), 445–453.

Bhardwaj, A.S., Deep, A.,Veeramani, D. and Zhou, S. (2021) A custom word embedding model for clustering of maintenance records. *IEEE Transactions on Industrial Informatics*, **18**(2), 816–826.

Blum, A. and Chawla, S. (2001) Learning from labeled and unlabeled data using graph mincuts, in *Proceedings of the Eighteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers Inc., San Francisco, CA, pp. 19–26.

Bo, L., Yuan, X. and Shi, M. (2020) Synthetic data generation based on local-foreground generative adversarial networks for surface defect detection. *Journal of Electronic Imaging*, **29**(1), 013016.

Byon, E., Shrivastava, A.K. and Ding, Y. (2010) A classification procedure for highly imbalanced class sizes. *IIE Transactions*, **42**(4), 288–303.

Ceglarek, D. and Shi, J. (1995) Dimensional variation reduction for automotive body assembly. *Manufacturing Review*, **8**(2), 139–154.

Ceglarek, D. and Shi, J. (1996) Fixture failure diagnosis for autobody assembly using pattern recognition. *ASME Transactions, Journal of Engineering for Industry*, **118**(1), 55–65.

Ceglarek, D. and Shi, J. (1999) Fixture failure diagnosis for sheet metal assembly with consideration of measurement noise. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **121**(4), 771–777.

Ceglarek, D., Shi, J. and Wu, S. (1994) A knowledge-based diagnostic approach for the launch of the auto-body assembly process. *ASME Transactions, Journal of Engineering for Industry*, **116**(4), 491–499.

Chaipradabgiat, T., Jin, J. and Shi, J. (2009) Optimal fixture locator adjustment strategies for multi-station assembly processes. *IIE Transactions*, **41**(9), 843–852.

Chapelle, O., Scholkopf, B.and Zien, A. (2006) *Semi-Supervised Learning*, MIT Press, Cambridge, MA.

Chapelle, O., Weston, J. and Schölkopf, B. (2003) Cluster kernels for semi-supervised learning. In S. Becker, S. Thrun, and K. Obermayer (ed.), *Advances in Neural Information Processing Systems* 15, pp. 585–592, MIT Press, Cambridge, MA.

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, **16**, 321–357.

Chen, Y., Ding, Y., Jin, J. and Ceglarek, D. (2006) Integration of process-oriented tolerancing and maintenance planning in design of multistation manufacturing processes. *IEEE Transactions on Automation Science and Engineering*, **3**(4), 440–453.

Cheng, L., Tsung, F. and Wang, A. (2017) A statistical transfer learning perspective for modeling shape deviations in additive manufacturing. *IEEE Robotics and Automation Letters*, **2**(4), 1988–1993.

Colosimo, B.M., Huang, Q., Dasgupta, T. and Tsung, F. (2018) Opportunities and challenges of quality engineering for additive manufacturing. *Journal of Quality Technology*, **50**(3), 233–252.

Deep, A., Zhou, S. and Veeramani, D. (2021a) Copula-based multi-event modeling and prediction using fleet service records. *IISE Transactions*, **53**(9), 1023–1036.

Deep, A., Zhou, S. and Veeramani, D. (2021b) A data-driven recurrent event model for system degradation with imperfect maintenance actions. *IISE Transactions*, **54**(3), 271–285.

Deng, X. and Jin, R. (2015) QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics*, **57**(3), 320–331.

Ding, Y. and Apley, D. (2007) Guidelines for placing additional sensors to improve variation diagnosis in assembly processes. *International Journal of Production Research*, **45**(23), 5485–5507.

Ding, Y., Ceglarek, D. and Shi, D. (2002) Design evaluation of multi-station assembly processes by using state space approach. *ASME Transactions, Journal of Mechanical Design*, **124**(3), 408–418.

Ding, Y., Ceglarek, D. and Shi, J. (2002) Fault diagnosis of multistage manufacturing processes by using state space approach. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **124**(2), 313–322.

Ding, Y., Jin, J., Ceglarek, D. and Shi, J. (2005) Process-oriented tolerancing for multi-station assembly systems. *IIE Transactions*, **37**(6), 493–508.

Ding, Y., Kim, P., Ceglarek, D. and Jin, J. (2003) Optimal sensor distribution for variation diagnosis in multistation assembly processes. *IEEE Transactions on Robotics and Automation*, **19**(4), 543–556.

Ding, Y., Shi, J. and Ceglarek, D. (2002) Diagnosability analysis of multi-station manufacturing processes. *ASME Transactions, Journal of Dynamic Systems, Measurement, and Control*, **124**(1), 1–13.

Ding, Y., Zeng, L. and Zhou, S. (2006) Phase I analysis for monitoring nonlinear profiles in manufacturing processes. *Journal of Quality Technology*, **38**(3), 199–216.

Ding, Y., Zhou, S. and Chen, Y. (2005) A comparison of process variance estimation methods for in-process dimensional control. *ASME Transactions*, *Journal of Dynamic Systems, Measurement, and Control*, **127**(1), 69–79.

Dong, T., Xue, F., Xiao, C. and Li, J. (2020) Task scheduling based on deep reinforcement learning in a cloud manufacturing environment. *Concurrency and Computation: Practice and Experience*, **32**(11), e5654.

Dornheim, J., Link, N. and Gumbsch, P. (2020) Model-free adaptive optimal control of episodic fixed-horizon manufacturing processes using reinforcement learning. *International Journal of Control, Automation and Systems*, **18**(6), 1593–1604.

Du, J., Yue, X., Hunt, J.H. and Shi, J. (2019) Optimal placement of actuators via sparse learning for composite fuselage shape control. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **141**(10), 101004.

Elhabashy, A.E., Wells, L.J. and Camelio, J.A. (2019) Cyber-physical security research efforts in manufacturing–a literature review. *Procedia Manufacturing*, **34**, 921–931.

Ferreira, R.S.B., Sabbaghi, A. and Huang, Q. (2019) Automated geometric shape deviation modeling for additive manufacturing systems via Bayesian neural networks. *IEEE Transactions on Automation Science and Engineering*, **17**(2), 584–598.

Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.J. and Ducoffe, M. (2020) Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence*, **92**, 103678.

Gahrooei, M.R., Yan, H., Paynabar, K. and Shi, J. (2021) Multiple tensor-on-tensor regression: An approach for modeling processes with heterogeneous sources of data. *Technometrics*, **63**(2), 147–159.

Gao, Z., Guo, W. and Yue, X. (2020) Optimal integration of supervised tensor decomposition and ensemble learning for in situ quality evaluation in friction stir blind riveting. *IEEE Transactions on Automation Science and Engineering*, **18**(1), 19–35.

Gates, D. (2007) Boeing finds 787 pieces aren't quite a perfect fit. Rapport Technique, Seattle Times, Available at https://www.seattletimes.com/business/boeing-finds-787-pieces-arent-quite-a-perfect-fit/ (accessed 15 December 2021).

Gebraeel, N.Z., Lawley, M.A., Li, R. and Ryan, J.K. (2005) Residual-life distributions from component degradation signals: A Bayesian approach. *IIE Transactions*, **37**(6), 543–557.

Guo, J., Yan, H., Zhang, C. and Hoi, S. (2020) Partially observable online change detection via smooth-sparse decomposition. *arXiv preprint arXiv:2009.10645*.

Guo, S., Wang, D., Chen, J., Feng, Z. and Guo, W. (2022) Predicting nugget size of resistance spot welds using infrared thermal videos with image segmentation and convolutional neural network. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **144**(2), 021009.

Izquierdo, L.E., Shi, J., Hu, S.J. and Wampler, C.W. (2007) Feedforward control of multistage assembly processes using programmable tooling. *Transactions of NAMRI/SME*, **35**, 295–302.

Jager, M., Knoll, C. and Hamprecht, F.A. (2008) Weakly supervised learning of a classifier for unusual event detection. *IEEE Transactions on Image Processing*, **17**(9), 1700–1708.

Jahani, S., Zhou, S., Veeramani, D. and Schmidt, J. (2021) Multioutput Gaussian process modulated Poisson processes for event prediction. *IEEE Transactions on Reliability*, **70**(4), 1569–1580.

Jin, J. and Ding, Y. (2004) Online automatic process control using observable noise factors for discrete-part manufacturing. *IIE Transactions*, **36**(9), 899–911.

Jin, J. and Shi, J. (1999a) Feature-preserving data compression of stamping tonnage information using wavelets. *Technometrics*, **41**(4), 327–339.

Jin, J. and Shi, J. (1999b) State space modeling of sheet metal assembly for dimensional control. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **121**(4), 756–762.

Jin, J. and Shi, J. (2000) Diagnostic feature extraction from stamping tonnage signals based on design of experiments. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **122**(2), 360–369.

Kang, P., Kim, D. and Cho, S. (2016) Semi-supervised support vector regression based on self-training with label uncertainty: An application to virtual metrology in semiconductor manufacturing. *Expert Systems with Applications*, **51**, 85–106.

Khan, A., Ceglarek, D., Shi, J., Ni, J. and Woo, T. (1999) Sensor optimization for fault diagnosis in single fixture systems: A methodology. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **121**(1), 109–117.

Kim, J., Huang, Q. and Shi, J. (2007) Latent variable based key process variable identification and process monitoring for forging. *Journal of Manufacturing Systems*, **26**(1), 53–61.

Kim, J., Huang, Q., Shi, J. and Chang, T.-S. (2006) Online multichannel forging tonnage monitoring and fault pattern discrimination using principal curve. *ASME Transactions*, *Journal of Manufacturing Science and Engineering*, **128**(4), 944–950.

Kim, M. and Liu, K. (2020) A Bayesian deep learning framework for interval estimation of remaining useful life in complex systems by incorporating general degradation characteristics. *IISE Transactions*, **53**(3), 326–340.

Kim, P. and Ding, Y. (2004) Optimal design of fixture layout in multistation assembly processes. *IEEE Transactions on Automation Science and Engineering*, **1**(2), 133–145.

Kim, P. and Ding, Y. (2005) Optimal engineering system design guided by data-mining methods. *Technometrics*, **47**(3), 336–348.

Koh, C.K., Shi, J. and Williams, W.J. (1995) Tonnage signature analysis using the Haar transform. *Transactions of NAMRI/SME*, **23**, 229–234.

Kolda, T.G. and Bader, B.W. (2009) Tensor decompositions and applications. *SIAM Review*, **51**(3), 455–500.

Kuhnle, A., Kaiser, J.-P., Theiß, F., Stricker, N. and Lanza, G. (2021) Designing an adaptive production control system using reinforcement learning. *Journal of Intelligent Manufacturing*, **32**(3), 855–876.

Kusiak, A. (2017) Smart manufacturing must embrace big data. *Nature*, **544**(7648), 23–25.

Kusiak, A. (2020) Convolutional and generative adversarial neural networks in manufacturing. *International Journal of Production Research*, **58**(5), 1594–1604.

Lee, K.B., Cheon, S. and. Kim, C.O. (2017) A convolutional neural network for fault classification and diagnosis in semiconductor manufacturing processes. *IEEE Transactions on Semiconductor Manufacturing*, **30**(2), 135–142.

Li, J. and Jin, J. (2010) Optimal sensor allocation by integrating causal models and set-covering algorithms. *IIE Transactions*, **42**(8), 564–576.

Li, J., Jin, J. and Shi, J. (2008) Causation-based T2 decomposition for multivariate process monitoring and diagnosis. *Journal of Quality Technology*, **40**(1), 46–58.

Li, J. and Shi, J. (2007) Knowledge discovery from observational data for process control using causal Bayesian networks. *IIE Transactions*, **39**(6), 681–690.

Li, W., Tsung, F., Song, Z., Zhang, K. and Xiang, D. (2021) Multi-sensor based landslide monitoring via transfer learning. *Journal of Quality Technology*, **53**(5), 474–487.

Li, Y., Shi, Z., Liu, C., Tian, W., Kong, Z. and Williams, C.B. (2021) Augmented time regularized generative adversarial network (ATR-GAN) for data augmentation in online process anomaly detection.

IEEE Transactions on Automation Science and Engineering, in-press and online published. https://doi.org/10.1109/TASE.2021.3118635

Libes, D., Lechevalier, D. and Jain, S. (2017) Issues in synthetic data generation for advanced manufacturing, in *Proceedings of the 2017 IEEE International Conference on Big Data*, IEEE Press, Piscataway, NJ, pp. 1746–1754,

Liu, C.Q., Ding, Y. and Chen, Y. (2005) Optimal coordinate sensor placements for estimating mean and variance components of variation sources. *IIE Transactions*, **37**(9), 877–889.

Liu, K., Gebraeel, N.Z. and Shi, J. (2013) A data-level fusion model for developing composite health indices for degradation modeling and prognostic analysis. *IEEE Transactions on Automation Science and Engineering*, **10**(3), 652–664.

Liu, K. and Huang, S. (2014) Integration of data fusion methodology and degradation modeling process to improve prognostics. *IEEE Transactions on Automation Science and Engineering*, **13**(1), 344–354.

Liu, K., Mei, Y. and Shi, J. (2015) An adaptive sampling strategy for online high-dimensional process monitoring. *Technometrics*, **57**(3), 305–319.

Liu, K. and Shi, J. (2013) Objective-oriented optimal sensor allocation strategy for process monitoring and diagnosis by multivariate analysis in a Bayesian network. *IIE Transactions*, **45**(6), 630–643.

Liu, K., Zhang, X. and Shi, J. (2013) Adaptive sensor allocation strategy for process monitoring and diagnosis in a Bayesian network. *IEEE Transactions on Automation Science and Engineering*, **11**(2), 452–462.

Liu, Y., Chen, Y. and Jiang, T. (2020) Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. *European Journal of Operational Research*, **283**(1), 166–181.

Mani, M., Lane, B.M., Donmez, M.A., Feng, S.C. and Moylan, S.P. (2017) A review on measurement science needs for real-time control of additive manufacturing metal powder bed fusion processes. *International Journal of Production Research*, **55**(5), 1400–1418.

Mason, R.L., Tracy, N.D. and Young, J.C. (1995) Decomposition of T2 for multivariate control chart interpretation. *Journal of Quality Technology*, **27**(2), 99–108.

Miao, H., Wang, A., Chang, T.S. and Shi, J. (2022) A product-oriented transformation of continuous streaming data for relationship mining in a hot rolling process. *ASME Transactions, Journal of Manufacturing Science and Engineering*, in-press. https://doi.org/10.1080/24725854.2021.2004626

Miao, H., Wang, A., Li, B., Chang, T.-S. and Shi, J. (2022) Process modeling with multi-level categorical inputs via variable selection and level aggregation. *IISE Transactions*, in-press and online published.

Miao, H., Wang, A., Li, B. and Shi, J. (2022) Structural tensor-on-tensor regression with interaction effects and its application to a hot rolling process. *Journal of Quality Technology*, in-press. https://doi.org/10.1080/00224065.2021.1973931

Mou, S., Biehler, M., Yue, X., Hunt, J.H. and Shi, J. (2021) SPAC: Sparse sensor placement based adaptive control for high precision fuselage assembly *IISE Transactions*, under review.

Mou, S., Wang, A., Zhang, C. and Shi, J. (2021) Additive tensor decomposition considering structural data information. *IEEE Transactions on Automation Science and Engineering*, in-press and online published. http://doi.org/10.1109/TASE.2021.3096964

Mozaffar, M., Paul, A., Al-Bahrani, R., Wolff, S., Choudhary, A., Agrawal, A., Ehmann, K. and Cao, J. (2018) Data-driven prediction of the high-dimensional thermal history in directed energy deposition processes via recurrent neural networks. *Manufacturing Letters*, **18**, 35–39.

Nabhan, M., Mei, Y. and Shi, J. (2021) Correlation-based dynamic sampling for online high dimensional process monitoring. *Journal of Quality Technology*, **53**(3), 289–308.

Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J. and Poor, H.V. (2021) Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, **23**(3), 1622–1658.

Niu, S., Li, B., Wang, X. and Lin, H. (2020) Defect image sample generation with GAN for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, **17**(3), 1611–1622.

Nodet, P., Lemaire, V., Bondu, A., Cornuéjols, A. and Ouorou, A. (2020) From weakly supervised learning to biquality learning: An introduction. *arXiv preprint arXiv:2012.09632.*

O'Donovan, P., Leahy, K., Bruton, K. and O'Sullivan, D.T. (2015) Big data in manufacturing: A systematic mapping study. *Journal of Big Data*, **2**(1), 1–22.

Okaro, I.A., Jayasinghe, S., Sutcliffe, C., Black, K., Paoletti, P. and Green, P.L. (2019) Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. *Additive Manufacturing*, **27**, 42–53.

Park, C. and Ding, Y. (2021). *Data Science for Nano Image Analysis*, Springer Nature, Cham, Switzerland.

Park, I.-B., Huh, J., Kim, J. and Park, J. (2019) A reinforcement learning approach to robust scheduling of semiconductor manufacturing facilities. *IEEE Transactions on Automation Science and Engineering*, **17**(3), 1420–1431.

Paynabar, K. and Jin, J. (2011) Characterization of non-linear profiles variations using mixed-effect models and wavelets. *IIE Transactions*, **43**(4), 275–290.

Paynabar, K., Jin, J. and Pacella, M. (2013) Monitoring and diagnosis of multichannel nonlinear profile variations using uncorrelated multilinear principal component analysis. *IIE Transactions*, **45**(11), 1235–1247.

Paynabar, K., Zou, C. and Qiu, P. (2016) A change-point approach for phase-I analysis in multivariate profile monitoring and diagnosis. *Technometrics*, **58**(2), 191–204.

Pourhabib, A., Mallick, B.K. and Ding, Y. (2015) Absent data generating classifier for imbalanced class sizes. *Journal of Machine Learning Research*, **16**(83), 2695–2724.

Qian, Y., Huang, J.Z. and Ding, Y. (2017) Identifying multi-stage nanocrystal growth using in situ TEM video data. *IISE Transactions*, **49**(5), 532–543.

Qian, Y., Huang, J.Z., Park, C. and Ding, Y. (2019) Fast dynamic nonparametric distribution tracking in electron microscopic data. *The Annals of Applied Statistics*, **13**(3), 1537–1563.

Rao, P.K., Liu, J.P., Roberson, D., Kong, Z.J. and Williams, C. (2015) Online real-time quality monitoring in additive manufacturing processes using heterogeneous sensors. *Journal of Manufacturing Science and Engineering*, **137**(6), 061007.

Ren, Y., Ding, Y. and Zhou, S. (2006) A data mining approach to study the significance of nonlinearity in multistation assembly processes. *IIE Transactions*, **38**(12), 1069–1083.

Rong, Q., Ceglarek, D. and Shi, J. (2000) Dimensional fault diagnosis for compliant beam structure assemblies. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **122**(4), 773–780.

Rong, Q., Shi, J. and Ceglarek, D. (2001) Adjusted least squares approach for diagnosis of ill-conditioned compliant assemblies. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **123**(3), 453–461.

Sabbaghi, A. and Huang, Q. (2018) Model transfer across additive manufacturing processes via mean effect equivalence of lurking variables. *The Annals of Applied Statistics*, **12**(4), 2409–2429.

Sergin, N.D. and Yan, H. (2021) Toward a better monitoring statistic for profile monitoring via variational autoencoders. *Journal of Quality Technology*, **53**(5), 454–473.

Settles, B. (2009) Active learning literature survey, Technical Report, University of Wisconsin-Madison Department of Computer Sciences. http://digital.library.wisc.edu/1793/60660

Shen, B., Wang, R., Law, A.C.C., Kamath, R., Choo, H. and Kong, Z. (2022) Super resolution for multi-sources image stream data using smooth and sparse tensor completion and its applications in data acquisition of additive manufacturing. *Technometrics*, **64**(1), 2–17.

Shi, J. (1996) CAREER: In-process quality improvement methodologies and implementation in manufacturing, https://www.nsf.gov/awardsearch/showAward?AWD_ID=9624402

Shi, J. (2006) *Stream of Variation Modeling And Analysis For Multistage Manufacturing Processes*, CRC Press, Taylor & Francis Group, Boca Raton, FL.

Shi, J. (2014) Stream of variations analysis, in *Encyclopedia of Systems and Control*, Springer London, pp. 1–6.

Shi, J. and Apley, D.W. (1998) A suboptimal N-step-ahead cautious controller for adaptive control applications. *ASME Transactions, Journal of Dynamic Systems, Measurement and Control*, **120**(3), 419–423.

Shi, J. and Zhou, S. (2009) Quality control and improvement for multi-stage systems: A survey. *IIE Transactions*, **41**(9), 744–753.

Shim, J., Kang, S. and Cho, S. (2021) Active cluster annotation for wafer map pattern classification in semiconductor manufacturing. *Expert Systems with Applications*, **183**, 115429.

Shiu, B., Ceglarek, D. and Shi, J. (1996) Multi-stations sheet metal assembly modeling and diagnostics. *Transactions of NAMRI/SME*, **24**, 199–204.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D. and Graepel, T. (2018) A general reinforcement learning algorithm that masters Chess, Shogi, and Go through self-play. *Science*, **362**(6419), 1140–1144.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M. and Bolton, A. (2017) Mastering the game of Go without human knowledge. *Nature*, **550**(7676), 354–359.

Sloan, J. (2020) Boeing conducts inspections of 787 composite inner fuselage skin., Available at https://www.compositesworld.com/news/boeingconducts-inspections-of-787-composite-innerfuselage-skin (accessed 15 December 2021).

Sun, H., Rao, P.K., Kong, Z.J. Deng, X. and Jin, R. (2017) Functional quantitative and qualitative models for quality modeling in a fused deposition modeling process. *IEEE Transactions on Automation Science and Engineering*, **15**(1), 393–403.

Tapia, G. and Elwany, A. (2014) A review on process monitoring and control in metal-based additive manufacturing. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **136**(6), 060801.

Tian, Q., Guo, S., Melder, E., Bian, L. and Guo, W.G. (2021) Deep learning-based data fusion method for in situ porosity detection in laser-based additive manufacturing. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **143**(4), 041011.

Tsirikoglou, A., Kronander, J., Wrenninge, M. and Unger, J. (2017) Procedural modeling and physically based rendering for synthetic data generation in automotive applications. *arXiv preprint arXiv: 1710.06270*.

Tsung, F., Zhang, K., Cheng, L. and Song, Z. (2018) Statistical transfer learning: A review and some extensions to statistical process control. *Quality Engineering*, **30**(1), 115–128.

Walonoski, J., Kramer, M., Nichols, J., Quina, A., Moesel, C., Hall, D., Duffett, C., Dube, K., Gallagher, T. and McLachlan, S. (2018) Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. *Journal of the American Medical Informatics Association*, **25**(3), 230–238.

Wang, A., Chang, T.S. and Shi, J. (2021) Multiple event identification and characterization by retrospective analysis of structured data streams. *IISE Transactions*, in-press and online published. https://doi.org/10.1080/24725854.2021.1970863

Wang, A. and Shi, J. (2021) Holistic modeling and analysis of multi-stage manufacturing processes with sparse effective inputs and mixed profile outputs. *IISE Transactions*, **53**(5), 582–596.

Wang, D., Liu, K. and Zhang, X. (2021) A generic indirect deep learning approach for multisensor degradation modeling. *IEEE Transactions on Automation Science and Engineering*, in-press and online published. https://doi.org/10.1109/TASE.2021.3072363

Wang, F., Gahrooei, M.R., Zhong, Z., Tang, T. and Shi, J. (2021) An augmented regression model for tensors with missing values. *IEEE Transactions on Automation Science and Engineering*, in-press and online available. https://doi.org/10.1109/TASE.2021.3102953

Wang, J., Ma, Y., Zhang, L., Gao, R.X. and Wu, D. (2018) Deep learning for smart manufacturing: Methods and applications. *Journal of Manufacturing Systems*, **48**, 144–156.

Wang, Y., Oyen, D., Guo, W.G., Mehta, A., Scott, C.B., Panda, N., Fernández-Godino, M.G., Srinivasan, G. and Yue, X. (2021) StressNet-deep learning to predict stress with fracture propagation in brittle materials. *npj Materials Degradation*, **5**, 6.

Wang, Y., Yue, X., Tuo, R., Hunt, J.H. and Shi, J. (2020) Effective model calibration via sensible variable identification and adjustment with application to composite fuselage simulation. *The Annals of Applied Statistics*, **14**(4), 1759–1776.

Wen, Y., Yue, X., Hunt, J.H. and Shi, J. (2018) Feasibility analysis of composite fuselage shape control via finite element analysis. *Journal of Manufacturing Systems*, **46**, 272–281.

Wen, Y., Yue, X., Hunt, J.H. and Shi, J. (2019) Virtual assembly and residual stress analysis for the composite fuselage assembly process. *Journal of Manufacturing Systems*, **52**, 55–62.

Wold, S., Esbensen, K. and Geladi, P. (1987) Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, **2**(1-3), 37–52.

Wu, C.J. and Hamada, M.S. (2011) *Experiments: Planning, Analysis, and Optimization*, John Wiley & Sons, Hoboken, NJ.

Xanthopoulos, A., Kiatipis, A., Koulouriotis, D.E. and Stieger, S. (2017) Reinforcement learning-based and parametric production-maintenance control policies for a deteriorating manufacturing system. *IEEE Access*, **6**, 576–588.

Yan, H., Grasso, M., Paynabar, K. and Colosimo, B.M. (2021) Real-time detection of clustered events in video-imaging data with applications to additive manufacturing. *IISE Transactions*, **54**(5), 464–480.

Yan, H., Liu, K., Zhang, X. and Shi, J. (2016) Multiple sensor data fusion for degradation modeling and prognostics under multiple operational conditions. *IEEE Transactions on Reliability*, **65**(3), 1416–1426.

Yan, H., Paynabar, K. and Pacella, M. (2010) Structured point cloud data analysis via regularized tensor regression for process modeling and optimization. *Technometrics*, **61**(3), 385–395.

Yan, H., Paynabar, K. and Shi, J. (2014) Image-based process monitoring using low-rank tensor decomposition. *IEEE Transactions on Automation Science and Engineering*, **12**(1), 216–227.

Yan, H., Paynabar, K. and Shi, J. (2017) Anomaly detection in images with smooth background via smooth-sparse decomposition. *Technometrics*, **59**(1), 102–114.

Yan, H., Paynabar, K. and Shi, J. (2018) Real-time monitoring of high-dimensional functional data streams via spatio-temporal smooth sparse decomposition. *Technometrics*, **60**(2), 181–197.

Yan, H., Paynabar, K. and Shi, J. (2020) AKM2D: An adaptive framework for online sensing and anomaly quantification. *IISE Transactions*, **52**(9), 1032–1046.

Yan, H., Sergin, N.D., Brenneman, W.A., Lange, S.J. and Ba, S. (2021) Deep multistage multi-task learning for quality prediction of multi-stage manufacturing systems. *Journal of Quality Technology*, **53**(5), 526–544.

Yan, H., Yeh, H.-M. and Sergin, N. (2019) Image-based process monitoring via adversarial autoencoder with applications to rolling defect detection, in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, pp. 311–316.

Yang, H., Li, W. and Wang, B. (2021) Joint optimization of preventive maintenance and production scheduling for multi-state production systems based on reinforcement learning. *Reliability Engineering & System Safety*, **214**, 107713.

Yue, X., Park, J.G., Liang, Z. and Shi, J. (2020) Tensor mixed effects model with application to nanomanufacturing inspection. *Technometrics*, **62**(1), 116–129.

Yue, X., Wang, K., Yan, H., Park, J.G., Liang, Z., Zhang, C., Wang, B. and Shi, J. (2016) Generalized wavelet shrinkage of inline Raman spectroscopy for quality monitoring of continuous manufacturing of carbon nanotube buckypaper. *IEEE Transactions on Automation Science and Engineering*, **14**(1), 196–207.

Yue, X., Wen, Y., Hunt, J.H. and Shi, J. (2018) Surrogate model-based control considering uncertainties for composite fuselage assembly. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **140**(4), 041017.

Yue, X., Yan, H., Park, J.G., Liang, Z. and Shi, J. (2017) A wavelet-based penalized mixed-effects decomposition for multichannel profile detection of in-line Raman spectroscopy. *IEEE Transactions on Automation Science and Engineering*, **15**(3), 1258–1271.

Zhang, C., Yan, H., Lee, S. and Shi, J. (2018) Weakly correlated profile monitoring based on sparse multi-channel functional principal component analysis. *IISE Transactions*, **50**(10), 878–891.

Zhang, C., Yan, H., Lee, S. and Shi, J. (2021) Dynamic multivariate functional data modeling via sparse subspace learning. *Technometrics*, **63**(3), 370–383.

Zhang, T. and Shi, J. (2016a) Stream of variation modeling and analysis for compliant composite part assembly—Part I: Single-station processes. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **138**(12), 121003.

Zhang, T. and Shi, J. (2016b) Stream of variation modeling and analysis for compliant composite part assembly—Part II: Multistation processes. *ASME Transactions, Journal of Manufacturing Science and Engineering*, **138**(12), 121004,

Zhang, W. and Mei, Y. (2020) Bandit change-point detection for real-time monitoring high-dimensional data under sampling control. *arXiv preprint arXiv:2009.11891*.

Zhong, J., Liu, J. and Shi, J. (2010) Predictive control considering model uncertainty for variation reduction in multistage assembly processes. *IEEE Transactions on Automation Science and Engineering*, **7**(4), 724–735.

Zhong, J., Shi, J. and Wu, J.C. (2009) Design of DOE-based automatic process controller with consideration of model and observation uncertainties. *IEEE Transactions on Automation Science and Engineering*, **7**(2), 266–273.

Zhong, Z., Paynabar, K. and Shi, J. (2022) Image-based feedback control using tensor analysis. *Technometrics,* accepted.

Zhong, Z., Wang, A., Kim, H., Paynabar, K. and Shi, J. (2021) Adaptive cautious regularized run-to-run controller for lithography process. *IEEE Transactions on Semiconductor Manufacturing*, **34**(3), 387–397.

Zhou, S., Ding, Y., Chen, Y. and Shi, J. (2003) Diagnosability study of multistage manufacturing processes based on linear mixed-effects models. *Technometrics*, **45**(4), 312–325.

Zhou, Z.H. (2018) A brief introduction to weakly supervised learning. *National Science Review*, **5**(1), 44–53.

Zhou, Z.H. and Li, M. (2010) Semi-supervised learning by disagreement. *Knowledge and Information Systems*, **24**(3), 415–439.

Zhu, X. (2008) Semi-supervised learning literature survey. Technical Report, University of Wisconsin-Madison Department of Computer Sciences. http://digital.library.wisc.edu/1793/60444