Data Fusion for In-Process Quality Improvement

Jianjun Shi

H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology
Atlanta, Georgia 30332-0205, jianjun.shi@isye.gatech.edu

ABSTRACT

This paper presents the concepts and achievements of data fusion for In-Process Quality Improvement (IPQI) in complex manufacturing systems. As opposed to traditional quality control concepts that emphasize process change detection, acceptance sampling, and off-line designed experiments, IPQI focuses on the integration of engineering domain knowledge with the in-process sensing data to achieve process monitoring, diagnosis, and control. The implementation of IPQI leads to root cause diagnosis (rather than change detection), automatic compensation (rather than off-line adjustment), and defect prevention (rather than defect inspection). The methodologies of IPQI have been developed and implemented in various manufacturing processes. Two topics, the Stream of Variation theory and the causation-based quality control, both of which play an essential role in enabling IPQI, will be discussed with more details in this paper. The prospect for future work, especially on how to address Big Data opportunities in advanced manufacturing processes, is discussed at the end of the paper.

Keywords: In-process quality improvement, manufacturing system, data fusion, big data, engineering-driven statistics

1. Introduction

The life cycle of a product typically goes through seven steps, including consumer needs and market analysis, product design, process design, manufacturing operations, quality control and assurance, product maintenance, and supply chain and logistics (Figure 1). In each of these steps, different types of data with heterogeneous formats are generated and collected. Though efforts have been made to model and analyze these data in each individual step, few researches have been made to use the data integrated from different steps in a unified manner to improve the overall manufacturing system performance. This paper discusses some recent efforts in the data fusion research, especially with a focus on the in-process quality improvements of manufacturing systems.

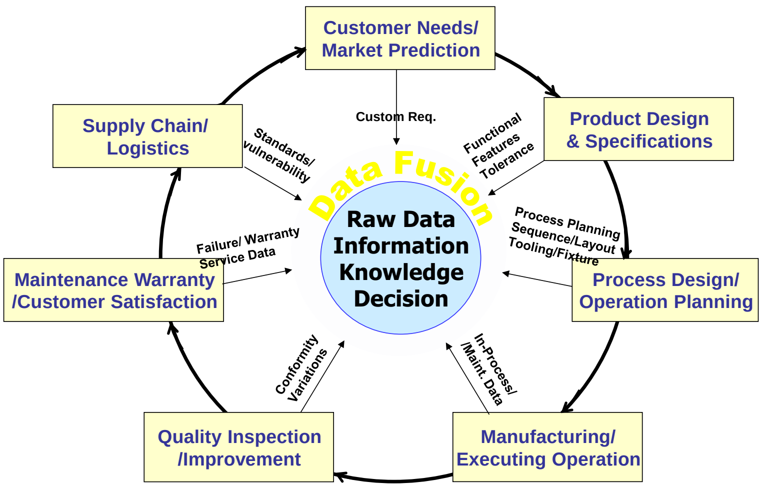

Figure 1. *The life cycle of a product and the needs of "Data Fusion"*

Conventional quality improvement methodologies have four major components: design of experiments (DoE), statistical process control (SPC), acceptance sampling, and quality management (Figure 2). DoE emphasizes how to design a product or a process that will be robust to disturbances in a manufacturing system; SPC mainly focuses on change detections; acceptance sampling makes lot sentencing decisions;

and quality management emphasizes policies and procedures of quality control in the whole organization. However, there are inherent limitations in these existing quality control methodologies. For example, the DoE assumes that the distributions of anticipated disturbances are known, in order to achieve robust design of the product/process; however, this assumption may not be true in practice. Meanwhile, SPC can detect process changes, but relies on the operator's effort to find root causes of the changes. 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 this framework is not designed to interface with the detailed levels that matter in problem solving. Therefore, there is one fundamental issue missing in the existing quality control framework: how to use the multiple in-site sensing signals, integrated with other data and knowledge generated in the life cycle of the product (Figure 1), to achieve in-process quality improvements.

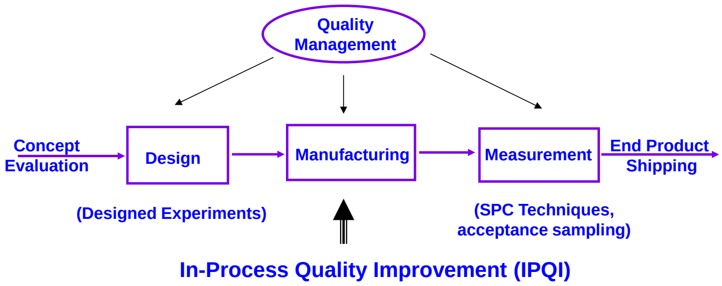

Figure 2. *Conventional Quality Control Methods vs. In-Process Quality Improvement*

In-process quality improvement emphasizes process monitoring, diagnosis, and control through analysis of data and information from all steps in the product realization. These data include product design, process design, in-process sensors, product quality measurement, and maintenance information. Data fusion is achieved by developing advanced statistical methods driven by engineering knowledge, and further enhanced with optimization and control theories (Figure 3). By implementing the IPQI, we can expect that the manufacturing system can achieve automatic root cause diagnosis (rather than change detection), on-line automatic control (rather than off-line adjustment), and defect prevention (rather than defect inspection).

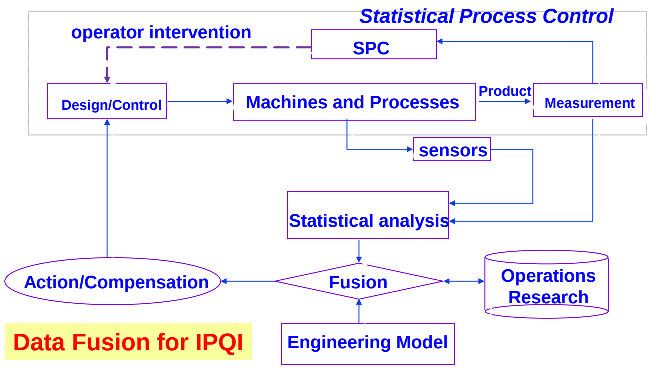

Figure 3. *Comparison of SPC vs. IPQI*

Since the introduction of the concepts of IPQI [1], various efforts have been made to develop methodologies and applications in different manufacturing processes, such as assembly, machining, forming, rolling, and nano-manufacturing scale up. In this paper, we will focus on IPQI for multistage manufacturing processes, and discuss in particular the development of the Stream of Variation theory and the causation-based quality control methods.

2. Selected topics in data fusion for in-process quality improvements

There are quite a few research achievements on the in-process quality improvements. This section is not intended to provide a comprehensive review of all these achievements, but instead tries to highlight some features of engineering-driven data fusion research for IPQI with two examples: Stream of Variation theory and causation-based quality control.

2.1 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. Multistage system is very common in modern manufacturing processes and service systems. In most cases, the final product or service quality of a multistage system is determined by complex interactions among multiple stages - 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.

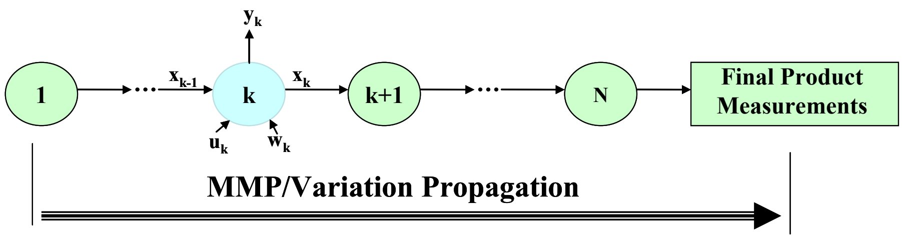

Figure 4. *Variation propagation and notations in the SoV modeling [reproduced from [2]]*

The concept of Stream of Variation (SoV) is 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 4).

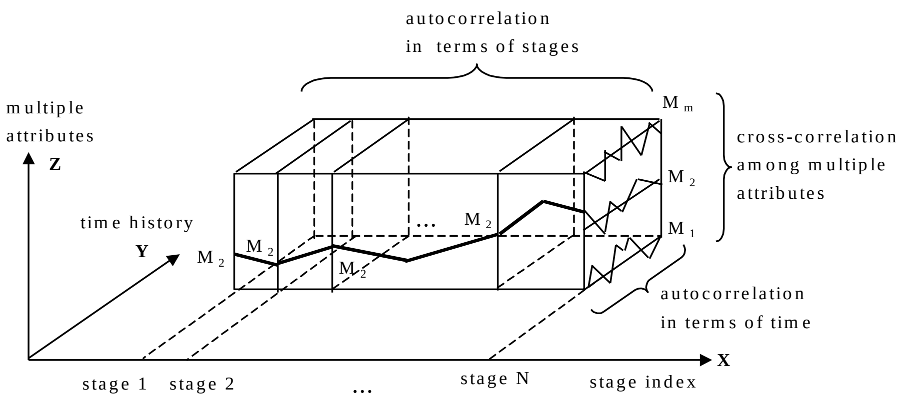

Figure 5. 3D diagram of the MMP illustrating the complex data relationships *[reproduced from [2]]*

An interpretation of the Stream of Variation reflects the complex data relationships in an MMP. As shown in Figure 5, the X axis represents the manufacturing stages; the Y axis represents the time; the Z axis represents the quality attributes; and M_i is the i^{th} quality attributes. 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 and shown as M_2 along the X axis; (*ii*) the quality attributes are cross-correlated among them within the same stage and represented as [M_1, M_2, …, M_m] at stage N 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 and represented as M_i (*i*=1, 2, …m) along the Y axis. These three types of correlations, observed as stream of data, introduce significant challenges in variation modeling, analysis, and control. Stream of Variation (SoV) methodology [2] is 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 that describes the deviation and its propagation in an *N*-station process (as shown in Figure 4), i.e.

$$\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k, \ k = 1, 2, \ldots, N, \quad (1)$$

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k, \ \{k\} \subset \{1, 2, \ldots, N\}. \quad (2)$$

where k = 1, 2, …, N and is the stage index. $\mathbf{x}_k$ is the state vector representing the key quality characteristics of the product (or intermediate work piece) after stage k. $\mathbf{u}_k$ is the control vector representing the tooling errors (e.g. tolerance when no faults occur, or deviation when there are failures occur on the tooling) at stage k. $\mathbf{y}_k$ is the measurement vector representing product quality measurements at stage k. $\mathbf{w}_k$ and $\mathbf{v}_k$ are modeling error and sensing error, respectively. The coefficient matrices $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$ are determined by product and process design information: $\mathbf{A}_k$ represents the impact of deviation transition from stage k-1 to stage k, $\mathbf{B}_k$ represents the impact of the local tooling deviation on the product quality at stage k, and $\mathbf{C}_k$ is the measurement matrix, which can be obtained from the defined quality features of the key product at stage k.

If we repeat the modeling efforts for each stage from k=1 to N, we will get the deviation and its propagation throughout the MMPs. By taking variances on both sides of (1) and (2), and with certain assumptions, we will obtain the "variation" and its propagation model for the MMP.

With the mathematical-based models (1) and (2), variation reduction can be achieved in both design and manufacturing stages through rigorous mathematical 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 are addressed in the SoV methodological research [2]. In more details, the Stream of Variation methodology addresses the following important questions for variation reduction and IPQI in a multistage manufacturing process:

- *How to model and represent the product and process design information for the variation reduction objective?* In [2], two basic methods were investigated, which were physical modeling method and data-driven modeling method. In the former method, the kinematics relationship between KCC and KPC is identified through a detailed physical analysis of the manufacturing process; while in the later method, the model is achieved through a statistical fitting based on historical process measurement data. Details of SoV modeling are discussed in Chapters 6, 7, and 8 of [2].

- *How to systematically find the root causes of variability in terms of which manufacturing station and what in the station introduce the variability?* During a continuous production, a larger product variation may occur at any stage of an MMP due to worn tooling, tooling breakage, and incoming part variation. The SoV monograph [2] 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 the on-line statistics. The variation model, developed from the design information, is used to link the product variation (quality attributes) with the tooling variation (or potential failure). During the production, the product features are measured and the data are used to conduct statistical analysis – based on the model (1) – to identify root causes. Advanced statistics and estimation theories are used in these efforts. The SoV monograph [2] presents two types of diagnostic techniques for root cause identification for the MMPs: variation pattern matching (Chapter 10) and estimation based diagnosis (Chapter 11).

- *How to distribute measurement gauges for an effective process control in an MMP by determining when, where, and what to measure on the final and intermediate work pieces?* One of the major tasks in variation reduction is to design gauging strategies to measure the product features in a manufacturing process. Most of the existing industrial practices focuses on the product coherence inspection (i.e., product-oriented measurements), which is effective for detecting product imperfection, but may not be effective to identify root causes of product variation. The SoV monograph [2] proposes a "process-oriented" measurement concept with a distributed sensing strategy. In this strategy, selected key control characteristics, as well as selected key product characteristics, will be measured in the selected stages for both detecting product defects and identifying their root causes. The diagnosability issue, which is the capability of identifying potential root causes of the process variation for a given measurement strategy, is studied in Chapter 9 of [2]. The related issues on optimal sensor placement and distribution are discussed in Chapter 12 of [2].

- *How to conduct design evaluation and tolerance synthesis 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. With the SoV model defined in (1), the following three tasks can be performed: (*i*) Tolerance analysis by allocating the part tolerance (x_0) and tooling tolerance (u_k) and then predicting the final product tolerance (x_N) by solving the difference equations; (*ii*) Tolerance synthesis by fixing the final product tolerance (x_N) and then assigning the tolerance for individual parts (x_0) and tooling components (u_k) with minimized cost objective functions; and (*iii*) Sensitivity study by identifying the critical parts (x_k) or tooling components (u_k) that have significant impacts on the final production variation through evaluation of the defined sensitivity indices. Details of these topics are discussed in Chapters 13, 14, and 15 in the monograph [2].

- *How to integrate product quality and production tooling reliability for an effective system design and maintenance decision making?* There is a complex, intrigued relationship between product quality and tooling reliability. A degraded (or failed) production tool 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. Thus, a "chain effect" between the product quality (Q) and tooling reliability (R) can be observed and thus noted as "QR Chain" effect. Modeling of the QR Chain is an integrated effort of the SoV model and the Semi-Markov process model. Chapter 16 of [2] discusses modeling of the QR Chain for MMPs. Chapter 17 of [2] investigates the applications of the QR Chain effect in reliability and maintenance decisions.

A comprehensive discussion on the Stream of Variation theory for MMP is summarized in a monograph [2]. In addition, Shi and Zhou [3] provides a survey of emerging methodologies for tackling various issues in multistage systems including modeling, analysis, monitoring, diagnosis, control, inspection, and design optimization.

2.2 Causation-based Quality Control

Due to rapid advancements in sensing and computation technologies, multiple types of sensors have been imbedded in manufacturing systems, on-line automatically collecting massive production information. Though this data-rich environment provides a great opportunity for more effective process control, the data analysis techniques currently used in the manufacturing area are limited. Most of the current quality control researches focus on *correlation* or *association*, which concerns how to reliably and accurately predict some features of a system from other features of that system. However, for effective process control, there is a need to identify the *cause-effect* relationships among variables which go beyond correlation or association. This idea leads to "causation-based quality control", which is built upon observational data, causal modeling and discovery, and causal inference and decision making.

Observational data is referred as the sensing data obtained from a manufacturing system during production, as well as other data generated in the product realization cycle (Figure 1). In general, they are passively observed, as contrasted with 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 conduction 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 a challenging task, and of interest to many researchers. The challenges in causal discovery involve an intricate interplay between assumptions on the data generating process, patterns of associations in the data, and aspects of causal processes that are consistent with the assumptions and can explain the patterns in the data. Though various research efforts have been made to develop generic causal modeling algorithms, most of the implementations and applications focus on the 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 difficulty for causal modeling of the underlying system. It is almost impossible to develop a universal causal modeling method without knowledge of the underlying manufacturing system. In the research, therefore, emphasis is placed on developing a 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 action and the topology of the physical system. The information related to key process and product features is evolving in the system following engineering principles. From product and process design, some engineering domain knowledge exists, which helps identify the key variables and potential causal relationships. Meanwhile, the data captured by in-process sensors records the process changes and interrelationship among variables in practice. By integrating those two sets of information (information flow and data), a causal model can be discovered from the observational data, and further used to develop effective process control strategies. In the past few years, concepts of causation-based quality control have been proposed. Several topics have been investigated, including:

- *Causation-based quality control for rolling processes [4]*: 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 6). 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 amount of surface defects and the process variation is measured by twenty-two variables, collected from two major manufacturing stages, continuous casting (pre-rolling) and rolling. With causal network representation, causal relationships among the variables can be identified both qualitatively and quantitatively. The results can further facilitate diagnosis, prediction, and development of control strategies.

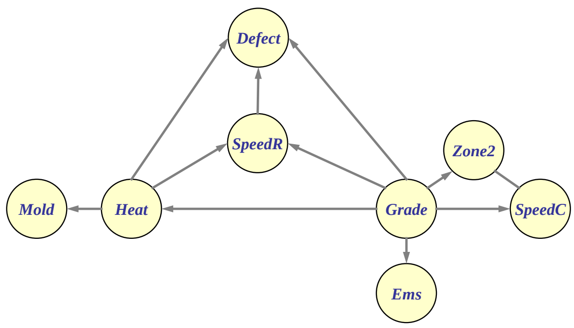

Figure 6. *The causal model developed from rolling process data (Reproduced from [4])*

- *Optimal sensor allocation by integrating causal models and optimization algorithms [5, 6]*: Optimal sensor allocation for system abnormality detection is an important research topic in quality engineering. A new method for optimal sensor allocation is developed in a distributed sensing network with the objective of timely detection of abnormalities in an underlying physical system. This method involves two steps: first, a Bayesian network (BN) is built to represent the causal relationships among the 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 satisfy a prescribed detectability requirement [5]. Recently, further studies were conducted to develop better sensor allocation strategies based on the causal relationships to achieve the optimal performance with the minimum sensor cost [6]. Case studies were performed on a hot forming process and a large-scaled 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 [7]:* Multivariate process monitoring and diagnosis is an important and challenging issue. The widely adopted Hotelling T^2control chart can effectively detect a change in a system but is not capable of diagnosing the root causes of the change. The MTY approach [8] 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 a large dimension of variables. The developed causation-based T^2 decomposition method integrates the causal relationships revealed by a Bayesian network 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.

3. Conclusion and future trend

The rapid advancements of 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 other complexities such as streaming data, spatio-temporal structures, network relationships, and so on. The term “Big Data” is being used to capture and characterize many challenges associated with collecting, managing, analyzing, and visualizing massive amounts of data. The two most promising opportunities for the manufacturing communities are the Advanced Manufacturing Programs and Big Data initiatives. The methodological developments presented in this paper are particularly relevant to the research issues at the interface of Advanced Manufacturing and Big Data.

With massive data readily available throughout the product realization, there is a pressing need to further develop advanced methodologies and associated tools that will enable and assist (i) the handling of the rich data streams communicated by the contemporary complex manufacturing systems, (ii) the extraction of pertinent knowledge about the environmental and operational dynamics driving these systems, and (iii) the exploitation of the acquired knowledge for more enhanced design, analysis, and control. Hence, the complexity of manufacturing systems requires a holistic system-level data fusion approach for effective quality control and performance improvement. Thus, the essence of the manufacturing system data fusion framework is the integration of theories, tools, and techniques from multiple disciplines such as industrial and systems engineering, statistics, mechanical engineering, and electrical engineering to achieve a full utilization of the wide spectrum of readily available information. Indeed, this has been an ever-growing trend in many academic disciplines. The computational power and the data availability have reached at an unprecedented level in recent years thanks to the information revolution. How to exploit these opportunities to establish transformative methodologies for solving engineering problems will be at the center stage of engineering research in the future. The research in manufacturing systems will follow the same trend. Specifically, the following research areas in manufacturing systems will grow significantly:

- First, a complex system is often not tractable by solely first principle based modeling approaches. The data-driven modeling and knowledge discovery for manufacturing systems will become more popular. The data here is broadly defined. It could come from system sensing data, product and process design data, marketing transactions, or the output of a high-fidelity numerical simulation model. Data mining techniques and surrogate modeling techniques will be developed specifically for complex manufacturing systems.

- Second, system design is always a critical issue in product and service realization. It has been promoted for many years to integrate product design with other factors within the product lifecycle. Even a new term, "Design for X" is coined for this effort. Manufacturing systems provide a quantitative framework for integrated product and process design. Some of the existing work in multistage systems [2, 3, 4], e.g., the work on process-oriented tolerancing, or the engineering driven causal discovery, has demonstrated this capability. It can be expected that more research will be conducted along this line. The key challenges in the integrated design such as design space characterization and high dimensional nonlinear optimization problems will be conquered in the near future.

- Third, as mentioned in the previous sections, the concept of multistage systems can be applied to a very broad range of systems, although the existing work mostly focuses on the quality control of multistage discrete manufacturing processes. The success of the multistage system framework in manufacturing processes will certainly stimulate the application of this framework to other systems. For example, monitoring and diagnosis of the abnormalities in the throughput, cycle time, and lead time of a multistage production system will be very promising application areas for the multistage system framework. Most service systems such as healthcare clinics and hospitals and transportation systems are inherently multistage. Quality control and improvement for such systems will definitely benefit from the framework of multistage systems.

In summary, the complexity and readily available massive amounts of data in manufacturing system demands IPQI research to achieve the best performance improvement. In isolation, neither conventional engineering-based modeling nor purely data-based empirical techniques can effectively address the challenges. The effective and seamless integration of physical and analytical models with data-based methodologies is critical for improving the design, analysis, and control of complex systems. In recent years, the term *System Informatics* has been coined to characterize such an integrated framework, and several groups of researchers have already engaged in research activities related to System Informatics. The author believes that this emerging research concentration represents an important future direction, and present numerous research and education opportunities for academic researchers and practitioners in the manufacturing field.

4. Acknowledgements

This work is partially supported by NSF grants: DMI-9624402, CMMI-0927574 and CMMI-1233143.

5. References

[1] Shi, J., 1996, "In-Process Quality Improvement: Concepts and Methodology", NSF report.

[2] Shi, J., *2006, "Stream of Variation Modeling and Analysis for Multistage Manufacturing Processes*", ISBN: 0-8493-2151-4, CRC Press, Taylor & Francis Group, 469pp.

[3] Shi, J. and Zhou, S, 2009, "Quality Control and Improvement for Multistage Systems: A Survey", *IIE Transactions on Quality and Reliability Engineering*, Vol. 41, pp744-753.

[4] Li, J., and Shi, J., 2007, "Knowledge Discovery from Observational Data for Process Control through Causal Bayesian Networks", *IIE Transactions,* Vol. 39, pp681-690.

[5] Li, J., and Jin, J., 2010, "Optimal Sensor Allocation by Integrating Causal Models and Set-Covering Algorithms," *IIE Transactions*, 42(8), 564-576.

[6] Liu, K. and Shi, J., 2013, "Objectives-Oriented Optimal Sensor Allocation Strategy for Process Monitoring and Diagnosis by Multivariate Analysis in a Bayesian Network", *IIE Transactions*, 45.

[7] Li, J., Jin, J., and Shi, J., 2008, "Causation-based T^2 Decomposition for Multivariate Process Monitoring and Diagnosis," *Journal of Quality Technology*, Vol. 40, No. 1, pp. 46-58

[8] Mason, R. L., Tracy, N. D., and Young, J. C. 1995. "Decomposition of T^2 for Multivariate Control Chart Interpretation." *Journal of Quality Technology* 27, pp. 99-108.