Diagnosability Study of Multistage Manufacturing Processes Based on Linear Mixed-Effects Models

Shiyu Zhou
Department of Industrial Engineering,
University of Wisconsin
Madison, Wisconsin 53706

Yu Ding*
Department of Industrial Engineering
Texas A&M University
College Station, TX 77843-3131

Yong Chen
Department of Industrial and Operations Engineering, University of Michigan
Ann Arbor, Michigan 48109

Jianjun Shi
Department of Industrial and Operations Engineering, University of Michigan
Ann Arbor, Michigan 48109

Abstract

Automatic in-process data collection techniques have been widely used in complicated manufacturing processes in recent years. The huge amounts of product measurement data created great opportunity for process monitoring and diagnosis. Given such product quality measurements, this article examines the diagnosability of the process faults in a multistage manufacturing process using a linear mixed-effect model. Fault diagnosability is defined in a general way that does not depend on specific diagnosis algorithms. The concept of a minimal diagnosable class is proposed to expose the "aliasing" structure among process faults in a partially diagnosable system. The algorithms and procedures necessary to obtain the minimal diagnosable class and to evaluate the system-level diagnosability are presented. The methodology, which can be used for any general linear input-output system, is illustrated using a panel assembly process and an engine-head machining process.

Key Words: Diagnosability Analysis, Fault Diagnosis, Multistage Manufacturing Process, Quality Control, Variance Components Analysis

* Corresponding author. Phone: (979)845-5448; Fax: (979)847-9005; Email: yuding@iemail.tamu.edu

1. Introduction

Automatic in-process sensing and data collection techniques have been widely used in complicated manufacturing processes in recent years (Apley and Shi 2001). For example, Optical Coordinate Measuring Machines (OCMM) are built into autobody assembly lines to obtain 100% inspection on product quality characteristics. In-process probes are also installed on machine tools to help assure the dimensional integrity of manufactured workpieces. The data collected by these tools create great opportunity not only for quality assurance and process monitoring, but also for process fault diagnosis of quality-related problems in manufacturing systems.

Statistical process control (SPC) (Montgomery and Woodall 1997 and Woodall and Montgomery 1999 for reviews) is the major technique used in practice for quality and process monitoring. After a process change is detected through SPC techniques, it is critical to determine the appropriate corrective actions toward restoring the manufacturing system to its normal condition. Because product quality is determined by the conditions of process tooling elements (such as cutting tool, fixture, and welding gun) in a manufacturing system, the appropriate corrective action is to fix the malfunctioning tooling elements that are responsible for the defective products. However, SPC methods provide little diagnostic capability – the diagnosis of malfunctioning tooling elements is left to human operators.

Consider the example of a 2-D panel assembly process (Figure 1) that is simplified from an autobody assembly process. In this process, three stations are involved to assemble four parts (marked as 1, 2, 3, 4, respectively, in Figure 1) and inspect the assembly: part 1 and part 2 are assembled at Station I; subassembly "1+2" is assembled with part 3 and part 4 at Station II; and the final assembly with four parts is inspected at Station III for surface finish, joint quality, and dimensional defects. Each part is restrained by a set of fixtures constituting of a 4-way locator, which controls motion in both *x*- and *z*-directions, and a 2-way locator, which controls motion only in the *z*-direction. A subassembly with several parts also needs a 4-way locator and a 2-way locator to completely control its degrees of freedom. The active locating points are marked as P_i, i = 1, ..., 8, in Figure 1.

The positioning accuracy of locators is one of the critical factors in determining the dimensional accuracy of the final assembly. Worn, broken, or improperly installed locators cannot provide desired positioning accuracy and the assembly will have excessive dimensional deviation or variation as a result. The malfunction of tooling elements (locators in this example) is called process fault, which is the root cause of product quality-related problems.

Directly measuring the position of locators during the production is costly, if not impossible. A practical way is to take measurements from the assembly (or subassembly). In this example, five coordinate sensors are installed on all three stations. Each coordinate sensor measures the position of a part feature, such as a corner, in two orthogonal directions (*x* and *z*). The measurement points are marked as $\{M_i, i=1\ldots5\}$ in Figure 1.

Measurements from M_1 to M_5 contain information regarding the accuracy of fixture locators, offering the possibility to diagnose locators' failure (i.e., process fault). However, the diagnosis of failing locators is not obvious since the out-of-control condition of a product feature at a downstream station *k* may be caused by a locator failure at an upstream station *i* ($i<k$). For example, if M_3 triggers an alarm, it could be caused by the failure of P_1 or P_4 on Station II. But it might also be caused by the failure of P_1, P_2, even that of P_3, P_4 on Station I.

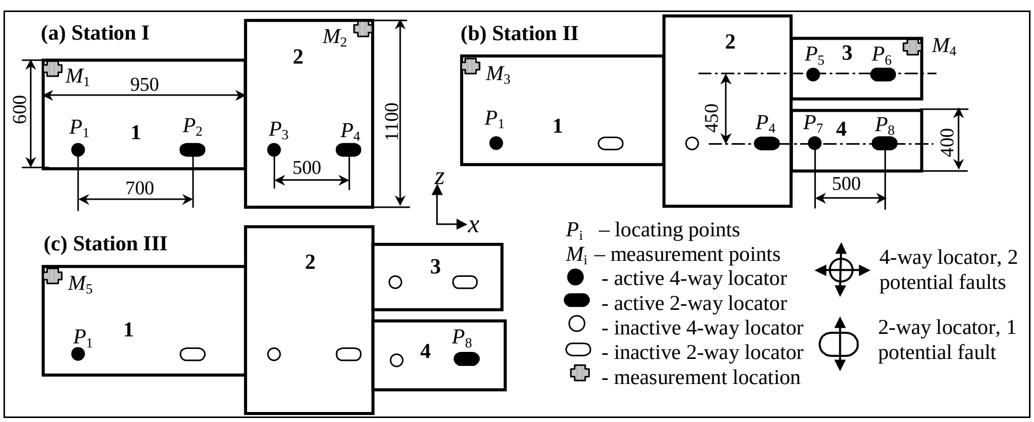


Figure 1. A Multistage 2-D Panel Assembly Process

In many other manufacturing processes, we encounter a similar situation: a tremendous amount of product measurements are available through in-process sensing devices, but the effective utilization of them beyond monitoring remains an interesting yet challenging problem. It is thus highly desirable to have the capability to diagnose process faults from product measurements.

Recent research has advanced toward this goal (Ceglarek and Shi 1996; Apley and Shi 1998; Chang and Gossard 1998; Rong, Ceglarek, and Shi 2000). There are two major components of the reported fault diagnosis methods: (1) a linear model linking product quality measurements to process faults; and (2) algorithms of extracting fault information based on the model. The linear model is often developed for particular processes considering the underlying physical laws. The model-based diagnosis algorithms can be further classified as either multivariate transformation such as the Principal Components Analysis followed by pattern recognition (Ceglarek and Shi 1996; Rong, Ceglarek, and Shi 2000), or least squares estimation followed by a hypothesis test (Apley and Shi 1998; Chang and Gossard 1998).

Limitations of the aforementioned work fall into two categories. First, the models used are developed for single-stage operations, where a manufacturing stage is defined as a group of operations that are conducted under the same workpiece setup. However, modern production systems often involve multiple stages to finish complex products. The fault-quality relationship in a multistage system is not a simple summation of single-stage models. The effect of a certain process fault on product quality could be altered by following operations, and different process faults could have the same manifestation on the final product. As we will see in Section 2, systematic modeling of the fault-quality relationship for multistage manufacturing systems is currently available. Exploring fault diagnosis problems explicitly for multistage systems is feasible and necessary.

Second, diagnosability analysis, which is a fundamental issue regarding fault diagnosis, has not been thoroughly studied. The issue of diagnosability refers to the problem of whether the product measurements contain enough information for the diagnosis of critical process faults, i.e., if process faults are diagnosable. In the abovementioned work, the diagnosability condition is implicitly specified in the pre-conditions required by specific diagnosis algorithms. No explicit discussion on diagnosability under a general framework is given in those papers.

The diagnosability issue is particularly relevant for a multistage system. First, it is challenging to evaluate diagnosability in a multistage system. As in Figure 1, the quality characteristic M_3 at Station II is affected by locators on both Station I and Station II. It is not obvious what kind of information can be obtained regarding those locators when M_3 is measured. Overall, are all process faults diagnosable, given five sensors measuring the current product features? If not, what is the "aliasing" structure among the coupled process faults? Second, even if it is technically feasible, it is not cost effective to install sensors or probes on every intermediate manufacturing stage. Therefore, the quantitative performance evaluation of a gauging system is very important. The proposed diagnosability analysis can provide the underlying analytical tools for this purpose.

Currently there is little reported research on diagnosability. Ding, Shi, and Ceglarek (2002) conducted a preliminary study. The diagnosability condition given in their paper is a special case of the diagnosability analysis presented in this paper. This relationship is clarified in Section 3. Furthermore, their paper does not expose the "aliasing" fault structure for coupled faults in a partially diagnosable system, which is another focus of this paper.

This paper focuses on developing a general framework of diagnosability analysis for the purpose of fault diagnosis in multistage manufacturing systems. We start with a linear state space model that links product quality measurements to process faults in a multistage system. The model can be reformulated into a mixed linear model used in statistical inference. The diagnosis problem is shown to be equivalent to the problem of variance components analysis (VCA). Following the concept of identifiability in VCA, we define diagnosability in a general sense, independent of specific diagnosis algorithms. Diagnosability, and especially partial diagnosability, is studied through the concept of minimal diagnosable class, which is developed to reveal the "aliasing" structure among coupled process faults. Three criteria for performance evaluation of gauging systems are proposed. The criteria benchmark the amount and the "quality" of information obtained through a gauging system, as well as the flexibility of the gauging system.

This paper is structured as follows. In Section 2, the fault-quality diagnostic model is formulated as a mixed linear model. Diagnosability analysis is presented in Section 3, including diagnosability criteria used to evaluate and compare gauging systems. The earlier example is re-visited in Section 4, together with another industrial case study, to illustrate the methodology. Conclusions are presented in Section 5.

2. Formulation of Fault-Quality Diagnostic Model

As mentioned in the previous section, the first step in diagnosability analysis is to develop a fault-quality diagnostic model that links process faults and product quality measurements. There are several linear fault-quality models available to describe the propagation of quality information in a multistage system. Mantripragada and Whitney (1999), Jin and Shi (1999), and Ding, Ceglarek, and Shi (2000) developed multistage fault-quality models for rigid-part assembly processes. Camelio, Hu, and Ceglarek (2001) modeled the variation propagation in multistage compliant-part assembly processes. Zhou, Huang, and Shi (2002) and Djurdjanovic and Ni (2001) provided linear fault-quality diagnostic models for multistage machining processes. All the above models are mechanism models, developed based on the physical laws of the processes. Lawless, Mackay, and Robinson (1999) and Agrawal, Lawless, and Mackay (1999) employed a data-driven AR(1) model to describe the variation transmission in both multistage assembly and machining processes. The parameters of their AR(1) model are estimated based on product measurements. All the aforementioned models adopt the same model structure, which is a linear state space representation. This linear state space model is used in this article to link product quality to individual process faults.

Figure 2 shows a manufacturing process with N stages. Variable k is the stage index. The product quality information (e.g., part dimensional deviations) at each stage is represented by the state vector $\mathbf{x}_k$. The process faults (e.g., the fixturing error, the machining error, and the thermal error) are included as the input $\mathbf{u}_k$. The process faults manifest themselves as the mean deviation and variance of $\mathbf{u}_k$. Natural variation and unmodeled errors in the process are represented by a noise input to the system, $\mathbf{w}_k$. We assume $\mathbf{w}_k$ is a zero mean and uncorrelated random vector. The product quality measurement is denoted by $\mathbf{y}_k$, where $\mathbf{y}_k$ is not necessarily available at every stage. The measurement noise is denoted by a zero mean and uncorrelated random vector $\mathbf{v}_k$.

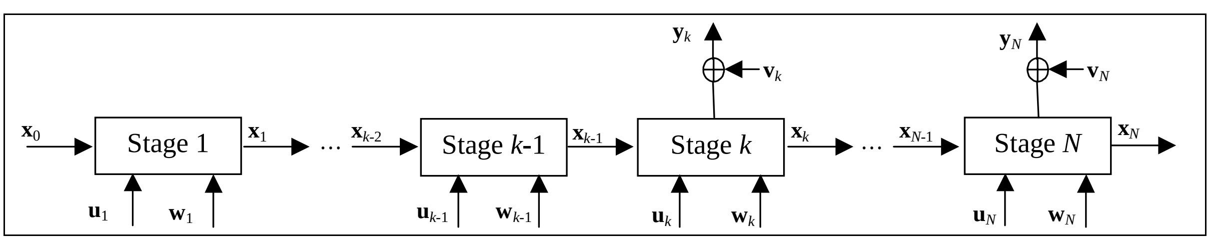


Figure 2. Diagram of a Multistage Manufacturing Process

Under the small error assumption, the linear state space model can be expressed as

$$\mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k \quad \text{and} \quad \mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k, \tag{1}$$

where $k = 1,2...N$, $\mathbf{A}_{k-1}\mathbf{x}_{k-1}$ represents the transformation of quality information from stage k-1 to stage k, $\mathbf{B}_k\mathbf{u}_k$ represents how the product quality is affected by the process faults at stage k, and $\mathbf{C}_k$ is the observation matrix that maps process states to measurements. System matrices $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$ are constant matrices. They are determined by the process/product design information.

The state space model can be transformed into a general mixed linear model as follows. First, it can be written in an input-output format as

$$\mathbf{y}_k = \Sigma_{i=1}^{k}\mathbf{C}_k\mathbf{\Phi}_{k,i}\mathbf{B}_i\mathbf{u}_i + \mathbf{C}_k\mathbf{\Phi}_{k,0}\mathbf{x}_0 + \Sigma_{i=1}^{k}\mathbf{C}_k\mathbf{\Phi}_{k,i}\mathbf{w}_i + \mathbf{v}_k \;, \tag{2}$$

where $\mathbf{\Phi}_{k,i} = \mathbf{A}_{k-1}\mathbf{A}_{k-2}\cdots\mathbf{A}_i$ for $k > i$ and $\mathbf{\Phi}_{k,k} = \mathbf{I}$. The quality characteristics $\mathbf{x}_0$ correspond to the initial condition of the product before it goes into the manufacturing line. If the measurement of $\mathbf{x}_0$ is available, $\mathbf{C}_k\mathbf{\Phi}_{k,0}\mathbf{x}_0$ can be moved to the left side of Equation (2), and the difference $\mathbf{y}_k - \mathbf{C}_k\mathbf{\Phi}_{k,0}\mathbf{x}_0$ can then be treated as a new measurement. If the measurement of $\mathbf{x}_0$ is not available, we can treat it as an additional process fault input. Without loss of generality, we set $\mathbf{x}_0$ to $\mathbf{0}$.

Define $\boldsymbol{\mu}_i$ as the expectation of $\mathbf{u}_i$ and $\tilde{\mathbf{u}}_i = \mathbf{u}_i - \boldsymbol{\mu}_i$. Combining all available measurements from station 1 to station N, we have

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \mathbf{\Gamma}\cdot\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{\mu}_N \end{bmatrix} + \mathbf{\Gamma}\cdot\begin{bmatrix} \tilde{\mathbf{u}}_1 \\ \tilde{\mathbf{u}}_2 \\ \vdots \\ \tilde{\mathbf{u}}_N \end{bmatrix} + \mathbf{\Psi}\cdot\begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_N \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_N \end{bmatrix}, \tag{3}$$

where

$$\mathbf{\Gamma} = \begin{bmatrix} \mathbf{C}_1\mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}_2\mathbf{\Phi}_{2,1}\mathbf{B}_1 & \mathbf{C}_2\mathbf{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_N\mathbf{\Phi}_{N,1}\mathbf{B}_1 & \mathbf{C}_N\mathbf{\Phi}_{N,2}\mathbf{B}_2 & \cdots & \mathbf{C}_N\mathbf{B}_N \end{bmatrix}, \quad \mathbf{\Psi} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{C}_2\mathbf{\Phi}_{2,1} & \mathbf{C}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{C}_N\mathbf{\Phi}_{N,1} & \mathbf{C}_N\mathbf{\Phi}_{N,2} & \cdots & \mathbf{C}_N \end{bmatrix},$$

$\boldsymbol{\mu}_i$ is an unknown constant vector, and $\tilde{\mathbf{u}}_i$, $\mathbf{w}_i$, $\mathbf{v}_k$ are uncorrelated random vectors with zero means. If no measurement is available at station k, the corresponding rows can be eliminated.

Let P denote the total number of potential faults ($[\boldsymbol{\mu}_1^T \;\; ... \;\; \boldsymbol{\mu}_k^T \;\; ... \;\; \boldsymbol{\mu}_N^T]^T$) and Q denote the number of system noises ($[\mathbf{w}_1^T \;\; ... \;\; \mathbf{w}_k^T \;\; ... \;\; \mathbf{w}_N^T]^T$) considered on all the stages. Let $\{\sigma_{u_i}^2\}_{i=1...P}$,

$\{\sigma^2_{w_i}\}_{i=1...Q}$, and σ^2_v be the variances of process faults, the variances of system noises, and the variance of observation noise, respectively. We assume that the variances of observation noise at different stages are the same. This assumption is reasonable if we use the same measurement devices on all measurement stages.

During production, multiple samples of the product are available at each stage. Assume we have M samples, and the samples can be stacked up as

$$\mathbf{Y} = (\mathbf{1}_M \otimes \mathbf{\Gamma})\mathbf{U} + (\mathbf{I}_M \otimes \mathbf{\Gamma})\tilde{\mathbf{U}} + (\mathbf{I}_M \otimes \mathbf{\Psi})\mathbf{W} + \mathbf{V}, \tag{4}$$

where $\mathbf{U}^T = [\boldsymbol{\mu}_1^T \quad ... \quad \boldsymbol{\mu}_k^T \quad ... \quad \boldsymbol{\mu}_N^T]$, $\mathbf{Y}^T = [\mathbf{Y}_1^T \quad \cdots \quad \mathbf{Y}_i^T \quad \cdots \quad \mathbf{Y}_M^T]$, $\mathbf{Y}_i^T = \left[\mathbf{y}_{1i}^T \quad ... \quad \mathbf{y}_{ki}^T \quad ... \quad \mathbf{y}_{Ni}^T\right]$ is the i^{th} sample measurement, and $\mathbf{y}_{ki}$ is the ith sample measurement at the kth stage. $\tilde{\mathbf{U}}$, $\mathbf{W}$, and $\mathbf{V}$ are defined in the similar way as $\mathbf{Y}$, $\otimes$ is the Kronecker matrix product (Scott 1997), $\mathbf{1}_M$ is the summing vector whose M elements equal unity, and $\mathbf{I}_M$ is an M by M identity matrix. Letting ":j" represent the j^{th} column of a matrix, Equation (4) can be re-organized as

$$\mathbf{Y} = (\mathbf{1}_M \otimes \mathbf{\Gamma})\mathbf{U} + \Sigma_{j=1}^{P}(\mathbf{I}_M \otimes \mathbf{\Gamma}_{:j})\tilde{\mathbf{U}}(j) + \Sigma_{j=1}^{Q}(\mathbf{I}_M \otimes \mathbf{\Psi}_{:j})\mathbf{W}(j) + \mathbf{V} \; . \tag{5}$$

where $\tilde{\mathbf{U}}(j) = [\tilde{u}_{j1} \quad \cdots \quad \tilde{u}_{ji} \quad \cdots \quad \tilde{u}_{jM}]^T$ and $\mathbf{W}(j) = [w_{j1} \quad \cdots \quad w_{ji} \quad \cdots \quad w_{jM}]^T$ are the collections of all samples of the jth fault and the jth system noise, respectively.

The diagnosability problem can then be restated: from M samples, can we identify the value of $\{\mu_i\}_{i=1...P}$ and $\{\sigma^2_{u_i}\}_{i=1..P}$? In the following section, this problem is studied using the framework of variance components analysis.

3. Diagnosability Analysis for Multistage Manufacturing Processes

3.1 Definition of Fault Diagnosability

The model in Equation (5) fits a general mixed linear model given by Rao and Kleffe (1988) as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \sum_{i=1}^{c}\boldsymbol{\xi}_i\mathbf{b}_i + \mathbf{e}, \tag{6}$$

where $\mathbf{y}$ is an $n_y \times 1$ observation vector; $\mathbf{X}$ is an $n_y \times l_x$ known constant matrix, $l_x \leq n_y$; $\boldsymbol{\alpha}$ is an $l_x \times 1$ vector of unknown constants; $\boldsymbol{\xi}_i$ is an $n_y \times m_i$ known constant matrix, $m_i \leq n_y$; $\mathbf{b}_i$ is an $m_i \times 1$ vector of independent variables with zero mean and unknown variance σ_i^2; $\mathbf{e}$ is an $n_y \times 1$ vector of independent variables with zero mean and unknown variance σ_e^2. The σ_i^2's and σ_e^2 are called "variance components."

A mixed model is used to describe both fixed and random effects. This model is often applied to biological and agricultural data. In designed experiments, the matrices **X** and $\{\boldsymbol{\xi}_i\}_{i=1\ldots c}$ are determined by designers. They often contain only 0s or 1s, depending upon whether the relevant effect contributes to the measurement. Given a mixed model, researchers are primarily interested in estimating the fixed effects and variance components. A large body of literature about VCA is available. Excellent overviews can be found in Rao and Kleffe (1988) and Searle, Casella, and McCulloch (1992).

We can establish a one-to-one corresponding relationship between terms in our fault-quality model (Equation (5)) and those in the mixed model (Equation (6)). In our fault diagnosis problem, however, the matrices **X**, $\{\boldsymbol{\xi}_i\}_{i=1\ldots c}$ are computed from system matrices $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$, k=1,…,N, which are determined by the process design information and measurement deployment information. The fixed effects are the mean values ($\boldsymbol{\mu}_i$) of process faults, and the random effects are the process faults and the process noise, $\tilde{\mathbf{u}}_i, \mathbf{w}_i, \mathbf{v}_k$. Fault diagnosis is thus equivalent to the problem of variance components estimation. The definition of *diagnosability* in this article follows the same concept of *identifiability* in VCA (Rao and Kleffe, 1988). The term "diagnosability" is used because it is more relevant in the context of our engineering applications.

Based on Equation (5), we have

$$E(\mathbf{Y}) = [\boldsymbol{\Gamma}^T \quad \ldots \quad \boldsymbol{\Gamma}^T \quad \ldots \quad \boldsymbol{\Gamma}^T]^T \, \mathbf{U} \tag{7}$$

$$Cov(\mathbf{Y}) = \mathbf{F}_1\sigma_{u_1}^2 + \ldots + \mathbf{F}_P\sigma_{u_P}^2 + \mathbf{F}_{P+1}\sigma_{w_1}^2 + \ldots + \mathbf{F}_{P+Q}\sigma_{w_Q}^2 + \mathbf{F}_{P+Q+1}\sigma_v^2 \tag{8}$$

where $E(\cdot)$ represents the expectation, $Cov(\cdot)$ represents the covariance matrix of a random vector,

$$\mathbf{F}_i = \begin{cases} \mathbf{I}_M \otimes (\boldsymbol{\Gamma}_{:i}\boldsymbol{\Gamma}_{:i}^T) & \text{when } 1 \le i \le P \\ \mathbf{I}_M \otimes (\boldsymbol{\Psi}_{:(i-P)}\boldsymbol{\Psi}_{:(i-P)}^T) & \text{when } P < i \le P+Q \end{cases},$$

and $\mathbf{F}_{P+Q+1}$ is an identity matrix with the appropriate dimension.

Define $[\sigma_{u_1}^2 \quad \ldots \quad \sigma_{u_P}^2 \quad \sigma_{w_1}^2 \quad \ldots \quad \sigma_{w_Q}^2 \quad \sigma_v^2]^T$ in Equation (8) as $\boldsymbol{\theta}$, E^U as the space containing all possible values of **U**, and E^S as the space containing all possible values of $\boldsymbol{\theta}$ (in the most general case, E^U is $\Re^{P\times 1}$ and E^S is a $(P+Q+1)\times 1$ space spanned by nonnegative real numbers).

Diagnosability is defined following the definition of *identifiability* in Rao and Kleffe (1988).

Definition 1 In model (5), a linear parametric function $\mathbf{p}^T\boldsymbol{\alpha}$, $\mathbf{p}\in\Re^{P\times 1}$, $\boldsymbol{\alpha}\in E^U$ is said to be diagnosable if, $\forall\,\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in E^U$,

$$\mathbf{p}^T\boldsymbol{\alpha}_1\neq\mathbf{p}^T\boldsymbol{\alpha}_2 \Rightarrow E(\mathbf{Y})|_{\mathbf{U}=\boldsymbol{\alpha}_1}\neq E(\mathbf{Y})|_{\mathbf{U}=\boldsymbol{\alpha}_2}. \quad (9)$$

A linear parametric function $\mathbf{f}^{\mathrm{T}}\boldsymbol{\theta}$, $\mathbf{f}\in\Re^{(P+Q+1)\times 1}$, $\boldsymbol{\theta}\in E^S$ is said to be diagnosable if, $\forall\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\in E^S$,

$$\mathbf{f}^T\boldsymbol{\theta}_1\neq\mathbf{f}^T\boldsymbol{\theta}_2 \Rightarrow Cov(\mathbf{Y})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}\neq Cov(\mathbf{Y})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_2}. \quad (10)$$

Remarks:

1. In model (5), we are only concerned about the mean and variance of process faults. Therefore, only the first and second order moments are considered in the definition.
2. The above definition means that a fault combination is called diagnosable if the change in the combined mean or variance causes a change in the mean or variance of observation **Y**. This definition does not depend on any specific diagnosis algorithm.
3. By selecting different **p** and **f**, the diagnosability of different fault combinations can be evaluated. For example, by selecting $\mathbf{p}$ or $\mathbf{f}=[1\quad 0\quad \ldots\quad 0]^T$, we can check if the mean or variance of the 1st fault is diagnosable. If yes, we say the mean or variance of this fault can be *uniquely* identified or diagnosed.

3.2 Criterion of Fault Diagnosability and Minimal Diagnosable Class

The necessary and sufficient condition of fault diagnosability in a linear system is given by Theorem 1. The proof can be found in Appendix A2.

Theorem 1 Define the range space of a matrix as $R(\cdot)$, and $\mathbf{D}=[\boldsymbol{\Gamma}\quad\boldsymbol{\Psi}]$. In model (5),

(i) $\mathbf{p}^{\mathrm{T}}\boldsymbol{\alpha}$ is diagnosable if and only if $\mathbf{p}\in R(\boldsymbol{\Gamma}^T)$.

(ii) $\mathbf{f}^{\mathrm{T}}\boldsymbol{\theta}$ is diagnosable if and only if $\mathbf{f}\in R(\mathbf{H})$, where **H** is symmetric and given as

$$\mathbf{H}=\begin{bmatrix} (\mathbf{D}_{:1}^T\mathbf{D}_{:1})^2 & \ldots & (\mathbf{D}_{:1}^T\mathbf{D}_{:i})^2 & \ldots & (\mathbf{D}_{:1}^T\mathbf{D}_{:(P+Q)})^2 & \mathbf{D}_{:1}^T\mathbf{D}_{:1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (\mathbf{D}_{:i}^T\mathbf{D}_{:1})^2 & \ldots & (\mathbf{D}_{:i}^T\mathbf{D}_{:i})^2 & \ldots & (\mathbf{D}_{:i}^T\mathbf{D}_{:(P+Q)})^2 & \mathbf{D}_{:i}^T\mathbf{D}_{:i} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (\mathbf{D}_{:(P+Q)}^T\mathbf{D}_{:1})^2 & \ldots & (\mathbf{D}_{:(P+Q)}^T\mathbf{D}_{:i})^2 & \ldots & (\mathbf{D}_{:(P+Q)}^T\mathbf{D}_{:(P+Q)})^2 & \mathbf{D}_{:(P+Q)}^T\mathbf{D}_{:(P+Q)} \\ \mathbf{D}_{:1}^T\mathbf{D}_{:1} & \ldots & \mathbf{D}_{:i}^T\mathbf{D}_{:i} & \ldots & \mathbf{D}_{:(P+Q)}^T\mathbf{D}_{:(P+Q)} & L \end{bmatrix} \quad (11)$$

where L is the length of $[\mathbf{y}_1^T \quad \mathbf{y}_2^T \quad \cdots \quad \mathbf{y}_N^T]^T$ in Equation (3).

Theorem 1 gives us a powerful tool to test if some combinations of faults are diagnosable. From Theorem 1, it is clear that the means of all the faults are uniquely diagnosable if and only if $\mathbf{\Gamma}^T\mathbf{\Gamma}$ is of full rank. The variances of all the faults are uniquely diagnosable if and only if **H** is of full rank.

For the above criterion, the diagnosability of the variance of process fault includes the effects of the modeling error **w** and the observation noise **v**. This means that even if a fault can be distinguished from other faults, it could still be non-uniquely diagnosable if it is tangled with the modeling error or the observation noise. In some cases, if the modeling error and the observation noise can be assumed small or their variance can be estimated from the normal working condition of a manufacturing process, we can ignore their effects when exploring the diagnosability of process faults. The testing matrix is revised accordingly by reducing $\boldsymbol{\theta}$ to include only $[\sigma_{u_1}^2 \quad \cdots \quad \sigma_{u_P}^2]$ and reducing the **H** matrix in Theorem 1 to $\mathbf{H}_r$, where $\mathbf{H}_r$ is a sub-block of **H**, i.e.,

$$\mathbf{H}_r = \begin{bmatrix} (\mathbf{\Gamma}_{:1}^T\mathbf{\Gamma}_{:1})^2 & \dots & (\mathbf{\Gamma}_{:1}^T\mathbf{\Gamma}_{:i})^2 & \dots & (\mathbf{\Gamma}_{:1}^T\mathbf{\Gamma}_{:P})^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (\mathbf{\Gamma}_{:i}^T\mathbf{\Gamma}_{:1})^2 & \dots & (\mathbf{\Gamma}_{:i}^T\mathbf{\Gamma}_{:i})^2 & \dots & (\mathbf{\Gamma}_{:i}^T\mathbf{\Gamma}_{:P})^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ (\mathbf{\Gamma}_{:P}^T\mathbf{\Gamma}_{:1})^2 & \dots & (\mathbf{\Gamma}_{:P}^T\mathbf{\Gamma}_{:i})^2 & \dots & (\mathbf{\Gamma}_{:P}^T\mathbf{\Gamma}_{:P})^2 \end{bmatrix}. \tag{12}$$

Remarks:

1. Under the setting that noises **w** and **v** are assumed negligible, the diagnosability matrix was defined in Ding, Shi, and Ceglarek (2002) as $\pi(\mathbf{\Gamma})$, where $\pi(\cdot)$ is a matrix transformation defined in their paper. The variances of process faults are considered fully diagnosable if and only if $\pi(\mathbf{\Gamma})^{\mathrm{T}}\pi(\mathbf{\Gamma})$ is of full rank. In fact, this condition is the same as what we derived here. It can be shown that $R(\mathbf{H}_r)=R(\pi(\mathbf{\Gamma})^{\mathrm{T}}\pi(\mathbf{\Gamma}))$. Therefore, their work can be considered as a special case of the general framework presented in this paper.
2. If noise terms are not included, Ding, Shi, and Ceglarek (2002) showed that the mean being diagnosable is a sufficient condition for variance being diagnosable. However, the converse is not true. This is illustrated in the case study of machining process in Section 4.

Theorem 1 alone is not very effective in analyzing a partially diagnosable system where not all faults are diagnosable. It is not obvious from Theorem 1 what are the other faults that we need to know before we can identify a non-uniquely diagnosable fault. To analyze the partial diagnosable system, we propose the concept of a *minimal diagnosable class*. We first introduce the concept of the *diagnosable class*, and then present the definition of the minimal diagnosable class.

Definition 2 A nonempty set of n faults $\{u_{i_1} \quad \ldots \quad u_{i_n}\}$ forms a *mean* or *variance diagnosable class* if a nontrivial linear combination of their means $\{\mu_{i_1} \quad \ldots \quad \mu_{i_n}\}$ or variances $\{\sigma^2_{i_1} \quad \ldots \quad \sigma^2_{i_n}\}$ are diagnosable. "Nontrivial" means at least one coefficient of the linear combination is nonzero.

Definition 3 A nonempty set of n faults $\{u_{i_1} \quad \ldots \quad u_{i_n}\}$ forms a *minimal mean* or *variance diagnosable class* if no strict subset of $\{u_{i_1} \quad \ldots \quad u_{i_n}\}$ is mean or variance diagnosable.

The diagnosability of the mean and the variance can be dealt with separately, and the testing procedures are very similar (the only difference is the testing matrix; it is $\mathbf{\Gamma}^{\mathrm{T}}$ for mean and $\mathbf{H}$ for variance). Hence, no distinction between mean or variance diagnosability will be made hereafter unless otherwise indicated.

The minimal diagnosable classes expose the inter-relationship between different faults. Intuitively, a minimal diagnosable class represents a set of faults that are closely coupled together. We can only identify a linear combination of them, but we cannot identify any strict subset. With this information, we can show the coupling relationship among faults and learn what additional information is needed to identify certain fault.

We found that the minimal diagnosable class can be generated from the Reduced Row Echelon Form (RREF) (Lay 1997) of the transpose of testing matrices. This result is stated in the following theorem. The proof of this theorem is given in Appendix A3.

Theorem 2 Given a testing matrix $\mathbf{G} \in \Re^{n \times m}$ ($\mathbf{G}$ is $\mathbf{\Gamma}^T$ or $\mathbf{H}$) and n faults $\boldsymbol{\theta} = [u_1 \quad \ldots \quad u_n]^T$ corresponding to $\mathbf{G}$, the fault set $\boldsymbol{\theta}[\mathbf{v}]$ is a minimal diagnosable class if $\mathbf{v}$ is a nonzero row of the RREF of $\mathbf{G}^T$, where $\boldsymbol{\theta}[\mathbf{v}]$ is a subset of $\boldsymbol{\theta}$ such that $\boldsymbol{\theta}(i)$ (the i^{th} element of $\boldsymbol{\theta}$) $\in \boldsymbol{\theta}[\mathbf{v}]$ if $\mathbf{v}(i)$ (the i^{th} element of $\mathbf{v}$) $\neq 0$.

When the RREF of $\mathbf{G}^T$ is calculated, we can obtain some of the minimal diagnosable classes.

The following corollary shows that by re-arranging the columns in $\mathbf{G}^T$, we can obtain all the possible minimal diagnosable classes (the proof is listed in Appendix A4). The re-arranging process is known as matrix permutation. The permuted matrix is defined as: if $\{\mathbf{c}_i\}_{i=1\ldots n}$ denote the column vectors of $\mathbf{G}^T$ and correspond to the faults $\boldsymbol{\theta} = [u_1 \quad \ldots \quad u_n]^T$, the column-wise permuted matrix $\mathbf{G}'^T = [\mathbf{c}_{i_1} \quad \ldots \quad \mathbf{c}_{i_n}]$ is called the *permuted matrix* corresponding to the fault permutation $\boldsymbol{\theta}' = [u_{i_1} \quad \ldots \quad u_{i_n}]^T$.

Corollary 1 Given a testing matrix $\mathbf{G} \in \Re^{n\times m}$ and if $\boldsymbol{\Theta} = \{u_{i_1}, \ldots, u_{i_s}\}$ is a minimal diagnosable class, then $\boldsymbol{\theta}[\mathbf{v}] = \boldsymbol{\Theta}$, where $\mathbf{v}$ is the last nonzero row of $\mathbf{G}_r'^T$. $\mathbf{G}_r'^T$ is the RREF of the permuted matrix of $\mathbf{G}^T$ corresponding to the fault permutation $u_{i_s+1}\ldots u_{i_n} u_{i_1}\ldots u_{i_s}$.

Corollary 1 tells that a complete list of minimal diagnosable classes can be obtained by thoroughly permuting $\mathbf{G}^T$. However, the number of permutations will explode if the number of faults is large. To handle this problem, we need the concept of the "connected fault class."

Given the RREF of $\mathbf{G}^T$, assume we can divide its nonzero rows into two sets of rows (C_1 and C_2) such that for any $\mathbf{v}_i \in C_1$ and $\mathbf{v}_j \in C_2$, $\mathbf{v}_i * \mathbf{v}_j = \mathbf{0}$, where * is the Hadamard product (Scott 1997). In other words, $\mathbf{v}_i$ does not share any common nonzero column positions with $\mathbf{v}_j$. Define symbol $\boldsymbol{\theta}[C]$ as the fault set of $\bigcup_k(\boldsymbol{\theta}[\mathbf{v}_k])$ for all $\mathbf{v}_k \in C$, where C is a set of rows. We can show that for an arbitrary minimal diagnosable class $\boldsymbol{\theta}[\mathbf{v}]$, either $\boldsymbol{\theta}[\mathbf{v}] \subseteq \boldsymbol{\theta}[C_1]$ or $\boldsymbol{\theta}[\mathbf{v}] \subseteq \boldsymbol{\theta}[C_2]$. From Theorem 1, $\mathbf{v}$ is in the space spanned by the rows of $\mathbf{G}^T$. Thus, $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2$, where $\mathbf{v}_1$ and $\mathbf{v}_2$ are in the space spanned by the rows in C_1 and C_2, respectively. However, if a_1 and a_2 are both nonzero, the facts that $\mathbf{v}_i * \mathbf{v}_j = \mathbf{0}$ and $\boldsymbol{\theta}[\mathbf{v}_1]$, $\boldsymbol{\theta}[\mathbf{v}_2]$ are both diagnosable will lead to the contradiction that $\boldsymbol{\theta}[\mathbf{v}]$ is not minimal. The implication is that the complete list of minimal diagnosable classes can be obtained by only permuting the faults within $\boldsymbol{\theta}[C_1]$ and $\boldsymbol{\theta}[C_2]$, respectively. Following the same rule, C_1 and C_2 can be further divided into smaller groups iteratively until they are no longer dividable. If C_i is an un-dividable set of rows, $\boldsymbol{\theta}[C_i]$ is called a "connected fault class." Following a similar argument, we know that the complete list of minimal fault classes can be obtained through permutations only within each connected fault class.

If there are many small connected fault classes in the system, the computational load required to find all minimal diagnosable classes can be significantly reduced. The worst case is that all faults are connected in a big fault class. However, that is usually not the situation in practice. For instance, one principle in manufacturing process design is to reduce the accumulation and propagation chain of process faults (Halevi and Weill, 1995). For many actual engineering systems, the entire fault set can often be partitioned into much smaller connected fault subsets, as we will see in the case studies in Section 4.

In summary, the algorithm obtaining all the minimal diagnosable classes is given as:

(a) Calculate the RREF of $\mathbf{G}^T$,

(b) Remove all the uniquely identifiable faults because each of them will form a minimal diagnosable class; remove the faults corresponding to zero columns because they are invisible to the measurement system and hence not diagnosable, and will not appear in any minimal diagnosable classes.

(c) Find the connected fault classes based on the RREF of $\mathbf{G}^T$.

(d) Permute the columns within the connected fault classes and obtain the minimal diagnosable classes based on the permuted matrices until all the possible permutations are visited.

The minimal diagnosable classes expose the "aliasing" structure among the faults in the system, revealing critical fault diagnosability information. For example, if a single fault forms a minimal diagnosable class, it is uniquely diagnosable. If a fault is not uniquely diagnosable and it forms a minimal diagnosable class with several other faults, this fault can be identified when all other faults are known. Thus, by looking at the minimal diagnosable classes, we can identify which fault can be identified from the measurements, and if not, what other faults need to be known to identify it.

Minimal diagnosable classes can be used to evaluate the performance of different gauging systems in terms of the diagnosability of the process faults. Consider the panel assembly process in Figure 1 as an example. Another gauging system implemented in this system is shown in Figure 3. Counting potential locator errors on all stations, we have a total of n=18 potential faults, which are assigned a serial number from 1 to 18 as shown in Figure 3. The difference between the two gauging systems in Figure 1 and in Figure 3 is the position of M_5. The problem of how to compare these two systems in terms of the diagnosability of all 18 potential faults is addressed in the next subsection.

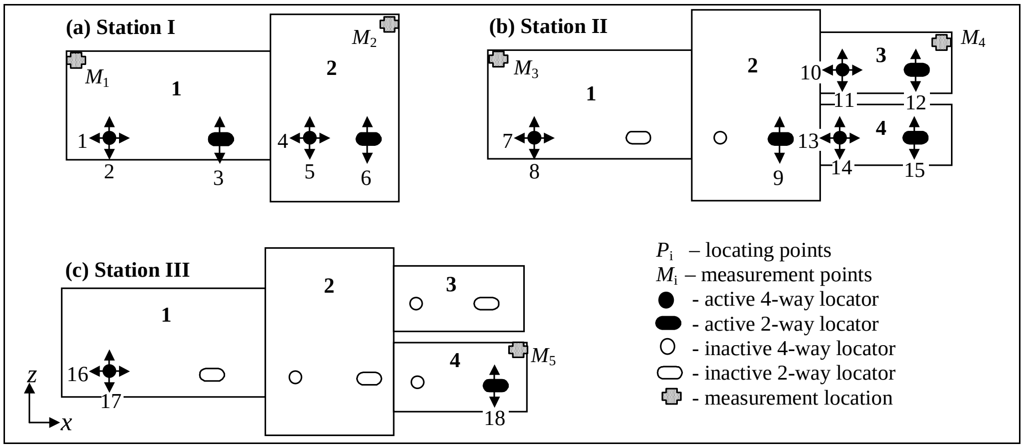


Figure 3. Gauging System 2 for the Multistage 2-D Panel Assembly Process

3.3 Gauging System Evaluation Based on Minimal Diagnosable Class

To evaluate a gauging system, we may need several easy-to-interpret indices to characterize the information obtained through the gauging system. We propose three criteria for evaluation of gauging systems: information *quantity*, information *quality*, and system *flexibility*.

The information quantity refers to the degree to which we know about process faults from the measurement data. When two gauging systems are used for the same manufacturing system, the number of potential faults is the same. However, for two different partially diagnosable systems, the number of faults that we need to know to ensure full diagnosability will often be different. This number can be used to quantify the amount of information obtained by different gauging systems. The following corollary indicates that the rank of the diagnosability testing matrix should be used to quantify the amount of measurement information.

Corollary 2 Given a testing matrix $\mathbf{G} \in \Re^{n \times m}$ and n faults $\boldsymbol{\theta} = [u_1 \quad \ldots \quad u_n]^T$ corresponding to **G**, if the rank of **G** is ρ, then *n*-ρ faults need to be known in order to uniquely identify all *n* faults.

The proof is omitted here. It will use the property of the RREF of a matrix. An intuitive understanding of this corollary is given as follows. The solvability condition of a linear system $\mathbf{Y} = \mathbf{AX}$ can be determined by analyzing the RREF of **A**. In such a linear system, *n*-ρ free variables need to be known before uniquely solving **X**, where n is the dimension of **X**, ρ=rank(**A**). If we consider the testing matrix **G** as if it is in a similar situation to matrix **A**, the result of Corollary 2 is not surprising.

The second criterion is information quality. Even if two gauging systems provide the same amount of information per the criterion developed above, the detailed information contents could be very different. In practice, it is always desirable to have unique identification of a fault so that a corrective action can be undertaken right away to eliminate the fault and to restore the system to its normal condition. The decision of corrective actions cannot be made for a fault coupled with others without further investigation or measurement. Thus, we use the number of uniquely identifiable faults to benchmark the quality of measurement information. The uniquely identifiable faults can be easily found by counting the number of the minimal diagnosable classes that contain only one single fault.

The third criterion is the flexibility provided by the current gauging system toward achieving the full diagnosability. Some gauging systems could be rigid in the sense that certain faults or fault combinations, which may be difficult to measure in practice, have to be known to achieve a fully diagnosable system. Some other gauging systems may provide information in a flexible way, i.e., many fault combinations can be selected to make the system fully diagnosable. This comparison needs the concept of *minimal complementary classes*. A minimal complementary class is a *minimal* set of faults such that if they are known, all the faults of the system can be uniquely identified. Consider a system with four faults and three minimal diagnosable classes as $\{u_1, u_2\}$, $\{u_1, u_3, u_4\}$, and $\{u_2, u_3, u_4\}$. One can verify that the minimal complementary classes for this system are $\{u_1, u_3\}$, $\{u_1, u_4\}$, $\{u_2, u_3\}$, $\{u_2, u_4\}$, and $\{u_3, u_4\}$. The number of minimal complementary classes is five. A system with more minimal complementary classes is considered to be more flexible.

In general, it is difficult to find the complete sets of minimal complementary classes by simply trying out different fault combinations, especially for a complex system with large fault number and intricate fault combinations. Corollary 3 facilitates the determination of minimal complimentary classes; its proof can be found in Appendix A5.

Corollary 3 A set of faults forms a minimal complementary class if and only if the set contains n-ρ faults but does not contain any minimal diagnosable class, where n is the total number of faults and ρ is the rank of the diagnosability testing matrix.

With Corollary 3, the complete minimal complementary classes can be found through a search among all fault sets with n-ρ faults. If the entire fault set can be partitioned into many smaller distinct connected fault classes, the load of searching the complete minimal complementary classes can be further substantially reduced. Corollary 3 can be applied to a connected fault class but n should be the total number of faults in the connected fault class and ρ is the rank of the space spanned by the associated row vectors in the RREF of the transpose of the testing matrix. Individual searches can be conducted within each connected fault class. The complete set of minimal complementary classes can then be obtained by joining the minimal complementary classes from each connected fault class and adding the non-diagnosable faults. An example will be given in Section 4 to illustrate this procedure.

The order of using the three criteria generally depends on the requirements of individual applications. In some cases, when the ultimate goal is to design a gauging system providing the full diagnosability, we can skip the second criterion and compare the number of minimal complementary classes directly. In some other cases, the second criterion can be used before the first criterion if the uniquely identified fault is highly desired. Based on our experience, using the three criteria in the sequence in which they were presented here is an effective way for gauging system evaluation in many industrial applications.

4. Case Study

4.1 Case Study of a Multistage Assembly Process

Consider the assembly processes shown in Figure 1 and Figure 3. The product state variable $\mathbf{x}_k$ is denoted by random deviations associated with the degrees of freedom (d.o.f.) of each part. Each 2-D part in this example has three d.o.f. (two translational and one rotational) and the size of $\mathbf{x}_k$ is 12×1 given that there are four parts. The state vector $\mathbf{x}_k$ is expressed as

$$\mathbf{x}_k = [\delta x_{1,k} \quad \delta z_{1,k} \quad \delta\alpha_{1,k} \;\vdots\; \cdots \;\vdots\; \cdots \;\vdots\; \delta x_{4,k} \quad \delta z_{4,k} \quad \delta\alpha_{4,k}]^T \tag{13}$$

where δ is the deviation operator, $\delta x_{i,k}, \delta z_{i,k}$, and $\delta \alpha_{1,k}$ are two translational and one rotational deviations of part i on station k, respectively. If part i has not yet appeared on station k, the corresponding $\delta x_{i,k}$, $\delta z_{i,k}$, and $\delta \alpha_{1,k}$ are zeros.

The input vector $\mathbf{u}_k$ represents the random deviations associated with fixture locators on station k. There are a total of 18 components of fixture deviations on three stations as indicated by the number 1-18 (i.e., the 18 faults) in Figure 3. Thus, we have $\mathbf{u}_1 = [\delta p_1 \quad \cdots \quad \delta p_6]^T$, $\mathbf{u}_2 = [\delta p_7 \quad \cdots \quad \delta p_{15}]^T$, $\mathbf{u}_3 = [\delta p_{16} \quad \delta p_{17} \quad \delta p_{18}]^T$, where δp_i is the deviation associated with fault i.

The measurement $\mathbf{y}$ contains positional deviations detected at M_i, i=1,…,5. In this 2-D case, each M_i can deviate in x- and/or z-directions. Hence, $\mathbf{y}_1 = [\delta M_1(x) \quad \delta M_1(z) \quad \delta M_2(x) \quad \delta M_2(z)]^T$, $\mathbf{y}_2 = [\delta M_3(x) \quad \delta M_3(z) \quad \delta M_4(x) \quad \delta M_4(z)]^T$, and $\mathbf{y}_3 = [\delta M_5(x) \quad \delta M_5(z)]^T$.

The state space representation of this process is shown as follows

$$\mathbf{x}_1 = \mathbf{B}_1\mathbf{u}_1 + \mathbf{w}_1 \text{ and } \mathbf{x}_k = \mathbf{A}_{k-1}\mathbf{x}_{k-1} + \mathbf{B}_k\mathbf{u}_k + \mathbf{w}_k,\ k=2, 3, \tag{14}$$

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k,\ k=1,2,3. \tag{15}$$

Matrices $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$ are determined by process design and sensor deployment. The $\mathbf{A}_k$ characterizes the change in product state when a product is transferred from station k to station k+1. Thus, $\mathbf{A}_k$ depends on the coordinates of fixture locators on two adjacent stations k and k+1. The $\mathbf{B}_k$ determines how fixture deviations affects product deviations on station k and is thus determined by the coordinates of fixture locators on station k. The $\mathbf{C}_k$ is determined by the coordinates of measurement points such as M_1 to M_5 in this example.

Following the model development presented in Jin and Shi (1999) and Ding, Ceglarek, and Shi (2000), we give the numerical expressions of **A**'s, **B**'s, and **C**'s of the assembly processes shown in Figures 1 and 3, respectively. The **A**'s, **B**'s, $\mathbf{C}_1$, and $\mathbf{C}_2$ are the same for these two processes since their fixture layouts are the same for all stations and the sensor deployments are the same for Station I and Station II.

$$\mathbf{A}_1=\left[\begin{array}{cccccc:c} 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & 0 & 0 & \\ 0 & 0.0007 & 1 & 0 & -0.0007 & -0.3497 & \\ -1 & 0 & 0 & 1 & 0 & 0 & \mathbf{0}^{6\times6} \\ 0 & -0.3497 & 0 & 0 & 0.3497 & -325.17 & \\ 0 & 0.0007 & 0 & 0 & -0.0007 & 0.6503 & \\ \hdashline & & & \mathbf{0}^{6\times6} & & & \mathbf{I}^{6\times6} \end{array}\right]_{12\times12} \quad \mathbf{A}_2=\left[\begin{array}{ccc:c:ccc} 0 & 0 & 0 & & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{0}^{3\times6} & 0 & 0 & 0 \\ 0 & 0.0005 & 1 & & 0 & -0.0005 & -0.2392 \\ \hdashline -1 & 0 & 0 & & 0 & 0 & 0 \\ 0 & -0.5550 & 0 & & 0 & -0.4450 & -222.49 \\ 0 & 0.0005 & 0 & \mathbf{I}^{6\times6} & 0 & -0.0005 & -0.2392 \\ -1 & -0.2153 & 0 & & 0 & 0.2153 & 107.655 \\ 0 & -0.2392 & 0 & & 0 & -0.7608 & -380.38 \\ 0 & 0.0005 & 0 & & 0 & -0.0005 & -0.2392 \\ \hdashline -1 & 0 & 0 & & 1 & -0.0005 & 0 \\ 0 & -0.2392 & 0 & \mathbf{0}^{3\times6} & 0 & 0.2392 & -380.38 \\ 0 & 0.0005 & 0 & & 0 & -0.0005 & 0.7608 \end{array}\right]_{12\times12}, \tag{16}$$

$$\mathbf{B}_1=\left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & -0.0014 & 0.0014 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & -0.002 & 0.002 \\ \hdashline & & \mathbf{0}^{6\times6} & & & \end{array}\right]_{12\times6} \quad \mathbf{B}_2=\left[\begin{array}{ccc:cccccc} 1 & 0 & 0 & & & & & & \\ 0 & 1 & 0 & & & & & & \\ 0 & -0.0007 & 0.0007 & & & \mathbf{0}^{6\times6} & & & \\ 1 & 0 & 0 & & & & & & \\ 0 & 0.3497 & 0.6503 & & & & & & \\ 0 & -0.0007 & 0.0007 & & & & & & \\ \hdashline & & & 1 & 0 & 0 & 0 & 0 & 0 \\ & & & 0 & 1 & 0 & 0 & 0 & 0 \\ & \mathbf{0}^{6\times3} & & 0 & -0.002 & 0.002 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 1 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 & 1 & 0 \\ & & & 0 & 0 & 0 & 0 & -0.002 & 0.002 \end{array}\right]_{12\times9} \quad \mathbf{B}_3=\left[\begin{array}{ccc} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -0.0005 & 0.0005 \\ 1 & 0 & 0 \\ 0 & 0.5550 & 0.4450 \\ 0 & -0.0005 & 0.0005 \\ 1 & 0.2153 & -0.2153 \\ 0 & 0.2392 & 0.7608 \\ 0 & -0.0005 & 0.0005 \\ 1 & 0 & 0 \\ 0 & 0.2392 & 0.7608 \\ 0 & -0.0005 & 0.0005 \end{array}\right]_{12\times3}, \tag{17}$$

$$\mathbf{C}_1=\left[\begin{array}{ccc:ccc:c} 1 & 0 & -550 & & \mathbf{0}^{2\times3} & & \mathbf{0}^{2\times6} \\ 0 & 1 & -100 & & & & \\ \hdashline & \mathbf{0}^{2\times3} & & 1 & 0 & -550 & \mathbf{0}^{2\times6} \\ & & & 0 & 1 & -630 & \end{array}\right]_{4\times12} \quad \mathbf{C}_2=\left[\begin{array}{ccc:c:ccc:c} 1 & 0 & -550 & \mathbf{0}^{2\times3} & & \mathbf{0}^{2\times3} & & \mathbf{0}^{2\times3} \\ 0 & 1 & -100 & & & & & \\ \hdashline & \mathbf{0}^{2\times3} & & \mathbf{0}^{2\times3} & 1 & 0 & -300 & \mathbf{0}^{2\times3} \\ & & & & 0 & 1 & -740 & \end{array}\right]_{4\times12}. \tag{18}$$

We use $\mathbf{C}_3^1$ and $\mathbf{C}_3^2$ to denote $\mathbf{C}_3$ of these two gauging systems, respectively. Their expressions are

$$\mathbf{C}_3^1=\left[\begin{array}{ccc:c} 1 & 0 & -550 & \mathbf{0}^{2\times9} \\ 0 & 1 & -100 & \end{array}\right]_{2\times12} \qquad \mathbf{C}_3^2=\left[\begin{array}{c:ccc} \mathbf{0}^{2\times9} & 1 & 0 & -200 \\ & 0 & 1 & 620 \end{array}\right]_{2\times12}. \tag{19}$$

For simplicity, we only discuss the variance diagnosability of fixture faults in this study. Thus, we use $\mathbf{H}_r$ in Equation (12) as the testing matrix. In order to use $\mathbf{H}_r$, we need to obtain $\mathbf{\Gamma}$ first. Substituting **A**'s, **B**'s and **C**'s in Equations (16) ~ (19) into Equation (3) yields

$$\mathbf{\Gamma}^1=\begin{bmatrix} 1 & 0.786 & -0.786 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.143 & -0.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.1 & -1.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.26 & -1.26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.401 & -0.786 & 0 & 0 & 0.385 & 1 & 0.385 & -0.385 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.073 & -0.143 & 0 & 0 & 0.070 & 0 & 1.070 & -0.070 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.6 & -0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.48 & -1.48 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.401 & -0.786 & 0 & 0 & 0.385 & 0 & 0.122 & -0.385 & 0 & 0 & 0 & 0 & 0 & 0.263 & 1 & 0.263 & -0.263 \\ 0 & 0.073 & -0.143 & 0 & 0 & 0.070 & 0 & 0.022 & -0.070 & 0 & 0 & 0 & 0 & 0 & 0.048 & 0 & 1.048 & -0.048 \end{bmatrix} \tag{20}$$

and

$$\mathbf{\Gamma}^2 = \begin{bmatrix} 1 & 0.786 & -0.786 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.143 & -0.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.1 & -1.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.26 & -1.26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.401 & -0.786 & 0 & 0 & 0.385 & 1 & 0.385 & -0.385 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.073 & -0.143 & 0 & 0 & 0.070 & 0 & 1.070 & -0.070 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.6 & -0.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2.48 & -1.48 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -0.096 & 0 & 0 & 0 & 0 & 1 & 0.4 & -0.304 & 1 & 0.096 & -0.096 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.057 & 0 & 0 & 0 & 0 & 0 & -0.24 & 0.183 & 0 & -0.057 & 1.057 \end{bmatrix} \qquad (21)$$

where the superscript 1 or 2 indicates which gauging system the $\mathbf{\Gamma}$ is associated with. Further, $\mathbf{H}_r$ can be obtained following its definition in Equation (12). Their expressions are

$$\mathbf{H}_r^1 = \begin{bmatrix} 1 & 0.617 & 0.617 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.617 & 5.089 & 2.050 & 0 & 0 & 0.102 & 0.161 & 0.08 & 0.102 & 0 & 0 & 0 & 0 & 0 & 0.012 & 0.161 & 0.033 & 0.012 \\ 0.617 & 2.050 & 3.661 & 0 & 0 & 0.390 & 0.617 & 0.307 & 0.390 & 0 & 0 & 0 & 0 & 0 & 0.046 & 0.617 & 0.127 & 0.046 \\ 0 & 0 & 0 & 1 & 1.21 & 1.21 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.21 & 39.91 & 16.46 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.102 & 0.390 & 1.21 & 16.46 & 9.63 & 0.148 & 0.073 & 0.093 & 0 & 0 & 0 & 0 & 0 & 0.011 & 0.148 & 0.030 & 0.011 \\ 0 & 0.161 & 0.617 & 0 & 0 & 0.148 & 1.0 & 0.148 & 0.148 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.08 & 0.307 & 0 & 0 & 0.073 & 0.148 & 1.711 & 0.073 & 0 & 0 & 0 & 0 & 0 & 0.001 & 0.015 & 0.003 & 0.001 \\ 0 & 0.102 & 0.390 & 0 & 0 & 0.093 & 0.148 & 0.073 & 0.093 & 0 & 0 & 0 & 0 & 0 & 0.011 & 0.148 & 0.03 & 0.011 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.36 & 0.36 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.36 & 42.39 & 16.24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.36 & 16.24 & 6.505 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.012 & 0.046 & 0 & 0 & 0.011 & 0 & 0.001 & 0.011 & 0 & 0 & 0 & 0 & 0 & 0.005 & 0.069 & 0.014 & 0.005 \\ 0 & 0.161 & 0.617 & 0 & 0 & 0.148 & 0 & 0.015 & 0.148 & 0 & 0 & 0 & 0 & 0 & 0.069 & 1 & 0.069 & 0.069 \\ 0 & 0.033 & 0.127 & 0 & 0 & 0.030 & 0 & 0.003 & 0.030 & 0 & 0 & 0 & 0 & 0 & 0.014 & 0.069 & 1.362 & 0.014 \\ 0 & 0.012 & 0.046 & 0 & 0 & 0.011 & 0 & 0.001 & 0.011 & 0 & 0 & 0 & 0 & 0 & 0.005 & 0.069 & 0.014 & 0.005 \end{bmatrix} \qquad (22)$$

and

$$\mathbf{H}_r^2 = \begin{bmatrix} 1 & 0.617 & 0.617 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.617 & 4.367 & 1.224 & 0 & 0 & 0.025 & 0.161 & 0.054 & 0.025 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.617 & 1.224 & 1.627 & 0 & 0 & 0.098 & 0.617 & 0.207 & 0.098 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.21 & 1.21 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.21 & 39.91 & 16.46 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.025 & 0.098 & 1.21 & 16.46 & 8.705 & 0.148 & 0.050 & 0.023 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.161 & 0.617 & 0 & 0 & 0.148 & 4.0 & 0.231 & 0.148 & 0 & 0 & 0 & 1 & 0.16 & 0.093 & 1 & 0.009 & 0.009 \\ 0 & 0.054 & 0.207 & 0 & 0 & 0.050 & 0.231 & 1.703 & 0.050 & 0 & 0 & 0 & 0.009 & 0.003 & 0.002 & 0.009 & 0.0002 & 0.005 \\ 0 & 0.025 & 0.098 & 0 & 0 & 0.023 & 0.148 & 0.05 & 0.023 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.36 & 0.36 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.36 & 42.39 & 16.24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.36 & 16.24 & 6.505 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.009 & 0 & 0 & 0 & 0 & 1 & 0.16 & 0.093 & 1 & 0.009 & 0.009 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.16 & 0.003 & 0 & 0 & 0 & 0 & 0.16 & 0.047 & 0.027 & 0.16 & 0.003 & 0.085 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.093 & 0.002 & 0 & 0 & 0 & 0 & 0.093 & 0.027 & 0.016 & 0.093 & 0.002 & 0.049 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0.009 & 0 & 0 & 0 & 0 & 1 & 0.16 & 0.093 & 1 & 0.009 & 0.009 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.009 & 0.0002 & 0 & 0 & 0 & 0 & 0.009 & 0.003 & 0.002 & 0.009 & 0.0002 & 0.005 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.009 & 0.005 & 0 & 0 & 0 & 0 & 0.009 & 0.085 & 0.049 & 0.009 & 0.005 & 1.271 \end{bmatrix}. \qquad (23)$$

The RREF of $\mathbf{H}_r$'s and the corresponding fault structures are compared in Table 1. For gauging system 1, fourteen rows have only one nonzero element, corresponding to fourteen uniquely identified faults and hence minimal diagnosable classes, {1},…,{12},{16},{17}. Two faults

(13,14) correspond to zero columns and hence they are not diagnosable. The 13th row has two nonzero elements, i.e., $[\mathbf{0}^{1\times 12} \mid 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1]$, indicating that {15, 18} is a minimal diagnosable class. The class {15, 18} is also a connected fault class and since it is already minimal, no further permutation is needed. Similarly, for gauging system 2, there are thirteen uniquely diagnosable classes, {1},…,{12}, {18}. Two minimal diagnosable classes, {13, 16} and {14, 15, 17}, correspond to the 13th and 14th row, respectively. No permutation of $\mathbf{H}_r$ is needed for gauging system 2, either.

For gauging system 1, in order to achieve a fully diagnosable system, at least n-ρ=3 faults need to be known. We first search fault set {15,18} with n=2 and ρ=1. It is clear that {15} and {18} are two minimal complementary fault classes for the connected fault class {15, 18}. Adding the non-diagnosable faults {13, 14}, we obtain the minimal complementary classes as {13, 14, 15} and {13, 14, 18}. The number of minimal complementary classes is two.

For gauging system 2, in order to find the minimal complementary class, we search the faults among {13, 16} with n=2 and ρ=1 and among {14,15,17} with n=3 and ρ=1. The search yields {13} and {16} for {13, 16} and {14,15}, {14,17}, and {15,17} for {14,15,17}. Joining these two fault groups together gives us $C_2^1 \cdot C_3^1 = 6$ minimal complementary classes, which are listed in Table 1. This analysis verifies that although engineering systems have many potential faults (18 faults in this case), they can often be partitioned into smaller connected fault classes.

Neither gauging systems provides full diagnosability since their $\mathbf{H}_r$'s are not of full rank. Ranks of $\mathbf{H}_r$'s are the same (ρ=15), suggesting that the amount of information obtained by both systems is the same. But gauging system 1 can uniquely identify 14 faults, which are faults 1-12, 16, and 17, while gauging system 2 can only uniquely identify 13 faults, which are faults 1-12 and 18. The information quality provided by gauging system 1 is considered better than that of gauging system 2. In this sense, gauging system 1 provides more valuable information. However, one may also notice that gauging system 2 can have six possible ways of measuring additional faults in achieving a fully diagnosable system, while gauging system 1 only has two possibilities. This difference indicates that gauging system 2 is more flexible. If the third criterion is in a higher priority, gauging system 2 is more favorable.

Table 1. Comparison of gauging systems 1 and 2

	Gauging System 1	Gauging System 2
RREF($\mathbf{H}_r$)	$\left[\begin{array}{c\|cccccc} \mathbf{I}^{12\times12} & & & \mathbf{0}^{12\times6} & & & \\ \hline & 0 & 0 & 1 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 1 & 0 & 0 \\ \mathbf{0}^{6\times12} & 0 & 0 & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]_{18\times18}$	$\left[\begin{array}{c\|cccccc} \mathbf{I}^{12\times12} & & & \mathbf{0}^{12\times6} & & & \\ \hline & 1 & 0 & 0 & 1 & 0 & 0 \\ & 0 & 1 & 0.58 & 0 & 0.06 & 0 \\ \mathbf{0}^{6\times12} & 0 & 0 & 0 & 0 & 0 & 1 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]_{18\times18}$
# of potential faults	18	18
Rank of testing matrix	15	15
Minimal diagnosable classes	{1},…,{12},{16},{17}, {15, 18}	{1}, …, {12}, {18}, {13,16}, {14, 15, 17}
# of uniquely identified faults	14	13
Minimal complementary classes	{13, 14, 15}, {13, 14, 18}	{13, 14, 15}, {13, 14, 17}, {13, 15, 17}, {16, 14, 15}, {16, 14, 17}, {16, 15, 17}
# of minimal complementary classes	2	6

4.2 Case Study of a Multistage Machining Process

The proposed evaluation criteria can also be applied to multistage machining processes. To machine a workpiece, we need first to fix the location of the workpiece in the space. Figure 4 shows a widely used 3-2-1 fixturing setup. If we require the workpiece to touch all the locating pads (L_1~L_3) and locating pins (P_1~P_3), the location of the workpiece in the machine coordinate system XYZ is fixed. The surface of the workpiece that touches the locating pads (L_1~L_3) (surface ABCD in Figure 4) is called "primary datum." Similarly, surface ADHE is called "secondary datum" and DCGH is called "tertiary datum" in Figure 4. Since the primary datum (surface ABCD) touches L_1~L_3, the translational motion in Z direction and the rotational motion in X and Y directions are restrained. Similarly, the secondary datum constrains the translational motion in X direction and the rotational motion in Z direction; the tertiary datum constrains the translational motion in the Y direction. Therefore, all six degrees of freedom associated with the workpiece are constrained by these three datum surfaces and the corresponding locating pins and pads.

The cutting tool-path is calibrated with respect to the machine coordinate system XYZ. Clearly, an error in the position of locating pads and pins will cause a geometric error in the machined feature. Suppose that we mill a slot on surface EFGH in Figure 4. If L_1 is higher than its nominal position, the workpiece will be tilted with respect to XYZ. However, the cutting tool path is still determined with respect to XYZ. Hence, the bottom surface of the finished slot will not be parallel to the primary datum (ABCD). Besides the fixture error, the geometric errors in the datum feature will also affect the workpiece quality. For example, if the primary datum (ABCD) is not perpendicular to the secondary datum (ADHE), the milled slot will not be perpendicular to the secondary datum, either.

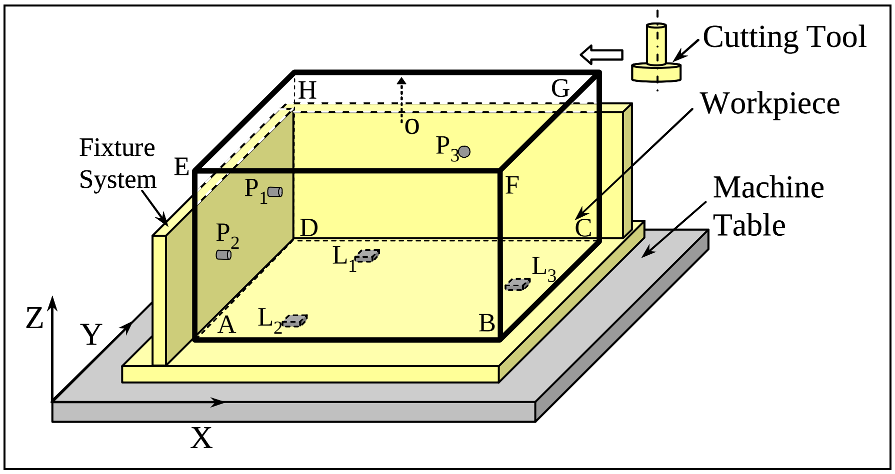
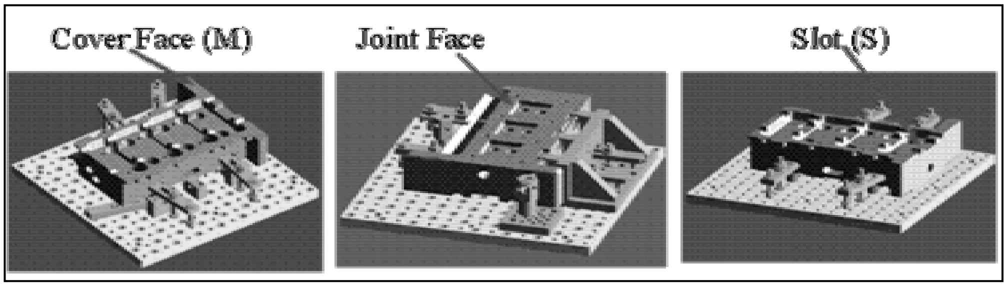


Figure 4. A Typical 3-2-1 Fixturing Configuration

A three-stage machining process using this 3-2-1 fixture setup is shown in Figure 5. The product is an automotive engine head. The features are the cover face (M), joint face, and the slot (S). The cover face, joint face, and the slot are milled at the 1st (Figure 5(a)), the 2nd (Figure 5(b)), and the 3rd (Figure 5(c)) stages, respectively.

We treat the positional errors of product features after stage k as state vector $\mathbf{x}_k$, the errors of fixture and cutting tool-path at stage k as input $\mathbf{u}_k$, and the measurements of positions and orientations of the machined product features as $\mathbf{y}_k$, which can be obtained by a Coordinate Measuring Machine (CMM). The state space model (Equation (1)) can be obtained through a similar (to the above panel assembly) but more complicated 3D kinematics analysis, where $\mathbf{A}_{k-1}\mathbf{x}_{k-1}$ is the error contributed by the errors of datum features (these features are produced in previous stages) and $\mathbf{B}_k\mathbf{u}_k$ is the error contributed by fixture and/or cutting tool at stage k. Details of this process and the corresponding state space model can be found in Zhou, Huang, and Shi (2002).



(a) (b) (c)

Figure 5. Process Layout at Three Stages

After the model in Equation (1) is obtained, the diagnosability study for the multistage machining process can be conducted following the theories in Sections 2 and 3. We focus on the fixture error in this case study. For a 3-2-1 fixture setup, there are 6 potential fixture errors at each stage (each locating pad and pin could have one error). Hence, there are 18 potential faults in the whole system, where faults 1-6 represent locator errors at the 1st stage, faults 7-12 at the 2nd stage, and faults 13-18 at the 3rd stage, respectively. Three gauging systems are used to measure slot S, cover face M, and the rough datum, respectively, where the rough datum is the primary datum at the 1st stage and can be seen from the joint face. The results of a fault diagnosability for the three systems are listed in Table 2.

The RREF of the testing matrices of gauging system 1 and 2 have a very simple structure. For the gauging system 3 (the 4th column in Table 2), the first 3 rows of RREF($\mathbf{\Gamma}$) share common nonzero column positions. The corresponding faults, {1, 2, 3, 7, 8, 9}, form a connected fault class regarding its mean diagnosability. By permuting the corresponding columns of RREF($\mathbf{\Gamma}$), we can generate 15 minimal diagnosable classes (each has four faults) within this connected fault class as shown in Table 2. The minimal complementary class of this connected fault class can be found by searching the class with n = 6, ρ = 3. We obtain C_6^3 = 20 minimal complementary classes for the connected class: {1, 7, 8}, {2, 7, 8}, {3, 7, 8}, {1, 7, 9}, {2, 7, 9}, {3, 7, 9}, {1, 8, 9}, {2, 8, 9}, {3, 8, 9}, {1, 2, 7}, {1, 2, 8}, {1, 2, 9}, {1, 3, 7}, {1, 3, 8}, {1, 3, 9}, {2, 3, 7}, {2, 3, 8}, {2, 3, 9}, {1, 2, 3}, {7, 8, 9}. Adding the non-diagnosable faults {4, 5, 6, 13~18}, we can obtain 20 minimal complementary classes for the system regarding the mean diagnosability. It is also interesting to see that while the faults {1, 2, 3, 7, 8, 9} form a connected fault class regarding its mean diagnosability, they are uniquely diagnosable regarding its variance diagnosability. This verifies our previous remark that mean diagnosability requires a stronger condition than variance diagnosability does.

Table 2. Comparison of gauging systems

Gauging System		System 1 (Slot S)	System 2 (Cover Face M)	System 3 (Rough Datum)
Mean Diagnosability: RREF($\boldsymbol{\Gamma}$)		$\left[\begin{array}{c\|c} \mathbf{0}^{6\times12} & \mathbf{I}^{6\times6} \\ \hline \mathbf{0}^{30\times12} & \mathbf{0}^{30\times6} \end{array}\right]_{36\times18}$	$\left[\begin{array}{cc\|ccc\|c} \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \begin{matrix}1&0&0\\0&1&0\\0&0&-1\\0&0&0\\0&0&0\\0&0&0\end{matrix} & & & \mathbf{0}^{6\times6} \\ \hline \mathbf{0}^{30\times3} & \mathbf{0}^{30\times6} & \mathbf{0}^{30\times3} & & & \mathbf{0}^{30\times6} \end{array}\right]_{36\times18}$	$\left[\begin{array}{cc\|ccc\|cc} \mathbf{I}^{3\times3} & \mathbf{0}^{3\times3} & \begin{matrix}-0.63&0.53&-0.90\\0.47&-0.57&-0.90\\-0.49&-0.48&-0.04\end{matrix} & & & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times6} \\ \hline \mathbf{0}^{3\times3} & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times3} & & & \mathbf{I}^{3\times3} & \mathbf{0}^{3\times6} \\ \mathbf{0}^{30\times3} & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times3} & & & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times6} \end{array}\right]$
Variance Diagnosability: RREF($\mathbf{H}_r$)		$\left[\begin{array}{c\|c} \mathbf{0}^{6\times12} & \mathbf{I}^{6\times6} \\ \hline \mathbf{0}^{12\times12} & \mathbf{0}^{12\times6} \end{array}\right]_{18\times18}$	$\left[\begin{array}{cc\|ccc\|c} \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \begin{matrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\\0&0&0\\0&0&0\end{matrix} & & & \mathbf{0}^{6\times6} \\ \hline \mathbf{0}^{12\times3} & \mathbf{0}^{12\times6} & \mathbf{0}^{12\times3} & & & \mathbf{0}^{12\times6} \end{array}\right]_{18\times18}$	$\begin{bmatrix} \mathbf{I}^{3\times3} & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times6} & \mathbf{0}^{3\times6} \\ \mathbf{0}^{6\times3} & \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \mathbf{0}^{6\times6} \\ \mathbf{0}^{9\times3} & \mathbf{0}^{9\times3} & \mathbf{0}^{9\times6} & \mathbf{0}^{9\times6} \end{bmatrix}$
# of potential faults		18	18	18
Rank of testing matrix	$\boldsymbol{\Gamma}$	6	6	6
	$\mathbf{H}_r$	6	6	9
Minimal diagnosable classes	mean	{13}, {14}, {15}, {16}, {17}, {18}	{7}, {8}, {9}, {4, 10}, {5, 11}, {6, 12}	{10}, {11}, {12}, {1, 7, 8, 9}, {2, 7, 8, 9}, {3, 7, 8, 9}, {1, 3, 8, 9}, {2, 3, 8, 9}, {1, 7, 3, 9}, {2, 7, 3, 9}, {1, 7, 8, 3}, {2, 7, 8, 3}, {1, 2, 8, 9}, {1, 7, 2, 9}, {1, 7, 8, 2}, {1, 2, 3, 9}, {1, 2, 8, 3}, {1, 7, 2, 3}
	variance	{13}, {14}, {15}, {16}, {17}, {18}	{7}, {8}, {9}, {4, 10}, {5, 11}, {6, 12}	{1}, {2}, {3}, {7}, {8}, {9}, {10}, {11}, {12}
# of uniquely identified faults	mean	6	3	3
	variance	6	3	9
# of minimal complementary classes	mean	1	8	20
	variance	1	8	1

5. Concluding Remarks

This paper studied the diagnosability of process faults given the product quality measurements in a complicated multistage manufacturing process. This study reveals that the diagnosis capability that a gauging system can provide strongly depends on sensor deployment in a multistage manufacturing system. A poorly designed gauging system is likely to result in the loss of diagnosability. On the contrary, a well-designed gauging system, which achieves the desired level of diagnosability, can not only monitor the process change but also quickly identify the process root causes of quality-related problems. The quick root cause identification will lead to product quality improvement, production downtime reduction, and hence a remarkable cost reduction in manufacturing systems.

This study was a model-based approach; a linear fault-quality model was used. The results can be used where a linear diagnostic model is available. Because the errors of tooling elements considered in quality control problems are often much smaller than the nominal parameters, most of manufacturing systems can be linearized and then represented by a linear model under the small error assumption. Many of the linear state space models reviewed in Section 2 were validated through comparison with either a commercial software simulation (Ding, Ceglarek, and Shi 2000) or with experimental data (Zhou, Huang, and Shi 2002; Djurdjanovic and Ni 2001). Thus, the small error assumption is not restrictive and the methodology presented in this article is generic and applicable to various manufacturing systems.

Another note on the applicability of the reported methodology is that for some poorly designed manufacturing system, a large number of process faults could possibly be coupled together and form a single huge connected fault class. As a result, it would be impractical to exhaust matrix column-permutation in finding the complete list of minimal diagnosable classes and the diagnosability study itself then becomes intractable.

The development of the diagnosis algorithm that can give the best estimation of process faults will follow this diagnosability study. This is our ongoing research.

Appendices

A1. Theorem 4.2.1 in Rao and Kleffe (1988).

Consider a general linear mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\beta}$ represents the fixed effects and $\boldsymbol{\varepsilon}$ is zero mean and $Cov(\boldsymbol{\varepsilon}) = \theta_1 \mathbf{V}_1 + ... + \theta_r \mathbf{V}_r$. Denote $\boldsymbol{\theta} = [\theta_1 \quad ... \quad \theta_r]^T$ as variance components. $\mathbf{p}^{\mathrm{T}}\boldsymbol{\beta}$ is identifiable if and only if $\mathbf{p} \in R(\mathbf{X}^T)$, $\mathbf{f}^{\mathrm{T}}\boldsymbol{\theta}$ is identifiable if and only if $\mathbf{f} \in R(\mathbf{H})$, $\mathbf{H}' = (tr(\mathbf{V}_i \mathbf{V}_j)), 1 \le i \le r, 1 \le j \le r$.

A2. Proof of Theorem 1

This theorem is an extension of the theorem (4.2.1) in Rao and Kleffe (1988) (it is listed in A1). From that theorem, $\mathbf{p}^{\mathrm{T}}\boldsymbol{\alpha}$ is diagnosable if and only if $\mathbf{p} \in R([\boldsymbol{\Gamma}^T \quad ... \quad \boldsymbol{\Gamma}^T])$. It is clear that $R([\boldsymbol{\Gamma}^T \quad ... \quad \boldsymbol{\Gamma}^T]) = R(\boldsymbol{\Gamma}^T)$. Therefore, (i) holds. For (ii), $\mathbf{f}^{\mathrm{T}}\boldsymbol{\theta}$ is diagnosable if and only if $\mathbf{f} \in R(\mathbf{H}')$, where $\mathbf{H}' = (tr(\mathbf{F}_i\mathbf{F}_j)),\ 1 \le i \le P+Q+1,\ 1 \le j \le P+Q+1$, and $\mathbf{F}_i$ and $\mathbf{F}_j$ are defined in Equation (8). It can be further shown that $\mathbf{H}' = M\mathbf{H}$. Since a constant coefficient does not affect the range space of a matrix, the result of (ii) follows.

A3. Proof of Theorem 2

Denote the row and column space of a matrix as Row(·) and Col(·), respectively, the RREF of $\mathbf{G}^T$ as $\mathbf{G}_r^T$, and the nonzero row vectors of $\mathbf{G}_r^T$ as $\{\mathbf{v}_i\}_{i=1\ldots\rho}$, where ρ is the rank of $\mathbf{G}_r^T$. Noticing that $\mathbf{G}_r^T$ is unique and Row($\mathbf{G}_r^T$) = Row($\mathbf{G}^T$) (Lay, 1997), we have $\mathbf{v}_i \in \mathrm{Col}(\mathbf{G})$. Hence, $\boldsymbol{\theta}[\mathbf{v}_i]$ is a diagnosable class.

We need to further prove that $\boldsymbol{\theta}[\mathbf{v}_i]$ is a minimal diagnosable class. From the algorithm to obtain RREF, the leftmost element of $\mathbf{v}_i$ is always a "leading 1." The position of such a "leading 1" in $\mathbf{v}_i$ is called the pivot position. Denote the set of all pivot positions contributed by the rows of $\mathbf{G}_r^T$ as $\boldsymbol{\Xi}$. It is known that (i) given an $i \in \{1,\ldots,\rho\}$, there is only one nonzero element in $\{\mathbf{v}_i(j)\}_{j\in\Xi}$. (ii) if $\{\mathbf{c}_i\}_{i=1\ldots n}$ are columns of $\mathbf{G}_r^T$, then there is only one nonzero element in $\mathbf{c}_i$ if $i \in \boldsymbol{\Xi}$. From (i), $\boldsymbol{\theta}[\mathbf{v}_i]$ must be in the form $\{u_{p_i}, u_{i_1}, ..., u_{i_k}\}$, $p_i \in \boldsymbol{\Xi},\ i_1 ... i_k \notin \boldsymbol{\Xi}$. Assume that $\boldsymbol{\theta}[\mathbf{v}_i]$ is not a minimal diagnosable class, we can then find a vector $\mathbf{v}_i'$ such that $\boldsymbol{\theta}[\mathbf{v}_i'] \subset \boldsymbol{\theta}[\mathbf{v}_i]$, $\mathbf{v}_i' \in \mathrm{Row}(\mathbf{G}_r^T)$ and hence $\mathbf{v}_i'$ can be written as $\mathbf{v}_i' = \sum_{j=1}^{\rho} a_j \mathbf{v}_j$. However, from (ii), if there is a j, a_j is nonzero, u_{p_j} must be in $\boldsymbol{\theta}[\mathbf{v}_i']$, where p_j is the pivot position of $\mathbf{v}_j$. Since $\boldsymbol{\theta}[\mathbf{v}_i]$ only contains one fault at the pivot position p_i, a_i is the only possible nonzero coefficient. Then, $\boldsymbol{\theta}[\mathbf{v}_i'] = \boldsymbol{\theta}[\mathbf{v}_i]$. This contradicts the assumption that $\boldsymbol{\theta}[\mathbf{v}_i'] \subset \boldsymbol{\theta}[\mathbf{v}_i]$, implying that $\boldsymbol{\theta}[\mathbf{v}_i]$ is a minimal diagnosable class.

A4. Proof of Corollary 1

Denote $\{\mathbf{v}_i\}_{i=1\ldots\rho}$ as the nonzero row vectors of $\mathbf{G}_r'^T$. We want to prove that the pivot position of the last row $\mathbf{v}_\rho$ must be n-s+1 (this position corresponds to u_{i_1}). First, suppose that the pivot position of $\mathbf{v}_\rho$ is larger than n-s+1. If so, $\boldsymbol{\theta}[\mathbf{v}_\rho] \subset \boldsymbol{\Theta}$. According to Theorem 2, however, $\boldsymbol{\theta}[\mathbf{v}_\rho]$ is a diagnosable class. This contradicts the fact that $\mathbf{\Theta}$ is minimal. Second, assume that the pivot position of $\mathbf{v}_\rho$ is smaller than n-s+1. If so, a fault among $\{u_{i_{s+1}},...,u_{i_n}\}$ must belong to $\boldsymbol{\theta}[\mathbf{v}_\rho]$. Since the pivot position of $\mathbf{v}_\rho$ is the largest among all the pivot positions of $\{\mathbf{v}_i\}_{i=1\ldots\rho}$, given any vector $\mathbf{v}_f = \Sigma_{j=1}^{\rho} a_j \mathbf{v}_j$ (defined as an arbitrary nontrivial linear combination of $\{\mathbf{v}_i\}_{i=1\ldots\rho}$), $\boldsymbol{\theta}[\mathbf{v}_f]$ contains at least one element among $\{u_{i_{s+1}},...,u_{i_n}\}$. According to Theorem 1, any diagnosable class should contain at least one element among $\{u_{i_{s+1}},...,u_{i_n}\}$ since $\mathbf{v}_f$ is an arbitrary vector in Row($\mathbf{G}_r'^T$). This contradicts the assertion that $\mathbf{\Theta} = \{u_{i_1},\ldots,u_{i_s}\}$ is a minimal diagnosable class. Therefore, the pivot position of $\mathbf{v}_\rho$ is at n-s+1, i.e., $\boldsymbol{\theta}[\mathbf{v}_\rho] \subseteq \mathbf{\Theta}$. Since $\boldsymbol{\theta}[\mathbf{v}_\rho]$ and $\mathbf{\Theta}$ are both minimal, $\boldsymbol{\theta}[\mathbf{v}_\rho] = \mathbf{\Theta}$.

A5. Proof of Corollary 3

From Corollary 2, it is clear that a minimal complementary class should contain exactly n-ρ faults. Assume that a minimal complementary class contains a minimal diagnosable class that includes n_1 faults. Since a minimal diagnosable class is diagnosable, we only need to know n_1-1 faults in the minimal diagnosable class to identify all the n_1 faults. Then, the number of faults in the minimal complementary class can be reduced by 1. Thus, a fault class is a minimal complementary class only if it does not contain any minimal diagnosable class. Now we need to prove that if a fault class with n-ρ elements does not include any minimal diagnosable class, it is a minimal complementary class. Assume that a fault class $\{u_{i_1}...u_{i_{n-\rho}}\}$ does not contain any minimal diagnosable class. Consider the RREF of the permuted matrix $\mathbf{G}'^T$ corresponding to the fault permutation $i_{n-\rho+1}\ldots i_n i_1 \ldots i_{n-\rho}$. Since $u_{i_1}...u_{i_{n-\rho}}$ do not include any minimal diagnosable class, the last n-ρ columns of the RREF should not include any pivot positions according to Corollary 1. However, since there are total ρ pivot positions, every column among the first ρ columns of the RREF should contain only a "leading 1". Hence, it is clear that all the faults can be uniquely identified if the n-ρ faults that correspond to the last n-ρ columns are known.

Acknowledgement

The authors gratefully acknowledge the financial support of the NSF Engineering Research Center for Reconfigurable Machining Systems (NSF Grant EEC95-92125), the NSF Award DMI-0217481, and the NSF CAREER Award DMI-9624402. The authors also appreciate the editor and

referees for their valuable comments and suggestions.

Reference

[1] Agrawal, R., Lawless, J.F. and Mackay, R.J. (1999), "Analysis of Variation Transmission in Manufacturing Processes – Part II," *Journal of Quality Technology*, 31, 143 – 154.

[2] Apley, D.W., and Shi, J. (2001), "A Factor-Analysis Methods for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, 43, 84-95.

[3] Apley, D.W., and Shi, J. (1998), "Diagnosis of Multiple Fixture Faults in Panel Assembly," ASME *Journal of Manufacturing Science and Engineering*, 120, 793-801.

[4] Camelio, A.J., Hu, S.J., and Ceglarek, D.J. (2001), "Modeling Variation Propagation of Multi-station Assembly Systems with Compliant Parts," *Proceedings of the 2001 ASME Design Engineering Technical Conference, Pittsburgh*, PA, September 9-12.

[5] Ceglarek, D., and Shi. J. (1996), "Fixture Failure Diagnosis for Autobody Assembly Using Pattern Recognition," ASME *Journal of Engineering for Industry*, 188, 55-65.

[6] Chang, M., and Gossard, D.C. (1998), "Computational method for diagnosis of variation-related assembly problem," *International Journal of Production Research*, 36, 2985-2995.

[7] Ding, Y., Ceglarek, D., and Shi, J. (2000), "Modeling and Diagnosis of Multistage Manufacturing Processes: Part I State Space Model," *Proceedings of the 2000 Japan/USA Symposium on Flexible Automation*, July 23-26, Ann Arbor, MI, 2000JUSFA-13146.

[8] Ding, Y., Shi, J., and Ceglarek, D. (2002), "Diagnosability Analysis of Multistage Manufacturing Processes," ASME *Journal of Dynamic Systems, Measurement and Control*, 124, 1-13.

[9] Djurdjanovic, D., and Ni, J. (2001), "Linear State Space Modeling of Dimensional Machining Errors," *Transactions of NAMRI/SME,* XXIX, 541-548.

[10] Halevi G. and Weill, R. D., (1995), "*Principles of Process Planning: a Logical Approach*," New York: Chapman & Hall.

[11] Jin, J., and Shi, J. (1999), "State Space Modeling of Sheet Metal Assembly for Dimensional Control," ASME *Journal of Manufacturing Science and Engineering*, 121, 756-762.

[12] Lawless, J.F., Mackay, R. J. and Robinson, J.A. (1999), "Analysis of Variation Transmission in Manufacturing Processes – Part I," *Journal of Quality Technology*, 31, 131 – 142.

[13] Lay, D.C. (1997), *Linear Algebra and Its Applications* (2nd ed.), New York: Addison-Wesley.

[14] Mantripragada, R., and Whitney, D.E. (1999), "Modeling and Controlling Variation Propagation in Mechanical Assemblies Using State Transition Models," *IEEE Transactions On Robotics and Automation*, 15, 124 – 140.

[15] Montgomery, D.C., and Woodall, W.H. (ed.) (1997), "A Discussion on Statistically-Based Process Monitoring and Control," *Journal of Quality Technology*, 29, 121-162.

[16] Rao, C.R., and Kleffe, J. (1988), *Estimation of Variance Components and Applications*, Amsterdam: North-Holland.

[17] Rong, Q., Ceglarek, D., and Shi, J. (2000), "Dimensional Fault Diagnosis for Compliant Beam Structure Assemblies," ASME *Journal of Manufacturing Science and Engineering*, 122, 773-780.

[18] Schott, J. R., (1997), *Matrix Analysis for Statistics*, New York: John Wiley & Sons.

[19] Searle, S.R., Casella, G., and McCulloch, C.E. (1992), *Variance Components*, New York: John Wiley & Sons.

[20] Woodall, W.H., and Montgomery, D.C. (1999), "Research Issues and Ideas in Statistical Process Control," *Journal of Quality Technology*, 31, 376-386.

[21] Zhou, S., Huang, Q., and Shi, J. (2002), "State Space Modeling for Dimensional Monitoring of Multistage Machining Process Using Differential Motion Vector," Accepted by *IEEE Transactions on Robotics and Automation*.