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

Shiyu ZHOU

Department of Industrial Engineering
University of Wisconsin
Madison, WI 53706
(*szhou@engr.wisc.edu*)

Yu DING

Department of Industrial Engineering
Texas A&M University
College Station, TX 77843
(*yuding@iemail.tamu.edu*)

Yong CHEN

Department of Mechanical and Industrial Engineering
The University of Iowa
Iowa City, IA 52242
(*yong-chen@uiowa.edu*)

Jianjun SHI

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109
(*shihang@umich.edu*)

Automatic in-process data collection techniques have been widely used in complicated manufacturing processes in recent years. The huge amounts of product measurement data have created great opportunities 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-effects 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 needed 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.

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 (OCMMs) 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 ensure 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; Woodall and Montgomery 1999) 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 (e.g., cutting tool, fixture, 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 two-dimensional panel assembly process (Fig. 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, and 4 in Fig. 1) and inspect the assembly: parts 1 and 2 are assembled at station I, subassembly "$1+2$" is assembled with parts 3 and 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 four-way locator, which controls motion in both x- and z-directions, and a two-way locator, which controls motion only in the z-direction. A subassembly with several parts also needs a four-way locator and a two-way locator to completely control its degrees of freedom. The active locating points are marked as P_i, $i = 1, \dots, 8$, 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 method is to

DOI 10.1198/004017003000000131

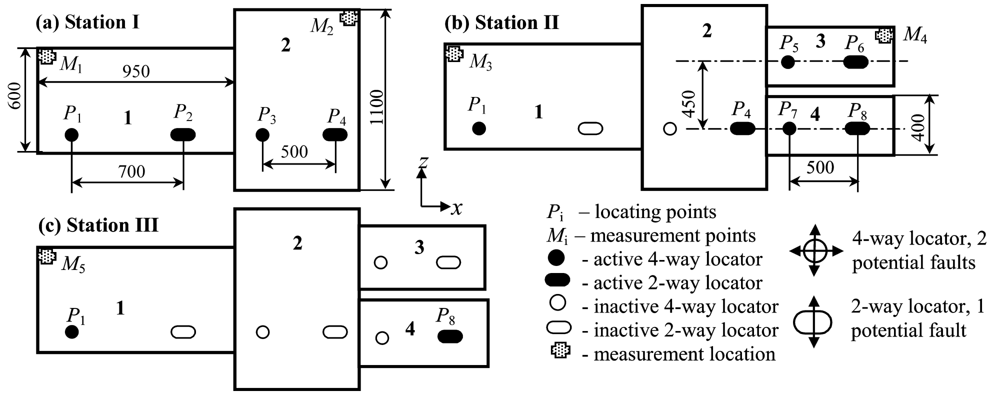

Figure 1. A Multistage Two-Dimensional Panel Assembly Process.

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 $\{M_i,\ i = 1, \ldots, 5\}$ in Figure 1.

Measurements from M_1 to M_5 contain information regarding the accuracy of fixture locators, offering the possibility of diagnosing locators' failure (i.e., process fault). However, the diagnosis of failing locators is not obvious, because 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.

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 et al. 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 et al. 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 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 discuss 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, 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, that is, if process faults are diagnosable. In the abovementioned work, the diagnosability condition is implicitly specified in the preconditions required by specific diagnosis algorithms. No explicit discussion on diagnosability under a general framework was given in those articles.

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, then 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 article is a special case of the diagnosability analysis presented in the present article. This relationship is clarified in Section 3. Furthermore, their article does not expose the "aliasing" fault structure for coupled faults in a partially diagnosable system, which is another focus of the present article.

This article 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

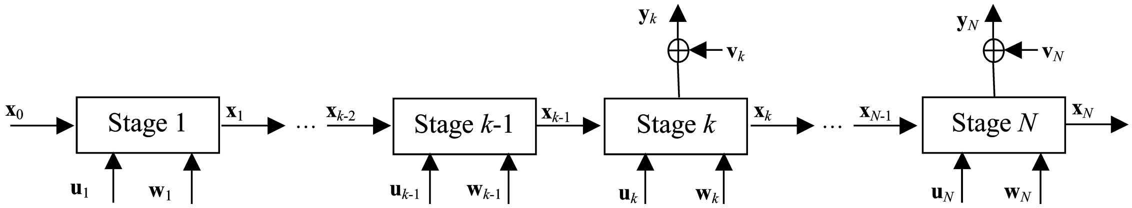

Figure 2. Diagram of a Multistage Manufacturing Process.

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. These criteria benchmark the amount and the "quality" of information obtained through a gauging system, as well as the flexibility of the gauging system.

The article 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 revisited in Section 4, together with another industrial case study, to illustrate the methodology. Conclusions are presented in Section 5.

2. FORMULATION OF THE 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. Several linear fault–quality models are 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 (2003) and Djurdjanovic and Ni (2001) provided linear fault–quality diagnostic models for multistage machining processes. All of these models are mechanism models, based on the physical laws of the processes. Lawless, Mackay, and Robinson (1999) and Agrawal, Lawless, and Mackay (1999) used 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 of the aforementioned models adopt the same model structure, 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 illustrates a manufacturing process with N stages. Variable k is the stage index. At the kth stage, several variables are involved in the variation propagation model: (1) the product quality information (e.g., part-dimensional deviations) at each stage, represented by the state vector $\mathbf{x}_k \in \Re^{n_x \times 1}$, where n_x is the dimension of $\mathbf{x}_k$; (2) the process variance sources (i.e., the process faults, such as the fixturing error, the machining error, and the thermal error), included as the input $\mathbf{u}_k \in \Re^{d_k \times 1}$, where d_k is the number of process variation sources at stage k; (3) background process noises and unmodeled errors, represented by $\mathbf{w}_k \in \Re^{n_x \times 1}$; (4) the product quality measurements, denoted by $\mathbf{y}_k \in \Re^{q_k \times 1}$, where q_k is the number of measurement features at stage k; (5) the measurement noise, denoted as a random vector $\mathbf{v}_k \in \Re^{q_k \times 1}$.

In this article we assume the following independence relationships. We assume that elements in $\mathbf{u}_k$ are independent to each other and also independent to any element in $\mathbf{u}_l$, $\forall l \neq k$, and assume the same independence relationship for $\mathbf{w}_k$ and $\mathbf{v}_k$. We further assume that elements in $\mathbf{u}_k$ are independent to elements in two other vectors, $\mathbf{w}_l$ and $\mathbf{v}_j$, $\forall k, l, j$, and likewise the same independent relationship for elements in $\mathbf{w}_l$ and $\mathbf{v}_j$. For $\mathbf{w}_k$, we assume that it is a zero-mean vector. For $\mathbf{v}_k$, we assume that elements in $\mathbf{v}_k$ are zero mean and have equal variance σ_v^2, that is, $\mathrm{cov}(\mathbf{v}_k) = \sigma_v^2 \mathbf{I}_{q_k}$, where $\mathrm{cov}(\cdot)$ represents the covariance matrix of a random vector, σ_v^2 is the variance of measurement noise, and $\mathbf{I}_{q_k}$ is a $q_k \times q_k$ identity matrix. Please note that here we assume that σ_v^2 is a constant for all stations; that is, measurement noise terms on all stations have the same variance. This assumption is reasonable if we use the same measurement devices on all measurement stages.

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

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

where $k = 1, 2, \ldots, N$, $\mathbf{A}_{k-1}\mathbf{x}_{k-1}$ 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, 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

$$\mathbf{y}_k = \sum_{i=1}^{k} \mathbf{C}_k \boldsymbol{\Phi}_{k,i} \mathbf{B}_i \mathbf{u}_i + \mathbf{C}_k \boldsymbol{\Phi}_{k,0} \mathbf{x}_0 + \sum_{i=1}^{k} \mathbf{C}_k \boldsymbol{\Phi}_{k,i} \mathbf{w}_i + \mathbf{v}_k, \quad (2)$$

where $\boldsymbol{\Phi}_{k,i} = \mathbf{A}_{k-1}\mathbf{A}_{k-2}\cdots\mathbf{A}_i$ for $k > i$ and $\boldsymbol{\Phi}_{k,k} = \mathbf{I}_{n_x}$. 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, then $\mathbf{C}_k\boldsymbol{\Phi}_{k,0}\mathbf{x}_0$ can be moved to the left side of (2), and the difference $\mathbf{y}_k - \mathbf{C}_k\boldsymbol{\Phi}_{k,0}\mathbf{x}_0$ can then be treated as a new measurement. If the measurement of $\mathbf{x}_0$ is not available, then it can be treated as an additional process fault input. Without loss of generality, we set $\mathbf{x}_0$ to $\mathbf{0}$.

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

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

where

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

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

$\boldsymbol{\mu}_k$ is an unknown constant vector, and $\tilde{\mathbf{u}}_k$, $\mathbf{w}_k$, and $\mathbf{v}_k$ are zero-mean random vectors. Because $\mathbf{y}_k$ is not necessarily available at every stage, if no measurement is available at station k, then the corresponding rows should be eliminated.

Let P denote the total number of potential faults (i.e., the length of $[\boldsymbol{\mu}_1^T \ \ldots \ \boldsymbol{\mu}_k^T \ \ldots \ \boldsymbol{\mu}_N^T]^T$) and let Q denote the number of system noises (i.e., the length of $[\mathbf{w}_1^T \ \ldots \ \mathbf{w}_k^T \ \ldots \ \mathbf{w}_N^T]^T$) considered on all of the stages. That is, $P = \sum_{k=1}^N d_k$ and $Q = N \cdot n_x$. We use the lower-case u_i, $i = 1, \ldots, P$, to represent the ith coordinate of the vector of $[\tilde{\mathbf{u}}_1^T \ \ldots \ \tilde{\mathbf{u}}_k^T \ \ldots \ \tilde{\mathbf{u}}_N^T]^T$ and denote u_i's variance as $\sigma_{u_i}^2$. Similarly, we use the lower-case w_i to represent the ith coordinate of the vector of $[\mathbf{w}_1^T \ \ldots \ \mathbf{w}_k^T \ \ldots \ \mathbf{w}_N^T]^T$ and denote w_i's variance as $\sigma_{w_i}^2$. With this notation, the variance components of process faults, the variance components of system noises, and the variance of measurement noises are represented by $\{\sigma_{u_i}^2\}_{i=1,\ldots,P}$, $\{\sigma_{w_i}^2\}_{i=1,\ldots,Q}$, and σ_v^2.

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

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

where $\mathbf{U}^T = [\boldsymbol{\mu}_1^T \ \ldots \ \boldsymbol{\mu}_k^T \ \ldots \ \boldsymbol{\mu}_N^T]$, $\mathbf{Y}^T = [\mathbf{Y}_1^T \ \ldots \ \mathbf{Y}_i^T \ \ldots \ \mathbf{Y}_M^T]$, $\mathbf{Y}_i^T = [\mathbf{y}_{1i}^T \ \ldots \ \mathbf{y}_{ki}^T \ \ldots \ \mathbf{y}_{Ni}^T]$ is the ith 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 (Schott 1997), $\mathbf{1}_M$ is the summing vector whose M elements equal unity. Letting "$:j$" represent the jth column of a matrix, (4) can be reorganized as

$$\mathbf{Y} = (\mathbf{1}_M \otimes \boldsymbol{\Gamma})\mathbf{U} + \sum_{j=1}^{P}(\mathbf{I}_M \otimes \boldsymbol{\Gamma}_{:j})\tilde{\mathbf{U}}(j) + \sum_{j=1}^{Q}(\mathbf{I}_M \otimes \boldsymbol{\Psi}_{:j})\mathbf{W}(j) + \mathbf{V}, \quad (5)$$

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

The process faults manifest themselves as the mean deviation and variance of $\mathbf{u}_k$. The diagnosability problem can then be restated: From M samples, can we identify the value of $\{\mu_i\}_{i=1,\ldots,P}$ and $\{\sigma_{u_i}^2\}_{i=1,\ldots,P}$? In the following section, this problem is studied using the framework of VCA.

3. DIAGNOSABILITY ANALYSIS FOR MULTISTAGE MANUFACTURING PROCESSES

3.1 Definition of Fault Diagnosability

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

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

where $\mathbf{y}$ is an $n_y \times 1$ observation vector; $\mathbf{X}$ is an $n_y \times l_x$ known constant matrix, $l_x \le n_y$; $\boldsymbol{\alpha}$ is an $l_x \times 1$ vector of unknown constants; $\boldsymbol{\xi}_i$ is an $n_y \times m_i$ known constant matrix, $m_i \le n_y$; $\mathbf{b}_i$ is an $m_i \times 1$ vector of independent variables with mean 0 and unknown variance σ_i^2; $\mathbf{e}$ is an $n_y \times 1$ vector of independent variables with mean 0 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 $\mathbf{X}$ and $\{\boldsymbol{\xi}_i\}_{i=1,\ldots,c}$ are determined by designers. They often contain only 0's or 1's, depending on whether the relevant effect contributes to the measurement. Given a mixed model, researchers are interested primarily in estimating the fixed effects and variance components. A large body of literature about VCA is available; excellent overviews have been given by 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 [eq. (5)] and those in the mixed model [eq. (6)]. In our fault diagnosis problem, however, the matrices $\mathbf{X}$, $\{\boldsymbol{\xi}_i\}_{i=1,\ldots,c}$ are computed from system matrices $\mathbf{A}_k$, $\mathbf{B}_k$, and $\mathbf{C}_k$, $k = 1, \ldots, N$, which are determined by the process design information and measurement deployment information. The fixed effects are the mean values ($\boldsymbol{\mu}_k$) of process faults, and the random effects are the process faults and the process noises, $\tilde{\mathbf{u}}_k$, $\mathbf{w}_k$, and $\mathbf{v}_k$. 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 (5), we have

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

and

$$\operatorname{cov}(\mathbf{Y}) = \mathbf{F}_1\sigma_{u_1}^2 + \cdots + \mathbf{F}_P\sigma_{u_P}^2 + \mathbf{F}_{P+1}\sigma_{w_1}^2 + \cdots + \mathbf{F}_{P+Q}\sigma_{w_Q}^2 + \mathbf{F}_{P+Q+1}\sigma_v^2 \quad (8)$$

where $E(\cdot)$ represents the expectation,

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

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

Define $[\sigma_{u_1}^2 \ \ldots \ \sigma_{u_P}^2 \ \sigma_{w_1}^2 \ \ldots \ \sigma_{w_Q}^2 \ \sigma_v^2]^T$ in (8) as $\boldsymbol{\theta}$, E^U as the space containing all possible values of $\mathbf{U}$, 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" of 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, $\forall\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2\in E^U$,

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

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

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

Remark 1. In model (5), we are concerned only about the mean and variance of process faults. Therefore, only the first- and second-order moments are considered in the definition.

Remark 2. The foregoing 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 $\mathbf{Y}$. This definition does not depend on any specific diagnosis algorithm.

Remark 3. By selecting different $\mathbf{p}$ and $\mathbf{f}$, the diagnosability of different fault combinations can be evaluated. For example, by selecting $\mathbf{p}$ or $\mathbf{f}=[1\ 0\ \ldots\ 0]^T$, we can check whether the mean or variance of the first fault is diagnosable. If it is, then 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 is given in Appendix A.2.

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

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

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

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

where L is the length of $[\mathbf{y}_1^T\ \mathbf{y}_2^T\ \ldots\ \mathbf{y}_N^T]^T$ in (3), that is, $L=\sum_{k=1}^N q_k$.

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

For the foregoing criterion, the diagnosability of the variance of process fault includes the effects of the modeling error $\mathbf{w}$ and the observation noise $\mathbf{v}$. This means that even if a fault can be distinguished from other faults, it still can be nonuniquely 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 to be small or their variance can be estimated from the normal working condition of a manufacturing process, then 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 $[\sigma_{u_1}^2\ \ldots\ \sigma_{u_P}^2]$ and reducing the $\mathbf{H}$ matrix in Theorem 1 to $\mathbf{H}_r$, where $\mathbf{H}_r$ is a subblock of $\mathbf{H}$, that is,

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

Remark 4. Under the setting where noises $\mathbf{w}$ and $\mathbf{v}$ are assumed to be negligible, the diagnosability matrix was defined by Ding et al. (2002) as $\pi(\boldsymbol{\Gamma})$, where $\pi(\cdot)$ is a matrix transformation that they defined. The variances of process faults are considered fully diagnosable if and only if $\pi(\boldsymbol{\Gamma})^T\pi(\boldsymbol{\Gamma})$ is of full rank. In fact, this condition is the same as what we have derived in the present article. It can be shown that $R(\mathbf{H}_r)=R(\pi(\boldsymbol{\Gamma})^T\pi(\boldsymbol{\Gamma}))$. Therefore, the work of Ding et al. (2002) can be considered a special case of the general framework presented in this article.

Remark 5. If noise terms are not included, Ding et al. (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 processes given in Section 4.

Theorem 1 alone is not very effective in analyzing a partially diagnosable system where not all faults are diagnosable. The other faults that we need to know before we can identify a nonuniquely diagnosable fault are not obvious from the theorem. 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 $\{u_{i_1}\ \ldots\ u_{i_n}\}$ forms a *mean* or *variance diagnosable class* if a nontrivial linear combination of their means $\{\mu_{i_1}\ \ldots\ \mu_{i_n}\}$ or variances $\{\sigma_{i_1}^2\ \ldots\ \sigma_{i_n}^2\}$ is diagnosable. ("Nontrivial" means that at least one coefficient of the linear combination is nonzero.)

Definition 3. A nonempty set of n faults $\{u_{i_1}\ \ldots\ u_{i_n}\}$ forms a *minimal mean* or *variance diagnosable class* if no strict subset of $\{u_{i_1}\ \ldots\ u_{i_n}\}$ 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 being the testing matrix, $\mathbf{\Gamma}^T$ for mean and $\mathbf{H}$ for variance). Hence no distinction between mean or variance diagnosability is made hereafter unless otherwise indicated.

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

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 which is given in Appendix A.3.

Theorem 2. Given a testing matrix $\mathbf{G} \in \Re^{n\times m}$ (where $\mathbf{G}$ is $\mathbf{\Gamma}^T$ or $\mathbf{H}$) and n faults $\boldsymbol{\theta} = [u_1 \ \dots \ u_n]^T$ 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 ith element of $\boldsymbol{\theta}$) $\in \boldsymbol{\theta}[\mathbf{v}]$ if $\mathbf{v}(i)$ (the ith 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 rearranging the columns in $\mathbf{G}^T$, we can obtain all of the possible minimal diagnosable classes. (The proof is given in Appendix A.4.) The rearranging process is known as *matrix permutation*. The *permuted matrix* is defined as follows: If $\{\mathbf{c}_i\}_{i=1,\dots,n}$ denote the column vectors of $\mathbf{G}^T$ and correspond to the faults $\boldsymbol{\theta} = [u_1 \ \dots \ u_n]^T$, then the columnwise permuted matrix $\mathbf{G}'^T = [\mathbf{c}_{i_1} \ \dots \ \mathbf{c}_{i_n}]$ is called the permuted matrix corresponding to the fault permutation $\boldsymbol{\theta}' = [u_{i_1} \ \dots \ u_{i_n}]^T$.

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

Corollary 1 shows 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 that 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 (Schott 1997). In other words, $\mathbf{v}_i$ does not share any common nonzero column positions with $\mathbf{v}_j$. Define $\boldsymbol{\theta}[C]$ 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. However, if a_1 and a_2 are both nonzero, then the fact 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 only by permuting the faults within $\boldsymbol{\theta}[C_1]$ and $\boldsymbol{\theta}[C_2]$. 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 undividable set of rows, then $\boldsymbol{\theta}[C_i]$ 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, then the computational load required to find all minimal diagnosable classes can be reduced significantly. The worst case is that all faults are connected in a big fault class. However, this 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 demonstrate in the case studies in Section 4.

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

1. Calculate the RREF of $\mathbf{G}^T$.
2. Remove all of 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.
3. Find the connected fault classes based on the RREF of $\mathbf{G}^T$.
4. Permute the columns within the connected fault classes and obtain the minimal diagnosable classes based on the permuted matrices until all of 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, then it is uniquely diagnosable. If a fault is not uniquely diagnosable and it forms a minimal diagnosable class with several other faults, then 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 or, 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 gauging systems shown in Figures 1 and 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 section.

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

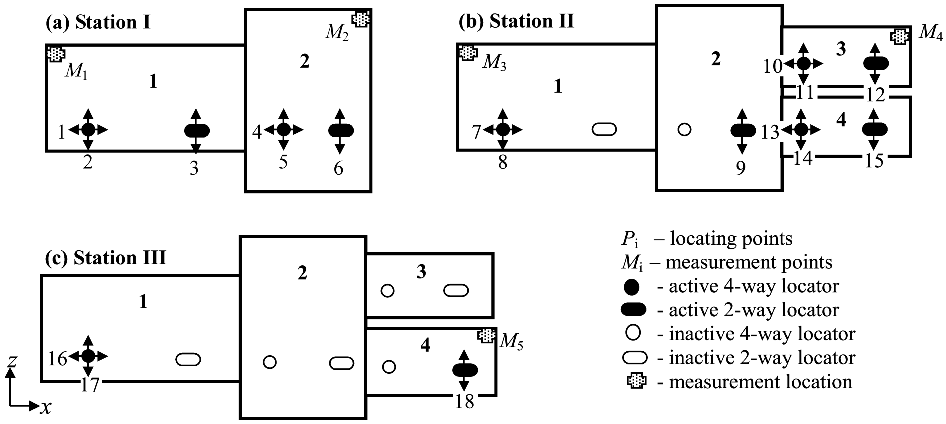

Figure 3. Gauging System 2 for the Multistage Two-Dimensional Panel Assembly Process.

through the gauging system. We propose three criteria for evaluating gauging systems: information *quantity*, information *quality*, and system *flexibility*.

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 $\boldsymbol{\theta} = [u_1 \ \dots \ u_n]^T$ corresponding to $\mathbf{G}$, if the rank of $\mathbf{G}$ is ρ, then $n - \rho$ faults need to be known to uniquely identify all n faults.

The proof is omitted here, it uses 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 $\mathbf{A}$. In such a linear system, $n - \rho$ free variables need to be known before uniquely solving $\mathbf{X}$, where n is the dimension of $\mathbf{X}$, $\rho = \text{rank}(\mathbf{A})$. If we consider the testing matrix $\mathbf{G}$ as if it is in a similar situation to matrix $\mathbf{A}$, then 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 earlier, the detailed information content could be very different. In practice, it is always desirable to have unique identification of a fault so that corrective action can be undertaken immediately eliminate the fault and restore the system to its normal condition. The decision on what corrective action to take 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 minimal diagnosable classes that contain only a single fault.

The third criterion is the flexibility provided by the current gauging system toward achieving full diagnosability. Some gauging systems can be rigid in the sense that certain faults or fault combinations, which may be difficult to measure in practice, must be known to achieve a fully diagnosable system. Other gauging systems may provide information in a flexible way; for example, 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, then all of the systems faults can be uniquely identified. Consider a system with four faults and three minimal diagnosable classes, $\{u_1, u_2\}$, $\{u_1, u_3, u_4\}$, 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 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 numerous faults and intricate fault combinations. Corollary 3 facilitates the determination of minimal complementary classes; its proof is given in Appendix A.5.

Corollary 3. A set of faults forms a minimal complementary class if and only if the set contains $n - \rho$ faults but does not contain any minimal diagnosable classes, 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 $n - \rho$ faults. If the entire fault set can be partitioned into many smaller distinct, connected fault classes, then the task of searching the complete minimal complementary classes can be further 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 nondiagnosable 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 that provides 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 of 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 Figures 1 and 3. The product state variable $\mathbf{x}_k$ is denoted by random deviations associated with the degrees of freedom (df) of each part. Each two-dimensional part in this example has three df (two translational and one rotational) and the size of $\mathbf{x}_k$ is 12×1 (i.e., $n_x = 12$) given that there are four parts. The state vector $\mathbf{x}_k$ is expressed as

$$\mathbf{x}_k = [\delta x_{1,k}\ \delta z_{1,k}\ \delta\alpha_{1,k} \mid \ldots \mid \ldots \mid \delta x_{4,k}\ \delta z_{4,k}\ \delta\alpha_{4,k}]^T, \quad (13)$$

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. 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 0's.

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 $\mathbf{u}_1 = [\delta p_1\ \ldots\ \delta p_6]^T$, $\mathbf{u}_2 = [\delta p_7\ \ldots\ \delta p_{15}]^T$, and $\mathbf{u}_3 = [\delta p_{16}\ \delta p_{17}\ \delta p_{18}]^T$, where δp_i is the deviation associated with fault i and the dimensions are $d_1 = 6$, $d_2 = 9$, $d_3 = 3$, and $P = 6 + 9 + 3 = 18$.

The measurement $\mathbf{y}$ contains positional deviations detected at M_i, $i = 1, \ldots, 5$. In this 2-D case, each M_i can deviate in x and/or z directions. Hence, $\mathbf{y}_1 = [\delta M_1(x)\ \delta M_1(z)\ \delta M_2(x)\ \delta M_2(z)]^T$, $\mathbf{y}_2 = [\delta M_3(x)\ \delta M_3(z)\ \delta M_4(x)\ \delta M_4(z)]^T$, and $\mathbf{y}_3 = [\delta M_5(x)\ \delta M_5(z)]^T$. The dimensions are $q_1 = 4$, $q_2 = 4$, $q_3 = 2$, and $L = 4 + 4 + 2 = 10$.

The state-space representation of this process is

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

and

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{v}_k, \qquad k = 1, 2, 3. \quad (15)$$

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 by Jin and Shi (1999) and Ding et al. (2000), we give the numerical expressions of $\mathbf{A}$'s, $\mathbf{B}$'s, and $\mathbf{C}$'s of the assembly processes shown in Figures 1 and 3. The $\mathbf{A}$'s, $\mathbf{B}$'s, $\mathbf{C}_1$, and $\mathbf{C}_2$ are the same for these two processes, because their fixture layouts are the same and the sensor deployments are the same for stations I and II.

$$\mathbf{A}_1 = \left[\begin{array}{cccccc|c}
0 & 0 & 0 & 0 & 0 & 0 & \\
0 & 0 & 0 & 0 & 0 & 0 & \\
0 & .0007 & 1 & 0 & -.0007 & -.3497 & \mathbf{0}^{6\times 6} \\
-1 & 0 & 0 & 1 & 0 & 0 & \\
0 & -.3497 & 0 & 0 & .3497 & -325.17 & \\
0 & .0007 & 0 & 0 & -.0007 & .6503 & \\
\hline
\multicolumn{6}{c|}{\mathbf{0}^{6\times 6}} & \mathbf{I}^{6\times 6}
\end{array}\right]_{12\times 12},$$

$$\mathbf{A}_2 = \left[\begin{array}{ccc|c|ccc}
0 & 0 & 0 & & 0 & 0 & 0 \\
0 & 0 & 0 & \mathbf{0}^{3\times 6} & 0 & 0 & 0 \\
0 & .0005 & 1 & & 0 & -.0005 & -.2392 \\
\hline
-1 & 0 & 0 & & 0 & 0 & 0 \\
0 & -.5550 & 0 & & 0 & -.4450 & -222.49 \\
0 & .0005 & 0 & \mathbf{I}^{6\times 6} & 0 & -.0005 & -.2392 \\
-1 & -.2153 & 0 & & 0 & .2153 & 107.655 \\
0 & -.2392 & 0 & & 0 & -.7608 & -380.38 \\
0 & .0005 & 0 & & 0 & -.0005 & -.2392 \\
\hline
-1 & 0 & 0 & & 1 & -.0005 & 0 \\
0 & -.2392 & 0 & \mathbf{0}^{3\times 6} & 0 & .2392 & -380.38 \\
0 & .0005 & 0 & & 0 & -.0005 & .7608
\end{array}\right]_{12\times 12}, \quad (16)$$

$$\mathbf{B}_1 = \left[\begin{array}{cccccc}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & -.0014 & .0014 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & -.002 & .002 \\
\hline
\multicolumn{6}{c}{\mathbf{0}^{6\times 6}}
\end{array}\right]_{12\times 6},$$

$$\mathbf{B}_2 = \left[\begin{array}{ccc|cccccc}
1 & 0 & 0 & & & & & & \\
0 & 1 & 0 & & & & & & \\
0 & -.0007 & .0007 & \multicolumn{6}{c}{\mathbf{0}^{6\times 6}} \\
1 & 0 & 0 & & & & & & \\
0 & .3497 & .6503 & & & & & & \\
0 & -.0007 & .0007 & & & & & & \\
\hline
 & & & 1 & 0 & 0 & 0 & 0 & 0 \\
 & & & 0 & 1 & 0 & 0 & 0 & 0 \\
\multicolumn{3}{c|}{\mathbf{0}^{6\times 3}} & 0 & -.002 & .002 & 0 & 0 & 0 \\
 & & & 0 & 0 & 0 & 1 & 0 & 0 \\
 & & & 0 & 0 & 0 & 0 & 1 & 0 \\
 & & & 0 & 0 & 0 & 0 & -.002 & .002
\end{array}\right]_{12\times 9},$$

$$\mathbf{B}_3 = \left[\begin{array}{ccc}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & -.0005 & .0005 \\
1 & 0 & 0 \\
0 & .5550 & .4450 \\
0 & -.0005 & .0005 \\
1 & .2153 & -.2153 \\
0 & .2392 & .7608 \\
0 & -.0005 & .0005 \\
1 & 0 & 0 \\
0 & .2392 & .7608 \\
0 & -.0005 & .0005
\end{array}\right]_{12\times 3}, \quad (17)$$

$$\mathbf{C}_1 = \left[\begin{array}{ccc|ccc|c} 1 & 0 & -550 & & \mathbf{0}^{2\times 3} & & \mathbf{0}^{2\times 6} \\ 0 & 1 & -110 & & & & \\ \hline & \mathbf{0}^{2\times 3} & & 1 & 0 & -550 & \mathbf{0}^{2\times 6} \\ & & & 0 & 1 & -630 & \end{array}\right]_{4\times 12},$$

$$\mathbf{C}_2 = \left[\begin{array}{ccc|c|ccc|c} 1 & 0 & -550 & \mathbf{0}^{2\times 3} & & \mathbf{0}^{2\times 3} & & \mathbf{0}^{2\times 3} \\ 0 & 1 & -100 & & & & & \\ \hline & \mathbf{0}^{2\times 3} & & \mathbf{0}^{2\times 3} & 1 & 0 & -300 & \mathbf{0}^{2\times 3} \\ & & & & 0 & 1 & -740 & \end{array}\right]_{4\times 12}. \tag{18}$$

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

$$\mathbf{C}_3^1 = \left[\begin{array}{ccc|c} 1 & 0 & -550 & \mathbf{0}^{2\times 9} \\ 0 & 1 & -100 & \end{array}\right]_{2\times 12}$$

and (19)

$$\mathbf{C}_3^2 = \left[\begin{array}{c|ccc} \mathbf{0}^{2\times 9} & 1 & 0 & -200 \\ & 0 & 1 & 620 \end{array}\right]_{2\times 12}.$$

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

$$\boldsymbol{\Gamma}^1 = \left[\begin{matrix} 1 & .786 & -.786 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.143 & -.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.1 & -1.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.26 & -1.26 & 0 & 0 & 0 & 0 \\ 0 & .401 & -.786 & 0 & 0 & .385 & 1 & .385 & -.385 & 0 \\ 0 & .073 & -.143 & 0 & 0 & .070 & 0 & 1.070 & -.070 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .401 & -.786 & 0 & 0 & .385 & 0 & .122 & -.385 & 0 \\ 0 & .073 & -.143 & 0 & 0 & .070 & 0 & .022 & -.070 & 0 \end{matrix}\right.$$
$$\left.\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .6 & -.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.48 & -1.48 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .263 & 1 & .263 & -.263 \\ 0 & 0 & 0 & 0 & .048 & 0 & 1.048 & -.048 \end{matrix}\right] \tag{20}$$

and

$$\boldsymbol{\Gamma}^2 = \left[\begin{matrix} 1 & .786 & -.786 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.143 & -.143 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1.1 & -1.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2.26 & -1.26 & 0 & 0 & 0 & 0 \\ 0 & .401 & -.786 & 0 & 0 & .385 & 1 & .385 & -.385 & 0 \\ 0 & .073 & -.143 & 0 & 0 & .070 & 0 & 1.070 & -.070 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -.096 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .057 & 0 & 0 \end{matrix}\right.$$
$$\left.\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .6 & -.6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2.48 & -1.48 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & .4 & -.304 & 1 & .096 & -.096 \\ 0 & 0 & 0 & -.24 & .183 & 0 & -.057 & 1.057 \end{matrix}\right], \tag{21}$$

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

$$\mathbf{H}_r^1 = \left[\begin{matrix} 1 & .617 & .617 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .617 & 5.089 & 2.050 & 0 & 0 & .102 & .161 & .08 & .102 & 0 \\ .617 & 2.050 & 3.661 & 0 & 0 & .390 & .617 & .307 & .390 & 0 \\ 0 & 0 & 0 & 1 & 1.21 & 1.21 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.21 & 39.91 & 16.46 & 0 & 0 & 0 & 0 \\ 0 & .102 & .390 & 1.21 & 16.46 & 9.63 & .148 & .073 & .093 & 0 \\ 0 & .161 & .617 & 0 & 0 & .148 & 1.0 & .148 & .148 & 0 \\ 0 & .08 & .307 & 0 & 0 & .073 & .148 & 1.711 & .073 & 0 \\ 0 & .102 & .390 & 0 & 0 & .093 & .148 & .073 & .093 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .36 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .36 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & .012 & .046 & 0 & 0 & .011 & 0 & .001 & .011 & 0 \\ 0 & .161 & .617 & 0 & 0 & .148 & 0 & .015 & .148 & 0 \\ 0 & .033 & .127 & 0 & 0 & .030 & 0 & .003 & .030 & 0 \\ 0 & .012 & .046 & 0 & 0 & .011 & 0 & .001 & .011 & 0 \end{matrix}\right.$$
$$\left.\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .012 & .161 & .033 & .012 \\ 0 & 0 & 0 & 0 & .046 & .617 & .127 & .046 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .011 & .148 & .030 & .011 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .001 & .015 & .003 & .001 \\ 0 & 0 & 0 & 0 & .011 & .148 & .03 & .011 \\ .36 & .36 & 0 & 0 & 0 & 0 & 0 & 0 \\ 42.39 & 16.24 & 0 & 0 & 0 & 0 & 0 & 0 \\ 16.24 & 6.505 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .005 & .069 & .014 & .005 \\ 0 & 0 & 0 & 0 & .069 & 1 & .069 & .069 \\ 0 & 0 & 0 & 0 & .014 & .069 & 1.362 & .014 \\ 0 & 0 & 0 & 0 & .005 & .069 & .014 & .005 \end{matrix}\right] \tag{22}$$

and

Table 1. Comparison of Gauging Systems 1 and 2

	Gauging system 1	*Gauging System 2*
RREF($\mathbf{H}_r$)	$\left[\begin{array}{c\|cccccc} \mathbf{I}^{12\times 12} & & & \mathbf{0}^{12\times 6} & & & \\ \hline \mathbf{0}^{6\times 12} & \begin{matrix} 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \end{array}\right]_{18\times 18}$	$\left[\begin{array}{c\|cccccc} \mathbf{I}^{12\times 12} & & & \mathbf{0}^{12\times 6} & & & \\ \hline \mathbf{0}^{6\times 12} & \begin{matrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & .58 & 0 & .06 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \end{array}\right]_{18\times 18}$
Number 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}
Number 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}
Number of minimal complementary classes	2	6

$$
\mathbf{H}_r^2 = \left[\begin{matrix}
1 & .617 & .617 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.617 & 4.367 & 1.224 & 0 & 0 & .025 & .161 & .054 & .025 & 0 \\
.617 & 1.224 & 1.627 & 0 & 0 & .098 & .617 & .207 & .098 & 0 \\
0 & 0 & 0 & 1 & 1.21 & 1.21 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1.21 & 39.91 & 16.46 & 0 & 0 & 0 & 0 \\
0 & .025 & .098 & 1.21 & 16.46 & 8.705 & .148 & .050 & .023 & 0 \\
0 & .161 & .617 & 0 & 0 & .148 & 4.0 & .231 & .148 & 0 \\
0 & .054 & .207 & 0 & 0 & .050 & .231 & 1.703 & .050 & 0 \\
0 & .025 & .098 & 0 & 0 & .023 & .148 & .05 & .023 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .36 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & .36 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & .009 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .16 & .003 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .093 & .002 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & .009 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .009 & .0002 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & .009 & .005 & 0 & 0
\end{matrix}\right.
$$

$$
\left.\begin{matrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & .16 & .093 & 1 & .009 & .009 \\
0 & 0 & .009 & .003 & .002 & .009 & .0002 & .005 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
.36 & .36 & 0 & 0 & 0 & 0 & 0 & 0 \\
42.39 & 16.24 & 0 & 0 & 0 & 0 & 0 & 0 \\
16.24 & 6.505 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & .16 & .093 & 1 & .009 & .009 \\
0 & 0 & .16 & .047 & .027 & .16 & .003 & .085 \\
0 & 0 & .093 & .027 & .016 & .093 & .002 & .049 \\
0 & 0 & 1 & .16 & .093 & 1 & .009 & .009 \\
0 & 0 & .009 & .003 & .002 & .009 & .0002 & .005 \\
0 & 0 & .009 & .085 & .049 & .009 & .005 & 1.271
\end{matrix}\right] \tag{23}
$$

The RREF of the $\mathbf{H}_r$'s and the corresponding fault structures are compared in Table 1. For gauging system 1, 14 rows have only 1 nonzero element, corresponding to 14 uniquely identified faults and hence minimal diagnosable classes, {1}, ..., {12}, {16}, {17}. Two faults (13, 14) correspond to 0 columns, and hence they are not diagnosable. The 13th row has two nonzero elements (i.e., $[\mathbf{0}^{1\times 12} \mid 0\ 0\ 1\ 0\ 0\ 1]$), indicating that {15, 18} is a minimal diagnosable class. The class {15, 18} is also a connected fault class, and because it is already minimal, no further permutation is needed. Similarly, for gauging system 2, there are 13 uniquely diagnosable classes, {1}, ..., {12}, {18}. Two minimal diagnosable classes, {13, 16} and {14, 15, 17}, correspond to the 13th and 14th rows. No permutation of $\mathbf{H}_r$ is needed for gauging system 2, either.

For gauging system 1, to achieve a fully diagnosable system, at least $n - \rho = 3$ faults must be known. We first search fault set {15, 18} with $n = 2$ and $\rho = 1$. It is clear that {15} and {18} are two minimal complementary fault classes for the connected fault class {15, 18}. Adding the nondiagnosable 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, to find the minimal complementary class, we search the faults among {13, 16} with $n = 2$ and $\rho = 1$ and among {14, 15, 17} with $n = 3$ and $\rho = 1$. 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, because their $\mathbf{H}_r$'s are not of full rank. Ranks of $\mathbf{H}_r$'s are the same ($\rho = 15$), 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, gauging system 2 can have six possible ways of measuring additional faults in achieving a fully diagnosable system, whereas gauging system 1 has only two possibilities. This difference indicates that gauging system 2 is more flexible. If the third criterion is in a higher priority, then gauging system 2 is more favorable.

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 first need to fix the location of the workpiece in the space. Figure 4 shows a widely used 3-2-1 fixturing setup. If we require that the workpiece touch all of the locating pads (L_1–L_3) and locating pins (P_1–P_3), then 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 Fig. 4) is called the "primary datum." Similarly, surface ADHE is the "secondary datum" and DCGH is the "tertiary datum" in Figure 4. Because the primary datum (surface ABCD) touches L_1–L_3, the translational motion in the z direction and the rotational motion in the x and y directions are restrained. Similarly, the secondary datum constrains the translational motion in the x direction and the rotational motion in the 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, then 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 workpiece quality. For example, if the primary datum (ABCD) is not perpendicular to the secondary datum (ADHE), then the milled slot also will not be perpendicular to the secondary datum.

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), the joint face, and the slot (S). The cover face, joint face, and the slot are milled at the first [Fig. 5(a)], second [Fig. 5(b)], and third [Fig. 5(c)] stages.

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

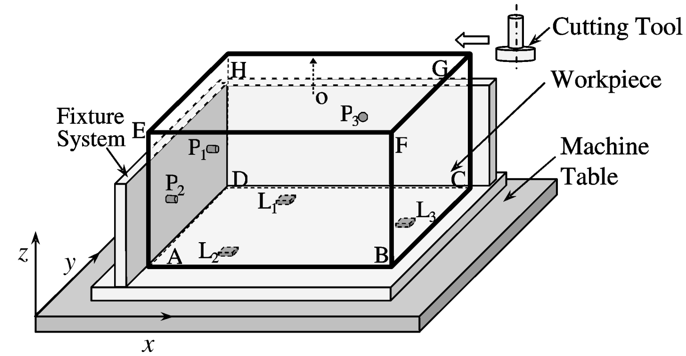

Figure 4. A Typical 3-2-1 Fixturing Configuration.

be obtained by a coordinate measuring machine (CMM). The state-space model [eq. (1)] can be obtained through a similar (to the foregoing panel assembly) but more complicated three-dimensional kinematics analysis, where $\mathbf{A}_{k-1}\mathbf{x}_{k-1}$ is the error contributed by the errors of datum features (with these features 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 have been given by Zhou et al. (2003).

After the model in (1) is obtained, the diagnosability study for the multistage machining process can be conducted following the theories proposed in Sections 2 and 3. We focus on the fixture error in this case study. For a 3-2-1 fixture setup, there are six 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 first stage, faults 7–12 represent locator errors at the second stage, and faults 13–18 represent locator errors at the third stage. Three gauging systems are used to measure slot S, cover face M, and the rough datum, where the rough datum is the primary datum at the first 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 systems 1 and 2 have a very simple structure. For the gauging system 3 (the fourth column in Table 2), the first three rows of RREF($\boldsymbol{\Gamma}$) 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($\boldsymbol{\Gamma}$), we can generate 15 minimal diagnosable classes (each with 4 faults) within this connected fault class, as shown in Table 2. The minimal complementary class of this connected fault class can be

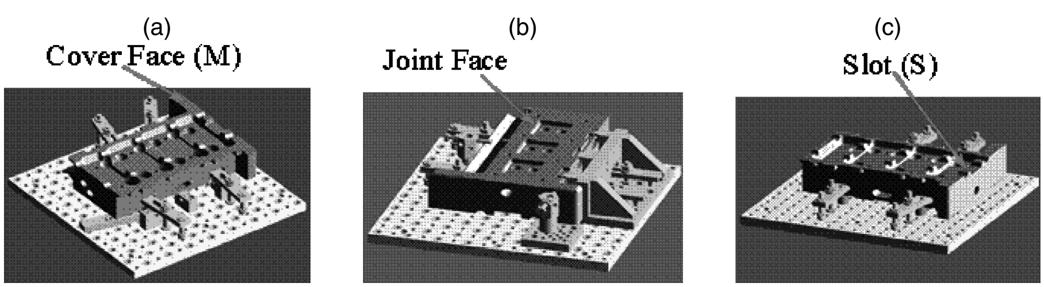

Figure 5. Process Layout at Three Stages.

Table 2. Comparison of Gauging Systems

Gauging system		*System 1 (slot S)*	*System 2 (cover face M)*	*System 3 (rough datum)*
Mean diagnosability: RREF($\boldsymbol{\Gamma}$)		$\left[\begin{array}{c\|c} \mathbf{0}^{6\times12} & \mathbf{I}^{6\times6} \\ \hline \mathbf{0}^{30\times12} & \mathbf{0}^{30\times6} \end{array}\right]_{36\times18}$	$\left[\begin{array}{cc\|c\|c} \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \mathbf{0}^{6\times6} \\ \hline \mathbf{0}^{30\times3} & \mathbf{0}^{30\times6} & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times6} \end{array}\right]_{36\times18}$	$\left[\begin{array}{cc\|c\|cc} \mathbf{I}^{3\times3} & \mathbf{0}^{3\times3} & \begin{matrix} -.63 & .53 & -.90 \\ .47 & -.57 & -.90 \\ -.49 & -.48 & -.04 \end{matrix} & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times6} \\ \hline \mathbf{0}^{3\times3} & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times3} & \mathbf{I}^{3\times3} & \mathbf{0}^{3\times6} \\ \mathbf{0}^{30\times3} & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times3} & \mathbf{0}^{30\times6} \end{array}\right]$
Variance diagnosability: RREF($\mathbf{H}_r$)		$\left[\begin{array}{c\|c} \mathbf{0}^{6\times12} & \mathbf{I}^{6\times6} \\ \hline \mathbf{0}^{12\times12} & \mathbf{0}^{12\times6} \end{array}\right]_{18\times18}$	$\left[\begin{array}{cc\|c\|c} \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{matrix} & \mathbf{0}^{6\times6} \\ \hline \mathbf{0}^{12\times3} & \mathbf{0}^{12\times6} & \mathbf{0}^{12\times3} & \mathbf{0}^{12\times6} \end{array}\right]_{18\times18}$	$\begin{bmatrix} \mathbf{I}^{3\times3} & \mathbf{0}^{3\times3} & \mathbf{0}^{3\times6} & \mathbf{0}^{3\times6} \\ \mathbf{0}^{6\times3} & \mathbf{0}^{6\times3} & \mathbf{I}^{6\times6} & \mathbf{0}^{6\times6} \\ \mathbf{0}^{9\times3} & \mathbf{0}^{9\times3} & \mathbf{0}^{9\times6} & \mathbf{0}^{9\times6} \end{bmatrix}$
Number 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}
Number of uniquely identified faults	Mean	6	3	3
	Variance	6	3	9
Number of minimal complementary classes	Mean	1	8	20
	Variance	1	8	1

found by searching the class with $n = 6$, $\rho = 3$. 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}, and {7, 8, 9}. Adding the nondiagnosable 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 although the faults {1, 2, 3, 7, 8, 9} form a connected fault class in regard to mean diagnosability, they are uniquely diagnosable regarding variance diagnosability. This verifies our previous remark that mean diagnosability requires a stronger condition than variance diagnosability.

5. CONCLUDING REMARKS

This article has reported on a study of the diagnosability of process faults given the product quality measurements in a complicated multistage manufacturing process. This study has revealed that the diagnosis capability that a gauging system can provide depends strongly on sensor deployment in a multistage manufacturing system. A poorly designed gauging system is likely to result in the loss of diagnosability. In contrast, a well-designed gauging system that achieves the desired level of diagnosability not only can monitor the process change, but also can 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 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 et al. 2000) or experimental data (Zhou et al. 2003; 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 systems, a numerous 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.

APPENDIX: PROOFS

A.1 Theorem 4.2.1 of Rao and Kleffe (1988)

Consider a general linear mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\boldsymbol{\beta}$ represents the fixed effects and $\boldsymbol{\epsilon}$ is mean 0 and $\mathrm{cov}(\boldsymbol{\epsilon}) = \theta_1\mathbf{V}_1 + \cdots + \theta_r\mathbf{V}_r$. Let $\boldsymbol{\theta} = [\theta_1 \ \ldots \ \theta_r]^T$ denote variance components. $\mathbf{p}^T\boldsymbol{\beta}$ is identifiable if and only if $\mathbf{p} \in R(\mathbf{X}^T)$, $\mathbf{f}^T\boldsymbol{\theta}$ is identifiable if and only if $\mathbf{f} \in R(\mathbf{H})$ and $\mathbf{H}' = (\mathrm{tr}(\mathbf{V}_i\mathbf{V}_j))$, $1 \le i \le r$, $1 \le j \le r$.

A.2 Proof of Theorem 1

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

A.3 Proof of Theorem 2

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

We also need to prove that $\boldsymbol{\theta}[\mathbf{v}_i]$ is a minimal diagnosable class. From the algorithm used to obtain the 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 Ξ. It is known that (a) given an $i \in \{1, \ldots, \rho\}$, there is only one nonzero element in $\{\mathbf{v}_i(j)\}_{j\in\Xi}$, and (b) 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 $i \in \Xi$. From (a), $\boldsymbol{\theta}[\mathbf{v}_i]$ must be in the form $\{u_{p_i}, u_{i_1}, \ldots, u_{i_k}\}$, $p_i \in \Xi$, $i_1, \ldots, i_k \notin \Xi$. Assuming 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 (b), if there is a j, then a_j is nonzero and u_{p_j} must be in $\boldsymbol{\theta}[\mathbf{v}'_i]$, where p_j is the pivot position of $\mathbf{v}_j$. Because $\boldsymbol{\theta}[\mathbf{v}_i]$ contains only 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.

A.4 Proof of Corollary 1

Let $\{\mathbf{v}_i\}_{i=1,\ldots,\rho}$ denote the nonzero row vectors of $\mathbf{G}'^T_r$. 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, then $\boldsymbol{\theta}[\mathbf{v}_\rho] \subset \boldsymbol{\Theta}$. According to Theorem 2, however, $\boldsymbol{\theta}[\mathbf{v}_\rho]$ is a diagnosable class, which contradicts the fact that $\boldsymbol{\Theta}$ is minimal. Second, assume that the pivot position of $\mathbf{v}_\rho$ is smaller than $n - s + 1$. If so, then a fault among $\{u_{i_{s+1}}, \ldots, u_{i_n}\}$ must belong to $\boldsymbol{\theta}[\mathbf{v}_\rho]$. Because the pivot position of $\mathbf{v}_\rho$ is the largest of all the pivot positions of $\{\mathbf{v}_i\}_{i=1,\ldots,\rho}$, given any vector $\mathbf{v}_f = \sum_{j=1}^{\rho} a_j\mathbf{v}_j$ (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 $\{u_{i_{s+1}}, \ldots, u_{i_n}\}$. According to Theorem 1, any diagnosable class should contain at least one element among $\{u_{i_{s+1}}, \ldots, u_{i_n}\}$, because $\mathbf{v}_f$ is an arbitrary vector in $\mathrm{row}(\mathbf{G}'^T_r)$. This contradicts the assertion that $\boldsymbol{\Theta} = \{u_{i_1}, \ldots, u_{i_s}\}$ is a minimal diagnosable class. Therefore, the pivot position of $\mathbf{v}_\rho$ is at $n - s + 1$, that is, $\boldsymbol{\theta}[\mathbf{v}_\rho] \subseteq \boldsymbol{\Theta}$. Because $\boldsymbol{\theta}[\mathbf{v}_\rho]$ and $\boldsymbol{\Theta}$ are both minimal, $\boldsymbol{\theta}[\mathbf{v}_\rho] = \boldsymbol{\Theta}$.

A.5 Proof of Corollary 3

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

ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support of National Science Foundation's Engineering Research Center for Reconfigurable Machining Systems (grant EEC95-92125), Awards DMI-0322147, DMI-0217481, and CAREER Award DMI-9624402. The authors also thank the editor and referees for their valuable comments and suggestions.

[Received November 2001. Revised April 2003.]

REFERENCES

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.

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.

——— (2001), "A Factor-Analysis Method for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, 43, 84–95.

Camelio, A. J., Hu, S. J., and Ceglarek, D. J. (2001), "Modeling Variation Propagation of Multi-Station Assembly Systems With Compliant Parts," in Proceedings of the 2001 ASME Design Engineering Technical Conference, Pittsburgh, PA, September 9–12.

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

Chang, M., and Gossard, D. C. (1998), "Computational Method for Diagnosis of Variation-Related Assembly Problem," *International Journal of Production Research*, 36, 2985–2995.

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

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.

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

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

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.

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.

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

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.

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

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

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.

Schott, J. R. (1997), *Matrix Analysis for Statistics*, New York: Wiley.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

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

Zhou, S., Huang, Q., and Shi, J. (2003), "State Space Modeling for Dimensional Monitoring of Multistage Machining Process Using Differential Motion Vectors," *IEEE Transactions on Robotics and Automation*, 19, 296–308.