

Robust Method of Multiple Variation Sources Identification in Manufacturing Processes For Quality Improvement

Zhiguo Li
Graduate Student

Shiyu Zhou¹

Assistant Professor
e-mail: szhou@engr.wisc.edu

Department of Industrial and Systems
Engineering,
University of Wisconsin,
Madison, WI 53706

Variation source identification is a critical step in the quality and productivity improvement of manufacturing processes and draws significant attention recently. In this article we present a robust pattern-matching technique for variation source identification. In this paper, a multiple variation sources identification technique is developed by adopting the linear relationship between variation sources and product quality characteristics, which is described by a coefficient matrix. The columns of the coefficient matrix are treated as the signatures of corresponding variation sources. The matching is conducted between the signature vectors and the eigenvectors of the sample covariance matrix of the product quality measurements. Multiple faults are allowed in the matching. Further, both the perturbation of unstructured noise and the sample uncertainties are considered in this matching method. A comprehensive case study illustrates the effectiveness of this method. This robust method can be used for root cause identification of manufacturing processes. The application of this method can significantly reduce the troubleshooting time and hence improve the quality and productivity of manufacturing processes.

[DOI: 10.1115/1.2117447]

1 Introduction

Due to the inevitable random disturbances on the operations and other inherent variability in equipments, etc., significant uncertainties and variations exist in the final outputs of a manufacturing process. These uncertainties and variations are inversely proportional to quality and productivity of the process [1]. Hence, variation reduction is essential to improving process efficiency and product quality in order to gain competitive advantage. Statistical process control (SPC) is widely used in practice for process monitoring and change detection. Although this method can give an alarm of the existence of process faults, it cannot pinpoint the fault or root cause of the alarm. Process faults could be roughly classified into two categories: process mean shift error and variation error [1]. Mean shift error refers to a deviation of the process mean from the target value, while variation error refers to the process variance increase to an unacceptable level. Some approaches have been developed in the past few years for root cause identification and fault diagnosis. May et al. [2] and Dunia and Qin [3] proposed some methods to identify the root causes of the process mean shift error. In this paper, we will focus on the root cause identification (also called "variation source identification") techniques for the variation increase errors.

Based on mathematical models linking variation sources and product quality, two types of variation source identification techniques have been developed. The first type is the direct estimation method. Different variance estimation methods such as maximum likelihood estimation, least squares estimation, were adopted in Chang and Gossard [4], Apley and Shi [5], Zhou et al. [6], and compared in Ding et al. [7] to estimate the variation source directly using the product quality measurement data. The second type is the pattern matching method based on a physical engineering model. Ceglarek and Shi [8] studied the fixture fault diagnosis

in automotive body assembly. They use the CAD information about the fixture geometry and location of the measurement points on the assembled part to obtain the variation pattern of the measurement data under a faulty fixture condition. Then, Principal Component Analysis (PCA) and the pattern recognition approach is used to map the real measurement data to those predetermined fault patterns. Their methodology works under the rigid part and single fault condition. This method is extended by Rong et al. [9] to the compliant beam structure assembly model. The single fault identification is implemented through comparing the diagnostic vectors derived from the assembly model and fault symptoms obtained from the measurement data. The sample properties of the principal eigenvector are considered in this paper, but the variance of the disturbance noise is assumed in a simple structure. Most recently, Ding et al. [10] studied the influence of measurement noises with a general covariance structure on the pattern matching of single fault under the large sample assumption. The fault pattern vectors are generated using a state space model for multistage manufacturing processes and a PCA-based algorithm similar to the one proposed in [8] is used to identify the root cause. Li et al. [11] considered the influence of both sample uncertainties and general structured measurement noise on the pattern matching algorithm for the single fault case because neither of the assumptions on the simple variance structure of noise or a large sample is always realistic in actual manufacturing environments. Compared with techniques proposed in [9,10], their approach has a higher identification probability when the above assumptions are not strictly satisfied.

All the above-mentioned pattern matching methods require a single fault condition, i.e., pattern matching can be conducted when only a single fault exists in the process. In this article, we focus on a robust pattern matching technique for *multiple* variation sources identification. The technique presented in this paper considers the influence of both sample uncertainty and general structured noise on the pattern matching for the *multiple* faults case. In other words, multiple variation sources can be identified simultaneously using this method, in which the principal angle between two subspaces spanned by the measurement symptom

¹Author whom correspondence should be addressed.

Contributed by the Manufacturing Engineering Division of ASME for publication in the JOURNAL OF MANUFACTURING SCIENCE AND ENGINEERING. Manuscript received August 25, 2004; final manuscript received July 21, 2005. Review conducted by S. J. Hu.

vectors and the signature vectors representing variation sources serves as a means of matching. Moreover, the pattern matching technique provides intuitive geometric illustration of the variation sources and it is easy to learn and be accepted by practitioners.

The remainder of the article is organized as follows. In Sec. 2, the problem formulation and the robust pattern matching algorithm are presented. In Sec. 3 we use a case study to illustrate the effectiveness of the developed technique. Finally, we conclude the paper in Sec. 4.

2 Robust Pattern Matching Technique for Multiple Variation Source Identification

2.1 Problem Formulation.

2.1.1 Problem Formulation and Assumptions. In this paper, we adopt the linear relationship between the process variation sources and the product quality measurements, which can be generally expressed as [12,13]:

$$\mathbf{y} = \mathbf{A}\mathbf{f} + \boldsymbol{\epsilon}. \quad (1)$$

In Eq. (1), \mathbf{y} is an $N \times 1$ vector of product quality measurements (it is actually the deviation of quality characteristics from their nominal values), \mathbf{A} is an $N \times n$ constant matrix determined by the physical principle of the process, \mathbf{f} is an $n \times 1$ random vector representing the process errors, and $\boldsymbol{\epsilon}$ is an $N \times 1$ random vector representing measurement and unmodeled noises in the process. Without causing confusion, the components of \mathbf{f} are also called variation sources or process faults in this paper.

In general, the impact of the process fault on the product quality is nonlinear, which is represented by a general function $g(\cdot)$ in Eq. (2),

$$\mathbf{y} = g(\mathbf{f}) + \boldsymbol{\epsilon}. \quad (2)$$

Although the relationship between \mathbf{y} and \mathbf{f} is, in general, nonlinear, $g(\mathbf{f})$ can be linearized through Taylor series expansion under nonrestrictive small error conditions. If we replace $g(\mathbf{f})$ with its Taylor series, this yields

$$\mathbf{y} \approx g(\mathbf{f}_0) + \left. \frac{\partial g(\mathbf{f})}{\partial \mathbf{f}} \right|_{\mathbf{f}=\mathbf{f}_0} \Delta \mathbf{f} + \boldsymbol{\epsilon}. \quad (3)$$

If \mathbf{f} is \mathbf{f}_0 , then no process fault occurs and $g(\mathbf{f}_0)$ is the nominal quality measurement value. If we denote $\Delta \mathbf{y}$ as $\mathbf{y} - g(\mathbf{f}_0)$, $\left. \frac{\partial g(\mathbf{f})}{\partial \mathbf{f}} \right|_{\mathbf{f}=\mathbf{f}_0}$ as \mathbf{A} , and merge the high-order nonlinear residuals into $\boldsymbol{\epsilon}$, then

$$\Delta \mathbf{y} = \mathbf{A} \cdot \Delta \mathbf{f} + \boldsymbol{\epsilon}. \quad (4)$$

Without causing confusion, the linear fault quality model as in (1) is obtained by replacing $\Delta \mathbf{y}$ by \mathbf{y} , and $\Delta \mathbf{f}$ by \mathbf{f} . Because $\Delta \mathbf{f}$ is often small in magnitude compared with \mathbf{f}_0 , the linear relation is a good approximation of the nonlinear relationship in many processes such as assembly [12] and machining [14].

The components of \mathbf{f} could have a constant bias and/or variance increase to cause errors in product quality characteristics. A constant bias in \mathbf{f} will cause a mean shift error and some approaches were developed to detect and diagnose this type of faults, as we stated in the Introduction. In this paper, we focus on the root cause identification of a variance increase. The vector \mathbf{f} is treated as a random vector and, further, because the mean of quality measurements can always be subtracted, \mathbf{f} and $\boldsymbol{\epsilon}$ are assumed to have a zero mean for the sake of simplicity.

Based on this linear model, we will formulate the problem of variation source identification as follows: Known matrix \mathbf{A} , given multiple observations of \mathbf{y} , how to identify which fault(s) occurred in the system or, namely, which component(s) of \mathbf{f} in (1) possess(es) significant variance?

From Eq. (1), we can obtain the following variational relationship

$$\boldsymbol{\Sigma}_y = \mathbf{A}\boldsymbol{\Sigma}_f \mathbf{A}^T + \boldsymbol{\Sigma}_\epsilon, \quad (5)$$

where $\boldsymbol{\Sigma}_y$ and $\boldsymbol{\Sigma}_f$ are the covariance matrices of \mathbf{y} and \mathbf{f} , respectively, and $\boldsymbol{\Sigma}_\epsilon$ is the rest disturbance term due to the existence of $\boldsymbol{\epsilon}$. For a process with insignificant nonlinearity, $\boldsymbol{\Sigma}_\epsilon$ mainly contains the variance of measurement noise. One assumption is that $\boldsymbol{\Sigma}_f$ is a diagonal matrix, because the process faults (i.e., the elements in \mathbf{f}) are often independent of each other.

Without losing generality, we would assume that there are p ($1 \leq p \leq n$) faults occurring in the system, that is, there would exist p out of n diagonal entries of $\boldsymbol{\Sigma}_f$ that are nonzero. By changing the sequence of the column vectors of matrices \mathbf{A} and $\boldsymbol{\Sigma}_f$, we can always get the following equation regarding $\boldsymbol{\Sigma}_f$:

$$\boldsymbol{\Sigma}_y = [\mathbf{a}_{(1)} \quad \mathbf{a}_{(2)} \quad \cdots \quad \mathbf{a}_{(p)} \quad \mathbf{a}_{(p+1)} \quad \cdots \quad \mathbf{a}_{(n)}] \cdot \begin{bmatrix} \sigma_{(1)}^2 & & & & & & \\ & \sigma_{(2)}^2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_{(p)}^2 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{a}_{(1)}^T \\ \mathbf{a}_{(2)}^T \\ \vdots \\ \mathbf{a}_{(p)}^T \\ \mathbf{a}_{(p+1)}^T \\ \vdots \\ \mathbf{a}_{(n)}^T \end{bmatrix} + \boldsymbol{\Sigma}_\epsilon, \quad (6)$$

where $\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$ are the column vectors that are associated with the nonzero process variations of $\sigma_{(1)}^2, \sigma_{(2)}^2, \dots, \sigma_{(p)}^2$, which are also known as the magnitudes of the fault. The number in parentheses in subscripts represents the rearranged sequence of these columns, as opposed to the original subscripts without parentheses.

Root cause identification is to identify which fault(s) occurs in the system, i.e., we should identify the column vectors

$\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$ ($1 \leq p \leq n$) in Eq. (6), which represent the process variation sources. Hereafter, we call the vectors $\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$ *fault signature vectors*. All of n column vectors in matrix $\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n]$ (or $\mathbf{A}_t = [\mathbf{a}_{(1)} \quad \cdots \quad \mathbf{a}_{(p)} \quad \mathbf{a}_{(p+1)} \quad \cdots \quad \mathbf{a}_{(n)}]$), defined as *potential* fault signature vectors, represent n potential process variation sources in total. Please notice that matrix \mathbf{A}_t is transformed from matrix \mathbf{A} according to Eq. (6). In practice, it is unreasonable to assume that

these n faults always occur simultaneously. Thus, it is allowable for the number of nonzero components of an \mathbf{f} vector change from sample to sample.

2.1.2 Basics of Multiple Variation Identification. If we assume the matrix Σ_{ϵ} have the simplest form as $\sigma_{\epsilon}^2 \cdot \mathbf{I}$, where σ_{ϵ}^2 is a scale indicating the magnitude of the noise and \mathbf{I} is an identity matrix with an appropriate dimension. This type of noise is called *structured noise*, referring to the special structure of Σ_{ϵ} . Under this assumption, it is known that for the model $\Sigma_{\mathbf{y}} = \mathbf{A}\Sigma_{\mathbf{f}}\mathbf{A}^T + \Sigma_{\epsilon}$ and $\mathbf{A} \neq \mathbf{0}$, the p eigenvectors of $\Sigma_{\mathbf{y}}$ that associate with the p largest eigenvalues (these eigenvectors are called *principal eigenvectors* in this paper) span the same space as the range of $\mathbf{A}_s \equiv [\mathbf{a}_{(1)} \ \mathbf{a}_{(2)} \ \dots \ \mathbf{a}_{(p)}]$, i.e., $\text{span}\{\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_p\} = \text{span}\{\mathbf{a}_{(1)} \ \mathbf{a}_{(2)} \ \dots \ \mathbf{a}_{(p)}\}$, where the vectors $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_p$ are the p principal eigenvectors of $\Sigma_{\mathbf{y}}$ [15,16]. If matrices $\Sigma_{\mathbf{y}}$ and \mathbf{A} are known in practice, then we can get the p principal eigenvectors of $\Sigma_{\mathbf{y}}$, and select combinations of p column vectors in matrix \mathbf{A} until we find the combination that meets the condition of $\text{span}\{\mathbf{h}_1 \ \mathbf{h}_2 \ \dots \ \mathbf{h}_p\} = \text{span}\{\mathbf{a}_{(1)} \ \mathbf{a}_{(2)} \ \dots \ \mathbf{a}_{(p)}\}$. Thus, we can claim that the process faults represented by $\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$ occur in the process. One point we would make clear here is that the conclusion will still hold without the presence of a disturbance term in the system, i.e., the system model will change to $\mathbf{y} = \mathbf{A}\mathbf{f}$ (this is right because we can let $\sigma_{\epsilon}^2 = 0$).

However, there are two difficulties in practice if we want to apply this simple pattern matching technique stated above. First, in reality, Σ_{ϵ} is often not in such a simple structure as $\sigma_{\epsilon}^2 \cdot \mathbf{I}$, which can probably represent well only the case where exactly the same measurement devices are used to measure quality features in a linear process. When different measurement devices are used, the variances of different components in ϵ will be different. Moreover, the noise term ϵ could comprise unmodeled factors in the system and high-order nonlinear residuals, that are likely to be correlated, and makes even a diagonal structure of Σ_{ϵ} unrealistic. For this reason, it is more practical to study the influence of Σ_{ϵ} in a general structure. With the presence of unstructured Σ_{ϵ} that is not in the simple structure, the space spanned by the p principal eigenvectors of $\Sigma_{\mathbf{y}}$ will deviate from that spanned by $\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$.

Second, in practice, the population covariance matrix $\Sigma_{\mathbf{y}}$ cannot be exactly known. We can only estimate $\Sigma_{\mathbf{y}}$ from the sample covariance matrix $\mathbf{S}_{\mathbf{y}}$ of measurement data. The principal eigenvectors of $\mathbf{S}_{\mathbf{y}}$ are random vectors, meaning that we may only know the distribution of *fault symptom vectors* (p principal eigenvectors of $\Sigma_{\mathbf{y}}$) instead of their exact values. In the following part of this paper, we will call the p principal eigenvectors of $\mathbf{S}_{\mathbf{y}}$ as *sample fault symptom vectors*. For this reason, instead of checking if the space spanned by the sample principal eigenvector is the same as that spanned by the theoretical fault signature in an exact and deterministic sense, we should establish a confidence boundary to see if they are the same in the statistical sense.

2.2 Robust Pattern Matching Technique. To deal with the above-mentioned two kinds of uncertainty, i.e., noise perturbation and sampling uncertainty, an effective technique is developed in this paper.

In the following sections, we represent the eigenvalue and eigenvector pairs of $\mathbf{A}\Sigma_{\mathbf{f}}\mathbf{A}^T$, $\Sigma_{\mathbf{y}}$, and $\mathbf{S}_{\mathbf{y}}$ are $\{m^{(i)}, \mathbf{v}^{(i)}\}$, $\{\lambda^{(i)}, \mathbf{h}^{(i)}\}$ and $\{l^{(i)}, \mathbf{q}^{(i)}\}$ ($i=1, 2, \dots, N$) respectively, where the superscript “ (i) ” means they are the i th *principal* eigenvalue and eigenvector pairs of $\mathbf{A}\Sigma_{\mathbf{f}}\mathbf{A}^T$, $\Sigma_{\mathbf{y}}$, and $\mathbf{S}_{\mathbf{y}}$, respectively. The basic approach is illustrated in Fig. 1.

The figure is for the case that two faults, \mathbf{a}_i and \mathbf{a}_j , occur in a process. From the theory stated above, the space spanned by the fault signature vectors \mathbf{a}_i and \mathbf{a}_j is the same as that spanned by the first two principal eigenvectors of $\mathbf{A}\Sigma_{\mathbf{f}}\mathbf{A}^T$, $\mathbf{v}^{(1)}$, and $\mathbf{v}^{(2)}$, i.e., $\text{span}\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}\} = \text{span}\{\mathbf{a}_i, \mathbf{a}_j\}$. From the geometric point of view,

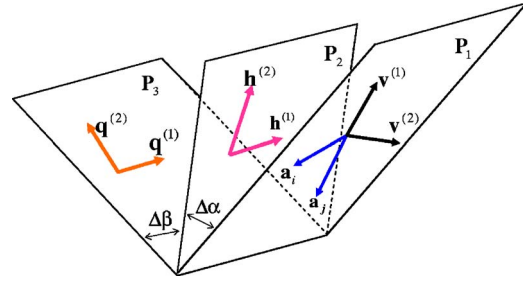


Fig. 1 The joint effects of two kinds of uncertainty

the two vectors \mathbf{a}_i and \mathbf{a}_j will form a plane that can also be formed by $\mathbf{v}^{(1)}$ and $\mathbf{v}^{(2)}$. The plane is denoted as \mathbf{P}_1 in Fig. 1. With the presence of unstructured noise, we can get the first two principal eigenvectors of $\Sigma_{\mathbf{y}}$, $\mathbf{h}^{(1)}$, and $\mathbf{h}^{(2)}$, which will also form a plane. However, due to the perturbation of Σ_{ϵ} , this plane, denoted as \mathbf{P}_2 , will deviate from \mathbf{P}_1 . The difference between these two planes is represented by an angle $\Delta\alpha$. In practice, we only can get the principal eigenvectors of $\mathbf{S}_{\mathbf{y}}$, $\mathbf{q}^{(1)}$, and $\mathbf{q}^{(2)}$, from the measurement data. These two vectors will span another plane, denoted as \mathbf{P}_3 , which will deviate from \mathbf{P}_2 by a angle of $\Delta\beta$. One point that needs to be mentioned is that \mathbf{P}_3 is actually random due to the sampling uncertainty.

If we can get the boundary of angles $\Delta\alpha$ and $\Delta\beta$, then the total boundary of $\Delta\alpha + \Delta\beta$ will serve as a means for fault identification. Knowing the plane formed by vectors $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$, we can pick up two vectors from matrix \mathbf{A} to form another plane and check the angle between these two planes. If we select the right column vectors from \mathbf{A} , the angle will be no larger than the total boundary of $\Delta\alpha + \Delta\beta$, we will claim that the process faults represented by these two vectors occur in the process. In our case, they are \mathbf{a}_i and \mathbf{a}_j . Now the problem of fault identification changes as to how to find the boundary.

In this article, we use the principal angle [17–19] between two subspaces as a measurement of the difference between two subspaces, which means that in above description, $\Delta\alpha$ is the principal angle between spaces \mathbf{P}_1 and \mathbf{P}_2 and $\Delta\beta$ is the principal angle between spaces \mathbf{P}_2 and \mathbf{P}_3 . We may note that the boundary of $\Delta\alpha$ is deterministic while the boundary of $\Delta\beta$ is stochastic.

In the subsequent sections we will present the confidence boundary for the principal angle between the space spanned by p principal eigenvectors of $\mathbf{S}_{\mathbf{y}}$ and that spanned by p fault signature vectors from matrix \mathbf{A} . We list our assumptions as follows: (a) the fault–quality relation can be adequately described by Eq. (1). (b) The system noises have a general structure, i.e. the disturbance matrix Σ_{ϵ} is of a general form. Σ_{ϵ} may be unknown, but the range of its eigenvalues is assumed known. (c) One fault or multiple faults occur in the system. In the following part we will assume that p faults occur in the system.

The rationale of assumption (b) is that ϵ is usually dominated by measurement noises, of which the range of eigenvalues represents the range of sensor accuracy that can be obtained through the sensor vendor’s specification. Even if ϵ comprises unmodeled faults and high-order nonlinear residuals, an offline calibration could provide the information regarding the Σ_{ϵ} ’s eigenvalue range.

2.2.1 Disturbance due to Unstructured Noises. In this part we focus on the perturbation of unstructured noises on the eigenvectors of $\Sigma_{\mathbf{y}}$. With the presence of unstructured noises the space spanned by principal eigenvector ($\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}$) of $\Sigma_{\mathbf{y}}$ differs from the space spanned by p fault signature vectors ($\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}$) and the boundary of this difference is related to the additive disturbance matrix Σ_{ϵ} . In order to characterize the perturbation boundary for this difference, we need the following

definition.

Principal Angle: Suppose \mathbf{S}_1 and \mathbf{S}_2 are two subspaces in \mathbf{R}^N , whose dimensions satisfy $r = \dim(\mathbf{S}_1) \geq \dim(\mathbf{S}_2) = q \geq 1$, then the *principal angles* $\theta_1, \theta_2, \dots, \theta_q \in [0, \pi/2]$ between \mathbf{S}_1 and \mathbf{S}_2 are defined recursively by $\cos(\theta_k) = \max_{\mathbf{u} \in \mathbf{S}_1} \max_{\mathbf{v} \in \mathbf{S}_2} \mathbf{u}^T \mathbf{v} = \mathbf{u}_k^T \mathbf{v}_k$, subject to: $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$, $\mathbf{u}^T \mathbf{u}_i = \mathbf{v}^T \mathbf{v}_i = 0$, $i = 1, 2, \dots, k-1$.

Note that the principal angles satisfy $0 \leq \theta_1 \leq \dots \leq \theta_q \leq \pi/2$. The *largest* principal angle between these two subspaces is the largest angle between an arbitrary vector in \mathbf{S}_1 and its closet vector in \mathbf{S}_2 . If \mathbf{S}_1 and \mathbf{S}_2 have the same dimension, the *largest* principal angle could be calculated based on the following proposition [18].

PROPOSITION. Suppose \mathbf{S}_1 and \mathbf{S}_2 are two subspaces with the same dimension p in \mathbf{R}^N , let \mathbf{L} and \mathbf{M} be the sets of p orthonormal bases of these two subspaces, respectively, the largest principal angle between \mathbf{S}_1 and \mathbf{S}_2 is given by $\cos(\sqrt{\lambda_p})$, where λ_p is the smallest eigenvalue of $\mathbf{L}^T \mathbf{M} \mathbf{M}^T \mathbf{L}$.

We use the matrix perturbation theory to evaluate the upper boundary on the perturbation of the subspace spanned by fault symptom vectors (p principal eigenvectors of Σ_y) due to the unstructured Σ_ϵ . An important result is given in the following Theorem.

THEOREM. Known $\mathbf{A} \Sigma_f \mathbf{A}^T$ and $\Sigma_y = \mathbf{A} \Sigma_f \mathbf{A}^T + \Sigma_\epsilon$ are $N \times N$ symmetric matrices, if there are p faults occurring in the system, we will have $m^{(1)} \geq m^{(2)} \geq \dots \geq m^{(p)} > 0 = m^{(p+1)} = \dots = m^{(N)}$, where $m^{(1)}, m^{(2)}, \dots, m^{(p)}$ are p principal eigenvalues of $\mathbf{A} \Sigma_f \mathbf{A}^T$. If $\|\Sigma_\epsilon\|_2 \leq m^{(p)}/5$, where $\|\cdot\|_2$ is the matrix 2-norm, the upper boundary of the principal angle between the subspaces $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$ and $\text{span}\{\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}\}$ is

$$\Delta \theta_1 \leq \dots \leq \Delta \theta_p \leq \sin^{-1} \left(\frac{4}{m^{(p)}} \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)} \right), \quad (7)$$

where $\Delta \theta_1 \leq \dots \leq \Delta \theta_p$ are the p principal angles between these two subspaces, $\lambda_{\max}(\Sigma_\epsilon)$ and $\lambda_{\min}(\Sigma_\epsilon)$ are the largest and smallest eigenvalues of Σ_ϵ , respectively.

The proof of this result is presented in Appendix A.

Remark 1. Under the perturbation of unstructured noise, the subspace $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$ will deviate from the subspace $\text{span}\{\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}\}$. Although in practice we cannot calculate the accurate principal angle between these two subspaces due to the unknown Σ_y , we can derive the boundary of principal angle based on the theorem.

Remark 2. The condition $\|\Sigma_\epsilon\|_2 \leq m^{(p)}/5$ of our result means that the largest eigenvalue of Σ_ϵ is five times smaller than the p th principal eigenvalue of $\mathbf{A} \Sigma_f \mathbf{A}^T$, which is usually very close to the smallest variance of the p process faults. This condition is not very restrictive in practice.

Remark 3. The above result tells us that the difference between two subspaces $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$ and $\text{span}\{\mathbf{a}_{(1)}, \mathbf{a}_{(2)}, \dots, \mathbf{a}_{(p)}\}$, represented by the principal angle $\Delta \theta_p$ between them (shown as the angle between two planes \mathbf{P}_1 and \mathbf{P}_2 in Fig. 1), is determined by the eigenvalue of $\mathbf{A} \Sigma_f \mathbf{A}^T$, and the extreme eigenvalues of Σ_ϵ . We would also like to point out that the boundary specified by (7) is a worst-case boundary.

Inequality (7) suggests that the perturbation boundary between these two subspaces depends on $m^{(p)}/\lambda_{\max}(\Sigma_\epsilon)$ and $\lambda_{\max}(\Sigma_\epsilon)/\lambda_{\min}(\Sigma_\epsilon)$. The value of $m^{(p)}/\lambda_{\max}(\Sigma_\epsilon)$ can be somewhat viewed as the signal-to-noise ratio, while $\lambda_{\max}(\Sigma_\epsilon)/\lambda_{\min}(\Sigma_\epsilon)$ indicates the imbalance in accuracies associated with different measurement devices. With the increase in the signal-to-noise ratio, the perturbation boundary will get smaller. For instance, if $\lambda_{\max}(\Sigma_\epsilon) = 2.25 \times 10^{-4}$, $m^{(p)}/\lambda_{\max}(\Sigma_\epsilon)$ is 50, and $\lambda_{\max}(\Sigma_\epsilon)/\lambda_{\min}(\Sigma_\epsilon)$ is 10, the perturbation boundary (i.e., the angle) will be around 4.6 deg. However, if $m^{(p)}/\lambda_{\max}(\Sigma_\epsilon)$ is around 150, meaning that the fault magnitude is three times larger than before, the perturbation angle will reduce to 1.5 deg or so.

On the other hand, with a higher imbalance among eigenvalues of Σ_ϵ , the perturbation boundary will get larger. If $\Sigma_\epsilon = \sigma_\epsilon^2 \mathbf{I}$, or there is no presence of noise in the system, we may notice that this boundary will be zero, which means these two subspaces are actually the same. This is consistent with the aforementioned result in Sec. 2.1.

Remark 4. When there are p faults occurring in the system, under the assumption that only the variations of p faults are larger than zero while those of the $(n-p)$ potential faults are zero, we have $m^{(p+1)} = \dots = m^{(N)} = 0$; thus the boundary is $\sin^{-1}((4/m^{(p)}) \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)})$. However, if the variance of one or more out of $(n-p)$ potential faults is not zero but very small (not beyond a preset value, in this case we would not claim this or these faults occur in the system), we will have $m^{(p+1)} > 0$, but it is of a very small value. In such a case, the boundary will be changed to $\sin^{-1}[4/(m^{(p)} - m^{(p+1)}) \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)}]$.

When we use Eq. (7) to get the boundary of the principal angle, we need the value of $m^{(p)}$, which is the p th principal eigenvalue of $\mathbf{A} \Sigma_f \mathbf{A}^T$ and is often not available. We can deal with this problem by finding the lower limit for $m^{(p)}$. Based on Theorem 8.1.5 in [17], we have

$$m^{(p)} + \lambda_{\min}(\Sigma_\epsilon) \leq \lambda^{(p)} \leq m^{(p)} + \lambda_{\max}(\Sigma_\epsilon). \quad (8)$$

Based on the above inequation, we can get the range of $m^{(p)}$ as follows:

$$\lambda^{(p)} - \lambda_{\max}(\Sigma_\epsilon) \leq m^{(p)} \leq \lambda^{(p)} - \lambda_{\min}(\Sigma_\epsilon). \quad (9)$$

Thus we have,

$$\sin(\Delta \theta_p) \leq \frac{4}{\lambda^{(p)} - \lambda_{\max}(\Sigma_\epsilon)} \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)}. \quad (10)$$

Since $l^{(p)}$, the p th principal eigenvalue of Σ_y , is the consistent estimate of $\lambda^{(p)}$, substituting $\lambda^{(p)}$ with $l^{(p)}$ in Eq. (10), we can estimate the boundary of the principal angle as follows,

$$\gamma_1 = \frac{4}{l^{(p)} - \lambda_{\max}(\Sigma_\epsilon)} \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)}. \quad (11)$$

Finally, the boundary of the principal angle can be selected as follows in practice,

$$\gamma_1 = \frac{4}{[l^{(p)} - \lambda_{\max}(\Sigma_\epsilon)] - [l^{(p+1)} - \lambda_{\min}(\Sigma_\epsilon)]} \sqrt{\lambda_{\max}(\Sigma_\epsilon) - \lambda_{\min}(\Sigma_\epsilon)}. \quad (12)$$

2.2.2 Disturbance due to Sampling Uncertainty. The confidence boundary between the subspace spanned by p fault symptom vectors, $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$, and the subspace spanned by p sample fault symptom vectors, $\text{span}\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(p)}\}$, is developed in this section.

In this paper, we use the Monte Carlo method similar to Krzanowski [18,19] to provide the boundary of the principal angle between $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$ and $\text{span}\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(p)}\}$.

The simulation is based on model (1). Known matrix \mathbf{A} , the problem is how to select the matrices Σ_f and Σ_ϵ . Since ϵ is mainly due to measurement noise in practice, we can get the range of standard deviation of components in ϵ based on the engineering analysis of the measurement equipments. Also, we assume that the range of the eigenvalues of Σ_ϵ is known. Thus, under these two prerequisites, we will generate Σ_ϵ randomly in the simulation. The parameters required for Σ_f are the number of the fault p and the variances of the faults $\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2$. It has been pointed out by some authors [20,21] that if the p th and $(p+1)$ th eigenvalues are close to each other, the associated eigenvectors of the sample covariance matrix, $\mathbf{q}^{(p)}$ and $\mathbf{q}^{(p+1)}$, are not stable. Based on this fact, we take the ratio $\sigma_p^2/\max(\sigma_\epsilon^2)$ as a parameter in the simulation, where $\max(\sigma_\epsilon^2)$ is the largest variance of components in ϵ .

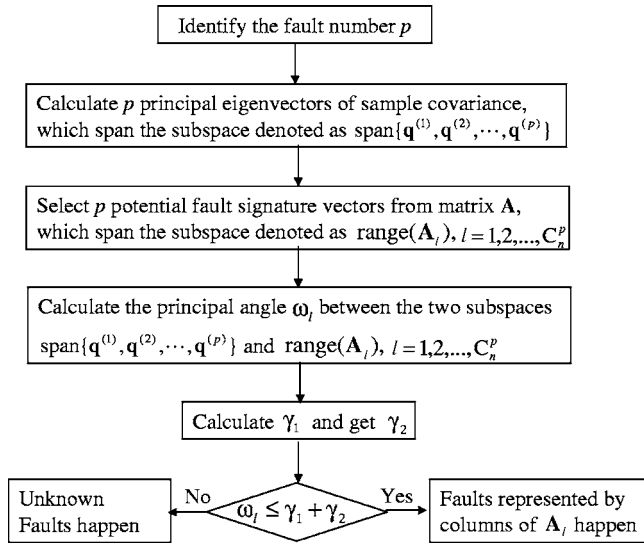


Fig. 2 The procedure of multiple process variation sources identification

Hereafter we would denote it as *variance ratio* $VR \equiv \sigma_p^2 / \max(\sigma_\epsilon^2)$ in this paper. As stated above, the largest noise variance can often be derived through a gauge capability analysis. Also, σ_p^2 can be regarded as the “tolerance” level on the variation source specified by design requirements. Thus, usually we can obtain the bound of this ratio. As Krzanowski pointed out in [19], the ratio $\sigma_1^2 / \text{trace}(\Sigma_y)$ will also impact the sample uncertainty of the eigenvectors. In our case, the process noise is unstructured. Thus, we would use the ratio $TR \equiv \sigma_1^2 / [(\sigma_1^2 + \dots + \sigma_p^2) + \text{trace}(\Sigma_\epsilon)]$ as a parameter in the simulation. When $p=1$, i.e., only one fault occurs, we will not specify this ratio because $\sigma_1^2 = \sigma_p^2 = VR \cdot \max(\sigma_\epsilon^2)$. When multiple faults occur, we will introduce the ratio TR in simulation to show its impact on the principal angle. Additionally, $\sigma_2^2, \dots, \sigma_{p-1}^2$ are selected randomly in the range of $\sigma_1^2 \sim \sigma_p^2$.

In summary, we select the parameters in simulation including sampling number N_s , the number of fault p , the ratio VR , and TR . Also, we would specify the number of replicates. Based on the Monte Carlo simulation, we can obtain the approximation of the boundary for principal angle $\Delta\beta$ between two subspaces of $\text{span}\{\mathbf{h}^{(1)}, \mathbf{h}^{(2)}, \dots, \mathbf{h}^{(p)}\}$ and $\text{span}\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(p)}\}$, which we denote as γ_2 in this paper. In the simulation illustrated in Sec. 3, we will present the estimated critical values of γ_2 at 90%, 95%, and 99% confidence level, respectively.

2.3 A Robust Pattern-Matching Procedure. By combining both effects of noise perturbation and sampling uncertainty, a robust pattern-matching technique can be developed. A diagram of this procedure is illustrated in Fig. 2.

First, we should estimate the number of faults based on the quality measurement, which equals the number of principal components that should be held in a PCA model to represent the N -variable dataset. Several methods have been developed to solve this problem. In this paper, we select the method presented in Jackson [22]. The criteria is called *IE (imbedded error)*,

$$IE(s) = \sqrt{\frac{\sum_{s=p+1}^N l^{(s)}}{N_s N(N-s)/s}}, \quad (13)$$

where N_s is the sample size, N is the dimension of the quality measurement, s is the number of faults, and $l^{(s)}$ is the s th eigenvalue of \mathbf{S}_y , respectively. To use this criteria, $IE(s)$ is evaluated for $s=1, 2, \dots, (N-1)$. The fault number p is the value of s that mini-

mizes $IE(s)$.

After we estimate the fault number p and obtain the parameters of \mathbf{S}_y , $\lambda_{\max}(\Sigma_\epsilon)$, $\lambda_{\min}(\Sigma_\epsilon)$ from samples of product quality measurements and engineering analysis, we can get the estimated boundary γ_1 for the principal angle between the subspace spanned by p principal eigenvectors of Σ_y and that spanned by p fault signature vectors. Based on Monte Carlo simulation, the approximate boundary γ_2 for the principal angle between the subspace spanned by p principal eigenvectors of Σ_y and that spanned by p principal eigenvectors of \mathbf{S}_y can be obtained.

The method of combining the overall effects of noise perturbation and sampling uncertainty is stated as follows. There are $k(k = C_n^p)$ combinations in total if we select p column vectors from all n potential fault signature vectors in matrix \mathbf{A} . For each combination we pick up a matrix $\mathbf{A}_l (l=1, 2, \dots, k)$ and get the largest principal angle ω_l between the subspace spanned by p principal eigenvectors of \mathbf{S}_y , $\text{span}\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(p)}\}$, and $\text{range}(\mathbf{A}_l)$. If $\omega_l \leq (\gamma_1 + \gamma_2)$, we would claim there are p faults, represented by the p column vectors of \mathbf{A}_l (fault signature vectors), occurring in the system; Otherwise, we have to check another combination.

The developed variation source identification method is summarized as follows.

- (i) S1: A fault-quality model $\mathbf{y} = \mathbf{A}\mathbf{f} + \boldsymbol{\epsilon}$ should be developed. Based on this model, potential fault signature vectors \mathbf{a}_i , $i=1, 2, \dots, n$, are obtained from matrix \mathbf{A} .
- (ii) S2: The multivariate measurements \mathbf{y} of product quality features are obtained during production. The sample size is N_s and the dimension of measurements is N . Based on the quality measurement, we can calculate the sample covariance matrix. From this matrix, the number of significant variation sources, p , can be estimated using Eq. (13).
- (iii) S3: Based on the sample covariance matrix \mathbf{S}_y , calculate the p pairs of principal eigenvalues and eigenvector $\{l^{(j)}, \mathbf{q}^{(j)}\} (j=1, 2, \dots, p)$ using the Principal Component Analysis (PCA). There are $k=C_n^p$ combinations in total if we select p columns vectors from all n potential fault signature vectors in matrix \mathbf{A} ($k=C_n^p$), denoted as $\mathbf{A}_l (l=1, 2, \dots, k)$,
- (iv) S4: Estimate $\lambda_{\max}(\Sigma_\epsilon)$ and $\lambda_{\min}(\Sigma_\epsilon)$ from an analysis of accuracy specifications of the measurement system. We can calculate γ_1 based on Eq. (11). Then, we can pick up one matrix \mathbf{A}_l and calculate the largest principal angles ω_l between $\text{span}\{\mathbf{q}^{(1)}, \mathbf{q}^{(2)}, \dots, \mathbf{q}^{(p)}\}$, and $\text{range}(\mathbf{A}_l)$, $l=1 \dots k$. If all of the angles are larger than γ_1 , go to the next step, otherwise go to step S6.
- (v) S5: Based on the Monte Carlo simulation stated in Sec. 2.2, select the 90%, 95%, or 99% critical values of principal angle as γ_2 . If all ω_l 's are larger than $(\gamma_1 + \gamma_2)$, we can claim that the fault of the unknown source occurs, otherwise go to the next step.
- (vi) S6: If $\omega_l \leq (\gamma_1 + \gamma_2)$, we can claim the p faults, represented by the column vectors of matrix \mathbf{A}_l , occur in the system.

In the following section, a comprehensive case study is presented to illustrate this procedure.

3 Case Study

3.1 A Machining Process. The manufacturing process considered here is a three-stage machining process. The product is a V-6 automotive engine head. Its key features and operation sequence are shown in Fig. 3.

The key features include the cover face, joint face, datum locating holes, the slot on the cover face, and extra holes on the cover and joint face. As shown in Fig. 3, the cover face is milled at the first stage. The next step is to mill the joint face and drill datum holes on the face. Last, we will tap holes and mill slot on

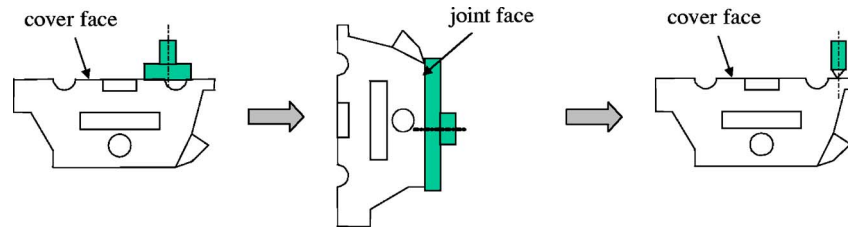


Fig. 3 Operation sequence of the final product

the cover face.

In this process, the key dimensions of the engine head are measured on a Coordinate Measurement Machine (CMM). A total of 15 and 16 points on the joint face and the cover face are measured to determine the quality of the machining operation. These measurements points are evenly distributed on these two surfaces. Therefore, \mathbf{y} will be a 31×1 vector (i.e., $N=31$), consisting of the deviations at 15 points on the cover face and 16 points on the joint face.

Three features, the rough casting datum, the joint face, and the cover face, are considered in this paper. The rough casting datum is a surface parallel and close to the joint face. Based on the CMM measurements of these three features from the machining operations, potential process faults are identified as the locating pin

position errors of the first and the second stage of the process. Fault \mathbf{f} is then a 6×1 vector (i.e., $n=6$), where the first three elements correspond to three pins at the first stage and the last three elements correspond to three pins at the second stage.

A fault-quality model bearing the same linear structure as Eq. (1) can be established from a sophisticated kinematic analysis (please refer to [14] for details). Matrix \mathbf{A} of this specific process is given as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}, \mathbf{u} \quad (14)$$

where $\mathbf{A}_1 = [\mathbf{A}_{11} \quad \mathbf{0}]$, $\mathbf{A}_2 = [\mathbf{0} \quad \mathbf{A}_{22}]$, and

$$\mathbf{A}_{11} = \begin{bmatrix} -0.5402 & -0.6455 & -0.2664 & 0.0683 & 0.5158 & 0.8555 & 0.9902 & 1.1863 & 1.3106 & 1.0177 & 0.6394 & 0.3403 & 0.0530 & -0.0820 & -0.2776 \\ 0.9690 & 0.5991 & 0.2203 & -0.1140 & -0.5612 & -0.9005 & -0.7058 & -0.4225 & -0.2428 & 0.1052 & 0.4597 & 0.7585 & 1.0455 & 0.8192 & 0.5367 \\ 0.5711 & 1.0464 & 1.0461 & 1.0458 & 1.0453 & 1.0450 & 0.7156 & 0.2362 & -0.0678 & -0.1229 & -0.0991 & -0.0988 & -0.0985 & 0.2628 & 0.7409 \end{bmatrix}^T$$

$$\mathbf{A}_{22} = \begin{bmatrix} -0.6349 & -0.4590 & -0.0332 & 0.2510 & 0.6955 & 1.0215 & 1.1372 & 1.1155 & 0.9048 & 0.6472 & 0.5864 & 0.3875 & 0.0777 & -0.1490 & -0.4063 & 0.0776 \\ 0.9602 & 0.6817 & 0.2556 & -0.0288 & -0.4736 & -0.5857 & -0.5031 & -0.3477 & -0.1368 & 0.1210 & -0.0142 & 0.3998 & 0.7098 & 0.9368 & 0.9339 & 0.4496 \\ 0.6747 & 0.7773 & 0.7776 & 0.7778 & 0.7781 & 0.5641 & 0.3659 & 0.2322 & 0.2320 & 0.2318 & 0.4279 & 0.2127 & 0.2124 & 0.2123 & 0.4725 & 0.4728 \end{bmatrix}^T.$$

The six column vectors of \mathbf{A} are the signature vectors of six potential faults.

In this example, noise $\boldsymbol{\epsilon}$ is dominated by errors associated with the measurement device. Since the orientations of a CMM probe are different when it is used to measure the joint face and the cover face, the variances of measurement errors are therefore different for measurement points on these two different surfaces. Based on an engineering analysis, the standard deviations of components in $\boldsymbol{\epsilon}$ are between 0.008 and 0.015 mm.

Furthermore, from the design specification of our fixture system, the locating tolerance is around 0.050 mm. If the standard deviation of components in \mathbf{f} is larger than this tolerance, we claim that the corresponding fault occurs.

3.2 Pattern Matching for Root Cause Identification in the Machining Process. Before we test the effectiveness of the proposed technique, the Monte Carlo simulation to get the estimated boundary γ_2 is implemented for $p=1, 2, 3$, and 4, $N_s=50, 100, 250, 500$, and 1000, $VR=50, 100$, and 150, and $TR=0.75$ and 0.50. Part of the results for $p=1, 2$ is summarized in Appendix B. For each case, 100 000 replicates are generated. $\bar{\beta}$ and σ_β are the mean and standard deviation of the principal angle, respectively. The critical values of the principal angle at confidence levels of 90%, 95%, and 99% are listed in the last three columns. We may use each one among these three values as γ_2 at a different confidence level. One point should be mentioned is that the parameters are selected for some typical engineering applications. Interpolation can be used to estimate the boundary for those values not

listed directly in these tables.

From the tables in the Appendices, the following observations can be made:

1. As predicted, the sample size N_s and the variance ratio $VR \equiv \sigma_p^2 / \max(\sigma_\epsilon^2)$ have a significant impact on the boundary of the principal angle. This simulation works well when VR is larger than 50. The simulation results show that the effects of VR on the boundary of the principal angle are monotone decreasing, i.e., if VR becomes larger, the boundary will decrease.
2. The number of faults impacts the boundary of the principal angle significantly. A larger number of faults will result in less accurate results. Therefore, we must be cautious when applying this method to the case with a large number of simultaneous faults. However, it is not a serious limitation in practice because the number of faults is usually limited even if there are quite a few potential faults in the system.
3. The parameter $TR \equiv \sigma_1^2 / [(\sigma_1^2 + \dots + \sigma_p^2) + \text{trace}(\boldsymbol{\Sigma}_\epsilon)]$ has no significant impact on the boundary of the principal angle. Also, the simulation results show that the effects of TR on the boundary of principal angle are monotone decreasing, i.e., if TR becomes larger, the boundary will decrease.

The range of VR stated in (1) is reasonable because the measurement error usually dominates in $\boldsymbol{\epsilon}$ for many processes such as machining and assembly; thus the VR value is around 100 to several hundred [1].

When we use the simulation results, given the fault number p

Table 1 Different cases considered (10⁻⁴ mm²)

Case #	<i>p</i>	σ_1^2	σ_2^2	σ_3^2	σ_4^2	σ_5^2	σ_6^2
1	0	10	10	10	10	10	10
2	1	10	10	300	10	10	10
3	2	200	10	350	10	10	10
4	3	200	10	10	400	10	300
5	4	250	350	10	10	300	400

and the sample size N_s , although VR and TR are unknown, we can pick up the largest value in the tables (corresponding to the smallest VR and TR), because the effects of VR and TR are consistently monotone decreasing. For example, given $p=3$ and $N_s=100$, we would get the 95% value as 6.04 deg, when VR=50 and TR=0.375. This will give a high detecting rate of this method when we try to identify the process faults.

3.2.1 Effectiveness of the Developed Technique. Given that the standard deviations of components in ϵ are between 0.008 and 0.015 mm, we randomly generate a positive-definite matrix with diagonal elements falling in the range of $64-225 \times 10^{-6}$ mm². Then, we use this matrix as the noise covariance matrix, whose largest and smallest eigenvalues, $\lambda_{\max}(\Sigma_\epsilon)$ and $\lambda_{\min}(\Sigma_\epsilon)$, can be easily computed. In this section, the parameters used are as follows: sample size $N_s=100$, $\max(\sigma_\epsilon^2)=1.97 \times 10^{-4}$ mm², $\min(\sigma_\epsilon^2)=1.27 \times 10^{-4}$ mm², $\lambda_{\max}(\Sigma_\epsilon)=2.50 \times 10^{-4}$ mm, and $\lambda_{\min}(\Sigma_\epsilon)=0.67 \times 10^{-4}$. To demonstrate the proposed variation source identification technique, we consider the following cases respectively. The variances of \mathbf{f} for these four cases are specified in Table 1.

In Table 1, $\sigma_1^2, \sigma_2^2, \dots, \sigma_6^2$ are the variances of f_1, f_2, \dots, f_6 , respectively. From this table, we can see in case (1), that all of the variances are very small, which represents the normal condition without faults. The remaining four cases represent the cases that there are $p=1, 2, 3, 4$ fault(s) occurring in the system, respectively. First, we can get the largest eigenvalues of S_y , as shown in Table 2.

From Table 2, we can see that the largest principal eigenvalue for case 1 is 0.214×10^{-2} mm², which is smaller than the process normal variance 0.25×10^{-2} mm², then we claim that there is no process fault in this case. For other cases, we may notice there are significant process faults in the system. The number of the significant principal eigenvalues may provide some hint to determine the number of faults first conducted using IE criterion. The results are listed in Table 3.

We list the values of IE(s), $s=1, 2, \dots, 9$ in this table. We use the value of s that can minimize IE as the estimated number of faults for each case. The estimated fault numbers are shown in the last column. From the result we can see the estimated numbers are

right.

Finally, the results of the principal angle (ω_l) between the subspace spanned by p principal eigenvectors of S_y and $\text{range}(A_l)$, $l=1, 2, \dots, C_p^n$, for cases (2)–(5) are listed in Table 4(a)–4(d). We also give the calculation result of γ_1 (the boundary of the principal angle due to noise perturbation) and simulation result γ_2 (the boundary of the principal angle due to sampling uncertainty).

For case (2), $p=1$, we then have $C_6^1=6$ matrices of A_l , $l=1, 2, \dots, 6$, which are vectors actually. From Table 4(a), we can see that the angle ω_3 is much smaller than other angles. Because $\gamma_1 < \omega_3 < (\gamma_1 + \gamma_2)$, we can claim that the process faults represented by $A_3=\mathbf{a}_3$ occur in the system, which is identified correctly from Table 1. The value of γ_2 is selected based on VR=50 in Table 7, Appendix B when sample size $N_s=100$.

For case 3, $p=2$ faults occur in the system. Hence, we will have $C_6^2=15$ matrices A_l , $l=1, 2, \dots, 15$. By checking the results in Table 4(b), we find that the angle ω_2 is much smaller than other values and $\gamma_1 < \omega_2 < (\gamma_1 + \gamma_2)$, then we would claim the process faults presented by $A_2=[\mathbf{a}_1, \mathbf{a}_3]$ occur. Similarly, γ_2 is selected based on VR=50, TR=0.50 in Table 8, Appendix B when sample size $N_s=100$.

It is shown in Table 4(c) there are $C_6^3=20$ principal angles ω_l , $l=1, 2, \dots, 20$ for case (4). The process faults represented by $A_9=[\mathbf{a}_1, \mathbf{a}_4, \mathbf{a}_6]$ are identified because $\gamma_1 < \omega_9 < (\gamma_1 + \gamma_2)$.

For the last case, we have $C_6^4=15$ principal angles. The results in Table 4(d) show that the process faults represented by $A_6=[\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_5, \mathbf{a}_6]$ occur in the system. Please note the difference between this case and cases (2)–(4) is that the principal angle $\omega_6 < \gamma_1$, which means that the principal angle falls in the boundary due to the noise perturbation.

3.2.2 Comparison of the Developed Technique and Other Methods. In this section, we would like to compare our developed fault identification technique combing the overall effects of noise perturbation and sampling uncertainty with the methods only considering one of these two factors. We call the method only considering the noise perturbation but not the sampling uncertainty (we denote it as method 1) and the one only considering the sampling uncertainty in the absence of unstructured noise (denoted as method 2). It can be imagined that our method will have a better detection rate.

It is assumed there are $p=2$ faults that occur in the system. We assume that faults \mathbf{a}_1 and \mathbf{a}_4 occur, i.e., the variances of f_1 and f_4 are larger than 0.05^2 mm². We conducted 10 000 replicates of fault detection using the three techniques for a set of combinations of sample N_s and the boundary of principal angle $\tilde{\gamma}_1$, which is the boundary calculated through Eq. (10). For each case, we also give the 95% value of the boundary γ_1 calculated through Eq. (11).

Table 2 The TEN largest eigenvalues of S_y for each case (10⁻² mm²)

Case #	<i>p</i>	$l^{(1)}$	$l^{(2)}$	$l^{(3)}$	$l^{(4)}$	$l^{(5)}$	$l^{(6)}$	$l^{(7)}$	$l^{(8)}$	$l^{(9)}$	$l^{(10)}$
1	0	0.214	0.189	0.127	0.103	0.067	0.040	0.036	0.033	0.028	0.027
2	1	3.034	0.157	0.146	0.135	0.068	0.055	0.037	0.035	0.030	0.027
3	2	2.563	1.858	0.197	0.111	0.074	0.040	0.038	0.034	0.031	0.025
4	3	5.601	2.578	2.313	0.128	0.096	0.056	0.037	0.034	0.029	0.028
5	4	5.750	4.152	1.860	1.403	0.113	0.054	0.043	0.038	0.033	0.028

Table 3 The IE testing results (10⁻³)

Case #	<i>p</i>	IE(1)	IE(2)	IE(3)	IE(4)	IE(5)	IE(6)	IE(7)	IE(8)	IE(9)	Estimated # of faults
2	1	0.313	0.410	0.407	0.459	0.477	0.502	0.522	0.544	0.560	1
3	2	0.531	0.413	0.444	0.469	0.489	0.518	0.539	0.555	0.567	2
4	3	0.772	0.812	0.475	0.500	0.515	0.537	0.562	0.582	0.600	3
5	4	0.925	0.921	0.821	0.511	0.519	0.543	0.563	0.580	0.594	4

Table 4 Results of variation sources identification for cases (2)–(5)

(a) Case 2 (deg) $\gamma_1=1.93$ deg, $\gamma_2=5.30$ deg										
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6					
83.14	84.29	3.63	87.93	89.00	89.10					
(b) Case 3 (deg) $\gamma_1=4.38$ deg, $\gamma_2=5.39$ deg										
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8			
89.52	5.93	79.36	88.05	89.08	63.80	83.50	87.76			
ω_9	ω_{10}	ω_{11}	ω_{12}	ω_{13}	ω_{14}	ω_{15}				
88.98	78.88	87.74	89.00	89.52	89.76	89.76				
(c) Case 4 (deg) $\gamma_1=3.46$ deg, $\gamma_2=6.04$ deg										
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8		ω_9	ω_{10}
89.27	87.59	88.36	89.12	88.63	89.65	88.69	49.57		4.71	42.08
ω_{11}	ω_{12}	ω_{13}	ω_{14}	ω_{15}	ω_{16}	ω_{17}	ω_{18}		ω_{19}	ω_{20}
89.10	89.97	88.88	63.58	63.63	63.62	84.59	84.62		84.59	88.53
(d) Case 5 (deg) $\gamma_1=4.23$ deg, $\gamma_2=12.03$ deg										
ω_1	ω_2	ω_3	ω_4	ω_5	ω_6	ω_7	ω_8			
89.76	88.72	88.68	51.57	45.14	4.19	86.76	86.75			
ω_9	ω_{10}	ω_{11}	ω_{12}	ω_{13}	ω_{14}	ω_{15}				
86.71	88.32	85.67	85.68	85.73	87.54	87.66				

Table 5 lists the resulting miss detection rates of different procedures, which are the rates of the number of miss detection cases over the total number of tests (i.e., 1,000 replicates). We denote e_1, e_2, e_3 the miss detection rates for method 1, method 2, and the technique proposed in this paper, respectively.

From the table the following observations can be made: (1) For method 1, the miss detection rate is very large when the sample size N_s is small, for example, e_1 is nearly 100% when N_s is less than 500 or the boundary $\tilde{\gamma}_1 < 1.41^\circ$. Given the sample size N_s , e_1 will decrease when $\tilde{\gamma}_1$ becomes larger. This is because $\tilde{\gamma}_1$ is the worst-case boundary, which will help offset disturbance from sampling uncertainty. (2) For method 2, we can find that the miss detection rate e_2 will become larger when the sample size N_s or $\tilde{\gamma}_1$ increases. (3) For the technique proposed in this paper, the miss detection rate is always smaller than the other two methods. In fact, except the case $\tilde{\gamma}_1=0$, the miss detection rate is quite close to zero. When $\tilde{\gamma}_1=0$, the miss detection rate is the same as e_2 . (4) As

the 95% value of boundary through Eq. (11), γ_1 is always larger than the exact value $\tilde{\gamma}_1$ calculated through Eq. (10). This will result in more conservative results when we use the boundary γ_1 in fault identification. Also, we may notice that when the sample size N_s increases, γ_1 will converge to $\tilde{\gamma}_1$, because $\lambda^{(p)}$ converges to $m^{(p)}$ when N_s increases.

Usually, a low miss detection rate is accompanied by a high false alarm rate, which is the probability of identifying a normal working condition as a faulty working condition. Since we use a conservative approach to identify the faults, it is not surprising that we will get a higher false alarm rate for the developed technique than the other ones. In this study, the same example is used, except we change the variances of f_1 and f_4 to a value smaller than 0.05^2 mm^2 . We list the results in Table 6, where the false alarm rate δ is the rate of the number of false alarms over the total number of tests (10 000 replicates). Similar to the notation used in the miss detection ratio, we denote δ_1, δ_2 , and δ_3 as the false

Table 5 Miss detection rates of three techniques

N_s		$\tilde{\gamma}_1$ (deg)							
		0	0.64	1.41	2.58	3.96	5.18	7.02	9.85
50	γ_1	0	1.01	2.22	4.10	6.36	8.37	11.06	16.84
	e_1	100%	100%	99.99%	78.78%	20.15%	2.78%	0.39%	0
	e_2	1.21%	1.23%	1.96%	2.01%	3.42%	4.23%	15.64%	32.21%
	e_3	1.21%	0.27%	0.15%	0.01%	0	0	0	0
100	γ_1	0	0.86	1.90	3.51	5.40	7.15	9.76	14.07
	e_1	100%	100%	99.82%	49.17%	4.42%	0.19%	0.01%	0
	e_2	1.60%	1.69%	2.34%	2.63%	4.53%	6.73%	20.86%	46.36%
	e_3	1.60%	0.31%	0.08%	0	0	0	0	0
250	γ_1	0	0.76	1.68	3.09	4.77	6.26	8.62	12.33
	e_1	100%	100%	93.10%	8.59%	0.05%	0	0	0
	e_2	3.11%	3.41%	3.92%	4.43%	6.77%	11.54%	36.2%	71.42%
	e_3	3.11%	0.26%	0.01%	0	0	0	0	0
500	γ_1	0	0.72	1.60	2.93	4.53	5.93	8.13	11.56
	e_1	100%	100%	70.11%	0.48%	0	0	0	0
	e_2	7.11%	6.99%	8.59%	9.31%	14.24%	23.04%	58.70%	92.73%
	e_3	7.11%	0.16%	0	0	0	0	0	0

Table 6 False alarm rates of three techniques

N_s		$\bar{\gamma}_1$ (deg)							
		0	0.64	1.41	2.58	3.96	5.18	7.02	9.85
50	δ_1	0	0	0	0	0.19%	3.43%	6.47%	6.67%
	δ_2	0.94%	1.19%	1.18%	1.11%	0.87%	0.65%	0.64%	0.38%
	δ_3	0.94%	3.41%	5.47%	6.47%	6.65%	6.67%	6.67%	6.67%
100	δ_1	0	0	0	0	1.55%	5.71%	6.67%	6.67%
	δ_2	0.59%	0.51%	0.49%	0.50%	0.42%	0.34%	0.21%	0.16%
	δ_3	0.59%	2.81%	5.75%	6.62%	6.67%	6.67%	6.67%	6.67%
250	δ_1	0	0	0	0.25%	6.19%	6.67%	6.67%	6.67%
	δ_2	0.25%	0.29%	0.26%	0.26%	0.17%	0.15%	0.17%	0.09%
	δ_3	0.25%	3.30%	6.31%	6.67%	6.67%	6.67%	6.67%	6.67%
500	δ_1	0	0	0	3.57%	6.65%	6.67%	6.67%	6.67%
	δ_2	0.11%	0.11%	0.11%	0.11%	0.07%	0.05%	0.03%	0.03%
	δ_3	0.11%	3.65%	6.59%	6.67%	6.67%	6.67%	6.67%	6.67%

alarm rates for method 1, method 2, and the technique developed in this paper. From the table we can see δ_3 is larger than the other two. However, the largest value of δ_3 is still quite small.

Combining the results from miss detection rates and false alarm rates, we can see that the developed technique is quite robust in fault identification because of more power in detection and only slightly more false alarms.

4 Concluding Remarks

In this article we present a robust variation source identification method for quality improvement of manufacturing processes. A linear relationship between variation sources (also called process faults) and product quality characteristics is adopted. The columns of the coefficient matrix of this relationship are taken as the unique signatures of the corresponding process faults. Then a method of matching the measurement symptom vectors and the signature vectors is developed. The salient features of this method are: (i) the matching can be applied to the condition when multiple faults simultaneously exist in the system, and (ii) the matching algorithms include the impacts of both the sample uncertainty and perturbation caused by general structured noise. A comprehensive case study illustrated the effectiveness of this method. This method provides a systematic methodology for the root cause identification of manufacturing processes and will significantly shorten the troubleshooting time and improve the quality and productivity of the process.

A very interesting open issue in the pattern matching for variation source identification is the diagnosability issue. In case the angle between two fault signature vectors (i.e., two column vectors of \mathbf{A}) is small, the method may not be able to distinguish them and single out the true fault source. The isolation condition under which all faults are uniquely identifiable is worth further studying. Further, it will be very interesting to compare the diagnosability of the direct estimation method and the pattern matching method. The results along this direction will be reported in the future.

Acknowledgment

The financial support of this work is provided by NSF Award No. DMI-0322147. The authors also appreciate the editor and referees for their valuable comments and suggestions.

Appendix A: Proof of Theorem

To prove the Theorem in Sec. 2.2, we need the following three definitions:

Distance: Suppose \mathbf{S}_1 and \mathbf{S}_2 are subspaces of \mathbf{R}^N , and they

have the same dimension; then the *distance* between these two subspaces is $\text{dist}(\mathbf{S}_1, \mathbf{S}_2) = \|\mathbf{B}_1 - \mathbf{B}_2\|_2$, where $\mathbf{B}_1, \mathbf{B}_2$ are the orthogonal projection to $\mathbf{S}_1, \mathbf{S}_2$ respectively.

Orthogonal projection: Let $\mathbf{S} \subseteq \mathbf{R}^N$ be a subspace, and $\mathbf{P} \in \mathbf{R}^{N \times N}$ is the *orthogonal projection* onto \mathbf{S} if $\text{range}(\mathbf{P}) = \mathbf{S}$, $\mathbf{P}^2 = \mathbf{P}$, and $\mathbf{P}^T = \mathbf{P}$.

It is known that if $\mathbf{v} \in \mathbf{R}^{N \times N}$, and $\mathbf{v}^T \mathbf{v} = \mathbf{I}$; then $\mathbf{P} = \mathbf{v} \mathbf{v}^T$ is the orthogonal projection onto $\mathbf{S} = \text{span}\{\mathbf{v}\}$, where $\text{span}\{\cdot\}$ represents the subspace spanned by \mathbf{v} .

Separation: Let $\mathbf{T}_1 \in \mathbf{R}^{p \times p}$ and $\mathbf{T}_2 \in \mathbf{R}^{(N-p) \times (N-p)}$. The *separation* between the matrices \mathbf{T}_1 and \mathbf{T}_2 is the number $\text{sep}_F(\mathbf{T}_1, \mathbf{T}_2)$, defined by

$$\text{sep}_F(\mathbf{T}_1, \mathbf{T}_2) = \min_{\mathbf{X} \neq \mathbf{0}} \frac{\|\mathbf{T}_1 \mathbf{X} - \mathbf{X} \mathbf{T}_2\|_F}{\|\mathbf{X}\|_F},$$

where $\|\cdot\|_F$ is the Frobenius norm.

Separation is a measure of the distance between eigenvalues of $\mathbf{T}_1, \lambda(\mathbf{T}_1)$, and the eigenvalues of $\mathbf{T}_2, \lambda(\mathbf{T}_2)$. If \mathbf{T}_1 and \mathbf{T}_2 are symmetric matrices, then $\text{sep}_F(\mathbf{T}_1, \mathbf{T}_2) = \min |\lambda(\mathbf{T}_1) - \lambda(\mathbf{T}_2)|$.

Also, to prove our theorem we need a fact as follows (Golub [17]).

Fact: Knowing $\mathbf{A} \Sigma_f \mathbf{A}^T$ and $\Sigma_y = \mathbf{A} \Sigma_f \mathbf{A}^T + \Sigma_\epsilon$ are $N \times N$ symmetric matrices and that $\mathbf{V} = [\mathbf{V}_1 \mathbf{V}_2]$ with $\mathbf{V}_1 = [\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(p)}] \in \mathbf{R}^{N \times p}$ and $\mathbf{V}_2 = [\mathbf{v}^{(p+1)}, \mathbf{v}^{(p+2)}, \dots, \mathbf{v}^{(N)}] \in \mathbf{R}^{N \times (N-p)}$, where $\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \dots, \mathbf{v}^{(N)}$ are eigenvectors of $\mathbf{A} \Sigma_f \mathbf{A}^T$. Partition the matrices $\mathbf{V}^T (\mathbf{A} \Sigma_f \mathbf{A}^T) \mathbf{V}$ and $\mathbf{V}^T \Sigma_\epsilon \mathbf{V}$ as follows:

$$\mathbf{V}^T (\mathbf{A} \Sigma_f \mathbf{A}^T) \mathbf{V} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix}, \quad \mathbf{V}^T \Sigma_\epsilon \mathbf{V} = \begin{bmatrix} \mathbf{E}_{11} & \mathbf{E}_{21}^T \\ \mathbf{E}_{21} & \mathbf{E}_{22} \end{bmatrix}$$

where $\mathbf{D}_1, \mathbf{E}_{11} \in \mathbf{R}^{p \times p}$, and $\mathbf{D}_2, \mathbf{E}_{22} \in \mathbf{R}^{(N-p) \times (N-p)}$. Thus

$$\mathbf{D}_1 = \begin{bmatrix} m^{(1)} & & & \\ & m^{(2)} & & \\ & & \ddots & \\ & & & m^{(p)} \end{bmatrix}$$

References

- [1] Montgomery, D. C., 2001, *Introduction to Statistical Quality Control*, 4th ed., Wiley, New York.
- [2] May, G. S., and Spanos, C. J., 1993, "Automated Malfunction Diagnosis of Semiconductor Fabrication Equipment: A Plasma Etch Application," *IEEE Trans. Semicond. Manuf.*, **6**, pp. 28–40.
- [3] Dunia, R., and Qin, S. J., 1998, "Subspace Approach to Multidimensional Fault Identification and Reconstruction," *AIChE J.*, **44**, pp. 1813–1831.
- [4] Chang, M., and Gossard, D. C., 1998, "Computational Method For Diagnosis of Variation-Related Assembly Problem," *Int. J. Prod. Res.*, **36**, pp. 2985–2995.
- [5] Apley, D. W., and Shi, J., 1998, "Diagnosis of Multiple Fixture Faults in Panel Assembly," *ASME J. Manuf. Sci. Eng.*, **120**, pp. 793–801.
- [6] Zhou, S., Chen, Y., and Shi, J., 2004, "Statistical Estimation and Testing for Variation Root-Cause Identification of Multistage Manufacturing Processes," *IEEE Transactions on Automation Science and Engineering*, **1**, pp. 73–83.
- [7] Ding, Y., Zhou, S., and Chen, Y., 2005, "A Comparison of Process Variance Estimation Methods for In-Process Dimensional Control," *ASME J. Dyn. Syst., Meas., Control*, **127**, pp. 69–79.
- [8] Ceglarek, D., and Shi, J., 1996, "Fixture Failure Diagnosis for Autobody Assembly Using Pattern Recognition," *ASME J. Eng. Ind.*, **118**, pp. 55–65.
- [9] Rong, Q., Ceglarek, D., and Shi, J., 2000, "Dimensional Fault Diagnosis for Compliant Beam Structure Assemblies," *ASME J. Manuf. Sci. Eng.*, **122**, pp. 773–780.
- [10] Ding, Y., Ceglarek, D., and Shi, J., 2002, "Fault Diagnosis of Multistage Manufacturing Processes by Using State Space Approach," *ASME J. Manuf. Sci. Eng.*, **124**, pp. 313–322.
- [11] Li, Z., Zhou, S., and Ding, Y., 2004, "Pattern Matching for Root Cause Identification in Manufacturing Processes With the Presence of Unstructured Noises," *IIE Trans.*, submitted for publication.
- [12] Jin, J., and Shi, J., 1999, "State Space Modeling of Sheet Metal Assembly for Dimensional Control," *ASME J. Manuf. Sci. Eng.*, **121**, pp. 756–762.
- [13] 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*, 23–26 July, Ann Arbor, MI, JUSFA–13146.
- [14] Zhou, S., Huang, Q., and Shi, J., 2001, "State Space Modeling for Dimensional Monitoring of Multistage Machining Process Using Differential Motion Vector," *IEEE Trans. Rob. Autom.*, **19**, pp. 296–309.
- [15] Johnson, R. A., and Wichern, D. W., 2002, *Applied Multivariate Statistical Analysis*, 5th ed., Prentice–Hall, Englewood Cliffs, NJ.
- [16] Apley, D. W., and Shi, J., 2001, "A Factor-Analysis Methods for Diagnosing Variability in Multivariate Manufacturing Processes," *Technometrics*, **43**, pp. 84–95.
- [17] Golub, G. H., and Van Loan, C. F., 1996, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD.
- [18] Krzanoski, W. J., 1979, "Between-Groups Comparison of Principal Components," *J. Am. Stat. Assoc.*, **74**, pp. 703–707.
- [19] Krzanoski, W. J., 1982, "Groups Comparison of Principal Components—Some Sampling Results," *J. Stat. Comput. Simul.*, **15**, pp. 141–154.
- [20] Anderson, T. W., 1963, "Asymptotic Theory for Principal Component Analysis," *Ann. Math. Stat.*, **34**, pp. 122–148.
- [21] Flury, B. N., 1987, "Two Generalizations of the Common Principal Component Model," *Biometrika*, **74**, pp. 59–69.
- [22] Jackson, J. E., 1991, *A User's Guide To Principal Components*, Wiley, New York.