Local ICA for Multivariate Statistical Fault Diagnosis in Systems with Unknown Signal and Error Distributions

Zhiqiang Ge and Lei Xie
Institute of Cyber Systems and Control, Zhejiang University, Hangzhou 310027, P.R. China

Uwe Kruger
Dept. of Chemical Engineering, The Petroleum Institute, Abu Dhabi, U.A.E.

Zhihuan Song
Institute of Cyber Systems and Control, Zhejiang University, Hangzhou 310027, P.R. China

DOI 10.1002/aic.12760
Published online October 12, 2011 in Wiley Online Library (wileyonlinelibrary.com).

Based on a noncausal data structure, this article develops a statistical-based monitoring scheme for diagnosing abnormal situations in complex systems. The recorded variables are assumed to exhibit Gaussian and non-Gaussian signal components, which are monitored using the statistical local approach. For diagnosing abnormal conditions, the paper introduces a regression-based technique that allows estimating the fault contribution from abnormal operating conditions. Application studies involving a simulation example and the analysis of recorded data from an industrial melter process demonstrate that the proposed diagnosis scheme is more sensitive in analyzing incipient fault conditions than existing approaches discussed in the literature. © 2011 American Institute of Chemical Engineers *AIChE J,* 58: 2357–2372, 2012
Keywords: statistical local approach, MLPCA, model plane, fault reconstruction, non-Gaussian signals

Introduction

Most recorded variables of chemical processes exhibit Gaussian and non-Gaussian signal components. The relevant statistical-based research literature has proposed the following linear time-invariant noncausal data structure to model the recorded variables[1–3]:

$$\mathbf{x}_0(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{e}(k) \tag{1}$$

where k is a sampling index, $\mathbf{A} \in \mathbb{R}^{N\times n}$ is a parameter matrix, $\mathbf{s}(k) \in \mathbb{R}^n$ is a vector of $n < N$ zero mean stochastic source signals representing common cause variation, $\mathbf{e}(k) \in \mathbb{R}^N$ is an error vector and $\mathbf{x}_0(k) = \mathbf{x}(k) - \bar{\mathbf{x}}$ is the analyzed data vector with $\bar{\mathbf{x}} \in \mathbb{R}^N$ being a mean vector and $\mathbf{x}(k) \in \mathbb{R}^N$ being the recorded data vector. The error vector is assumed to follow the multivariate Gaussian distribution $\mathbf{e}(k) \sim \mathcal{N}\{\mathbf{0}, \mathbf{S}_{ee}\}$, and is statistically independent of the source signals. The source vector can be divided into $\mathbf{s}^T(k) = (\mathbf{s}_1^T(k)\, \mathbf{s}_2^T(k))$, where $\mathbf{s}_1(k) \in \mathbb{R}^m$ and $\mathbf{s}_2(k) \in \mathbb{R}^{n-m}$, $m \leq n$, are non-Gaussian and Gaussian vectors, respectively.

Narasimhan and Shah[2] showed that the use of maximum likelihood principal component analysis (MLPCA) yields a consistent estimation of the column space of $\mathbf{A}$ by applying a Cholesky decomposition to $\mathbf{S}_{ee} = E\{\mathbf{e}\mathbf{e}^T\} \in \mathbb{R}^{N\times N}$, which is of full rank. Moreover, the reference also introduced an algorithm for simultaneously estimating $\mathbf{S}_{ee}$ and the column space of $\mathbf{A}$. A revised version of this algorithm including the estimation of the number of source signals has recently been proposed by Feital et al.[4] Liu et al.,[3] showed that the source signals are encapsulated in the first n dominant principal components (PCs) and proposed to extract the non-Gaussian components from these PCs using independent component analysis (ICA). A direct application of ICA to the recorded data has also been investigated in recent years.[5–7] However, the geometric simplicity of principal component analysis (PCA) is not maintained by directly applying ICA.[3]

To monitor non-Gaussian signal components, several different approaches have been discussed including kernel density estimation (KDE),[8] support vector data description (SVDD)[3,9] or charting individual independent components in an *ad hoc* fashion.[5] Using a KDE and a SVDD is problematic (i) as they depend on the number of reference samples and (ii) the identification of confidence limits and regions can be time consuming. Moreover, a sparse or clustered data distribution can present difficulties in extracting such limits, as discussed by Chen et al.[8] for a KDE.

This article addresses the problem of efficiently extracting confidence limits/regions for statistics that involve non-Gaussian distributed variables by applying the statistical local approach.[10] The local approach relies on the central limit theorem and transforms the problem of detecting faults in multivariate processes to that of monitoring the mean of a Gaussian vector that is constructed from recorded process data. More precisely, rather than performing a direct analysis of the recorded variables, the statistical local approach examines parametric changes in the data structure of Eq. 1.

The second contribution of this article is the development of a fault diagnosis scheme that is based on the statistics

Correspondence concerning this article should be addressed to L. Xie at leix@iipc.zju.edu.cn and U. Kruger at ukruger@pi.ac.ae.

obtained from the local approach. This scheme relies on a regression approach to determine the magnitude of the fault condition upon the Gaussian and non-Gaussian signal components and is similar to the work discussed in Ref. 11. To isolate the impact of a fault condition from the residual-based monitoring statistic, conventional projection and regression based methods can be utilized.[12,13]

Compared with existing work on monitoring systems that exhibit Gaussian and non-Gaussian signal components, the framework proposed in this article is conceptually simpler for constructing univariate monitoring statistics and diagnosis charts. An additional advantage of incorporating the local approach is that monitoring statistics are based on functions constructed from data within a moving window. This yields an increased sensitivity in detecting incipient fault conditions, as the multivariate local approach is similar in approach to univariate cumulative sum charts,[14,15] which are commonly used in statistical process control applications. These benefits are exemplified by a simulation example and the analysis of recorded data from an industrial melter process.

The rest of this article is organized as follows. Details concerning PCA are presented next. Fault Detection Using a Local ICA Monitoring Approach Section and Implementation of Fault Detection Scheme Section develop monitoring functions for the local approach to construct univariate monitoring statistics that asymptotically follow a Gaussian distribution and introduce the regression-based method for extracting the fault signature from the non-Gaussian signal components, respectively. Simulation Example Section and Application Study to a Melter Process Section then summarize the results of two application studies involving a simulation example and the analysis of recorded data from a glass melter process, respectively. Finally, Concluding Summary Section provides a concluding summary of this article.

Estimating Covariance Matrices, Model Plane, and Signal Components

This section gives a brief overview of PCA and how to simultaneously estimate the column space of $\mathbf{A}$ and the error covariance matrix $\mathbf{S}_{ee}$. MLPCA[2,4] relies on an estimate of the covariance matrix $\hat{\mathbf{S}}_{x_0x_0} = \frac{1}{K-1}\sum_{k=1}^{K}\mathbf{x}_0(k)\mathbf{x}_0^T(k) \in \mathbb{R}^{N\times N} \approx E\{\mathbf{x}_0\mathbf{x}_0^T\}$ and yields estimates of $\hat{\mathbf{S}}_{ee}$ and the column space of $\mathbf{A}$, which is $\mathbf{L}\tilde{\mathbf{P}}$ where $\tilde{\mathbf{P}} \in \mathbb{R}^{N\times n}$ stores the n dominant eigenvectors of $\hat{\tilde{\mathbf{S}}}_{x_ox_o} = \mathbf{L}^{-1}\hat{\mathbf{S}}_{x_ox_o}\mathbf{L}^{-T}$ and $\hat{\mathbf{S}}_{ee} = \mathbf{L}\mathbf{L}^T$ is a Cholesky decomposition.

Feital et al.,[4] highlighted that the column vectors of $\mathbf{L}\tilde{\mathbf{P}}$ do not produce PCs that possess a maximum variance and a diagonal covariance matrix, and showed that a constraint NIPALS algorithm can determine loading vectors that represent a solution to the optimization problem $\hat{\mathbf{p}}_i = \arg\max_{\mathbf{p}_i}\{\mathbf{p}_i^T\hat{\tilde{\mathbf{S}}}_{x_ox_o}\mathbf{p}_i\}$, $1 \le i \le n$, subject to the following constraints $\mathbf{p}_i^T\mathbf{p}_i - 1 = 0$, $\mathbf{p}_i^T[\hat{\mathbf{p}}_1 \cdots \hat{\mathbf{p}}_{i-1}] = \mathbf{0}^T$ and $\mathbf{p}_i - \mathbf{L}\tilde{\mathbf{P}}\boldsymbol{\alpha} = \mathbf{0}$, $\boldsymbol{\alpha} \in \mathbb{R}^n$. An alternate is the computation of the eigendecomposition of $\hat{\tilde{\mathbf{S}}}_{x_ox_o} = \hat{\mathbf{S}}_{x_ox_o} - \hat{\mathbf{S}}_{ee}$,[16,17] which yields n nonzero eigenvalues and $N - n$ zero eigenvalues. The resultant eigenvectors associated with the n nonzero eigenvalues span the PCA model plane, which is the same as the column space of $\mathbf{L}\tilde{\mathbf{P}}$.

Liu et al.,[3] outlined that the source signals are encapsulated within the retained components $\hat{\mathbf{t}}(k) \in \mathbb{R}^n$, $\hat{\mathbf{t}}(k) = \hat{\mathbf{P}}^T\mathbf{x}_0(k)$. Here, $\hat{\mathbf{P}} \in \mathbb{R}^{N\times n}$ stores the n eigenvectors in descending order, i.e., the first one is associated with the largest eigenvalue. This reference also showed that the application of ICA extracts the non-Gaussian components from $\hat{\mathbf{t}}(k)$. Defining $\hat{\mathbf{B}} \in \mathbb{R}^{n\times m}$ as an estimated separation matrix that possesses mutually orthonormal columns, the extracted non-Gaussian signal components $\hat{\mathbf{s}}_1(k) = \hat{\mathbf{B}}^T\boldsymbol{\Lambda}^{-1/2}\hat{\mathbf{t}}(k)$ maximize non-Gaussianity, which is expressed by a differential entropy criterion.[16,18] The diagonal matrix $\boldsymbol{\Lambda}$ stores the eigenvalues of the data covariance matrix. An estimation of the Gaussian components is given by $\hat{\mathbf{s}}_2(k) = [\hat{\mathbf{B}}^\perp]^T\boldsymbol{\Lambda}^{-1/2}\hat{\mathbf{t}}(k)$, where $\hat{\mathbf{B}}^\perp \in \mathbb{R}^{n\times(n-m)}$ is an orthogonal complement of $\hat{\mathbf{B}}$, i.e. $\hat{\mathbf{B}}^T\hat{\mathbf{B}}^\perp = \mathbf{0} \in \mathbb{R}^{m\times(n-m)}$. The estimates of $\mathbf{s}_1(k)$ and $\mathbf{s}_2(k)$ approximate these original sequences up to a similarity transformation:

$$\mathbf{s}_1(k) = \boldsymbol{\Xi}_1\hat{\mathbf{s}}_1(k) - \boldsymbol{\varepsilon}_1(k) \quad \text{and} \quad \mathbf{s}_2(k) = \boldsymbol{\Xi}_2\hat{\mathbf{s}}_2(k) - \boldsymbol{\varepsilon}_2(k) \tag{2}$$

The abbreviated matrices and vectors are: $\boldsymbol{\Xi}_1 = [\hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{A}_1]^{-1} \in \mathbb{R}^{m\times m}$, $\boldsymbol{\varepsilon}_1(k) = \boldsymbol{\Xi}_1\hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{e}(k) \in \mathbb{R}^m$, $\boldsymbol{\Xi}_2 = [[\hat{\mathbf{B}}^\perp]^T\hat{\mathbf{P}}^T\mathbf{A}_2]^{-1} \in \mathbb{R}^{(n-m)\times(n-m)}$, $\boldsymbol{\varepsilon}_2(k) = \boldsymbol{\Xi}_2[\hat{\mathbf{B}}^\perp]^T\hat{\mathbf{P}}^T\mathbf{e}(k) \in \mathbb{R}^{n-m}$, $\mathbf{A} = [\mathbf{A}_1\ \mathbf{A}_2]$ and $\mathbf{A}_1 \in \mathbb{R}^{n\times m}$ and $\mathbf{A}_2 \in \mathbb{R}^{n\times(n-m)}$ are submatrices that correspond to the non-Gaussian and Gaussian signal components, respectively.

The discarded eigenvectors of $\hat{\tilde{\mathbf{S}}}_{x_0x_0}$ are orthogonal to the column space of $\mathbf{L}\tilde{\mathbf{P}}$, maximize the cost function $\hat{\mathbf{p}}_i = \arg\max_{\mathbf{p}_i}\{\mathbf{p}_i^T\hat{\tilde{\mathbf{S}}}_{x_0x_0}\mathbf{p}_i\}$, $n+1 \le i \le N$, and are subject to the following constraints $\mathbf{p}_i^T\mathbf{p}_i - 1 = 0$, $\mathbf{p}_i^T[\hat{\mathbf{p}}_1 \cdots \hat{\mathbf{p}}_n\ \hat{\mathbf{p}}_{n+1} \cdots \hat{\mathbf{p}}_{i-1}] = \mathbf{0}^T$. These eigenvectors, provided that there are no zero values for $\mathbf{p}_i^T\hat{\mathbf{S}}_{x_0x_0}\mathbf{p}_i$, span a complementary residual subspace that describes the distance between $\mathbf{x}_0(k)$ and its orthogonal projection onto the model plane $\tilde{\mathbf{x}}_0(k) = \hat{\mathbf{P}}\hat{\mathbf{P}}^T\mathbf{x}_0(k)$, i.e. $\boldsymbol{e}(k) = \mathbf{x}_0(k) - \tilde{\mathbf{x}}_0(k)$.

Fault Detection Using a Local ICA Monitoring Approach

The first contribution of this article is the utilization of the statistical local approach to monitor the estimated non-Gaussian signal components. Given that these components are extracted using ICA, we refer to this technique as the local ICA, which yields a univariate statistic that asymptotically follows a Gaussian distribution. Hence, it circumvents the application of KDE and SVDD, which can be time consuming. The next two subsections detail the derivation of monitoring functions and the development of the monitoring statistic. Appendix A gives a brief introduction of ICA.

Derivation of primary residuals

The derivative of primary residuals, which are subsequently used to construct monitoring functions, commences with a review of the ICA cost function. Next, following a summary of the assumptions imposed on primary residuals, Derivation of primary residuals third subsection defines primary residuals, based on the ICA cost functions, and shows that they meet the assumptions for primary residuals.

ICA Cost Function. Defining $\mathbf{y}(k) = \hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\mathbf{t}}(k) = \boldsymbol{\Lambda}^{-1/2}\hat{\mathbf{P}}^T\mathbf{x}_0(k) \in \mathbb{R}^n$ and $\hat{\mathbf{b}}_i \in \mathbb{R}^n$ as the ith column of the separation matrix $\hat{\mathbf{B}}_i$, Hyvarinen[16,18] showed that the negentropy of $\hat{\mathbf{b}}_i^T\mathbf{y}$ can be approximated by $[E\{G(\hat{\mathbf{b}}_i^T\mathbf{y})\} - E\{G(\upsilon)\}]^2$, where $G(\cdot)$ is a nonquadratic function and υ is a Gaussian variable. Incorporating the constraints $\hat{\mathbf{b}}_i^T\hat{\mathbf{b}}_i - 1 = 0$ and $\hat{\mathbf{b}}_i^T[\hat{\mathbf{b}}_1 \cdots \hat{\mathbf{b}}_{i-1}] = \mathbf{0}^T$, the cost function becomes:

$$\hat{\mathbf{b}}_i = \arg\max_{\mathbf{b}_i}\{[E\{G(\mathbf{b}_i^T\mathbf{y})\} - E\{G(\upsilon)\}]^2 - (\mathbf{b}_i^T\mathbf{b}_i - 1)\beta_i - \mathbf{b}_i^T[\hat{\mathbf{b}}_1 \cdots \hat{\mathbf{b}}_{i-1}]\gamma_i\} \tag{3a}$$

$$\hat{\mathbf{b}}_i = \arg\left\{ \frac{\partial}{\partial \mathbf{b}_i} \{[E\{G(\hat{\mathbf{b}}_i^T\mathbf{y})\} - E\{G(\upsilon)\}]^2 - (\mathbf{b}_i^T\mathbf{b}_i - 1)\beta_i - \mathbf{b}_i^T[\hat{\mathbf{b}}_1 \cdots \hat{\mathbf{b}}_{i-1}]\gamma_i\}\right\} = \mathbf{0} \quad (3b)$$

Here, $\beta_i \in \mathbb{R}$ and $\gamma_i \in \mathbb{R}^{i-1}$ are Lagrangian multipliers. Hyvarinen[16,18] introduced a deflation procedure that subtracts the contribution of the ith independent component from the original set of non-Gaussian variables prior to the determination of the $(i + 1)$th component. This guarantees that the column vectors of $\hat{\mathbf{B}}$ are mutually orthogonal. Hence, the orthogonality constraint $\mathbf{b}_i^T[\hat{\mathbf{b}}_1 \cdots \hat{\mathbf{b}}_{i-1}] = \mathbf{0}^T$ can be removed from Eq. 3b, which consequently becomes:

$$\hat{\mathbf{b}}_i = \arg\{E\{\boldsymbol{\phi}_i(\mathbf{b}_i, \mathbf{y})\}\} = \mathbf{0} \quad (4)$$

Here, $E\{\boldsymbol{\phi}_i(\mathbf{b}_i, \mathbf{y})\} = [E\{G(\mathbf{b}_i^T\mathbf{y})\} - E\{G(\upsilon)\}]\frac{\partial G(\mathbf{b}_i^T\mathbf{y})}{\partial(\mathbf{b}_i^T\mathbf{y})}\mathbf{y} - \beta_i\mathbf{b}_i \in \mathbb{R}^n$. Eq. 4 allows the definition of the following monitoring function $\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})$.

Assumptions for Primary Residuals. Basseville[10] formulated two basic requirements for the definition of primary residuals. The first one is the existence of a vector-valued function, which must be differentiable with respect to a parameter vector. Eq. 4 presents such a vector-valued function for $\mathbf{b}_i$. The second basic requirement is the existence of a neighborhood for $\hat{\mathbf{b}}_i$, defined by $\boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)$, for which

$$\begin{aligned} E\{\boldsymbol{\phi}_i(\mathbf{b}_i, \mathbf{y})\} &= 0 \quad \text{if} \quad \mathbf{b}_i = \hat{\mathbf{b}}_i \\ E\{\boldsymbol{\phi}_i(\mathbf{b}_i, \mathbf{y})\} &\neq 0 \quad \text{if} \quad \mathbf{b}_i = \boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)\backslash\hat{\mathbf{b}}_i. \end{aligned} \quad (5)$$

The symbol $\backslash$ implies that the element $\hat{\mathbf{b}}_i$ is not a member of the set $\boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)$. In the presence of a fault condition, the parameter vector $\mathbf{b}_i$ will depart from its estimated value $\hat{\mathbf{b}}_i$. Formulating this change by $\mathbf{b}_i = \hat{\mathbf{b}}_i + \Delta\mathbf{b}_i/\mathcal{K}$, $\mathcal{K}$ being the number of reference samples, allows devising improved residuals that asymptotically follow a Gaussian distribution with a mean value of $\mathbf{M}(\hat{\mathbf{b}}_i)\Delta\mathbf{b}_i$, where

$$\mathbf{M}(\hat{\mathbf{b}}_i) = -E\left\{\partial\boldsymbol{\phi}(\mathbf{b}_i, \mathbf{y})/\partial\mathbf{b}_i\big|_{\mathbf{b}_i=\hat{\mathbf{b}}_i}\right\}. \quad (6)$$

Definition of improved residuals Subsection details the definition of improved residuals for $\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})$. An additional requirement is that the matrix $\mathbf{M}(\hat{\mathbf{b}}_i)$ has full rank.

Definition of Primary Residuals. Defining $\boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i) \in \mathbb{R}^n$ as the vicinity of the estimated parameter vector $\hat{\mathbf{b}}_i$ with $\hat{\mathbf{b}}_i \notin \boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)$, Eq. 5 implies that $E\{\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})|_{\mathbf{b}_i=\hat{\mathbf{b}}_i}\} = \mathbf{0}$ and $E\{\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})|_{\mathbf{b}_i\in\boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)}\} \neq \mathbf{0}$. Moreover, $\boldsymbol{\phi}_i(\hat{\mathbf{b}}_i,\mathbf{y})$ is differentiable w.r.t. $\mathbf{b}_i$ and $\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})$ exists $\forall\mathbf{b}_i \in \boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)$. A monitoring function can be obtained for each of the m parameter vectors $\hat{\mathbf{b}}_1,\ldots,\hat{\mathbf{b}}_m$ and storing them in a vector yields:

$$\boldsymbol{\Phi}^T(\mathbf{B}, \mathbf{y}) = (\boldsymbol{\phi}_1^T(\mathbf{b}_1, \mathbf{y}) \ \cdots \ \boldsymbol{\phi}_m^T(\mathbf{b}_m, \mathbf{y})) \in \mathbb{R}^{mn} \quad (7)$$

Here, $E\{\boldsymbol{\phi}(\mathbf{B},\mathbf{y})|_{\mathbf{B}=\hat{\mathbf{B}}}\} = 0$ and $\mathrm{E}\{\boldsymbol{\phi}(\mathbf{B},\mathbf{y})|_{\mathbf{B}\in\boldsymbol{\Omega}(\hat{\mathbf{B}})}\} \neq \mathbf{0}$, $\boldsymbol{\Omega}(\hat{\mathbf{B}}) \in \mathbb{R}^{mn}$. Given that $\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})$ is differentiable w.r.t. $\mathbf{b}_i$ for $1 \leq i \leq m$ and that $\boldsymbol{\phi}_i(\mathbf{b}_i,\mathbf{y})$ exists within the sphere $\boldsymbol{\Omega}_i(\hat{\mathbf{b}}_i)$, these properties remain for $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$ and $\boldsymbol{\Omega}(\hat{\mathbf{B}})$, respectively. To reduce the dimension of $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$, it can be simplified by premultiplying it with $\mathbf{b}_i^T$, which gives rise to:

$$\mathbf{b}_i^T\boldsymbol{\phi}_i(\mathbf{b}_i, \mathbf{y}) =: \theta_i(\mathbf{b}_i, \mathbf{y}) \quad (8)$$

Since $\hat{s_1}_i = \mathbf{b}_i^T\mathbf{y}$, and $\mathbf{b}_i^T\mathbf{b}_i = 1$, Eq. 8 becomes:

$$\theta_i(\mathbf{b}_i, \mathbf{y}) = \hat{s}_{1i}g(\hat{s}_{1i}) - \beta_1 \quad (9)$$

Here, $\hat{s_1}_{,i}$ is the ith element of $\hat{\mathbf{s}}_1$ and $g(\hat{s}_{1i}) = [E\{G(\hat{s}_{1i})\} - E\{G(\upsilon)\}]\frac{\partial G(\hat{s}_{1i})}{\partial(\hat{s}_{1i})} \in \mathbb{R}$. Similar to Eq. 7, an augmented monitoring vector can be formulated as follows:

$$\boldsymbol{\theta}^T(\mathbf{B}, \mathbf{y}) = (\theta_1(\mathbf{b}_1, \mathbf{y}) \cdots \theta_m(\mathbf{b}_m, \mathbf{y})) \in \mathbb{R}^m \quad (10)$$

According to Eq. 8, $\boldsymbol{\theta}^T(\mathbf{B},\mathbf{y})$ is a linear combination of $\boldsymbol{\phi}^T(\mathbf{B},\mathbf{y})$. Consequently, $E\{\boldsymbol{\theta}(\mathbf{B})|_{\mathbf{B}=\hat{\mathbf{B}}}\} = \mathbf{0}$, $E\{\boldsymbol{\theta}(\mathbf{B},\mathbf{y})|_{\mathbf{B}\in\boldsymbol{\Omega}(\hat{\mathbf{B}})}\} \neq \mathbf{0}$, $\boldsymbol{\theta}(\hat{\mathbf{B}},\mathbf{y})$ is differentiable w.r.t. $\mathbf{B}$ and $\boldsymbol{\theta}(\mathbf{B},\mathbf{y})$ exists if $\mathbf{B} \in \boldsymbol{\Omega}(\hat{\mathbf{B}})$. Consequently, $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$ and $\boldsymbol{\theta}(\mathbf{B},\mathbf{y})$ meet the required properties for constructing primary residuals.[10] Given that the dimensions of $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$ and $\boldsymbol{\theta}(\mathbf{B},\mathbf{y})$ are mn and m, respectively, it is important to examine whether the use of the higher dimensional vector $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$ is required or whether the utilization of the lower dimensional $\boldsymbol{\theta}(\mathbf{B},\mathbf{y})$ is sufficient in detecting incipient fault conditions. This gives rise to Theorem 1, which is proven in Appendix B.

Theorem 1. Both primary residuals, $\boldsymbol{\phi}(\mathbf{B},\mathbf{y})$ and $\boldsymbol{\theta}(\mathbf{B},\mathbf{y})$, are sensitive in detecting whether $\mathbf{B} \in \boldsymbol{\Omega}(\hat{\mathbf{B}})$.

Definition of improved residuals

Defining $\boldsymbol{\theta}(\hat{\mathbf{B}},\mathbf{y})$ as the vector-valued primary residual function that represents the process *in-statistical-control* gives rise to the following definition of the improved residual vector[10]:

$$\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, \mathcal{K}) = \frac{1}{\sqrt{\mathcal{K}}}\sum_{k=1}^{\mathcal{K}}\boldsymbol{\theta}\big(\hat{\mathbf{B}}, \mathbf{y}(k)\big) \quad (11)$$

where $\mathcal{K}$ is the number of samples and $\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, \mathcal{K})$ is the improved residual. Here, the subscript *ng* refers to "non-Gaussian." According to the central limit theorem, the distribution function of the improved residual vector $\boldsymbol{\zeta}_{ng}(\cdot)$ follows asymptotically a multivariable Gaussian distribution:

$$\lim_{\mathcal{K}\to\infty}\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, \mathcal{K}) \sim \mathcal{N}\{\mathbf{0}, \mathbf{S}_{\boldsymbol{\theta}\boldsymbol{\theta}}\} \quad (12)$$

Here, $\mathbf{S}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ is the covariance matrix of $\boldsymbol{\zeta}_{ng}(\cdot)$. If the process represents an *out-of-statistical-control* situation, the parameter matrix $\hat{\mathbf{B}} \to \mathbf{B}$, with $\mathbf{B} = \hat{\mathbf{B}} + \Delta\mathbf{B}$, and has the distribution function $\boldsymbol{\zeta}_{ng}(\mathbf{B}, \mathcal{K}) \sim \mathcal{N}\{\mathbf{M}(\hat{\mathbf{B}})\Delta\mathbf{B}, \mathbf{S}_{\boldsymbol{\theta}\boldsymbol{\theta}}\}$ with $\mathbf{M}(\hat{\mathbf{B}}) = -\frac{\partial\boldsymbol{\zeta}(\mathbf{B},\mathcal{K})}{\partial\mathbf{B}}\Big|_{\mathbf{B}=\hat{\mathbf{B}}}$.[10] More precisely, a small incipient change in the parameter, $\mathbf{B} = \hat{\mathbf{B}} + \Delta\mathbf{B}$, translates into a change of $E\{\boldsymbol{\zeta}_{ng}(\mathbf{B}, \mathcal{K})\}$ from $\mathbf{0}$ to $\mathbf{M}(\hat{\mathbf{B}})\Delta\mathbf{B} \neq \mathbf{0}$ and is therefore detectable. It is also interesting to note that the distribution function of $\boldsymbol{\zeta}_{ng}(\cdot)$ has the same covariance matrix irrespective of whether $\mathbf{B} = \hat{\mathbf{B}}$ or $\mathbf{B} = \hat{\mathbf{B}} + \Delta\mathbf{B}$.

Definition of monitoring statistics

For detecting an *out-of-statistical-control* situation it is therefore sufficient to monitor a Hotelling's T^2 statistic that is constructed from the improved residual vector $\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, \mathcal{K})$:

$$T_{ng}^2(\boldsymbol{\zeta}_{ng}, \mathcal{K}) = \boldsymbol{\zeta}_{ng}^T(\boldsymbol{\theta}, \mathcal{K})\mathbf{S}_{\boldsymbol{\theta}\boldsymbol{\theta}}^{-1}\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, \mathcal{K}). \quad (13)$$

In practice, the improved residuals can be constructed from primary residuals within a moving window. This increases the sensitivity of the statistical local approach, which may reduce if the number of data samples becomes large. Defining the size of the moving window by K, the improved residual for the kth sample is as follows:

$$\boldsymbol{\zeta}_{ng}(\boldsymbol{\theta}, k) = \frac{1}{\sqrt{K}}\sum\nolimits_{j=k-K+1}^{k}\boldsymbol{\theta}\big(\hat{\mathbf{B}}, \mathbf{y}(j)\big) \quad (14)$$

Using $\zeta_{ng}(\theta,k)$ in Eq. 14 allows constructing the univariate T^2 monitoring statistic.

Incorporating $\hat{\mathbf{s}}_2(k) = [\hat{\mathbf{B}}^T]^{\perp}\mathbf{y}(k)$ into the statistical local approach gives rise to the definition of the Hotelling's T^2 statistic $T_g^2(k) = \zeta_g^T(\vartheta,k)\mathbf{S}_{\vartheta\vartheta}^{-1}\zeta_g(\vartheta,k)$. Here, $\hat{\mathbf{S}}_{\vartheta\vartheta}$ is the estimated covariance matrix of $\vartheta = (\vartheta_1 \cdots \vartheta_{n-m})^T$, $\vartheta_i(\hat{\mathbf{b}}_i^{\perp}, \mathbf{y}(j)) = (\mathbf{y}^T(j)\hat{\mathbf{b}}_i^{\perp})^2 - E\{(\mathbf{y}^T\hat{\mathbf{b}}_i^{\perp})^2\}$ and $\zeta_g(\vartheta,k) = \frac{1}{\sqrt{K}}\sum_{j=k-K+1}^{k} \vartheta(\hat{\mathbf{B}}^{\perp}, \mathbf{y}(j))$. The main benefit of relying on a Hotelling's T^2 statistic based on the statistical local approach instead utilizing the Gaussian signal components is the increased sensitivity for detecting incipient fault conditions.[11] The univariate statistics T_{ng}^2 and T_g^2 monitor common cause variation and therefore the estimated source signals $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$, respectively.

The MLPCA residuals enable the definition of a third univariate statistic, $T_e^2(k) = \zeta_e^T(\psi,k)\hat{\mathbf{S}}_{\psi\psi}^{-1}\zeta_e(\psi,k)$, which monitors the minimal distance of the recorded samples from the model plane on the basis of the statistical local approach. The primary residual for the residual-based T_e^2 statistic is $\psi = (\psi_1 \cdots \psi_{N-n})^T$, $\psi_{i\_n}(\hat{\mathbf{p}}_i,\mathbf{x}_0(j)) = (\hat{\mathbf{p}}_i^T\mathbf{x}_0(j))^2 - \lambda_i$, where λ_i is the ith largest discarded eigenvalue, $n+1 \le i \le N$. The improved residual vector is then $\zeta_e(\psi,k) = \frac{1}{\sqrt{K}}\Sigma_{j=k-K+1}^{k}\psi(\hat{\mathcal{P}},\mathbf{x}_0(j))$. With respect to the discussion by Tracey et al.,[19] the above univariate monitoring statistics follows an F-distribution if the estimates of the covariance matrices $\hat{\mathbf{S}}_{\theta\theta}^{-1}$, $\hat{\mathbf{S}}_{\vartheta\vartheta}^{-1}$, and $\hat{\mathbf{S}}_{\psi\psi}^{-1}$ are statistically independent of the improved residuals $\zeta_{ng}(\theta,k)$, $\zeta_g(\vartheta,k)$, and $\zeta_e(\psi,k)$, respectively. The confidence limits can therefore be easily obtained for a significance of 1 or 5%. With regards to the estimation of the covariance matrices, $\hat{\mathbf{S}}_{\theta\theta}^{-1}$, $\hat{\mathbf{S}}_{\vartheta\vartheta}^{-1}$, and $\hat{\mathbf{S}}_{\psi\psi}^{-1}$, it is important to note that the estimation variance of the elements of the covariance matrix can be considerable for small sample sets. This, however, is a side issue, as recorded data from industrial process systems in the chemical industry are usually data rich, which is reported in numerous articles over the past few decades. Recent work[20,21] involve the estimation of covariance matrix for small sample sets.

Implementation of fault detection scheme

The implementation of the proposed monitoring scheme involves the following steps.

(1) Record a sufficiently large set of normal operating data from the process (not describing any fault scenario);

(2) Divide the data into two sets (data sets 1 and 2) before mean centering and scaling each set;

(3) Obtain a MLPCA model[2,4] using data set 1;

(4) Validate the MLPCA model using data set 2;

(5) Identify an ICA model [3,4,16,18,22] using the MLPCA scores obtained from data set 1;

(6) Construct the primary residuals for the non-Gaussian, the Gaussian and the MLPCA residuals as discussed in Derivation of Primary Residuals Section and Definition of monitoring statistics Section for data sets 1 and 2;

(7) Compute the improved residuals for the non-Gaussian, the Gaussian and the MLPCA residuals from the primary residuals, as discussed in Definition of Improved Residuals Section and Definition of monitoring statistics Section for data sets 1 and 2;

(8) Estimate the covariance matrices for each of the primary residuals from data set 2;

(9) Set up univariate monitoring statistics for the non-Gaussian, the Gaussian and the MLPCA residuals by determining the confidence limits from and F-distribution and utilizing the estimated covariance matrices from Step 8.

Fault Diagnosis Using a Local ICA Monitoring Approach

If some abnormal event has been successfully detected, the next step is to determine its root cause. This section summarizes the mathematics of the proposed fault reconstruction scheme for the non-Gaussian signal components and introduces an associated fault diagnosis index.

Fault reconstruction scheme

For introducing the proposed fault reconstruction method for non-Gaussian signal components, we assume that the fault subspace is known a priori. Each of the J possible fault conditions, denoted here by $\mathcal{F}_j, j = 1,2,\cdots,J$, is described by such a fault subspace. If recorded data describing specific fault conditions is available, Yue and Qin[23] discussed the application of a singular value decomposition to estimate the fault subspace. If such data is not available, however, a detailed understanding of the process is required to estimate the fault directions for specific fault conditions. Such understanding can be provided by experienced plant personnel, which is exemplified in Application Study to a Melter Process Section. On the basis of the fault subspace, the uncorrupted data $\mathbf{x}_0$ can be reconstructed from the corrupted one $\mathbf{x}_0^*$:

$$\hat{\mathbf{x}}_0(k) = \mathbf{x}_0^*(k) - \mathfrak{F}_j f_i \tag{15}$$

where $\hat{\mathbf{x}}_0(k)$ is the reconstruction of $\mathbf{x}_0$, $\mathfrak{F}_j$ is the corresponding fault subspace of $\mathcal{F}_j$ and f_j is the magnitude of the fault. Appendix C shows the derivative of an iterative algorithm that allows the determination of the fault magnitude f_j Eq. 16 shows the resultant formula for one iteration step. The next subsection illustrates how to diagnose whether the assumed fault condition represents the root cause of an abnormal event.

$$\hat{f}_j(t+1) = \frac{\zeta^T(\theta,K)\mathbf{S}_{\theta\theta}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}(g(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_j\hat{f}_j(t)) + \hat{s}_{11}^*(k)g'(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_j\hat{f}_j(t)))\mathbf{w}_1^T\mathfrak{F}_j \\ (g(\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_j\hat{f}_j(t)) + \hat{s}_{12}^*(k)g'(\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_j\hat{f}_j(t)))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ (g(\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_j\hat{f}_j(t)) + \hat{s}_{1m}^*(k)g'(\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_j\hat{f}_j(t)))\mathbf{w}_m^T\mathfrak{F}_j\end{bmatrix}}{\zeta^T(\theta,K)\mathbf{S}_{\theta_0\theta_0}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}\mathfrak{F}_j^T\mathbf{w}_1 g'(\hat{s}_{11}^*(k) - \mathbf{w}_1\mathfrak{F}_j\hat{f}_j(t))\mathbf{w}_1^T\mathfrak{F}_j \\ \mathfrak{F}_j^T\mathbf{w}_2 g'(\hat{s}_{12}^*(k) - \mathbf{w}_2\mathfrak{F}_j\hat{f}_j(t))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ \mathfrak{F}_j^T\mathbf{w}_m g'(\hat{s}_{1m}^*(k) - \mathbf{w}_m\mathfrak{F}_j\hat{f}_j(t))\mathbf{w}_m^T\mathfrak{F}_j\end{bmatrix}} \tag{16}$$

Fault diagnosis

The estimation of f_j is followed by assessing the impact of the corresponding fault subspace upon the measured vector $\mathbf{x}_0(k)$. More precisely, the diagnosis of the detected fault condition requires an assessment metric to determine the most likely root cause for the anomalous event. After reconstructing each of the potential fault subspaces (assumed conditions), the improved residuals can be computed along with the corresponding non-Gaussian statistic T_{ng}^2. This allows an inspection of how significant the jth fault condition affects the non-Gaussian signal components:

$$\hat{T}_{ng_j}^2(\hat{\zeta}_j,k) = \hat{\zeta}_j^T(\theta,k)\mathbf{S}_{\theta\theta}^{-1}\hat{\zeta}_j(\theta,k) \tag{17}$$

Here, $\hat{\boldsymbol{\zeta}}_j(\boldsymbol{\theta},k)$ is the reconstructed improved residual vector w.r.t. the jth fault condition $\mathcal{F}_j$ and $\hat{T}^2_{ng_j}$ the associated reconstructed Hotelling's T^2 statistic for the non-Gaussian signal components. If this univariate statistic reduces significantly in value and confirms an *in-statistical-control* situation the fault was successfully reconstructed and the jth fault condition $\mathcal{F}_j$ is the root cause of the detected process abnormality. In contrast, if $\hat{T}^2_{ng_j}$ still shows a statistically significant number of violations the assumed fault condition may not be the actual root cause.

To determine a metric that quantifies whether an assumed fault condition is the actual root case, the reconstructed value for $\hat{T}^2_{ng_j}(\hat{\boldsymbol{\zeta}}_j,k)$ can be divided by the confidence limit for the T^2_{ng} statistic to produce:

$$\eta_{ng}(k,j)=\frac{\hat{T}^2_{ng_j}(\hat{\boldsymbol{\zeta}}_j,k)}{T^2_{ng,lim}},\quad j=1,2,\cdots,J \tag{18}$$

If $\eta_{ng}(k,j) \leq 1$ the improved residual vector $\hat{\boldsymbol{\zeta}}_j(\boldsymbol{\theta},k)$ corresponds to an *in-statistical-control* situation and vice versa. In a similar fashion, the performance indices $\eta_g(k,j)$ and $\eta_e(k,j)$ can be established:

$$\eta_g(k,j)=\frac{\hat{T}^2_{g_j}(\hat{\mathbf{s}}_{2_j}(k))}{T^2_{g,lim}}\quad \eta_e(k,j)=\frac{\hat{T}^2_{e_j}(e(k))}{T^2_{e,lim}},\quad j=1,2,\cdots,J \tag{19}$$

The next two sections show application studies demonstrating the working of the proposed monitoring scheme.

Simulation Example

This section presents a simulation example involving a total of six process variables, which are linear combinations of three source variables, corrupted by Gaussian distributed measurement noise:

$$\mathbf{x}_0=\mathbf{A}\mathbf{s}+\mathbf{e}$$

$$\mathbf{A}=\begin{bmatrix} 0.145 & 0.946 & 0.210\\ 0.932 & -0.125 & -0.135\\ 0.466 & 0.203 & 0.305\\ -0.419 & 0.672 & 0.013\\ 0.146 & 0.392 & 0.256\\ 0.683 & -0.482 & 0.105 \end{bmatrix}$$

$$\mathbf{s}=\begin{pmatrix} s_1\\ s_2\\ s_3 \end{pmatrix}\qquad \mathbf{e}\sim\mathcal{N}\left\{\begin{pmatrix} 0\\0\\0\\0\\0\\0 \end{pmatrix},\begin{bmatrix} 0.05 & 0 & 0 & 0 & 0 & 0\\ 0 & 0.045 & 0 & 0 & 0 & 0\\ 0 & 0 & 0.05 & 0 & 0 & 0\\ 0 & 0 & 0 & 0.05 & 0 & 0\\ 0 & 0 & 0 & 0 & 0.07 & 0\\ 0 & 0 & 0 & 0 & 0 & 0.05 \end{bmatrix}\right\} \tag{20}$$

The first two source variables, s_1 and s_2, follow a uniform (non-Gaussian) distribution of zero mean. The third one follows a zero mean Gaussian distribution of unity variance. To contrast the proposed fault diagnosis scheme with traditional ICA[5–7] and ICA-SVDD,[3] a total of 2000 samples were generated on the basis of Eq. 20.

After determining an MLPCA model and extracting the two independent components, an appropriate window size for constructing the improved residual vector $\boldsymbol{\zeta}(\boldsymbol{\theta},k)$ was determined to be $K = 100$. Choosing a smaller window length produced improved residuals that significantly departed from a Gaussian distribution. In contrast, selecting a larger size may have rendered the improved residuals less sensitive in detecting abnormal process behavior, which follows from the discussion on page 2585 by Kruger and Dimitriadis.[11] As highlighted in Fault diagnosis using a local ICA monitoring approach Subsection, the moving window approach was applied to the non-Gaussian, the Gaussian and the MLPCA residual components constructing the T^2_{ng}, the T^2_g, and the T^2_e statistics, respectively.

The SVDD approach relates to a nonlinear transformation of the non-Gaussian source variables into a feature space.[24,25] The nonlinear mapping between these spaces is such that the transformed score variables fall within a hypersphere of minimal radius with a likelihood of 95 or 99%. This mapping is determined by parameterized kernel functions. Tax and Duin[24] advocated the use of Gaussian kernel functions for which the associated scaling parameter was determined to be 20 in this example. The squared distance of the transformed variables from the center of the hypersphere defines a univariate statistic denoted here as the D^2 statistic for which the radius of the hypersphere is the confidence limit to test the null hypothesis that the process is *in-statistical-control*. The Gaussian source signal components are utilized here to define a standard Hotelling's T^2 statistic.[19] The residuals remaining after deflating the contribution of the estimated source signals form a third univariate squared prediction error (SPE). Confidence limits for these three monitoring statistics can be obtained.[19,26,27] Different from the local ICA approach, the ICA-SVDD monitoring statistics do not rely on the statistical local approach.

Monitoring this simulated process using conventional ICA relied on a univariate monitoring statistics constructed from the extracted non-Gaussian signal components and the residuals generated after deflating the contribution of the

Table 1. Summary of Critical Values of Univariate Statistics Used for Each Method

	Non-Gaussian		Gaussian		Residual	
Statistic Method	Notation	Critical Value	Notation	Critical Value	Notation	Critical Value
Local ICA	T_{ng}^2	F-distribution[21]	T_g^2	F-distribution[21]	T_e^2	F-distribution[21]
ICA-SVDD	D^2	Squared radius[28]	T^2	F-distribution[21]	SPE	Approximate χ^2-distribution[23]
ICA	I^2	Estimated distribution[5]	N/A	N/A	SPE	Approximate χ^2-distribution[24]

non-Gaussian components. The univariate statistic established from the non-Gaussian signal components and the residuals are denoted here by I^2 and SPE, respectively. In contrast to the ICA-SVDD approach, the conventional ICA technique extracts the non-Gaussian components directly from the recorded variable set $\mathbf{x}_0$.[5–7] The residuals, therefore, incorporate Gaussian source signal components as well as a contribution of the error variables. The problem of determining a confidence limit for univariate statistics constructed from stationary but non-Gaussian variables has been discussed by Martin and Morris[28] by estimating the probability density function. Utilizing the estimate of this function, a confidence limit can be obtained by a numerical integration. Table 1 summarizes the calculation of the univariate monitoring statistics and the associated confidence limits for each of the methods.

Each monitoring model was determined using the first 1000 samples as reference data. The first step was to establish a MLPCA model for the local ICA (LICA) and ICA-SVDD approaches, followed by the application of ICA to three retained scores yielding two independent components and one Gaussian component. For the ICA monitoring method a direct application of ICA to the reference data suggested, as expected, the inclusion of two independent components. The next step was the determination of the confidence limits, or critical values, for each statistic, as outlined in Table 1. The covariance matrices for the T_{ng}^2, the T_g^2, and the T_e^2 statistics of the LICA approach and the T^2 statistic for the ICA-SVDD approach were estimated using the second 1000 samples of the simulated data set to ensure that these statistics follow an F-distribution.[19] The control limits for the SPE statistics for the ICA-SVDD and the ICA approach were also determined on the basis of the latter half of the simulated data using the approximations.[26,27] Finally, the probability density function of the I^2 statistic has been estimated using the values produced from the second 1000 samples of the generated data set followed by numerically determining the critical value for a significance of 1%.

To test the performance of each method, a second data set of 2000 samples was simulated using Eq. 20 and altered to produce the following three cases:

Case 1. No manipulation, so that $\mathbf{x}_{0_1} = \mathbf{As} + \mathbf{e}$ with $\mathbf{x}_{0_1}$ being the first test set;

Case 2. Rotating the mixing matrix by 15° to affect the non-Gaussian source signal components, such that:

$$\mathbf{A}_{rot} = \mathbf{A}\begin{bmatrix} 0.966 & -0.259 & 0 \\ 0.259 & 0.966 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{21}$$

and $\mathbf{x}_{0_2} = \mathbf{A}_{rot}\mathbf{s} + \mathbf{e}$ where $\mathbf{x}_{0_2}$ represents the second half of the manipulated set; and

Case 3. A superimposed sensor bias of 1.2 to the first process variable, i.e.

$$\mathbf{x}_{0_3} = \mathbf{As} + \mathbf{e} + (1.2 \;\; 0 \;\; 0 \;\; 0 \;\; 0 \;\; 0)^T = \mathbf{As} + \mathbf{e} + \mathbf{f} \tag{22}$$

where $\mathbf{x}_{0_2}$ describes this sensor bias to the latter half of the third test set.

Figure 1 shows the results of applying each monitoring model to the first case, i.e., the normal operation that did not describe any fault condition. As expected, none of the univariate statistics showed a statistically significant number of violations for this set.

Figure 2 summarizes the results of applying the LICA (upper left plot), the ICA-SVDD (upper right plot) and the ICA (lower left plot) methods to the data set representing Case 2. The 15° rotation for constructing the non-Gaussian source signals can be detected by the T_{ng}^2 statistic of the proposed LICA approach. However, none of the other statistics were sensitive to this condition.

The geometric interpretation of these results is that the projections of the manipulated data set $\mathbf{x}_{0_2}$ onto the model plane did not generate an increase in the mismatch between the original and projected variables, since none of the residual statistics were sensitive. This finding is correct, since no manipulations of the error signals were made and the conducted rotation of the column space of $\mathbf{A}$ did not change the orientation of the model plane either. However, the projections of the variable set $\mathbf{x}_{0_2}$ described different directions as the first two columns of the matrix $\mathbf{A}$ were different. Hence, the principal directions representing a maximum covariance were different to those identified by the MLPCA model. More precisely, the change in the principal directions is governed by the column space of $\hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{A}_1 \neq \hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{A}_{1,rot}$, which directly follows from Eq. 21. That the direction describing the Gaussian signal components remained unchanged results from $\hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{A}_2 = \hat{\mathbf{B}}^T\hat{\mathbf{P}}^T\mathbf{A}_{2,rot}$, which, again, follows from Eq. 21.

In contrast to the T_{ng}^2 statistic, the D^2 and I^2 statistics did not show a statistically significant number of violations. For the SVDD approach, the nonlinear transformation of the estimated non-Gaussian source signals into the feature space did not produce transformed scores that fell outside the constructed sphere and hence, this fault remained undetected. Following from the discussion by Kruger et al.,[22] a slight rotation of the base vectors spanning the PCA model plane may not be detectable by score-based statistics. That the rotation by 15° cannot be detected by the independent components extracted from the ICA-SVDD and the conventional ICA models confirms the finding by Kruger et al.,[22] for non-Gaussian variables.

Incorporating the statistical local approach, however, detected this change. This follows from the fact that the direction of each column vector of the estimated separation matrix is monitored. In contrast, the ICA-SVDD and the conventional ICA approaches only monitor the projections of the PCs and the original variable set, respectively, which may not produce a significant number of violations if the rotation angle is small. Increasing the angle of rotation to 20° or more resulted in significant violations of the D^2 and I^2 statistics. The application of the each method to Case 2 confirmed that the incorporation of the statistical local

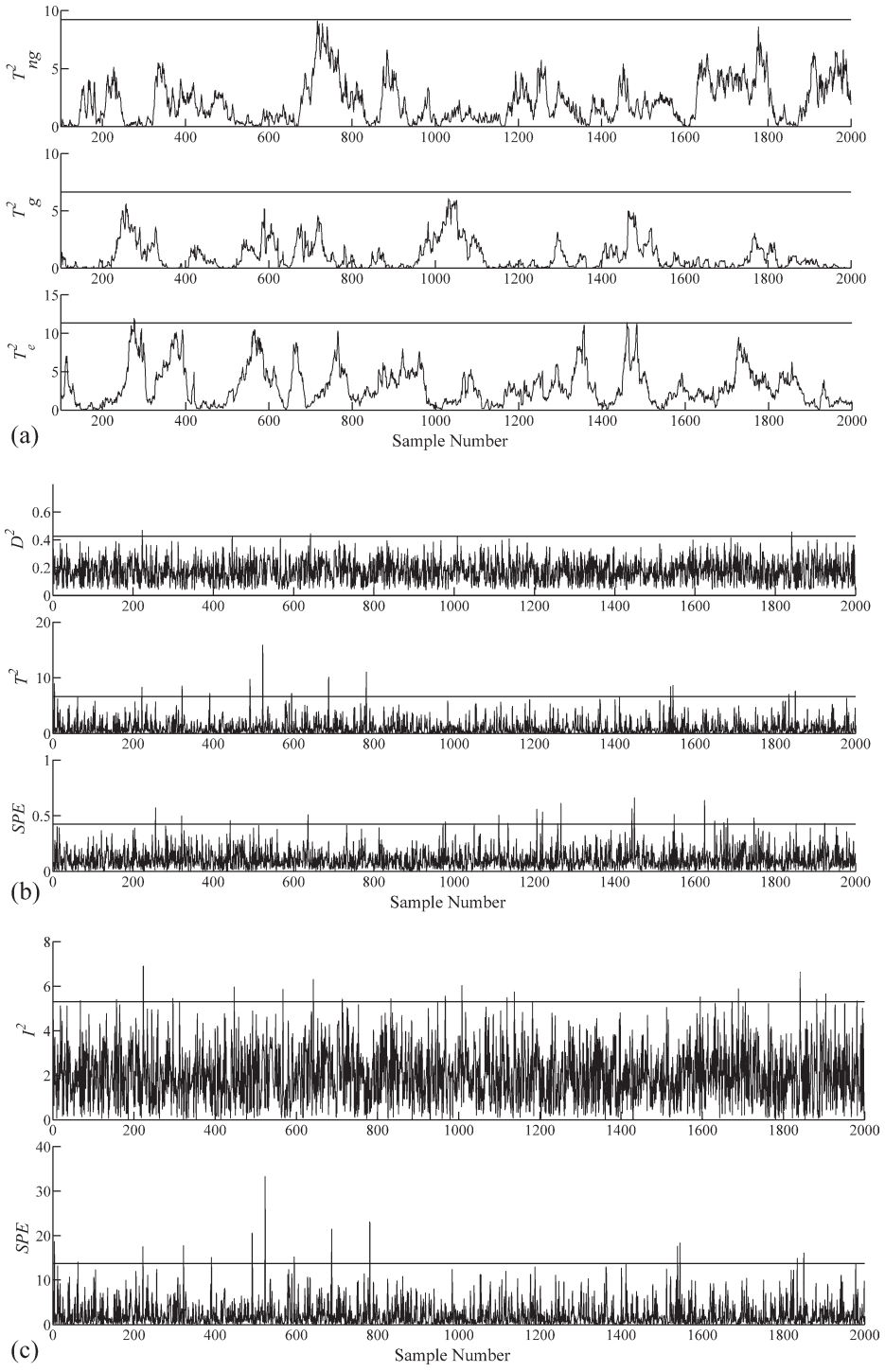

Figure 1. Monitoring results for normal process behavior; (a) Local ICA, (b) ICA-SVDD, and (c) ICA.

approach yields monitoring statistics that are more sensitive in detecting incipient fault conditions.

Figure 3 summarizes the results of applying each of the three monitoring models to the data set representing Case 3, describing a sensor bias in the first sensor (latter half of the second simulated data set $\mathbf{x}_{0_3}$). Different to Case 2, this time, each of the monitoring models detected this event. More precisely, the T^2_{ng}, the T^2_e statistics (LICA), the D_2 and *SPE* statistics (ICA-SVDD) and the I^2 and *SPE* statistics (ICA) produced statistically significant violations of their confidence limits.

Following from the discussion of Case 2, the LICA statistics were the most affected statistics, whilst the other statistics showed a less significant response to this event. That the T^2_g and T^2 statistics are insensitive to this event follows from the calculation of the Gaussian component:

$$\hat{s}_3 = (\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T(\mathbf{a}_3 s_2 + \mathbf{e} + \mathbf{f}) = (\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T(\mathbf{a}_3 s_3 + \mathbf{e}) + (\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T\mathbf{f} = \hat{s}_3^{(0)} + (\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T\mathbf{f}. \quad (23)$$

In fact, the matrix-vector product $(\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T = (0.0809\ 0.9226\ -0.5611\ 0.3272\ -0.5160\ -0.4134)$ shows a small value for

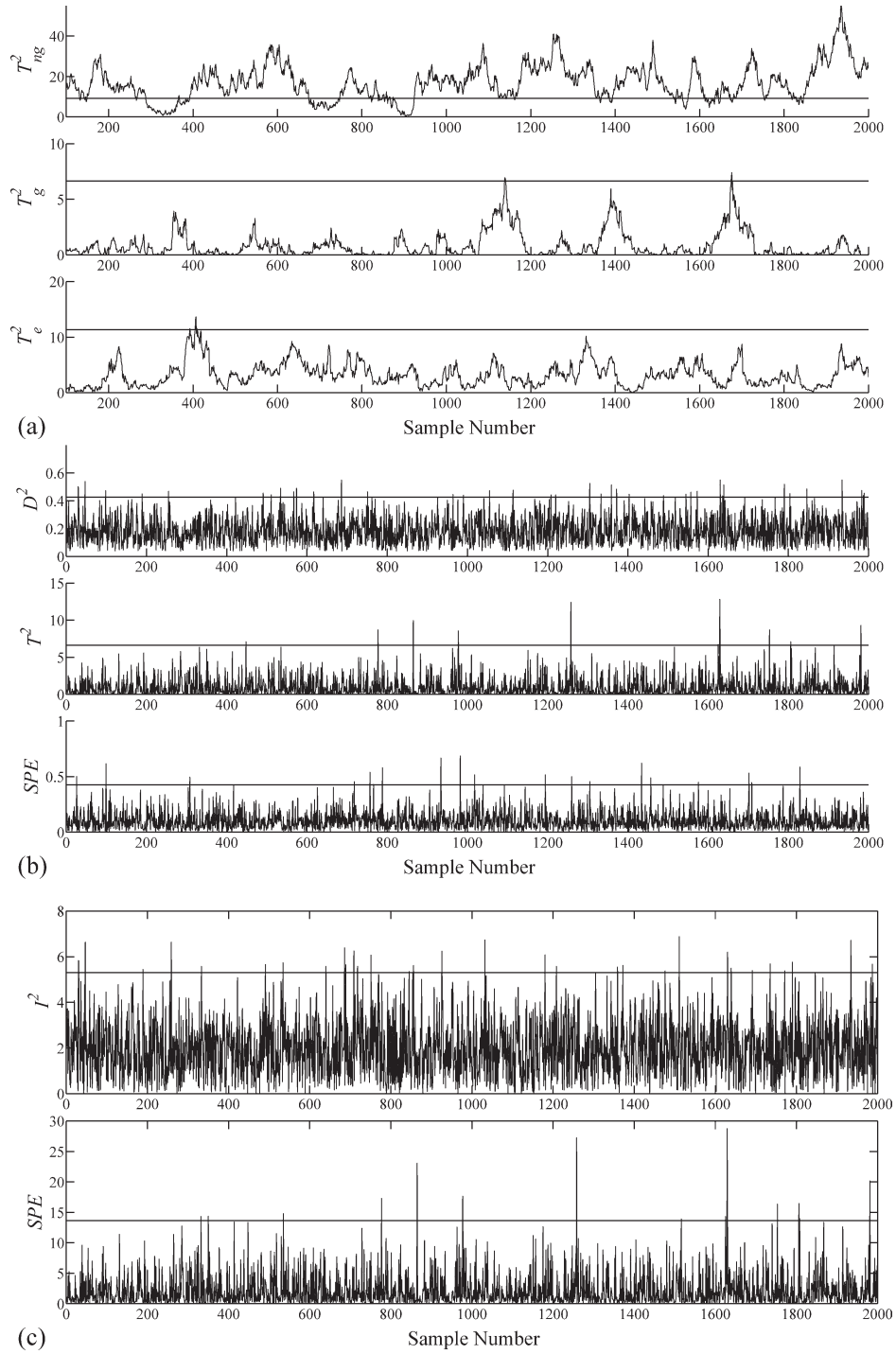

Figure 2. Monitoring results for rotation of non-Gaussian source variables by 15???; (a) Local ICA, (b) ICA-SVDD, and (c) ICA.

the first element. Consequently, the product $(\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T\mathbf{f} = 0.0971$ and is a marginal offset to the Gaussian distributed zero mean variable $s_3^{(0)}$. The univariate statistics T_g^2 and T^2 are therefore not affected by a sensor bias to the first recorded variable. In contrast, a bias in the second variable would have increased the offset term $(\hat{\mathbf{b}}^{\perp})^T\hat{\mathbf{P}}^T\mathbf{f}$ significantly to 1.1071 and resulted in statistical significant violations of both statistics.

That the *SPE* statistic for the ICA-SVDD approach was more sensitive than that of conventional ICA is a result of their construction. The ICA-SVDD residual statistic was based on the three discarded PCs of the MLPCA model. In contrast, the *SPE* statistic for ICA relied on the residuals of the original variable set $\mathbf{x}_0$ after deflating the two independent components. More precisely, the Gaussian source signal, s_3, contributed to the calculation of the ICA-based *SPE* statistic, whilst the estimated contribution of s_3 formed the T^2 statistic of the ICA-SVDD approach.

Figure 4 summarizes the results of applying the fault identification (left plots) and isolation (right plots) method,

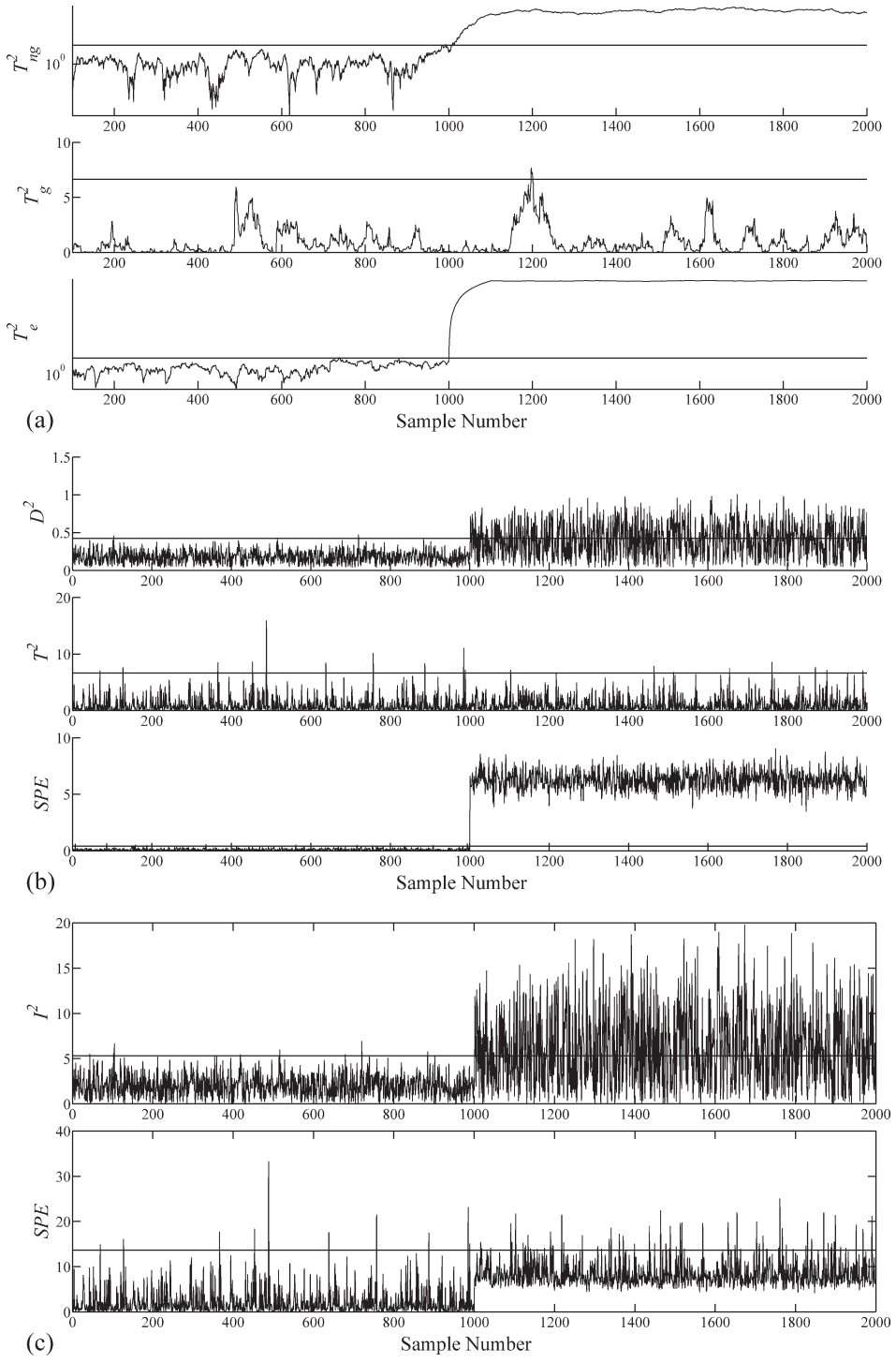

Figure 3. Monitoring results for measurement bias of 1.2 in sensor 1; (a) Local ICA, (b) ICA-SVDD, and (c) ICA.

developed in Implementation of Fault Detection Scheme Section. For the proposed LICA approach, the T^2_{ng}, the T^2_e statistics were sensitive. This required the application of the fault reconstruction scheme, detailed in Fault diagnosis using a local ICA monitoring approach Subsection, for the T^2_{ng} statistic and the reconstruction scheme discussed in Kruger and Dimitriadis.[11] The upper left plot in Figure 4 shows the fault isolation index defined in Eq. 18. The assumed fault conditions here represent the reconstruction of each of the six sensors, i.e., assumed fault condition 1 represents the reconstruction of the first sensor and so on. The height of each bar represents an average value for the first 100 samples after the event was detected. By inspection, only the first fault condition produced a value below the critical value of 1. It can therefore be concluded that the detected fault condition affected the first recorded variable.

The lower left plot in Figure 4 confirms this conclusion on the basis reconstructing the T^2_e statistic. For reconstructing

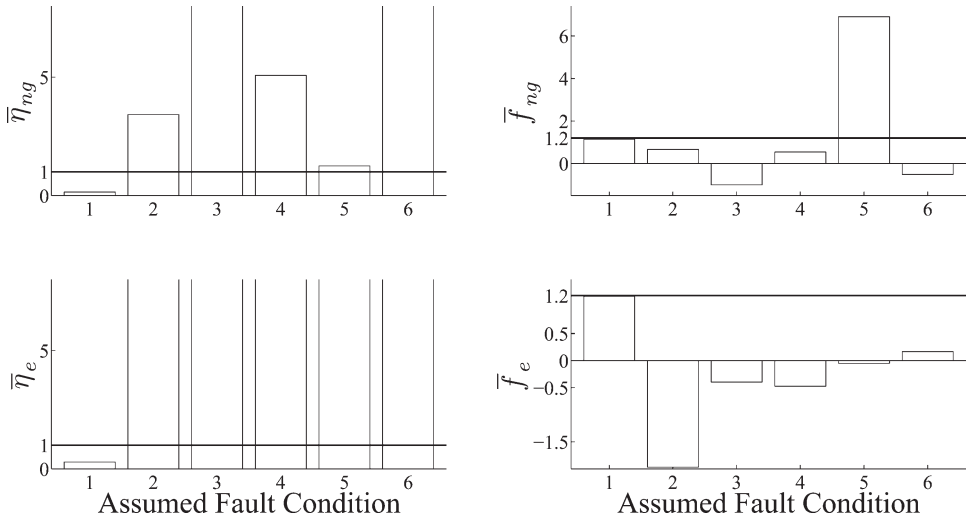

Figure 4. Fault identification chart for measurement bias in sensor 1.

each of the recorded variables in turn, only the reconstruction of the first variable produced average values below the critical value of 1. The upper right plot in Figure 4 highlights that applying the iterative fault isolation technique, described in Eq. 16, yielded a fault magnitude of around 1.2. The same fault magnitude was identified when applying the least squares approach to the T_e^2 statistic.[11] For Case 3, the sensor bias could therefore be detected and correctly diagnosed by the proposed LICA method. Case 2 therefore highlighted that the main advantages of the proposed LICA approach is the increased sensitivity in detecting incipient fault conditions. This advantage is further elaborated upon in the next section, which describes the analysis of data from an industrial glass melter process.

Application Study to a Melter Process

This section contrasts the proposed LICA approach with ICA-SVDD and conventional ICA applied to recorded data from an industrial melter process. As part of a disposal procedure, this process clad waste material in form of powder by molten glass. The vessel is continuously filled with powder and raw glass is discretely introduced in the form of glass frit. This binary composition is heated by four induction coils. The process of heating and filling continues until the desired height of the liquid column is reached, at which stage the molten mixture is poured out through an exit funnel. After the vessel has been emptied, the next cycle of filling and heating begins. This process has been described and analyzed by Liu et al.,[3] using ICA-SVDD approach. The presented work in this section, however, concentrates on recorded data from a different melter unit.

The recorded variable set includes eight temperature readings, denoted by $M_1 - M_8$, the power in the four induction coils, $A_1 - A_4$, the viscosity of the molten glass, A_5, and the voltage supplied to the induction coils, A_6. The heating and filling cycles produced non-Gaussian source signal contributions to the recorded variables, similarly to that observed by Liu et al..[3] Samples for each of the recorded variables were taken at a sampling interval of 5 min. Given that the sampling period is considerably longer than the response time of this process a linear steady state relationship, described in Eq. 1, can be assumed. From this process, a reference set of $K = 7{,}500$ samples (625 h) to identify a monitoring model for each method, and two additional data sets to validate these models and to diagnose a fault condition were recorded.

The first step for establishing monitoring models was to identify a MLPCA model[2–4] for the LICA and ICA-SVDD approaches, which yielded the retention of 9 PCs. A subsequent analysis confirmed that the discarded components did not encapsulate any non-Gaussian components. The application of the Jarque-Bera test suggested that seven of the nine PCs seven were non-Gaussian. A direct application of ICA to the recorded reference set revealed seven non-Gaussian components.

To construct Gaussian distributed improved residuals for the LICA technique; a window size of 100 samples was selected. A considerably shorter window size led to significant departures from a Gaussian distribution, whilst a much larger size is undesired as it may compromise sensitivity.[10,11] The nonlinear transformation required for the ICA-SVDD approach and the probability density function of the I^2 statistic was identified as discussed in the previous section.

Figure 5 shows the performance of each method to the data set representing a recorded fault condition. The upper left plot outlines that the T_{ng}^2 statistic violated its confidence limit after 600 samples into the data set. Sporadic violations were also noticeable from the T_e^2 statistic around the 700th and 800th sample. In contrast, the upper right and lower left plot highlight, respectively, that the ICA-SVDD and ICA approaches were not as sensitive. More precisely, whilst sporadic and isolated violations of the D^2 statistic between 650 and 950 samples into the data set, none of the other statistics were sensitive. This, in turn, implies that the ICA approach did not detect this event and the ICA-SVDD method was only sensitive at few short periods after 650 samples were recorded.

The diagnosis of this event relied on the T_{ng}^2 statistic, as the most dominantly affected one. For applying Eqs. 16 and 18, a total of 18 different fault directions, $\mathfrak{F}_1 - \mathfrak{F}_{18}$, were selected. The first 14 fault subspaces represented a sensor bias to one of 14 recorded process variables. The remaining

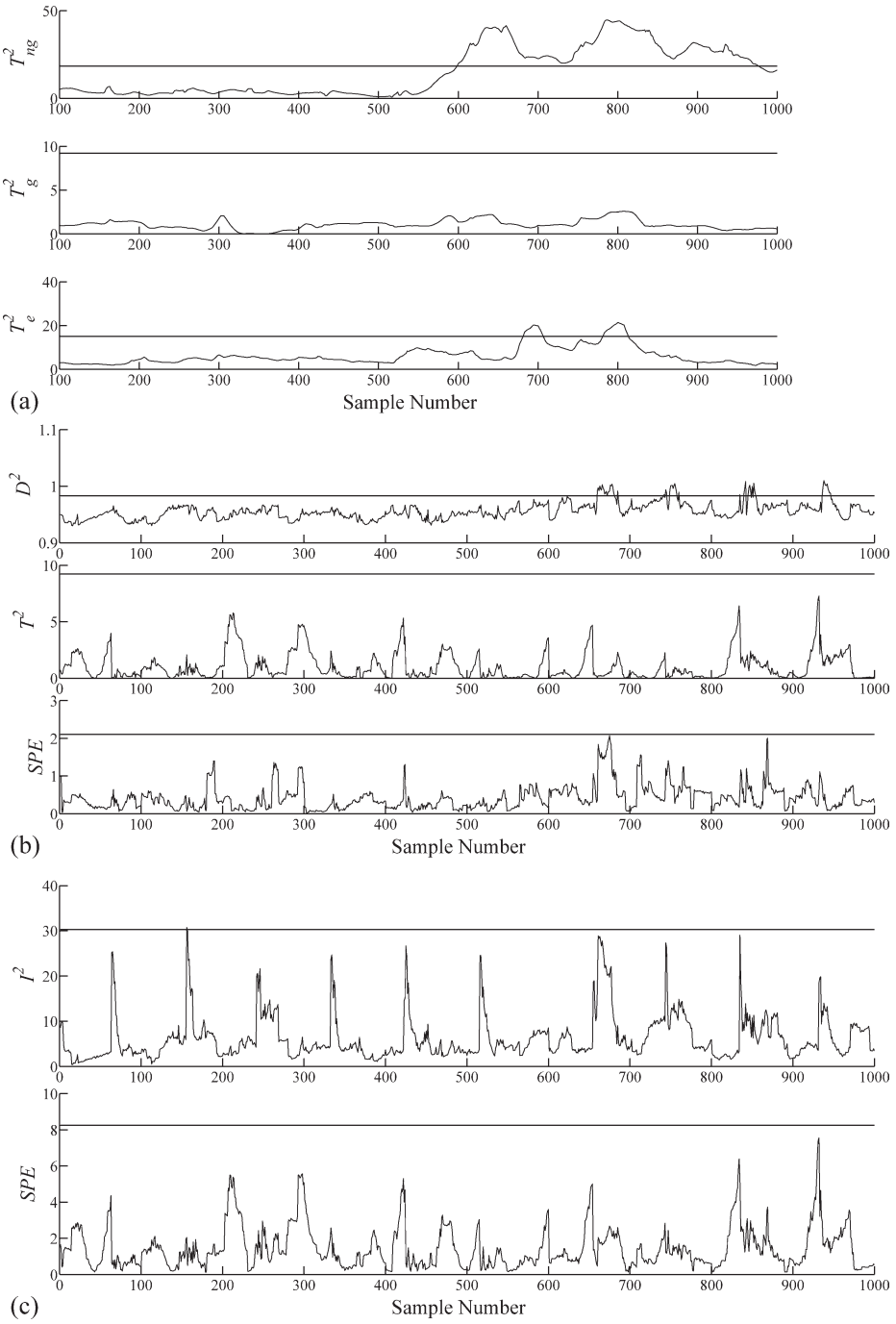

Figure 5. Monitoring results for abnormal process behavior; (a) Local ICA, (b) ICA-SVDD, and (c) ICA.

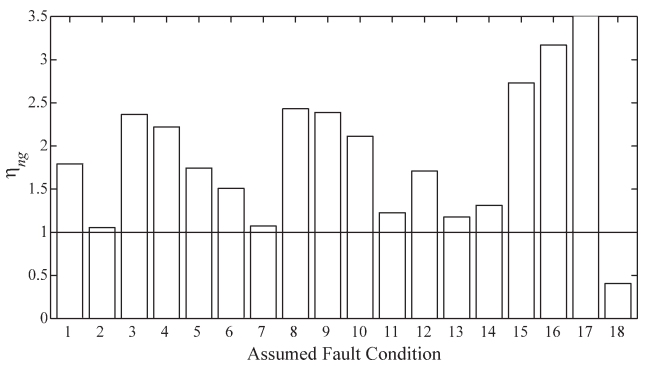

Figure 6. Fault isolation chart for detected fault condition.

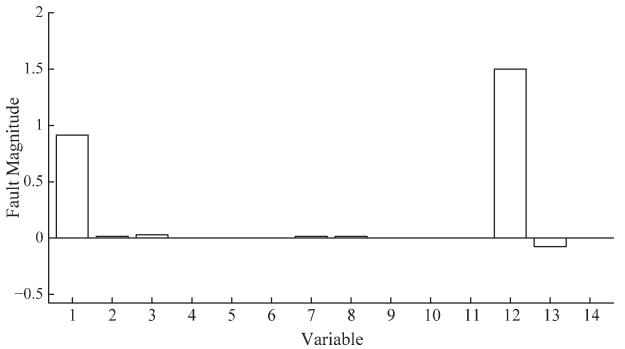

Figure 7. Fault identification chart for detected fault condition.

four fault subspaces described a fault in one of the induction coils, which were estimated from existing data describing these fault conditions. The estimated fault subspaces are as follows:

$$\begin{aligned}
\mathfrak{F}_{15} &= (0.72 \quad 0.74 \quad 0.41 \quad 0.38 \quad 0.19 \quad 0.22 \quad 0.08 \quad 0.11 \quad 1 \quad 0 \quad 0 \quad 0 \quad -0.04 \quad 0)^T \\
\mathfrak{F}_{16} &= (0.01 \quad 0.42 \quad 0.61 \quad 0.58 \quad 0.39 \quad 0.32 \quad 0.21 \quad 0.17 \quad 0 \quad 1 \quad 0 \quad 0 \quad -0.05 \quad 0)^T \\
\mathfrak{F}_{17} &= (0 \quad 0.02 \quad 0.48 \quad 0.59 \quad 0.01 \quad 0.64 \quad 0.30 \quad 0.03 \quad 0 \quad 0 \quad 1 \quad 0 \quad -0.03 \quad 0)^T \\
\mathfrak{F}_{18} &= (0.01 \quad 0 \quad 0.39 \quad 0 \quad 0 \quad 0 \quad 0.61 \quad 0.01 \quad 0 \quad 0 \quad 0 \quad 1 \quad -0.05 \quad 0)^T
\end{aligned} \tag{24}$$

Fault subspaces $\mathfrak{F}_{15}$, $\mathfrak{F}_{16}$, $\mathfrak{F}_{17}$, and $\mathfrak{F}_{18}$ describe the affect of an abnormal power increase/reduction in the top, upper middle, lower middle and bottom induction coil upon the 14 recorded variables, respectively. Given that sufficient records of data describing the impact of faults in individual induction coils was not available, the application of a singular value decomposition could not be relied upon in order determine the corresponding fault subspaces. In contrast, fault subspaces $\mathfrak{F}_{15}$, $\mathfrak{F}_{16}$, $\mathfrak{F}_{17}$, and $\mathfrak{F}_{18}$ were constructed through detailed discussions with experienced plant personnel. More precisely, fault subspaces $\mathfrak{F}_{15}$, $\mathfrak{F}_{16}$, $\mathfrak{F}_{17}$, and $\mathfrak{F}_{18}$ were constructed by incorporating estimates of the impact of a 1, 3, and 5% increase/reduction in the top, the upper middle, the lower middle and the bottom input coil, respectively, upon the temperatures and the velocity of the molten glass. The estimates of increases/decreases in °C for the temperature readings and *poise* for the molten glass in relation to power alterations in the induction coils were provided by plant personnel and scaled to become the entries in Eq. 24.

Figure 6 shows the corresponding fault isolation index for each of the 18 assumed fault conditions and highlights that fault subspace $\mathfrak{F}_{18}$ yielded a fault isolation index that is below the critical value of 1. This, in turn, implies that this recorded fault condition was an abnormal behavior of the induction coil mounted at the bottom of the melter vessel. In contrast, each of the other fault subspaces violated the critical value. Finally, Figure 7 shows the contribution of each variable to the power disturbance. This contribution chart depicts the elements of $\mathfrak{F}_{18}f_{18}$ vs. their respective process variable. The main contributing variables are the first thermocouple, M_1, and the power of the bottom induction coil, A_4. Minor contributions can be noticed from thermocouples M_2, M_3, M_7, and M_8, and from the viscosity reading A_5.

The estimated fault magnitude of 1.52 implies that the power supplied to the bottom induction coil was larger than expected and increased the temperature in its vicinity, i.e., M_1, and those in the lower part of the melter, which are M_2, M_3, M_7, and M_8. On the other hand, this increase in temperature resulted in a reduction of the viscosity. This information would have provided valuable information to an experienced operator offering a clear guidance to inspect the bottom induction coil.

Concluding Summary

This article has studied the incorporation of the statistical local approach into the monitoring of complex multivariate processes. The recorded variable sets are described by the noncausal data structure $\mathbf{x}_0 = \mathbf{As} + \mathbf{e}$, where the recorded variables $\mathbf{x}_0$ are driven by the source signals $\mathbf{s}$, which represent common cause variation of the process, and corrupted by an error term $\mathbf{e}$. The variance contribution of this error term, however, is assumed to be significantly smaller than that of the source signals $\mathbf{As}$. The application of MLPCA allows the extraction of PCs that encapsulate the source signal variation. The subsequent use of ICA then separates the non-Gaussian from the Gaussian source signals. The article has proposed that these signal components, the Gaussian and non-Gaussian source signals as well as the MLPCA residuals, form the basis for improved residuals that follow, asymptotically, a Gaussian distribution function.

Besides the benefit of constructing monitoring statistics that follow known standard parametric distribution functions, another advantage of the proposed LICA method is the increased sensitivity in detection incipient fault conditions. The article has demonstrated this on the basis of a simulation example where a minor alteration to the data structure $\mathbf{x}_0 = \mathbf{As} + \mathbf{e}$ could be detected by the proposed LICA method but remained undetected by other competitive methods that were developed in the research literature. Furthermore, the increased sensitivity of the LICA method has also been shown by analyzing recorded data of an industrial melter process. Whilst the LICA technique could detect an abnormality in one of the induction coils, this event went unnoticed by conventional ICA and could only be sporadically detected by ICA-SVDD.

To enhance the capability of the proposed LICA method, the paper has developed a fault identification and isolation technique to diagnose abnormal process behavior. This entails the identification of the correct fault subspace and estimation of the fault magnitude. The application of the fault diagnosis scheme to a sensor bias (simulation example) and an abnormal behavior in one of the induction coils (melter process) could correctly identify both root causes and their magnitude.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No. 60904039, 61134007) and the 111 project (Grant No. B07031).

Notation

$\mathbf{A}$ = Parameter matrix of noncausal data structure
$\mathbf{B}$ = Projection matrix for extracting non-Gaussian components
β, γ = Lagrangian scaling factors
$E\{\cdot\}$ = Expectation operator
$\mathbf{e}$ = Error vector of noncausal data structure
η_{ng}, η_g and η_e, = Fault identification index for non-Gaussian, Gaussian and error signal contribution
$\mathcal{F}_j$, $\mathfrak{F}_j$ and f_j = jth fault condition, fault subspace and fault magnitude
k = Sampling index
K,$\mathcal{K}$ = Number of samples
$\mathbf{L}$ = Lower triangular matrix of Cholesky decomposition
$\mathbf{\Lambda}$ = Matrix of eigenvalues of data covariance matrix
n = Number of source signals
N = Number of recorded process variables
$\mathbf{P}$ and $\mathcal{P}$ = Matrix of retained and discarded eigenvectors of data covariance matrix

$\mathbf{\Omega}(\mathbf{b})$ = Sphere defining the vicinity of a vector $\mathbf{b}$
$\mathbf{s}$, $\mathbf{s}_1$ and $\mathbf{s}_2$ = Vector of source signals, non-Gaussian (1) and Gaussian (2) source signals
$\mathbf{S}_{xx}$ = Covariance matrix of variable set $\mathbf{x}$
$\mathbf{t}$ = Vector of PCs
T^2_{ng}, T^2_g and T^2_e = Hotelling's T^2 statistic for non-Gaussian, Gaussian and error signal contributions
$\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}$ = Primary residual for ith source signal and concatenated residual functions
θ_i and $\boldsymbol{\theta}$ = Scaled primary residual for ith source signal and concatenated scaled residual functions
ϑ and $\boldsymbol{\psi}$ = Scaled primary residual functions for Gaussian and error signal contributions
$\mathbf{x}_0$ = Measured data vector (mean centered)
$\mathbf{y}$ = Scaled vector of PCs
$\boldsymbol{\zeta}_{ng}$, $\boldsymbol{\zeta}_g$ and $\boldsymbol{\zeta}_e$ = Improved residual vectors for non-Gaussian, Gaussian and error signal contribution
$\hat{}$ = Estimated variable
$\tilde{}$ = Scaled variable

Literature Cited

1. Tipping ME, Bishop CM. Probabilistic principal component analysis. *J R Stat Soc Series B*. 1999;61:611–622.
2. Narasimhan S, Shah SL. Model identification and error covariance matrix estimation from noisy data using PCA. *Control Eng Pract*. 2008;16:146–155.
3. Liu X, Xie L, Kruger U, Littler T, Wang S. Statistical-based monitoring of multivariate non-Gaussian systems. *AIChE J*. 2008;54:2379–2391.
4. Feital T, Kruger U, Xie L, Schubert U, Lima EL, Pinto JC. A unified statistical framework for monitoring multivariate systems with un-known source and error signals. *Chemometr Intell Lab Syst*. 2010;104:223–232.
5. Kano M, Hasebe S, Hashimoto I, Ohno H. Evolution of multivariate statistical process control: application of independent component analysis and external analysis. *Comp Chem Eng*. 2004;28:1157–1166.
6. Lee JM, Yoo CK, Lee IB. Statistical process monitoring with independent component analysis. *J Process Contr*. 2004;14:467–485.
7. Lee JM, Qin SJ, Lee IB. Fault detection and diagnosis based on modified independent component analysis. *AIChE J*. 2006;52:3501–3514.
8. Chen Q, Kruger U, Leung AYT. Regularised kernel density estimation for clustered process data. *Control Eng Pract*. 2004;12:267–274.
9. Ge Z, Xie L, Kruger U, Lamont L, Song Z, Wang S. Sensor fault identification and isolation for mutltivariate non-Gaussian processes. *J Process Control*. 2009;19:1707–1715.
10. Basseville M. On-board component fault detection and isolation using the statistical local approach. *Automatica*. 1998;34:1391–1415.
11. Kruger U, Dimitriadis D. Diagnosis of process faults in chemical systems using the local partial least squares approach. *AIChE J*. 2008;54:2581–2596.
12. Dunia R, Qin SJ. Subspace approach for to multidimensional fault identification and reconstruction. *AIChE J*. 1998;44:2797–2812.
13. Lieftucht D, Kruger U, G.W. Irwin. Improved reliability in diagnosing faults using multivariate statistics. *Comp Chem Eng*. 2006;30:901–912.
14. Hawkins DM. Cumulative sum control charting: an underutilized SPC tool. *Qual Eng*. 1993;5:463–477.
15. Hawkins DM, Olwell DH. *Cumulative Sum Charts and Charting for Quality Improvements*. New York: Springer Verlag, 1989.
16. Hyvarinen A. Gaussian moments for noisy independent component analysis. *IEEE Signal Process Lett*. 1999;6:145–147.
17. Yang YM, Guo CH. Gaussian moments for noisy unifying model. *Neurocomputing*. 2008;71:3656–3659.
18. Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Network*. 2000;13:411–430.
19. Tracey ND, Young JC, Mason RL. Multivariate control charts for individual observations. *J Qual Technol*. 1992;24:88–95.
20. Jensen AC, Berge A, Solberg AS. Regression approaches to small sample inverse covariance matrix estimation for hyperspectral image classification. *IEEE Trans Geosci Remote Sens*. 2008;46:2814–2822.
21. D.C. Hoyle. Automatic PCA dimension selection for high dimensional data and small sample sizes. *J Mac Learn Res*. 2008;9:2733–2759.
22. Kruger U, Kumar S, Littler T. Improved principal component monitoring using the local approach. *Automatica*. 2007;43:1532–1542.
23. Yue HH, Qin SJ. Reconstruction-based fault identification using a combined index. *Ind Eng Chem Res*. 2001;40:4403–4414.
24. Tax DMJ, Duin RPW. Support vector domain description. *Pattern Recognit Lett*. 1999;20:1191–1199.
25. Tax DMJ, Duin RPW. Support vector data description. *Mac Learn*. 2004;54:45–66.
26. Jackson JE, Mudholkar G. Control procedures for residuals associated with principal component analysis. *Technometrics*. 1979;21:341–349.
27. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics*. 1995;37:41–59.
28. Martin EM, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J Process Control*. 1996;6:349–358.

Appendix A: Independent Component Analysis

With respect to Eq. 1, the ICA model assumes the existence of independent source variables $\boldsymbol{s}(k) \in \mathbb{R}^r$ that generate the variable set stored in $\mathbf{x}_0(k)$. The aim of ICA is to find a separating matrix $\mathbf{W} \in \mathbb{R}^{r \times N}$, such that

$$\hat{\mathbf{s}}(k) = \mathbf{W}\mathbf{x}_0(k) = \mathbf{W}(\mathbf{A}\mathbf{s}(k) + \mathbf{e}(k)) \approx \mathbf{s}(k) \tag{A1}$$

Including a whitening procedure, $\mathbf{y}(k) = \mathbf{Q}\mathbf{x}_0(k)$, $\mathbf{Q} \in \mathbb{R}^{N \times N}$ being the whitening matrix, the separation matrix can be defined as follows:

$$\hat{\mathbf{s}}(k) = \mathbf{W}\mathbf{x}_0(k) = \mathbf{W}\mathbf{Q}^{\dagger}\mathbf{y}(k) = \mathbf{B}^T\mathbf{y}(k) \tag{A2}$$

where $[\cdot]^{\dagger}$ is a generalized inverse and $\mathbf{B} \in \mathbb{R}^{N \times r}$ is determined to maximize the non-Gaussianity of $\hat{\boldsymbol{s}} = \mathbf{B}^T\mathbf{y}$ under the constraint that the columns of $\mathbf{B}$ are mutually orthonormal and determined by maximizing $J(t)$, $t = \mathbf{b}^T\mathbf{y}$ measuring non-Gaussianity. For $J(t)$, the negentropy is usually used, which relies on the information-theoretic quantity of differential entropy, defined as $H(t) = -\int f(t) \log (f(t))dt$, where t and $f(\cdot)$ is a random variable and its density function, respectively. A Gaussian variable, v, has the largest entropy among all random variables of equal variance, and allows the definition of $J(t) = H(v) - H(t)$, which can be approximated by[27]:

$$J(\mathrm{t}) \approx [E\{G(t)\} - E\{G(v)\}]^2 \tag{A3}$$

Here, $G(\cdot)$ is the nonquadratic function.

Appendix B: Proof of Theorem 1

Suppose an abnormal behavior affects the relationship between the source variables and the recorded variables in the data structure of Eq. 1, i.e., a change in the parameter matrix $\mathbf{A}^{\#} = \mathbf{A} + \Delta\mathbf{A}$. Accordingly, such a change has a direct effect on the orientation of the model plane, which is defined by the column space of $\mathbf{A}$. This, in turn, alters the underlying data model to describe the recorded process variables to become $\mathbf{x}_0^{\#} = \mathbf{x}_0 + \Delta\mathbf{x}_0$ and hence, $\mathbf{y}^{\#} = \Lambda^{-1/2}\mathbf{P}^T\mathbf{x}_0^{\#} = \mathbf{y} + \Delta\mathbf{y}$. The primary residual vector for $\mathbf{y}^{\#} = \mathbf{y} + \Delta\mathbf{y}$ is given by:

$$\begin{aligned}\boldsymbol{\phi}_i^{\#}(\mathbf{b}_i, \mathbf{y}^{\#}) &= \mathbf{y}^{\#} g(\mathbf{b}_i^T\mathbf{y}^{\#}) - \beta_i\mathbf{b}_i \\ &= (\mathbf{y} + \Delta\mathbf{y}) g(\mathbf{b}_i^T(\mathbf{y} + \Delta\mathbf{y})) - \beta_i\mathbf{b}_i\end{aligned} \tag{B1}$$

It should be noted that the expectation for Eq. B1 is equal to zero if and only if $\Delta\mathbf{y} = \mathbf{0}$, which follows from Eq. 4. More precisely, the expectation of Eq. B1 is given by:

$$\begin{aligned} E\{\boldsymbol{\phi}_i^{\#}(\mathbf{b}_i, \mathbf{y}^{\#})\} &= E\{\mathbf{y}g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y})) - \beta_i\mathbf{b}_i + \Delta\mathbf{y}g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} \\ E\{\boldsymbol{\phi}_i^{\#}(\mathbf{b}_i, \mathbf{y}^{\#})\} &\approx E\left\{\mathbf{y}g(\mathbf{b}_i^T\mathbf{y}) - \beta_i\mathbf{b}_i + \mathbf{y}\left(\frac{\partial g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))}{\partial\Delta\mathbf{y}}\Big|_{\Delta\mathbf{y}=0}\Delta\mathbf{y}\right) + \Delta\mathbf{y}g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\right\} \\ E\{\boldsymbol{\phi}_i^{\#}(\mathbf{b}_i, \mathbf{y}^{\#})\} &\approx E\left\{\mathbf{y}\left(\frac{\partial g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))}{\partial\Delta\mathbf{y}}\Big|_{\Delta\mathbf{y}=0}\Delta\mathbf{y}\right) + \Delta\mathbf{y}g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\right\} \\ E\{\boldsymbol{\phi}_i^{\#}(\mathbf{b}_i, \mathbf{y})\} &\approx E\{\Delta\mathbf{y}g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} = \Delta\mathbf{y}E\{g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} \end{aligned} \tag{B2}$$

By evaluating

- $E\{g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\}$
$= E\left\{(E\{G(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} - E\{G(v)\})\frac{dG(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))}{d\mathbf{b}_i^T\mathbf{y}}\right\};$
- $\frac{dG(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))}{d(\mathbf{b}_i^T\mathbf{y})} \neq 0 \quad \forall\mathbf{y} \neq \Delta\mathbf{y}$, since $G(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y})) = (-1/a)\exp(-a^{(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))^2}/2)$, $a \in \mathbb{R}$;[9] and
- $E\{G(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} - E\{G(\upsilon)\} \neq 0$, as $\|\Delta\mathbf{y}\|$ is assumed to be small and $(E\{G(\mathbf{b}_i^T\mathbf{y})\} - E\{G(\upsilon)\})^2$ is a minimum.

it follows that $\Delta\mathbf{y}E\{g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} \neq \mathbf{0}$. Therefore, the shift in mean by the primary residual vector can be detected. Premultiplying Eq. B2 by $\mathbf{b}_i^T$ yields:

$$E\{\theta_i^{\#}(\mathbf{b}_i, \mathbf{y}^{\#})\} = \mathbf{b}_i^T\Delta\mathbf{y}E\{g(\mathbf{b}_i^T(\mathbf{y}+\Delta\mathbf{y}))\} \neq 0 \tag{B3}$$

It can therefore be concluded that the expectations in Eqs. B2 and B3 are not equal to zero and hence, any incipient change introduced by $\Delta\mathbf{y} \neq \mathbf{0}$ is detectable by both monitoring functions.

Appendix C: Derivative of Eq. 16

Generally, the optimal reconstruction can be obtained by minimizing $\|\hat{\mathbf{x}}_0(k) - \mathbf{x}_0(k)\|$, where $\|\cdot\|$ is the norm of a vector. However, a direct analytical solution cannot be obtained, as $\mathbf{x}_0(k)$ is unknown and the approximation of the negentropy value for each independent component relies on nonquadratic functions, as discussed in Derivation of primary residuals Subsection. Based on the developed monitoring function, an iterative reconstruction scheme can be obtained that minimizes the following function:

$$\hat{f}_j = \underset{f_j}{\operatorname{argmin}}\left\|\mathbf{S}_{\theta\theta}^{-\frac{1}{2}}\boldsymbol{\zeta}(\boldsymbol{\theta}, k)\right\|^2 = \underset{f_j}{\operatorname{argmin}}\left\|\mathbf{S}_{\theta\theta}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\boldsymbol{\theta}\left(\hat{\mathbf{B}}, \hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\mathbf{P}}^T\hat{\mathbf{x}}_0(k)\right)\right\|^2$$

$$\hat{f}_j = \underset{f_j}{\operatorname{argmin}}\left\|\mathbf{S}_{\theta\theta}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\boldsymbol{\theta}\left(\hat{\mathbf{B}}, \hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\mathbf{P}}^T\left(\mathbf{x}_0^*(k) - \mathfrak{F}f_j\right)\right)\right\|^2$$

$$\hat{f}_j = \underset{f_j}{\operatorname{argmin}}\left\|\mathbf{s}_{\theta\theta}^{-\frac{1}{2}}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}\hat{s}_{11}(k)g(\hat{s}_{11}) - \beta_1 \\ \hat{s}_{12}(k)g(\hat{s}_{12}) - \beta_2 \\ \vdots \\ \hat{s}_{1m}(k)g(\hat{s}_{1m}) - \beta_m\end{bmatrix}\right\|^2$$

$$\hat{f}_j = \underset{f_j}{\arg\min}\left\|\mathbf{S}_{\theta\theta}^{-1/2}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_jf_j)g(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_jf_j) - \beta_1 \\ (\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_jf_j)g(\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_jf_j) - \beta_2 \\ \vdots \\ (\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_jf_j)g(\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_jf_j) - \beta_m\end{bmatrix}\right\|^2 \tag{C1}$$

where $\mathbf{w}_i = \hat{\mathbf{P}}\hat{\boldsymbol{\Lambda}}^{-1/2}\hat{\mathbf{b}}_i \in \mathbb{R}^N$, $\hat{s}_{1i}^*(k) = \mathbf{w}_i^T\mathbf{x}_0^*(k) \in \mathbb{R}$, $i = 1,2,\cdots,m$, and $\hat{s}_{1i}(k) = \hat{s}_{1i}^*(k) - \mathbf{w}_i^T\mathfrak{F}_jf_j \in \mathbb{R}$. It is important to note that the reconstruction entails the estimation of the optimal fault magnitude, $\hat{f}_j$, relative to the assumed fault subspace $\mathfrak{F}_j$. The next subsection details how the optimal fault condition can be obtained from the J potential ones. Defining the following cost function:

$$\psi = \left\|\mathbf{S}_{\theta\theta}^{-1/2}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_jf_j)g(\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_jf_j) - \beta_1 \\ (\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_jf_j)g(\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_jf_j) - \beta_2 \\ \vdots \\ (\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_jf_j)g(\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_jf_j) - \beta_m\end{bmatrix}\right\|^2, \tag{C2}$$

that is the squared length of the scaled improved residual vector, and taking the first derivative of ψ w.r.t. f_j yields:

$$\nabla_{f_j}\psi = \frac{\partial\psi}{\partial f_j} = 2\boldsymbol{\zeta}^T(\boldsymbol{\theta}, K)\mathbf{S}_{\theta\theta}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix}-\frac{\partial(\hat{s}_{11}(k)g(\hat{s}_{11}(k)) - \beta_1)}{\partial\hat{s}_{11}(k)}\mathbf{w}_1^T\mathfrak{F}_j \\ -\frac{\partial(\hat{s}_{12}(k)g(\hat{s}_{12}(k)) - \beta_2)}{\partial\hat{s}_{12}(k)}\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ -\frac{\partial(\hat{s}_{1m}(k)g(\hat{s}_{1m}(k)) - \beta_m)}{\partial\hat{s}_{1m}(k)}\mathbf{w}_m^T\mathfrak{F}_j\end{bmatrix} \tag{C3}$$

and allows $\hat{f_j}$ to be determined iteratively. According to Eq. 4, this article discusses two different forms for $G(\cdot)$: i) $G_1(\hat{s}) = -e^{\frac{-(\hat{s})^2}{2}}$ and (ii) $G_2(\hat{s}) = -\log(\cosh \hat{s})$. These have the following first and second derivatives: (i) $G_1^{'}(\hat{s}) = \hat{s}e^{\frac{-(\hat{s})^2}{2}}$, $G_1^{''}(\hat{s}) = e^{\frac{-(\hat{s})^2}{2}} - \hat{s}^2 e^{\frac{-(\hat{s})^2}{2}}$ and (ii) $G_2^{'}(\hat{s}) = \tanh(\hat{s})$, $G_2^{''}(\hat{s}) = 1 - \tanh^2(\hat{s})$, respectively. Equation C3 can therefore be rewritten to become:

$$\nabla_{f_j}\psi = 2\boldsymbol{\zeta}^{\boldsymbol{T}}(\boldsymbol{\theta},K)\mathbf{S}_{\theta\theta}^{-1}\tfrac{1}{\sqrt{K}}\textstyle\sum_{k=1}^{K}\begin{bmatrix} -\frac{\partial(\hat{s}_{11}(k)g(\hat{s}_{11}(k))-\beta_1)}{\partial\hat{s}_{11}(k)}\mathbf{w}_1^T\mathfrak{F}_j \\ -\frac{\partial(\hat{s}_{12}(k)g(\hat{s}_{12}(k))-\beta_2)}{\partial\hat{s}_{12}(k)}\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ -\frac{\partial(\hat{s}_{1m}(k)g(\hat{s}_{1m}(k))-\beta_m)}{\partial\hat{s}_{1m}(k)}\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix}$$

$$\nabla_{f_j}\psi = 2\boldsymbol{\zeta}^{\boldsymbol{T}}(\boldsymbol{\theta},K)\mathbf{S}_{\theta\theta}^{-1}\tfrac{1}{\sqrt{K}}\textstyle\sum_{k=1}^{K}\begin{bmatrix} -(g(\hat{s}_{11}(k)) + \hat{s}_{11}(k)g'(\hat{s}_{11}(k)))\mathbf{w}_1^T\mathfrak{F}_j \\ -(g(\hat{s}_{12}(k)) + \hat{s}_{12}(k)g'(\hat{s}_{12}(k)))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ -(g(\hat{s}_{1m}(k)) + \hat{s}_{1m}(k)g'(\hat{s}_{1m}(k)))\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix} \tag{C4}$$

$$\nabla_{f_j}\psi = 2\boldsymbol{\zeta}^{\boldsymbol{T}}(\boldsymbol{\theta},K)\mathbf{S}_{\theta\theta}^{-1}\tfrac{1}{\sqrt{K}}\textstyle\sum_{k=1}^{K}\begin{bmatrix} -(g(\hat{s}_{11}(k)) + (\hat{s}_{11}^*(k) - \mathbf{w}_1^T\mathfrak{F}_j f_j)g'(\hat{s}_{11}(k)))\mathbf{w}_1^T\mathfrak{F}_j \\ -(g(\hat{s}_{12}(k)) + (\hat{s}_{12}^*(k) - \mathbf{w}_2^T\mathfrak{F}_j f_j)g'(\hat{s}_{12}(k)))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ -(g(\hat{s}_{1m}(k)) + (\hat{s}_{1m}^*(k) - \mathbf{w}_m^T\mathfrak{F}_j f_j)g'(\hat{s}_{1m}(k)))\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix}$$

Now, setting $\nabla_{f_i}\psi = 0$ yields:

$$\begin{aligned}\boldsymbol{\zeta}^T(\boldsymbol{\theta},k)\mathbf{S}_{\theta\theta}^{-1}\frac{1}{\sqrt{K}}\Sigma_{k=1}^{K}&\begin{bmatrix} (g(\hat{s}_{11}(k)) + \hat{s}_{11}^*(k)g'(\hat{s}_{11}(k)))\mathbf{w}_1^T\mathfrak{F}_j \\ (g(\hat{s}_{12}(k)) + \hat{s}_{12}^*(k)g'(\hat{s}_{12}(k)))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ (g(\hat{s}_{1m}(k)) + \hat{s}_{1m}^*(k)g'(\hat{s}_{1m}(k)))\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix} \\ &= \left\{\boldsymbol{\zeta}^T(\boldsymbol{\theta},k)\mathbf{S}_{\theta\theta}^{-1}\frac{1}{\sqrt{K}}\Sigma_{k=1}^{K}\begin{bmatrix} \mathfrak{F}_j^T\mathbf{w}_1 g'(\hat{s}_{11}(k))\mathbf{w}_1^T\mathfrak{F}_j \\ \mathfrak{F}_j^T\mathbf{w}_2 g'(\hat{s}_{12}(k))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ \mathfrak{F}_j^T\mathbf{w}_m g'(\hat{s}_{1m}(k))\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix}\right\} f_j\end{aligned} \tag{C5}$$

The estimated fault magnitude $\hat{f_j}$ finally becomes:

$$\hat{f_j} = \frac{\boldsymbol{\zeta}^T(\boldsymbol{\theta},k)\mathbf{S}_{\theta\theta}^{-1}\tfrac{1}{\sqrt{K}}\Sigma_{k=1}^{K}\begin{bmatrix} (g(\hat{s}_{11}(k)) + \hat{s}_{11}^*(k)g'(\hat{s}_{11}(k)))\mathbf{w}_1^T\mathfrak{F}_j \\ (g(\hat{s}_{12}(k) + \hat{s}_{12}^*(k))g'(\hat{s}_{12}(k)))\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ (g(\hat{s}_{1m}(k) + \hat{s}_{1m}^*(k))g'(\hat{s}_{1m}(k)))\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix}}{\boldsymbol{\zeta}^T(\boldsymbol{\theta},k)\mathbf{S}_{\theta\theta}^{-1}\tfrac{1}{\sqrt{K}}\Sigma_{k=1}^{K}\begin{bmatrix} \mathfrak{F}_j^T\mathbf{w}_1 g'(\hat{s}_{11}(k)\mathbf{w}_1^T\mathfrak{F}_j \\ \mathfrak{F}_j^T\mathbf{w}_2 g'(\hat{s}_{12}(k)\mathbf{w}_2^T\mathfrak{F}_j \\ \vdots \\ \mathfrak{F}_j^T\mathbf{w}_m g'(\hat{s}_{1m}(k)\mathbf{w}_m^T\mathfrak{F}_j \end{bmatrix}} \tag{C6}$$

Equation C6, however, does not present an analytical solution since the right hand side contains the estimated fault magnitude. The form of Eq. C6 gives rise to an iterative formulation for which the $(t+1)$ th iteration step is given in Eq. C7. Upon convergence, Eq. 15 allows the determination of the reconstructed measurement vector $\hat{\mathbf{x}}_0(k)$. It is important to note that this iterative algorithm is similar to that developed in Ref. 25. However, different from the iteration procedure of Eq. C7, the algorithm in Ref. 25 (i) does not rely on a moving window formulation, (ii) does not estimate the model plane, and (iii) applies ICA directly to the recorded variables.

$$\hat{f}_j(t+1) = \frac{\boldsymbol{\zeta}^{\boldsymbol{T}}(\boldsymbol{\theta},K)\mathbf{S}_{\theta\theta}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix} (g(\hat{s}_{11}^{*}(k)-\mathbf{w}_1^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))+\hat{s}_{11}^{*}(k)g'(\hat{s}_{11}^{*}(k)-\mathbf{w}_1^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t)))\mathbf{w}_1^T\boldsymbol{\mathfrak{F}}_j \\ (g(\hat{s}_{12}^{*}(k)-\mathbf{w}_2^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))+\hat{s}_{12}^{*}(k)g'(\hat{s}_{12}^{*}(k)-\mathbf{w}_2^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t)))\mathbf{w}_2^T\boldsymbol{\mathfrak{F}}_j \\ \vdots \\ (g(\hat{s}_{1m}^{*}(k)-\mathbf{w}_m^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))+\hat{s}_{1m}^{*}(k)g'(\hat{s}_{1m}^{*}(k)-\mathbf{w}_m^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t)))\mathbf{w}_m^T\boldsymbol{\mathfrak{F}}_j \end{bmatrix}}{\left\{\boldsymbol{\zeta}^{\boldsymbol{T}}(\boldsymbol{\theta},K)\mathbf{S}_{\theta_0\theta_0}^{-1}\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\begin{bmatrix} \boldsymbol{\mathfrak{F}}_j^T\mathbf{w}_1 g'(\hat{s}_{11}^{*}(k)-\mathbf{w}_1^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))\mathbf{w}_1^T\boldsymbol{\mathfrak{F}}_j \\ \boldsymbol{\mathfrak{F}}_j^T\mathbf{w}_2 g'(\hat{s}_{12}^{*}(k)-\mathbf{w}_2^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))\mathbf{w}_2^T\boldsymbol{\mathfrak{F}}_j \\ \vdots \\ \boldsymbol{\mathfrak{F}}_j^T\mathbf{w}_m g'(\hat{s}_{1m}^{*}(k)-\mathbf{w}_m^T\boldsymbol{\mathfrak{F}}_j\hat{f}_j(t))\mathbf{w}_m^T\boldsymbol{\mathfrak{F}}_j \end{bmatrix}\right\}} \tag{C7}$$

It is also important to note that reconstruction schemes for the Gaussian and residual components can be obtained from the work by Dunia and Qin.[12]

Manuscript received Oct. 19, 2010, revision received Jun. 13, 2011, and final revision received Aug. 17, 2011.