AIChE

Statistical-Based Monitoring of Multivariate Non-Gaussian Systems

Xueqin Liu and Lei Xie
State Key Lab of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, P.R. China, and School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Ashby Building, Stranmillis Road, Antrim BT9 5AH, U.K.

Uwe Kruger
Dept. of Electrical Engineering, The Petroleum Institute, P.O. Box 2533, Abu Dhabi, U.A.E.

Tim Littler
School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Ashby Building, Stranmillis Road, Antrim BT9 5AH, U.K.

Shuqing Wang
State Key Lab of Industrial Control Technology, Institute of Advanced Process Control, Zhejiang University, Hangzhou 310027, P.R. China

DOI 10.1002/aic.11526
Published online July 14, 2008 in Wiley InterScience (www.interscience.wiley.com).

The monitoring of multivariate systems that exhibit non-Gaussian behavior is addressed. Existing work advocates the use of independent component analysis (ICA) to extract the underlying non-Gaussian data structure. Since some of the source signals may be Gaussian, the use of principal component analysis (PCA) is proposed to capture the Gaussian and non-Gaussian source signals. A subsequent application of ICA then allows the extraction of non-Gaussian components from the retained principal components (PCs). A further contribution is the utilization of a support vector data description to determine a confidence limit for the non-Gaussian components. Finally, a statistical test is developed for determining how many non-Gaussian components are encapsulated within the retained PCs, and associated monitoring statistics are defined. The utility of the proposed scheme is demonstrated by a simulation example, and the analysis of recorded data from an industrial melter. © 2008 American Institute of Chemical Engineers *AIChE J,* 54: 2379–2391, 2008
Keywords: multivariate systems, non-Gaussian variables, independent component analysis, support vector data description, process monitoring, fault detection

Introduction

Because of their simplicity, multivariate projection-based techniques, such as PCA, have gained attention for monitoring complex processes, such as those found in the chemical industry.[1] PCA exploits the correlation within the typically large number of recorded variables by defining a reduced set of score variables for constructing a Hotelling's T^2 statistic.[2] The mismatch between the recorded variables and their reconstruction using these score variables leads to the definition of the Q statistic.[3]

Changes in throughput, emptying and filling cycles, the presence of unmeasured disturbances and plant recycle loops,

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

however, may produce process variables that do not follow a Gaussian distribution, as assumed for multivariate statistical-based monitoring. This, in turn, implies that a statistical inference can no longer rely on simple parametric distribution functions for the T^2 statistic. More precisely, the assumption that the T^2 statistic follows an F-distribution may no longer be true. Hence, applying an incorrect distribution can have an undesired effect on the number of Type I and II errors.

ICA has recently been applied for extracting linear combinations of non-Gaussian variables. For modeling chemical systems, Li and Wang[4] utilized ICA to identify dynamic trends in process data. However, their work did not exploit these trends for process monitoring. This was addressed by Kano et al.[5,6] which proposed charting individual independent components (ICs) with a time base. These references showed through application studies that ICs can be more sensitive in detecting faults and reduce the number of Type II errors. However, a theoretical rationale for these results was not provided. Furthermore, the confidence limits for the ICs were obtained in an *ad hoc* fashion, and the number of charts increases with the number of ICs, which hampers the practical usefulness of this approach for large variable sets.

Similar to PCA, Lee et al.[7] utilized the set of ICs for reconstructing the recorded variables. The ICs are divided with respect to $s_i = \mathbf{w}_i^T \mathbf{z}$ where $s_i \in \mathbb{R}$ is the i^{th} IC, $\mathbf{w}_i$ and $\mathbf{z} \in \mathbb{R}^N$ are a parameter and a data vector, respectively, such that $\boldsymbol{w}_i$ vectors producing large $\|\mathbf{w}_i\|_2$ values present dominant ICs. The aim is to extract dominant ICs from the computed ICA decomposition to produce a total of three univariate statistics for process monitoring. These are the I^2, the I_e^2, and the Q statistic, that are based on the dominant ICs, the remaining ICs and the residuals of the ICA decomposition. Confidence limits for these statistics and bivariate scatter diagrams were then determined using a kernel density estimation (KDE).[8,9]

The use of a KDE, however, (1) is problematic if the data distribution is sparse or clustered,[10] and (2) is computationally expensive for determining confidence limits/regions. Furthermore, the division into *dominant* and *remaining* ICs is not based on a statistical interpretation of the recorded data, since inspecting $\|\mathbf{w}_i\|_2$ is not related to their importance in terms of reconstructing the original data nor their degree of non-Gaussianity.

It is also important to note that the geometric simplicity of PCA is not maintained when ICA is used instead. PCA determines base vectors, and uses these to project the data vectors onto the model plane and the complementary residual subspace. ICA, however does not offer such a simple geometric interpretation. Finally, Lee et al.[11] showed that selecting the initial components as PCs produces a unique and repeatable solution of the ICA cost function, which is, however, not guaranteed to be globally optimal.

Despite the reported progress in monitoring non-Gaussian systems over the past few years, the aforementioned summary highlights that a number of issues still remain: (1) how to geometrically interpret ICs, (2) how to estimate confidence limits for ICs effectively, (3) how to evaluate the importance of ICs, and (4) source signals may contain both, Gaussian and non-Gaussian components.

This article addresses these issues by assuming that the underling data structure can be described by $\mathbf{z} = \boldsymbol{\Gamma}\mathbf{u} + \mathbf{f}$, where $\mathbf{u} \in \mathbb{R}^n$ is the set of source signals that contain both Gaussian and non-Gaussian components, $\mathbf{f} \in \mathbb{R}^N$ are residuals that follow a zero mean Gaussian distribution with covariance Σ_f, $\mathbf{f} \sim \mathcal{N}\ \{\mathbf{0}, \Sigma_f\}$ and Γ is a parameter matrix. The source signals are first separated from $\mathbf{z}$ set through the application of PCA. Non-Gaussian components within the retained PCs are then extracted by applying ICA.

Incorporating ICA into PCA-based monitoring maintains the simple geometric interpretation of PCA and relates to the first issue. The separation of the retained score variables into Gaussian and non-Gaussian components leads to the construction of two univariate statistics and addresses the fourth issue. The discarded components are then utilized to establish a PCA Q statistic.

This article addresses the second issue by determining confidence limits for ICs using a support vector data description (SVDD). This technique relies on the support vector machine concept applied to the transformed ICs into a feature space. This space is constructed such that each of the feature variables falls within a small hypersphere. Transformed ICs that produce points outside the sphere indicate anomalous process behavior and vice versa. Since a SVDD relies on a quadratic programming cost function, it is computationally efficient compared to the KDE approach, advocated by Lee et al.[7]

The final contribution is the development of a testing procedure for non-Gaussian variables, and relates to the third issue. The proposed test is designed to rank ICs according to their degree of non-Gaussianity. This, in turn, is used to categorize the importance of ICs, since this test is directly linked to the ICA cost function. ICs that produce a significant test statistic possess a non-Gaussian distribution and vice versa. This article demonstrates that the widely used JB test[12] is a special case of the proposed test. To determine a globally optimal ICA solution, the article uses recent work on particle swarm optimization (PSO).[13]

This article is organized as follows. Preliminaries, including PCA-based monitoring, ICA and SVDD are given next, followed by a proof that PCA applied to the data structure $\mathbf{z} = \boldsymbol{\Gamma}\mathbf{u} + \mathbf{f}$ projects the source signals $\mathbf{u}$, onto the model plane. Then the non-Gaussianity test is introduced, which is followed by detailing the proposed monitoring scheme, and then presents a simulation example and an application study to recorded data from an industrial melter system, respectively. A concluding summary is given at the end of this article.

Preliminaries

This section gives a summary of PCA-based monitoring and the principles of ICA and SVDD.

Principal component analysis

PCA relies on a coordinate transformation of $\mathbf{z}$ to produce a reduced set of PCs[14]

$$\mathbf{z} = \mathbf{P}\mathbf{t} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{t} \in \mathbb{R}^n$, $n \leq N$, is a score vector storing n coordinates of the new coordinate system $\mathbf{P} \in \mathbb{R}^{N \times n}$ stores the base vectors of the system, and $\boldsymbol{\varepsilon} \in \mathbb{R}^N$ is a residual vector. The columns of $\mathbf{P}$ span a model plane capturing significant variation

encapsulated in $\mathbf{z}$. A complementary residual subspace, spanned by the columns of $\mathbf{R} \in \mathbb{R}^{N\times(N-n)}$, $\mathbf{R}^T\mathbf{P} = \mathbf{0}$ describes $\boldsymbol{\varepsilon}$.

The geometric simplicity of the PCA decomposition allows the construction of two univariate statistics, a Hotelling's T^2 statistic that represents significant variation of the recorded data, and is associated with the PCA model plane, and a Q statistic describing the mismatch between the original variables and their projections onto the model plane

$$T^2 = \mathbf{t}^T\boldsymbol{\Lambda}\mathbf{t} = \mathbf{z}^T\mathbf{P}\boldsymbol{\Lambda}\mathbf{P}^T\mathbf{z}$$
$$Q = \boldsymbol{\varepsilon}^T\boldsymbol{\varepsilon} = \mathbf{z}^T[\mathbf{I} - \mathbf{P}\mathbf{P}^T]\mathbf{z} = \mathbf{z}^T\mathbf{R}\mathbf{R}^T\mathbf{z}, \quad (2)$$

for which confidence limits can be obtained as discussed by Jackson.[14] The confidence limits, however, are obtained under the assumption that the recorded variables are Gaussian.

Independent component analysis

ICA is designed to extract ICs from a variable set $\mathbf{z}$ that is a linear combination of non-Gaussian variables $\mathbf{s} \in \mathbb{R}^m$, $m \leq N$:

$$\mathbf{z} = \mathbf{A}\mathbf{s} + \mathbf{e} \quad (3)$$

where $\mathbf{A} \in \mathbb{R}^{N\times m}$ contains *mixture coefficients,* and $\mathbf{e}$ is a zero mean Gaussian residual vector with covariance Σ_e, $\mathbf{e} \sim \mathcal{N}\{\mathbf{0}, \Sigma_e\}$. The aim of ICA is to find a separating matrix $\mathbf{W} \in \mathbb{R}^{m\times N}$, such that

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{z} = \mathbf{W}\mathbf{A}\mathbf{s} \approx \mathbf{s}, \quad (4)$$

Including a whitening procedure[15] $\mathbf{x} = \mathbf{Q}\mathbf{z}$, $\mathbf{Q} \in \mathbb{R}^{N\times N}$, being the whitening matrix, Eq. 4 becomes

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{z} = \mathbf{W}\mathbf{Q}^{\dagger}\mathbf{x} = \mathbf{B}^T\mathbf{x}, \quad (5)$$

where $[\cdot]^{\dagger}$ is a generalized inverse, and $\mathbf{B} \in \mathbb{R}^{N\times m}$ is determined to maximize the non-Gaussianity of $\hat{\mathbf{s}} = \mathbf{B}^T\mathbf{x}$ under the constraint that the columns of $\mathbf{B}$ are mutually orthogonal and determined by maximizing $J(y)$, $y = \mathbf{b}^T\mathbf{x}$, which measures non-Gaussianity. For $J(y)$ the negentropy is usually employed, which relies on the information-theoretic quantity of differential entropy, defined as $H(y) = -\int f(y)\log(f(y))\,dy$, where y 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 $J(y) = H(v) - H(y)$, which can be approximated by

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2. \quad (6)$$

Here, $G(\cdot)$ is the nonquadratic function.[15] In the presence of outliers, $G(\cdot)$ is preferred to be $-1/a_2(\exp(-a_2y^2/2))$.[16]

Existing work on computing an ICA solution relates to gradient descent techniques, e.g., Lee et al,[11] which may not produce a global optimum. This article employs the recently proposed and numerically efficient PSO-ICA technique,[13] where a PSO search algorithm is utilized to compute the non-Gaussian components, and yields a globally optimal solution.

Support vector data description

Inspired by the support vector classifier (SVC) concept,[17] SVDD was derived by Tax and Duin.[18,19] Designed as a classification tool, it produces a description of a reference set for detecting whether new samples resemble the properties of that set. Applications have shown a high-generalization performance if large reference sets with very few abnormal samples are available.[20]

The core idea is to envelope the data within a feature space by a minimal spherical volume. A simple sphere (dashed line), as shown in Figure 1a, could envelope the original data.

This sphere, however, is not flexible enough to give an adequate description of the data, compared to the region determined by a more complex nonlinear support vector approach.

To maintain the numerical efficiency in determining a hypersphere, SVDD performs the transformation $\boldsymbol{\zeta} = \Phi(\mathbf{z})$ into the *feature space*. Using the *kernel trick* of SVC, kernel functions can be established to simplify the determination of $\Phi(\cdot)$. This article uses Gaussian kernel functions $\mathcal{K}(\mathbf{z}_i,\mathbf{z}_j) = \exp(-\|\mathbf{z}_i - \mathbf{z}_j\|^2/\sigma^2)$, where σ is a scaling factor, as they are preferable to other kernel types.[18] Figure 1b illustrates this transformation and shows that the feature space (1) produces an alignment of the transformed data onto a surface, and (2) that a hypersphere can be constructed to appropriately envelope the transformed data.

The hypersphere in the feature space is constructed by maximizing the cost function $\mathcal{J}$[18]

$$\mathcal{J} = \arg\max_{\boldsymbol{\alpha}} 1 - \sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_i\alpha_j\mathcal{K}(\mathbf{z}_i,\mathbf{z}_j) \quad \sum_{i=1}^{K}\alpha_i = 1; \alpha_i \in [0, C] \quad (7)$$

where $\boldsymbol{\alpha}^T = (\alpha_1, \ldots, \alpha_K)$. The center $\mathbf{a}$, and the radius R, of the hypersphere are given by

$$\mathbf{a} = \sum_{i=1}^{K}\alpha_i\Phi(\mathbf{z}_i)$$

$$R = \sqrt{1 - 2\sum_{i=1}^{K}\alpha_i\mathcal{K}(\mathbf{z}_S,\mathbf{z}_i) + \sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_i\alpha_j\mathcal{K}(\mathbf{z}_i,\mathbf{z}_j)} \quad (8)$$

Most of the α_i coefficients are zero. The sample point $\mathbf{z}_S$, corresponding to a positive $\alpha_i < C$, is referred to as a support vector. The parameter C represents a trade-off between the volume of the hypersphere and the significance level, while σ determines the resolution at which the data is considered in the original variable space.[20]

After constructing the hypersphere in the kernel space, the hypothesis that a new object $\tilde{\zeta} = \Phi(\tilde{\mathbf{z}})$ is a normal sample is accepted if the squared distance $d(\Phi(\tilde{\mathbf{z}})) = \|\Phi(\tilde{\mathbf{z}}) - \mathbf{a}\|^2 \leq R^2$:

$$d(\Phi(\tilde{\mathbf{z}})) = 1 + \sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_i\alpha_j\mathcal{K}(\mathbf{z}_i,\mathbf{z}_j) - 2\sum_{i=1}^{K}\alpha_i\mathcal{K}(\tilde{\mathbf{z}},\mathbf{z}_i) \leq R^2. \quad (9)$$

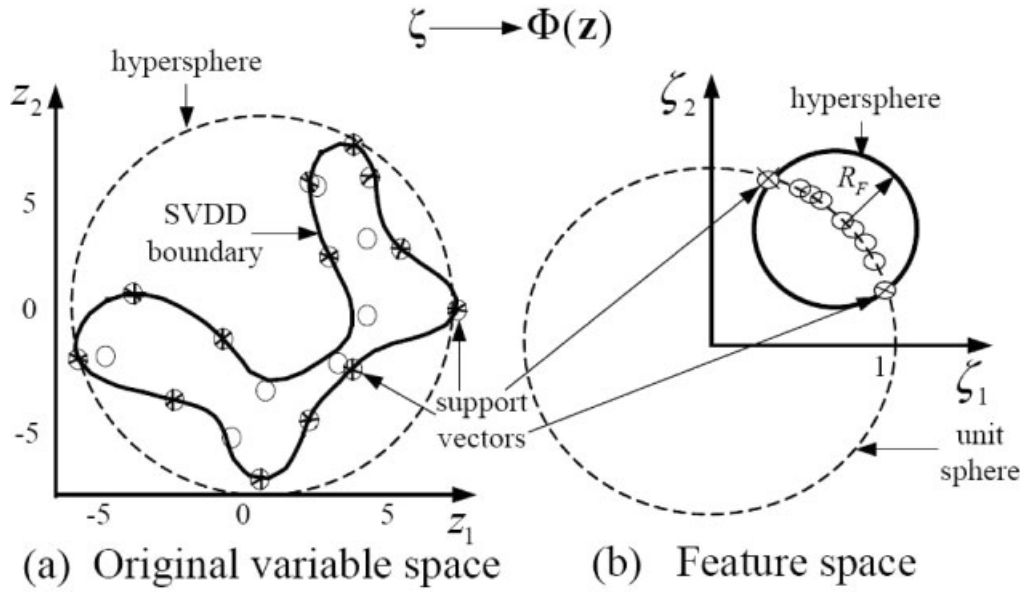

Figure 1. (a) Circular hypersphere and estimated boundary using support vectors from original data; (b) circular SVDD hypersphere of transformed data.

Source Signals Extraction Using Principal Component Analysis

This section provides a proof that the application of PCA projects the source signals onto the model plane only, if the process variables have the data structure $\mathbf{z} = \mathbf{\Gamma u} + \mathbf{f}$. Under the assumption that the process exhibits a total of $n < N$ source variables $\mathbf{u} \in \mathbb{R}^n$, among which $m \leq n$ are non-Gaussian, the following relationship can be established

$$\begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{\Gamma}_1 \\ \mathbf{\Gamma}_2 \end{pmatrix} \mathbf{u} + \mathbf{f} = \begin{bmatrix} \mathbf{I} \\ \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1} \end{bmatrix} \mathbf{\Gamma}_1\mathbf{u} + \mathbf{f}, \tag{10}$$

where $\mathbf{\Gamma}_1 \in \mathbb{R}^{n\times n}$ and of full rank, and $\mathbf{\Gamma}_2 \in \mathbb{R}^{(N-n)\times n}$. Under the assumption that the residuals $\mathbf{f}$ follow a Gaussian distribution $\mathbf{f} \sim \mathcal{N}\{\mathbf{0}, \Sigma_f\}$, Eq. 10 represents an error-in-variable, or total least-squares problem. This follows from $\mathbf{z}_2^{(0)} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}\mathbf{z}_1^{(0)}$, with $\mathbf{z}_1^{(0)}$ and $\mathbf{z}_2^{(0)}$ being the uncorrupted recordings of $\mathbf{z}_1$ and $\mathbf{z}_2$, respectively. Wentzell et al.[21] showed that maximum likelihood PCA can be employed to estimate $\Theta = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$ for uncorrelated residuals in Table 1 (page 348), and for correlated errors in Table 2 (page 350). The tight equivalence between total least-squares and maximum likelihood PCA was highlighted by Schuermans.[22] To address cases where Σ_f is not known *a priori*, Narasimhan and Shah[23] introduced a maximum likelihood PCA approach for simultaneous model identification and residual covariance matrix estimation.

Rewriting Eq. 10 in the form of

$$\mathbf{z} = \begin{bmatrix} \mathbf{I} \\ \mathbf{\Theta} \end{bmatrix} \mathbf{P}_1\mathbf{\Gamma}_1^*\mathbf{u} + \mathbf{f}, \tag{11}$$

where $\mathbf{P}_1 \in \mathbb{R}^{n\times n}$, $\mathbf{\Gamma}_1^* \in \mathbb{R}^{n\times n}$, $\mathbf{\Gamma}_1^* = \mathbf{P}_1^{-1}\mathbf{\Gamma}_1$, and comparing it with that of a maximum likelihood PCA model of $\mathbf{z}$

$$\mathbf{z} = \mathbf{Pt} + \varepsilon = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \end{bmatrix} \mathbf{t} + \varepsilon = \begin{bmatrix} \mathbf{I} \\ \mathbf{P}_2\mathbf{P}_1^{-1} \end{bmatrix} \mathbf{P}_1\mathbf{t} + \varepsilon \tag{12}$$

the matrix expression $\hat{\mathbf{\Theta}} = \mathbf{P}_2\mathbf{P}_1^{-1}$ converges to $\mathbf{\Theta} = \mathbf{\Gamma}_2\mathbf{\Gamma}_1^{-1}$ as the number of samples, K, converges to infinity under the assumption that the residual covariance matrix is known. This allows writing the asymptotic solution of Eq. 12 as follows

$$\mathbf{z} = \mathbf{\Gamma}\mathbf{\Gamma}_1^{*-1}\mathbf{t} + \boldsymbol{\varepsilon}, \tag{13}$$

Using the aforementioned derivative, the projection of $\mathbf{z}$ onto the model plane and residual subspace is given by

Table 1. Calculation of Confidence Limits for Univariate Monitoring Statistics

Variable	Statistic	Confidence limit
$\mathbf{s}$	$D^2 = \|\Phi(\mathbf{s}) - \mathbf{a}\|^2$	R^2
τ	$T^2 = \tau^T\tau$	$\frac{(n-m)(K^2-1)}{K(K-n+m)}F_{n-m,K-n+m}$
ε	$Q = \varepsilon^T\varepsilon$	$g\chi^2(h)$

Table 2. Number of Type I errors for T^2_{PCA}, Q, D^2 and T^2 Statistics

	95% Control limit	99% Control limit
T^2_{PCA}	1.850%	0.200%
Q	5.000%	1.050%
D^2	5.050%	1.150%
T^2	4.900%	1.100%

$$\mathbf{t} = \mathbf{P}(\mathbf{\Gamma u} + \mathbf{f}) = \mathbf{P\Gamma u} + \mathbf{Pf},$$
$$\boldsymbol{\varepsilon} = \mathbf{R}(\mathbf{\Gamma u} + \mathbf{f}) = \underbrace{\mathbf{R}^T\mathbf{P}}_{0}\mathbf{\Gamma}_1^*\mathbf{u} + \mathbf{R}^T\mathbf{f}. \quad (14)$$

which implies that the $n \geq m$ source signals are projected onto the model plane along with some variation of the residuals **f,** and that the PCA residuals $\boldsymbol{\varepsilon}$ are linear combinations of **f**. More precisely, the source signals, including Gaussian and non-Gaussian components, are projected onto the model plane as $K \to \infty$. This allows the extraction of the non-Gaussian and Gaussian components from the score variables of a maximum likelihood PCA model, which is discussed in the ICA-SVDD monitoring scheme section.

Negentropy-Based Non-Gaussianity Test

The Preliminaries section has shown how a multivariate non-Gaussian data set can be modeled (ICA), and how these sequences can be transformed to determine a simple spherical confidence region that can be used for statistical inference. The question of how many ICs need to be extracted to capture the non-Gaussian signal components, however, has not been adequately addressed in the research literature. This, however, is of fundamental importance to achieve a good and robust performance of the monitoring scheme. For example, the retention of too many ICs may cause an amplification of noise.[7] Cheung and Xu[24] proposed the selection of retained ICs by applying a two-step procedure:

Step 1. List all the ICs in an appropriate order; and

Step 2. Select the first few ICs as dominant ones.

Hyvarinen[25] suggested two different methods to rank the ICs:

Method 1: Order the ICs with respect to column norm of **A**. This is based on the assumption that the ICs corresponding to the column with the largest norm have the greatest contribution to the variance of the variable set **z**. However, the ICs are determined to maximize $J(y)$, and it is, therefore, not guaranteed that ICs corresponding to the largest column norms of **A** have the most significant variance contribution to the recorded variable set. Moreover, this method does not provide any means for determining how many ICs need to be extracted.

Method 2: This approach is based on the ordering of ICs according to their non-Gaussianity. This can be accomplished by applying standard non-Gaussianity tests, for example the well known Jarque-Bera (JB) test.[12]

Given the aforementioned analysis, Method 2 is more appropriate for determining how many ICs must be included to capture the non-Gaussian signal components of recorded variables. It should be noted, however, that the determination of ICs relies on non-Gaussianity, computed by the negentropy expression of Eq. 6. Moreover, the negentropy function is a well established concept in statistical theory, and is an optimal estimator of non-Gaussianity.[15] Therefore, through the computation of the ICs this measure is available, which, in turn, can be taken advantage of in deciding whether to include a particular IC.

This section introduces a new non-Gaussianity test that is based on the available negentropy values. We also show later that the JB test is a special case of the proposed non-Gaussianity test. This test is based on Theorem 1, which is proven in Appendix A.

Theorem 1 *The negentropy function of a variable y that follows a standard Gaussian distribution* $y \in \mathcal{N}(0,1)$, *has the following distribution if* $K \to \infty$

$$K \cdot J(y_1, y_2, \ldots, y_k) \sim \text{var}\{G(v)\}\chi^2(1), \quad (15)$$

In the aforementioned theorem, v is a Gaussian variable, statistically independent of y, and of zero mean and unit variance, and var $\{\cdot\}$ denotes variance. Equation 15 can alternatively be rewritten, such that the term $\frac{K}{\text{var}\{G(v)\}} \cdot J(y_1, y_2, \ldots, y_K)$, asymptotically follows a χ^2 distribution with one degree of freedom. Moreover, $J(y_1, y_2, \ldots, y_K) = (\frac{1}{K}\sum_{i=1}^{K} G(y_i) - E(G(v)))^2$.

With regards to Theorem 1, the following holds true.

Remark 1 *In order to determine whether y follows a Gaussian distribution, a confidence limit for the statistic* $\frac{K}{\text{var}\{G(v)\}} J(y_1, y_2, \ldots, y_K)$. *can be computed on the basis of a* χ^2 *distribution with one degree of freedom, and a significance level of* 0.05 *or* 0.01, *for example*.

Theorem 2 *The conventional JB-test can be formulated to be a special case of the negentropy-based non-Gaussianity test.*

Theorem 2 is proven in Appendix B.

ICA-SVDD Monitoring Scheme

Previous sections discussed details of ICA, SVDD, and how to determine ICs that significantly depart from a Gaussian distribution. This section introduces a monitoring scheme that extracts non-Gaussian signal components from the score variables of the PCA decomposition. The proposed ICA-SVDD monitoring approach is motivated in the Monitoring scheme section, followed by a summary of steps for implementing this scheme.

Monitoring scheme

The success of PCA for monitoring complex multivariate process systems lies in its simplicity.[1] The fact that conventional PCA is only applicable for processes that present Gaussian distributed variables, however, presents a significant limitation. In contrast to existing work on monitoring non-Gaussian processes, the technique proposed here maintains the simplicity of the PCA decomposition. As discussed in the Preliminaries section, PCA decomposes the original variable space into an n-dimensional model plane, $n \geq m$, capturing dominant process variation in a typically much reduced dimensional space, i.e., $n < N$, and a complementary residual subspace, describing the mismatch between the original process variables and the reduced dimensional data description.

ICA, briefly revised in the Preliminaries section, is then applied to extract non-Gaussian signal components from the retained score variables. Unlike PCs, which are statistically independent up to second-order statistical information if the original variables serially uncorrelated and normally distributed, ICs do not possess such appealing features. In contrast to the work by Lee et al.[7] which advocates the use of the KDE technique, the method proposed in this article relates to the use of a SVDD, discussed in the Preliminaries section. In

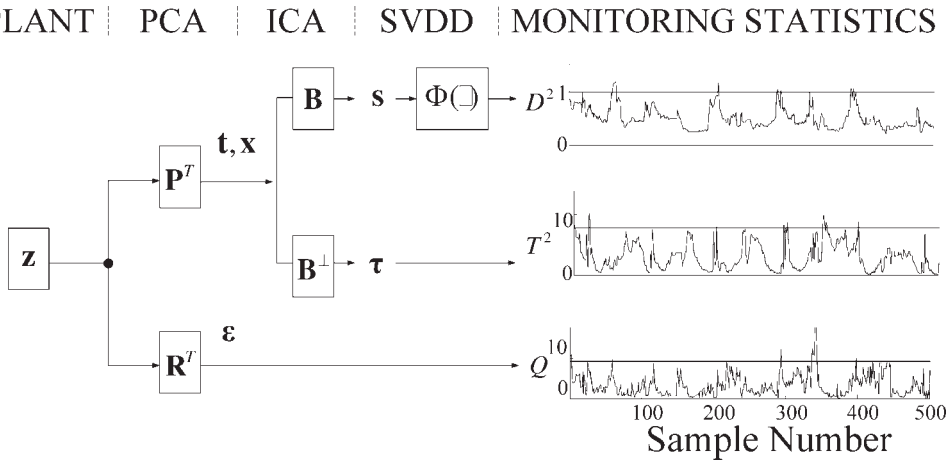

Figure 2. Implementation of the online monitoring approach.

contrast to the leave-one-out cross-validation approach, required to extract a KDE of the ICs, the SVDD method reduces to a quadratic programming problem that produces a unique analytical solution.

However, not all of the ICs, extracted from the retained PCs, present significant non-Gaussian distributions. This can be assessed by applying the non-Gaussianity test developed in the previous section. In a similar fashion to the work by Lee et al.[7] this gives rise to the construction of up to three univariate monitoring statistics, as defined in the following. Two of these relates to the retained PCs, while the remaining third is associated with the discarded ones. More precisely, the retained scores may be separated between those that show a significant non-Gaussian distribution function, and those that can be approximated by a Gaussian distribution, which Eq.16 summarizes

$$\mathbf{z}\longrightarrow\begin{cases}\mathbf{t}=\mathbf{P}^T\mathbf{z}\rightarrow\mathbf{x}=\boldsymbol{\Lambda}^{-1/2}\mathbf{t}\rightarrow\begin{cases}\mathbf{s}=\mathbf{B}^T\mathbf{x}\\ \boldsymbol{\tau}=(\mathbf{B}^{\perp})^T\mathbf{x}\end{cases}\\ \boldsymbol{\varepsilon}=\mathbf{R}^T\mathbf{z}\end{cases}\qquad(16)$$

where $\boldsymbol{\tau}$ and $\mathbf{s}$ are non-Gaussian variable sets, extracted from the retained PCs, respectively, $\mathbf{B}\in\mathbb{R}^{n\times m}$ is the parameter matrix of ICA models (Eq. 15), and $\mathbf{B}^{\perp}\in\mathbb{R}^{n\times(n-m)}$ is the orthogonal complements of $\mathbf{B}$.

The decomposition in Eq. 16 gives rise to the construction of the following three univariate statistics:

1. A statistic representing the variation of the non-Gaussian components in the principal component model plane $\mathbf{s}$, defined here as D^2;

2. a nonnegative squared statistic relating to the Gaussian components in the model plane τ, denoted as T^2, and

3. a statistic that characterizes the variation of the PCA residuals, ε in the residual subspace, referred to as Q.

In line with conventional PCA, the T^2 statistic follows an F-distribution,[14] and the statistic can be approximated by a central χ^2-distribution.[26] As discussed in the Preliminaries section, the confidence limit for the statistic is the squared radius of the hypersphere in the respective feature space. Table 1 summarizes the calculation of the confidence limits, where $g=\rho^2/2\mu$, $h=2\mu^2/\rho^2$ and, μ and ρ are the sample mean and variance of the Q statistic.

For online process monitoring, a statistically significant number of violations of these limits, or at least one of them, is indicative of abnormal process behavior.

Figure 2 illustrates the steps for implementing the ICA-SVDD monitoring scheme. From the left to the right, the data vector is passed onto the first block. After mean-centering and scaling, $\mathbf{z}$ is projected onto the PCA model plane (upper block) and residual subspace (lower block). After scaling the retained PCs, the middle block extract the non-Gaussian ICs. Next, the SVDD transformations are carried out for the ICs of the model plane, producing finally up to three univariate statistics, those are the D^2, the T^2, and the Q statistics to the right.

Implementation of the monitoring scheme

The implementation of the ICA-SVDD monitoring scheme requires the following steps:

1. Record a reference data set, including K samples, from the process to be monitored;

2. Mean center and scale the data;

3. Obtain a maximum likelihood PCA model by using the work by Wentzell et al.[21] if the residuals covariance matrix is known *a priori* or the technique by Narasimhan[22] if an estimation of the residual covariance matrix is required;

4. Compute the n retained score variables $\mathbf{t}=\mathbf{P}^T\mathbf{z}$, and the PCA residuals $\boldsymbol{\varepsilon}=\mathbf{R}^T\mathbf{z}$;

5. Select a function to approximate Eq. 6, $G(\cdot)$;

6. Iteratively determine the ICs by benchmarking each of the associated negentropy value against the confidence limit of the test statistic as discussed in the Negentropy-based non-Gaurssianity test section. Terminate the iteration when the first associated negentropy value is below this limit. The number of negentropy values above the confidence limit then represents the number of ICs, m;

7. After determining the separating matrices for the score variable sets $\mathbf{B}$, construct its orthogonal complements $\mathbf{B}^{\perp}$;

8. Compute an SVDD model for the ICs, and establish the confidence limits for the ICs in the feature spaces by determining the radius R.

9. Calculate the confidence limits for the T^2 and Q statistic as discussed by Jackson[14] and Nomikos,[26] respectively.

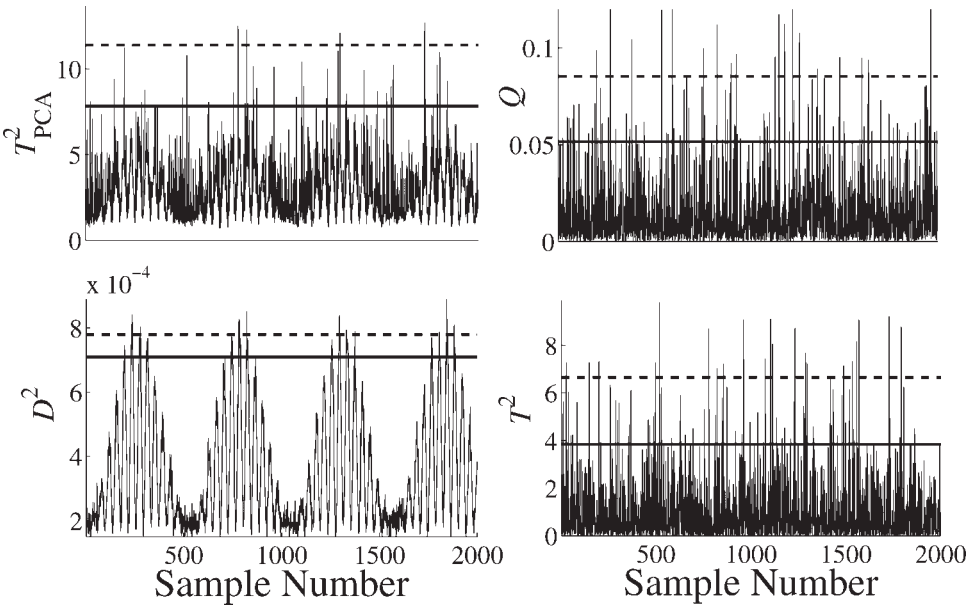

Figure 3. Univariate monitoring statistics for reference data; T^2_{PCA} statistic (upper left plot), Q statistic (upper right plot), D^2 statistic (lower left plot), and T^2 statistic (lower right plot).

The next section shows how to apply the aforementioned steps to establish an online monitoring model for a simulated example.

Simulation Example

The simulated process is a modified version of that used by Lee et al.[11] where a total of three "source" variables are generated as follows

$$\begin{cases} u_1(k) = 2\cos(0.08k)\sin(0.06k) \\ u_2(k) = \text{sign}[\sin(0.03k) + 9\cos(0.01k)] \\ u_3(k) \sim \mathcal{N}(0, 0.25), \end{cases} \tag{17}$$

where k is a sampling index. A total of five process variables $\mathbf{z}$, are then constructed as linear combinations of these source variables $\mathbf{u}^T = (u_1\ u_2\ u_3)$, $\mathbf{z}^{(0)} = \mathbf{\Gamma u}$, with

$$\mathbf{\Gamma} = \begin{bmatrix} 0.860 & 0.790 & 0.670 \\ -0.550 & 0.650 & 0.460 \\ 0.170 & 0.320 & -0.280 \\ -0.330 & 0.120 & 0.270 \\ 0.890 & -0.970 & -0.740 \end{bmatrix} \tag{18}$$

and noise sequences that were superimposed to $\mathbf{z}^{(0)}$, such that the recorded variables are $\mathbf{z} = \mathbf{z}^{(0)} + \mathbf{f}$, where $E\{\mathbf{f}\} = 0$, $E\{\mathbf{f}\mathbf{f}^T\} = 0.0025\mathbf{I}$ and $E\{\mathbf{f}(\mathbf{z}^{(0)})^T\} = \mathbf{0}$.

From the aforementioned process, a total of 2,000 samples were simulated for identifying an ICA-SVDD monitoring model. The reference data set of the process variables produced the following covariance matrix

$$\mathbf{S}_{zz} = \begin{bmatrix} 1.495 & 0.097 & 0.354 & -0.154 & -0.093 \\ 0.097 & 0.796 & 0.079 & 0.299 & -1.230 \\ 0.354 & 0.079 & 0.153 & -0.038 & -0.102 \\ -0.154 & 0.299 & -0.038 & 0.149 & -0.474 \\ -0.093 & -1.230 & -0.102 & -0.474 & 1.912 \end{bmatrix} \tag{19}$$

which had the following eigenvalues

$$\lambda_1 = 2.8369 \quad \lambda_2 = 1.5932 \quad \lambda_3 = 0.0692$$
$$\lambda_4 = 0.0025 \quad \lambda_5 = 0.0025. \tag{20}$$

As expected, the first three PCs correspond to interrelationships between the five process variables, while the remaining two refer to the noise variance. The negentropy values of the first three ICs were 9.500×10^{-3}, 5.109×10^{-4}, and 4.626×10^{-5}, while the 95% confidence limit is 1.484×10^{-4}. To confirm that the discarded PCs were Gaussian, the application of the ICA to the discarded PCs highlighted that the first IC followed a Gaussian distribution, since its negentropy value was 3.146×10^{-5}.

The application of the JB test with a confidence of 95% to the 5 PCs confirmed that the first two PCs were non-Gaussian, while the remaining three were Gaussian. Consequently, the online monitoring of this simulated process required the use of the D^2 statistic to represent the non-Gaussian signal components, a T^2 statistic to describe Gaussian signal components, both of which monitored process variation of the PCA model plane, and the Q statistic to monitor process variation in the PCA residual subspace.

To obtain the $\boldsymbol{\alpha}$ coefficients for the SVDD of the two extracted ICs, the optimal parameters for the Gaussian kernel functions, and the penalizing factor were found to be $\sigma = 2.8$, and $C = 0.02$, to produce a 95% confidence limit, and $\sigma = 2.5$, and $C = 0.15$, to generate a 99% limit.

For the simulated 2,000 samples, Table 2 summarizes the Type I error, or false alarm rate, of the T^2_{PCA}, the Q statistics (conventional PCA), and the D^2 and T^2 statistics of the ICA-SVDD model for a confidence of 95 and 99%.

Figure 3 shows the computed univariate statistics, and graphically demonstrated the impact of the lower-than-expected numbers of Type I errors for the T^2_{PCA} statistic,

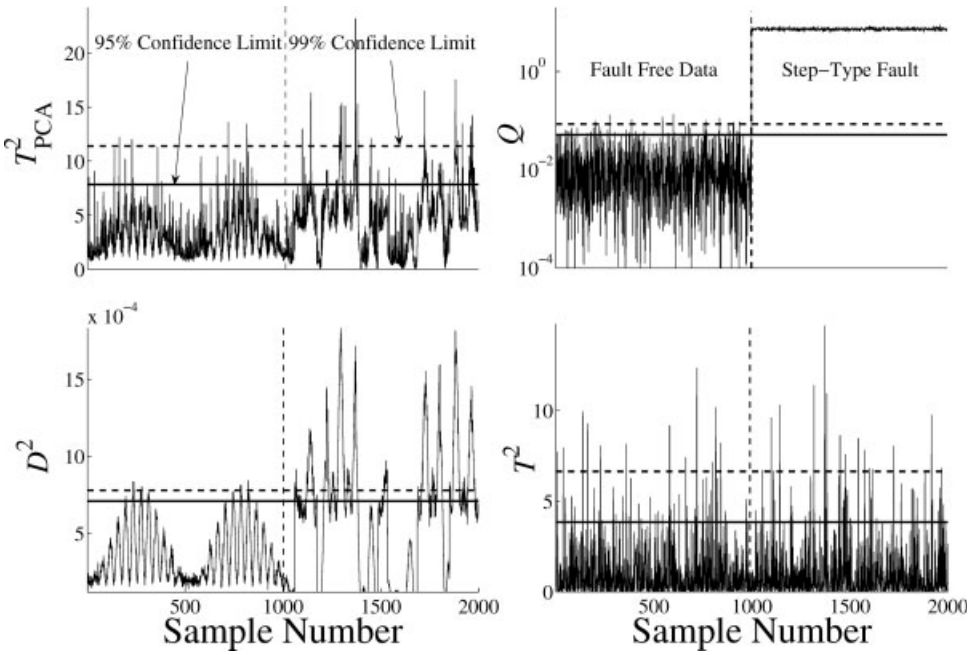

Figure 4. Univariate monitoring statistics describing step-type fault on process variable z_2; T^2_{PCA} statistic (upper left plot), Q statistic (upper right plot), D^2 statistic (lower left plot) and T^2 statistic (lower right plot).

which is later shown to result in a decrease in sensitivity for detecting anomalous behavior.

This can be attributed to the assumption that PCs follow a Gaussian distribution, which was not the case for the first two retained PCs.

To analyze a fault condition, the aforementioned monitoring model was next applied to a second data set, which also included 2,000 simulated samples using Eqs.17 and 18. Injecting a sensor bias to variable z_2, that took the form of a step of magnitude 3.0, then represented a fault condition. Figure 4 shows the monitoring charts for these data set, and indicates that the D^2 and the Q statistics were most affected by this event.

In contrast, the conventional T^2_{PCA} statistic only showed a sporadic response to the sensor bias. This, in turn, is in line with the previous observation that substantially fewer Type I errors than expected were noticed for the reference data. This example is, therefore, an indication that utilizing an incorrect distribution function to obtain confidence limits may (1) render the monitoring statistic insensitive, or (2) increase the number of Type I errors (false alarms).

To demonstrate that the decrease in sensitivity may render a fault undetectable, a second fault scenario was simulated. This time, the first "source signal" was augmented by superimposing a step-type fault of magnitude 1.0 to the latter half of a third data set that also contained 2,000 samples. In this case, the residual subspace of the PCA decomposition cannot be affected by this fault condition, which Figure 5 confirms.

The figure shows that T^2_{PCA} statistic only showed a sporadic response, which makes this event almost undetectable. In contrast, the D^2 statistic constructed from the ICA-SVDD approach detected this event in the latter half of the data set. The increased sensitivity of the ICA-SVDD approach is clearly related to the more accurate representation of the non-Gaussian trends encapsulated in the retained PCs.

Application to an Industrial Melter Process

This system is part of a disposal procedure, where a powder (waste material) is clad in glass. The melter 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, positioned around the vessel. Resulting from this heating procedure, the glass becomes molten homogeneously. The process of filling and heating 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 to the height of the nozzle, the next cycle of filling and heating begins.

Measurements of eight temperatures, the power in four induction coils, the viscosity of the molten glass, and the voltage were taken every 5 min. The heating and emptying cycles produced non-Gaussian readings of the recorded variables, which our analysis confirmed. Two data sets were recorded, a reference set of $K = 6{,}000$ samples (500 h), and a second set of 360 samples (30 h) for testing the performance of the ICA-SVDD monitoring approach. Given the sampling period of five minutes, there are steady-state relationships between the recorded variables.

A maximum likelihood PCA model was established next with the aid of the non-Gaussianity test introduced in the Negentropy-based non-Gaussianity test section. After the retention of 11 PCs the maximum likelihood PCA model produced residuals ε that did not show any remaining non-Gaussianity trends. The next step involved the determination of the number of ICs to be included in the D^2 statistic.

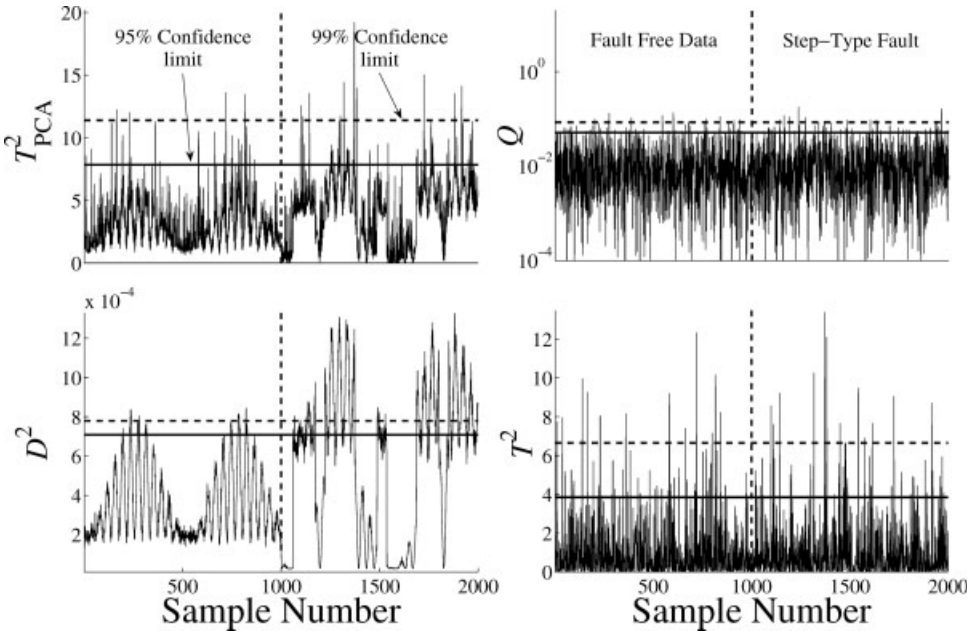

Figure 5. Univariate monitoring statistics describing step-type fault on "source" variable u_2; T^2_{PCA} statistic (upper left plot), Q statistic (upper right plot), D^2 statistic (lower left plot), and T^2 statistic (lower right plot).

Table 3 shows the calculated value of each IC and the 95% confidence limit.

Marking significant ICs in bold, it follows from Table 3 that a total of $m = 9$ ICs need to be included to extract the non-Gaussian components. This highlights that remaining components can be approximated by a Gaussian distribution, and confirms that applying the data model $\mathbf{z} = \mathbf{\Gamma u} + \mathbf{f}$, the source signals for melter process are both Gaussian and non-Gaussian.

To determine the 95% confidence limit for the D^2 statistic, the parameters σ and C were found to be 2.5 and 0.167. Using the three univariate monitoring statistics of the proposed monitoring scheme, and the conventional PCA-based statistic T^2, the left upper plot in Figure 6 shows that an excessive number of Type I errors were produced.

More precisely, the number of Type I errors for the T^2_{PCA}, D^2, the T^2, and the Q statistics were determined to be 12.430%, 4.910%, 5.020% and 4.930%, respectively. In contrast, considerably different numbers of Type I errors relative to the significance level of 5% were not experienced using the ICA-SVDD monitoring scheme. The lower plots in Figure 6 confirm this. While the simulation example in the previous section illustrated that the use of an incorrect distribution function may render the monitoring statistics insensitive, the analysis of the melter data are, therefore, an example of an excessive number of Type I errors or false alarms.

Table 3. Negentropy Value for ICs of Retained and Discarded PCs

95% Confidence limit	4.945×10^{-5}
ICs	Negentropy
IC_1	0.029
IC_2	0.012
IC_3	0.005
IC_4	0.005
IC_5	0.003
IC_6	0.002
IC_7	0.001
IC_8	6.570×10^{-4}
IC_9	5.806×10^{-4}
IC_{10}	1.842×10^{-5}

Figure 7 shows that the second data set represented a significant disturbance in two of the induction coils, A15 and A16, in the last third of the recording period.

The exact root cause of this event could not be determined, but it was noticeable that an increase in voltage (A44) arose at the same time. Figure 8 summarizes the application of PCA and the ICA-SVDD monitoring approach to this second data set.

Each of the univariate statistics showed a significant response to the disturbance; however, the conventional T^2_{PCA} and Q ones (upper charts in Figure 8) detected this event 17.75 h into the data set. Furthermore, two emptying cycles produced false alarms after about 3 and 10 h. This compares to the two statistics constructed from the ICA-SVDD approach (lower plots in Figuere 8), where no false alarms were noticeable during the first 17 h. In addition, the D^2 statistic (lower left plot) detected this event already after 17.2 h, i.e., 35 min earlier. This implies that filtering the non-Gaussian components from the retained PCs increased the sensitivity for fault detection in this example plus circumvented the production of false alarms, resulting from using an incorrect distribution function.

Concluding Summary

This article has studied the monitoring of multivariate systems that exhibit non-Gaussian processes behavior, which is a phenomenon that is often encountered in the chemical industry. Existing work in this area relies on the use of ICA

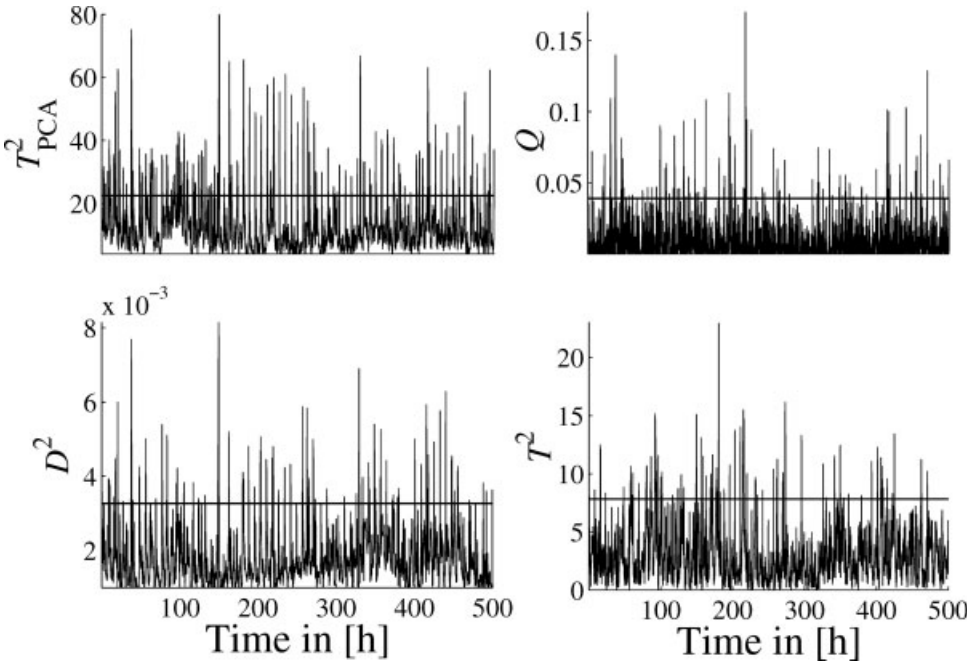

Figure 6. Univariate monitoring statistics describing reference data of melter process; T^2_{PCA} statistic (upper left plot), Q statistic (upper right plot), D^2 statistic (lower left plot), T^2 statistic (lower right plot).

to extract non-Gaussian components from the recorded variables. However, existing work has not addressed the following issues: (1) how to geometrically interpret ICs, (2) how to estimate confidence limits for ICs effectively, (3) how to evaluate the importance of ICs, and (4) how to deal with source variables that contain both Gaussian and non-Gaussian signal components.

Assuming that recorded data can be described by the data structure $\mathbf{z} = \mathbf{\Gamma u} + \mathbf{f}$, where $\mathbf{u}$ can contain both Gaussian and non-Gaussian components, and $\mathbf{f}$ are Gaussian residuals

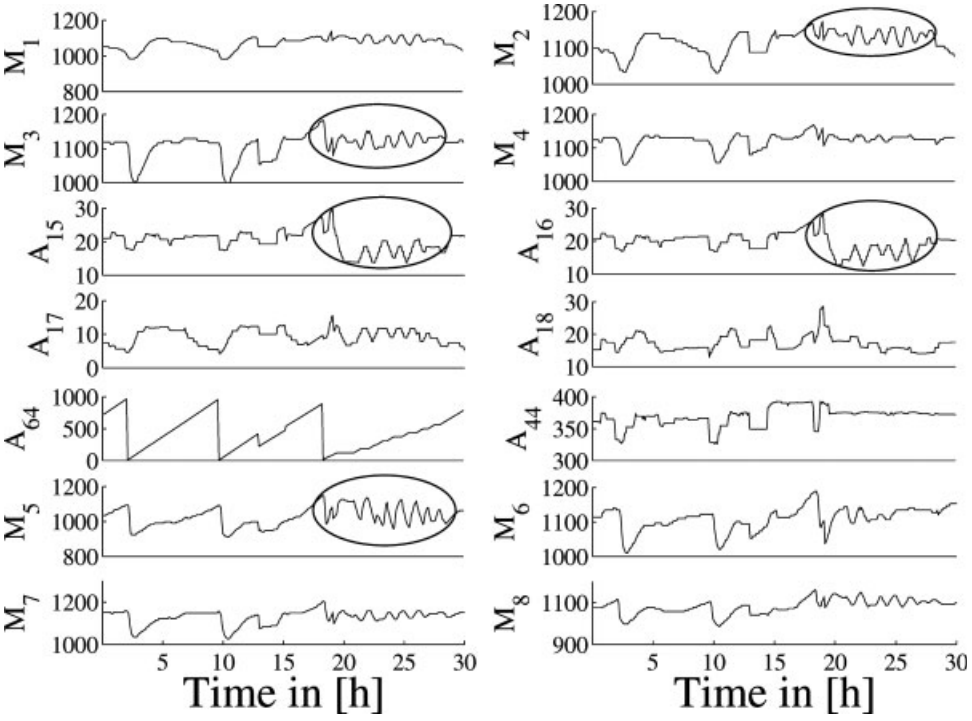

Figure 7. Recorded data of melter process showing the effect of a significant disturbance in power signals A15 and A16, and its effect on the temperatures, most notably M2, M3 and M5.

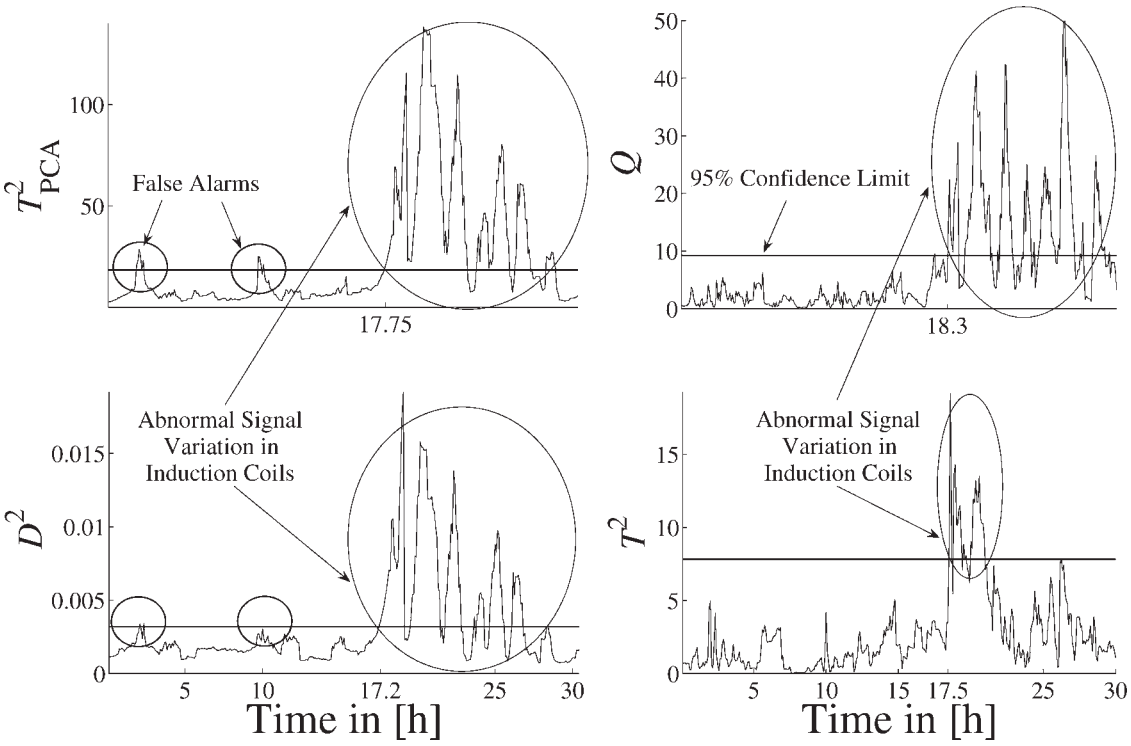

Figure 8. Univariate monitoring statistics describing disturbance of induction coils; T^2_{PCA} statistic (upper left plot), Q statistic (upper right plot), D^2 statistic (lower left plot), T^2 statistic (lower right plot).

of zero mean and covariance Σ_f, these issues have been studied in this article. The contribution of this article commences by providing a proof that the application of maximum likelihood PCA can establish a model plane captures the variation within the source signals **u**, plus some variation from the residuals **f**. Moreover, the complementary residual subspace does not reflect any variation of the source signals **u**, but linear combinations of the residuals **f**. This implies that the non-Gaussian components can be extracted from the retained PCs.

To address the first of the aforementioned issue, ICA has been applied to the retained PCs to maintain the geometric interpretation of the PCA decomposition. Existing works on determining confidence limits for ICs rely on *ad hoc* approaches or utilize the computationally expensive KDE. This article proposes the use of a SVDD. As the calculation of a mapping function to a feature space, and the determination of the hypersphere to envelop the transformed ICs reduce to a numerically efficient quadratic programming problem, the SVDD approach does not suffer from a high computational load.

The third issue has been tackled by introducing a non-Gaussianity test for determining how many ICs need to be included, and how severe, i.e., how important, is their non-Gaussianity. A theoretical analysis of the proposed test has yielded that it is a generalization of the well known Jarque-Bera test. Finally, the article has introduced a monitoring scheme by incorporating ICA, SVDD, and the new non-Gaussianity test into the PCA-based monitoring framework. The utility of this scheme has been demonstrated by two application studies, a simulation example and the analysis of recorded data from an industrial melter process.

Both application studies demonstrated that using conventional PCA either rendered the established monitoring approach insensitive or produced an excessive number of Type I errors. In contrast, applying the proposed monitoring scheme removed these undesired impacts and produced a more sensitive monitoring approach to detect fault conditions in both application studies. Further work on studying ICA and SVDD for process monitoring, which is beyond the scope of this article, will involve the integration of the proposed scheme in a dynamic multivariate statistical, as well as a nonlinear context.

Acknowledgments

The work was partly supported by the National Natural Science Foundation of China, grant number 60721062, and the National High Technology Research and Development Program of China (863 Program), grant number 2007AA04Z162. Xueqin Liu is also grateful for the financial support from the Engineering and Physical Science Research Council (EPSRC), grant number GR/S84354/01.

Literature Cited

1. Venkatasubramanian V, Rengaswamy R, Kavuri SN, Yin K. A review of process fault detection and diagnosis Part III: Process history based methods. *Comp Chem Eng*. 2003;27(3):327–346.
2. Wise B, Gallagher NB. The process chemometrics approach to process monitoring and fault detection. *J Proc Contr*. 1996;6(6):329–348.
3. Jackson JE, Mudholkar GS. Control procedures for residuals associated with principal component analysis. *Technometrics*. 1979;21: 341–349.
4. Li RF, Wang XZ. Dimension reduction of process dynamic trends using independent component analysis. *Comp Chem Eng*. 2002;26(3): 467–473.

5. Kano M, Tanaka S, Hasebe S, Hashimoto I, Ohno H. Monitoring independent components for fault detection. *AIChE J*. 2003;49(4):969–976.
6. 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(6–7):1157–1166.
7. Lee JM, Yoo C, Lee IB. Statistical process monitoring with independent component analysis. *J Proc Contr*. 2004;14(5):467–485.
8. Silverman BW. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall; 1986.
9. Martin EB, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J Proc Contr*. 1996;6(6):349–358.
10. Chen Q, Kruger U, Leung AYT. Regularised kernel density estimation for clustered process data. *Contr Eng Practice*. 2004;12(3):267–274.
11. Lee JM, Qin SJ, Lee IB. Fault detection and diagnosis based on modified independent component analysis. *AIChE J*. 2006;52(10):3501–3514.
12. Jarque CM, Bera AK. A test for normality of observations and regression residuals. *Intl Statist Rev*. 1987;55(2):1–10.
13. Xie L, Wu J. Global optimal ica and its application in meg data analysis. *Neurocomp*. 2006;69(16–18):2438–2442.
14. Jackson JE. *A Users Guide to Principal Components*. New York: Wiley; 1991.
15. Hyvarinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks*. 2000;13(4–5):411–430.
16. Hyvarinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Networks*. 1999;10(3):626–634.
17. Vapnik VN. *Statistic Learning Theory*. New York: Wiley; 1998.
18. Tax DMJ, Duin RPW. Support vector domain description. *Pattern Recog Letts*. 1999;20:1191–1199.
19. Tax DMJ, Duin RPW. Support vector data description. *Mach Learn*. 2004;54:4–66.
20. Tax DMJ. *One-class classification*., Delft University of Technology; 2001. Available at: http://www-ict.ewi.tudelft.nl/~davidt/oneclass.html. PhD thesis.
21. Wentzell PD, Andrews DT, Hamilton DC, Faber K, Kowalski BR. Maximum likelihood principal component analysis. *J Chemomet*. 1997;11(4):339–366.
22. Schuermans M, Markovsky PD, Wentzell PD, van Huffel S. On the equivalence between total least squares and maximum likelihood PCA. *Anal Chim Acta*. 2005;544(1–2):254–267.
23. Narasimhan S, Shah SL. Model identification and error covariance estimation from noisy data using PCA. *Contr Eng Prac*. 2008;16(1):146–155.
24. Cheung Y, Xu L. Independent component ordering in ICA time series analysis. *Neurocomp*. 2001;41(1–4):145–152.
25. Hyvarinen A. Survey on independent component analysis. *Neural Comp Surv*. 1999;2:94–128.
26. Nomikos P, MacGregor JF. Multivariate SPC charts for monitoring batch processes. *Technometrics*. 1995;37(1):41–59.
27. Stuart A, Ord JK. *Kendall's Advanced Theory of Statistics*, vol 1. Oxford: Oxford University Press; 1987.
28. Pearson ES, D'Agostino RB, Bowman KO. Tests for departure from normality comparison of powers. *Biometrika*. 1977;64(2):231–246.
29. Pearson ES. Notes on tests for normality. *Biometrika*. 1931;22(3–4):423–424.
30. Urzua CM. On the correct use of omnibus tests for normality. *Econ Letts*. 1996;53(3):247–251.

Appendix A: Proof of Theorem 1

Without restricting generality, $G(y)$ is given here by $G(y) = y^4$, where $y,v \in \mathcal{N}(0,1)$ and $y_1,y_2, \ldots ,y_K$ are K samples from y. The mean $E\{G(v)\}$, and variance $\mathrm{var}\{G(v)\}$ is given by

$$\begin{cases} E\{G(v)\} = E\{v^4\} = 3 \\ \mathrm{var}\{G(v)\} = E\{G^2(v)\} - (E\{G(v)\})^2 = E\{v^8\} \\ \quad - (E\{v^4\})^2 = 96 \end{cases} \tag{A1}$$

Utilizing the central-limit theorem for $K \to \infty$, it follows that

$$\lim_{K\to\infty} \sqrt{\frac{K}{\mathrm{var}\{G(v)\}}} \left(\frac{1}{K} \sum_{i=1}^{K} G(y_i) - 3 \right) \sim \mathcal{N}(0,1). \tag{A2}$$

So $K{\cdot}J(y_1,y_2, \ldots ,y_K) \sim 96\ \chi^2(1)$.

Appendix B: Equivalence of JB- and Negentropy-Based Tests

It follows from Theorem 1 that

$$\lim_{K\to\infty} \frac{K}{\mathrm{var}\{G(v)\}} \left(\frac{1}{K} \sum_{i=1}^{K} G(y_i) - E\{G(v)\} \right)^2 \sim \chi^2(1). \tag{B1}$$

By defining $G_1(y) = y^3$, $E\{G(v)\} = E\{v^3\} = 0$, $\mathrm{var}\{G(v)\} = \mathrm{var}\{v^3\} = E\{v^6\} - (E\{v^3\})^2 = E\{v^6\} = 15$, and utilizing Eq. B1 produces

$$\lim_{K\to\infty} K \frac{\left(\frac{1}{K} \sum_{i=1}^{K} y^3 \right)^2}{15} \sim \chi^2(1) \tag{B2}$$

Next, defining $G_2(y) = y^4$, $E\{G(v)\} = E\{v^4\} = 3$, $\mathrm{var}\{G(v)\} = \mathrm{var}\{v^4\} = 96$, and again, using Eq. B1 gives rise to

$$\lim_{K\to\infty} K \frac{\left(\frac{1}{K} \sum_{i=1}^{K} y^4 - 3 \right)^2}{96} \sim \chi^2(1). \tag{B3}$$

Using Eqs. B2 and B3, and further defining $\sqrt{b_1'} = \gamma_3'$ and $b_2' = \gamma_4'$, where $\gamma_i' = \frac{1}{K} \sum_{j=1}^{K} y_j^i$, is the ith moment[27] for $E\{y\} = 0$ yields, as a sum, a χ^2 distribution with two degrees of freedom

$$\lim_{K\to\infty} K \left(\frac{\left(\sqrt{b_1'}\right)^2}{15} + \frac{\left(b_2' - 3\right)^2}{96} \right) \sim \chi^2(2). \tag{B4}$$

It should be noted that Eq. B2 is based on the fact that $\sqrt{b_1'}$ and b_2' are nearly independent for large K.[28]

In a similar fashion to Eq. B2, the widely used Jarque-Bera test statistic follows a χ^2 distribution with two degrees of freedom[12]

$$\lim_{K\to\infty} K \left(\frac{\left(\sqrt{b_1}\right)^2}{6} + \frac{(b_2 - 3)^2}{24} \right) \sim \chi^2(2). \tag{B5}$$

where $\sqrt{b_1} = \gamma_3/\gamma_2^{3/2}$ is the sample skewness, $b_2 = \gamma_4/\gamma_2^2$ is the sample kurtosis, $\gamma_i = \frac{1}{K} \sum_{j=1}^{K} (y_j - \bar{y})^i$ is the i^{th} sample moment,[27] and $\bar{y}$ is the sample mean. Asymptotically $\sqrt{b_1}$ and b_2 converge to 0 and 3, and their asymptotic variances are $6/K$ and $24/K$, respectively.[29] For the JB-test, the sample mean $\bar{y}$ and variance γ_2 for determining the skewness $\sqrt{b_1}$ and b_2 kurtosis need to be estimated by the samples. For the negentropy-based test, however, the mean and variance are already known.

Under the assumption that $y_1,y_2, \ldots , y_K$ are samples of a Gaussian distributed population y i.e., $y \in \mathcal{N}(0,1)$, we show that Eq. B5 converges to Eq. B4. For the negentropy-based test, the sample mean and variance are defined as 0 and 1,

respectively, that is $\bar{y} = 0$, $\gamma_2 = 1$ (or $\mathrm{var}\{y\} = 1$). This, however, implies that $\sqrt{b_1} = \sqrt{b_1'}$ and $b_2 = b_2'$. We now examine the distribution of $\sqrt{b_1'}$ and b_2'.

The asymptotic means and variance of $\sqrt{b_1'}$ are[27]

$$E\left\{\sqrt{b_1'}\right\} = \gamma_3' = E\{y\} + \left((E\{y\})^2 + 3\ \mathrm{var}\{y\}\right) = 0,$$

$$\mathrm{var}\left\{\sqrt{b_1'}\right\} = \frac{1}{K}\left(\gamma_6' - (\gamma_3')^2\right) = \frac{15}{K}. \quad \text{(B6)}$$

where γ_i' is the moment of order i and $\gamma_{2i}' = \frac{(\mathrm{var}\{y\})^{2i}}{2^i}\frac{2i!}{i!}$. Similarly, the asymptotic means and variance of b_2' are[27]

$$E\{b_2'\} = \gamma_4' = E\{y\})^4 + 6(E\{y\})^2\mathrm{var}\{y\} + 3(\mathrm{var}\{y\})^2 = 3,$$

$$\mathrm{var}\{b_2'\} = \frac{1}{K}\left(\gamma_8' - (\gamma_4')^2\right) = \frac{96}{K}. \quad \text{(B7)}$$

Urzua[30] showed that Eq. B8

$$\lim_{K\to\infty}\left(\frac{\left(\sqrt{b_1}\right)^2}{\mathrm{var}\left\{\sqrt{b_1}\right\}} + \frac{(b_2 - E\{b_2\})^2}{\mathrm{var}\{b_2\}}\right) \sim \chi^2(2). \quad \text{(B8)}$$

produces

$$\lim_{K\to\infty} K\left(\frac{\left(\sqrt{b_1'}\right)^2}{15} + \frac{\left(b_2' - 3\right)^2}{96}\right) \sim \chi^2(2). \quad \text{(B9)}$$

by using Eqs. B7 and 8.

Equation B9 is the same as Eq. B4. Thus, the Jarque-Bera test statistic can be formulated on the basis of the negentropy-based test statistic by taking $G(\cdot)$ to be y^3 and y^4, and, subsequently, add both expressions. If other forms for $G(\cdot)$ are used, a different test can be obtained. Consequently, the proposed test can be seen as a generalization of the JB-test.

Manuscript received Aug. 23, 2007, revision received Dec. 31, 2007, and final revision received Apr. 24, 2008.