A Quality Relevant Non-Gaussian Latent Subspace Projection Method for Chemical Process Monitoring and Fault Detection

Junichi Mori and Jie Yu
Dept. of Chemical Engineering, McMaster University, Hamilton, Ontario L8S 4L7, Canada

DOI 10.1002/aic.14261
Published online October 28, 2013 in Wiley Online Library (wileyonlinelibrary.com)

Partial least-squares (PLS) method has been widely used in multivariate statistical process monitoring field. The goal of traditional PLS is to find the multidimensional directions in the measurement-variable and quality-variable spaces that have the maximum covariances. Therefore, PLS method relies on the second-order statistics of covariance only but does not takes into account the higher-order statistics that may involve certain key features of non-Gaussian processes. Moreover, the derivations of control limits for T^2 and squared prediction error (SPE) indices in PLS-based monitoring method are based on the assumption that the process data follow a multivariate Gaussian distribution approximately. Meanwhile, independent component analysis (ICA) approach has recently been developed for process monitoring, where the goal is to find the independent components (ICs) that are assumed to be non-Gaussian and mutually independent by means of maximizing the high-order statistics such as negentropy instead of the second-order statistics including variance and covariance. Nevertheless, the IC directions do not take into account the contributions from quality variables and, thus, ICA may not work well for process monitoring in the situations when the quality variables have strong influence on process operations. To capture the non-Gaussian relationships between process measurement and quality variables, a novel projection-based monitoring method termed as quality relevant non-Gaussian latent subspace projection (QNGLSP) approach is proposed in this article. This new technique searches for the feature directions within the measurement-variable and quality-variable spaces concurrently so that the two sets of feature directions or subspaces have the maximized multidimensional mutual information. Further, the new monitoring indices including I^2 and SPE statistics are developed for quality relevant fault detection of non-Gaussian processes. The proposed QNGLSP approach is applied to the Tennessee Eastman Chemical process and the process monitoring results of the present method are demonstrated to be superior to those of the PLS-based monitoring method. © 2013 American Institute of Chemical Engineers *AIChE J*, 60: 485–499, 2014

Keywords: quality relevant process monitoring, fault detection, non-Gaussian latent subspace projection, partial least squares, independent component analysis, multidimensional mutual information

Introduction

Process monitoring, fault detection and diagnosis are gaining significant attention for the rapid detection of abnormal operation, process upsets, equipment malfunctions, sensor failures, and other special events in industrial plants to improve plant safety, product quality, energy efficiency, and profit margin.[1–3] Recently, owing to the fast development of measurement, automation, and advanced computing technologies, a huge number of process variables can be frequently measured and recorded in industrial plant historians, which make it possible to conduct data-driven large-scale process monitoring and fault diagnosis. Meanwhile, effective process monitoring plays a critical role in ensuring product quality, operation safety, and manufacturing sustainability.

The approaches to process monitoring, fault detection and diagnosis fall into the two categories, which are the model-based and the data-driven techniques.[4] Model-based process monitoring methods may be applicable only if the accurate mechanistic models of processes can be developed.[5–7] However, those first-principle models require in-depth knowledge about processes and also it is difficult and time consuming to build precise mechanistic models for large-scale complex industrial plants. Conversely, the data-driven monitoring techniques have become increasingly attractive because they do not require in-depth fundamental knowledge and mechanistic models but instead depend on historical process data only. Traditionally, univariate statistical process control (SPC) has been applied for process monitoring. Nevertheless, most SPC methods are based on control charts of individual or noncorrelated process variables and, thus, the highly correlated process variables in industrial plants can cause the failure of conventional SPC methods.[8]

Multivariate statistical process monitoring (MSPM) techniques have been developed to extract useful information from large number of highly correlated process variables and historical data sets.[9–15] Two popular latent variables methods in MSPM field are principal component analysis (PCA) and

Correspondence concerning this article should be addressed to J. Yu at jieyu@mcmaster.ca.

partial least squares or projection to latent structure (PLS).[16–24] The major advantages of these methods include their strong capability to deal with the colinearity among different process variables and identify the statistical model within a lower-dimensional latent subspace that well retains the multivariable correlation structure. Then the statistics such as T^2 and squared prediction error (SPE) are proposed for extracting the critical features of process data for fault detection and diagnosis.[25,26] Furthermore, the total projection to latent structures (T-PLS) method is proposed for process monitoring.[27] Compared to conventional PLS method that is based on regression models, T-PLS approach can separate the orthogonal and correlated parts to the quality variable and, thus, is proven to be more suitable for process monitoring and fault detection. The conventional PCA/PLS-based process monitoring techniques are based on the second-order statistics of covariance only but do not take into account the higher-order statistics. Hence, they may not effectively extract the non-Gaussian process features that are characterized by the higher-order statistics, even though PCA and PLS models do not require Gaussian distribution explicitly. Moreover, the derivations of T^2 and SPE control limits in PCA/PLS-based process monitoring methods are based on the assumption that the process data follow a multivariate Gaussian distribution approximately.[28] In industrial processes, however, operating condition shifts are often encountered due to the changes of various factors such as feedstock, product specification, set points, and manufacturing strategy.[29] Such operating condition changes often result in non-Gaussian probability distribution of process data. As an alternative solution, eigenvalue decomposition on the covariance matrices of process measurement variables is utilized to determine the dissimilarity factor between the normal and the monitored data sets.[30] Nevertheless, this method suffers from the same issue as PCA/PLS-based monitoring methods that only the second-order statistics are taken into account and, thus, the non-Gaussian process features may not be efficiently extracted.

To deal with non-Gaussian processes, independent component analysis (ICA) has been applied to project multivariate process data into latent subspace of statistically independent components (IC).[31–37] ICs are assumed to be non-Gaussian and mutually independent based on high-order statistics and they retain the non-Gaussian process features that may not be effectively extracted in traditional PCA/PLS methods. Moreover, the ICA-based statistics like I^2 and SPE are developed for detecting faulty operation.[38] More recently, multidimensional mutual information is adopted to measure the statistical independency between IC subspaces and further determine the dissimilarity factors between the normal benchmark and the monitored data sets for process monitoring and fault detection.[39] This method takes into consideration not only the high-order statistics but also the time-varying process dynamics. However, if the variations in the process measurement variables are most influential on product quality variables, the above ICA-based monitoring techniques may not be well suited because only the process measurement variables are utilized in the developed statistical models while the product quality variables are excluded. In other words, they do not take into account the quality variables as in the PLS-based monitoring methods. Another non-Gaussian process monitoring technique is based on Gaussian mixture models (GMM) that decompose the process data into multiple Gaussian components with different means and covariances corresponding to various operational conditions and modes.[40] In this way, the globally non-Gaussian process data can be characterized as mixture models of different Gaussian components and then the Bayesian inference strategy can be developed to incorporate multiple local models for fault detection.[37] Furthermore, Gaussian mixture model can be updated by adopting particle filter strategy to take into account the dynamic changes of operating scenarios.[41] In addition, an ensemble clustering-based process pattern construction method and multiple ICA-PCA model-based multimode process monitoring technique are developed for operating mode identification and fault detection.[42] Due to the clustering algorithm as well as the integration of PCA and ICA methods, both Gaussian and non-Gaussian process features in the multimode operating data can be captured. However, the non-Gaussian process monitoring methods still do not utilize output variables especially product quality variables and, thus, the detected abnormal operating events may not be relevant to any degradations of product quality or losses of other operational objectives such as energy efficiency and sustainability.

Alternately, supervised learning techniques such as Fisher discriminant analysis (FDA) and support vector machine (SVM) have been developed for chemical process monitoring.[43,44] FDA approach can identify multiple classes with both maximized between-class separation and minimized within-class scattering. FDA may become well suited only if the subset of data in each class do not have significant within-class multimodality. To overcome this limitation, the localized Fisher discriminant analysis (LFDA) has recently been proposed for process monitoring and fault detection.[45] Nevertheless, the performance of LFDA depends on the way of calculating a similarity matrix and, thus, the best selection of similarity matrix in LFDA algorithm is very important. Conversely, SVM can perform nonlinear classification by maximizing separating margin between support vector hyperplanes. However, all these methods are based on supervised learning models and, thus, require known class labels of all the training samples, which may not be realistic for industrial applications. To overcome this limitation, support vector clustering (SVC)-based probabilistic approach is proposed for unsupervised process monitoring.[46] Different from SVM, SVC has ability to classify the unlabeled training samples and, thus, known class labels are not needed in advance. Nevertheless, the above supervised and unsupervised monitoring methods typically do not include the output quality variables in the classification models and, thus, are not quality relevant either.

In this study, a novel output quality variable relevant non-Gaussian latent subspace projection method is proposed to monitor complex chemical processes that follow non-Gaussian distributions. Both the process measurement and product quality variables are used to extract the non-Gaussian subspaces for monitoring the abnormal behaviors in process operations that have significant influence on product quality. The basic idea is to estimate the non-Gaussian loading matrices of both process measurement and product quality variables, respectively, so that the mutual information between latent scores of measurement and quality variables is maximized. In this way, the proposed method can identify the feature directions in the measurement-variable and quality-variable spaces concurrently to retain the maximized statistical dependency between two latent subspaces. With

the quality relevant non-Gaussian latent variable model, I^2 and SPE indices are further proposed and compared for process monitoring and fault detection. In contrast to PLS-based monitoring methods, the proposed approach utilizes the high-order statistics of mutual information instead of the second-order statistics of covariance and, thus, can well extract the non-Gaussian features from process data. Meanwhile, compared to ICA, GMM or supervised learning-based monitoring approaches, both process measurement and output quality variables are included in the non-Gaussian model of the presented method so that the latent directions within input and output spaces are concurrently searched for with optimized mutual information.

The remainder of the article is organized as follows. "Review of PLS and ICA-Based Process Monitoring Methods" section briefly reviews the PLS and ICA-based monitoring technique and discusses the issues of these conventional methods. "Quality Relevant non-Gaussian Latent Subspace Projection Approach for Process Monitoring" section describes the proposed quality relevant non-Gaussian latent subspace projection (QNGLSP) approach and the corresponding monitoring indices for capturing abnormal process operations. "Application Example" section demonstrates the utility and performance of the new process monitoring approach through the application example of the Tennessee Eastman Chemical process and its comparison to the PLS-based monitoring method. Finally, the conclusions of this work are summarized in "Conclusions" section.

Review of PLS and ICA-Based Process Monitoring Methods

PLS-*based process monitoring method*

PLS handles high-dimensional correlated data by finding the multidimensional latent directions in both the measurement-variable and quality-variable spaces with the maximum covariance. Given an input matrix $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m]\in\Re^{n\times m}$ consisting of n samples along m process measurement variables and an output matrix $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_k]\in\Re^{n\times k}$ along k quality variables, they are decomposed onto low-dimensional subspaces as follows

$$\mathbf{X}=\mathbf{T}\mathbf{P}^T+\mathbf{E} \tag{1}$$

$$\mathbf{Y}=\mathbf{T}\mathbf{Q}^T+\mathbf{F} \tag{2}$$

where $\mathbf{T}=[\mathbf{t}_1,\mathbf{t}_2,\ldots,\mathbf{t}_d]\in\Re^{n\times d}$ is the score matrix, $\mathbf{P}=[\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_d]\in\Re^{m\times d}$ is the loading matrix for $\mathbf{X}$, $\mathbf{Q}=[\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_d]\in\Re^{k\times d}$ is the loading matrix for $\mathbf{Y}$, $\mathbf{E}\in\Re^{n\times m}$ denotes the residual matrix for $\mathbf{X}$, $\mathbf{F}\in\Re^{n\times k}$ represents the residual matrix for $\mathbf{Y}$, and d is the selected number of latent variables in PLS model. The basic idea of PLS is to determine the score matrix $\mathbf{T}$ and the loading matrices $\mathbf{P}$ and $\mathbf{Q}$ from $\mathbf{X}$ and $\mathbf{Y}$ through the nonlinear iterative partial least-squares (NIPALS) algorithm.[47]

Given a new test sample $\mathbf{x}$, its corresponding prediction and residual vectors are given as follows

$$\text{Score}:\mathbf{t}=\mathbf{x}\mathbf{P} \tag{3}$$

$$\text{Prediction}:\hat{\mathbf{x}}=\mathbf{x}\mathbf{P}\mathbf{P}^T \tag{4}$$

$$\text{Residual}:\mathbf{e}=\mathbf{x}\left(\mathbf{I}_m-\mathbf{P}\mathbf{P}^T\right) \tag{5}$$

where $\mathbf{I}_m$ is a $m\times m$ identity matrix. The following PLS-based T^2 and SPE statistics are used as the measures of variations in the latent variable and residual subspaces for process monitoring

$$T^2=\mathbf{t}\left\{\frac{1}{n-1}\mathbf{T}^T\mathbf{T}\right\}^{-1}\mathbf{t}^T \tag{6}$$

$$\text{SPE}=\mathbf{e}\mathbf{e}^T \tag{7}$$

where the confidence limits for T^2 and SPE statistics can be estimated from F and χ^2 distributions, respectively.[14]

ICA-*based process monitoring method*

PLS is based on the covariance between the score vectors of the measurement and quality variables, respectively. However, it does not take into account the high-order statistics and, thus, may not be well suited in extracting the non-Gaussian features from process data. In contrast, ICA is developed for non-Gaussian process monitoring on the basis of high-order statistics. It is essentially a multivariate statistical technique for computing ICs that are assumed to be non-Gaussian and mutually independent.[48] Given the input matrix $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m]\in\Re^{n\times m}$, all the process measurement variables are assumed to be generated as linear combinations of m unknown ICs

$$\mathbf{X}^T=\mathbf{A}\mathbf{S}^T \tag{8}$$

where $\mathbf{A}=[\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_d]\in\Re^{m\times d}$ is unknown mixing matrix and $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_d]\in\Re^{n\times d}$ represents the IC matrix. The solution is equivalent to finding a demixing matrix $\mathbf{W}=[\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_d]^T\in\Re^{d\times m}$ as follows

$$\mathbf{S}^T=\mathbf{W}\mathbf{X}^T \tag{9}$$

where the ICs $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_d]$ have the maximized statistical independency in terms of negentropy among each other.[49,50]

Given a new test sample vector $\mathbf{x}$, its corresponding IC score, prediction, and residual vectors are given below

$$\text{IC Score}:\mathbf{y}=\mathbf{x}\mathbf{W}^T \tag{10}$$

$$\text{Prediction}:\hat{\mathbf{x}}=\mathbf{x}\mathbf{W}^T\mathbf{A}^T \tag{11}$$

$$\text{Residual}:\mathbf{e}=\mathbf{x}\left(\mathbf{I}_m-\mathbf{W}^T\mathbf{A}^T\right) \tag{12}$$

Further, the I^2 and SPE statistics can be defined as follows for process monitoring[38]

$$I^2=\mathbf{y}\mathbf{y}^T \tag{13}$$

$$\text{SPE}=\mathbf{e}\mathbf{e}^T \tag{14}$$

where the confidence limits for I^2 and SPE statistics can be estimated through kernel density estimation (KDE).[51]

Quality Relevant Non-Gaussian Latent Subspace Projection Approach for Process Monitoring

The basic idea of PLS approach is to optimize the loading matrices $\mathbf{P}$ and $\mathbf{Q}$ from the input and output data matrices $\mathbf{X}$ and $\mathbf{Y}$ by means of the NIPALS algorithm. The embedded optimization problem in PLS is defined as

$$\begin{aligned}\max\quad & \psi\left(\mathbf{w}_i^{\text{PLS}},\mathbf{c}_i^{\text{pls}}\right)=\text{cov}\left(\mathbf{t}_i^{\text{PLS}},\mathbf{u}_i^{\text{PLS}}\right)=\text{cov}\left(\mathbf{X}_i\mathbf{w}_i^{\text{PLS}},\mathbf{Y}_i\mathbf{c}_i^{\text{PLS}}\right)\\ \text{s.t.}\quad & ||\mathbf{w}_i^{\text{PLS}}||=1,||\mathbf{c}_i^{\text{PLS}}||=1\end{aligned} \tag{15}$$

where $\psi\left(\mathbf{w}_i^{\text{PLS}},\mathbf{c}_i^{\text{PLS}}\right)$ is the objective function $\mathbf{w}_i^{\text{PLS}}$ and $\mathbf{c}_i^{\text{PLS}}$ are the weighting vectors, and $\text{cov}\left(\mathbf{t}_i^{\text{PLS}},\mathbf{u}_i^{\text{PLS}}\right)$ denotes the

covariance between the score vectors $\mathbf{t}_i^{\mathrm{PLS}}$ and $\mathbf{u}_i^{\mathrm{PLS}}$. It should be noted that the weighting vector $\mathbf{w}_i^{\mathrm{PLS}}$ also corresponds to the i-th eigenvector of the matrix $\mathbf{X}^T\mathbf{Y}\mathbf{Y}^T\mathbf{X}$.[52] If process data follow Gaussian distribution, the joint Gaussian density function p_{G} of $\mathbf{v}=[v_1,\ldots,v_M]$ can be described as follows

$$p_{\mathrm{G}}(\mathbf{v})=f(\mathbf{v}|\mu,\Sigma)=\frac{1}{(2\pi)^{M/2}|\Sigma|^{1/2}}\exp\left(-\frac{1}{2}(\mathbf{v}-\mu)^T\Sigma^{-1}(\mathbf{v}-\mu)\right) \tag{16}$$

where $\mathbf{v}$ represents the combined input and output variables, μ denotes a M-dimensional mean vector, and Σ is a $M\times M$ covariance matrix. As Eq. 16 indicates, if data are normalized to zero-mean, $p_{\mathrm{G}}(\mathbf{v})$ is parameterized by the second-order statistic of covariance only. Therefore, PLS is able to capture the characteristic relationship between process measurement and quality variables if process data follow Gaussian distribution approximately.

However, PLS-based monitoring methods may not be well suited if process data follow significantly non-Gaussian distribution because the joint density function cannot be adequately characterized by the second-order statistics of covariance only. Specifically, the joint density function $p(\mathbf{v})$ of non-Gaussian data with up to fifth-order statistics can be expressed through Edgeworth expansion as[53]

$$p(\mathbf{v})\approx p_{\mathrm{G}}(\mathbf{v})\left(1+\frac{1}{3!}\sum_{i,j,k}\kappa^{i,j,k}h_{ijk}(\mathbf{v})+\frac{1}{4!}\sum_{i,j,k,l}\kappa^{i,j,k,l}h_{ijkl}(\mathbf{v})+\frac{1}{72}\sum_{i,j,k,l,p,q}\kappa^{i,j,k}\kappa^{l,p,q}h_{ijklpq}(\mathbf{v})\right) \tag{17}$$

where $p_G(\mathbf{v})$ denotes the Gaussian density function with the same mean and covariance as $p(\mathbf{v})$, (i,j,k), (i,j,k,l), and $(i,j,k,l,p,q)\in\{1,\ldots,M\}$ are the input dimensions, h_{ijk}, h_{ijkl}, and h_{ijklpq} are the ijk-th, $ijkl$-th, and $ijklpq$-th Hermite polynomials $\kappa^{i,j,k}=\frac{\kappa^{ijk}}{\sigma_i\sigma_j\sigma_k}$ is the standardized cumulant with κ^{ijk} being the cumulant for input dimension (i,j,k), and $\kappa^{i,j,k,l}=\frac{\kappa^{ijkl}}{\sigma_i\sigma_j\sigma_k\sigma_l}$ is the standardized cumulant with κ^{ijkl} being the cumulant for input dimension (i,j,k,l). Since non-Gaussian probability density function $p(\mathbf{v})$ includes the higher-order statistics instead of covariance only, PLS may not efficiently extract the non-Gaussian process features of measurement variables that contain sufficient information on product quality variables. Thus, PLS-based monitoring methods may not be effective in detecting abnormal events of non-Gaussian processes.

In ICA method, ICs are calculated by using the mutual information between the measurement variables and the high-order statistics are taken into account for extracting non-Gaussian process features. However, it does not incorporate output quality variables in data analysis and, thus, may not specifically isolate the abnormal variations of process measurement variables that have significant influence on product quality variables.

Due to the above technical challenges, the new QNGLSP method is developed for non-Gaussian process monitoring with output quality variables incorporated. The basic idea of QNGLSP approach is to find the multidimensional latent directions in the measurement-variable and quality-variable spaces concurrently so that the maximized multidimensional mutual information between measurement and quality spaces is obtained. It should be noted that mutual information is a quantitative measure of statistical dependency between two random variables and can be estimated from information entropy. Compared to covariance, it is essentially high-order statistics and, thus, is able to extract the non-Gaussian process features.

Given an input matrix $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m]\in\Re^{n\times m}$ with n samples and m process measurement variables and an output matrix $\mathbf{Y}=[\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_k]\in\Re^{n\times k}$ with k quality variables, the data matrices are first normalized to zero-mean and unit-variance and then decomposed onto low-dimensional subspaces as follows

$$\mathbf{X}=\mathbf{S}\mathbf{P}^T+\mathbf{E} \tag{18}$$

$$\mathbf{Y}=\mathbf{S}\mathbf{Q}^T+\mathbf{F} \tag{19}$$

where $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_d]\in\Re^{n\times d}$ denotes the score matrix, $\mathbf{P}=[\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_d]\in\Re^{m\times d}$ is the loading matrix for $\mathbf{X}$, $\mathbf{Q}=[\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_d]\in\Re^{k\times d}$ is the loading matrix for $\mathbf{Y}$, $\mathbf{E}\in\Re^{n\times m}$ is the residual matrix for $\mathbf{X}$, $\mathbf{F}\in\Re^{n\times k}$ is the residual matrix for $\mathbf{Y}$, and d is the number of latent variables. The initial objective in the proposed QNGLSP algorithm is to find weighting vectors $\mathbf{w}$ and $\mathbf{c}$ from the deflated $\mathbf{X}$ and $\mathbf{Y}$ for each pair of score vectors through the following constrained optimization problem

$$\max I(\mathbf{s}_i,\mathbf{u}_i)=I(\mathbf{X}_i\mathbf{w}_i,\mathbf{Y}_i\mathbf{c}_i) \tag{20}$$

$$\text{subject to } ||\mathbf{w}||_i=1,\ ||\mathbf{c}||_i=1 \tag{21}$$

where $I(\mathbf{s}_i,\mathbf{u}_i)$ represents the mutual information between the score vectors $\mathbf{s}_i$ and $\mathbf{u}_i$. The mutual information $I(\mathbf{s}_i,\mathbf{u}_i)$ can be expressed as

$$\begin{aligned}I(\mathbf{s}_i,\mathbf{u}_i)&=H(\mathbf{u}_i)-H(\mathbf{u}_i|\mathbf{s}_i)\\&=H(\mathbf{s}_i)-H(\mathbf{s}_i|\mathbf{u}_i)\\&=H(\mathbf{s}_i,\mathbf{u}_i)-H(\mathbf{u}_i|\mathbf{s}_i)-H(\mathbf{s}_i|\mathbf{u}_i)\end{aligned} \tag{22}$$

where $H(\mathbf{u}_i)$ is the marginal entropy, $H(\mathbf{u}_i|\mathbf{s}_i)$ is the conditional entropy, and $H(\mathbf{s}_i,\mathbf{u}_i)$ is the joint entropy defined as

$$H(\mathbf{u}_i)=-\int_{\mathbf{u}_i}f(u)\log f(u)du \tag{23}$$

$$H(\mathbf{u}_i|\mathbf{s}_i)=-\int_{\mathbf{s}_i}\int_{\mathbf{u}_i}f(s,u)\log\frac{f(u|s)}{f(s,u)}dsdu \tag{24}$$

$$H(\mathbf{u}_i,\mathbf{s}_i)=-\int_{\mathbf{s}_i}\int_{\mathbf{u}_i}f(s,u)\log f(u,s)dsdu \tag{25}$$

The above complex integrals for mutual information are difficult to calculate analytically. Therefore, a numerical optimization method termed as Nelder–Mead algorithm is instead adopted to solve this problem through nonconstraint nonlinear heuristic optimization procedure.[54] It should be noted that the normalization step of $\mathbf{w}_i$ and $\mathbf{c}_i$ are added in the numerical iterations to handle the constraints in Eq. 21. Moreover, the objective function in the mutual information-based optimization problem may have strong nonlinearity and multipeak feature, which can potentially lead to local optimal solution instead of global optimum. To overcome this issue, the multistart optimization strategy is used.

After the extraction of the weighting vectors $\mathbf{w}_i$ and $\mathbf{c}_i$, the score vectors $\mathbf{s}_i$ and $\mathbf{u}_i$ can be computed as follows

$$\mathbf{s}_i=\mathbf{X}_i\mathbf{w}_i \tag{26}$$

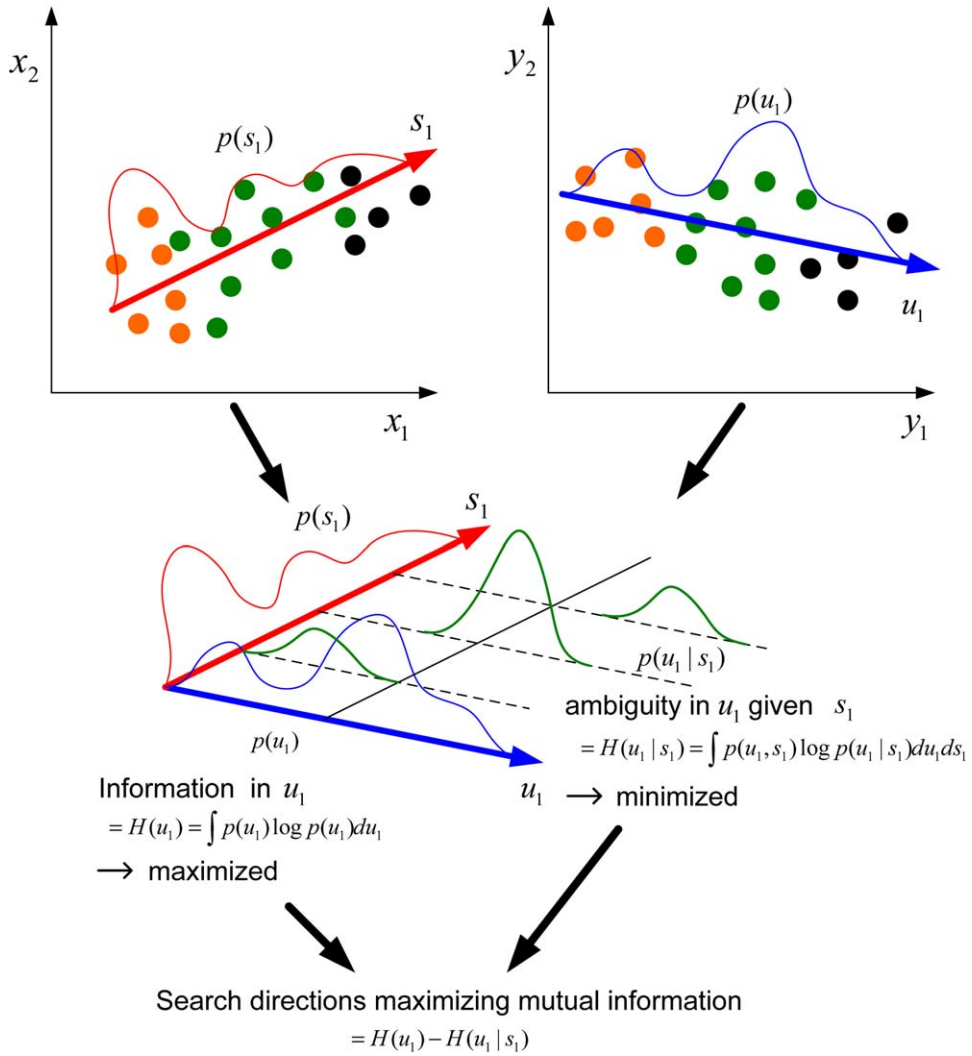

Figure 1. Illustration of the proposed QNGLSP method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\mathbf{u}_i = \mathbf{Y}_i \mathbf{c}_i \tag{27}$$

Then, loading vectors $\mathbf{p}_i$ and $\mathbf{q}_i$ are estimated as

$$\mathbf{p}_i = \mathbf{X}_i^T \mathbf{s}_i / \mathbf{s}_i^T \mathbf{s}_i \tag{28}$$

$$\mathbf{q}_i = \mathbf{Y}_i^T \mathbf{s}_i / \mathbf{s}_i^T \mathbf{s}_i \tag{29}$$

With the obtained loading vectors, the matrices $\mathbf{X}$ and $\mathbf{Y}$ can be deflated as follows

$$\mathbf{X}_{i+1} = \mathbf{X}_i - \mathbf{s}_i \mathbf{p}_i^T \text{ and } \mathbf{Y}_{i+1} = \mathbf{Y}_i - \mathbf{s}_i \mathbf{q}_i^T \tag{30}$$

However, $\mathbf{w}_i$ does not relate $\mathbf{s}_i$ to the original input data matrix $\mathbf{X}$ directly. Thus, a decomposition matrix $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_d]$ is defined below

$$\mathbf{r}_1 = \mathbf{w}_1 \tag{31}$$

and

$$\mathbf{r}_i = \prod_{j=1}^{i-1} \left(\mathbf{I}_m - \mathbf{w}_j \mathbf{p}_j^T \right) \mathbf{w}_i \; i \geq 2 \tag{32}$$

Then, score matrix $\mathbf{S}$ can be computed from original input matrix $\mathbf{X}$ as follows

$$\mathbf{S} = \mathbf{X}\mathbf{R} \tag{33}$$

The searching strategy of latent directions in the proposed QNGLSP method is illustrated in Figure 1. It can be observed that both the input and output data are projected onto the first feature directions within the input and output spaces to obtain the scores $\mathbf{s}_1$ and $\mathbf{u}_1$. Both directions are searched in such a way that the marginal entropy $H(\mathbf{u}_1)$ that equals the amount of information in $\mathbf{u}_1$ is maximized while the conditional entropy that equals the amount of ambiguity in $\mathbf{u}_1$ given $\mathbf{s}_1$ is minimized. Equivalently, the mutual information $I(\mathbf{s}_1; \mathbf{u}_1)$ between the score vectors $\mathbf{s}_1$ and $\mathbf{u}_1$ is maximized. The remaining score vectors can be estimated in the same fashion through iterative procedure.

After all the loading and score vectors are obtained, it is necessary to sort the two sets of latent directions corresponding to the input and output spaces, respectively. The marginal entropies of score vectors are used to rearrange the column vectors of the score and decomposition matrices $\mathbf{S}$ and $\mathbf{R}$. With all the sorted latent variables, the number of components $\mathbf{p}_i$ and $\mathbf{q}_i$ for concurrent subspace projections needs to be selected to achieve the best monitoring performance. If the numbers of components are too small, the projected subspaces

do not contain sufficient non-Gaussian features for quality relevant fault detection. On the contrary, if too many components are chosen, then the formed subspaces may include irrelevant or redundant information that can degrade the sensitivity of monitoring statistics to faults. In QNGLSP method, the numbers of latent variables d is chosen so that the first d column vectors of the full score and decomposition matrices $\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_m]$ and $\mathbf{R}=[\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_m]$ satisfy the following marginal entropy-based criteria

$$\frac{\sum_{i=1}^{d} H(\mathbf{s}_i)}{\sum_{i=1}^{m} H(\mathbf{s}_i)} > \epsilon \tag{34}$$

and

$$\frac{\sum_{i=1}^{d} H(\mathbf{r}_i)}{\sum_{i=1}^{m} H(\mathbf{r}_i)} > \epsilon \tag{35}$$

where ϵ is the predefined threshold value and set to 0.95 in this work. Thus, the loading matrices $\mathbf{P}$ and $\mathbf{Q}$, score matrix $\mathbf{S}$, and decomposition matrix $\mathbf{R}$ consisting of d latent variables can be extracted as

$$\mathbf{P}=[\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_d]\in\Re^{m\times d} \tag{36}$$

$$\mathbf{Q}=[\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_d]\in\Re^{k\times d} \tag{37}$$

$$\mathbf{R}=[\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_d]\in\Re^{m\times d} \tag{38}$$

$$\mathbf{S}=[\mathbf{s}_1,\mathbf{s}_2,\ldots,\mathbf{s}_d]\in\Re^{n\times d} \tag{39}$$

Given a new test sample vector $\mathbf{x}$, the corresponding score, prediction, and residual vectors are computed as follows

$$\text{Score}:\mathbf{s}=\mathbf{xR} \tag{40}$$

$$\text{Prediction}:\hat{\mathbf{x}}=\mathbf{xRP}^T \tag{41}$$

$$\text{Residual}:\mathbf{e}=\mathbf{x}\left(\mathbf{I}-\mathbf{RP}^T\right) \tag{42}$$

In multivariate statistical process monitoring, two types of statistics are widely used for fault detection. One is the D statistic for monitoring the systematic part of process variations, whereas the other is the Q statistic for monitoring the residual part of process variations. As described in "Review of PLS and ICA-Based Process Monitoring Methods," section PLS-based fault detection methods use T^2 and SPE indices, whereas ICA-based methods adopt I^2 and SPE statistics. In QNGLSP-based monitoring method, I^2 and SPE indices are proposed for quality relevant fault detection as follows

$$I^2=\mathbf{sL}^{-1}\mathbf{s}^T \tag{43}$$

$$\begin{aligned}\text{SPE}&=\mathbf{ee}^T\\&=\mathbf{x}\left(\mathbf{I}-\mathbf{RP}^T\right)\left(\mathbf{I}-\mathbf{PR}^T\right)\mathbf{x}^T\end{aligned} \tag{44}$$

where $\mathbf{L}$ is the diagonal matrix with the variances of different column vectors of $\mathbf{S}$ being the diagonal entries.

As it is assumed that the latent variables may follow non-Gaussian distribution, the control limits of the proposed indices are estimated from kernel density estimation strategy.[51] Let $(D_1,D_2,\ldots,D_n)$ be a set of observations from an unknown probability density function f. Then f can be estimated by kernel density estimator as follows

Table 1. Step-by-Step Procedure of the Proposed Quality Relevant Non-Gaussian Latent Subspace Projection Method

(1) Form $\mathbf{X}$ and $\mathbf{Y}$ by filling missing entries with zeros and then scale $\mathbf{X}$ and $\mathbf{Y}$ to zero mean and unit variance.
(2) Set counter $i \Leftarrow 1$
(3) Set $\mathbf{X}_i \Leftarrow \mathbf{X}$ and $\mathbf{Y}_i \Leftarrow \mathbf{Y}$
(4) Take random initial vectors $\mathbf{w}_i$ and $\mathbf{c}_i$ of unit norm
(5) Solve the following constrained nonlinear optimization problem (Details of optimization algorithm are given in Table 2)

$$\begin{aligned}&\text{maximize} && I(\mathbf{s}_i,\mathbf{u}_i)=I(\mathbf{X}_i\mathbf{w}_i,\mathbf{Y}_i\mathbf{c}_i)\\&\text{subject to} && ||\mathbf{w}||_i=1,||\mathbf{c}||_i=1\end{aligned}$$

(6) Calculate $\mathbf{s}_i$ and $\mathbf{u}_i$

$$\begin{aligned}\mathbf{s}_i&=\mathbf{X}_i\mathbf{w}_i\\\mathbf{u}_i&=\mathbf{Y}_i\mathbf{c}_i\end{aligned}$$

(7) Calculate $\mathbf{p}_i$ and $\mathbf{q}_i$

$$\begin{aligned}\mathbf{p}_i&=\mathbf{X}_i^T\mathbf{s}_i/\mathbf{s}_i^T\mathbf{s}_i\\\mathbf{q}_i&=\mathbf{Y}_i^T\mathbf{s}_i/\mathbf{s}_i^T\mathbf{s}_i\end{aligned}$$

(8) Residual deflation for the available entries only:

$$\begin{aligned}\mathbf{X}_{i+1}&\Leftarrow\mathbf{X}_i-\mathbf{s}_i\mathbf{p}_i^T\\\mathbf{Y}_{i+1}&\Leftarrow\mathbf{Y}_i-\mathbf{s}_i\mathbf{q}_i^T\end{aligned}$$

(9) Set i $\Leftarrow i+1$ and return to (4) until $i=i_{\max}$
(10) Calculate the decomposition vector $\mathbf{r}_i$

$$\mathbf{r}_i=\mathbf{w}_1(i=1)$$

$$\mathbf{r}_i=\prod_{j=1}^{i-1}\left(\mathbf{I}_m-\mathbf{w}_j\mathbf{p}_j^T\right)\mathbf{w}_i\quad(i\geq 2)$$

$$\hat{f}(D)=\frac{1}{nh}\sum_{i=1}^{n}K\left\{\frac{D-D_i}{h}\right\} \tag{45}$$

where D represents the I^2 or SPE index, h is the kernel window width, and K denotes the Gaussian kernel function

$$K(u)=\frac{1}{\sqrt{2\pi}}e^{\left(-\frac{1}{2}u^2\right)} \tag{46}$$

with u being an arbitrary data point. After a probability density function is estimated, the corresponding point with cumulative density function value at $1-\alpha$ is the control limit under the confidence level of $(1-\alpha)\times 100\%$.

The step-by-step numerical procedure of the proposed QNGLSP method is listed in Table 1, whereas the constrained nonlinear optimization algorithm for maximizing the mutual information between input and output latent variables is shown in Table 2.

Application Example

Tennessee *Eastman chemical process*

In this study, the Tennessee Eastman Chemical process is used to examine the effectiveness of the proposed quality

Table 2. Step-by-Step Procedure of Constrained Nonlinear Optimization Algorithm for Searching Non-Gaussian Latent Directions

(1)	Set multi-loop counter $l \Leftarrow 1$
(2)	Make a simplex which is a special polytope of $m+k+1$ vertices corresponding to
(3)	$\begin{bmatrix} \mathbf{w}_i^{(1)(l)} & \mathbf{w}_i^{(2)(l)} & \dots & \mathbf{w}_i^{(v)(l)} & \dots & \mathbf{w}_i^{(m+k+1)(l)} \\ \mathbf{c}_i^{(1)(l)} & \mathbf{c}_i^{(2)(l)} & \dots & \mathbf{c}_i^{(v)(l)} & \dots & \mathbf{c}_i^{(m+k+1)(l)} \end{bmatrix}$
	Normalize $\mathbf{w}_i^{(v)(l)}$ and $\mathbf{c}_i^{(v)(l)}$
(3)	$\mathbf{w}_i^{(v)(l)} = \mathbf{w}_i^{(v)(l)} / \lvert\lvert \mathbf{w}_i^{(v)(l)} \rvert\rvert \quad \forall v$
(3)	$\mathbf{c}_i^{(v)(l)} = \mathbf{c}_i^{(v)(l)} / \lvert\lvert \mathbf{c}_i^{(v)(l)} \rvert\rvert \quad \forall v$
(4)	Determine the updated vertex $v\prime$ and its steps $\Delta\mathbf{w}$, $\Delta\mathbf{c}$ by the Nelder–Mead method
	Update $\mathbf{w}_i^{(v\prime)(l)}$ and $\mathbf{c}_i^{(v\prime)(l)}$
(5)	$\mathbf{w}_i^{(v\prime)(l)} = \mathbf{w}_i^{(v\prime)(l)} + \Delta\mathbf{w}$
(5)	$\mathbf{c}_i^{(v\prime)(l)} = \mathbf{c}_i^{(v\prime)(l)} + \Delta\mathbf{c}$
	Return to (3) until $\mathbf{w}_i^{(v)(l)}$ and $\mathbf{c}_i^{(v)(l)}$ are converged. If converged, set $\mathbf{w}_i^{(l)} = \mathbf{w}_i^{(v)(l)}$, $\mathbf{c}_i^{(l)} = \mathbf{c}_i^{(v)(l)}$ and go to (6)
(6)	Set $l \Leftarrow l + 1$ and return to (2) until $l = l_{\max}$
(7)	Choose the optimal $\mathbf{w}_i$ and $\mathbf{c}_i$ as follows:
	$l_{\text{opt}} = \text{argmax}_l I\left(\mathbf{X}_i\mathbf{w}_i^{(l)}, \mathbf{Y}_i\mathbf{c}_i^{(l)}\right)$
	$\mathbf{w}_i = \mathbf{w}^{l_{\text{opt}}}$
	$\mathbf{c}_i = \mathbf{c}^{l_{\text{opt}}}$

relevant non-Gaussian process monitoring method. The process flow diagram of the Tennessee Eastman Chemical process is shown in Figure 2 and this process includes five major unit operations, which are a chemical reactor, a product condenser, a vapor-liquid separator, a recycle compressor, and a product stripper.[55] This process produces two liquid products, G and H, along with a byproduct of F from four gaseous reactants A, C, D, and E. An inert B is fed into the chemical reactor where G and H are formed. There are total 41 measurement variables and 12 manipulated variables

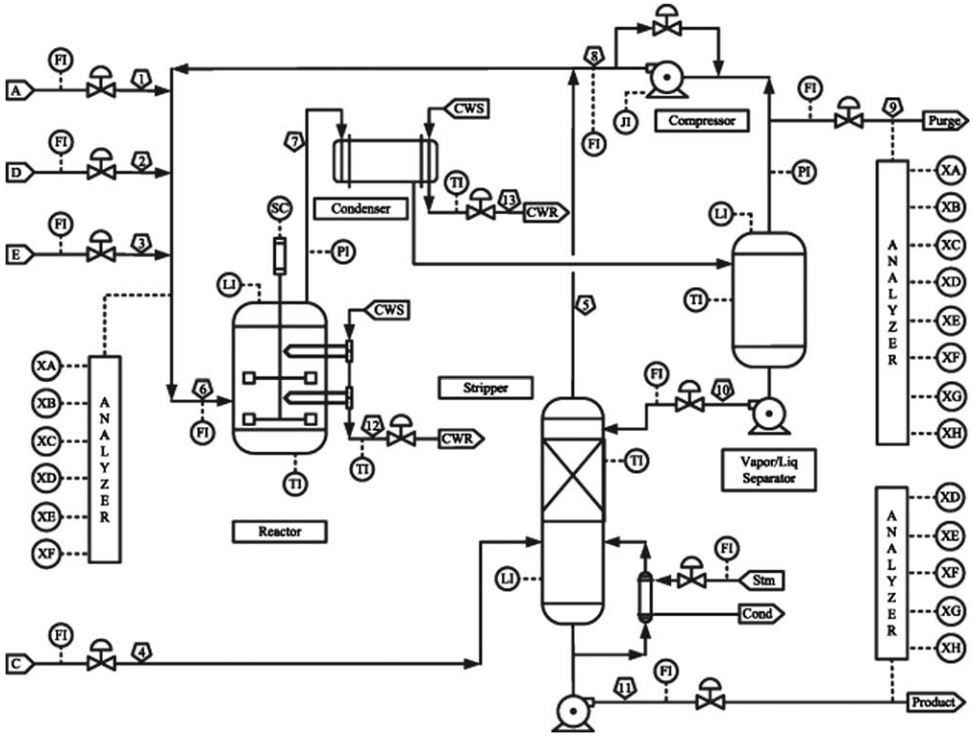

Figure 2. Process flow diagram of the Tennessee Eastman Chemical process.

Table 3. Input Variables of the Tennessee Eastman Chemical Process

Variable No.	Variable Description
1	A Feed (stream 1)
2	D feed (stream 2)
3	E feed (stream 3)
4	A and C feed (stream 4)
5	Recycle flow (stream 8)
6	Reactor feed RATE (stream 6)
7	Reactor pressure
8	Reactor level
9	Reactor temperature
10	Purge rate (stream 9)
11	Product Sep Temp
12	Product Sep level
13	Product Sep pressure
14	Product Sep underflow (stream 10)
15	Stripper level
16	Stripper pressure
17	Stripper underflow (stream 11)
18	Stripper Temp
19	Stripper steam flow
20	Compressor work
21	Reactor coolant Temp
22	Separator coolant Temp
23	D feed flow (stream 2)
24	E feed flow (stream 3)
25	A feed flow (stream 1)
26	A and C feed flow (stream 4)
27	Purge value (stream 9)
28	Separator pot liquid flow (stream 10)
29	Stripper liquid product flow (stream 11)
30	Reactor cooling water flow
31	Condenser cooling water flow

in the process. Moreover, there are 20 predefined abnormal operating events and six different operating conditions in the Tennessee Eastman Chemical process, as shown in Tables 5 and 6. The process involves a plant-wide decentralized control implementation with different feedback control loops.[56]

For process monitoring purpose, 22 continuous measurement variables and nine manipulated variables are selected as input variables, which are listed in Table 3. Meanwhile, as listed in Table 4, 19 composition variables that are measured through either off-line lab analysis or on-line analyzers are used as output quality variables in the process monitoring framework. The sampling time of both input and output variables are set to 0.25 h. Two subsets of training data with 1440 normal samples in each set are generated from operating modes 1 and 3, respectively. Then the combined normal data set of 2880 samples is used to build the QNGLSP model for process monitoring and fault detection. In this work, the normalized multivariate kurtosis of process data is used to quantitatively measure process non-Gaussianity. The value of normalized multivariate kurtosis of the training data set generated from different operating modes is 747, which is significantly larger than zero. Therefore, it can be inferred that the process data generated from two different operating modes follow a non-Gaussian probability distribution.

Furthermore, three test cases containing various types of predefined process faults are designed to compare the monitoring performance of the PLS and the proposed QNGLSP methods. It should be noted that the other existing techniques such as PCA, ICA, GMM, and supervised classification methods are not chosen for methodology comparison in the application example because all these monitoring methods do not include product quality variables as output variables in model development and data analysis while PLS and QNGLSP approaches take into account both process measurement and product quality variables concurrently. The detailed test scenarios are shown in Table 7 and all these

Table 4. Output Quality Variables of the Tennessee Eastman Chemical Process

Variable No.	Variable Description
1	Component to A to reactor
2	Component to B to reactor
3	Component to C to reactor
4	Component to D to reactor
5	Component to E to reactor
6	Component to F to reactor
7	Component A in purge
8	Component B in purge
9	Component C in purge
10	Component D in purge
11	Component E in purge
12	Component F in purge
13	Component G in purge
14	Component H in purge
15	Component D in product
16	Component E in product
17	Component F in product
18	Component G in product
19	Component H in product

Table 5. Predefined Faults of the Tennessee Eastman Chemical Process

Fault ID.	Fault Description
IDV(1)	Step in A/C feed ratio, B composition constant
IDV(2)	Step in B composition, A/C ratio constant
IDV(3)	Step in D feed temperature (stream 2)
IDV(4)	Step in reactor cooling water inlet temperature
IDV(5)	Step in condenser cooling water inlet temperature
IDV(6)	A feed loss (step change in stream 1)
IDV(7)	C header pressure loss (step change in stream 4)
IDV(8)	Random variation in A+C feed composition (stream 4)
IDV(9)	Random variation in D feed temperature (stream 2)
IDV(10)	Random variation in C feed temperature (stream 4)
IDV(11)	Random variation in reactor cooling water inlet temperature
IDV(12)	Random variation in condenser cooling water inlet temperature
IDV(13)	Slow drift in reaction kinetics
IDV(14)	Sticking reactor cooling water valve
IDV(15)	Sticking condenser cooling water valve
IDV(16)	Unknown disturbance
IDV(17)	Unknown disturbance
IDV(18)	Unknown disturbance
IDV(19)	Unknown disturbance
IDV(20)	Unknown disturbance

Table 6. Six Operation Modes of the Tennessee Eastman Chemical Process

Operating Mode	G/H Mass Ratio	Production Rate (stream 11)
1	50/50	7038 kg/h G and 7038 kg/h H
2	10/90	1408 kg/h G and 12669 kg/h H
3	90/10	10,000 kg/h G and 1111 kg/h H
4	50/50	Maximum
5	10/90	Maximum
6	90/10	Maximum

Table 7. Three Test Cases of the Tennessee Eastman Chemical Process

Case No.	Test Scenario
1	Normal operation from the 1-st to 40-th samples C header pressure loss from the 41-st to 100-th samples Normal operation from the 101-st to 160-th samples Random variation in condenser cooling water inlet temperature from the 161-st to 200-th samples
2	Normal operation from the 1-st to 55-th samples Step change in condenser cooling water inlet temperature from the 56-th to 100-th samples Normal operation from the 101-st to 130-th samples Sticking reactor cooling water valve from the 131-st to 200-th samples
3	Normal operation from the 1-st to 70-th samples C header pressure loss from the 71-st to 100-th samples Normal operation from the 101-st to 140-th samples Random variation in condenser cooling water inlet temperature from the 141-st to 200-th samples Normal operation from the 201-st to 250-th samples Sticking valve of reactor cooling water flow from the 251-st to 300-th samples

cases include both normal and different type of faulty operations. In the first test case, the process begins with normal operating condition from the first through the 40-th samples. Then the fault of C header pressure loss occurs at the 41-st sample and remains until the 100-th sample, after which the process returns to normal operation. From the 161-st sample, the fault of increased random variation in condenser cooling water inlet temperature takes place with the duration of 40 samples. For the second test scenario, the process fault of step change in condenser cooling water inlet temperature happens after the initial period of normal operation and lasts 45 samples. Then the process operation is back to normal

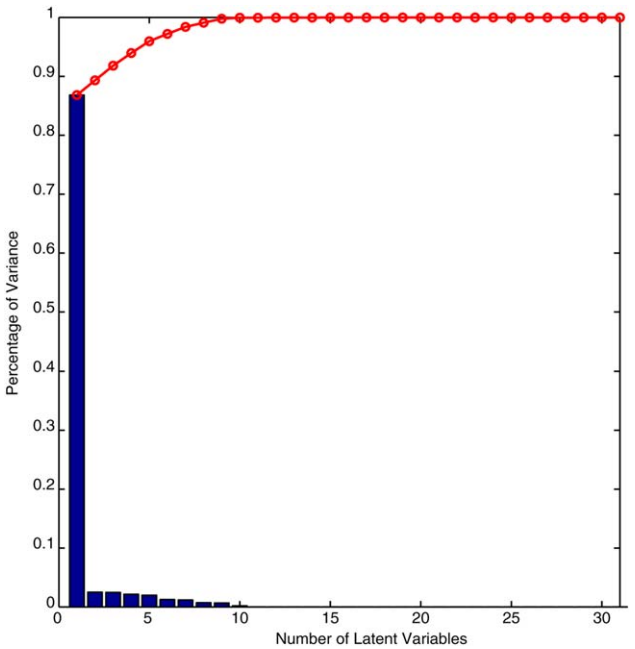

Figure 3. Trend plot of the individual percentage (bar) and cumulative percentage (solid line) of variance of PLS-based latent variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

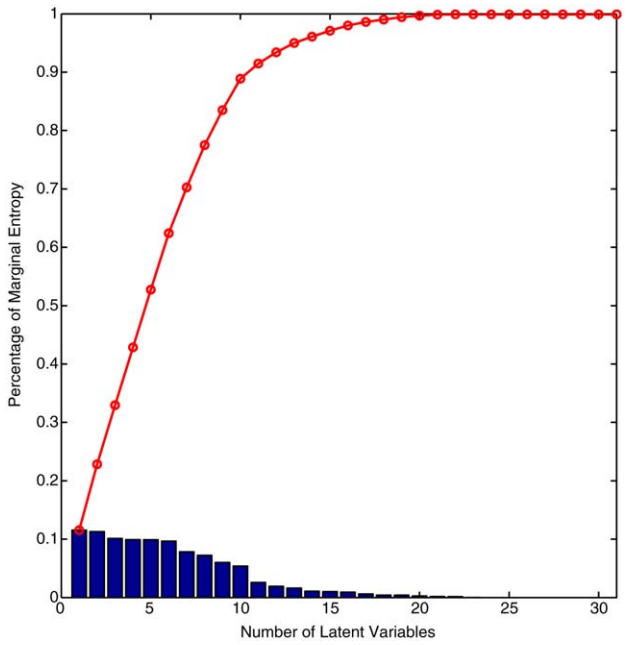

Figure 4. Trend plot of the individual percentage (bar) and cumulative percentage (solid line) of marginal entropy of QNGLSP-based latent variables.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

condition and remains for 30 samples before a new fault of sticking valve of reactor cooling water flow occurs. In the more complex third test case, the faulty operation of C header pressure loss happens from the 71-st through 100-th samples. Then the operational status returns to normal before the second fault of increased random variations of condenser cooling water inlet temperature takes place from the 141-st through 200-th samples. After the process is operated under normal condition for another 50 samples, the third fault of sticking valve of reactor cooling water flow occurs between the 251-st and 300-th samples. The proposed I^2 and SPE indices in the QNGLSP method are compared to the T^2 and SPE statistics of the PLS method for fault detection in the three test cases.

Comparison of *process monitoring results*

After the QNGLSP model is built from normal training data with both input and output variables, it is important to select the number of non-Gaussian latent variables for input and output subspace projections. In the proposed approach, marginal entropy-based strategy is utilized to determine the best number of latent directions to be retained. For the training set, the individual and cumulative percentages of the marginal entropy of the sorted latent variables vs. the number of variables are shown Figure 4. Based on the proposed selection criteria, total 13 latent variables that contain over 95% of the marginal entropy is chosen. As shown in Figure 3, total 5 variables that cover over 95% of the variance are selected in PLS model. Figures 5 and 6 show the scatter plots of process data along the first vs. the second and the first vs. the third latent variables in the PLS and QNGLSP methods. It can be readily observed that the process data do not follow Gaussian distribution. In addition, as the training

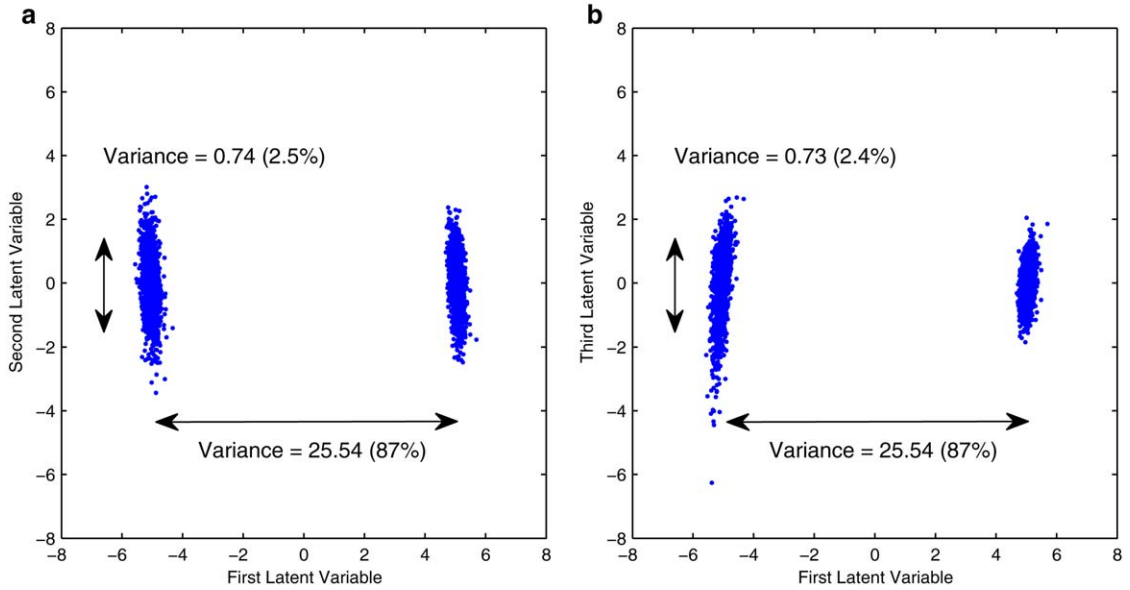

Figure 5. Scatter plots of process data along (a) the first vs. the second and (b) the first vs. the third latent variables of the PLS method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

data are generated from multimode process operation, most of the variations of the training data are due to different operating modes and captured by the first latent variable. Therefore, the variances of the remaining latent variables become significantly smaller than that of the first latent variable, as shown in Figures 3 and 5.

In the first test case, the normal operation of the plant is mixed with two different types of faults, which are C header pressure loss from the 41-st to 100-th samples and increased random variation in condenser cooling water inlet temperature from the 161-st to 200-th samples. The process monitoring results of the proposed QNGLSP method including I^2 and SPE indices and the PLS method including T^2 and SPE statistics are compared in Figure 7. Meanwhile, the fault detection rates and false alarms rates of different indices in the two methods are listed in Tables 8 and 9. It can be seen that the QNGLSP-based I^2 and SPE indices detect abnormal operating events with fairly high fault detection rates of over 99.0% while low false alarm rates of only 1.25% and 0.50%, respectively. In contrast, the PLS-based monitoring method does not result in satisfactory performance as the T^2 index can capture only 39.75% of faulty samples, although the

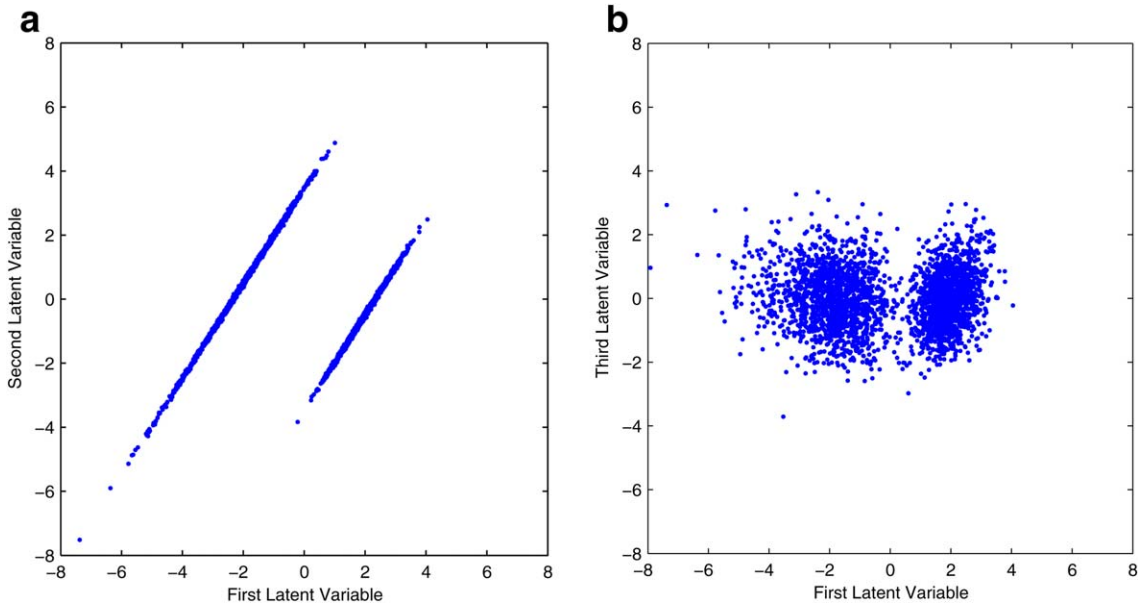

Figure 6. Scatter plots of process data along (a) the first vs. the second and (b) the first vs. the third latent variables of the proposed QNGLSP method.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

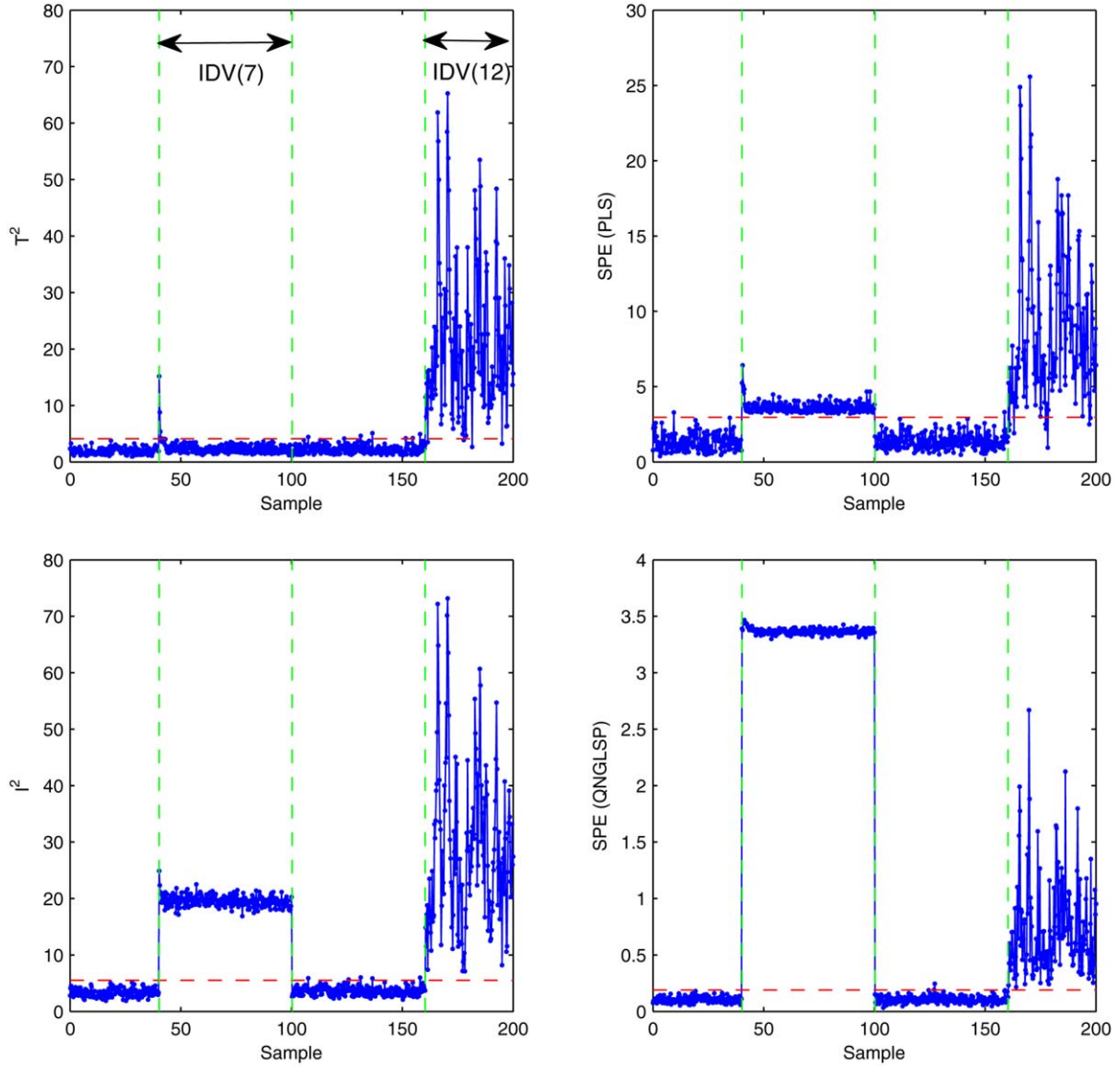

Figure 7. Monitoring results of PLS and QNGLSP methods in the first test case of the Tennessee Eastman Chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

SPE index leads to 97.75% fault detection rate. Such results reveal that the proposed QNGLSP method can well extract non-Gaussian process features from measurement and quality variables and, thus, the proposed I^2 and SPE indices are able to detect process faults accurately with minimum number of false alarms triggered. The specific comparison between PLS and QNGLSP monitoring results shows that the I^2 index has much higher fault detection accuracy than the T^2 index while the QNGLSP-based SPE statistic has a litter better fault detection rate than the PLS-based SPE index. As for the overall fault detection rates, both PLS and QNGLSP methods lead to the detection of over 99% of the abnormal operations with either T^2 or SPE index exceeds the corresponding control limit.

Table 8. Comparison of Fault Detection Rates (%) of Three Test Cases in the Tennessee Eastman Chemical Process

Method	Fault Detection Rate (%)					
	PLS				QNGLSP	
Number of Latent Variables	5 (95% of Variance)		8 (99% of Variance)		13 (95% of Entropy)	
Statistics	T^2	SPE	T^2	SPE	I^2	SPE
Case 1	39.75	97.75	41.25	87.50	99.75	99.75
Case 2	85.00	64.13	89.13	45.65	98.91	98.26
Case 3	64.64	83.57	71.79	78.39	98.57	96.79
Average	63.13	81.82	67.39	70.51	99.08	98.27

Table 9. Comparison of False Alarm Rates (%) of Three Test Cases in the Tennessee Eastman Chemical Process

Method	Fault Detection Rate (%)					
	PLS				QNGLSP	
Number of Latent Variables	5 (95% of Variance)		8 (99% of Variance)		13 (95% of Entropy)	
Statistics	T^2	SPE	T^2	SPE	I^2	SPE
Case 1	1.75	0.50	1.00	0.75	1.25	0.50
Case 2	0.59	0.59	0.29	1.17	1.17	0.59
Case 3	1.09	0.47	0.94	0.62	1.25	1.09
Average	1.14	0.52	0.74	0.85	1.01	0.73

In the second test case, the plant operation includes normal condition along with two different types of faults, which are a step change in condenser cooling water inlet temperature from the 56-th to 100-th samples and increased random variation in sticking valve of reactor cooling water flow from the 131-st to 200-th samples. The process monitoring results of PLS and QNGLSP methods are shown in Figure 8, and Tables 8 and 9. It can be seen that the QNGLSP-based I^2 and SPE indices are able to accurately alarm the faulty operations with the high fault detection rates of 98.91 and 98.26%, respectively. In comparison, the PLS-based T^2 and SPE statistics lead to the fault detection rates of only 85.00 and 64.13%, respectively. Therefore, the QNGLSP method has significantly stronger capability to capture different types of process faults than the PLS method. The main reason of the superior performance of QNGLSP method is due to its

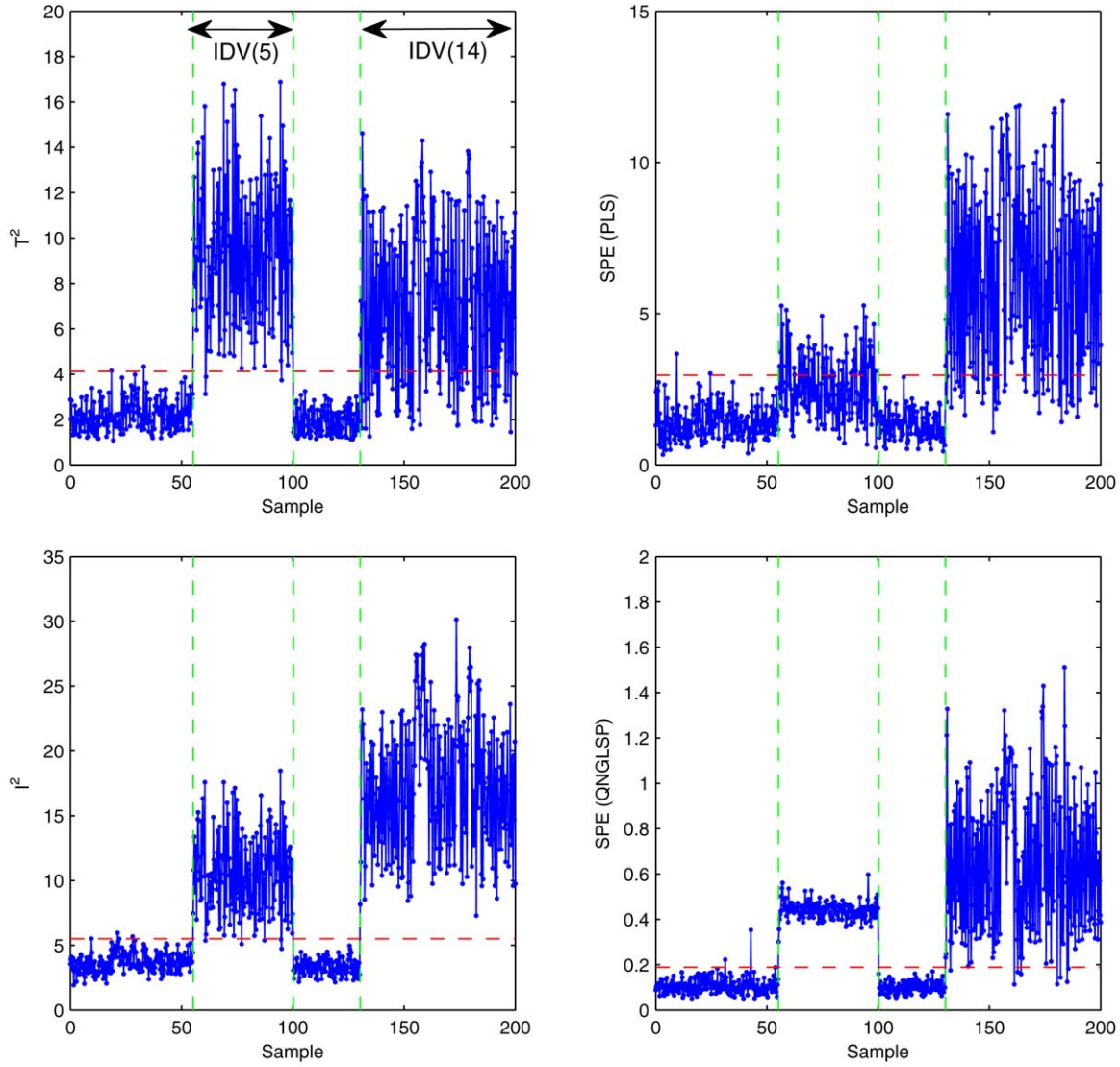

Figure 8. Monitoring results of PLS and QNGLSP methods in the second test case of the Tennessee Eastman Chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

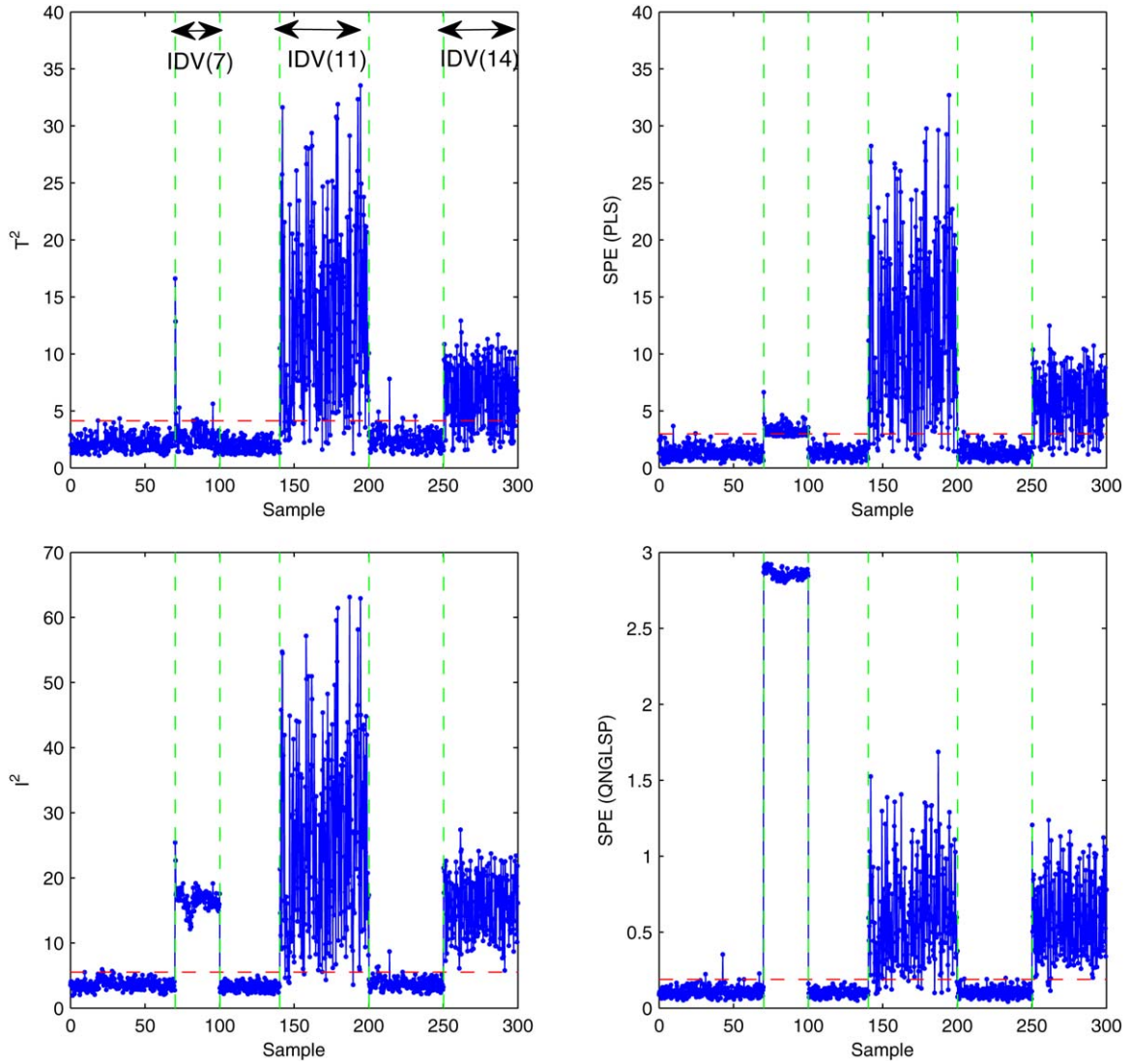

Figure 9. Monitoring results of PLS and QNGLSP methods in the third test case of the Tennessee Eastman Chemical process.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

mutual information rather than covariance-based objective function in searching for the optimal latent directions. In this way, the obtained latent subspace can efficiently extract and retain the non-Gaussian features for enhanced fault detection capacity. It should be noted that, compared to the first test case, the PLS-based T^2 statistic is more sensitive but the SPE index is less sensitive to the fault in this case. It implies that this fault affects the systematic part of process variations more than the residual part in the PLS model while it affects both the systematic and residual parts of process variations in the proposed QNGLSP model. In addition, the overall fault detection rate of the proposed QNGLSP method reaches 100.00% as either I^2 or SPE index exceeds its corresponding control limit, whereas the overall detection rate of PLS method is only 89.78%.

The last test case includes a complex operating scenario with three types of process faults that are mixed into normal operation. The abnormal operating events are C header pressure loss from the 71-st to the 100-th samples, increased random variation in condenser cooling water inlet temperature from the 141-st to 200-th samples and sticking valve of reactor cooling water flow from the 251-st to 300-th samples. The fault detection results of the various statistics from QNGLSP and PLS methods are shown in Figure 9. Meanwhile, the quantitative comparison of fault detection rates and false alarm rates are given in Tables 8 and 9. One can readily observe that the fault detection rate of T^2 statistic of PLS method is only 64.64%, which is lower than that of the QNGLSP-based I^2 index (98.57%). Meanwhile, the PLS-based SPE index also yields lower fault detection rate (83.57%) than that of the QNGLSP-based SPE statistic (96.79%). These comparisons verify that the proposed QNGLSP method has better monitoring performance and fault detection capability than the conventional PLS method. Furthermore, the overall fault detection rate of the PLS method is only 85.36%, whereas the proposed QNGLSP method leads to the overall detection rate of 99.11%, where either I^2 or SPE statistic exceeds its corresponding control limit.

In the above comparison, five latent variables that contain over 95% of the variance are selected in the PLS model. To investigate the effectiveness of an increase of the number of

latent variables in PLS model, eight latent variables that contain over 99% of the variance are also selected and examined. The quantitative comparison of fault detection and false alarm rates are shown in Tables 8 and 9. It can be seen that the fault detection rates of T^2 and SPE statistics of PLS model that contains over 99% variance are still worse than those of the proposed QNGLSP method in all test cases.

Conclusions

In this article, a new QNGLSP method is proposed for chemical process monitoring and fault detection by taking into account both process measurement and quality variables. To capture the non-Gaussian features and relationships between input and output variables, the proposed QNGLSP method adopts the high-order statistics of mutual information instead of the second-order statistics of covariance for searching the latent directions within input and output spaces, respectively. Then, the multistart optimization procedure is designed to identify the optimal feature directions iteratively with nonlinear multipeak function handling capability. Furthermore, I^2 and SPE indices are developed to detect process faults within non-Gaussian latent variable and residual subspaces. Different from the PCA or ICA-based monitoring techniques, the presented QNGLSP method has the inherent model structure of combining process measurement and quality variables. Meanwhile, this new approach relies on mutual information-based objective function and, thus, can effectively extract the non-Gaussian features in latent subspaces, which cannot be achieved in the PLS-based monitoring method.

The proposed QNGLSP method is compared to the conventional PLS method in the three test cases of the Tennessee Eastman Chemical process with different operating modes. The monitoring results demonstrate that the QNGLSP-based I^2 and SPE indices are superior to the PLS-based T^2 and SPE indices in terms of more accurate fault detection. Future research will focus on extending the new QNGLSP approach for nonlinear batch or semibatch processes monitoring as well as taking into account the dynamic nature of process data.

Literature Cited

1. Piovoso M, Hoo K. Multivariate statistics for process control. *IEEE Control Syst Mag*. 2002;22:8–9.
2. Venkatasubramanian V, Rengaswamy R, Yin K, Kavuri SN. A review of process fault detection and diagnosis: part I: quantitative model-based methods. *Comput Chem Eng*. 2003;27:293–311.
3. Yu J. Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach. *Chem Eng Sci*. 2012; 82:22–30.
4. Gertler J. Survey of model-based failure detection and isolation in complex plants. *IEEE Control Syst Mag*. 1988;12:3–11.
5. Pons M, Rajab A, Flaus J, Engasser J, Cheruy A. Comparison of estimation methods for biotechnological processes. *Chem Eng Sci*. 1988;8:1909–1914.
6. Bastin G, Dochain D. On-Line Estimation and Adaptive Control of Bioreactors. Amsterdam, Netherlands: Elsevier, 1990.
7. Doyle F. Nonlinear inferential control for process applications. *J Process Control*. 1998;8:339–353.
8. MacGregor JF, Kourti T. Statistical process control of multivariate processes. *Control Eng Practice*. 1995;3(3):403–414.
9. Nomikos P, MacGregor JF. Monitoring of batch processes using multi-way principal component analysis. *AIChE J*. 1994;40:1361–1375.
10. MacGregor JF, Jaeckle C, Kiparissides C, Koutoudi M. Process monitoring and diagnosis by multiblock PLS methods. *AIChE J*. 1994;40(5):826–838.
11. Kano M. Comparison of statistical process monitoring methods: application to the Eastman challenge problem. *Comput Chem Eng*. 2000;24:175–181.
12. Chiang L, Russell E, Braatz R. Fault Detection and Diagnosis in Industrial Systems. Advanced Textbooks in Control and Signal Processing. London, Great Britain: Springer-Verlag, 2001.
13. Ündey C, Ertunç S, Çinar A. Online batch/fed-batch process performance monitoring, quality prediction, and variable-contribution analysis for diagnosis. *Ind Eng Chem Res*. 2003;42:4645–4658.
14. Qin SJ. Statistical process monitoring: basics and beyond. *J Chemom*. 2003;17:480–502.
15. Cinar A, Palazoglu A, Kayihan F. *Chemical Process Performance Evaluation*. Boca Raton: CRC Press, 2007.
16. Mejdell T, Skogestad S. Estimation of distillation compositions from multiple temperature measurements using partial least squares regression. *Ind Eng Chem Res*. 1991;30(12):2543–2555.
17. Kresta J, Marlin T, MacGregor J. Development of inferential process models using PLS. *Comput Chem Eng*. 1994;18:597–611.
18. Kosanovich K, Dahl K, Piovoso M. Improved process understanding using multiway principal component analysis. *Ind Eng Chem Res*. 1996;35:138–146.
19. Kano M, Miyazaki K, Hasebe S, Hashimoto I. Inferential control system of distillation compositions using dynamic partial least squares regression. *J Process Control*. 2000;10:157–166.
20. Lennox B, Montague G, Hiden H, Kornfeld G, Goulding P. Process monitoring of an industrial fed-batch fermentation. *Biotechnol Bioeng*. 2001;74:125–135.
21. Zhang H, Lennox B. Integrated condition monitoring and control of fed-batch fermentation processes. *J Process Control*. 2004;14:41–50.
22. Zamprogna E, Barolo M, Seborg D. Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. *J Process Control*. 2005;15:39–52.
23. Zhao SJ, Zhang J, Xu YM. Performance monitoring of processes with multiple operating modes through multiple PLS models. *J Process Control*. 2006;16:763–772.
24. Lin B, Recke B, Knudsen J, Jorgensen S. A systematic approach for soft sensor development. *Comput Chem Eng*. 2007;31:419–425.
25. Ku W, Storer R, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom Intell Lab Syst*. 1995;30:179–196.
26. Bakshi BR. Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J*. 1998;44(7):1596–1610.
27. Zhou D, Li G, Qin S. Total projection to latent structures for process monitoring. *AIChE J*. 2010;56(1):168–178.
28. Martin EB, Morris AJ. Non-parametric confidence bounds for process performance monitoring charts. *J Process Control*. 1996;6:349–358.
29. Hwang DH, Han C. Real-time monitoring for a process with multiple operating modes. *Control Eng Practice*. 1999;7:891–902.
30. Kano M, Nagao K, Hasebe S, Hashimoto I, Ohno H. Statistical process monitoring based on dissimilarity of process data. *AIChE J*. 2002;48(6):1231–1240.
31. Albazzaz H, Wang XZ. Statistical process control charts for batch operations based on independent component analysis. *Ind Eng Chem Res*. 2004;43:6731–6741.
32. Yan W, Shao H, Wang X. Soft sensing modeling based on support vector machine and Bayesian model selection. *Comput Chem Eng*. 2004;28:1489–1498.
33. Chen A, Song Z, Li P. Soft sensor modeling based on DICA-SVR. *Lect Notes Comput Sci*. 2005;3644:868–877.
34. Desai K, Badhe Y, Tambe S, Kulkarni B. Soft-sensor development for fed-batch bioreactors using support vector regression. *Biochem Eng J*. 2006;27(3):225–239.
35. Ge Z, Song Z. Process monitoring based on independent component analysis-principal component analysis (ICA-PCA) and similarity factors. *Ind Eng Chem Res*. 2007;46:2054–2063.
36. Kaneko H, Arakawa M, Funatsu K. Development of a new soft sensor method using independent component analysis and partial least squares. *AIChE J*. 2009;55:87–98.
37. Yu J. A nonlinear kernel Gaussian mixture model based inferential monitoring approach for fault detection and diagnosis of chemical processes. *Chem Eng Sci*. 2012;68:506–519.

38. Lee JM, Yoo CK, Lee IB. Statistical process monitoring with independent component analysis. *J Process Control*. 2004;14:467–485.
39. Rashid M, Yu J. A new dissimilarity method integrating multidimensional mutual information and independent component analysis for non-Gaussian dynamic process monitoring. *Chemom Intell Lab Syst*. 2012;115:44–58.
40. Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J*. 2008;54:1811–1829.
41. Yu J. A particle filter driven dynamic Gaussian mixture model approach for complex process monitoring and fault diagnosis. *J Process Control*. 2012;22(4):778–788.
42. Zhu Z, Song Z, Palazoglu A. Process pattern construction and multimode monitoring. *J Process Control*. 2012;22:247–262.
43. Chiang L. Fault diagnosis based on Fisher discriminant analysis and support vector machines. *Comput Chem Eng*. 2004;28(8):1389–1401.
44. Yu J. A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. *Comput Chem Eng*. 2012;41:134–144.
45. Yu J. Localized Fisher discriminant analysis based complex process monitoring. *AIChE J*. 2011;57:1817–1828.
46. Yu J. A support vector clustering-based probabilistic method for unsupervised fault detection and classification of complex chemical processes using unlabeled data. *AIChE J*. 2012;59(2):407–419.
47. Geladi P, Kowalski BR. Partial least-squares regresion: a tutorial. *Anal Chim Acta*. 1986;185:1–17.
48. Hyvarinen A, Karhunen J, Oja E. *Independent Component Analysis*. New York: John Wiley & Sons, 2001.
49. Hyvrinen A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Trans Neural Netw*. 1999;10(3):626–634.
50. Hyvrinen A. Independent component analysis: algorithms and applications. *IEEE Trans Neural Netw*. 2000;13:411–430.
51. Bishop C. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006.
52. Hoskuldsson A. PLS regression methods. *J Chemom*. 1988;2:211–228.
53. Hulle MV. Edgeworth approximation of multivariate differential entropy. *Neural Comput*. 2005;17(9):1903–1910.
54. Nelder A, Mead R. A simplex method for function minimization. *Comput J*. 1965;7(4):308–313.
55. Downs JJ, Vogel EF. Plant-wide industrial process control problem. *Comput Chem Eng*. 1993;17:245–255.
56. Ricker NL. Decentralized control of the Tennessee Eastman challenge process. *J Process Control*. 1996;6:205–221.

Manuscript received Aug. 15, 2013, and revision received Sept. 15, 2013.