I&EC research
Industrial & Engineering Chemistry Research
Article
Cite This: Ind. Eng. Chem. Res. 2019, 58, 11213−11223
pubs.acs.org/IECR

Supervised Diagnosis of Quality and Process Faults with Canonical Correlation Analysis

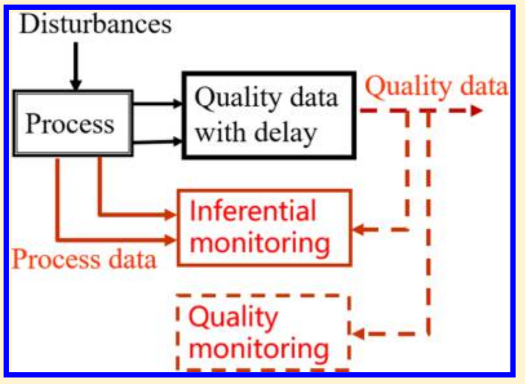
Qinqin Zhu[†] and S. Joe Qin[*,†,‡,§]

[†]Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, California 90089, United States

[‡]Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, United States

[§]School of Science and Engineering, The Chinese University of Hong Kong, Shenzhen, Guangdong 518172, China

ABSTRACT: Concurrent monitoring schemes that achieve simultaneous process and quality-relevant monitoring have recently attracted much attention. In this Article, we formulate a *supervised fault diagnosis* framework based on canonical correlation analysis (CCA) with regularization, which includes quality-relevant and quality-irrelevant fault diagnosis. Monitoring indices based on regularized concurrent CCA models are introduced to perform quality-relevant, potentially quality-relevant, and quality-irrelevant monitoring. Additionally, contribution plots and generalized reconstruction-based contribution methods are developed, along with their implications for the diagnosis based on the various monitoring indices. Finally, the Tennessee Eastman process is used to illustrate the supervised monitoring and diagnosis of quality-relevant and quality-irrelevant disturbances, and the 15 known disturbances are classified into two categories based on whether they have an impact on product quality variables.



1. INTRODUCTION

Process monitoring and diagnosis as a major approach of process data analytics has been one of the most active research areas in process systems engineering (PSE) over the past three decades.[1−4] With the availability of high dimensional process data, data-driven latent variable methods, such as principal component analysis (PCA) and partial least-squares (PLS), have been the preferred methods. The recent resurgence of interest in machine learning and big data in other disciplines has revived similar interest in the PSE area.[5−7] It is hopeful that data analytics will bring significant benefits for a wide range of chemical engineering applications.

With a large amount of process data, PCA has been widely used to analyze variations in process variables and capture the normal region of variability. If real-time data from the process fall outside the predefined normal region, an alarm is generated to signal a potential anomaly. If the alarm persists, intervention and troubleshooting of the process operation are recommended. In practice, however, anomalies detected in process variables alone do not always result in an anomaly in product quality because of corrective actions by feedback controllers. Nuisance alarms for situations where the quality is unaffected tend to reduce the trustworthiness of the fault detection methods.

Supervised learning methods, such as PLS and neural networks, model the relation between process variables and quality variables, which can reduce the occurrence of nuisance alarms and improve the credibility of the monitoring system. By maximizing the covariance between the scores of quality data and those of the process data, PLS decomposes the original spaces into principal and residual subspaces, which can be monitored by T^2 and Q statistics, respectively.[8] However, since PLS maximizes covariance instead of correlation that is done in canonical correlation analysis (CCA),[9] it usually requires many latent dimensions to model even one output variable. The excessive dimensions make a large portion of the latent space irrelevant to the output to be modeled. Therefore, not all variations in the PLS latent space are relevant to the output variations.

Recently developed algorithms have been devoted to overcoming these issues, including total PLS (T-PLS),[10−12] concurrent PLS (CPLS),[13,14] and concurrent CCA (CCCA) with regularization to deal with potential collinearity in the process and quality variables.[15,16] These supervised learning methods simultaneously exploit covariations between process data and quality data, and are robust to collinearity. In addition, the concurrent CCA algorithm[15,16] further monitors quality-irrelevant variations in the input space.

Special Issue: Sirish Shah Festschrift

Received: January 18, 2019
Revised: March 9, 2019
Accepted: April 25, 2019
Published: April 25, 2019

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

Quality-relevant process monitoring has received increased attention since the early work on T-PLS.[10,11] Li et al.[17] proposed a fault diagnosis method based on T-PLS, where the diagnosis of process data variations is performed in four subspaces. Liu et al.[18] and Ge and Chen[19] developed quality-relevant fault detection by taking into account of dynamics in the data. Huang and Yan[20] proposed a two-block monitoring scheme based on mutual information (MI) and kernel PCA. In their scheme, process variables were divided into two sub-blocks based on the MI values with quality variables. Ma et al.[21] put forth a new robust Gaussian mixture model based quality-relevant fault detection for multimode processes, where quality-relevant faults are detected with a modified Mahalanobis distance. A Bayesian inference-based contribution index was also designed to diagnose faulty variables that are relevant to quality. Qin and Zhao[22] designed a monitoring method with consideration of quality information and process dynamics for closed-loop manufacturing processes, where quality-relevant process variations are separated by maximizing the correlation between latent variables and quality variables. In addition, nonlinear kernel and multiblock extensions are developed and applied to industrial continuous annealing processes.[23,24]

Diagnosis of a detected fault is necessary to further inspect the fault situation. Contribution plots, as an early approach, has been employed to diagnose a fault by determining the contribution of each variable to a fault detection index.[25,26] Westerhuis et al.[27] showed that contribution plots has smearing effects, which can lead to misleading results. To avoid ambiguity, reconstruction-based diagnosis methods have been proposed, where rigorous diagnosability can be analyzed for a given fault direction.[28] The advantage of the reconstruction based method is that faults with known fault directions can be diagnosed without ambiguity, but it requires prior knowledge of fault directions. To avoid this requirement, Alcala and Qin[29] proposed a reconstruction-based contribution (RBC) method. If fault data are available for a particular fault, the fault direction can be extracted from the fault data using singular value decomposition.[30] With the knowledge of fault directions, a generalized RBC (GRBC) is proposed for multidimensional fault diagnosis.[17] Li et al.[31] further applied the generalized RBC to reconstruct along variable directions, alleviating the need to derive fault directions.

Unlike diagnosis based on unsupervised learning models where one can at most analyze abnormal changes in process variables, quality-relevant diagnosis refers to the use of supervised learning models for the diagnosis of a quality anomaly due to relevant abnormal process variations. Since this approach uses supervised learning models to relate quality variations to process data, it is referred to as *supervised monitoring and diagnosis*. In this approach quality anomalies can be traced to relevant process variations with a supervised model. Therefore, it is inferential and can avoid the usual measurement delays of the quality variables. The reliability of supervised diagnosis depends on (i) the accuracy of the supervised learning model, and (ii) the variability of the data to train the supervised model. After quality data are measured, actual diagnosis of quality variables can be performed to further validate the inferential diagnosis results.

Since anomalies in product quality are of critical concern in industrial processes, this Article focuses on the supervised diagnosis of quality faults based on concurrent CCA with regularization.[16,32] Other supervised models, such as CPLS, can also be used. We consider the typical situation where quality measurements are sparse comparing to process data and have significant measurement delays. Therefore, in this paper, diagnosis is performed when a quality-relevant fault is detected. When a quality-irrelevant process fault is detected, diagnosis can also be performed, but with a much lower level of attention. Quality monitoring is subsequently performed on the quality data with delayed measurements. The Tennessee Eastman process data is used to illustrate the supervised diagnosis scheme by first categorizing the 15 disturbances based on their impact on quality variables and then testing the effectiveness of the proposed monitoring and diagnosis method on all disturbance cases.

The main contributions of this Article are (1) the proposal of a supervised diagnosis scheme for early process root-cause diagnosis before the quality is measured and actual quality monitoring after quality is measured; (2) the development of the supervised monitoring method based on regularized CCCA for monitoring quality-relevant and quality-irrelevant abnormal variations; (3) derivation of the necessary and sufficient conditions for supervised fault detectability; and (4) discussion of common misuses of Tennessee Eastman process (TEP) in the literature and categorizing the TEP disturbances based on quality relevance.

The rest of this Article is organized as follows: Section 2 defines the supervised monitoring scheme based on supervised models to detect quality-relevant faults and quality-irrelevant disturbances. Supervised fault detection, detectability, and diagnosis are discussed in Section 3 using CCCA models. The generalized RBC method is adopted to diagnose quality-relevant and quality-irrelevant process faults. In Section 4, the Tennessee Eastman process and quality data are employed for two objectives: one objective is to categorize the TEP disturbances based on whether they have an impact on the quality variables or not; another objective is to test the effectiveness of the proposed monitoring method in detecting and diagnosing quality-relevant and quality-irrelevant disturbances. Conclusions are drawn in the last section.

2. SUPERVISED MODELING AND MONITORING

We discuss and analyze in detail the inferential quality monitoring based on supervised models, the actual monitoring based on quality data, as well as quality-irrelevant process monitoring. The inferential quality monitoring used in conjunction with actual quality monitoring can facilitate early detection and root-cause diagnosis of relevant process anomalies, and actual confirmation of quality faults.

2.1. Inferential and Actual Quality Monitoring.

With data analytics tools such as PCA, PLS, CCA, and other machine learning methods, monitoring schemes of process and quality anomalies can be configured based on the nature of the models and data. Depending on the availability of quality data, either supervised learning models or unsupervised learning models can be built. Figure 1 depicts the inferential quality modeling, process monitoring, and quality monitoring schemes to be discussed in this section. The manufacturing process collects fast sampling process data, which are often used for feedback or feedforward control. The quality measurement data are typically slow and infrequent with large measurement delays. Although infrequent, these quality data are more critical than process variables as they indicate whether the product is good or not. The supervised monitoring includes inferential quality monitoring based on a supervised model, but it also monitors quality-irrelevant faults when they happen.

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

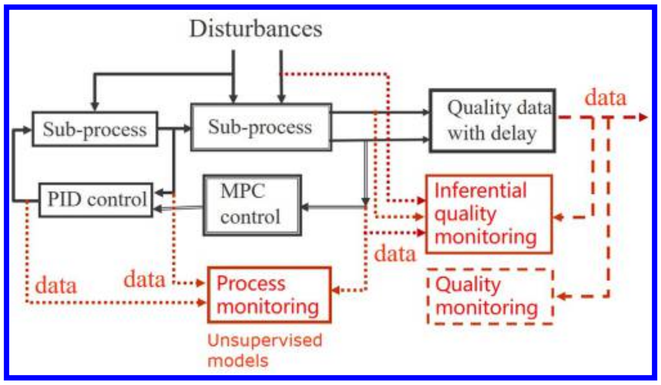


Figure 1. Supervised monitoring includes inferential quality monitoring, quality monitoring, and quality-irrelevant process monitoring. Dotted line: Fast sampling process data. Long dashed line: Slow and sparse quality data with measurement delays.

From Figure 1, we see that the product quality depends on controlled variables (CV), disturbance variables (DV), manipulated variables (MV), and potential faults (F). In general, this dependence can be expressed with the following relation,

$$\mathbf{y} = \mathbf{g}(\mathrm{CV}, \mathrm{DV}, \mathrm{MV}, F) + \text{noise} \tag{1}$$

where $\mathbf{y}$ can include quality variables, as well as other critical variables. The disturbance variables include both measured and unmeasured disturbances, and both of them can trigger MVs to act to reject them due to feedback or feedforward control. When the disturbance changes are compensated well by these controllers, the quality variables $\mathbf{y}$ will be little affected. Therefore, a supervised learning model that uses data from $\mathbf{y}$ will neglect these DV and MV changes. These disturbance changes are quality-irrelevant process variations. On the other hand, if any of these changes or potential faults (F) result in an abnormal change in $\mathbf{y}$, they are referred to as *quality-relevant process variations*. Within this setting it is clear that process disturbances are different from faults as they do not always lead to any effect on product quality.

Supervised monitoring includes the following scenarios based on quality relevance.

(1) *Inferential monitoring* (IM). IM refers to the detection and diagnosis of quality faults that can be inferred or predicted from process variables. IM includes quality-relevant monitoring (QrM) of abnormal variations that exist in the training data. In some cases, it should include potentially quality-relevant monitoring (PQrM) of variations that do not exist in the training data, but appear in real time data.[12,13,15]

(2) *Quality-irrelevant process monitoring* (QiPM). QiPM refers to the monitoring of faults in a subspace of the training data that have significant variations orthogonal to the quality-relevant subspace. QiPM should receive lower attention than IM.

(3) *Quality monitoring* (QM). QM focuses on the monitoring of variations in quality variables.[33] Typically the Hotelling's T^2 and Q statistics are used to detect the abnormal situations. However, QM cannot pinpoint to which process variables contribute to the quality anomalies since it does not build a supervised model to correlate quality variables with process variables.

These monitoring schemes are collectively referred to as *supervised monitoring* based on supervised models. Since IM provides the monitoring of quality variables using an inferential model based on process measurements, it can be executed as frequently and as soon as the process measurements are available. In this sense, IM is predictive. On the other hand, QM is the final authority to determine whether product quality is indeed abnormal or not, although it usually involves long time delays and long sampling intervals. Since the quality variables cannot be perfectly predicted from process variables, there can still be false alarms in inferential monitoring. Nevertheless, IM can significantly reduce nuisance alarms that would be caused by unsupervised monitoring alone.

Monitoring that uses unsupervised models of process variables alone is subject to false alarms caused by process disturbances. Therefore, it is inappropriate to treat process disturbances equally as process faults. For instance, the Tennessee Eastman process has been popularly used as a benchmark to demonstrate the detection rates of various methods.[34,35] Numerous publications try to show an incrementally high detection rate for a disturbance with their methods, but it is actually a high false alarm rate in the case of quality-irrelevant disturbances, which should be of no concern at all.

2.2. Supervised Modeling. In the modeling phase, the input matrix $\mathbf{X} \in \mathbb{R}^{n\times m}$ consisting of n samples with m process variables and the output matrix $\mathbf{Y} \in \mathbb{R}^{n\times p}$ with p quality variables are collected under normal conditions. PLS, for instance, can be used to decompose the scaled and mean-centered $\mathbf{X}$ and $\mathbf{Y}$ as

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathrm{T}} + \mathbf{E} \tag{2}$$

$$\mathbf{Y} = \mathbf{T}\mathbf{Q}^{\mathrm{T}} + \mathbf{F} \tag{3}$$

where the columns of $\mathbf{T}$ are latent score vectors and $\mathbf{P}$ and $\mathbf{Q}$ are loading matrices for $\mathbf{X}$ and $\mathbf{Y}$, respectively. $\mathbf{E}$ and $\mathbf{F}$ are residuals for $\mathbf{X}$ and $\mathbf{Y}$, and the number of PLS factors can be determined by cross-validation.[36,37] Although T^2 and Q statistics are suggested for PLS-based monitoring with the implication that T^2 is quality relevant and Q is process relevant,[26] this scheme has two major problems as pointed out by Li et al.,[10] Zhou et al.,[12] and Qin and Zheng.[13]

Considering the efficiency of CCA over PLS in predicting the output, Zhu et al.[16] proposed a regularized concurrent CCA algorithm to realize a comprehensive decomposition of the original space. The algorithm includes a regularized CCA, which is given in Appendix A, followed by a PCA decomposition of the residuals. The regularization parameters κ_1 and κ_2 are employed to control near-zero eigenvalues in strongly collinear cases, and they are determined by cross validation.[16] The decomposition by the regularized CCA is

$$\mathbf{X} = \mathbf{T}_c\mathbf{R}_c^{\dagger} + \tilde{\mathbf{X}}_c \tag{4}$$

where $\mathbf{T}_c = \mathbf{X}\mathbf{R}_c \in \mathbb{R}^{n\times l_c}$ and $\mathbf{R}_c \in \mathbb{R}^{m\times l_c}$ are the correlation scores and loadings from the regularized CCA model, which is referred to as the correlation subspace (CRS), and the number of principal components l_c is determined by the cross validation. $\mathbf{R}_c^{\dagger} = (\mathbf{R}_c^{\mathrm{T}}\mathbf{R}_c)^{-1}\mathbf{R}_c^{\mathrm{T}}$ is the Moore−Penrose pseudoinverse. The residuals $\tilde{\mathbf{X}}_c$ are decomposed further with the PCA model

$$\tilde{\mathbf{X}}_c = \mathbf{T}_x\mathbf{P}_x^{\mathrm{T}} + \tilde{\mathbf{X}} \tag{5}$$

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

where $\mathbf{P}_x \in \mathbb{R}^{m\times l_x}$ and $\mathbf{T}_x \in \mathbb{R}^{n\times l_x}$ are PCA loadings and scores in the orthogonal complement of CRS, which is referred to as the process-principal subspace (PPS), and $\tilde{\mathbf{X}}$ are the corresponding PCA residuals, which lies in the process-residual subspace (PRS).

Combining the above two models gives the following relation:

$$\mathbf{X} = \mathbf{T}_c\mathbf{R}_c^{\dagger} + \mathbf{T}_x\mathbf{P}_x^{T} + \tilde{\mathbf{X}} \tag{6}$$

A particular row of eq 6 represents the model relation for one sample of the data, which is

$$\mathbf{x}^T = \mathbf{t}_c^T\mathbf{R}_c^{\dagger} + \mathbf{t}_x^T\mathbf{P}_x^{T} + \tilde{\mathbf{X}}^T$$

or

$$\mathbf{x} = \mathbf{R}_c^{\dagger T}\mathbf{t}_c + \mathbf{P}_x\mathbf{t}_x + \tilde{\mathbf{X}} \tag{7}$$

where

$\mathbf{t}_c = \mathbf{R}_c^T\mathbf{x}$ is the correlation score vector

$\mathbf{t}_x = \mathbf{P}_x^T(\mathbf{I} - \mathbf{R}_c\mathbf{R}_c^{\dagger})\mathbf{x}$ is the principal score vector

$\tilde{\mathbf{X}} = (\mathbf{I} - \mathbf{P}_x\mathbf{P}_x^T)(\mathbf{I} - \mathbf{R}_c\mathbf{R}_c^{\dagger})\mathbf{x}$

3. SUPERVISED MONITORING AND DIAGNOSIS

3.1. Supervised Monitoring. The corresponding monitoring indices for each of the three parts in eq 7 are summarized in Table 1, where $\mathbf{\Lambda}_c = (n-1)^{-1}\mathbf{T}_c^T\mathbf{T}_c$ is the covariance matrix of the scores $\mathbf{t}_c$. Since $\mathbf{T}_c$ are the canonical variate scores, which are orthogonal in CCA, $\mathbf{\Lambda}_c$ is proportional to an identity matrix. Therefore, we can simply let $T_c^2 = \mathbf{t}_c^T\mathbf{t}_c$. $\mathbf{\Lambda}_x$ contains the principal eigenvalues of $\tilde{\mathbf{X}}_c$.

Table 1. Monitoring Indices and Control Limits Based on CCCA Decomposition[a]

index	scheme	control limit
$T_c^2 = \mathbf{t}_c^T\mathbf{\Lambda}_c^{-1}\mathbf{t}_c$	QrM	$\tau_c^2 = \{[l_c(n^2-1)]/[n(n-l_c)]\}F_{l_c,n-l_c,\alpha}$
$Q_x = \tilde{\mathbf{x}}^T\tilde{\mathbf{x}}$	PQrM	$\delta_x^2 = g_x\chi_{hx,\alpha}^2$
$T_x^2 = \mathbf{t}_x^T\mathbf{\Lambda}_x^{-1}\mathbf{t}_x$	QiPM	$\tau_x^2 = \{[l_x(n^2-1)]/[n(n-l_x)]\}F_{l_x,n-l_x,\alpha}$

[a]The control limits are derived in Zhu et al.[16]

When a fault occurs, it usually affects more than one fault detection index. Therefore, we have the following scenarios to prioritize the fault detection alarms:

1. If the T_c^2 index alarms, the fault is quality-relevant regardless of the other detection indices.
2. If the T_c^2 index does not alarm but Q_x alarms, the fault is potentially quality-relevant.
3. If the T_x^2 index alarms only, the fault is quality-irrelevant.

The fault detection indices T_c^2 and Q_x can be combined as an index for inferential monitoring[38]

$$\phi_x = \frac{T_c^2}{\tau_c^2} + \frac{Q_x}{\delta_x^2} = \mathbf{x}^T\mathbf{\Phi}_x\mathbf{x} \tag{8}$$

where

$$\mathbf{\Phi}_x = \frac{\mathbf{R}_c\mathbf{\Lambda}_c^{-1}\mathbf{R}_c^T}{\tau_c^2} + \frac{(\mathbf{I}-\mathbf{R}_c\mathbf{R}_c^{\dagger})(\mathbf{I}-\mathbf{P}_x\mathbf{P}_x^T)(\mathbf{I}-\mathbf{R}_c\mathbf{R}_c^{\dagger})}{\delta_x^2}$$

Since ϕ_x is a quadratic function of $\mathbf{x}$, its control limit can be easily obtained.[38,39] The three monitoring indices based on $\mathbf{x}$ can be unified as

$$\mathrm{Index}(\mathbf{x}) = \mathbf{x}^T\mathbf{M}\mathbf{x} \tag{9}$$

where $\mathbf{M}$ is given in Table 2.

Table 2. Formulations of M

index	$\mathbf{M}$
T_c^2	$\mathbf{R}_c\mathbf{\Lambda}_c^{-1}\mathbf{R}_c^T$
Q_x	$(\mathbf{I}-\mathbf{R}_c\mathbf{R}_c^{\dagger})(\mathbf{I}-\mathbf{P}_x\mathbf{P}_x^T)(\mathbf{I}-\mathbf{R}_c\mathbf{R}_c^{\dagger})$
ϕ_x	$\mathbf{\Phi}_x$

For quality monitoring after the quality data are measured, the following QM index can be used with a PCA model of the quality data:

$$\phi_y = \frac{T_y^2}{\tau_y^2} + \frac{Q_y}{\delta_y^2} = \mathbf{y}^T\mathbf{\Phi}_y\mathbf{y} \tag{10}$$

ϕ_y is the associated combined index.[38] The QM index provides a confirmation of the quality fault, but it is infrequent and subject to measurement delays.

3.2. Supervised Fault Detectability. When a fault occurs, the faulty sample vector can be represented as

$$\mathbf{x} = \mathbf{x}^* + \boldsymbol{\xi}_i\mathbf{f} \tag{11}$$

where $\mathbf{x}^*$ is the sample vector under normal operating conditions, and $\boldsymbol{\xi}_i\mathbf{f}$ is the fault part added to $\mathbf{x}^*$. In eq 11, $\boldsymbol{\xi}_i \in \Re^{m\times A_f}$ is an orthonormal matrix that spans the fault subspace with dimension A_f, and $\mathbf{f}$ is the magnitude of the fault. When $A_f = 1$, $\boldsymbol{\xi}_i$ reduces to a vector with unit norm, and it is classified as a unidimensional fault.

Substituting eq 11 into eq 9, we obtain

$$\begin{aligned}\mathrm{Index}(\mathbf{x}) &= \|\mathbf{M}^{1/2}\mathbf{x}\|^2 = \|\mathbf{M}^{1/2}\mathbf{x}^* + \mathbf{M}^{1/2}\boldsymbol{\xi}_i\mathbf{f}\|^2\\ &= \|\bar{\mathbf{x}}^* + \bar{\boldsymbol{\xi}}_i\mathbf{f}\|^2\end{aligned} \tag{12}$$

where $\bar{\mathbf{x}}^* = \mathbf{M}^{1/2}\mathbf{x}^*$, and $\bar{\boldsymbol{\xi}}_i = \mathbf{M}^{1/2}\boldsymbol{\xi}_i$. The matrix $\mathbf{M}$ can be any case in Table 2. Although $\boldsymbol{\xi}_i$ has full column rank, $\bar{\boldsymbol{\xi}}_i$ may not have full column rank. After applying singular value decomposition on $\bar{\boldsymbol{\xi}}_i$, we can get

$$\begin{aligned}\bar{\boldsymbol{\xi}}_i &= \mathbf{U}_i\mathbf{D}_i\mathbf{V}_i^T\\ &\equiv \boldsymbol{\xi}_i^{\circ}\mathbf{D}_i\mathbf{V}_i^T\end{aligned} \tag{13}$$

where $\mathbf{D}_i$ contains nonzero singular values and $\boldsymbol{\xi}_i^{\circ} \equiv \mathbf{U}_i$.

Then, eq 12 can be rearranged into

$$\begin{aligned}\mathrm{Index}(\mathbf{x}) &= \|\bar{\mathbf{x}}^* + \boldsymbol{\xi}_i^{\circ}\mathbf{D}_i\mathbf{V}_i^T\mathbf{f}\|^2\\ &= \|\bar{\mathbf{x}}^* + \boldsymbol{\xi}_i^{\circ}\mathbf{f}^{\circ}\|^2\end{aligned} \tag{14}$$

where $\mathbf{f}^{\circ} = \mathbf{D}_i\mathbf{V}_i^T\mathbf{f}$. Depending on the rank of $\bar{\boldsymbol{\xi}}_i$ and how the fault occurs relative to the null space of $\bar{\boldsymbol{\xi}}_i^T$, the quality-relevant fault detectability can have three possible scenarios, which are summarized in *Lemma 1*.

Lemma 1. The detectability condition of quality-relevant faults can be summarized as follows:

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

(1) If $\text{rank}(\bar{\xi}_i) = 0$, then the fault is not detectable no matter what $\mathbf{f}$ is.
(2) If $0 < \text{rank}(\bar{\xi}_i) < A_f$, then the fault is not detectable if $\mathbf{f} \in \mathbb{N}(\mathbf{V}_i)$, where $\mathbb{N}(\cdot)$ denotes the null space.
(3) If $\text{rank}(\bar{\xi}_i) = A_f$, then the fault is detectable for $\mathbf{f} \neq 0$.

Lemma 1 is only the necessary condition for a fault to be detectable. However, to guarantee fault detection, the magnitude of the fault should be large enough such that $\text{Index}(\mathbf{x}) > \zeta^2$, where ζ^2 is the control limit. The sufficient detectability condition is shown in *Lemma 2*.

Lemma 2. When $\mathbf{f} \notin \mathbb{N}(\mathbf{V}_i)$, the quality-relevant fault is guaranteed to be detected if $\|\mathbf{f}\| > 2\zeta/d_{\min}$, where $d_{\min}$ is the minimum nonzero singular values of $\bar{\xi}_i$.

The proof of *Lemma 1* and *Lemma 2* is given in Appendix B.

3.3. Supervised Diagnosis. To perform diagnosis once they are detected by a monitoring index, Zhu et al.[32] derive the following contributions for eq 9:

$$c_i^{\mathbf{M}} = (\boldsymbol{\xi}_i^{\mathrm{T}}\mathbf{M}^{1/2}\mathbf{x})^2 \tag{15}$$

$\boldsymbol{\xi}_i$ is a column vector of the identity matrix, and the following reconstruction-based contributions (RBC)

$$\text{RBC}_i^{\mathbf{M}} = \mathbf{x}^{\mathrm{T}}\mathbf{M}\boldsymbol{\xi}_i(\boldsymbol{\xi}_i^{\mathrm{T}}\mathbf{M}\boldsymbol{\xi}_i)^{-1}\boldsymbol{\xi}_i^{\mathrm{T}}\mathbf{M}\mathbf{x} \tag{16}$$

where $\boldsymbol{\xi}_i$ is a fault direction matrix, which can have multiple columns. In the case that $\boldsymbol{\xi}_i$ is a column vector

$$\text{RBC}_i^{\mathbf{M}} = \frac{(\boldsymbol{\xi}_i^{\mathrm{T}}\mathbf{M}\mathbf{x})^2}{\boldsymbol{\xi}_i^{\mathrm{T}}\mathbf{M}\boldsymbol{\xi}_i}$$

which shows the similarity between RBC and contributions. For the case of the Q_x index, $\mathbf{M}^2 = \mathbf{M}$. Therefore, $\mathbf{M}^{1/2} = \mathbf{M}$. In this case, RBC is exactly scaled contributions. Equations 15 and 16 have quadratic forms, so their control limits can be approximated using Chi-square distributions.[39]

For multidimensional faults, we adopt the generalized RBC approach.[31,32] The generalized RBC reconstructs along a variable direction with the largest RBC and augments the reconstruction subspace with the next largest contributing variable, until the fault detection index returns to the normal region. The effectiveness of the generalized RBC will be demonstrated in the next section using the TEP data.

4. SUPERVISED DIAGNOSIS OF THE TENNESSEE EASTMAN PROCESS

The Tennessee Eastman process[40] was created to provide a challenging benchmark for the purpose of developing, studying and evaluating plant-wide control strategies. The process makes two products (G and H) and a byproduct (F) from four reactants (A, C, D, and E) with the following reactions:

$$\text{A(g)} + \text{C(g)} + \text{D(g)} \rightarrow \text{G(l)}$$

$$\text{A(g)} + \text{C(g)} + \text{E(g)} \rightarrow \text{H(l)}$$

$$\text{A(g)} + \text{E(g)} \rightarrow \text{F(l)}$$

$$3\text{D(g)} \rightarrow 2\text{F(l)}$$

The TEP process diagram along with feedback control loops can be found in Lyman and Georgakis.[41]

The TEP simulation benchmark was first used for process monitoring by Ku et al.[42] and Raich and Cinar.[43] Later Chiang et al.[34] used it again for fault diagnosis and classification, which generated data for a disturbance-free period and data for each of the 15 known types of disturbances. Since then, based on a recent Scopus search, over 600 papers have been published by the end of 2018. It has been observed that most of the recently published papers tried to increase the detection rates of these disturbances, but many of the disturbances should not be of concern since they are compensated well by feedback controllers. Other inappropriate uses are summarized in Zhu et al.[32] If a disturbance is well compensated with feedback and feedforward controllers and thus has no impact on the product quality, any alarms based on it would be nuisance and would be annoying to the operation personnel. Therefore, one objective of this session is to categorize these disturbances based on whether they have an impact on the quality variables or not. Another objective is to test the effectiveness of the proposed supervised method in diagnosing quality-relevant and quality-irrelevant disturbances.

4.1. Quality Monitoring of TEP Disturbances. The main products G and H are extracted in stream 11, which are measured every 15 min with a time delay and are main quality variables of concern. Additionally, the byproduct F is purged from the separator and flowed into stream 9, which should also be concerned if it is affected by a disturbance. We perform quality monitoring on the quality data of these products to determine whether these disturbances have an impact on the quality variables or not.

The descriptions of the 15 known disturbances in TEP are shown in the first three columns of Table 3. Figure 2 shows the quality monitoring results for one disturbance-free case and 15 disturbance cases using the combined index ϕ_y in eq 10. From this figure, we can see that seven of the 15 disturbance cases have no impact on the quality variables. Among the eight cases that have impact on the quality variables, IDV(5) (condenser cooling water inlet temperature) and IDV(7) (C header pressure reduced in stream 4) have only temporary impact on quality shortly after the fault occurred. The classification of these disturbances in terms of quality impact is summarized in the fourth column of Table 3. For the seven quality unimpacted cases, five of them are temperature changes that are compensated well by feedback controllers. The other two

Table 3. TEP Disturbances and Their Impact on Quality

	disturbance description	type	QM impact
IDV(1)	*A*/*C* feed ratio (stream 4)	step	**yes**
IDV(2)	*B* composition (stream 4)	step	**yes**
IDV(3)	*D* feed temperature (stream 2)	step	no
IDV(4)	reactor cooling water inlet temperature	step	no
IDV(5)	condenser cooling water inlet temperature	step	**yes**
IDV(6)	*A* feed loss (stream 1)	step	**yes**
IDV(7)	*C* header pressure loss—reduced (stream 4)	step	**yes**
IDV(8)	*A*, *B*, *C* feed composition (stream 4)	random	**yes**
IDV(9)	*D* feed temperature (stream 2)	random	no
IDV(10)	*C* feed temperature (stream 4)	random	no
IDV(11)	reactor cooling water inlet temperature	random	no
IDV(12)	condenser cooling water inlet temperature	random	**yes**
IDV(13)	reaction kinetics	slow drift	**yes**
IDV(14)	reactor cooling water valve	sticking	no
IDV(15)	condenser cooling water valve	sticking	no

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

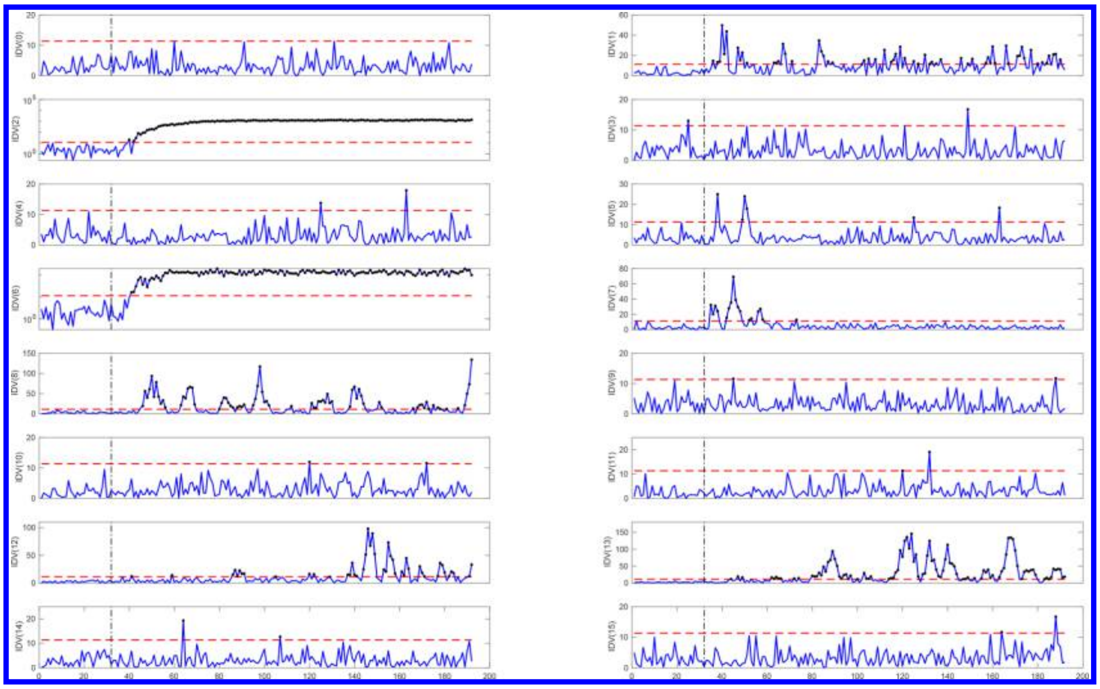


Figure 2. Quality monitoring results on the disturbance free case and each of the 15 disturbance cases using ϕ_y.

cases are sticking valves of cooling water, which do not yield an impact on quality. They should be better monitored with valve stiction detection algorithms.[44]

4.2. Supervised Monitoring for TEP. For monitoring based on process data, CCCA, PLS, and PCA are performed to compare the effectiveness of quality relevant monitoring. The selected process variables are XMEAS(1−22) and XMV(1−11) and the quality variables are XMEAS(41, 42, 34) for products G, H, and F. After handling the unequal sampling rates and measurement delays in the quality data, 100 normal samples are used to build the models. The number of components used in the models are determined by cross-validation: for CCCA, $l_c = 2$, $l_x = 19$, $l_y = 3$, $\kappa_1 = 0.021$, and $\kappa_2 = 0.002$, where κ_1 and κ_2 are the regularization terms for process and quality data respectively; for PLS, $l = 10$ and for PCA, $l = 19$. The confidence level for the control limits is chosen as 99%.

The quality-relevant index T_c^2 with CCCA for the quality impacted cases and unimpacted cases are summarized in Tables 4 and 5, respectively. Table 4 includes all eight quality impacted cases. From Figure 2, it is seen that the QM control limits are exceeded intermittently for most of the quality impacted cases. Therefore, we use the number of QM-alarmed samples as the denominator to calculate the inferential detection rates. We see from Table 4 that the T_c^2 index with CCCA successfully detects the quality-impacted faults. For comparison, we also applied PLS and PCA on these cases, which are shown in Table 4 as well. It is seen that T_c^2 with CCCA outperformed the other methods. On the other hand, for the quality-irrelevant cases shown in Table 5, T_c^2 with CCCA based monitoring shows very low percentage of alarms, which indicates that these disturbances are not quality-relevant. However, PLS and PCA give much higher alarm rates, which are nuisance alarms.

Table 4. Quality-Relevant Alarm Rates for the Quality-Impacted Cases (%)[a]

disturbance	CCCA	PLS	PCA
IDV(1)	**100**	**100**	99.36
IDV(2)	**100**	**100**	**100**
IDV(5)	**93.26**	72.73	75.76
IDV(6)	**100**	**100**	**100**
IDV(7)	**100**	98.85	95.40
IDV(8)	**100**	**100**	**100**
IDV(12)	**100**	**100**	**100**
IDV(13)	**100**	**100**	99.56

[a]The higher the better.

Table 5. Quality-Relevant Alarm Rates for the Quality Unimpacted Cases (%)[a]

disturbance	CCCA	PLS	PCA
IDV(3)	**3.17**	14.29	15.34
IDV(4)	**4.74**	57.37	83.68
IDV(9)	**2.63**	11.58	16.84
IDV(10)	**8.85**	44.27	44.27
IDV(11)	**8.38**	50.79	65.97
IDV(14)	**13.23**	83.60	84.13
IDV(15)	**2.66**	11.70	9.04

[a]The lower the better.

4.3. Supervised Diagnosis for TEP. To perform diagnosis for faults and disturbances in Tables 4 and 5, CCCA-based contribution plots, RBC, and GRBC are performed. Following the order of priority of T_c^2, Q_x, and T_x^2, the combined index ϕ_x is used for diagnosis when T_c^2 alarms. If T_c^2 does not alarm but Q_x does, Q_x is used for diagnosis. For the limited space, we only illustrate IDV(1) as an example of the quality-impacted cases and IDV(3) and IDV(4) as examples of the quality-unimpacted cases. Figure 3 gives a preview of the quality-relevant monitoring results of the IDV(1) and IDV(3) cases based on T_c^2, which clearly show that IDV(1) is quality relevant while IDV(3) is not.

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

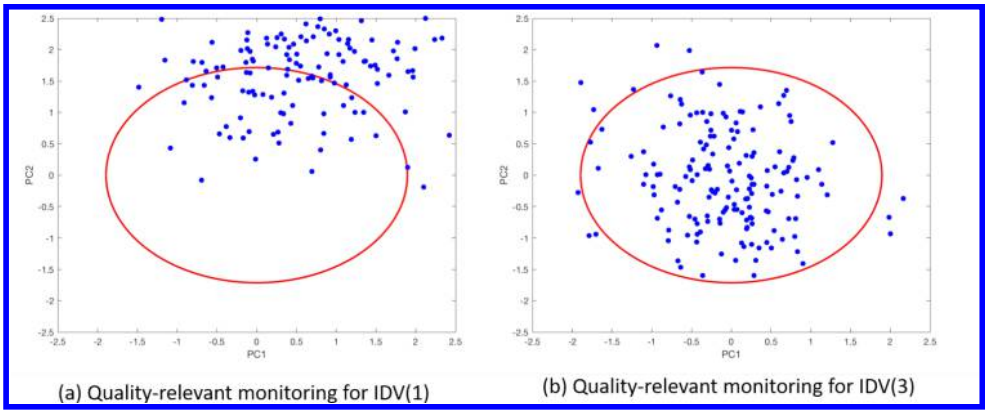


Figure 3. Quality-relevant monitoring results for IDV(1) and IDV(3) based on T_c^2.

4.4. Diagnosis of the Impact of IDV(1). The supervised monitoring results for IDV(1) based on regularized CCCA are shown in Figure 4. It is observed that the QrM index T_c^2 agrees well with the QM index T_y^2, which shows its effectiveness to detect quality-relevant faults. It is also observed that the T_c^2 index detects the fault faster than the QM index T_y^2 does.

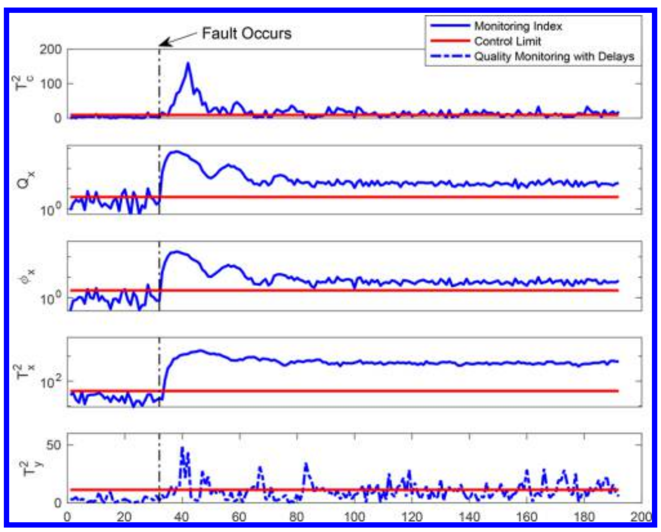


Figure 4. Monitoring results for IDV(1) based on regularized CCCA. The QrM index T_c^2 agrees well with the QM index T_y^2.

To diagnose IDV(1) as a quality-relevant fault, we apply contribution plots and RBC for T_c^2, Q_x, or ϕ_x after these indices exceed the control limits as if they would be done in real time. As shown in Figure 4, right after the disturbance is introduced, the Q_x and ϕ_x indices alarmed immediately. The T_c^2 index is marginally above the control limit initially for samples 33−36, then goes above the control limit significantly. Therefore, we pick the samples 33−36 and 42 as examples to diagnose the progression of the fault based on ϕ_x, which is shown in Figure 5. It is noted that the contribution plots and RBC are scaled by their control limits for easy visualization. From Figure 5, on the basis of the magnitudes of both contribution plots and RBCs, we observe that variables 4 (*A*/*C* feed in stream 4) and 20 (compressor work) are persistent high contributors through these samples. At sample 42, variable 16 (stripper pressure) shows a high contribution. Since IDV(1) is a step change in *A*/*C* feed ratio, variable 4 (*A*/*C* feed in stream 4) is correctly diagnosed. We can see that the early diagnosis of samples 33−36 is predictive before the quality is significantly impacted.

In addition, the RBC makes these variables more standing out than the contributions do, but they all point to multiple high contributions since the fault impact is multidimensional.

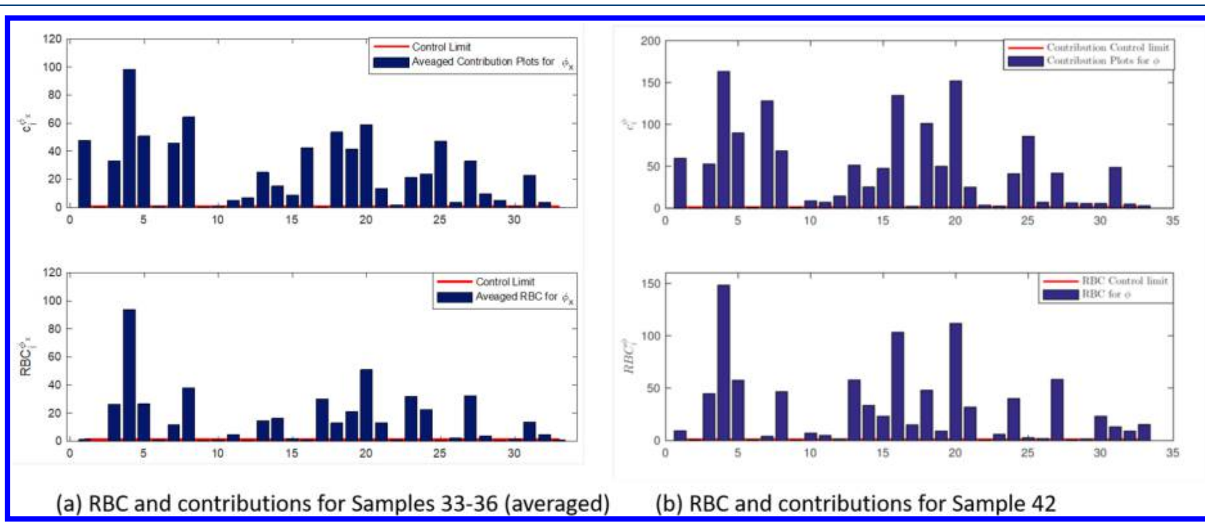


Figure 5. Diagnosis results of IDV(1) using contributions (upper) and RBCs (lower) based on ϕ_x.

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

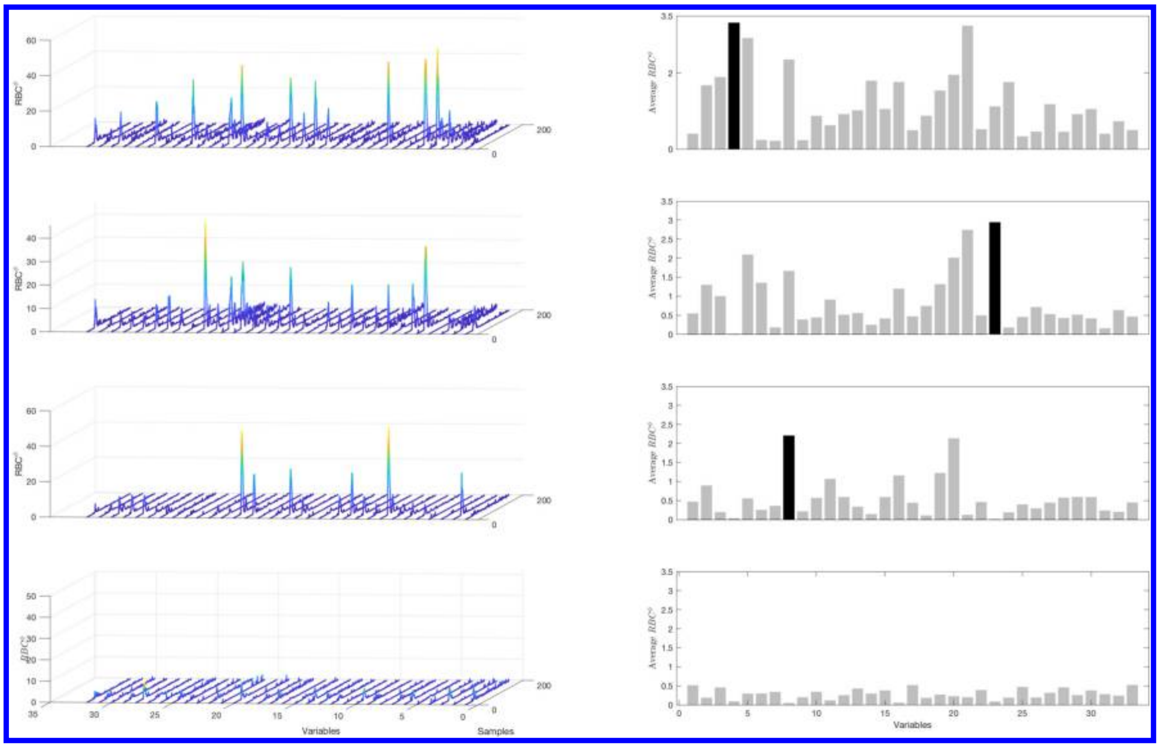


Figure 6. Generalized RBC diagnosis result for IDV(1) for all samples (left) and average RBC (right) . First row: Initial RBCs. Second row: RBCs after reconstructing along variable 4. Third row: RBCs after reconstructing along variables 4 and 23. Fourth row: RBCs after reconstructing along variables 4, 23, and 8.

Therefore, the generalized RBC method is applied next. Figure 6 shows the results of GRBC by successively including the next highest contributing variable in the reconstruction iterations. The left charts show the contributions after including the next variable for reconstruction, while the right charts shows the next highest-contributing variable on average to be included. With GRBC the extracted variables are variable 4 (A/C feed in stream 4), variable 23 (D feed flow in stream 2), and variable 8 (reactor level), successively.

For IDV(1), a step change is introduced in A/C feed ratio. Variable 4, which is the A/C feed in stream 4, is affected immediately. Then, because of the mass balance required for the reactions, the feed flow of reactant D in stream 2 (variable 23) decreases, which leads to a decrease of the throughput of product G. Since G is a liquid, the reactor level (variable 8) is also affected. Variables 5, 16, and 20 have high RBCs before any reconstruction, but they disappeared after reconstructing along variables 4, 23, and 8, as shown in Figure 7, which means that they are dependent on variables 4, 23, and 8.

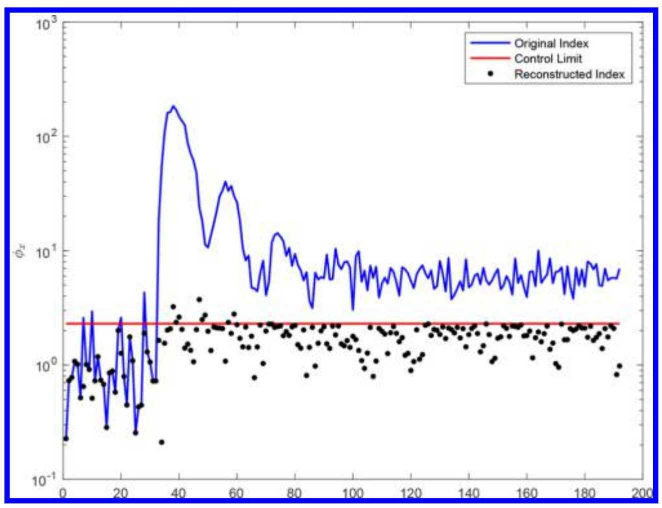


Figure 7. Combined index ϕ_x for IDV(1) before and after reconstruction along variables 4, 23, and 8 identified by generalized RBC.

4.5. Diagnosis of the Impact of IDV(3) and IDV(4).

Since the supervised monitoring can be executed as soon as the process measurement data are available, the related indices are calculated once every 3 min for IDV(3) and IDV(4), while the quality monitoring index T_y^2 is calculated once every 15 min. The case of IDV(4) is chosen since in this case the proposed monitoring method reduces nuisance alarms significantly.

For IDV(3), which is a step change in the D feed temperature, the fault detection results are shown in Figure 8. It is clear that this case is a quality-irrelevant disturbance, since the T_c^2 index is within control. There are only a small fraction of samples where Q_x is marginally above the control limit. For the sake of completeness, the T_c^2 contribution plots and RBC for the 250th process sample are shown in Figures 9. It is observed that the contributions and RBC for T_c^2 are small across all variables, which confirmed that the case is quality-irrelevant.

For IDV(4), which is a step change in the reactor cooling water inlet temperature, the fault detection results are shown in Figure 10. It is again a case of quality-irrelevant disturbance, since the T_c^2 index is within control except for the sample right after the disturbance occurred. The Q_x index is above its control limit. Therefore, we pick the 250th process sample to calculate the contributions and RBC of the Q_x index for diagnosis, which is shown in Figure 11. It is seen that variable 32 has a high contribution for Q_x. Variable 32 is the reactor

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

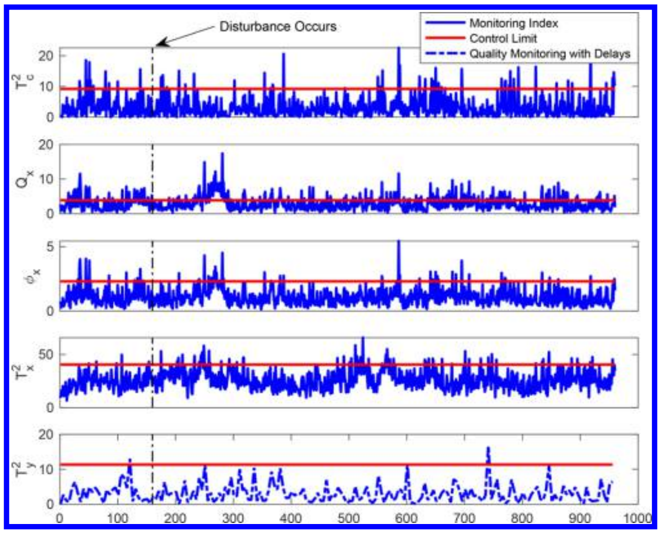


Figure 8. Monitoring results for IDV(3) based on regularized CCCA. The QrM index T_c^2 index agrees well with the QM index T_y^2.

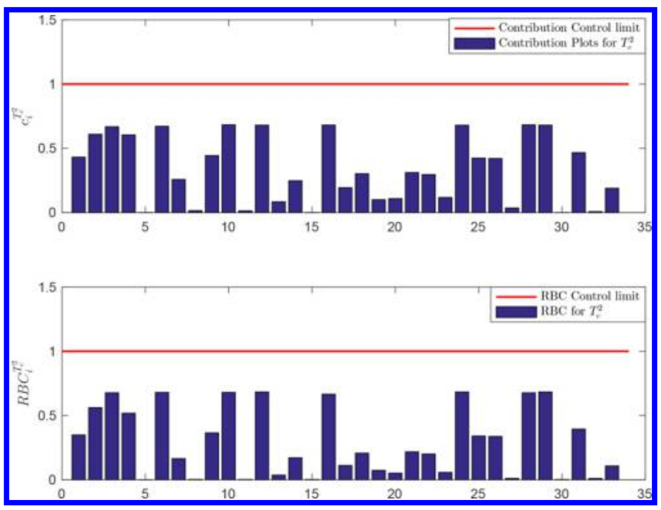


Figure 9. Diagnosis results for the 250th process sample of IDV(3) using contributions (upper) and RBC (lower) with T_c^2.

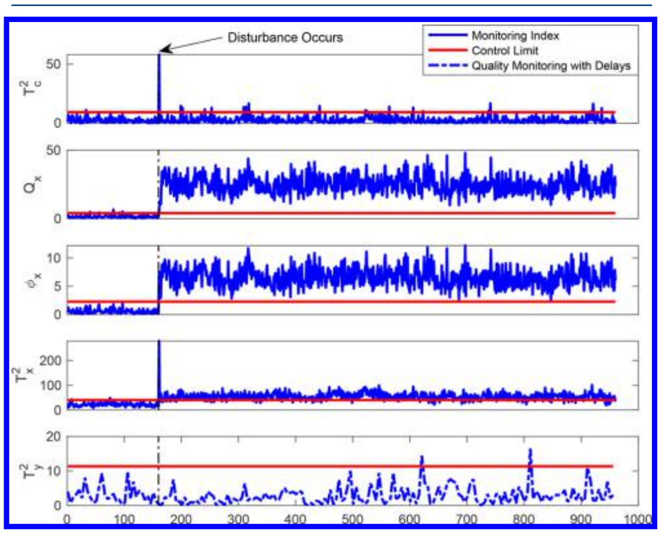


Figure 10. Monitoring results for IDV(4) based on regularized CCCA.

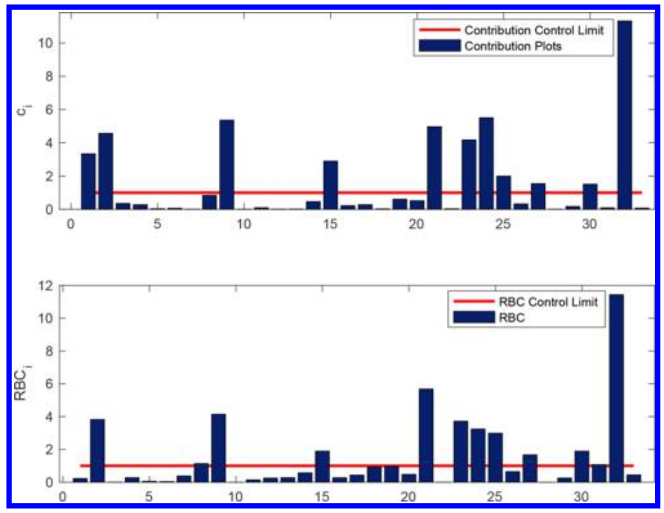


Figure 11. Diagnosis results for the 250th process sample of IDV(4) using contributions (upper) and RBC (lower) with Q_x.

cooling water flow, which acted to compensate for the step change in the reactor cooling water inlet temperature. To conclude, the proposed supervised monitoring method clearly indicates that IDV(3) and IDV(4) are quality-irrelevant.

5. CONCLUSIONS

In this Article, a supervised monitoring and diagnosis framework is formulated to deal with the detection and diagnosis of quality-relevant and quality-irrelevant faults. Concurrent CCA with regularization is employed to calculate the quality-relevant, potentially quality-relevant, and quality-irrelevant monitoring indices. The supervised monitoring indices can be calculated as soon as the process data are measured, which provide predictive monitoring for product quality. The generalized RBC approach is shown to be more effective than RBC and contributions for multidimensional inferential fault diagnosis. Finally, the performance of supervised monitoring and diagnosis based on regularized CCCA is successfully demonstrated using the TEP simulation data, where the 15 disturbances are classified into quality-impacted faults and quality unimpacted disturbances based on the quality monitoring of two products and one byproduct.

■ APPENDIX A

Regularized CCA Algorithm

This regularized CCA algorithm is adapted from Zhu et al.[16]

1 Scale the data $\mathbf{X}$, $\mathbf{Y}$ to zero mean and unit variance.
2 Perform eigenvalue decomposition on the matrices $\mathbf{X}^T\mathbf{X}$ and $\mathbf{Y}^T\mathbf{Y}$ to calculate the square roots

$$\mathbf{X}^T\mathbf{X} = \mathbf{V}_x\mathbf{D}_x\mathbf{V}_x^T$$

$$\mathbf{Y}^T\mathbf{Y} = \mathbf{V}_y\mathbf{D}_y\mathbf{V}_y^T$$

$$\mathbf{\Sigma}_{xx}^{-1/2} = \mathbf{V}_x(\mathbf{D}_x + \kappa_1\mathbf{I}))$$

$$\mathbf{\Sigma}_{yy}^{-1/2} = \mathbf{V}_y(\mathbf{D}_y + \kappa_2\mathbf{I})^{-1/2}\mathbf{V}_y^T$$

where κ_1 and κ_2 are regularization weights.
3 Perform SVD and calculate the weight matrix $\mathbf{R}_c$

$$\mathbf{\Sigma}_{xx}^{-1/2}\mathbf{X}^T\mathbf{Y}\mathbf{\Sigma}_{yy}^{-1/2} = \mathbf{USV}^T$$

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

$$\mathbf{R}_c = \mathbf{\Sigma}_{xx}^{-1/2}\mathbf{U}$$

where $\mathbf{S}$ contains the non-zero singular values only.

4 Obtain the canonical correlation scores $\mathbf{T}_c = \mathbf{X}\mathbf{R}_c$.

APPENDIX B

Proof of Lemmas 1 and 2

If $\text{rank}(\bar{\boldsymbol{\xi}}_i) = 0$ (i.e., $\bar{\boldsymbol{\xi}}_i = 0$), according to eq 12, the fault cannot be detected.

If $\mathbf{f} \in \mathbb{N}(\mathbf{V}_i)$, then

$$\mathbf{f}^\circ = \mathbf{D}_i\mathbf{V}_i^T\mathbf{f} = 0 \tag{17}$$

This makes the second term of eq 14 equal to zero and $\text{Index}(\mathbf{x})$ always within ζ^2. Thus, in this case, the fault is not detectable as well.

If $\text{rank}(\bar{\boldsymbol{\xi}}_i) = A_f$, that is, $\bar{\boldsymbol{\xi}}_i$ has full column rank, the fault is detectable for a large enough fault magnitude.

If $\mathbf{f} \notin \mathbb{N}(\mathbf{V}_i)$, in order for the fault to be guaranteed detected

$$\text{Index}(\mathbf{x}) = \|\bar{\mathbf{x}}^* + \bar{\boldsymbol{\xi}}_i\mathbf{f}\|^2 \geq (\|\bar{\mathbf{x}}^*\| - \|\bar{\boldsymbol{\xi}}_i\mathbf{f}\|)^2 > \zeta^2 \tag{18}$$

Since the fault-free portion $\|\bar{\mathbf{x}}^*\| < \zeta$, it is sufficient to require that $\|\bar{\boldsymbol{\xi}}_i\mathbf{f}\| > 2\zeta$. Therefore,

1. If $\text{rank}(\bar{\boldsymbol{\xi}}_i) = A_f$, $\bar{\boldsymbol{\xi}}_i$ is a full column-rank matrix and all the singular values are greater than 0. Then $\|\bar{\boldsymbol{\xi}}_i\mathbf{f}\| \geq d_{\min}\|\mathbf{f}\| > 2\zeta$.
2. If $0 < \text{rank}(\bar{\boldsymbol{\xi}}_i) < A_f$, then some of the singular values are zero. Then the minimum singular value $d_{\min}$ is adopted among the nonzero ones, which gives $d_{\min}\|\mathbf{f}\| > 2\zeta$.

Thus, when $\|\mathbf{f}\| > 2\zeta/d_{\min}$, the fault can be guaranteed to be detected.

AUTHOR INFORMATION

Corresponding Author

*E-mail: sqin@usc.edu.

ORCID

S. Joe Qin: 0000-0001-7631-2535

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the Natural Science Foundation of China (61490704), the Fundamental Disciplines Program of the Shenzhen Committee on Science and Innovations (20160207, 20170155), and the Texas−Wisconsin−California Control Consortium.

REFERENCES

(1) MacGregor, J. F.; Kourti, T. Statistical process control of multivariate processes. *Control Engineering Practice* **1995**, *3*, 403−414.

(2) Wise, B.; Gallagher, N. The process chemometrics approach to process monitoring and fault detection. *J. Process Control* **1996**, *6*, 329−348.

(3) Qin, S. J. Statistical process monitoring: basics and beyond. *J. Chemom.* **2003**, *17*, 480−502.

(4) Qin, S. J. Survey on data-driven industrial process monitoring and diagnosis. *Annual Reviews in Control* **2012**, *36*, 220−234.

(5) Qin, S. J. Process data analytics in the era of big data. *AIChE J.* **2014**, *60*, 3092−3100.

(6) McClure, K. S.; Gopaluni, R. B.; Chmelyk, T.; Marshman, D.; Shah, S. L. Nonlinear Process Monitoring Using Supervised Locally Linear Embedding Projection. *Ind. Eng. Chem. Res.* **2014**, *53*, 5205−5216.

(7) Venkatasubramanian, V. The promise of artificial intelligence in chemical engineering: Is it here, finally? *AIChE J.* **2019**, *65*, 466−478.

(8) MacGregor, J. F.; Jaeckle, C.; Kiparissides, C.; Koutoudi, M. *AIChE J.* **1994**, *40*, 826−838.

(9) Hotelling, H. Relations between two sets of variates. *Biometrika* **1936**, *28*, 321−377.

(10) Li, G.; Qin, S. J.; Ji, Y.; Zhou, D. Total PLS Based Contribution Plots for Fault Diagnosis. *Acta Automatica Sinica* **2009**, *35*, 759−765.

(11) Li, G.; Qin, S.; Zhou, D. Output relevant fault reconstruction and fault subspace extraction in total projection to latent structures models. *Ind. Eng. Chem. Res.* **2010**, *49*, 9175−9183.

(12) Zhou, D.; Li, G.; Qin, S. J. Total projection to latent structures for process monitoring. *AIChE J.* **2009**, *56*, 168−178.

(13) Qin, S. J.; Zheng, Y. Quality-relevant and process-relevant fault monitoring with concurrent projection to latent structures. *AIChE J.* **2013**, *59*, 496−504.

(14) Zhao, Z.; Li, Q.; Huang, M.; Liu, F. Concurrent PLS-based process monitoring with incomplete input and quality measurements. *Comput. Chem. Eng.* **2014**, *67*, 69−82.

(15) Zhu, Q.; Liu, Q.; Qin, S. J. Concurrent Canonical Correlation Analysis Modeling for Quality-Relevant Monitoring. *IFAC-PapersOnLine* **2016**, *49*, 1044−1049.

(16) Zhu, Q.; Liu, Q.; Qin, S. J. Concurrent quality and process monitoring with canonical correlation analysis. *J. Process Control* **2017**, *60*, 95−103.

(17) Li, G.; Alcala, C. F.; Qin, S. J.; Zhou, D. Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the Tennessee Eastman process. *IEEE transactions on control systems technology* **2011**, *19*, 1114−1127.

(18) Liu, Q.; Qin, S. J.; Chai, T. Quality-relevant monitoring and diagnosis with dynamic concurrent projection to latent structures. *IFAC Proceedings Volumes* **2014**, *47*, 2740−2745.

(19) Ge, Z.; Chen, X. Supervised linear dynamic system model for quality related fault detection in dynamic processes. *J. Process Control* **2016**, *44*, 224−235.

(20) Huang, J.; Yan, X. Quality relevant and independent two block monitoring based on mutual information and KPCA. *IEEE Transactions on Industrial Electronics* **2017**, *64*, 6518−6527.

(21) Ma, L.; Dong, J.; Peng, K. Root cause diagnosis of quality-related faults in industrial multimode processes using robust Gaussian mixture model and transfer entropy. *Neurocomputing* **2018**, *285*, 60−73.

(22) Qin, Y.; Zhao, C. A Quality-Relevant Monitoring Method for Closed-Loop Industrial Processes with Dual Consideration of Static and Dynamic Analysis. *2018 10th International Conference on Modelling*; Identification and Control (ICMIC), 2018; pp 1−6.

(23) Sheng, N.; Liu, Q.; Qin, S. J.; Chai, T. Comprehensive monitoring of nonlinear processes based on concurrent kernel projection to latent structures. *IEEE Transactions on Automation Science and Engineering* **2016**, *13*, 1129−1137.

(24) Liu, Q.; Zhu, Q.; Qin, S. J.; Chai, T. Dynamic concurrent kernel CCA for strip-thickness relevant fault diagnosis of continuous annealing processes. *J. Process Control* **2018**, *67*, 12−22.

(25) Miller, P.; Swanson, R. E.; Heckler, C. E. Contribution plots: a missing link in multivariate quality control. *Applied Mathematics and Computer Science* **1998**, *8*, 775−792.

(26) Nomikos, P.; MacGregor, J. F. Multivariate SPC charts for monitoring batch processes. *Technometrics* **1995**, *37*, 41−59.

(27) Westerhuis, J. A.; Gurden, S. P.; Smilde, A. K. Generalized contribution plots in multivariate statistical process monitoring. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 95−114.

(28) Dunia, R.; Qin, S. J. Subspace approach to multidimensional fault identification and reconstruction. *AIChE J.* **1998**, *44*, 1813−1831.

(29) Alcala, C. F.; Qin, S. J. Reconstruction-based contribution for process monitoring. *Automatica* **2009**, *45*, 1593−1600.

(30) Qin, S. J.; Valle, S.; Piovoso, M. J. On unifying multiblock analysis with application to decentralized process monitoring. *J. Chemom.* **2001**, *15*, 715−742.

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223

(31) Li, G.; Qin, S. J.; Chai, T. *Multi-directional Reconstruction Based Contributions for Root-Cause Diagnosis of Dynamic Processes*; American Control Conference (ACC), 2014; pp 3500−3505.
(32) Zhu, Q.; Liu, Q.; Qin, S. J. Concurrent Monitoring and Diagnosis of Process and Quality Faults with Canonical Correlation Analysis. *Proc. of IFAC Congress* **2017**, *50*, 7999−8004.
(33) Jackson, J. E. *A User's Guide to Principal Components*; Wiley-Interscience: New York, 1991; Vol. *587*.
(34) Chiang, L.; Russell, E.; Braatz, R. Fault Diagnosis and Fisher Discriminant Analysis, Discriminant Partial Least Squares, and Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **2000**, *50*, 243−252.
(35) Yin, S.; Ding, S. X.; Haghani, A.; Hao, H.; Zhang, P. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *J. Process Control* **2012**, *22*, 1567−1581.
(36) Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* **1986**, *185*, 1−17.
(37) Höskuldsson, A. PLS regression methods. *J. Chemom.* **1988**, *2*, 211−228.
(38) Yue, H. H.; Qin, S. J. Reconstruction-based fault identification using a combined index. *Ind. Eng. Chem. Res.* **2001**, *40*, 4403−4414.
(39) Box, G. E.; et al. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Stat.* **1954**, *25*, 290−302.
(40) Downs, J. J.; Vogel, E. F. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, *17*, 245−255.
(41) Lyman, P.; Georgakis, C. Plant-wide control of the Tennessee Eastman problem. *Comput. Chem. Eng.* **1995**, *19*, 321−331.
(42) Ku, W.; Storer, R.; Georgakis, C. Disturbance detection and isolation by dynamic principal component analysis. *Chemom. Intell. Lab. Syst.* **1995**, *30*, 179.
(43) Raich, A.; Cinar, A. Statistical process monitoring and disturbance diagnosis in multivariate continuous processes. *AIChE J.* **1996**, *42*, 995−1009.
(44) Choudhury, S. M. A. A.; Shah, S. L.; Thornhill, N. F. *Diagnosis of Process Nonlinearities and Valve Stiction—2008 Data Driven Approaches*; Springer, 2008.

DOI: 10.1021/acs.iecr.9b00320
Ind. Eng. Chem. Res. 2019, 58, 11213−11223