Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnline CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 55-2 (2022) 258–263

Enhancing Fault Detection with Clustering and Covariance Analysis

Ethan Gallup * **Titus Quah** * **Derek Machalek** *
Kody M. Powell *,**

* *Department of Chemical Engineering, University of Utah, 50 Central Campus Dr, Salt Lake City, UT 84112, USA (Tel: +1-801-581-3957; email: kody.powell@utah.edu*
** *Department of Mechanical Engineering, University of Utah, 1495 E 100 S, Salt Lake City, UT 84112, USA*

Abstract: Fault detection plays an important role in identifying abnormalities in high-cost, large-scale industrial processes. Clustering in combination with dimensionality reduction is a common practice in data analysis and anomaly detection but is not well explored in the field of industrial fault detection. In this paper, we apply correlation clustering before and after dimensionality reduction to enhance fault detection on the Tennessee Eastman Process. The reduction techniques employed are principal component analysis (PCA) and dynamic inner principal component analysis (DiPCA). This paper also introduces a novel index ($\delta_{\mathbf{S}^2}$) for monitoring the covariance of principal and residual components. Adding the novel index increases fault detection rates of PCA by more than 20% on 5 faults, and clustering before PCA increases fault detection by over 10% on 5 faults. This paper demonstrates two points: our clustering method can considerably increase fault detection rates, and the novel $\delta_{\mathbf{S}^2}$ index can boost performance on noisy faults, which were previously considered to be difficult to detect.

Keywords: Fault Detection, Clustering, Data-driven, Machine Learning, Covariance Analysis

1. INTRODUCTION

Fault detection in industrial processes aims to identify faults before they become serious enough to potentially cause damage. Several different methods may be employed to detect these faults. Recent research heavily focuses on applying technologies such as deep learning to fault detection (Zhang et al. (2020)). Though the use of principal component analysis is relatively old, it is still plays a role in fault detection research today (Márquez-Vera et al. (2021)). Clustering is another technique that is still utilized in the field of fault detection (Musa et al. (2019)). These two techniques have often been used together to enhance their performance. Performing PCA on correlation-based clusters shows promise in increasing the fault detection rate (FDR) and fault isolation accuracy (Smith and Powell (2019)). It has been shown that preprocessing data with clustering can also improve sensitivity (Zhang et al. (2019)). Clustering is also performed after PCA for data analysis and anomaly detection (Yeung and Ruzzo (2001)). However, much remains to be explored concerning the use of clustering in combination with PCA for industrial fault detection as there are many ways to combine the two methods.

The goal of this paper is to explore several new methods for combining correlation-based clustering with PCA as well as to present a novel fault statistic inspired by performing correlation-based clustering to a set of data prepared using PCA. This paper will compare the effects of performing clustering before PCA and clustering after PCA with traditional PCA techniques. A novel statistic to monitor covariance of the principal and residual spaces ($\delta_{\mathbf{S}^2}$) is also developed. This statistic was used to enhance basic PCA and PCA with clustering after.

To test if combining clustering with dimensionality reduction provides increased FDRs, we combine clustering with one of two reduction methods; PCA and DiPCA. Clustering is either performed before or after these reduction methods. We test the effectiveness of both, using the Tennessee Eastman Process, and include Canonical Variate Analysis (CVA) and Independent Component Analysis (ICA) as benchmarks.

The rest of the paper is organized as follows: Section 2 explains the clustering procedure, while section 3 describes dimensionality reduction techniques. Section 4 covers the statistics used and their upper control limits derived through Kernel Density Estimations. Section 5 describes the case study, and the results are presented and discussed in section 6. Finally, the work is concluded in section 7.

2. CLUSTERING

In this section, the agglomerative clustering method is introduced. Then the details of clustering before and after dimensionality reduction techniques are described.

2.1 Agglomerative Clustering of Correlation Matrix

Agglomerative clustering aims to produce a hierarchy of agglomerates that can be partitioned down to arbitrarily fine granularity. This is achieved by recursively grouping the pair of data points or clusters that are closest to one another according to a predetermined distance metric. Af-

10.1016/j.ifacol.2022.04.203

ter calculating a pairwise distance matrix(the correlation matrix), the algorithm is as follows:

(1) Locate the two points or groups that have the smallest distance between them.
(2) Define the location of the new node, halfway between the two objects.
(3) Update the distance matrix by removing the two grouped objects and replacing them with the calculated distance of their cluster from each of the other objects.
(4) Repeat the process until all of the objects are grouped into one agglomerate.

Step three requires a method (linkage type) to estimate the distance of the formed cluster to the other objects in the distance matrix. Ward linkage aims to minimize the total within-cluster variance when choosing which objects are grouped. This is accomplished by minimizing the sum squared difference between each point in the newly formed clusters.

Clustering was performed on the correlation matrix of the data sets. The correlation matrix is calculated by finding the Pearson correlation coefficient, r_{xy}, of each pair of variables (x and y) using Eq. 1.

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

2.2 Clustering before dimensionality Reduction (CB)

To perform fault detection using CB, the data is clustered, and dimensionality reduction is performed individually on each cluster. Since variables are clustered based on correlation, the data sets for each cluster will have maximal variance in one direction but little variance in any direction perpendicular to the trend. This means that the variance in each cluster will be more easily explained with fewer principal components. This allows reduction techniques to capitalize on smaller features of the plant that would have been lost, had dimensionality reduction been performed on the plant as a whole.

2.3 Clustering after dimensionality reduction (CA)

To perform CA, dimensionality reduction is performed, and clustering is done based on the variables in the eigenvector matrix, $\mathbf{V} \in R^{m\times m}$, produced by the decomposition of the covariance matrix shown in Eq. 2.

$$\mathbf{V} = \begin{array}{c} \\ \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_m \end{array} \begin{array}{c} \begin{array}{cccccccc} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_a & \tilde{\mathbf{p}}_1 & \tilde{\mathbf{p}}_2 & \cdots & \tilde{\mathbf{p}}_{m-a} \end{array} \\ \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{a1} & \tilde{p}_{11} & \tilde{p}_{21} & \cdots & \tilde{p}_{m-a1} \\ p_{12} & p_{22} & \cdots & p_{a2} & \tilde{p}_{12} & \tilde{p}_{22} & \cdots & \tilde{p}_{m-a2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{1m} & p_{2m} & \cdots & p_{am} & \tilde{p}_{1m} & \tilde{p}_{2m} & \cdots & \tilde{p}_{m-am} \end{bmatrix} \end{array} \tag{2}$$

Each $\mathbf{v}$ represents one of the m variables, each $\mathbf{p}$ represents one of the a principal vectors, and each $\tilde{\mathbf{p}}$ represents one of the m-a residual vectors. Clustering is performed separately on both the principal and residual components in order to maintain enough correlation to form valid clusters.

3. DIMENSIONALITY REDUCTION TECHNIQUES

In this section, the PCA and DiPCA algorithms are described.

3.1 Principal components analysis (PCA)

Given a normalized dataset $\mathbf{X} \in R^{n\times m}$ with m variables and n samples, PCA finds the set of vectors necessary to maximize the variance of the data along each principal component. To perform PCA, eigenvalue decomposition is done on the covariance matrix $\mathbf{S}$ as shown in Eq. 3.

$$\mathbf{S} = \frac{1}{n-1}\mathbf{X}^T\mathbf{X} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T \tag{3}$$

The diagonal matrix $\mathbf{\Lambda} \in R^{m\times m}$ contains the non-negative eigenvalues in decreasing order. The principal component matrix $\mathbf{P}$ is formed by choosing a eigenvectors in $\mathbf{V}$ associated with the a largest eigenvalues. The a largest eigenvalues are stored in the matrix $\mathbf{\Lambda}_a$. Thus, $\mathbf{P}$ and $\mathbf{\Lambda}_a$ can be used to transform $\mathbf{X}$ into a lower dimensionality.

Hotelling's T^2 statistic, Squared Prediction Error, and the novel covariance statistic are used to monitor faults. Thresholds are estimated using Kernel Density Estimation (KDE).

3.2 Dynamic inner PCA (DiPCA)

While PCA aims to maximize the variance in each extracted principal component, PCA only extracts static cross-correlations. DiPCA accounts for the dynamics in the data by extracting dynamic latent variables which maximize auto-covariance (Dong and Qin (2018)). The residual will have little to no auto-covariance and can be monitored using traditional PCA.

Denote $\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_{n+s}]^T$, where s is the dynamic order of the model. Using $\mathbf{X}$, Eqs 4 and 5 form $\mathbf{X_i}$ and $\mathbf{Z_s}$.

$$\mathbf{X_i} = [\mathbf{x}_i\ \mathbf{x}_i+1\ \cdots\ \mathbf{x}_{n+i-1}]^T \text{ for } i = 1, 2, \cdots, s+1 \tag{4}$$

$$\mathbf{Z_s} = [\mathbf{X}_1\ \mathbf{X}_1\ \cdots\ \mathbf{X}_s] \tag{5}$$

Extracting the latent variables that maximize the auto-covariance can be formulated as Eqs 6 and 7.

$$\mathbf{w}, \boldsymbol{\beta} = \arg\max_{\mathbf{w},\boldsymbol{\beta}} \quad \mathbf{w}^T\mathbf{X}_{s+1}^T\mathbf{Z}_s(\boldsymbol{\beta} \otimes \mathbf{w}) \tag{6}$$

$$\text{s.t.} \quad ||\mathbf{w}|| = 1,\ ||\boldsymbol{\beta}|| = 1 \tag{7}$$

Eqs 6 and 7 can be solved using the algorithm given in Dong and Qin (2018). After obtaining $\mathbf{w}$, the dynamic latent variable $\mathbf{t}$ and loading vector $\mathbf{p}$ can be calculated with Eqs 8 and 9. $\mathbf{X}$ can then be deflated with Eq 10.

$$\mathbf{t} = \mathbf{X}\mathbf{w} \tag{8}$$

$$\mathbf{p} = \mathbf{X}^T\mathbf{t}/\mathbf{t}^T\mathbf{t} \tag{9}$$

$$\mathbf{X} := \mathbf{X} - \mathbf{t}\mathbf{p}^T \tag{10}$$

After solving Eqs 6 and 7 and deflating $\mathbf{X}$ with Eq. 10 l times, l latent variables and loading vectors are extracted. Let $\mathbf{T} = [\mathbf{t_1}\ \mathbf{t_2}\ \cdots\ \mathbf{t_l}]^T$ where $\mathbf{t}_j$, $j = 1, 2, \cdots, l$, is the jth latent variable. $\mathbf{T_i}$ can then be formed from $\mathbf{T}$ similarly how $\mathbf{X_i}$ is formed from $\mathbf{X}$. An estimate for the vector autoregressive (VAR) model can be formulated using Eqs. 11, 12, and 13.

$$\bar{\mathbf{T}}_s = [\mathbf{T_1}\ \mathbf{T_2}\ \cdots\ \mathbf{T_s}] \tag{11}$$

$$\hat{\mathbf{\Theta}} = \left(\bar{\mathbf{T}}_s^T\bar{\mathbf{T}}_s\right)^{-1}\bar{\mathbf{T}}_s^T\mathbf{T}_{s+1} \tag{12}$$

$$\hat{\mathbf{T}}_{s+1} = \bar{\mathbf{T}}_s\hat{\mathbf{\Theta}} \tag{13}$$

Let $\mathbf{W} = [\mathbf{w_1}\ \mathbf{w_2}\ \cdots\ \mathbf{w_l}]$ and $\mathbf{P} = [\mathbf{p_1}\ \mathbf{p_2}\ \cdots\ \mathbf{p_l}]$. With $\mathbf{W}$, $\mathbf{P}$, and the VAR model, dynamic residuals $\mathbf{v_k}$ and

prediction errors $\mathbf{e_k}$ can be calculated using Eqs. 14, 15, 16, and 17.

$$\mathbf{R} = \mathbf{W}\left(\mathbf{PW}\right)^{-1} \tag{14}$$

$$\mathbf{t_k} = \mathbf{R}^T\mathbf{x}_k \tag{15}$$

$$\mathbf{v_k} = \mathbf{t_k} - \hat{\mathbf{t}}_k;\ \hat{\mathbf{t}}_k = \sum_{i=1}^{s} \hat{\mathbf{\Theta}}_i^T \mathbf{t}_{k-i} \tag{16}$$

$$\mathbf{e_k} = \mathbf{x_k} - \mathbf{P}\hat{\mathbf{t_k}} \tag{17}$$

$\hat{\mathbf{\Theta}}_i$ is formed from $\hat{\mathbf{\Theta}}$ similarly how $\mathbf{X_i}$ is formed from $\mathbf{X}$.

Standard PCA is used to monitor $\mathbf{v}_k$ and $\mathbf{e}_k$. The combined index defined in Yue and Qin (2001) is used to monitor $\mathbf{v}_k$ and is shown in Eq. 18 and 19.

$$\phi_v = \mathbf{v}_k^T \mathbf{\Phi}_v \mathbf{v}_k \tag{18}$$

$$\mathbf{\Phi}_v = \frac{\mathbf{P}_v \boldsymbol{\lambda}_v^{-1} \mathbf{P}_v^T}{\chi_v^2} + \frac{\mathbf{I} - \mathbf{P}_v\mathbf{P}_v^T}{\delta_v^2} \tag{19}$$

$\mathbf{P}_v$ and $\boldsymbol{\lambda}_v$ are retained loading vectors and retained eigenvalues, respectively; χ_v^2 and δ_v^2 are the control limits for T^2 and Q, respectively. $\mathbf{e}_k$ is monitored using the T_r^2 and Q_r statistics. Control limits are determined using KDE.

4. FAULT STATISTICS

This section presents the statistics and distributions used to determine if a fault has occurred.

4.1 Fault indices

Hotelling's T^2. Hotelling's T^2 is a multivariate statistical method used to determine the likelihood that a sample was drawn from a population by finding the square Mahalanobis distance between their multivariate means as shown in Eq. 20.

$$t^2 = (\bar{\mathbf{x}} - \boldsymbol{\mu})^T \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \tag{20}$$

The loadings matrix obtained from PCA represents the vector means of the data. Thus, the T^2 statistic for any data point can be determined with Eq. 21

$$t^2 = \mathbf{x}^T \mathbf{P} \mathbf{\Lambda}_a^{-1} \mathbf{P}^T \mathbf{x} \tag{21}$$

Squared prediction error. The squared prediction error estimates how far into the residual space a data point lies. Projecting the observed variables onto the principle space like in Eq. 22 gives the latent variables.

$$\mathbf{z} = \mathbf{P}^T\mathbf{x} \tag{22}$$

$\mathbf{x}$ can be reconstructed as $\hat{\mathbf{x}}$ from $\mathbf{z}$ using Eq. 23

$$\hat{\mathbf{x}} = \mathbf{Pz} = \mathbf{PP}^T\mathbf{x} \tag{23}$$

The prediction error, r, is the difference between the reconstructed and observed variables as shown in Eq.24.

$$\mathbf{r} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} - \mathbf{PP}^T\mathbf{x} = \left(I - \mathbf{PP}^T\right)\mathbf{x} \tag{24}$$

The square of the prediction error is used to account for negatives as in Eq. 25.

$$\begin{aligned} Q &= \mathbf{r}^T\mathbf{r} = \mathbf{x}^T\left(I - \mathbf{PP}^T\right)^T\left(I - \mathbf{PP}^T\right)\mathbf{x} \\ &= \mathbf{x}^T\left(I - \mathbf{PP}^T\right)\mathbf{x} \end{aligned} \tag{25}$$

Covariance statistic. The covariance statistic, $\delta_{\mathbf{S}^2}$, calculates the square difference between the covariance of the transformed variables of the training or past data $\mathbf{X}$ and the incoming (testing) data $\mathbf{Y}$.

$$\delta_{S^2} = \left(\frac{1}{q-1}\mathbf{Y}^T\mathbf{Y} - \frac{1}{n-1}\mathbf{X}^T\mathbf{X}\right)^2 \tag{26}$$

$\mathbf{Y} \in R^{q\times a}$ represents the PCA-transformed data from the past q timesteps of the incoming data and $\mathbf{X} \in R^{n\times a}$ represents the PCA-transformed training data. This is done on both the principle and residual space separately. If a fault generates a notable increase in noise, it will register even if an individual point would not.

4.2 Kernel density estimation (KDE)

Upper control limits (UCL) serve as thresholds between faulty and normal data. In general, UCLs for fault detection are determined analytically. This method relies on the assumption that the underlying process produces data that follows a Gaussian distribution. These solutions are not valid for nonlinear processes under varying conditions. Kernel Density Estimation (KDE) is used to estimate the distribution to accommodate for varying processes as it can provide a more adaptive fit to the data (Pilario and Cao (2018); Odiowei and Cao (2009)).

The kernel used for this paper is the Gaussian kernel.

$$K(g) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{g^2}{2}\right) \tag{27}$$

Finding the UCL such that $P(x < UCL) = \alpha$ is equivalent to finding the upper limit to the integral representing area under the distribution such that the area is equal to α.

$$P(x < UCL) = \int_{-\infty}^{UCL} \frac{1}{Mh} \sum_{k=1}^{M} K\left(\frac{x - x_k}{h}\right) dx = \alpha \tag{28}$$

x_k, $k = 1, 2, 3 \ldots, M$ are the samples of x, and h is the kernel bandwidth.

For online monitoring, the statistics are continuously computed from real-time data, and a fault is registered whenever one of the statistics has a value higher than the UCL.

5. APPLICATION AND EVALUATION

The Tennessee Eastman Process (TEP) dataset was created to provide a realistic industrial process to evaluate process control and monitoring methods (Chiang et al. (2000)). The process contains 5 units: a reactor, a condenser, a compressor, a separator, and a stripper. The process converts four reactants into two products and a byproduct in the presence of an inert, for a total of eight components denoted as A, B, C, D, E, F, G, and H. The process explored is operated under closed-loop control and has 52 variables. 41 of these variables are measured, while the remaining 11 are manipulated variables. Of the 41 measured variables, 22 are process measurements with a sample interval of 3 minutes. The remaining measured variables are composition measurements and have sample intervals that vary from six to fifteen minutes. The indices and names for these variables can be found in Li et al. (2010). A detailed process description can be found in Chiang et al. (2000).

5.1 Evaluated algorithms and metrics

To benchmark the developed methods, PCA and DiPCA are applied using a combined Hotelling's T^2 and SPE statistic to the TEP three times: once without clustering,

Table 1. FDRs (%) and FARs (%) of combined T^2 and *SPE* statistics for detection methods with no clustering (NC), clustering before (CB), clustering after (CA), and novel $\delta_{\mathbf{S}^2}$ statistic.

	PCA					DiPCA			ICA	CVA
	NC	NC+$\delta_{\mathbf{S}^2}$	CB	CA	CA+$\delta_{\mathbf{S}^2}$	NC	CB	CA	NC	NC
Fault 1	**100**	**100**	99.88	99.75	**100**	**100**	**100**	89.00	**100**	**100**
Fault 2	98.63	98.63	98.50	98.63	98.63	98.49	98.87	96.75	98.38	**98.87**
Fault 3	3.00	4.50	11.00	10.25	11.25	3.52	13.60	**22.88**	1.50	2.26
Fault 4	**100**	**100**	**100**	**100**	**100**	**100**	**100**	62.25	**100**	**100**
Fault 5	**100**	**100**	**100**	**100**	**100**	29.18	**100**	99.75	**100**	**100**
Fault 6	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Fault 7	**100**	**100**	**100**	**100**	**100**	99.75	**100**	68.00	**100**	**100**
Fault 8	97.88	97.88	98.50	98.38	98.38	98.87	**98.99**	93.13	97.88	98.49
Fault 9	3.25	9.25	10.50	10.25	16.13	3.90	17.88	**24.13**	1.00	2.52
Fault 10	70.63	91.63	87.13	83.75	**92.88**	58.11	81.86	66.75	85.38	90.44
Fault 11	81.38	95.25	88.25	85.25	**96.38**	84.53	94.84	66.25	78.00	92.96
Fault 12	99.13	**100**	**100**	99.50	**100**	99.75	**100**	98.38	99.88	**100**
Fault 13	95.50	95.50	95.75	95.00	95.00	94.47	95.09	95.25	95.25	**95.97**
Fault 14	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
Fault 15	5.13	6.75	15.13	13.13	14.63	3.65	16.12	**31.25**	2.88	22.64
Fault 16	67.75	90.00	93.25	86.13	**95.88**	61.26	82.24	95.75	89.13	92.08
Fault 17	97.13	97.88	97.13	96.75	97.75	89.43	97.36	75.25	96.63	**98.11**
Fault 18	90.63	90.63	91.38	91.00	91.00	89.81	91.56	**93.50**	90.00	90.94
Fault 19	68.63	96.13	92.88	41.38	92.75	20.75	81.74	69.50	91.88	**100**
Fault 20	68.25	90.63	88.63	66.75	91.00	51.95	90.43	76.63	90.88	**91.70**
Fault 21	47.63	**87.75**	52.63	48.75	**87.75**	7.04	45.21	46.38	48.38	48.05
FAR	2.40	3.02	9.70	5.02	5.23	2.40	14.29	14.81	**1.46**	2.09

once with CB, and once with CA. The novel $\delta_{\mathbf{S}^2}$ statistic is also applied to PCA with no clustering and clustering after. Independent Component Analysis (ICA) and Canonical Variate Analysis (CVA) are also evaluated as baselines. These methods are evaluated based on their capability to detect 21 faults whose indices and descriptions can be found in Yin et al. (2012). For this case study, the training data is gathered from Rieth et al. (2017) and the testing data is from Chiang et al. (2000). The 22 process measurements and 11 manipulated variables are used as $\mathbf{X}$.

The methods are evaluated on FDRs and FARs. Since there are many test statistics, the condition for a fault is if one test statistic exceeds the threshold.

6. RESULTS

6.1 Performance Analysis

Table 1 shows the FDRs and FARs of the novel methods, which includes PCA with no clustering and the novel $\delta_{\mathbf{S}^2}$ statistic (PCA-NC+$\delta_{\mathbf{S}^2}$), PCA with clustering before (PCA-CB), PCA with clustering after (PCA-CA), PCA with clustering after and the novel $\delta_{\mathbf{S}^2}$ statistic (PCA-CA+$\delta_{\mathbf{S}^2}$), DiPCA with clustering before (DiPCA-CB), and DiPCA with clustering after (DiPCA-CA). PCA (PCA-NC), DiPCA (DiPCA-NC), CVA, and ICA with no clustering are included for comparison.

The $\delta_{\mathbf{S}^2}$ statistic notably increases FDRs for PCA with no clustering and clustering after. For PCA-NC, adding the novel $\delta_{\mathbf{S}^2}$ statistic increases FDRs by more than 20% for faults 10, 16, 19, 20, and 21. The novel $\delta_{\mathbf{S}^2}$ statistic only increases the FAR by 0.62%.

PCA-CA+$\delta_{\mathbf{S}^2}$ is competitive with CVA as PCA-CA+$\delta_{\mathbf{S}^2}$ raises FDRs on faults 10, 11, and 16 by at least 2% and increases the FDR for fault 21 by nearly 40 %. Excluding faults 15 and 19, PCA-CA+$\delta_{\mathbf{S}^2}$ only decreases FDRs by 1% compared to CVA. However, the FAR does increase by 3%. While PCA-CA+$\delta_{\mathbf{S}^2}$ performs relatively well on faults 3, 9, and 15, the FARs in these fault datasets increase, suggesting PCA-CA amplifies noise for these datasets. We note that PCA-CA+$\delta_{\mathbf{S}^2}$ combination has higher FDRs compared to PCA-NC+$\delta_{\mathbf{S}^2}$ for nine of the faults. This suggests that PCA-CA has difficulty detecting faults detected by the $\delta_{\mathbf{S}^2}$ statistic as compared to PCA-NC. Thus, when combined, PCA-CA+$\delta_{\mathbf{S}^2}$ has higher FDRs than PCA-NC+$\delta_{\mathbf{S}^2}$.

While clustering before typically increases FDRs, it also greatly increases the FAR. As seen with PCA-CB, the FDRs for six faults increase by more than 5% compared to PCA-NC, but the FAR increases to 9.7%. Similarly, in DiPCA, DiPCA-CB increases FDRs for eight faults by more than 5 % as compared to DiPCA-NC but increases the FAR to almost 15%.

For PCA, clustering after increases FDRs as PCA-CA's FDRs for faults 10 and 16 are more than 10% greater than PCA-NC's FDRs. The FAR only increases by 2.6%. However, all other increases are marginal, and for five faults, the FDRs drop. For DiPCA, clustering after is overall detrimental to fault detection as the FAR increases by 12.4% and decreases the FDR for a third of the faults.

6.2 Novel covariance statistic appreciably increases FDRs for PCA

As shown in Fig. 1, the novel $\delta_{\mathbf{S}^2}$ statistic can considerably increase the FDR of more subtle faults where the fault does not directly influence the position of each individual point but increases noise in the data.

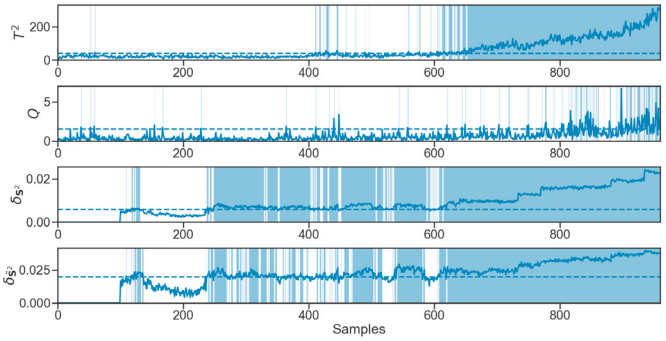

Fig. 1. Fault detection charts for fault 21 with PCA and novel covariance statistics

6.3 Analysis of clustering before

Clustering before yields groupings that correspond to plant relations. Agglomerative clustering of the TEP variables based on correlation is shown in Fig. 2. Fig. 2a shows the dendrogram, which shows how dissimilar the variables are. The lower the branch, the more similar the variables are. This corresponds to a brighter spot in the heatmap, as shown in Fig. 2b. Cluster 1 contains variables relating to the units' pressures, the compressor work and recycle valve, and inlet streams 1 and 3 (components A and E, respectively). Cluster 2 contains variables relating to the reactor's temperature and inlet stream 2 (component D). Cluster 3 contains the variables relating to the remaining streams to the reactor (streams 4, 6, 8, and 10), the reactor level, stripper temperature, the purge stream (stream 9), and the separator temperature. Cluster 4 has variables relating to the condenser cooling water flow, the separator level, and the stripper level. While cluster 1 relates to variables that affect pressure and cluster 2 relates to variables that affect the reactor temperature, the last two clusters are less obvious. Fig. 2a shows cluster 3 has two smaller clusters captured inside and correspond to reactor level, and feed streams and stripper and separator temperatures and cluster 4 is composed of 3 small clusters.

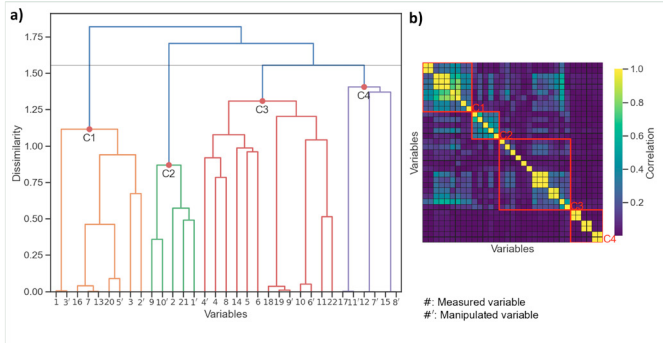

Fig. 2. Agglomerative clustering of process variables based on correlation

Clustering before increases FDR of residual space faults. Clustering decreases the number of variables that affect the location of a data point, and thus, each of the variables in the cluster has a stronger impact on the behavior of the cluster. There are two effects that stem from this phenomenon: fault migration to the principal space and amplified scores of faulty data. Faults 10 and

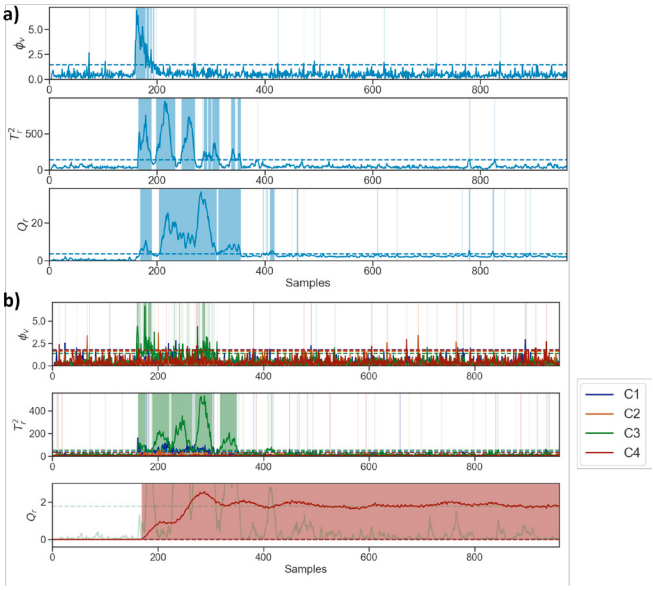

Fig. 3. Fault detection charts for fault 5 with a) DiPCA and no clustering, and b) DiPCA and clustering before.

16 are examples of faults that are generally detected in the residual space, but clustering moved them to the principal space. Fig. 3 shows the amplification of fault 5 in the residual space due to a decrease in the number of variables. In Fig. 3a, DiPCA-NC is shown to miss the fault since the Q_r index is just below the limit. When clustering before is applied, as seen in Fig 3b, the fault is again amplified and thus, clearly registered as a fault in cluster 4.

6.4 Analysis of clustering after

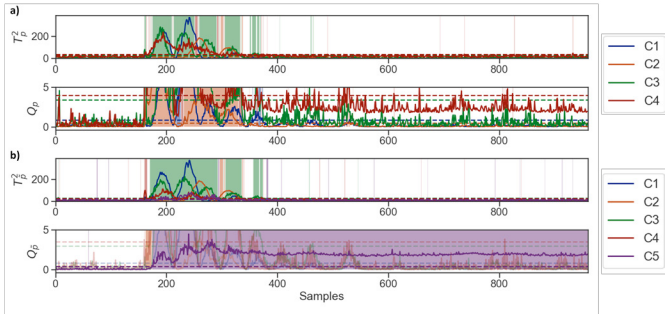

Fig. 4. Fault detection charts for fault 5 with PCA and clustering after based on a) principal components and b) residual components.

CA only shows benefits in the residual space. Aside from a few special cases, clustering after on the principal components gave results that were equal to or worse than PCA alone. Any benefit from clustering after comes from clustering on the residual matrix. The residual components have the lowest associated eigenvalues and thus, have the highest potential for correlation. The clusters attained from the residual space even showed a higher correlation to actual processes in the plant than those taken from the principal space. Fig. 4 shows the performance of each statistic when clusters were performed on the principal and residual space. When

clusters were obtained using the residual space, the Q statistic shows a clear fault in the correct cluster that is sustained above the threshold for the entirety of the fault. Clusters formed using the principal components show the fault being mostly missed as the last cluster had too many variables.

7. CONCLUSION

In this paper, we investigate the effects of combining clustering with dimensional reduction techniques. We also augment fault detection methods with a novel $\delta_{\mathbf{S}^2}$ statistic to monitor changes to the covariance. Our results show that (1) including the novel $\delta_{\mathbf{S}^2}$ statistic increases fault detection rates (FDRs) considerably with small increases in false alarm rates (FARs) for PCA and PCA-CA, (2) Clustering before PCA and DiPCA increases FDRs by intensifying the indices' response to faults, and (3) Clustering after PCA increases FDRs over PCA alone, but only when clustering is done on the residual components.

(1) The novel $\delta_{\mathbf{S}^2}$ statistic increases FDRs. This first finding is supported by the tests performed on the TEP. Adding the $\delta_{\mathbf{S}^2}$ statistic increases six FDRs for PCA-NC by more than 10% and seven FDRs for PCA-CA by more than 5%. The FARs were not seriously increased as they increased 0.62% and 0.21% for PCA-NC and PCA-CA, respectively.

(2) PCA and DiPCA with clustering before intensify indices responses to faults. This finding is supported by the FDRs of PCA-CB and DiPCA-CB on the TEP and fault detection charts for fault 5. Tests on the TEP show five indices from PCA and ten indices from DiPCA increase by at least 10% with clustering before compared to no clustering. The PCA-CB's fault detection chart for fault 5 shows the ratio between the index and the threshold is drastically increased.

(3) PCA with clustering after has higher FDRs when clustering based on the residual component. This last finding is supported by PCA-CA's fault detection chart for fault 5. In general, clustering the principle components decreases the FDR while clustering on the residual gives an FDR higher than that of PCA alone.

This paper's covariance monitoring results support findings in Musa et al. (2019), that monitoring the covariance increases how many varieties of faults are consistently registered as faults. Our results for PCA-CB and DiPCA-CB agree with Smith and Powell (2019), who further show clustering before reduces contribution smearing for fault isolation. Our paper shows that monitoring the principal and residual covariances helps detect specific types of faults and may be applicable for other dimensional reduction techniques. While our clustering after method only marginally increases FDRs, it offers a new approach to fault detection and has many options for improvement.

Future work includes: using different metrics to cluster after PCA, and clustering based on PCA-transformed data rather than the loadings matrix.

ACKNOWLEDGEMENTS

This work was supported by funding from the Undergraduate Research Opportunities Program at the University of Utah.

REFERENCES

Chiang, L.H., Russell, E.L., and Braatz, R.D. (2000). Fault diagnosis in chemical processes using fisher discriminant analysis, discriminant partial least squares, and principal component analysis. *Chemometrics and intelligent laboratory systems*, 50(2), 243–252.

Dong, Y. and Qin, S.J. (2018). A novel dynamic pca algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67, 1–11.

Li, G., Alcala, C.F., Qin, S.J., and Zhou, D. (2010). Generalized reconstruction-based contributions for output-relevant fault diagnosis with application to the tennessee eastman process. *IEEE transactions on control systems technology*, 19(5), 1114–1127.

Márquez-Vera, M.A., López-Ortega, O., Ramos-Velasco, L.E., Rodríguez-Romero, A., Ramos-Fernández, J.C., and Hernández-Salazar, J.A. (2021). Adaptive threshold pca for fault detection and isolation. *Journal of Robotics and Control (JRC)*, 2(3), 119–125.

Musa, M.H., He, Z., Fu, L., and Deng, Y. (2019). A covariance indices based method for fault detection and classification in a power transmission system during power swing. *International Journal of Electrical Power & Energy Systems*, 105, 581–591.

Odiowei, P.E.P. and Cao, Y. (2009). Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations. *IEEE Transactions on Industrial Informatics*, 6(1), 36–45.

Pilario, K.E.S. and Cao, Y. (2018). Canonical variate dissimilarity analysis for process incipient fault detection. *IEEE Transactions on Industrial Informatics*, 14(12), 5308–5315.

Rieth, C.A., Amsel, B.D., Tran, R., and Cook, M.B. (2017). Additional Tennessee Eastman Process Simulation Data for Anomaly Detection Evaluation. Harvard Dataverse, 1. doi:10.7910/DVN/6C3JR1. URL `https://doi.org/10.7910/DVN/6C3JR1`. (Accessed 22 May 2021).

Smith, A.J. and Powell, K.M. (2019). Fault detection on big data: A novel algorithm for clustering big data to detect and diagnose faults. *IFAC-PapersOnLine*, 52(10), 328–333.

Yeung, K.Y. and Ruzzo, W.L. (2001). Principal component analysis for clustering gene expression data. *Bioinformatics*, 17(9), 763–774.

Yin, S., Ding, S.X., Haghani, A., Hao, H., and Zhang, P. (2012). A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark tennessee eastman process. *Journal of process control*, 22(9), 1567–1581.

Yue, H.H. and Qin, S.J. (2001). Reconstruction-based fault identification using a combined index. *Industrial & engineering chemistry research*, 40(20), 4403–4414.

Zhang, H., Chen, H., Guo, Y., Wang, J., Li, G., and Shen, L. (2019). Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on pca and preprocessed by combined clustering. *Applied Thermal Engineering*, 160, 114098.

Zhang, X., Zhou, J., and Chen, W. (2020). Data-driven fault diagnosis for pemfc systems of hybrid tram based on deep learning. *International Journal of Hydrogen Energy*, 45(24), 13483–13495.