Contribution Plots based Fault Diagnosis of a Multiphase Flow Facility with PCA-enhanced Canonical Variate Analysis

Ruomu Tan and Yi Cao
School of Water, Energy and Environment
Cranfield University
Cranfield, UK MK10 0AL
Emails: R.Tan@cranfield.ac.uk; Y.Cao@cranfield.ac.uk

***Abstract*—Process monitoring plays a vital role in order to sustain optimal operation and maintenance of the plant in process industry. As an essential stage in process monitoring, data-driven fault detection and diagnosis techniques have evolved quickly owing to the prosperity of multivariate feature extraction methods. In addition to the application of basic feature extraction methods, hybrid algorithms combining different methods have also been invented for better monitoring performance. In the meantime, little study has been done towards the fault diagnosis techniques under this 2-stage feature extraction framework. To deal with complex faults which will have impact on multiple process variables and the relationships among them, the Principal Component Analysis (PCA) enhanced Canonical Variate Analysis (CVA) based fault detection and diagnosis algorithm is investigated in this paper. PCA is used to pre-process the raw measurements and extracts the principal components as better indicators of process condition; CVA is conducted sequentially to further project the principal components to canonical variate space and the detection statistics are calculated based on these canonical variates. When a fault has been detected, the contributions of original process variables in monitoring statistics are derived to identify influential variables and locate the fault. To validate, along with other multivariate statistical monitoring techniques, this PCA-enhanced CVA algorithm is applied to a benchmark data set collected from an industrial scale multiphase flow facility in Cranfield University for performance evaluation.**

Keywords*—*multivariate process monitoring, fault detection and diagnosis, contribution plots, feature extraction

I. Introduction

Multivariate statistical feature extract method has been the trend in data-driven process monitoring for over a decade and is still prosperous nowadays. Apart from the most commonly used Principal Component Analysis (PCA) and its kin, Canonical Variate Analysis (CVA) has also been studied intensively and applied to various aspects of process monitoring [1]–[3]. CVA is a powerful tool with not only the ability of dynamic process monitoring but also establishing state space model from data. Since monitoring algorithms based on PCA and its dynamic extensions suffer from the inaccuracy of the model they build, introducing CVA will compensate and improve model quality. Simultaneously, canonical variates and residuals generated from principal components instead of raw data can be better representations of the process dynamics as well as random errors. Previously, Samuel and Cao [4] have established a PCA-enhanced latent variable CVA based fault detection method and applied it to the Tennessee Eastman challenge process.

In addition to fault detection, this PCA-enhanced CVA algorithm is also anticipated to reinforce fault diagnosis performance and provide better insights for process operation and maintenace. However, unlike the rapid development of hybrid methods in fault detection, a knowledge gap is identified that these methods lack a commonly acknowledged approach for fault identification and diagnosis. As an example of data-driven fault diagnosis methods, contribution plots [5] have been extensively applied to identifying variables associating to a certain fault and locating the fault. Rooting in the idea that the process variable that the fault has a significant impact on is supposed to have larger contribution to the monitoring statistics, the study of contribution plots is persistent and profound. To fill in the gap, this work utilizes contribution plots and establishes a general structure of contribution propagation for fault diagnosis based on 2-stage feature extraction process monitoring techniques. Moreover, the contribution propagation solution to this PCA-enhanced CVA algorithm is derived and validated via identifying influential variables in an industrial case study. It is reasonable to infer that contribution of influential variables will be properly emphasized if contributions transmitted through multiple levels of feature extraction methods have been quantitatively analysed.

The remainder of this paper is organized as follows. The fault detection algorithm of PCA-enhanced CVA monitoring approach is revisited in Section II. For fault diagnosis, a general formulation of contribution propagation is proposed for contribution plots calculation of a type of 2-stage feature extraction based monitoring techniques, to which the PCA-enhanced CVA algorithm belongs, in Section III in order to identify the process variable with large contribution with respect to the monitoring statistics. The contribution plots for the PCA-enhanced CVA monitoring technique are derived under this framework. Section IV presents a case study on the benchmark data set collected from the multiphase flow facility. In this case study, the fault detection and diagnosis performance of PCA-enhanced CVA is compared with various linear multivariate statistical methods. The applicability of PCA-enhanced CVA algorithm is further verified by its application to the data from large scale real-life process in addition to simulated benchmark data sets. Section V summarizes the findings in this work and illustrates potential directions of extension in future study.

II. PCA-ENHANCED CVA PROCESS MONITORING REVISIT

A. 2-stage PCA-enhanced CVA Feature Extraction Method

According to [4], CVA fault detection algorithm is enhanced by using latent variables extracted by PCA as its input instead of the original measured variables. Firstly, PCA projects the original measurement data in v-dimensional variable space to a reduced r-dimensional principal component space with maximum explanation of variations in original variables. The model structure of PCA is illustrated as follows:

$$Y = XP \tag{1}$$

where $X \in \mathbb{R}^{n\times v}$ with zero mean and unit variance is the standardized original data set, $Y \in \mathbb{R}^{n\times r}$ is the extracted principal components and $P \in \mathbb{R}^{v\times r}$ is the projection matrix. The projection matrix P is obtained by eigenvalue decomposition of sample covariance matrix X^TX. Hence the principal component vector $\boldsymbol{y}$ is linear projection of original variable vector $\boldsymbol{x}$.

CVA is a linear dynamic feature extraction method from which the canonical variates with maximum correlation between past and future vectors can be acquired. Instead of using original data matrix X, PCA-enhanced CVA algorithm adopts the principal components Y extracted by PCA as the input to CVA and get canonical variates $Z \in \mathbb{R}^{n\times d}$ and residuals $E \in \mathbb{R}^{n\times r}$.

At certain time stamp t, past and future vectors $\boldsymbol{y_p}(t)$ and $\boldsymbol{y_f}(t)$ are formed by Eqn 2 with fixed vector lengths p and f:

$$\begin{aligned} \boldsymbol{y_p}(t) &= [\tilde{\boldsymbol{y}}^T(t-1), \tilde{\boldsymbol{y}}^T(t-2), ..., \tilde{\boldsymbol{y}}^T(t-p)]^T \\ \boldsymbol{y_f}(t) &= [\tilde{\boldsymbol{y}}^T(t), \tilde{\boldsymbol{y}}^T(t+1), ..., \tilde{\boldsymbol{y}}^T(t+f)]^T \end{aligned} \tag{2}$$

where $\tilde{\boldsymbol{y}}_(t) = \boldsymbol{y}(t) - \bar{\boldsymbol{y}}$ such that $\bar{\boldsymbol{y}}$ is the mean of principal components $\boldsymbol{y}$ over time.

Furthermore, past and future Henkel matrices $Y_p \in \mathbb{R}^{rp\times m}$ and $Y_f \in \mathbb{R}^{rf\times m}$ comprise of $m = n - p - f + 1$ past/future vector pairs, making the time lagged data matrices for feature extraction starting at time stamp p (the minimal initial time point for constructing past vector):

$$\begin{aligned} Y_p &= \boldsymbol{y_p}(p), \boldsymbol{y_p}(p+1), ..., \boldsymbol{y_p}(p+m-1)] \\ Y_f &= [\boldsymbol{y_f}(p), \boldsymbol{y_f}(p+1), ..., \boldsymbol{y_f}(p+m-1)] \end{aligned} \tag{3}$$

Analogically to PCA, the quasi-covariance matrix H is defined by covariance and cross-covariance matrices of Y_p and Y_f:

$$\Sigma_{pp} = Y_p^TY_p; \quad \Sigma_{ff} = Y_f^TY_f; \quad \Sigma_{fp} = Y_f^TY_p \tag{4}$$

$$H = \Sigma_{ff}^{-\frac{1}{2}}\Sigma_{fp}\Sigma_{pp}^{-\frac{1}{2}} \tag{5}$$

Consequentially, The projection matrices J and L are the normalized results of singular value decomposition result of H:

$$H = U\Lambda V^T \tag{6}$$

$$J = V_d\Sigma_{pp}^{-\frac{1}{2}}; \quad L = (I_r - V_dV_d^T)\Sigma_{pp}^{-\frac{1}{2}} \tag{7}$$

The canonical variate vector $\boldsymbol{z}$ and residual vector $\boldsymbol{e}$ are both linear projections of past vector $\boldsymbol{y}_p$ at time t:

$$\boldsymbol{z}(t) = J\boldsymbol{y}_p(t); \quad \boldsymbol{e}(t) = L\boldsymbol{y}_p(t) \tag{8}$$

B. Fault Detection with PCA-enhanced CVA

After features have been extracted from original data, monitoring statistics are to be calculated using these features and compared with their control limits for fault detection. Qin [6] has studied a variety of monitoring metrics in data-driven process monitoring. The most widely used ones among all are the T^2 statistics for detection of systematic variation and Q statistics for random error. For CVA, T^2 and Q statistics are calculated in the canonical variate space and residual space, respectively.

$$T^2(t) = \boldsymbol{z}^T(t)\boldsymbol{z}(t); \quad Q(t) = \boldsymbol{e}^T(t)\boldsymbol{e}(t) \tag{9}$$

Based on normal data, upper control limits with confidence level α, i.e. $T^2_{UCL}(\alpha)$ and $Q_{UCL}(\alpha)$, of statistics in Eqn 9 are defined as:

$$P(T^2 > T^2_{UCL}(\alpha)) = \alpha; \quad P(Q > Q_{UCL}(\alpha)) = \alpha \tag{10}$$

Due to the potential non-Gaussianity of the process variables, the distribution functions in Eqn 10 and corresponding control limits are estimated via Kernel Density Estimation [7]. In online fault detection, monitoring statistics calculated with the real-time measurements are compared with these control limits to determine the fault occurrence based on the condition shown in Eqn 11.

$$\left(T^2(t) > T^2_{UCL}\right) || \left(Q(t) > Q_{UCL}\right) \tag{11}$$

III. CONTRIBUTION PLOTS BASED FAULT DIAGNOSIS OF PCA-ENHANCED CVA

This section discusses the general formulation of contribution plots under 2-stage feature extraction framework and derives the solution to PCA-enhanced CVA method.

A. Contribution Plots for 2-stage Feature Extraction Based Monitoring Techniques

Eqn 12 formulates the structure of a general 2-stage feature extraction based fault detection method.

$$
\begin{aligned}
\boldsymbol{y} = & g(\boldsymbol{x}; P) \\
\boldsymbol{z} = f_1(\boldsymbol{y}; J); & \quad \boldsymbol{e} = f_2(\boldsymbol{y}; L) \\
T^2 = \boldsymbol{z}^T\boldsymbol{z} & \quad Q = \boldsymbol{e}^T\boldsymbol{e}
\end{aligned} \tag{12}
$$

where

- $\boldsymbol{x}$: original measured variables
- P, g: first layer parameters and model structure g is not used here.
- $\boldsymbol{y}$: intermediate features from first layer
- J, f_1: second layer parameters and model structure for feature variables representing systematic error
- L, f_2: second layer parameters and model structure for residuals representing random error
- $\boldsymbol{z}, \boldsymbol{e}$: features from second layer representing systematic and random error
- T^2, Q: testing statistics for systematic and random error

In this formulation, the original data set $X = [\boldsymbol{x}(1), \boldsymbol{x}(2), ..., \boldsymbol{x}(n)]^T$ is initially processed by the first feature extraction method to obtain the intermediate feature data $Y = [\boldsymbol{y}(1), \boldsymbol{y}(2), ..., \boldsymbol{y}(n)]^T$. Y is further processed by the second layer of feature extraction method to attain the feature variables $Z = [\boldsymbol{z}(1), \boldsymbol{z}(2), ..., \boldsymbol{z}(n)]^T$ and residual variables $E = [\boldsymbol{e}(1), \boldsymbol{e}(2), ..., \boldsymbol{e}(n)]^T$. The final monitoring statistics are based on $\boldsymbol{z}$ and $\boldsymbol{e}$. It is obvious that aforementioned PCA-enhanced CVA algorithm falls into this category.

The objective of contribution plots based fault diagnosis is to gain the contribution of original process variables to the final monitoring statistics such as T^2 and Q. In order to do so, Figure 1 illustrates the propagation of variable contributions under this 2-stage feature extraction framework.

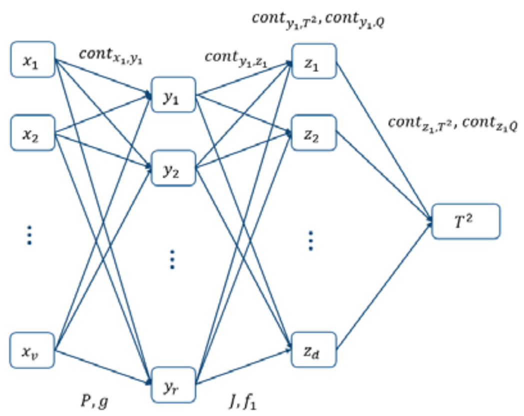


Fig. 1. Illustration of 2-stage contribution propagation in T^2

The following equations hold for individual variables $x_i \in \boldsymbol{x}$, $y_j \in \boldsymbol{y}$ and $z_k \in \boldsymbol{z}$: $\sum_{i=1}^{v} cont_{x_i,y_j} = y_j$; $\sum_{j=1}^{r} cont_{y_j,z_k} = z_k$; $cont_{z_k,T^2} = z_k^T z_k$. Analogy can be made for the contribution plots to Q statistics. The general philosophy behind is to calculate the weighted combination of the contributions of intermediate features to the final statistics ($cont_{y_j,T^2}$ and $cont_{y_j,Q}$), in which the weighting coefficients are the contribution of original variable to the intermediate features ($cont_{x_i,y_j}$).

In general, the individual contributions of single variables are displayed in Eqn 13, where $\{g_{(j)}; P_{(j)}\}$ and $\{f_{1(k)}; J_{(k)}\}$ are subspace model structures and parameters with respect to y_j and z_k, respectively.

$$
\begin{aligned}
cont_{x_i,y_j} &= g_{(j)}(x_i; P_{(j)}) \\
cont_{y_j,z_k} &= f_{1(k)}(y_j; J_{(k)}) \\
cont_{x_i,z_k} &= \sum_{j=1}^{r} f_{1(k)}(y_j; J_{(k)}) g_{(j)}(x_i; P_{(j)})
\end{aligned} \tag{13}
$$

Determined by the specific feature extraction methods adopted in two layers, the model structures g, f_1 and f_2 are related to the complexity of contribution plots calculation. Both layers are linear in PCA-enhanced CVA; hence a close form solution to the complete contribution plots of original variables X to monitoring statistics can be derived.

B. PCA-enhanced CVA-based Contribution Plots Calculation

Following the general formulation proposed previously, the contribution plots of this PCA-enhanced CVA algorithm can be derived in 3 steps.

1) Contribution of the original process variable vector $\boldsymbol{x}$ to the principal components $\boldsymbol{y}$: the general process model of PCA is:

$$\boldsymbol{y}^T = \boldsymbol{x}^T P = \sum_{i=1}^{v} x_i \boldsymbol{p}_i \tag{14}$$

where $\boldsymbol{x} = [x_1, x_2, ..., x_v]^T$ and x_i is the i^{th} process variable, whilst $\boldsymbol{p}_i \in \mathbb{R}^r$ is the i^{th} row vector of P. Therefore, for each variable x_i, its contribution to the entire principal components $\boldsymbol{y}$ is calculated by Eqn 15:

$$cont_{x_i,\boldsymbol{y}} = x_i \boldsymbol{p}_i \tag{15}$$

2) Contribution of principal components $\boldsymbol{y}$ to testing statistics: the contribution of j^{th} principal component y_j to the T^2 and Q statistics at time t defined by Eqn 16 resembles the CVA-based contribution plots in [3].

$$cont_{y_j,T^2}(t) = \sum_{l=1}^{p} \left| \boldsymbol{z}^T(t) J_{j_l} y_j(t-l) \right| \tag{16}$$

$$cont_{y_j,Q}(t) = \sum_{l=1}^{p} \left| \boldsymbol{e}^T(t) L_{j_l} y_j(t-l) \right| \tag{17}$$

where $\boldsymbol{z}(t)$ and $\boldsymbol{e}(t)$ are canonical variates and residuals obtained by CVA at time t.

3) Contribution of original variables to testing statistics: noticing that based on Eqn 15, the contribution of x_i to y_j at time $t-l$ is $cont_{x_i,y_j}(t-l) = x_i(t-l)p_{i,j}$, the

contribution of variable x_i to T^2 statistics at time t can be obtained by Eqn 18:

$$
\begin{aligned}
cont_{x_i,T^2}(t) &= \sum_{j=1}^{r}\sum_{l=1}^{p}\left|\boldsymbol{z}^T(t)J_{j_l}cont_{x_i,y_j}(t-l)\right| \\
&= \left|\boldsymbol{z}^T(t)\right|\left|J\right|\left|C_{i,p}(t)\right|
\end{aligned} \tag{18}
$$

$$
\begin{aligned}
cont_{x_i,Q}(t) &= \sum_{j=1}^{r}\sum_{l=1}^{p}\left|\boldsymbol{e}^T(t)L_{j_l}cont_{x_i,y_j}(t-l)\right| \\
&= \left|\boldsymbol{e}^T(t)\right|\left|L\right|\left|C_{i,p}(t)\right|
\end{aligned} \tag{19}
$$

where the $C_{i,p}(t) \in \mathbb{R}^{pf}$ is the t^{th} column vector of past Hankel matrix $C_{i,p}$ of $x_i\boldsymbol{p}_i$ constructed the same way as Y_p.

Similarly to other data-driven fault detection and diagnosis techniques, the online monitoring procedure of PCA-enhanced CVA is follows:

1) Retrieve the new sample vector $\boldsymbol{x}^*$ and calculate its principal vector $\boldsymbol{y}^*$;
2) construct a past data vector $\boldsymbol{y}_p^*$ with length p for $\boldsymbol{y}^*$;
3) calculate the canonical variate vector $\boldsymbol{z}^*$ and $\boldsymbol{e}^*$;
4) calculate the testing statistics T^{2*} and Q^* and compare with upper control limits;
5) if T^{2*} and/or Q exceed the limits, construct a past vector $\boldsymbol{x}_p^*$ and obtain the contribution plots of all x_i^* so as to identify the influential variables.

IV. Case Study

In this section, the benchmark data set collected by experiments on a multiphase flow facility with two types of faults is used to validate the proposed fault detection and diagnosis algorithm.

A. Process Description

The multiphase flow facility in the Process System Engineering lab of Cranfield University is a unique industrial scale rig for researches and experiments on measuring, monitoring and control of multiphase flows. Water, oil and air are supplied from individual pipelines; by converging and intersection of pipelines, 3-phase flows are mixed, making a multiphase flow with liquid and gas. The multiphase flow is transported, measured, separated and recycled successively afterwards. Being fully automated, this facility can operate in multiple normal operating conditions as well as simulate various faulty scenarios with manually seeded faults. It is also well equipped with measurement instrumentations which contain both regular process variables such as pressure and temperature, and mechanical condition variable such as pump current. All measurement data are collected in real-time and recorded by DeltaV system for further analysis. A more detailed description of this benchmark case study and previous work on statistical monitoring of it can be found in [8] and [9].

The schematic with the layout of measurement instrumentations of this facility is shown in Figure 2. A total of 23 process variables are measured and recorded in the benchmark data set and variable descriptions are provided by Table IV-A.

To validate its capability in for fault detection and identification of influenced variables, aforementioned PCA-enhanced CVA monitoring algorithm is applied to the data sets collected in presence of two types of faults in this facility and compared with CVA, Dynamic PCA (DPCA) and Dynamic PLS (DPLSs). All the contribution plots presented here are accumulated results during the faulty period detected. The monitoring performance metrics are the detection rate $\frac{N(\text{detected sample})}{N(\text{faulty sample})}$ and false alarm rate $\frac{N(\text{false alarm})}{N(\text{normal sample})}$.

B. Fault Detection and Diagnosis Results

1) Fault 1: top separator input blockage: In practice, pipeline blockage is commonly an incipient fault and accumulates over time. To mimic it, the control valve on the input pipeline to top riser (VC404) is turned off gradually and the measurements recorded in this procedure constitute the faulty data set. The fault detection results presented in Table IV-B1 indicate that this fault can be easily detected by different dynamic monitoring methods.

Figure 3 and 4 compare the contribution plots of all variables obtained by PCA-enhanced CVA and ordinary CVA. Noticing that the valve opening of VC404 is not involved as process variable, the pressure drop over this valve will be the proper indicator of pipeline blockage. PCA-enhanced CVA successfully identifies the differential pressure over VC404 (variable 7) as the most influential variable while ordinary CVA is distracted by the riser top pressure (variable 3) and claims that it is also relevant. Therefore, one can conclude that for Fault 1, the PCA-enhanced CVA based contribution plots will provide a more confined location of fault and improve the insight produced for diagnosis and maintenance comparing to the ordinary CVA even without any prior process knowledge. It is also promising for locating other single variable fault such as sensor failure.

2) Fault 2: slugging condition: Slugging is an undesirable phenomenon existing in multiphase flow transportation which is commonly a result of insufficient supply of water and air flows, which will reduce production efficiency and impair the equipment if severe [10]. From process data perspective, the slugging fault will cause large fluctuations in process variables, such as pressure, flow rate and density, at the riser top. Such complexity aggravates the difficulty of detection and identification of slugging fault. In this study, the slugging condition is simulated by reducing input water and air flow rates to obtain the faulty data set. As shown in Table IV-B2, PCA-enhanced CVA algorithm improves the detection of slugging situation with a reasonable level of false alarm rate, which is mainly due to the transient periods before and after slugging happens.

The fault diagnosis results of regular and PCA-enhanced CVA under slugging condition are shown in Figure 5 and 6. Referring to Table IV-A, both the T^2 and Q contribution plots obtained by PCA-enhanced CVA algorithm suggest that

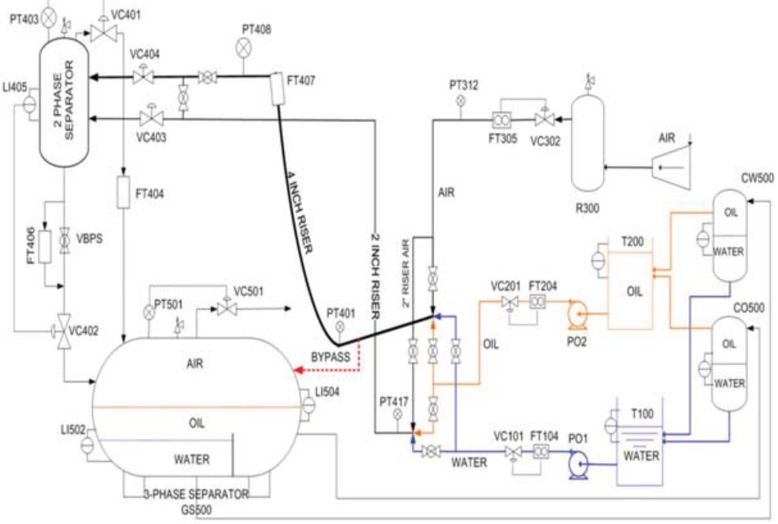


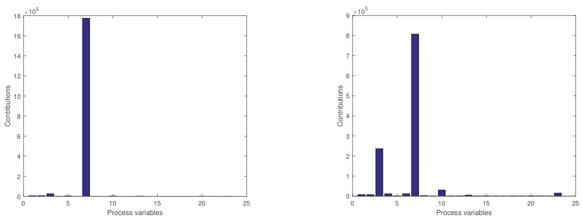
Fig. 2. Schematic of the multiphase flow facility

TABLE I. MEASURED VARIABLES IN MULTIPHASE FLOW FACILITY

No.	Description	Location	No.	Description	Location
1	Air delivery pressure	PT312	13	Top riser density	FT407
2	Riser bottom pressure	PT401	14	Top separator output density	FT406
3	Riser top pressure	PT408	15	Input water density	FT104
4	Top separator pressure	PT403	16	Top riser temperature	FT407
5	3 phase separator pressure	PT501	17	Top separator output temperature	FT406
6	Differential pressure (PT401-PT408)	PT408	18	Input water temperature	FT104
7	Differential pressure over VC404	PT403	19	3 phase separator gas-liquid level	LI504
8	Input air flow rate	FT305	20	Valve position of VC501	VC501
9	Input water flow rate	FT104	21	Valve position of VC302	VC302
10	Top riser flow rate	PT403	22	Valve position of VC101	VC101
11	Top separator level	LI405	23	Water pump current	PO1
12	Top separator output flow rate	FT406			

TABLE II. FAULT DETECTION PERFORMANCES OF FAULT 2

	PCA-CVA	CVA	DPCA	DPLS
Detection rate	98.86	98.69	97.81	96.72
False alarm rate	6.99	2.13	10.83	1.81

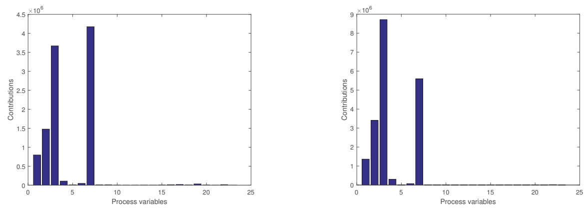


(a) Contributions in T^2 (b) Contributions in Q

Fig. 3. Contribution plots obtained by PCA-enhanced CVA for Fault 1

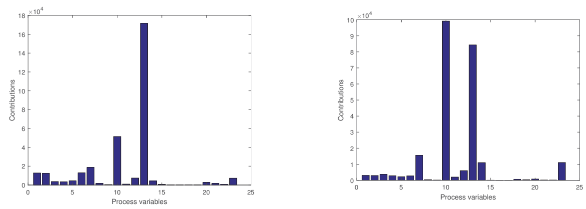
TABLE III. FAULT DETECTION PERFORMANCES OF FAULT 2

	PCA-CVA	CVA	DPCA	DPLS
Detection rate	85.93	43.17	66.49	52.77
False alarm rate	19.39	0.48	2.68	3.10



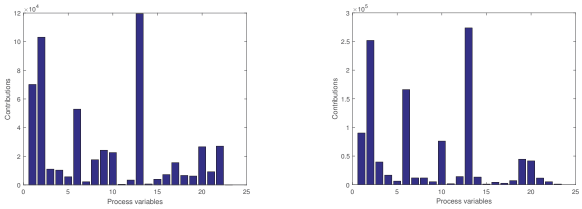
(a) Contributions in T^2 (b) Contributions in Q

Fig. 4. Contribution plots obtained by ordinary CVA for Fault 1



(a) Contributions in T^2 (b) Contributions in Q

Fig. 5. Contribution plots obtained by PCA-enhanced CVA for Fault 2



(a) Contributions in T^2 (b) Contributions in Q

Fig. 6. Contribution plots obtained by ordinary CVA for Fault 2

the most influential variables are flow rate (variable 10) and density (variable 13) measured at the riser top by FT407; while CVA determines the air delivery pressure (variable 1) and bottom riser pressure (variable 2) also have contributed to the T^2 statistics. Since the major influence of slugging is on the riser top flow, it should be located at the riser top and variables in the vicinity are anticipated as influential variables; hence the PCA-enhanced CVA locates the fault more accurately. Furthermore, riser top flow rate and density display a similar trend with respect to the monitoring statistics according to Figure 7.

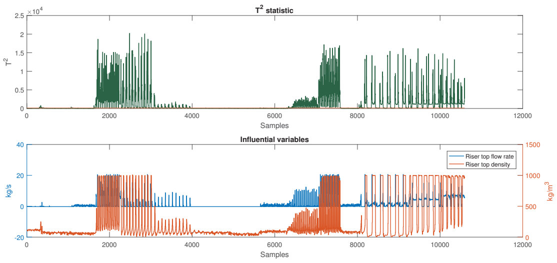


Fig. 7. Trend comparison of testing statistics and influential variables

C. Discussions

The case study of two faults demonstrates that PCA-enhanced CVA algorithm maintains a satisfactory detection performance of simple incipient fault and improves the performance of the complex one. As for fault identification, the contribution plots based on this method also can refine the contributions of different variables, restrict the scope of influential variables, and facilitate the successive maintenance operations accordingly. On the other hand, there still exists potential of improvement for detection and diagnosis of the slugging fault.

V. Conclusion and Future Work

The findings in this work are concluded as follows: 1) extended the contribution plots for fault diagnosis to 2-stage feature extraction based monitoring techniques and derived the corresponding solution for the PCA-enhanced CVA method; 2) validated this monitoring method using a benchmark data set obtained from an industrial scale multiphase flow facility in presence of different types of faults and the superiority of proposed algorithm over other methods in both fault detection and diagnosis has been testified.

For monitoring of processes and faults with extra complexity, advanced fault detection methods are emerging while the gap of contribution plots based fault diagnosis still exists in this 2-stage contribution propagation framework due to the mathematical complexity of kernel transformation. Therefore, it is worthwhile considering the propagation of contribution plots in "kernelized" and other advanced feature extraction methods so as to provide a general solution to the contribution plots of variables for different monitoring techniques in the future.

Acknowledgment

The authors would like to acknowledge the funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675215 - PRONTO - H2020-MSCA-ITN-2015.

References

[1] C. Ruiz-Cárcel, V. Jaramillo, D. Mba, J. Ottewill, and Y. Cao, "Combination of process and vibration data for improved condition monitoring of industrial systems working under variable operating conditions," *Mechanical Systems and Signal Processing*, vol. 66, pp. 699–714, 2016.

[2] R. T. Samuel and Y. Cao, "Nonlinear process fault detection and identification using kernel pca and kernel density estimation," *Systems Science & Control Engineering*, vol. 4, no. 1, pp. 165–174, 2016.

[3] B. Jiang, D. Huang, X. Zhu, F. Yang, and R. D. Braatz, "Canonical variate analysis-based contributions for fault identification," *Journal of Process Control*, vol. 26, pp. 17–25, 2015.

[4] R. T. Samuel and Y. Cao, "Dynamic latent variable modelling and fault detection of tennessee eastman challenge process," in *Industrial Technology (ICIT), 2016 IEEE International Conference on*. IEEE, 2016, pp. 842–847.

[5] P. Miller, R. E. Swanson, and C. E. Heckler, "Contribution plots: A missing link in multivariate quality control," *Applied mathematics and computer science*, vol. 8, no. 4, pp. 775–792, 1998.

[6] S. Joe Qin, "Statistical process monitoring: basics and beyond," *Journal of Chemometrics*, vol. 17, no. 8-9, pp. 480–502, 2003.

[7] P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimations," *Computer Aided Chemical Engineering*, vol. 27, pp. 1557–1562, 2009.

[8] C. Ruiz-Cárcel, Y. Cao, D. Mba, L. Lao, and R. Samuel, "Statistical process monitoring of a multiphase flow facility," *Control Engineering Practice*, vol. 42, pp. 74–88, 2015.

[9] Y. Cao, "A benchmark case for statistical process monitoring - Cranfield multiphase flow facility," MATLAB File Exchage, 2015. [Online]. Available: http://uk.mathworks.com/matlabcentral/ fileexchange/50938-a-benchmark-case-for-statistical-process-monitoring-cranfield-multiphase-flow-facility

[10] A. I. Ogazi, Y. Cao, H. C. Yeung, L. Lao *et al.*, "Slug control with large valve openings to maximize oil production," *SPE Journal*, vol. 15, no. 03, pp. 812–821, 2010.