Improved Kernel Canonical Variate Analysis for Process Monitoring

Raphael T. Samuel
Oil and Gas Engineering Centre,
School of Energy, Environment and Agrifood (SEEA),
Cranfield University,
Cranfield, Bedford, MK43 0AL, UK.
Email: r.t.samuel@cranfield.ac.uk

Yi Cao
Oil and Gas Engineering Centre,
School of Energy, Environment and Agrifood (SEEA),
Cranfield University,
Cranfield, Bedford MK43 0AL, UK.
Email: y.cao@cranfield.ac.uk

***Abstract*—This paper proposes a kernel canonical variate analysis (KCVA) approach for process fault detection. The technique employs the kernel principle to map the original process observations to a high dimensional feature space on which canonical variate analysis is performed. The aim is to obtain an effective monitoring technique that accounts for non-linearity and process dynamics simultaneously. The kernel principle accounts for non-linearity while the CVA accounts for serial correlations widely encountered in dynamic processes. The kernel CVA algorithm proposed in this work is based on QR decomposition in order to avoid singularity problems associated with kernel matrices which require a regularisation step. The technique is evaluated using the Tennessee Eastman Challenge process. Tests show the effectiveness of the proposed kernel CVA approach.**

I. Introduction

Using multivariate statistical techniques to monitor chemical plants have gained much research interest in the last few decades. These techniques depend mainly on process history data and are therefore relatively easier to employ in large scale processes (i.e. processes involving several dozens or higher number of measured variables) compared to the classical approaches based on rigorous process models derived from first principles. Process plants generate large amounts of data from measuring several variables during normal operations. This is possible due to advancement in instrumentation and automation technology. The data acquired are easily stored and/or explored to extract useful information about the process. Improvement in data analysis applications and increase in computer power have also contributed immensely in providing the stage for data-driven techniques to thrive. Two early examples of multivariate statistical methods which are very widely used are Principal Component Analysis (PCA) [1] and Canonical Correlation Analysis (CCA) [2]. Both PCA and CCA are eigenvalue problems, however, PCA is used in cases involving a single collection of variables while CCA is used when considering two sets of variables.

Canonical correlation analysis attempts to find the existing relations between two multivariate data sets. This is achieved by obtaining linear combinations of each of the original sets of variables and determining the pairwise correlations of the linear combinations of the two sets of variables. The linear combinations are called canonical variates while the pairwise correlations are known as canonical correlations. The strength of the association between the two sets of variables is measured by the canonical correlations. If correlation is considered to be the main determinant of information in the original two blocks of variables, then CCA can be used to obtain a reduced dimensional set of variables from the original data sets by discarding the canonical variate pairs with very low correlations. However, both PCA and CCA are static and linear techniques. They are therefore deficient in capturing relations in data rich in dynamic and non-linear characteristics.

Many complex chemical industry processes exhibit both non-linear and dynamic behaviour. Therefore, to effectively monitor such processes, the techniques employed are expected to capture these characteristics. Otherwise abnormal process conditions may be detected long after they have occurred or may not be detected at all. Both of these situations can compromise process safety, operational efficiency and consistent product quality, which are extremely important in chemical process industries. Ineffective monitoring can also lead to less than optimal maintenance practices leading to frequent equipment break down, longer downtimes and higher operational cost.

Several methods have been proposed to address either non-linearity or process dynamics separately but not many studies have addressed tackling both properties simultaneously. One of the few studies reported in the literature which address both of these characteristics was conducted by Choi and Lee [3]. They proposed the dynamic kernel PCA approach and tested it on a simulated non-linear process as well as a wastewater treatment process. This approach employed a kernel function to capture the non-linear relations and a time lag-data extension of the original observations to describe the dynamics of the process. They reported that the method provided better monitoring evidenced in lower missing alarms and smaller detection times compared to PCA and KPCA. The kernel methods are the preferred techniques for capturing non-linear relations compared to methods based on neural network and principal curves because they do not require solving a non-linear optimization problem. However, in the DKPCA technique proposed by Choi and Lee, the kernel approach was combined with an extension of the PCA (dynamic PCA)

which has limitation in capturing process dynamics [4]. This is likely to limit the performance of this technique especially in faults that are not easily detectable. In other words, although the approach has a good technique for describing non-linear relations, the method employed for accounting for process dynamics, being an extension of a linear algorithm (i.e. PCA), leaves room for improvement.

Canonical variate analysis (CVA) is a state-space based technique which is widely reported as an appropriate methodology for monitoring dynamic processes [5]–[8]. Like the CCA, the CVA finds relations between two sets of variables but the two sets of variables are obtained from expanding an observation at a given time instant to p past and f future measurements, in order to account for serial correlations.

To improve the monitoring of non-linear dynamic processes, Odiowei and Cao [4] proposed the CVA with KDE technique. In their work, the CVA was associated with kernel density-based upper control limits derived from the estimated probability density functions of the monitoring indices instead of determining the control limits based on the Gaussian assumption. Nevertheless, this approach does not directly address non-linear problems in a dynamic process. Considering the successful application of kernel methods in several application domains and the CVA technique in describing non-linear and dynamic behaviour respectively, a combination of these two approaches should make an appropriate scheme for describing non-linear and dynamic relations simultaneously in data-driven process monitoring. However, not much is reported on kernel CVA in the literature even though many studies involving kernel extensions to the CCA exist [9]–[11]. A management system based on kernel CVA to monitor and diagnose smart homes is reported by Giantomassi and others in [12] but application of this technique in the chemical process industry is not well investigated. Also, the work mentioned above does not provide a comparison of the technique with other approaches. This makes it difficult to assess how its performance compares with other known techniques. Furthermore, since the past and future kernel matrices generated are singular, regularisation of these matrices is needed to perform the matrix inversion step required to implement their CVA algorithm.

The objective of this paper is to implement KCVA using QR decomposition to preclude the need for regularising the kernel matrices generated and to investigate the performance of the proposed approach in monitoring a chemical process. The paper also provides a comparison of the effectiveness of the technique with the kernel CVA approach based on the regularisation of singular kernel matrices. Assessment of both techniques was done by applying them to simulation data obtained from the Tennessee Eastman benchmark process.

The rest of the paper is organised as follows: Section II summarises the KCVA procedure adopted. Section III shows how to compute the upper control limits of the monitoring statistics using the KDE method. The proposed kernel CVA based process monitoring procedure is presented in Section IV. Section V describes the application to the Tennessee Eastman process while conclusions reached are presented in Section VI.

II. Kernel Canonical Variate Analysis

The idea of kernel CVA is to extract state variables that also capture non-linear characteristics in the observed data using non-linear kernel transformation and CVA. A brief description of the kernel CVA technique adopted is given in this section. Detailed discussion including mathematical procedure on non-linear mapping based on a kernel function and the CVA algorithm can be found in [4], [13]–[15].

To account for time correlations, each observation vector $\mathbf{x}$ is expanded at a given time point t to obtain information from the past (p) and future (f) measurements each containing d variables using (1):

$$\mathbf{x}_{(p,t)} = \begin{bmatrix} \mathbf{x}_{(t-1)} \\ \mathbf{x}_{(t-2)} \\ \vdots \\ \mathbf{x}_{(t-p)} \end{bmatrix} \in \Re^{dp} \quad \text{and} \quad \mathbf{x}_{(f,t)} = \begin{bmatrix} \mathbf{x}_{(t)} \\ \mathbf{x}_{(t+1)} \\ \vdots \\ \mathbf{x}_{(t+f-1)} \end{bmatrix} \in \Re^{df} \tag{1}$$

The various components are mean-centred as follows:

$$\hat{\mathbf{x}}_{(p,t)} = \mathbf{x}_{(p,t)} - \bar{\mathbf{x}}_{(p,t)} \quad \text{and} \quad \hat{\mathbf{x}}_{(f,t)} = \mathbf{x}_{(f,t)} - \bar{\mathbf{x}}_{(f,t)} \tag{2}$$

where $\bar{\mathbf{x}}_{(p,t)}$ and $\bar{\mathbf{x}}_{(f,t)}$ are the sample means of $\mathbf{x}_{(p,t)}$ and $\mathbf{x}_{(f,t)}$ respectively. The past and future vectors are then arranged together in columns to obtain the corresponding past and and future matrices, $\mathbf{X}_p$ and $\mathbf{X}_f$ respectively.

$$\mathbf{X}_p = \left[\hat{\mathbf{x}}_{(p,p+1)},\, \hat{\mathbf{x}}_{(p,p+2)}, \ldots \hat{\mathbf{x}}_{(p,p+M)}\right] \in \Re^{dp \times M} \tag{3}$$

$$\mathbf{X}_f = \left[\hat{\mathbf{x}}_{(f,p+1)},\, \hat{\mathbf{x}}_{(f,p+2)}, \ldots \hat{\mathbf{x}}_{(f,p+M)}\right] \in \Re^{df \times M} \tag{4}$$

where the columns of the truncated Hankel matrices for N observations is $M = N - f - p + 1$.

To apply the kernel principle, non-linear mappings, Φ_1 and Φ_2 are used to map $\Re^{dp}$ and $\Re^{df}$ into a high dimensional feature space, $\Phi_1 : \Re^{dp} \longrightarrow F$ and $\Phi_2 : \Re^{df} \longrightarrow F$ respectively. Kernel matrices ($\mathbf{K}_p$ and $\mathbf{K}_f$) are obtained using the kernel trick, ([13], [14]):

$$\mathbf{K}_p = \langle \Phi_1(\mathbf{X}_p), \Phi_1(\mathbf{X}_p) \rangle, \tag{5}$$

$$\mathbf{K}_f = \langle \Phi_2(\mathbf{X}_f), \Phi_2(\mathbf{X}_f) \rangle \tag{6}$$

where the elements of these kernel matrices are defined as

$$(\mathbf{K}_p)_{ji} = \left\langle \Phi_1\left(\hat{\mathbf{x}}_{(p,p+j)}\right), \Phi_1\left(\hat{\mathbf{x}}_{(p,p+i)}\right) \right\rangle$$

$$(\mathbf{K}_f)_{ji} = \left\langle \Phi_2\left(\hat{\mathbf{x}}_{(f,p+j)}\right), \Phi_1\left(\hat{\mathbf{x}}_{(f,p+i)}\right) \right\rangle$$

for all $j, i = 1 \ldots, M$. These kernel matrices are mean-centred as follows:

$$\mathbf{K}_{cp} = \mathbf{K}_p - \mathbf{B}\mathbf{K}_p - \mathbf{K}_p\mathbf{B} + \mathbf{B}\mathbf{K}_p\mathbf{B} \tag{7}$$

$$\mathbf{K}_{cf} = \mathbf{K}_f - \mathbf{B}\mathbf{K}_f - \mathbf{K}_f\mathbf{B} + \mathbf{B}\mathbf{K}_f\mathbf{B} \tag{8}$$

where $\mathbf{K}_{cp}$ and $\mathbf{K}_{cf}$ are the past and future mean-centred kernel matrices, $\mathbf{B}$ is an $M \times M$ matrix in which each element is equal to $\frac{1}{M}$.

Kernel CVA seeks to find weights which make the linear combinations of $\mathbf{K}_{cp}$ and $\mathbf{K}_{cf}$ have maximal correlations.

Since kernel matrices are ill-conditioned and suffer computational instabilities, a regularisation step is normally needed so that the matrix inversion required by the CVA algorithm can be carried out. However, such a regularisation step reduces the accuracy of the model which makes the monitoring performance poor. In this paper, the mean-centred past and future kernel matrices were factorised using QR decomposition as follows:

$$\mathbf{K}_{cp} = \mathbf{Q}_p \mathbf{R}_p \quad \text{and} \quad \mathbf{K}_{cf} = \mathbf{Q}_f \mathbf{R}_f, \tag{9}$$

where $\mathbf{Q}_p$ and $\mathbf{Q}_f$ are orthogonal matrices and $\mathbf{R}_p$ and $\mathbf{R}_f$ are upper triangular matrices. Though $\mathbf{K}_{cp}$ and $\mathbf{K}_{cf}$ are not full rank, they were managed by using the MATLAB backlash operator which makes them equivalent to pseudo-inverses. The product of the orthogonal matrix pair was computed and canonical variates were obtained by performing singular value decomposition (SVD):

$$\mathbf{W} = \mathbf{Q}_f^T \mathbf{Q}_p = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{10}$$

where T denotes transpose, $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices, while $\mathbf{S}$ is a diagonal matrix whose entries on the main diagonal (singular values) show the degree of correlation between pairs of $\mathbf{U}$ and $\mathbf{V}$. This procedure precludes the computational problems associated with obtaining the scaled Hankel matrix for performing SVD when covariance and cross-covariance matrices are used in a KCVA-based methodology as proposed in [12] and is a major strength of the proposed approach.

The normalised left and right singular vectors, $(\mathbf{U}^*)$ and $(\mathbf{V}^*)$ respectively are obtained using the following:

$$\mathbf{U}^* = \mathbf{R}_f^+ \mathbf{U} \left(M-1\right)^{\frac{1}{2}} \quad \text{and} \quad \mathbf{V}^* = \mathbf{R}_p^+ \mathbf{V} \left(M-1\right)^{\frac{1}{2}}, \tag{11}$$

where the superscript in $\mathbf{R}_f^+$ and $\mathbf{R}_p^+$ represent pseudo-inverse. Sorting the normalised singular values and the columns of the singular vectors associated with them in descending order makes $\mathbf{V}_n^*$ (i.e. the first n columns of $\mathbf{V}^*$), the most dominant pairwise correlations with those of $\mathbf{U}^*$. Thus, the transformation matrices for determining the n-dimensional state variables and residuals are obtained as:

$$\mathbf{J} = \mathbf{V}_n^* \in \Re^{M \times n} \quad \text{and} \quad \mathbf{L} = \left(I - \mathbf{J}\mathbf{J}^T\right) \in \Re^{M \times M} \tag{12}$$

The state space $\mathbf{Z}$ and residual space $\mathbf{E}$ are computed using (13):

$$\mathbf{Z} = \mathbf{J} \cdot \mathbf{K}_{cp} \in \Re^{n \times M} \quad \text{and} \quad \mathbf{E} = \mathbf{L} \cdot \mathbf{K}_{cp} \in \Re^{M \times M} \tag{13}$$

Hotellings T^2 and the Q statistic or squared prediction error (SPE) are also used in kernel CVA as the monitoring statistics. The Hotellings T^2 monitors the changes in the state space while the Q statistic monitors the changes in the residual space. They are determined using (14)

$$T_k^2 = \sum_{i=1}^{n} z_{i,k}^2 \quad \text{and} \quad Q_k = \sum_{i=1}^{M} e_{i,k}^2, \tag{14}$$

where n is the number of states retained, $z_{i,k}$ and $e_{i,k}$ are $(i,k)^{\text{th}}$ the entries of $\mathbf{Z}$ and $\mathbf{E}$ matrices respectively.

III. Computation of Upper Control Limits

To correct the Gaussian assumption, the kernel density estimation (KDE) technique was used to estimate the probability density functions (PDF) of the monitoring indices. Upper control limits were computed from the estimated PDFs instead of using parametrically obtained control limits.

Given the probability density function g(x), the probability of x to be less than c at a specified confidence level α is given by (15):

$$P(x < c) = \int_{-\infty}^{c} g(x)\,\mathrm{dx} = \alpha \tag{15}$$

The control limits of the monitoring statistics (T^2 and Q) were determined using (16).

$$\int_{-\infty}^{T_\alpha^2} p(T^2)\,\mathrm{d}T^2 = \alpha \quad \text{and} \quad \int_{-\infty}^{Q_\alpha} p(Q)\,\mathrm{d}Q = \alpha \tag{16}$$

A more comprehensive account of the KDE technique including the importance of selecting the bandwidth H and methods of obtaining an optimum value can be found in [16] and [17].

IV. Fault Detection Procedure for Kernel Canonical Variate Analysis

Similar to other multivariate statistical process monitoring methodologies, the fault detection strategy of kernel CVA involves two phases: off-line training and on-line monitoring or testing. The off-line training phase involves development of the process model, calculation of the monitoring indices and their upper control limits using the normal operation data. Conversely, on-line monitoring involves computing the monitoring indices using faulty or test data and comparing their values with the control limits obtained in the off-line training phase to determine the status of the process. The steps involved in the proposed kernel CVA technique for the training and monitoring phases are outlined below:

A. Off-line Training

1) Obtain observation vector.
2) Expand observation vector at each time point t to obtain information from the past (p) and future (f) measurements using (1).
3) Form kernel matrices of the past and future measurements.
4) Mean-centre the past and future kernel matrices. Factorise the mean-centred past and future kernel matrices using QR decomposition to obtain pairs of upper triangular and orthogonal matrices.
5) Compute the product of the orthogonal matrix pair from step 4 and perform singular value decomposition. Normalise the canonical coefficients.
6) Determine states and residuals.
7) Compute monitoring indices T^2 and Q at each time point as the sum of the squared state variables and residuals respectively.

B. On-Line Monitoring

1) Acquire test data and define past and future matrices and arrange data similar to training data.
2) Form kernel matrices of the past and future measurements using the same function and parameters used in the training stage and mean-centre.
3) Calculate states and residuals of test data.
4) Compute T^2 and Q of test data.
5) Monitor process by comparing value of T^2 and Q against their control limits. A fault is detected if both monitoring indices exceed their control limits.

V. Application Study

In this section kernel CVA via QR decomposition and kernel CVA with regularisation were applied to the Tennessee Eastman (TE) process for monitoring performance evaluation. Faults 3, 9, and 15 of the TE process were considered for the application study. Faults 3 and 9 are step change and random variation in D feed temperature respectively, while Fault 15 is condenser cooling water sticking. These faults are usually more difficult to detect in the TE process.

A. Tennessee Eastman Process

The TE process is a simulation of a real industrial plant manifesting both non-linear and dynamic properties. It is widely used as a benchmark process for evaluating process monitoring and control approaches [18]. It is made up of five major units: separator, compressor, reactor, stripper and condenser, and eight components coded A to H. A schematic diagramme of the processes is presented in Figure 1.

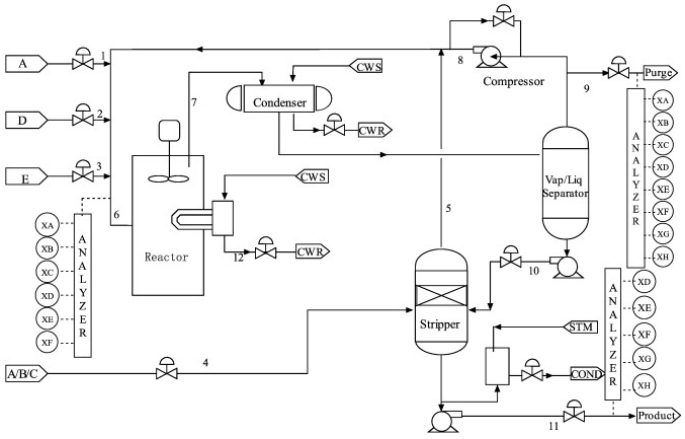


Fig. 1: Schematic diagramme of TE process

The TE process consists of 960 samples and 52 variables which include 22 continuous variables, 19 composition measurements sampled by 3 composition analysers and 12 manipulated variables. All the 22 continuous variables and 11 manipulated variables were used in this study. Variables 1 to 22 in Table I are measurement variables while 23 to 33 are manipulated variables. The agitation speed of the reactor's stirrer (the $12th$ manipulated variable) was not included because it is constant. Details of the 33 variables are shown in Table I.

TABLE I: Monitoring variables for the TE process

No.	Description	No.	Description
1	A feed (stream 1)	18	Stripper temperature
2	D feed (stream 2)	19	Stripper stream flow
3	E feed (stream 3)	20	Compressor work
4	Total feed (stream 4)	21	Reactor cooling water outlet temp
5	Recycle flow (stream 8)	22	Separator cooling water outlet temp
6	Reactor feed rate (stream 6)	23	D feed flow (stream 2)
7	Reactor pressure	24	E feed flow (stream 3)
8	Reactor level	25	A feed flow (stream 1)
9	Reactor temperature	26	Total feed flow (stream 4)
10	Purge rate	27	Compressor recycle valve
11	Separator temperature	28	Purge valve (stream 9)
12	Separator level	29	Separator pot liquid flow (stream 10)
13	Separator pressure	30	Stripper liquid product flow
14	Separator under flow	31	D Stripper stream valve
15	Stripper level	32	Reactor cooling water flow
16	Stripper pressure	33	Condenser cooling water flow
17	Stripper under flow		

B. Parameters Selection

Some key parameters were determined to optimise the approach. These include the kernel used and its bandwidth, the length of lag used, and the number of states retained. The radial basis kernel which is a common choice in previous studies [14], [19] was used in this paper. The value of the kernel parameter c can be determined using cross validation or by using the relation $c = Wn\sigma^2$, where W is a constant which is dependent on the data being used, n and σ^2 are the dimension and variance of the input space respectively [14], [20]. The latter method was adopted in this study with a value of c = 20.

The length of lag represents the number of past or future observations that correlate significantly with an observation at a particular time point. For the CVA algorithm, the optimal number of lags can be obtained by considering the summed squares of all process measurements [4]. Consequently, a lag of 15 was adopted.

According to Negiz and Cinar [21], the states to retain can be selected based on the number of dominant singular values. Fig. 2 shows the normalised singular values of the training data. It can be observed from Fig. 2 that the singular values decrease very slowly. Therefore, choosing how many states to retain based on dominant singular values will not give a realistic model [4]. Furthermore, the number of states retained does not really matter in this case because the monitoring indices were used jointly for fault detection due to their complementary nature. This means that fault detection is acknowledged if either the T^2 or Q statistic detects a fault since detectable process variation may not always occur in both the model space and the residual space at the same time. Consequently, 16 states were retained to curtail the false alarm rate.

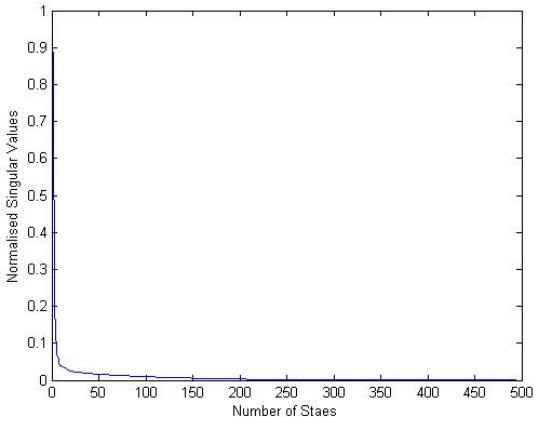


Fig. 2: Normalised singular values of training data

C. Results and Discussion

Monitoring performance was based on fault detection rates (FDR), false alarm rate (FAR), and detection time. FDR is the percentage of faulty observations identified correctly. It was computed as

$$FDR = \frac{\theta_{\text{fc}}}{\theta_{\text{tf}}} \times 100 \tag{17}$$

where θ_{fc} denotes the number of fault samples identified correctly and θ_{tf} is the total number of fault samples. FAR is the percentage of observations identified as abnormal under normal operating mode. It was calculated as,

$$FAR = \frac{\theta_{\text{nf}}}{\theta_{\text{tn}}} \times 100 \tag{18}$$

where θ_{nf} is number of normal observations reported as faulty and θ_{tn} represent the total number of normal observations. Detection delay was computed as the amount of time that passes before a fault is detected after it has occurred.

Table IIa shows the rate of fault detection for Faults 3, 9 and 15 using the QR decomposition approach while Tables IIb, IIc, and IId show results obtained at different values of regularisation for the same faults. The detection rates at a regularisation value of 10^{-2} were the lowest (51.25, 78.13, and 85.38 percent) for Faults 3, 9 and 15 respectively. At a very small regularization value of 10^{-8}, the detection rates improved but the values were still lower than the results obtained via QR decomposition. Also, the detection delay for the QR-based approach for all three faults was 15 seconds while the corresponding rates for the technique based on regularisation were 54/45, 84/63, and 54/45 seconds for Faults 3, 9 and 15 respectively for the worst and best detection time delays. In all faults considered, the QR based detection times were better than the best detection times obtained via the regularisation approach. The regularisation approach also had higher FARs (which makes it relatively poorer) except for the smallest regularisation value.

Fig. 3 shows the monitoring statistics for Fault 15. The poor monitoring performance arising from choosing a poor regularisation value is shown in Fig 3b. It can be seen that the monitoring index (the solid signal) did not fully go above the control limit (dash-dot horizontal line) most of the time which shows that fault detection performance was poor.

TABLE II: Detection performance (a) KCVA with QR decomposition (Faults 3, 9 and 15), (b) KCVA with regularisation (Fault 3), (c) KCVA with regularisation (Fault 9), and (d) KCVA with regularisation (Fault 15)

(a)

	Fault 3	Fault 9	Fault 15
FDR (%)	98.25	97.50	98.25
FAR	0.0382	0.0382	0.0382
Detection delay, s	15	15	15

(b)

Regularisation value	10^{-2}	10^{-5}	10^{-8}
FDR (%)	51.25	98.13	98.13
FAR	0.0458	0.0840	0.0076
Detection delay, s	54	45	45

(c)

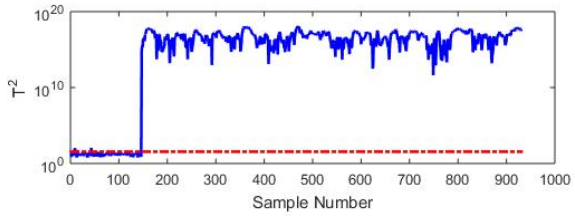
Regularisation value	10^{-2}	10^{-5}	10^{-8}
FDR (%)	78.13	97.38	97.38
FAR	0.0458	0.0840	0
Detection delay, s	84	63	63

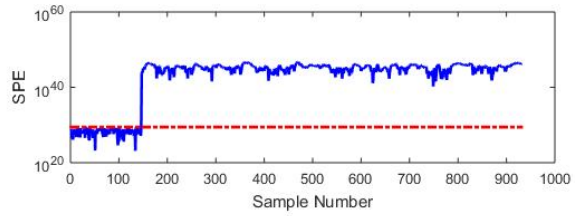
(d)

Regularisation value	10^{-2}	10^{-5}	10^{-8}
FDR (%)	85.38	98.13	98.13
FAR	0.0458	0.0840	0
Detection delay, s	54	45	45

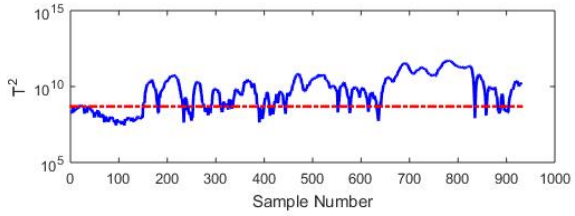
VI. Conclusions

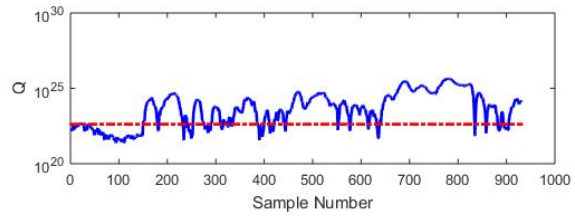
The existing kernel CVA technique is improved in this paper for detecting process faults. Problems of non-linearity and dynamism in processes were accounted for simultaneously by employing the kernel principle followed by the CVA technique. In the proposed method the product matrix obtained from the past and future kernel matrices, on which singular value decomposition was performed at the CVA stage, was obtained via QR decomposition to avoid singularity problems associated with kernel matrices, such that there was no need to carry out regularisation of the generated kernel data. Results obtained from applying this technique to the Tennessee Eastman process were compared with results based on different values of regularisation. The results show that the proposed technique outperformed the KCVA based on regularisation in both monitoring rate and the time taken to detect faults. This supports the effectiveness of the proposed method in enhancing process monitoring performance. Avoiding the need to determine an optimum regularisation value reduces the parameters required for implementing kernel-based CVA by



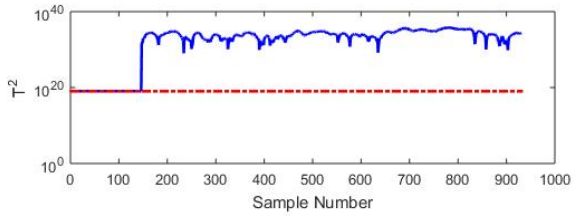


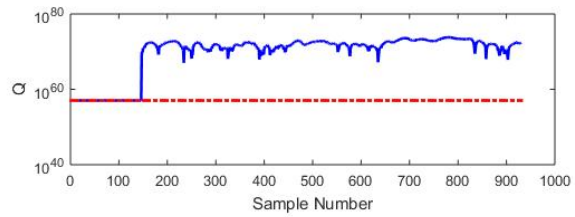
(a)





(b)





(c)

Fig. 3: Monitoring statistics of Fault 15. (a) KCVA with QR, (b) KCVA with regularisation (10^{-2}), (c) KCVA with regularisation (10^{-8})

one. This is desired because a poorly chosen regularisation parameter gives poor monitoring results.

REFERENCES

[1] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, vol. 24, pp. 417–441, 498–520, 1933.

[2] ——, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 321-377, 1936.

[3] S. W. Choi and I.-B. Lee, "Nonlinear dynamic process mornitoring based on dynamic kernel pca," *Chemical Engineering Science*, vol. 59, pp. 5897–5908, 2004.

[4] P.-E. P. Odiowei and Y. Cao, "Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimation," *IEEE Transaction on Industrial Informatics*, vol. 6, no. 1, pp. 36–44, 2010a.

[5] E. L. Russell, L. H. Chiang, and R. D. Braatz, "Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 51, pp. 81–93, 2000.

[6] B. C. Juricek, D. E. Seborg, and W. E. Larimore, "Fault detection using canonical variate analysis," *Industrial & Engineering Chemistry Research*, vol. 43, pp. 458–474, 2004.

[7] B. Jiang, X. Zhu, D. Huang, J. A. Paulson, and R. D. Braatz, "A combined canonical variate analysis and fisher discriminant analysis (cva-fda) approach for fault diagnosis," *Computers and Chemical Engineering*, vol. 77, pp. 1–9, 2015.

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

[9] X. Zhu, Z. Huang, H. T. Shen, J. Cheng, and C. Xu, "Dimensionality reduction by mixed kernel canonical correlation analysis," *Pattern Recognition*, vol. 45, no. 8, pp. 3003–3016, 2012.

[10] S.-Y. Huang, M.-H. Lee, and C. K. Hsiao, "Nonlinear measures of association with kernel canonical correlation analysis and applications," *Journal of Statistical Planning and Inference*, vol. 139, pp. 2162–2174, 2009.

[11] S. Tan, F. Wang, Y. Chang, W. Chen, and J. Xu, "Fault detection and diagnosis of nonlinear processes based on kernel ica-kcca," in *Chinese Control and Decision Conference*. Xuzhou: IEEE, May 2010, pp. 3869–3874.

[12] A. Giantomassi, F. Ferracuti, S. Iarlori, S. Longhi, A. Fonti, and G. Comodi, "Kernel canonical variate analysis based management system for monitoring and diagnosing smart homes," in *2014 International Joint Conference on Neural Networks, IJCNN 2014, Beijing, China, July 6-11, 2014*, 2014, pp. 1432–1439. [Online]. Available: http://dx.doi.org/10.1109/IJCNN.2014.6889821

[13] B. Scholkopf, A. Smola, and K. Muller, "Non-linear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299–1399, 1998.

[14] J. M. Lee, C. K. Yoo, S. W. Choi, P. A. Vanrolleghem, and I. B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical and Engineering Science*, vol. 59, pp. 223–234, 2004.

[15] R. T. Samuel and Y. Cao, "Fault detection in a multivariate process based on kernel pca and kernel density estimation," in *International Conference on Automation and Computing*. Cranfield, UK: IEEE, September 2014.

[16] X.-W. Chen, "An improved branch and bound algorithm for feature selection," *Pattern Recognition Letters*, vol. 24, pp. 1925–1933, 2003.

[17] L. Liang, "Multivariate statistical process monitoring using kernel density estimation," *Development in Chemical Engineering and Mineral Processing*, vol. 13, pp. 185–192, 2005.

[18] L. Chiang, E. Russell, and R. Braatz, *Fault Detection and Diagnosis in Industrial Systems*. London: Springer, 2001.

[19] G. Stefatos and A. B. Hamza, "Statistical process control using pca," in *Proceedings of 15th Mediterranean Conference on Control and Automation*. Anthens, Greece: IEEE, July 2007, pp. 1418–1423.

[20] S. Mika, B. Scholkopf, A. Smola, K. R. Muller, M. Scholz, and G. Ratsch, "Kernel pca and de-noising in feature spaces," in *Advances in Neural Information Processing System*, vol. 11, 1999, pp. 536–542.

[21] A. Negiz and A. Cinarl, "Monitoring of multivariable dynamic processes and sensor auditing," *Journal of Process Control*, vol. 8, no. 5-6, pp. 375–380, 1998.