Preprints of the
9th International Symposium on Advanced Control of Chemical Processes
The International Federation of Automatic Control
June 7-10, 2015, Whistler, British Columbia, Canada

Kernel Canonical Variate Analysis for Nonlinear Dynamic Process Monitoring ⋆

Raphael T. Samuel * **Yi Cao** **

* *School of Energy, Environment and Agrifood, Cranfield University, UK (e-mail: r.t.samuel@ cranfield.ac.uk).*
** *(e-mail: y.cao@cranfield.ac.uk)*

Abstract: Effective monitoring of industrial processes provides many benefits. However, for dynamic processes with strong nonlinearity many existing techniques still cannot give satisfactory monitoring performance. This is evidenced by the well known Tennessee Eastman (TE) benchmark process, where some faults, *e.g.* Faults 3 and 9, have not been comfortably detected by almost all data-driven approaches published in the literature. This is because most data driven approaches, such as the principal component analysis (PCA) are linear. In recent years, powerful nonlinear analysis tools using kernel principles have been proposed. However, these tools have not been successfully applied to dynamic systems due to enormous dimensionality and complexity issues. This paper proposes nonlinear dynamic process monitoring based on kernel canonical variate analysis (KCVA). The proposed technique performs the traditional canonical variate analysis with KDE (CVA-KDE) in the kernel space generated from kernel PCA. The kernel PCA accounts for the nonlinearity in the process data while the CVA captures the process dynamics. The approach was tested on the TE benchmark problem for fault detection. The results obtained showed that KCVA detected faults at a higher rate and much earlier than CVA especially in the more difficult faults such as Faults 3 and 9 in the TE process which cause very little variation in the measured variables.

Keywords: Canonical variate analysis, kernel methods, fault detection, multivariate statistics, process monitoring.

1. INTRODUCTION

Statistical monitoring of chemical processes have gained prominence in process control research over the last few decades. Methods such as principal component analysis (PCA), and partial least squares (PLS) have been used extensively to obtain useful information from chemical process history data for detecting and diagnosing process faults (abnormal process deviations), (Chiang et al., 2001). Due to the data-based nature of these methods, they are applied with relative ease to large complex processes, compared to other methods based on systems theory or rigorous process models (Qin, 2003). It is also easy to acquire, store and process large amount of data nowadays due to advances in automation, data mining technologies and higher computer processing power. Examples of large scale industries that have embraced their use include: petrochemical, fertilizer and cement; metal extraction/processing; power systems and power grids; air traffic control and railway regulation, as well as complicated instruments/equipments (Yin et al., 2015). However, traditional statistical process monitoring techniques are based on assumptions of static and linear process behaviour which may not be valid in many practical situations. This makes static and linear approaches inefficient in monitoring many real industrial processes because they provide incomplete representation of such processes. This is critical because the essence of monitoring a process is to timely detect and accurately diagnose abnormal conditions so that remedial actions can be taken while the system is still running. Adopting the linear and static assumptions in complex processes exhibiting non-linear and dynamic characteristics will defeat this objective. This raises the need for developing more effective monitoring approaches for non-linear dynamic processes.

Ku et al. (1995) proposed dynamic PCA by carrying out PCA on lagged variables to account for time correlation. Similarly, a dynamic version of PLS was proposed by Komulainen et al. (2004). Although dynamic PCA and dynamic PLS performed better than PCA and PLS respectively, they have limited capacity in representing system dynamics (Odiowei and Cao, 2010a; Chiang et al., 2001). Canonical varaiate analysis (CVA), a state-space based method, was therefore proposed as a more effective method for monitoring dynamic processes (Russell et al., 2000; Lu and Liu, 2006). However, like PCA and PLS, CVA is a linear technique.

The linear correlations employed in CVA may not be able to adequately describe the association between variables when nonlinear relations exist. This naturally calls for the need for approaches that can explore and exploit nonlinear relationships in addition to accounting for dynamic process behaviour. To address the non-Gaussian distribution issue associated with process nonlinearities, Odiowei and Cao

⋆ The first author's PhD studentship at Cranfield University was sponsored by the Bayelsa State Scholarship Board of Nigeria under the State Government's Oversees Scholarship Scheme.

(2010a) adopted the method of kernel density estimation (KDE) to determine the control limits when applying the CVA for monitoring nonlinear dynamic processes. The results of their work showed that the CVA with KDE approach outperformed other dynamic methods like dynamic PCA and dynamic PLS and significantly improved both fault detection rate and time in a nonlinear dynamic process. Monitoring performance was further improved by applying the independent component analysis (ICA) to the state-space obtained through the CVA. This led to the so called state-space ICA (SSICA) (Odiowei and Cao, 2010b). Nevertheless, both the CVA with KDE and SSICA approaches do not directly address the nonlinear issue associated with dynamic processes. Therefore, fault detection performance based on these techniques is still not satisfactory for some cases.

On the other hand, kernel based approaches such as the kernel PCA (KPCA) and kernel canonical correlation analysis (KCCA) have been successfully applied to nonlinear process monitoring which shows that kernel methods are efficient tools to deal with nonlinearity (Lee et al., 2004; Choi et al., 2005; Tan et al., 2010). Unfortunately, the effort to apply kernel approaches to address nonlinearity associated with dynamic systems is not very successful. Although, both KCCA and CVA algorithms have been well known for many years, directly applying the KCCA to dynamic systems may result in a singular kernel matrix, which requires regularization in order to avoid potential computational instabilities (Huang et al., 2009; Scholkopf and Smola, 2002). Furthermore, such an approach often leads to poor detection performance.

In this work, we propose a novel kernel CVA (KCVA) technique by performing the traditional CVA in the kernel space generated from kernel PCA. Through the compression of KPCA, the singularity issue associated with the KCCA approach is avoided. The new KCVA approach was successfully tested on the Tennessee Eastman (TE) benchmark problem for fault detection. The results obtained showed that KCVA detected faults at a higher rate and much earlier than CVA with KDE especially in the more difficult faults such as Faults 3 and 9 in the TE process which cause very little variation in the measured variables.

The paper is organized as follows: The implementation strategy of the proposed technique is explained in Section 2. Its application to process monitoring is illustrated using the TE process as a case study in Section 3 while conclusions from the study are drawn in Section 4.

2. IMPLEMENTATION STRATEGY OF KCVA

The basic idea of KCVA is to extract state variables that also capture the nonlinearity in the process measurements. Firstly, the observed data are nonlinearly mapped into a high dimensional feature space. This is followed by extracting nonlinear principal components in the feature space using the kernel trick, and the implementation of CVA in the kernel component space. The KPCA and CVA algorithms are sketched in Sections 2.1 and 2.2. More detailed derivations of these approaches can be found in (Scholkopf et al., 1998; Lee et al., 2004; Samuel and Cao, 2014; Odiowei and Cao, 2010a).

2.1 Kernel Principal Component Analysis

Given a data set $\mathbf{x_k} \in \Re^M, k = 1, \ldots, N$, where N is the number of observations, a nonlinear mapping $\Phi : \Re^M \longrightarrow \boldsymbol{H}$, maps $\mathbf{x}$ in the input space to a high dimensional feature space $\boldsymbol{H}$, where the data structure is more likely to be linear (Haykin, 1999). The covariance matrix in the feature space is given by

$$C_H = \frac{1}{N}\sum_{j=1}^{N}\langle \Phi(x_j), \Phi(x_j)\rangle \tag{1}$$

where $\Phi(x_k)$, for $k = 1, \ldots N$ is assumed to have a mean of 0 and a variance of 1. To diagonalize the covariance matrix, we solve the eigenvalue problem in the feature space as

$$\lambda \mathbf{w} = \mathbf{C}_{\mathrm{H}}\mathbf{w} \tag{2}$$

where λ is an eigenvalue of $\mathbf{C}_{\mathrm{H}}$, and $\mathbf{w}$ is the corresponding eigenvector satisfying the condition $\lambda \geq 0$ and $\mathbf{w} \neq 0$ respectively.

Expressing the eigenvector as a linear combination of the mapped data points we have:

$$\mathbf{w} = \sum_{i=1}^{N}\alpha_i \Phi(\mathbf{x}_i) \tag{3}$$

Using $\Phi(\mathbf{x}_k)$ to multiply both sides of (2) gives:

$$\lambda\langle \Phi(\mathbf{x}_k), \mathbf{w}\rangle = \langle \Phi(\mathbf{x}_k), \mathbf{C}_H\mathbf{w}\rangle \tag{4}$$

Substituting (1) and (3) in (4) we have

$$\lambda\sum_{i=1}^{N}\alpha_i\langle \Phi(\mathbf{x}_k), \Phi(\mathbf{x}_i)\rangle = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\left\langle \Phi(\mathbf{x}_k), \sum_{j=1}^{N}\Phi(\mathbf{x}_j)\right\rangle\langle \Phi(\mathbf{x}_j), \Phi(\mathbf{x}_i)\rangle \tag{5}$$

To avoid carrying out the nonlinear mapping and the explicit computation of inner products in the feature space before eigen-decomposition, principal components can be obtained using the kernel trick by defining an $N \times N$ matrix (Scholkopf et al., 1998; Lee et al., 2004):

$$[\mathbf{K}]_{ij} = K_{ij} = \langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j)\rangle = k(\mathbf{x}_i, \mathbf{x}_j) \tag{6}$$

for all $i, j = 1, \ldots, N$. The kernel matrix is centred as follows:

$$\mathbf{K}_c = \mathbf{K} - \mathbf{R}\mathbf{K} - \mathbf{K}\mathbf{R} + \mathbf{R}\mathbf{K}\mathbf{R} \tag{7}$$

where $\mathbf{R}$ is an $N \times N$ matrix in which each element is equal to $\frac{1}{N}$. With the centred kernel matrix, we can re-write (5) as

$$\lambda\sum_{i=1}^{N}\alpha_i\mathbf{K}_{ki} = \frac{1}{N}\sum_{i=1}^{N}\alpha_i\sum_{j=1}^{N}\mathbf{K}_{kj}\mathbf{K}_{ji} \tag{8}$$

This can be expressed as

$$N\lambda\boldsymbol{\alpha} = \mathbf{K}_c\boldsymbol{\alpha} \tag{9}$$

The $N \times N$ centred kernel matrix, $\mathbf{K}_c$ is symmetric, which has N eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_N$ associated with N orthogonal eigenvectors, $\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^N$ satisfying

$$\left\langle \boldsymbol{\alpha}^i, \boldsymbol{\alpha}^j\right\rangle = \delta_{i,j}, (i, j = 1, 2, \ldots, N) \tag{10}$$

Let $\mathbf{S} \in \Re^{N\times r}$ contain eigenvectors corresponding to the first r eigenvalues. In order to avoid the singularity problem in the CVA, the reduced kernel matrix $\mathbf{Z}$ is defined as follows:

$$\mathbf{Z} = \mathbf{S}^T\mathbf{K}_c \in \Re^{r\times N} \tag{11}$$

We then applied CVA on $\mathbf{Z}$ by treating it as a collected $r \times N$ data to extract state variables and residuals used to compute the monitoring statistics.

2.2 CVA in Kernel Principal Component Space

Given $\mathbf{z}_k$ as the k^{th} column vector of $\mathbf{Z}$, information from the past (p) and future (f) data series are obtained from (12),

$$\mathbf{z}_{p,k} = \begin{bmatrix} \mathbf{z}_{k-1} \\ \mathbf{z}_{k-2} \\ \vdots \\ \mathbf{z}_{k-p} \end{bmatrix} \in \Re^{rp} \quad \text{and} \quad \mathbf{z}_{f,k} = \begin{bmatrix} \mathbf{z}_{t} \\ \mathbf{z}_{k+1} \\ \vdots \\ \mathbf{z}_{k+f-1} \end{bmatrix} \in \Re^{rf} \tag{12}$$

Each component is then normalized to have a mean of 0 as follows:

$$\hat{\mathbf{z}}_{p,k} = \mathbf{z}_{p,k} - \bar{\mathbf{z}}_{p,k} \quad \text{and} \quad \hat{\mathbf{z}}_{f,k} = \mathbf{z}_{f,k} - \bar{\mathbf{z}}_{f,k} \tag{13}$$

where $\bar{\mathbf{z}}_{p,k}$ and $\bar{\mathbf{z}}_{f,k}$ denote the sample means of $\mathbf{z}_{p,k}$ and $\mathbf{z}_{f,k}$ respectively. To obtain the past and and future Hankel matrices, $\mathbf{G}_p$ and $\mathbf{G}_f$ respectively, the corresponding past and future vectors are arranged together in columns.

$$\mathbf{G}_p = [\hat{\mathbf{z}}_{p,p+1}, \hat{\mathbf{z}}_{p,p+2}, \dots \hat{\mathbf{z}}_{p,p+B}] \in \Re^{rp \times B} \tag{14}$$

$$\mathbf{G}_f = [\hat{\mathbf{z}}_{f,p+1}, \hat{\mathbf{z}}_{f,p+2}, \dots \hat{\mathbf{z}}_{f,p+B}] \in \Re^{rf \times B} \tag{15}$$

where the columns of the truncated Hankel matrices for N observations is given by $B = N - f - p + 1$. The sample covariances and cross-covariances of the past and future matrices are estimated as show below:

$$\Sigma_{pp} = \frac{1}{B-1}\mathbf{G}_p\mathbf{G}_p^T \tag{16}$$

$$\Sigma_{ff} = \frac{1}{B-1}\mathbf{G}_f\mathbf{G}_f^T \tag{17}$$

$$\Sigma_{fp} = \frac{1}{B-1}\mathbf{G}_f\mathbf{G}_p^T \tag{18}$$

To avoid the singularity of Σ_{pp} and Σ_{ff}, parameters r, p and f have to satisfy $\{rp, rf\} < N - p - f + 1$. The canonical variates can then be obtained by performing Singular Value Decomposition (SVD) on the scaled Hankel matrix, $\mathbf{H}_m$

$$\mathbf{H}_m = \Sigma_{ff}^{-1/2}\Sigma_{fp}\Sigma_{pp}^{-1/2} = \mathbf{U}\boldsymbol{\Delta}\mathbf{V}^T \tag{19}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices and $\boldsymbol{\Delta}$ is a diagonal matrix whose singular values indicate the degree of correlation between pairs of $\mathbf{U}$ and $\mathbf{V}$. By sorting the singular values in descending order and reordering the columns of the associated singular vectors, the first q columns of $\mathbf{V}$ can be considered as having the top pairwise correlation with those of $\mathbf{U}$. This generates a new matrix $\mathbf{V}_q$ of a smaller dimension such that $(q < rp)$.

The transformation matrices $\mathbf{C}$ and $\mathbf{D}$ used to convert the rp-dimensional past matrices to the q-dimensional state variables and the residuals respectively are computed as follows:

$$\mathbf{C} = \mathbf{V}_q^T\Sigma_{pp}^{-1/2} \in \Re^{q \times rp} \tag{20}$$

$$\mathbf{D} = \left(I - \mathbf{V}_q\mathbf{V}_q^T\right)\Sigma_{pp}^{-1/2} \in \Re^{rp \times rp} \tag{21}$$

The state space $\mathbf{Z}^*$ and residual space $\mathbf{E}$ are computed using (22):

$$\mathbf{Z}^* = \mathbf{C} \cdot \mathbf{G}_p \in \Re^{q \times B} \quad \text{and} \quad \mathbf{E} = \mathbf{D} \cdot \mathbf{G}_p \in \Re^{rp \times B} \tag{22}$$

Similar to the traditional CVA, the Hotellings T^2 and the Q statistic or squared prediction error (SPE) and their control limits are used in KCVA-based process monitoring. The Hotellings T^2 is used to monitor variations inside the state space while the Q statistic is used to monitor the variations in the residual space. They are computed using (23)

$$T_k^2 = \sum_{i=1}^{q} \mathbf{z}_{i,k}^{*2} \quad \text{and} \quad Q_k = \sum_{i=1}^{rp} \mathbf{e}_{i,k}^2 \tag{23}$$

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

Since measurements in a nonlinear process do not follow the Gaussian distribution, the control limits are better determined from the actual probability density functions of the monitoring indices using a non-parametric technique such as kernel density estimation (Odiowei and Cao, 2010a).

Assuming a set of data points $\mathbf{x}_i, i = 1, 2, \dots, N$, the kernel density estimate at point x is defined by:

$$\hat{g}(x) = \frac{1}{NH}\sum_{i=1}^{N} K\left(\frac{x - x_i}{H}\right) \tag{24}$$

where K and H are the kernel function and bandwidth respectively and x_i is each of the data points. The control limit c at a given confidence level α is given by

$$P(x < c) = \int_{-\infty}^{c} g(x)\,\mathrm{d}x = \alpha \tag{25}$$

Hence, the control limits for T^2 and Q for a given α can be computed such that $P\left(T^2 < T_\alpha^2\right) = \alpha$ and $P(Q < Q_\alpha) = \alpha$ using (26)

$$\int_{-\infty}^{T_\alpha^2} p(T^2)\,\mathrm{d}T^2 = \alpha \quad \text{and} \quad \int_{-\infty}^{Q_\alpha} p(Q)\,\mathrm{d}Q = \alpha \tag{26}$$

2.3 Summary of KCVA process monitoring procedure

Off-line training

(1) Acquire normal operation condition data, construct kernel matrix and determine reduced kernel component data according to (11).
(2) Compute past and future data series from the reduced kernel component space using (12).
(3) Compute Hankel matrices and obtain their covariances and cross-covariance from (14) to (18)
(4) Perform SVD on scaled Hankel matrix using (19) and determine number of states to retain.
(5) Determine state variables and residuals using (22).
(6) Compute monitoring indices using (23) and their control limits using (26) respectively.

On-line monitoring

(1) Acquire test data, construct kernel matrix and arrange data similar to training data.
(2) Calculate state and residual of real time test data by multiplying the transformation matrices by the past component space vectors, i.e. $\mathbf{z}^* = \mathbf{C}.\hat{\mathbf{z}}_{p,k}$ and $\mathbf{e} = \mathbf{D}.\hat{\mathbf{z}}_{p,k}$ respectively .
(3) Compute T^2 and Q of test data using (23).

(4) 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.

3. CASE STUDY

The newly proposed KCVA approach was compared against the CVA with KDE approach by applying them to the TE process.

3.1 Overview of Tennessee Eastman process

The TE process is a simulation of an actual nonlinear dynamic industrial plant (Downs and Vogel, 1993). It is commonly used for benchmark process monitoring and control studies involving the evaluation and comparison of newly developed approaches (Chiang et al., 2001). It has five main units (separator, compressor, reactor, stripper and condenser) and eight components (A to H). A schematic diagram of the process is shown in Fig. 1. The process involves a total of $N = 960$ observations and

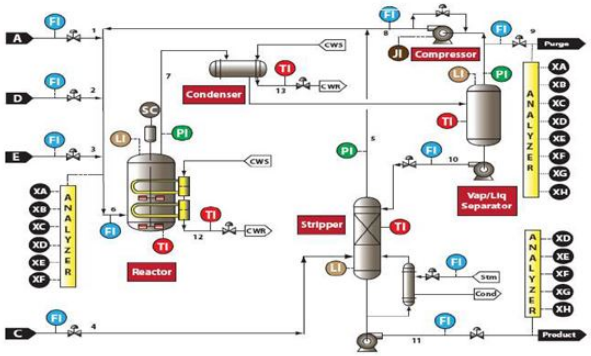

Fig. 1. Schematic diagram of the TE process

52 variables which include 12 manipulated variables, 22 measured variables with a sampling interval of 3 minutes, and 19 composition measurements. Each of the 21 pre-programmed faults is introduced into the process at sample number 161.

3.2 Important implementation details

The data obtained under normal operating condition were used as the training data set while the data obtained under each of the faulty operating conditions were used as test data. All the 22 measured and 11 manipulated variables, and Faults 1 to 20 (Chiang et al., 2001) were included in this study. The agitation speed of the reactor's stirrer which is the 12^{th} manipulated variable was not considered because it is constant.

The radial basis function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = e^{\|\mathbf{x}-\mathbf{y}\|^2/c}$ was used in this study. The width of the kernel function was empirically set at 1720 after several tests to strike a reasonable balance between detection and false alarm rates.

Since the effective rank of a kernel matrix is much lower than its size which give rise to an ill-conditioned optimization problem (Huang et al., 2009), to obtain a suitable kernel form and avoid the need for regularisation, the dimension of the kernel component space was set to the number of principal components whose cumulative eigenvalue sum accounted for over 99% of the total sum of eigenvalues to ensure little loss of information. This set the KPCA space to a dimension of $r = 60$ which was used as the reduced kernel component space for applying the CVA.

To successfully, develop the CVA and KCVA models to characterize the variability of the off-line data requires that the number of time lags for the past and future measurements and the number of states to be retained are determined. The lag order represents the number of past measurements that are significantly correlated with a measurement at a given time instant. In this study, $p = f = 15$ lags were adopted after several trials. This satisfies $rp = rf = 900 < N - p - f + 1 = 931$.

Two methods commonly suggested in the literature for determining the number of states to retain are based on the dominant singular values (Negiz and Cinarl, 1998) and the Akaike Information Criterion (AIC) (Chiang et al., 2001). Fig. 2 shows the normalized singular values obtained from the scaled Hankel matrix. Since the singular values decrease very slowly in this case, fixing the number of states to retain on the basis of the dominant singular values will give an unrealistic model (Odiowei and Cao, 2010a). Moreover, the number of states retained is not critical in this study because the Hotelling's T^2 and SPE (Q) statistics were used jointly for fault detection. This implies that abnormal deviations not captured in the model space could be captured in the residual space and vice versa. Hence, the 26 states reported in (Odiowei and Cao, 2010a) was adopted for both CVA and KCVA in this work. Control charts based on the T^2 and and SPE were constructed for process monitoring at 99% confidence level. All control limits were computed based on kernel density estimation.

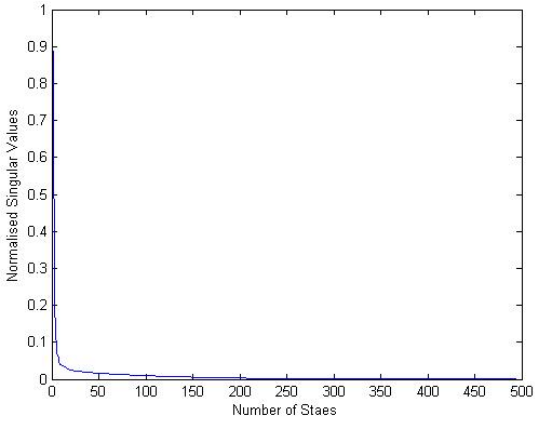

Fig. 2. Normalised singular values of training data

Monitoring performance was based on three indices; fault detection rates (FDRs), false alarm rates (FARs) and detection delay. FDR is the percentage of fault samples identified correctly. It was computed as

$$FDR = \frac{N_{fc}}{N_{Tf}} \times 100 \tag{27}$$

where N_{fc} denotes the number of fault samples identified correctly and N_{Tf} is the total number of fault samples.

Table 1. FDRs (%) and Detection Delay (min) for CVA and KCVA

Fault	CVA with KDE		KCVA	
	FDR	Detection Delay	FDR	Detection Delay
1	99.88	0	100	0
2	99.75	6	99.88	3
3	76.88	9	100	0
4	100	0	100	0
5	100	0	100	0
6	100	0	100	0
7	100	0	100	0
8	99.00	24	99.25	18
9	92.38	30	98.75	30
10	96.75	81	97.50	60
11	99.25	12	99.63	9
12	99.63	9	99.75	6
13	96.25	90	96.50	84
14	100	0	100	0
15	99.75	6	99.88	3
16	99.38	15	99.50	15
17	98.38	39	98.50	36
18	99.38	15	99.50	12
19	100	0	100	0
20	97.75	57	97.50	60

A technique with a higher FDR is acknowledged to be better than one with a lower value. FAR was calculated as the percentage of normal samples identified as faults (or abnormal) during the normal operation of the plant.

$$FAR = \frac{N_{nf}}{N_{Tn}} \times 100 \tag{28}$$

where N_{nf} represents the number of normal samples identified as faults and N_{Tn} is the total number of normal samples.

Detection delay is the time that elapses before a fault is detected after its introduction. In this study, fault detection for the purpose of computing the detection delay was based on the criterion that a fault is detected when a monitoring index value exceeds its control limit in at least three consecutive samples.

3.3 Results and discusion

The FARs for CVA with KDE based T^2 and SPE statistics were 0.0068 and 0.0137 respectively while KCVA based FARs were 0.0068 and 0 for T^2 and SPE respectively. The FAR of KCVA based SPE had the lowest false alarm rate.

Table 1 shows the detection rates and detection delay results for CVA with KDE and KCVA respectively for all 20 faults studied. The results, show that KCVA has better overall performance than CVA-KDE in terms of both FDRs and detection delay. Faults 3 snd 9 are among the faults that are more difficult to detect in the TE process because they cause very little variation in the measured process variables (Odiowei and Cao, 2010a). Fault 3 is a step change in the D feed temperature. The monitoring charts for this fault are shown in Figs. 3 and 4. It can be seen that KCVA gave comparable results in both the T^2 and SPE respectively. All the monitoring indices are clearly above the horizontal line (the control limit), indicating the presence of a fault. However, the CVA-based T^2 is almost completely below the control limit. If the fault detection criterion used for detection delay is applied strictly, the CVA-based T^2 will completely miss detecting Fault 3. The SPE gave a detection rate of 76.88% which is far lower than KCVA's 100% detection rate. Fault 9 is a

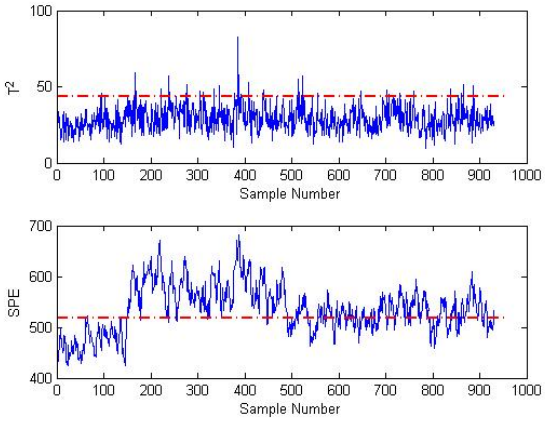

Fig. 3. CVA-based monitoring charts for Fault 3

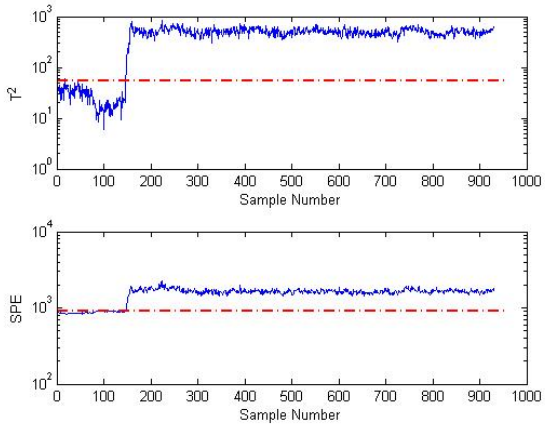

Fig. 4. KCVA-based monitoring charts for Fault 3

random variation of the feed temperature of component D. The monitoring charts for this fault are shown in Figs. 5 and 6. Again, though the CVA based SPE detected the fault, the SPE performed badly. Conversely, all the KCVA indices detected the fault very clearly.

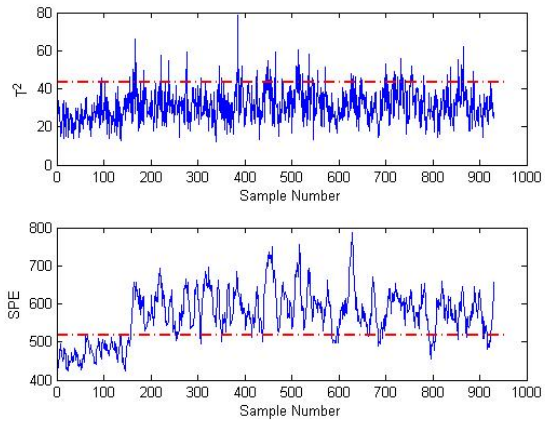

Fig. 5. CVA-based monitoring charts for Fault 9

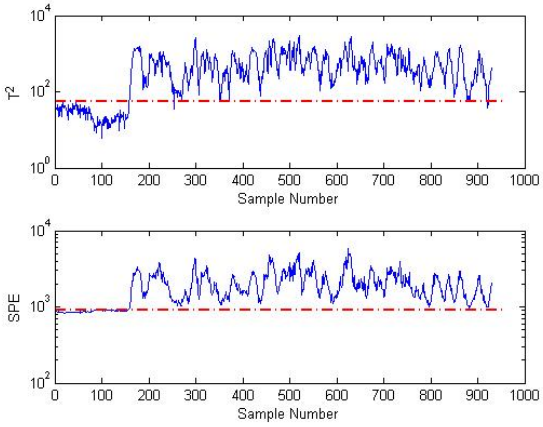

Fig. 6. KCVA-based monitoring charts for Fault 9

4. CONCLUSION

The CVA was kernelized in this study to obtain the novel kernel CVA. The proposed technique consists of two steps: kernel PCA and CVA. The kernel PCA addresses process nonlinearities while the CVA accounts for the dynamics of the process. Testing the technique on the Tennessee Eastman challenge problem for fault detection showed that it outperformed the CVA with KDE approach in all three indices investigated: fault detection rate, false alarms, and detection delay, especially in faults that are more difficult to detect, despite the latter's reported superior monitoring performance over other dynamic methods such as dynamic PCA and dynamic PLS. Furthermore, placing the CVA in a kernel generated feature space makes the implementation of the proposed technique significantly simple. It does not require regularization of the kernel data to avoid numerical instability which is usually done in optimisation problems involving covariance matrices based on kernel data.

Although, the results of this study clearly show the superior performance of the proposed technique over other nonlinear dynamic monitoring approaches, some things require further attention. Principally, design parameters such as the type of kernel function and the kernel window width selected, how many components are retained in the kernel space, the number of time lags employed and the number of states retained are likely to affect the monitoring performance of the technique. Therefore, the sensitivity of the technique to changes in these parameters and more rigorous ways of selecting their optimum values should be considered in future studies. Furthermore, it will also be worthwhile to address the application of the technique in fault identification. Results of more monitoring approaches also need to be captured to provide more data for comparison.

REFERENCES

Chiang, L., Russell, E., and Braatz, R. (2001). *Fault detection and diagnosis in industrial systems*. Springer, London.

Choi, S.W., Lee, C., Lee, J.M., Park, J.H., and Lee, I.B. (2005). Fault detection and identification of nonlinear processes based on kernel pca. *Chemometrics and Intelligent Laboratory Systems*, 75, 55–67.

Downs, J. and Vogel, E. (1993). A plant-wide industrial process control problem. *Computers and Chemical Engineering*, 17(3), 245–255.

Haykin, S. (1999). *Neural Networks*. Prentice Hall International, Inc., New Jersey.

Huang, S.Y., Lee, M.H., and Hsiao, C.K. (2009). Nonlinear measures of association with kernel canonical correlation analysis and applications. *Journal of Statistical Planning and Inference*, 139, 2162–2174.

Komulainen, T., Sourander, M., and Jms-Jounela, S.L. (2004). An online application of dynamic pls to a dearomatization process. *Computer and Chemical Engineering*, 28, 2611–2619.

Ku, W., Storer, R.H., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 30, 179–196.

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

Lu, J. and Liu, C.L. (2006). Statistical modeling of dynamic multivariate process using canonical variate analysis. In *Proceedings of International Conference on Information and Automation*, 218–221. IEEE, Shandong, China.

Negiz, A. and Cinarl, A. (1998). Monitoring of multivariable dynamic processes and sensor auditing. *Journal of Process Control*, 8(5-6), 375–380.

Odiowei, P.E.P. and Cao, Y. (2010a). Nonlinear dynamic process monitoring using canonical variate analysis and kernel density estimation. *IEEE Transaction on Industrial Informatics*, 6(1), 36–44.

Odiowei, P. and Cao, Y. (2010b). State-space independent component analysis for nonlinear dynamic process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 103, 59–65.

Qin, S.J. (2003). Statistical process monitoring: basic and beyond. *Journal of Chemometrics*, 17, 480–502.

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

Samuel, R.T. and Cao, Y. (2014). Fault detection in a multivariate process based on kernel pca and kernel density estimation. In *International Conference on Automation and Computing*. IEEE, Cranfield, UK.

Scholkopf, B., Smola, A., and Muller, K. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10, 1299–1399.

Scholkopf, B. and Smola, A.J. (2002). *Learning with Kernels Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press Cambridge, Massachusetts London, England.

Tan, S., Wang, F., Chang, Y., Chen, W., and Xu, J. (2010). Fault detection and diagnosis of nonlinear processes based on kernel ica-kcca. In *Chinese Control and Decision Conference*, 3869–3874. IEEE, Xuzhou.

Yin, S., , Li, X., , Gao, H., and Kaynak, O. (2015). Data-based techniques focused on modern industry: An overview. *IEEE Transactions on Industrial Electronics*, 62(1), 657–667.