IEEE Transactions on Industrial Informatics, Volume 6, Issue 1, February 2010, Pages 36-45

Nonlinear Dynamic Process Monitoring using Canonical Variate Analysis and Kernel Density Estimations

P.P. Odiowei and Y. Cao*

School of Engineering, Cranfield University, Bedford, MK43 0AL, UK

Abstract

The Principal Component Analysis (PCA) and the Partial Least Squares (PLS) are two commonly used techniques for process monitoring. Both PCA and PLS assume that the data to be analysed are not self-correlated i.e. time-independent. However, most industrial processes are dynamic so that the assumptions of time-independence made by the PCA and the PLS are invalid in nature. Dynamic extensions to PCA and PLS, so called DPCA and DPLS, have been developed to address this problem, however, unsatisfactorily. Nevertheless, the Canonical Variate Analysis (CVA) is a state-space based monitoring tool, hence is more suitable for dynamic monitoring than DPCA and DPLS. The CVA is a linear tool and traditionally for simplicity, the upper control limit (UCL) of monitoring metrics associated with the CVA is derived based on a Gaussian assumption. However, most industrial processes are non-linear and the Gaussian assumption is invalid for such processes so that CVA with a UCL based on this assumption may not be able to correctly identify underlying faults. In this work, a new monitoring technique using the CVA with UCLs derived from the estimated probability density function through kernel density estimations (KDE) is proposed

* Corresponding author, email address: y.cao@cranfield.ac.uk

and applied to the simulated nonlinear Tennessee Eastman Process Plant. The proposed CVA with KDE approach is able to significantly improve the monitoring performance and detect faults earlier when compared to other methods also examined in this study.

Keywords: Canonical Variate Analysis, Probability Density Function, Kernel Density Estimation, Process Monitoring

1. INTRODUCTION

Process monitoring is essential to maintain high quality products as well as process safety. Widely applied process monitoring techniques like the PCA and the PLS rely on static models, which assume that the observations are time independent and follow a Gaussian distribution. However, the assumptions of time-independence and normality are invalid for most chemical processes because variables driven by noise and disturbances are strongly auto-correlated and most plants are nonlinear in nature. Therefore, the static PCA and PLS based approaches are inappropriate to monitor such nonlinear dynamic processes.

To extend PCA applications to dynamic systems, Ku et al.[1] presented a study of PCA on lagged variables to develop dynamic models and Multivariate Statistical Process Monitoring (MSPM) tools for dynamic continuous processes. In this so called Dynamic PCA (DPCA) approach, Ku et al.[1] used parallel analysis to determine the number of time-lagged value for the process variables as well as the number of principal components to retain in the DPCA model. Although dynamic models are developed in DPCA and faults are detected, diagnosis of abnormal behaviour is more

complicated with DPCA given that lagged variables are involved[2]. It is also reported that principal components extracted in this way are not necessarily the minimal dynamic representations[3]. Furthermore, Komulainen[4] extended PLS applications to dynamic systems, in a similar way to the DPCA, for the monitoring of an online industrial dearomatization process. The extended PLS approach is known as the Dynamic PLS (DPLS). Although the DPLS technique was reported to be efficient for fault detection, like the DPCA, the capability of the DPLS to identify dynamic faults is still questionable because the way of the DPCA and DPLS to represent a dynamic system is not efficient and may not be able to capture some important dynamic behaviours of the system. .

More recently, monitoring techniques based on Canonical Variate Analysis (CVA) have been developed with UCLs derived based on the Gaussian assumption[5,6,7]. CVA was first introduced in 1936 by Hotelling[7], adopted for use in dynamic systems for a limited class of processes by Akaike in 1975[7,8] and adapted to general linear systems by Larimore in 1983[8]. CVA is a state space based MSPM method, hence is more appropriate for dynamic process monitoring.

Norvalis et al.[7] developed a process monitoring and fault diagnosis tool that combined canonical variate state space (CVSS) models with knowledge based systems (KBS) for monitoring multivariate process operations. Faults were detected using the CVSS models and then UCLs derived based on the Gaussian assumption while diagnosis was based on the KBS. The efficiency of the technique was illustrated by monitoring simulated data of a polymerisation reactor system.

Juan and Fei[6] employed CVA for fault detection based on Hotelling's T^2 charts to monitor a chemical separation plant. The results from the study illustrated a good performance of the statistical model based on CVA. Furthermore, it was demonstrated that the precision of the CVA model improved with an increase in the length of the data employed for the CVA analysis.

Different from the above mentioned studies, Chiang et al.[5] employed canonical variate analysis to include the input and output variables for the estimation of the state space variable. From the estimated state space variable, UCLs of T^2 and Q metrics were determined to judge whether or not those processes were in-control.

The T^2 and Q metrics are widely employed with various MSPM techniques[1,3,5,9,10,11,12]. For linear MSPM techniques, such as PCA, PLS and CVA, traditionally, UCLs of the T^2 and Q metrics are estimated based on an assumption that the latent or state variables follow a Gaussian distribution. However, most industrial processes are nonlinear. For such processes, although the distribution of stochastic sources might be Gaussian, such as measurement noises and normally distributed disturbances, the distribution of process variables, in general, will be non-Gaussian. In such a case, the UCL estimated based on the Gaussian assumption is unable to correctly identify underlying faults.

The problem of monitoring non-Gaussian processes can be addressed by directly estimating the underlying probability density function (PDF) of the T^2 and Q metrics through the kernel density estimation (KDE) to derive the correct UCL[13,14].

Martin and Morris[13] presented an overview of multivariate process monitoring techniques using the PCA and the PLS with T^2 and M^2 metrics for process monitoring. The control limit of M^2 metric was estimated based on the PDF, combining techniques of standard bootstrap and kernel density estimations to overcome the limitations of the T^2 metric mentioned above. Both methodologies were applied to a continuous polyethylene reactor and a polymerisation reactor to demonstrate the efficiencies of both methodologies and the M^2 metric was reported to be a more efficient process monitoring tool than the T^2 metric.

Chen et al.[14] adopted several KDE approaches in association with PCA for process monitoring. A gas melter process was used as the case study and it was demonstrated that the KDEs could obtain nonparametric empirical density function as a tool for a more efficient process monitoring. Their emphasis was to demonstrate the efficiencies of three different density estimators which were verified based on the misclassification rates at given confidence intervals.

In order to use the linear dynamic tools, such as the CVA to monitor nonlinear dynamic processes, the limitation of the Gaussian assumption based T^2 and Q metrics mentioned above has to be addressed. In this paper, KDE is employed in association with the CVA resulting in a new extension of the CVA algorithm, the 'CVA with KDE' for process monitoring. To achieve this, a CVA model is firstly estimated from the so called past and future variables constructed from the collected process data. From the estimated CVA model, the T^2 and Q metrics are then calculated and the KDE is employed to estimate the PDF of these T^2 and Q metrics calculated. UCLs are then determined based on the estimated PDF for a given confidence bound. For

comparison, different monitoring algorithms; DPCA and DPLS with and without KDE as well as CVA with and without KDE have been applied to the simulated nonlinear Tennessee Eastman Process Plant in the present study. Results show that the monitoring performance is significantly improved by using the 'CVA with KDE' approach compared with other monitoring algorithms aforementioned. Although the CVA is a linear model, in this study, the CVA is employed to monitor a nonlinear dynamic process plant. Hence, this study is described as nonlinear dynamic process monitoring.

The rest of the paper is organised as follows: Section 2 explains the CVA model while section 3 describes monitoring metrics and their UCLs derived through Kernel Density Estimations. The procedure of CVA with KDE is then summarised in section 4. Section 5 describes the case study whilst the results of the case study are presented and discussed in section 6. Finally, the work is concluded in section 7.

2. CANONICAL VARIATE ANALYSIS

Canonical Variate Analysis (CVA) is a linear dimension reduction technique to construct a minimum state space model for dynamic process monitoring. This section applies the linear CVA algorithm to a nonlinear dynamic plant for identifying state variables directly from the process measurements.

Assume the nonlinear dynamic plant under consideration represented as follows.

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{f}(\mathbf{x}_k) + \mathbf{w}_k \\ \mathbf{y}_k &= \mathbf{g}(x_k) + \mathbf{v}_k \end{aligned} \tag{1}$$

where $\mathbf{x_k} \in R^n$ and $\mathbf{y_k} \in R^m$ are state and measurement vectors respectively, $\mathbf{f}(\cdot)$ and $\mathbf{g}(\cdot)$ are unknown nonlinear functions, while w_k and v_k are plant disturbances and measurement noise vectors respectively. It is clear that such an unknown nonlinear dynamic system is generally difficult to deal with for monitoring. However, at a stable normal operating point, the nonlinear plant can be approximated by a linear stochastic state space model as follows;

$$\begin{aligned} \mathbf{x_{k+1}} &= \mathbf{A}\mathbf{x_k} + \boldsymbol{\varepsilon}_k \\ \mathbf{y_k} &= \mathbf{C}\mathbf{x_k} + \boldsymbol{\eta}_k \end{aligned} \tag{2}$$

where $\mathbf{A}$ and $\mathbf{C}$ are unknown state and output matrices respectively while $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are collective modelling errors partially due to the underlying nonlinearity of the plant which has not been included in the linear model, as well as associated with process disturbance and measurement noise, $\mathbf{w}_k$ and $\mathbf{v}_k$ respectively. Due to the unknown nonlinearity, the collective modelling errors, $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ generally will be non-Gaussian although $\mathbf{w}_k$ and $\mathbf{v}_k$ might be normally distributed processes. This is the main difference of this work from other CVA based approaches reported in literature. Instead of dealing with the unknown nonlinear system (1) directly, in this work, the approximated linear state space model given in (2) is considered through the standard CVA approach. Although the linear model (2) is easier to deal with than the nonlinear system (1), the collective errors $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ have to be treated as non-Gaussian processes. This leads to the direct PDF estimation of the associated T^2 and Q metrics through the KDE approach explained in section 3.

In the CVA approach, firstly, the measurement vector $\mathbf{y}_k$ is expanded by q past and future measurements to give the past and future observation vectors $\mathbf{y}_{p,k}$ and $\mathbf{y}_{f,k}$ respectively.

$$\mathbf{y}_{p,k} = \begin{bmatrix} \mathbf{y}_{k-1} \\ \mathbf{y}_{k-2} \\ \vdots \\ \mathbf{y}_{k-q} \end{bmatrix} \in R^{mq}, \quad \tilde{\mathbf{y}}_{p,k} = \mathbf{y}_{p,k} - \overline{\mathbf{y}}_{p,k} \tag{3}$$

$$\mathbf{y}_{f,k} = \begin{bmatrix} \mathbf{y}_{k} \\ \mathbf{y}_{k+1} \\ \vdots \\ \mathbf{y}_{k+q-1} \end{bmatrix} \in R^{mq}, \quad \tilde{\mathbf{y}}_{f,k} = \mathbf{y}_{f,k} - \overline{\mathbf{y}}_{f,k} \tag{4}$$

where $\overline{\mathbf{y}}_{p,k}$ and $\overline{\mathbf{y}}_{f,k}$ are the sample means of $\mathbf{y}_{p,k}$ and $\mathbf{y}_{f,k}$ respectively, and the products of mq represents the lengths of the past and future observation vectors respectively. The length of the past and future observations can be determined by checking the autocorrelation of the square sum of the process variables such that the correlation can be neglected when the time distance is larger than the number of lags determined.

These past and future observations are stochastic processes. Their sample-based covariance and cross-covariance matrices can be estimated through the truncated Hankel matrices as follows;

$$\mathbf{\Sigma}_{pp} := (M-1)^{-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{p,k} \tilde{\mathbf{y}}_{p,k}^T = \mathbf{Y}_p \mathbf{Y}_p^T (M-1)^{-1} \tag{5}$$

$$\mathbf{\Sigma}_{ff} := (M-1)^{-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{f,k} \tilde{\mathbf{y}}_{f,k}^T = \mathbf{Y}_f \mathbf{Y}_f^T (M-1)^{-1} \tag{6}$$

$$\boldsymbol{\Sigma}_{fp} := (M-1)^{-1} \sum_{k=q+1}^{q+M} \widetilde{\mathbf{y}}_{f,k} \widetilde{\mathbf{y}}_{p,k}^{T} = \mathbf{Y}_f \mathbf{Y}_p^T (M-1)^{-1} \tag{7}$$

where $\mathbf{Y}_p$ and $\mathbf{Y}_f$ are past and future truncated M-column Hankel matrices respectively, and defined as follows.

$$\mathbf{Y}_p = [\widetilde{\mathbf{y}}_{p,q+1} \quad \widetilde{\mathbf{y}}_{p,q+2} \quad \cdots \quad \widetilde{\mathbf{y}}_{p,q+M}] \in R^{mq \times M} \tag{8}$$

$$\mathbf{Y}_f = [\widetilde{\mathbf{y}}_{f,q+1} \quad \widetilde{\mathbf{y}}_{f,q+2} \quad \cdots \quad \widetilde{\mathbf{y}}_{f,q+M}] \in R^{mq \times M} \tag{9}$$

For a set of measurements with total N observations, the last element of $\mathbf{y}_{p,q+1}$ in (3) is $\mathbf{y}_1$, whilst the last element of $\mathbf{y}_{f,q+M}$ in (4) should be $\mathbf{y}_N$. Therefore, the maximum number of columns of these Hankel matrices is

$$M = N - 2q + 1 \tag{10}$$

The CVA aims to find the best linear combinations, $\mathbf{a}^T(\widetilde{\mathbf{y}}_{f,k})$ and $\mathbf{b}^T(\widetilde{\mathbf{y}}_{p,k})$ of the future and past observations so that the correlation between these combinations is maximised. The correlation can be represented as follows:

$$\rho_{fp}(\mathbf{a},\mathbf{b}) = \frac{\mathbf{a}^T \boldsymbol{\Sigma}_{fp} \mathbf{b}}{\left(\mathbf{a}^T \boldsymbol{\Sigma}_{ff} \mathbf{a}\right)^{1/2} \left(\mathbf{b}^T \boldsymbol{\Sigma}_{pp} \mathbf{b}\right)^{1/2}} \tag{11}$$

Let $\mathbf{u} = \boldsymbol{\Sigma}_{ff}^{1/2} \mathbf{a}$ and $\mathbf{v} = \boldsymbol{\Sigma}_{pp}^{1/2} \mathbf{b}$. The optimization problem can be casted as:

$$\begin{aligned} \max_{\mathbf{u},\mathbf{v}} \quad & \mathbf{u}^T \left(\boldsymbol{\Sigma}_{ff}^{-1/2} \boldsymbol{\Sigma}_{fp} \boldsymbol{\Sigma}_{pp}^{-1/2}\right) \mathbf{v} \\ \text{s.t.} \quad & \mathbf{u}^T \mathbf{u} = 1 \\ & \mathbf{v}^T \mathbf{v} = 1 \end{aligned} \tag{12}$$

According to linear algebra theory, the solution, $\mathbf{u}$ and $\mathbf{v}$ are left and right singular vectors of the scaled Hankel matrix, $\mathbf{H} := \boldsymbol{\Sigma}_{ff}^{-1/2} \boldsymbol{\Sigma}_{fp} \boldsymbol{\Sigma}_{pp}^{-1/2}$ and the maximal correlation

$\sigma = \max_{\mathbf{a},\mathbf{b}} \rho_{fp}(\mathbf{a},\mathbf{b})$ is the corresponding singular value of $\mathbf{H}$. If the rank of the scaled Hankel matrix, $\mathbf{H}$ is r, then there are r non-zero singular values, σ_i, $i = 1,2,\ldots r$ in the descending order and correspondingly r pairs of the left and right singular vectors, $\mathbf{u}_i$ and $\mathbf{v}_i$ for $i = 1,2,\ldots,r$. Singular values and vectors can be collected in the following matrix form of the singular value decomposition (SVD).

$$\mathbf{H} = \boldsymbol{\Sigma}_{ff}^{-1/2} \boldsymbol{\Sigma}_{fp} \boldsymbol{\Sigma}_{pp}^{-1/2} = \mathbf{U}\mathbf{D}\mathbf{V}^T \tag{13}$$

where

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \in R^{mq \times r}, \quad \mathbf{V} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} \in R^{mq \times r}, \quad \mathbf{D} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \in R^{r \times r}$$

Furthermore, the canonical variates can be directly estimated from the past observation vector $\tilde{\mathbf{y}}_{p,k}$ as illustrated in (14).

$$\mathbf{z}_k = \begin{bmatrix} \mathbf{b}_1^T \\ \mathbf{b}_2^T \\ \vdots \\ \mathbf{b}_r^T \end{bmatrix} \tilde{\mathbf{y}}_{p,k} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_r^T \end{bmatrix} \boldsymbol{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{p,k} = \mathbf{V}^T \boldsymbol{\Sigma}_{pp}^{-1/2} \tilde{\mathbf{y}}_{p,k} = \mathbf{J}\tilde{\mathbf{y}}_{p,k} \tag{14}$$

where $\mathbf{J} = \mathbf{V}^T \boldsymbol{\Sigma}_{pp}^{-1/2} \in R^{r \times mq}$ is the transformation matrix, which transforms the mq-dimensional past measurements to the r-dimensional canonical variates. These canonical variates are normalised with a unit sample covariance.

$$\frac{1}{M-1} \sum_{k=q+1}^{q+M} \mathbf{z}_k \mathbf{z}_k^T = \mathbf{V}^T \boldsymbol{\Sigma}_{pp}^{-1/2} \left(\frac{1}{M-1} \sum_{k=q+1}^{q+M} \tilde{\mathbf{y}}_{p,k} \tilde{\mathbf{y}}_{p,k}^T \right) \boldsymbol{\Sigma}_{pp}^{-1/2} \mathbf{V} = \mathbf{V}^T \boldsymbol{\Sigma}_{pp}^{-1/2} \boldsymbol{\Sigma}_{pp} \boldsymbol{\Sigma}_{pp}^{-1/2} \mathbf{V} = \mathbf{V}^T \mathbf{V} = \mathbf{I}$$

From equation (14), the canonical variate space spanned by all the estimated canonical variates can be separated into the state space and the residual space based on the order of the system. According to the magnitude of the singular values, the first *n* dominant singular values are determined and the corresponding *n* canonical variates

retained as the state variables where $n < r$. In addition, the remaining $(r-n)$ canonical variates are said to be in the residual space. Equation (15) below shows the entire canonical variate space $(\mathbf{z}_k \in R^r)$ spanned by the state variables $(\mathbf{x}_k \in R^n)$ and the residual canonical variates $(\mathbf{d}_k \in R^{r-n})$.

$$\mathbf{z}_k = [\mathbf{x}_k^T \ \mathbf{d}_k^T]^T \tag{15}$$

The state variables $(\mathbf{x}_k)$ are a subset of the canonical variates $(\mathbf{z}_k)$ estimated in (14). Hence the state variable like the canonical variates is defined as a linear combination of the past observation vector $\tilde{\mathbf{y}}_{p,k}$, $\mathbf{x}_k = \mathbf{J}_x \tilde{\mathbf{y}}_{p,k}$, where $\mathbf{J}_x = \mathbf{V}_x^T \boldsymbol{\Sigma}_{pp}^{-1/2}$ with $\mathbf{V}_x$ consisting of the first n columns of $\mathbf{V}$ defined in (13).

Like the canonical variates, the state variables also have the unit covariance. Once the states of the system are determined, the state and output matrices, $\mathbf{A}$ and $\mathbf{C}$ can then be estimated through linear least squares regression. However, the determination of the state and output matrices $\mathbf{A}$ and $\mathbf{C}$ will be omitted from the rest of the paper since these matrices will not be used in this work.

The variation of state variables can be represented by the T^2 metric. Another commonly used monitoring metric is the Q metric which measures the total sum of square errors of the variations in the residual space. The estimation and use of the T^2 and Q metrics are explained in the next section.

3. CONTROL LIMIT THROUGH KERNEL DENSITY ESTIMATIONS

Traditionally, it was assumed that $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ are normally distributed, as well as the state, measurement and residual vectors, $\mathbf{x}_k$, $\mathbf{y}_k$ and $\mathbf{e}_k$ since a linear combination of multivariate Gaussian variables is also normally distributed.

For N samples of data, the number of samples of the states available is M, given in (10). For the normally distributed n-dimensional state vector, $\mathbf{x}$, with M samples, $\mathbf{x}_k$, $k = 1,2,\ldots,M$, the T^2 statistic defined in (16) can be used to test whether the mean $\boldsymbol{\mu}$ of $\mathbf{x}$ is at the desired target $\boldsymbol{\tau}$.

$$T_k^2 = (\mathbf{x}_k - \boldsymbol{\tau})^T \mathbf{S}^{-1} (\mathbf{x}_k - \boldsymbol{\tau}), \qquad k = 1,2,\ldots,M \tag{16}$$

where $\mathbf{S}$ is the estimated covariance of $\mathbf{x}$. If $\boldsymbol{\mu} = \boldsymbol{\tau}$, then $CT^2 \sim F(n, M-n)$, where $C = M(M-n)/(M-1)(M+1)n$. Therefore, the system (2) can be monitored by plotting T_k^2 against time, k, along with a UCL, $T_{\mathrm{UCL}}^2(\alpha)$ corresponding to a significance level, α, that has the probability, $P\left(T_k^2 > T_{\mathrm{UCL}}^2(\alpha)\right) = \alpha$.

Equation (16) can be simplified as the state covariance matrix, $\mathbf{S} = \mathbf{I}$. Furthermore, since the past and future observations, $\tilde{\mathbf{y}}_{p,k}$ and $\tilde{\mathbf{y}}_{f,k}$ have zero means, the desired target for the state is $\boldsymbol{\tau} = \mathbf{0}$. With these simplifications in place, the T^2 metric for the state space is represented in (17).

$$T_k^2 = \mathbf{x}_k{}^T \mathbf{x}_k \tag{17}$$

The corresponding UCL $T_{UCL}^2(\alpha)$ for a significance level α is derived as follows

$$T_{CUL}^2(\alpha) = \frac{n(M-1)^2}{M(M-n)} F_{n,M-n}(\alpha) \tag{18}$$

where $F_{a,b}(\alpha)$ is the critical value of the F-distribution with a and b degrees of freedom for a significance level α. By comparing T_k^2 against $T_{UCL}^2(\alpha)$ in real-time, an abnormal condition is then determined when $T_k^2 > T_{UCL}(\alpha)$.

The Q metric is introduced to test the significance level of the prediction error represented in the scaled past observation space. According to (14), the prediction error for the scaled past measurement and the corresponding Q metric are then defined in (19) and (20) respectively.

$$\mathbf{e}_k = \left(\mathbf{I} - \mathbf{V}_x \mathbf{V}_x^T\right)\boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k} = \mathbf{F}\tilde{\mathbf{y}}_{p,k} \tag{19}$$

$$Q_k = \mathbf{e}_k^T \mathbf{e}_k \tag{20}$$

Given a level of significance, α, also based on the assumption of normality, the threshold, $Q_{UCL}(\alpha)$ of the Q-metric for the PCA is estimated by Jackson and Mudholkar[15] as

$$Q_{UCL}(\alpha) = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{21}$$

where $\theta_i = \sum_{j=n+1}^{r} \lambda_j^i$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, and c_α is the normal deviate corresponding to (1 – α) percentile. For the PCA, in (21), λ_j is the jth eigenvalue of the covariance of the measured data. For the CVA error represented in (19), it should be the covariance of the scaled past observations, $\boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k}$, i.e.

$$\lambda_j = \lambda\left(\sum_{k=q+1}^{q+M} (\boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k})(\boldsymbol{\Sigma}_{pp}^{-1/2}\tilde{\mathbf{y}}_{p,k})^T \right) = 1, \qquad j = 1,2,\ldots,r$$

Therefore, the calculation of $Q_{UCL}(\alpha)$ can be simplified by letting $\theta_i = (r-n)$ and $h_0 = 1/3$ in (21). By comparing Q_k against $Q_{UCL}(\alpha)$ in real-time, an abnormal condition is determined when $Q_k > Q_{UCL}(\alpha)$.

Both control limits in (18) and (21) are based on the assumptions that the state variables and prediction errors are Gaussian. However, when the collective modelling errors, $\boldsymbol{\varepsilon}_k$ and $\boldsymbol{\eta}_k$ of the system (2) are non-Gaussian processes, this assumption is not valid. Hence, $T^2_{\mathrm{UCL}}(\alpha)$ and $Q_{UCL}(\alpha)$ derived above can no longer be used as control limits for real-time monitoring. One solution to this issue is to estimate the PDF directly for these T^2 and Q metrics through a non-parametric approach[13,14]. Amongst various PDF estimating approaches, the kernel density estimation (KDE) approach[13,14] is selected for this work. The KDE is a well established approach to estimate the PDF particularly for univariate random processes[16]. Therefore, it is particularly suitable for the T^2 and Q metrics which are univariate although the underlying processes are multivariate. Assume x is a random variable and its density function is denoted by $p(x)$. This means that

$$P(x < b) = \int_{-\infty}^{b} p(x)dx \tag{22}$$

Therefore, by knowing $p(x)$, an appropriate control limit can be determined for a specific confidence bound, α using (22). The estimation of the probability density function $\hat{p}(x)$ at point x through the kernel function, $K(\cdot)$ is defined as follows.

$$\hat{p}(x) = \frac{1}{Mh}\sum_{k=1}^{M} K\left(\frac{x - x_k}{h}\right) \tag{23}$$

where x_k, $k = 1,2,\ldots,M$ are samples of x and h is the bandwidth. The bandwidth selection in KDE is an important issue because selecting a bandwidth too small will result in the density estimator being too rough, a phenomenon known as under-smoothed while selecting a bandwidth too big will result in the density estimator being too flat. There is no single perfect way to determine the bandwidth. However, a rough estimation of the optimal bandwidth h_{opt} subject to minimising the approximation of the mean integrated square error can be derived in (24), where σ is the standard deviation[17].

$$h_{\text{opt}} = 1.06\sigma N^{-1/5} \tag{24}$$

By replacing x_k with T_k^2 and Q_k obtained in equations (17) and (20) respectively, the above KDE approach is able to estimate the underlying PDFs of the T^2 and Q metrics. The corresponding control limits, $T_{\text{UCL}}^2(\alpha)$ and $Q_{UCL}(\alpha)$ can then be obtained from the PDFs of the T^2 and Q metrics for a given confidence bound, α by solving the following equations respectively.

$$\begin{aligned} \int_{-\infty}^{T_{UCL}^2(\alpha)} p(T^2)dT^2 &= \alpha \\ \int_{-\infty}^{Q_{UCL}(\alpha)} p(Q)dQ &= \alpha \end{aligned} \tag{25}$$

The T^2 and Q metrics are complementary. A fault may cause a significant deviation in the state space but not necessary results in a similar level of significance in the error space, vice versa. Therefore, in this work, a fault is then identified ($F_k = 1$) if either $T_k^2 > T_{UCL}^2(\alpha)$ or $Q_k > Q_{UCL}(\alpha)$ conditions are satisfied, i.e.

$$F_k = \left(T_k^2 > T_{UCL}^2(\alpha)\right) \oplus \left(Q_k > Q_{UCL}(\alpha)\right) \tag{26}$$

where $\oplus$ represents a logical OR operation. By using the fault detection condition (26), the monitoring performance becomes insensitive to the number of states, n

since any ignored variances in the T^2 metric by reducing n will be recovered by Q metric.

4. CVA with KDE Algorithm

By summarising the analysis presented in the previous sections, a new extension of CVA using KDEs for nonlinear dynamic process monitoring is proposed to identify underlying faults subject to non-Gaussian processes. The step by step procedure of the proposed CVA with KDE algorithm is illustrated in the flowchart presented in Figure 1.

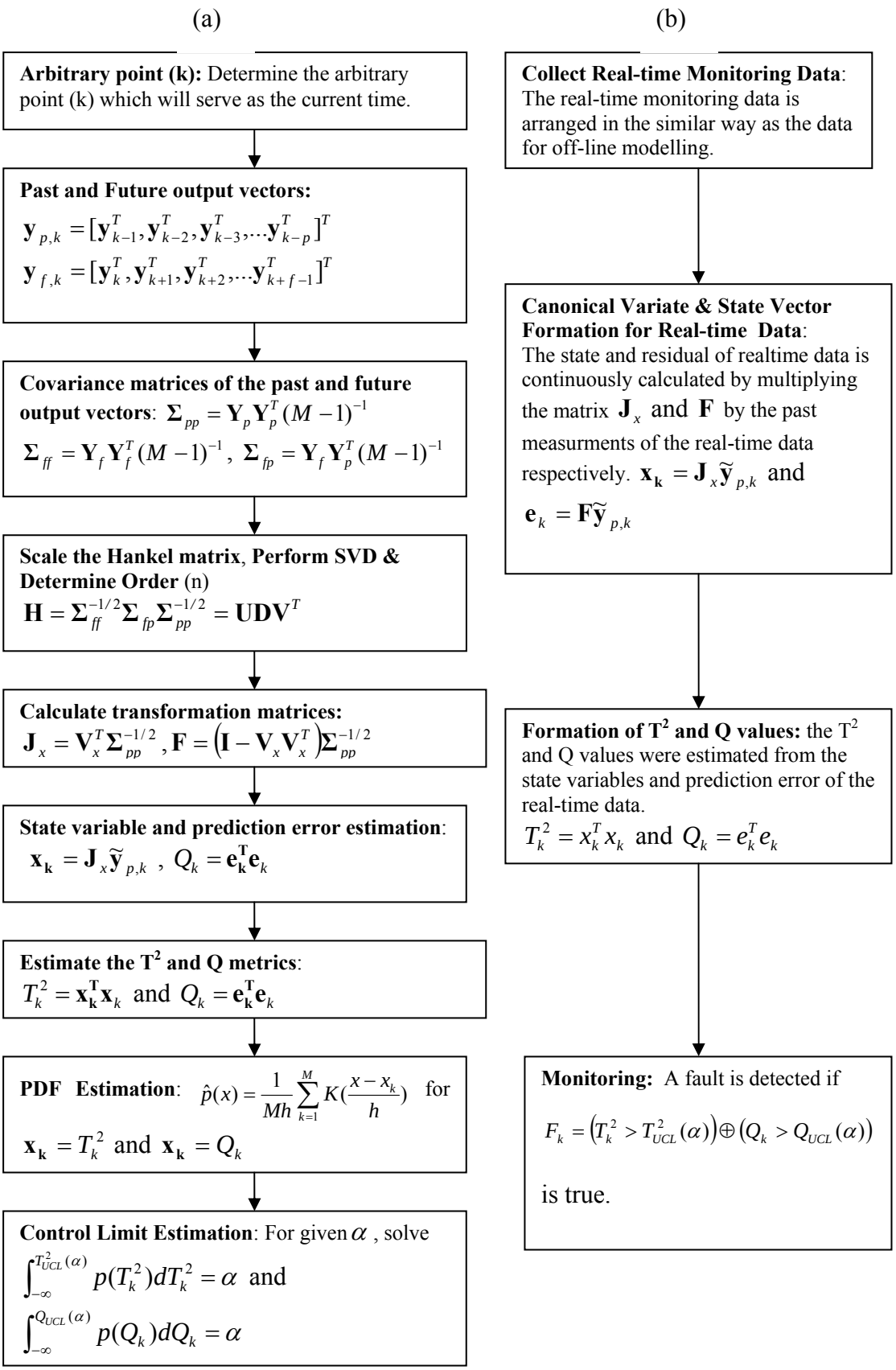

Figure 1: Flowchart of the CVA with KDE algorithm, (a) Off-line modelling procedure, (b) Real-time monitoring procedure

5. Case Study - Tennessee Eastman Process Plant

The Tennessee Eastman Process (TEP) plant[18] has 5 main units which are the reactor, condenser, separator, stripper and compressor[5,18]. Streams of the plant consists of 8 components; A, B, C, D, E, F, G and H. Components A, B and C are gaseous reactants which were fed to the reactor to form products G and H. The TEP data used for this work consists of two blocks; the training and test data blocks. Each block has 21 data sets corresponding to the normal operation (Fault 0) and 20 fault operations (Fault 1 – Fault 20). The sampling time for most of the process variables in the TEP plant is 3 minutes. A total of 52 measurements are collected for each data set of length, N=960 representing 48-hour operation with a sampling rate of 3 minutes. However, 19 of the 52 measurements, 14 of them sampled at 6 minute interval and 5 of them sampled in every 15 minutes, have not been included in this study due to the measurement time delay. Different from the work reported by Chiang[15], 11 manipulated variables are treated the same as other measured variables because under feedback control, these variables are not independent any more. The simulation time of each operation run in the test data block is 48 hours and the various faults are introduced only after 8 hours. This means that for each of the faults, the process is in-control for the first 8 simulation hours before the process gets out of control at the introduction of the fault. All twenty faults have been studied in this work. Also in this paper, the normal operating process data will be referred to as the training data. A graphical description of the TEP Plant is shown in Figure 2 while a brief description of the twenty TEP Faults is presented in Table 1.

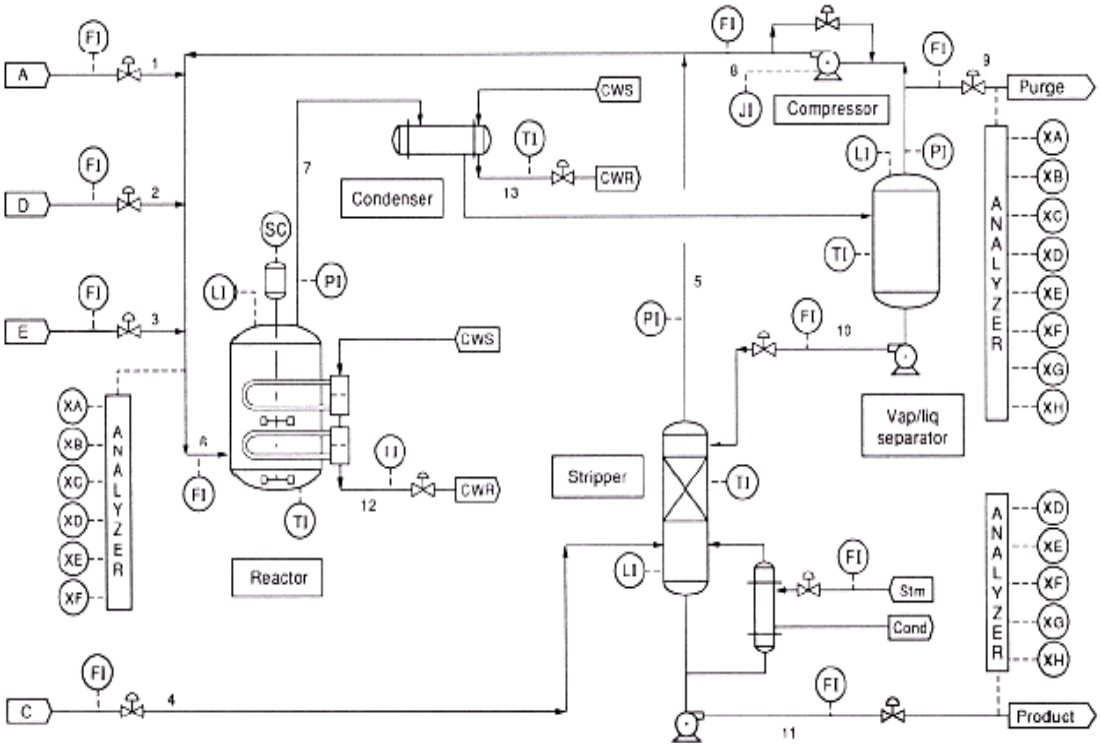

Figure 2 Graphical Description of the TEP Plant

Table 1. Brief Description of TEP Plant Faults

Fault ID	Description	Type
1	A/C Feed Ratio, B Composition Constant (Stream 4)	Step
2	An increase in B while A/C Feed ratio is constant (stream 4)	Step
3	D Feed Temperature (Stream 2)	Step
4	Reactor Cooling Water Inlet Temperature	Step
5	Condenser Cooling Water Inlet Temperature	Step
6	A loss in Feed A (stream 1)	Step
7	C Header Pressure Loss – Reader Availability (Stream 4)	Step
8	A,B,C Feed Composition (Stream 4)	Random
9	D Feed Temperature (Stream 2)	Random
10	C Feed Temperature (Stream 4)	Random
11	Reactor Cooling Water Inlet Temperature	Random
12	Condenser Cooling Water Inlet Temperature	Random
13	Reaction Kinetics	Drift
14	Reaction Cooling Water Valve	Sticking
15	Condenser Cooling Water Valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown

6. Monitoring Performance

The monitoring performance in this study is assessed based on the percentage reliability which is defined as the percentage of the samples outside the control limits[19] within the last 40 hour faulty operation. Hence a monitoring technique is said to be better than another technique if the percentage reliability of this technique is numerically higher than the percentage reliability of another. Also, the monitoring performance is assessed by the detection delay which is the time period it takes to detect a fault after the introduction of the fault. The false alarm rate was also investigated. The monitoring performance of the proposed CVA with KDE is compared with the performance of the DPCA and DPLS with and without KDE as well as CVA without KDE using all twenty faults described above. The 99% confidence interval is adopted in this study.

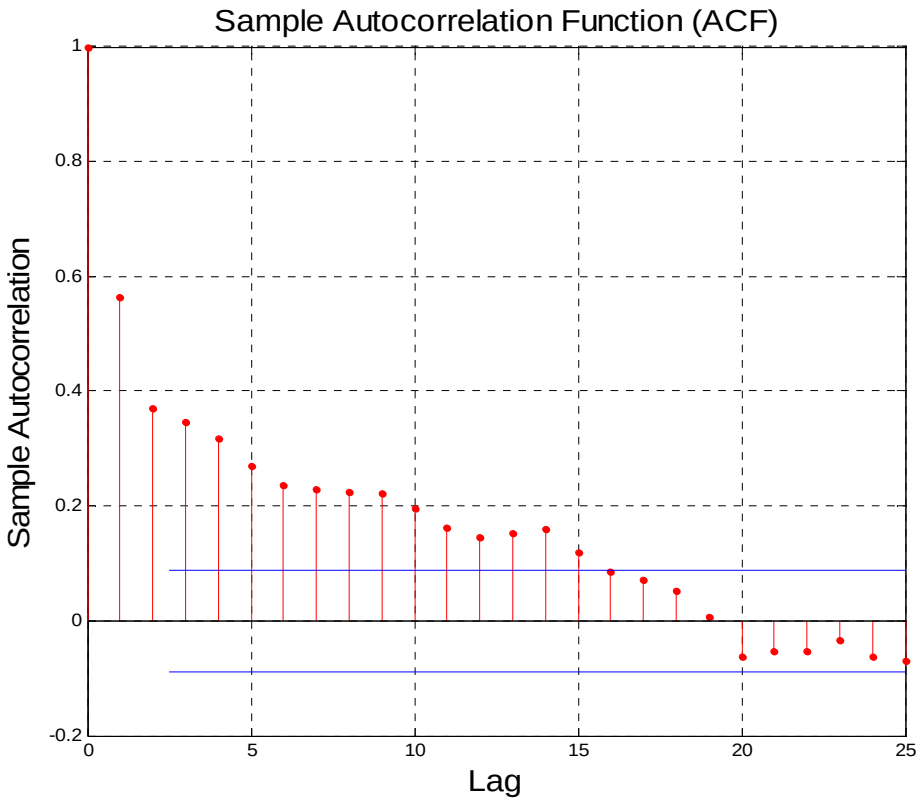

Figure 3. Autocorrelation function of the summed squares of all measurements.

The variability of the training data is characterised by the extracted canonical variate state space model. Firstly, the number of time lags for past and future observations is determined from the autocorrelation function of the summed squares of all measurements as shown in Figure 3 against ±5% confidence bounds. The autocorrelation function indicates that the maximum number of significant lags in this study is 16. Hence both p and f are set to 16. The length of the past and future observations (mq) is 528 according to (3) and (4). The number of columns of the truncated Hankel matrices according to (10) is $M = 929$. The singular value decomposition is then performed on the scaled Hankel matrix as in (13).

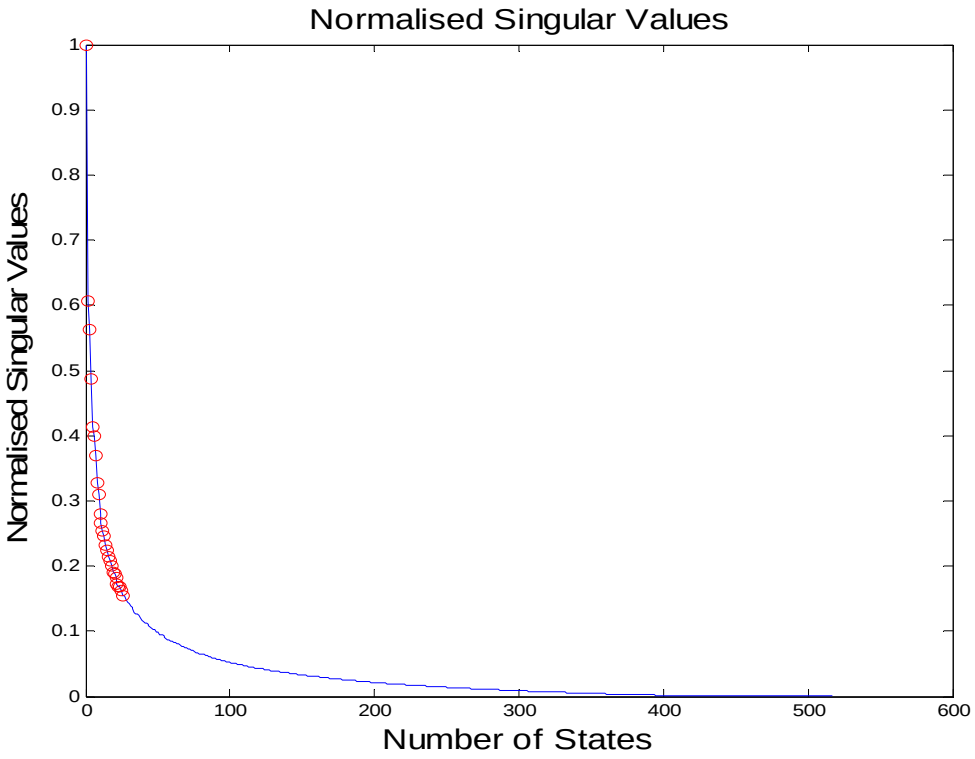

Figure 4 Normalised Singular Values from the Scaled Hankel Matrix

Several ways have been suggested to determine the order (n) of the system for CVA based approaches amongst which the dominant singular values[3,5] and the Akaike

Information Criterion (AIC)[6] are most widely adopted. The former method was adopted in this study to determine the order of the system. The singular values from the scaled Hankel were normalised to have the values ranging between 0 and 1 and then the order determined based on the dominant normalised singular values. For the TEP case study, it is noticed that the singular values of the scaled Hankel matrix H in (13) decrease slowly. If n is determined from these singular values, it will be unrealistically large as indicated in Figure 4 (a), which shows the normalised sum of squares of residual singular values against the number of states. As mentioned in section 3, the value of n is not important to monitoring performance for this work due to the fault detection condition (26) adopted. Hence, a more realistic number of singular values, n = 26 represented by circles in Figure 4 are employed to represent the model space. Also, to make a fair comparison of the proposed technique with the other techniques considered, the process variables, the number of lag and the order to determine the dimension of the latent variables are the same for all the approaches compared. The monitoring criterion mentioned above is applied to all the other methods considered.

6.1 Reliability Comparison

The superiority of the CVA with KDE over other techniques considered in this paper is demonstrated in Table 2. Over all the faults compared, the CVA achieves the best performance in terms of reliability. Both CVA techniques are able to improve the monitoring performance for most TEP faults comparing with the DPCA, DPCA with KDE, DPLS and DPLS with KDE techniques. Nevertheless, the proposed CVA with KDE technique is able to further improve the reliability for faults that are more difficult to detect such as Faults 3 and 9. Faults 3 and 9 are more difficult to detect

because these faults have very little effect on the corresponding process measurements. For such faults, the performance of the CVA with KDE is significantly better than that of the CVA. All KDE approaches achieve the reliability higher than or the same as their non-KDE counterparts as indicated in Table 2. This is due to the nonlinear and non-Gaussian features of the plant, which justify the necessity of this work.

Table 2. Performance based on % Reliability of all Algorithms (99%)

Faults	CVA+KDE (%)	CVA (%)	DPCA+KDE (%)	DPCA (%)	DPLS+KDE (%)	DPLS (%)
1	99.75	99.75	99.38	99.25	99.25	99.25
2	99.5	98.5	98	97.88	98.13	98
3	73.03	37.2	0	0	0.2497	0
4	99.88	99.88	99.88	99.88	99.88	99.88
5	99.88	99.88	29.09	27.84	28.21	26.47
6	99.88	99.88	99.88	99.88	99.88	99.88
7	99.88	99.88	99.88	99.88	99.88	99.88
8	98.88	98.75	97.25	97.13	97	97
9	92.26	75.28	0.2497	0	0.2497	0
10	96.63	96.25	39.08	28.21	36.83	29.46
11	99.38	99.38	98.88	98.63	97.88	97.75
12	99.5	99.5	98.13	98.13	98	98
13	96.13	96.13	95.01	95.01	94.76	94.76
14	99.88	99.75	99.75	99.75	99.75	99.75
15	99.5	99.5	0.1248	0	0.1248	0
16	99.13	99.13	35.83	26.22	26.97	21.6
17	98.13	98.13	97.75	97.75	97.75	97.75
18	99.25	99.25	98.63	98.5	98.63	98.5
19	99.88	99.88	90.51	87.02	84.64	79.28
20	97.63	97.25	79.15	76.9	73.91	71.41

6.2 Detection Delay Comparison

The detection delays for the CVA with KDE and other techniques considered are presented in Table 3. As shown in Table 3, the CVA with KDE approach is able to detect most of these faults earlier than other techniques. This means operators have

more time to take safety measures to counteract occurring faults if the proposed CVA with KDE approach is adopted. Again, all KDE associated approach achieve detection delay less than or the same as their non-KDE counterparts due to the same reason aforementioned.

Also investigated is the false alarm rates for all the faults and no false alarm has been observed for all faults and all approaches studied.

Table 3. Detection Delay for all the Algorithms

Faults	CVA +KDE	CVA	DPCA+KDE	DPCA	DPLS+KDE	DPLS
1	9	9	18	21	21	21
2	15	15	51	54	48	51
3	15	39	-	-	1125	-
4	6	6	6	6	6	6
5	6	6	12	12	12	12
6	6	6	6	6	6	6
7	6	6	6	6	6	6
8	30	33	69	72	75	75
9	33	45	2115	-	1125	-
10	84	93	210	210	219	219
11	18	18	24	24	24	24
12	15	15	48	48	51	51
13	96	96	123	123	129	129
14	6	9	9	9	9	9
15	15	15	1140	-	1125	-
16	24	24	111	111	216	219
17	48	48	57	57	57	57
18	21	21	36	39	36	39
19	6	6	36	39	36	42
20	60	69	120	123	123	123

6.3 Monitoring Chart Comparison of Fault 9

To appreciate the superior performance achieved by the new CVA with KDE approach, the T^2 and Q monitoring charts of all approaches for Fault 9 are presented in Figure 5. In Figure 5, sub-figures in the left column and the right column are for the T^2 and Q charts respectively; whilst the first, second and third rows are for CVA,

DPCA and DPLS approaches respectively. Upper control limits obtained based on the Gaussian assumption are represented as dashed lines, whilst the UCLs determined by the KDE approach are shown in dash-dot lines.

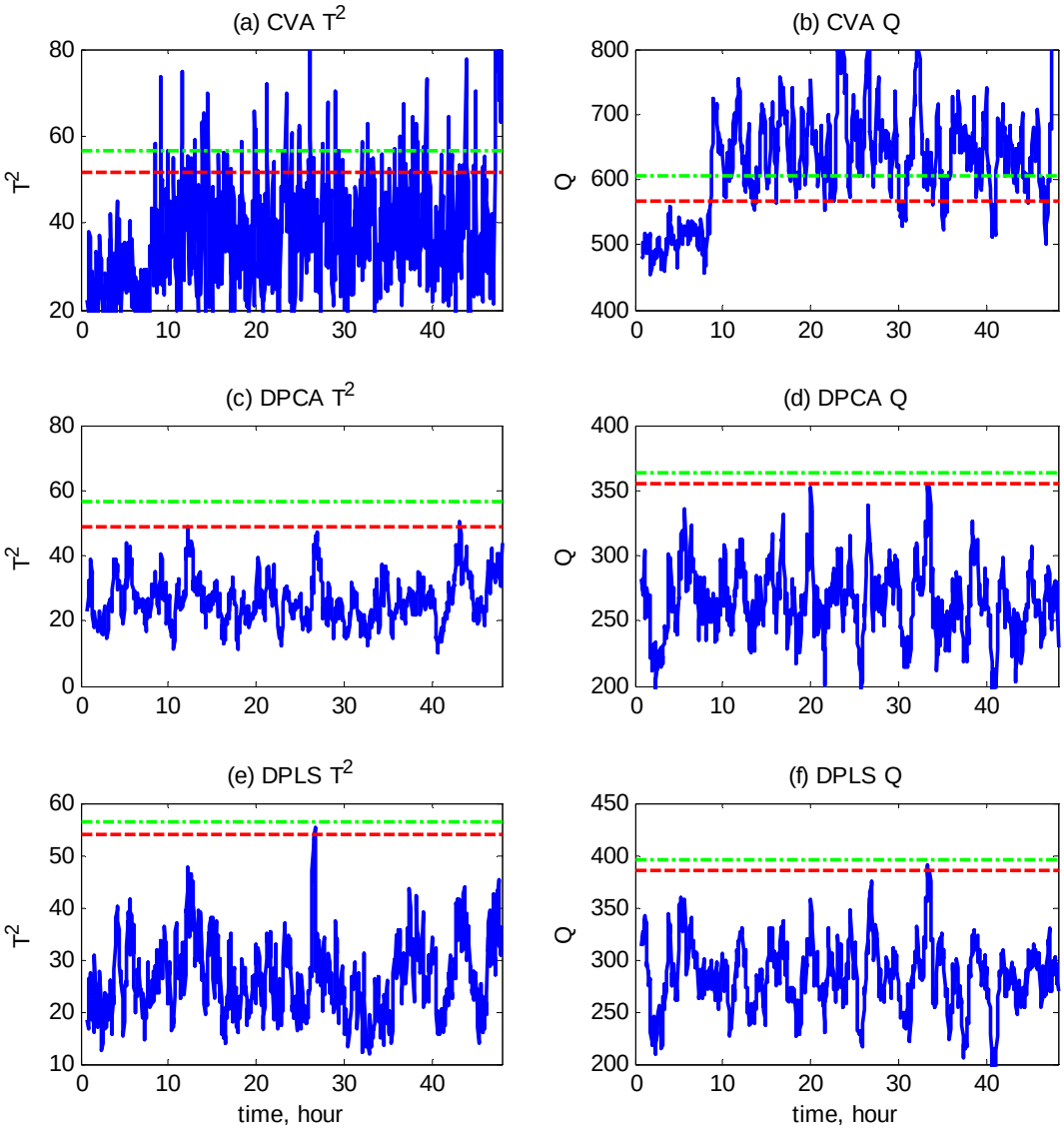

Figure 5. Fault 9 monitoring chats, solid: metrics, dashed: KDE based UCL, dash-dot: Gaussian assumption based UCL.

Figure 5 clearly indicates that only the CVA model is able to reveal the difference in dynamic behaviour between the normal operation and the operation with fault 9. Both

T2 and Q metrics produced by the DPCA and the DPLS approaches have no identifiable difference between the normal and faulty operations. Furthermore, the CVA with KDE approach gives tighter UCLs for both metrics resulting in a higher percentage of reliability and earlier fault detection than the traditional CVA approach.

7 Conclusions

To deal with fault monitoring for nonlinear dynamic processes, the linear state-space model based CVA approach is extended by directly estimating the underlying PDF of the associated T^2 and Q metrics to derive more appropriate control limits for these monitoring metrics. This leads to the new CVA with KDE algorithm proposed for nonlinear dynamic process monitoring. The proposed approach is applied to the Tennessee Eastman Process. The monitoring performance of the proposed CVA with KDE is compared with that of the DPCA and DPLS with and without KDE as well as CVA without KDE techniques. The percentage reliability and the detection delays were adopted to assess and compare the monitoring performance of the proposed approach with that of all other techniques considered in this study. Although some of the faults are commonly detected by all the techniques considered, the outstanding superiority of the CVA with KDE is demonstrated in those faults that are not easily detectable. For such faults, the proposed CVA with KDE has higher percentage reliability than other techniques considered. In addition, the proposed CVA with KDE is able to detect faults earlier than other techniques considered. Hence, the CVA with KDE is a more efficient tool than the DPCA and the DPLS with and without KDE as well as the CVA without KDE for nonlinear dynamic process monitoring.

References

1. W. F. Ku, H. R. Storer and C. Georgakis. Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems,* 1995, pp 179 - 196.
2. T. J. Richard, K. Uwe and E. C. Jonathan. Dynamic Multivariate Statistical Process Control Using Subspace Identification. *Journal of Process Control* Vol. 14, 2004, pp 279 - 292.
3. A. Negiz and A. Cinar. Monitoring of Multivariable Dynamic Processes and Sensor Auditing. *Journal of Process Control,* Vol. 8, No. 56, 1998, pp 375 - 380.
4. T. Komulainen, M. Sourander and S. Jamsa-Jounela. An Online Application of Dynamic PLS to a Dearomatization Process. *Computers and Chemical Engineering*, Vol. 28, 2004, pp 2611 - 2619.
5. L. H Chiang, E. L. Russell and R. D. Braatz. *Fault Detection and Diagnosis in Industrial Systems*. Springer, London, 2001, pp 85-98, 103-109.
6. L. Juan and L. Fei. Statistical Modelling of Dynamic Multivariate Process Using Canonical Variate Analysis. *Proceedings of IEEE International Conference of Information and Automation*, Colombo, Sri Lanka, 15 -17, December 2006, pp 218 - 221.
7. A. Norvalis, A. Negiz, J. DeCicco and A. Cinar. Intelligent Process Monitoring by Interfacing Knowledge-Based systems and Multivariate Statistical Monitoring. *Journal of Process Control,* Vol. 10, 2000, pp 341 - 350.

8. C. D. Schaper, W. E. Larimore, D. E. Seborg and D. A. Mellichamp. Identification of Chemical Processes Using Canonical Variate Analysis. *Computers Chemical Engineering,* Vol. 18, No. 1, 1994, pp 55 - 69.

9. A. Negiz and A. Cinar. PLS, Balanced and Canonical Variate Realization Techniques for Identifying VARMA models in State Space. *Chemometrics and Intelligent Laboratory Systems,* Vol. 38, 1997, pp 209 - 221.

10. A. Chiuso and G. Picci. Asymptotic Variance of Subspace Estimates. *Proceedings of the 48th IEEE Conference on Decision and Control*, Vol. 4, Dec 2001, Orlando, Florida USA, pp 3910 - 3915.

11. N. F. Hunter Jr. Comparing CVA and ERA in Transfer Function Measurement for Lithography Applications. *Proceedings of the American Control Conference*, Vol. 2, 2-4 June 1999, San Diego, California, USA pp 1171 - 1175.

12. W .E Larimore. Statistical Optimality and Canonical Variate Analysis System Identification. *Signal Processing*, Vol. 52, 1996, pp 131 - 144.

13. E. B. Martin and A. J. Morris. Non-Parametric Confidence Bounds for Process Performance Monitoring Charts. *Journal of Process Control*, Vol. 6, No. 6, 1996, pp 349 - 358.

14. Q. Chen, P. Goulding, D. Sandoz and R. Wynne. Application of Kernel Density Estimates to Condition Monitoring for Process Industries. *Proceedings of the American Control Conference*, Vol. 6, 21-26 June 1998, Philadelphia, PA, USA, pp 3312 - 3316.

15. J. E. Jackson and G. S. Modholkar. Control Procedures for Residuals Associated with Principal Component Analysis. *Technometric*, Vol. 21, 1979, pp 341 - 349.

16. A. W. Bowman and A. Azzalini. Applied Smoothing Techniques for Data Analysis, The Kernel Approach with S-Plu Illustrations, Clarendon Press, Oxford, 1997.

17. S. Xiaoping and A. Sonali. Kernel Density Estimation For an Anomaly Based Intrusion Detection System. *Proceedings of the 2006 World Congress in Computer Science, Computer Engineering and Applied Computing*, Las Vegas Nevada 26 – 29 June, 2006, pp 161 - 167.

18. J. J. Downs and E. F. Vogel. A Plant-wide Industrial Process Control Problem. *Computers and Chemical Engineering*, Vol. 17, No. 3, 1993, pp 245 - 255.

19. M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss and B. R. Bakshi. Comparison of Multivariate Statistical Process Monitoring Methods with Applications to the Eastman Challenge Problem. *Computers and Chemical Engineering,* Vol. 26, 2002, pp 161 - 174.

20. A. Simoglou, E. B. Martin and A. J. Morris. Statistical Performance Monitoring of Dynamic Multivariate Processes Using State Space Modelling. *Computers and Chemical Engineering,* Vol. 26, 2002, pp 909 - 920.

21. W. E. Larimore. System Identification Reduced Order Filtering and Modelling via Canonical Correlation Analysis. Proceedings of the American Control Conference, San Francisco, USA, 1983, pp 445 - 451.

22. B. C. Juricek, D. E. Seborg and W. E. Larimore .W .E. Fault Detection Using Canonical Variate Analysis. *Ind. Eng. Chem. Res,* Vol. 43, 2004, pp 458 - 474.

23. T. Chen**,** J. Morris and E. Martin. Probability Density Estimation

via an Infinite Gaussian Mixture Model: Application to Statistical Process Monitoring. *Appl. Statist,* Vol. 55, Part 5, 2006, pp. 699 - 715.

24. A. Simoglou, E. B. Martin and A. J. Morris. A Comparison of Canonical Variate Analysis and Partial Least Squares for the Identification of Dynamic Processes. *Proceedings of the American Control Conference*, 2-4 June, 1999, San Diego, California, USA, pp 832 - 837.

25. E. Tatara and A. Cinar. An Intelligent System for Multivariate Statistical Process Monitoring and Diagnosis. *Instrumentation, Systems and Automation Society,* Vol. 41, 2002, pp 255 - 270.

26. A. C. Rencher. *Methods of Multivariate Analysis*. A Wiley-Interscience Publication. John Wiley and Sons, INC. New York, 1934, pp 132 - 146.

27. M. Kano, K. Nagao, S. Hasebe, I. Hashimoto, H. Ohno, R. Strauss and B. R. Bakshi. Comparison of Multivariate Statistical Process Monitoring Methods with Applications to the Eastman Challenge Problem. *Computers and Chemical Engineering,* Vol. 26, 2002, pp 161 - 174.