FAULT DETECTION USING CORRESPONDENCE ANALYSIS: APPLICATION TO TENNESSEE EASTMAN CHALLENGE PROBLEM

Detroja K. P.[1], Gudi R. D.[2*], Patwardhan S. C.[2]

[1] Interdisciplinary Programme on Systems and Control Engineering,
[2] Department of Chemical Engineering
Indian Institute of Technology Bombay, Powai, Mumbai – 400076, India.
* Email: ravigudi@che.iitb.ac.in

Abstract: This paper presents an approach based on the use of correspondence analysis (CA) for the task of fault detection and diagnosis. Unlike other tools (PCA / DPCA) that are used for this latter task, CA is shown to use a different metric to represent the information content in the data matrix **X**. Decomposition of the information represented in the metric is shown to yield superior performance from the viewpoints of data compression, discrimination and classification as well as early detection of faults. We demonstrate these performance improvements over PCA and DPCA on the Tennessee Eastman problem, which is a representative benchmark problem used in the literature. CA is shown to yield vastly superior performance for the monitoring of the TE problem, when compared with PCA and DPCA.

Keywords: Correspondence Analysis (CA), Principal Components Analysis (PCA), Dynamic PCA (DPCA), Fault Detection (FD)

1. INTRODUCTION

Early detection of the occurrence of an abnormal event in an operating plant is very important for plant safety and maintaining product quality. Tremendous advancements in the area of advanced instrumentation have made it possible to measure hundreds of variables every few seconds. These measurements bring in useful signatures about the status of the plant operation. A wide variety of techniques, for detecting faults, have been proposed in the literature. These techniques can be broadly classified into model based methods and historical data based methods. While model based methods can be used to detect and isolate signals indicating abnormal operation, such quantitative (or qualitative) cause-effect models may be difficult to develop from the first principles.

Historical data based methods for fault detection attempt to extract maximum information out of the archived data and require minimum physical knowledge of the plant. Due to the high dimensionality and correlation amongst the variables of the plant data, multivariate statistical tools, which take correlation amongst variables into account, are better suited for this task. Dimensionality reduction is also a very important aspect of historical data based methods.

Generally, the information content in a data matrix $\mathbf{X}$ can be quantified in terms of a number of criteria or metrics. The most commonly used metric, the variance or the multivariate analysis of the variance (MANOVA), usually yields a wealth of knowledge from the information embedded in the matrix $\mathbf{X}$. Multivariate statistical tools, such as PCA, are based on decomposition of the variances and address issues related to correlation along the column <u>or</u> the row spaces. PCA determines the lower dimensional representation of the data, in terms of capturing the data directions that have the most variance. This is done via singular value decomposition (SVD) of a suitably scaled (mean centered and variance scaled) data matrix ($\mathbf{X}$) and retaining those principal components that have significant singular values. PCA achieves dimensionality reduction in the column space by considering the correlation amongst the variables. The statistical model thus built, characterizes the normal plant operation. PCA has been used for fault detection using statistical control limits Q (Squared Prediction Error) and/ or T^2 statistics (Nomikos and MacGregor, 1995). Once a fault is detected using either Q or T^2 statistics, contribution plots (Miller *et al.*, 1998) have been used to help fault isolation. One of the drawbacks of PCA, however, is that it is representation-oriented and not discrimination-oriented. As shown in Chiang *et al.* (2000), there are other algorithms such as multiple discriminant analysis that can better discriminate between the normal and abnormal operating regions in the data and hence yield smaller misclassification rates during on-line monitoring.

An important aspect that also needs to be considered is that the variance need not be the best metric for capturing cause and effect relationships. Usually, such cause and effect relationships are dynamic and can be more effectively analyzed by assessing the row (sample) versus column (variable) associations. In PCA or in the multiple discriminant analysis (MDA) approach, such dynamic relationships require expanding of the column space to generate a static map of the dynamic relationships. This latter strategy has drawbacks in terms of larger matrix and data sizes and increasing computational intensity.

This paper proposes to address the above problems using an approach that is based on CA for the task of FDD. Correspondence Analysis (CA) (Greenacre, 1984; Greenacre, 1993; Hardle and Simar, 2003) is a powerful multivariate statistical tool, which is based on generalized SVD (GSVD). CA is a dual analysis, as it simultaneously analyzes dependencies in column, row and the joint row-column space in a dual lower dimensional space. Thus, dynamic correlation can be represented relatively easily without having to expand and deal with larger data sizes. CA primarily uses a measure of the row-column association and decomposes it to obtain directions in the lower dimension space which discriminate as well as compress information. Unlike its earlier counterparts such as PCA and MDA, it represents the cause-effect relationships in terms of a chi-square (χ^2) value, that measures row-column associations. Since, decomposition of the χ^2 value takes joint row-column association into account; it can be expected to perform better than conventionally used variance decomposition based methods, such as Principal Components Analysis (PCA).

In this paper, we show how Correspondence Analysis (CA) is superior to PCA and MDA and can be used for the purpose of fault detection and have also defined statistics based on CA that can be used for online process monitoring. It has also been found that the performance of statistics based on CA is better as compared to conventional PCA. The dimensionality reduction achieved using CA is more effective as it takes joint row-column association into account. Also, due the special kind of scaling it employs, CA is also shown to be able to cluster and aggregate the data more effectively (Ding *et al.*, 2002).

The objective of this paper is (i) to demonstrate the usefulness of CA for fault detection, (ii) to define new statistics which are equivalent to Q and T^2 statistics for PCA and (iii) to evaluate and compare performance of PCA, DPCA and CA for detecting faults in a realistic chemical process simulation. We show here that the proposed statistic performs better than the existing PCA and DPCA statistics when applied to the Tennessee Eastman process. The paper is organized as follows. First, PCA and DPCA are briefly presented. Then, CA is described followed by the proposed approach to fault detection using CA based statistics. Finally, PCA, DPCA and CA are applied to the data collected from the Tennessee Eastman process simulator. We conclude with comparative study of results.

2. Principal Components Analysis (PCA)

Any matrix $\mathbf{X}_{m\times n}$ consisting of *m*-observations and *n*-variables, collected from an operating plant has a wealth of information regarding the health of the plant. PCA decomposes the variance in the data, based on dependencies along the columns, to achieve dimensionality reduction. PCA computes a set of new orthogonal principal directions, called loading vectors. Loading vectors are obtained by solving an optimization problem involving maximization of variance explained in the data matrix by each direction. For example, the first direction is obtained as a solution of the optimization problem in the space of the first linear combination $\mathbf{t}_1 = \mathbf{X}\mathbf{p}_1$ as,

$$\max_{\mathbf{p}_1}\left(\mathbf{t}_1^T\mathbf{t}_1\right) = \mathbf{p}_1^T\mathbf{X}^T\mathbf{X}\mathbf{p}_1 \qquad (1)$$

Such that $\mathbf{p}_1^T\mathbf{p}_1 = 1$.

It has been shown that the singular vector corresponding to the largest singular value provided by the SVD of $\mathbf{X}$, is the solution to the above optimization problem. Because of correlation amongst variables, only first k (substantially smaller than n) loading vectors may explain most of the variance in the data. Thus, PCA decomposes the matrix $\mathbf{X}$ as,

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \qquad (2)$$

where, $\mathbf{P}$ contains only first k ($k << n$) loading vectors. The matrix $\mathbf{T}$ is called the scores matrix. The matrix $\mathbf{E}$ contains the component of variance of matrix $\mathbf{X}$, such as noise, which can not be explained by $\mathbf{T}\mathbf{P}^T$, and is also known as residual matrix.

2.1. *Fault detection using PCA*

The statistical model developed using PCA, from the normal operating data, can be used for the purpose of online monitoring and fault detection. When employed online, new scores are obtained by projecting the new measurements onto the loading vectors. Normal operation of the plant can be characterized by Hotelling's T^2 statistic (Equation (3)), based on the first *k* loading vectors (principal components) retained. The status of the plant is considered normal if the value of T^2 static stays within its control limit.

$$\mathrm{T}^2 = \mathbf{x}^T \mathbf{P} \mathbf{\Lambda}^{-1} \mathbf{P}^T \mathbf{x} \quad (3)$$

where, $\mathbf{x}$ is the new measurement vector and $\mathbf{\Lambda}$ is a diagonal matrix containing first k eigen values of the covariance matrix of $\mathbf{X}$.
The control limit (threshold) for the T^2 statistic T^2_α can be calculated from Equation (4) (Ku et al., 1995). A value of T^2 statistic greater than the control limit (T^2_α) indicates occurrence of a fault.

$$\mathrm{T}^2_\alpha = \frac{(m-1)k}{(m-k)} F_\alpha(k, m-k) \quad (4)$$

where, $F_\alpha(k, m-k)$ is the upper $100\alpha\%$ critical point of F-distribution with k and $m-k$ degrees of freedom.

However, monitoring only T^2 statistic is not sufficient, as it only detects variation in the direction of the first k PCs. Variation in the space corresponding to $(n-k)$ PCs (having smallest associated singular values) can also be monitored using Q statistic (Jackson and Mudholkar, 1979). The value of Q statistic and its control limit can be calculated as follows:

$$Q = \left[\left(\mathbf{I} - \mathbf{P}\mathbf{P}^T\right)\mathbf{x}\right]^T \left[\left(\mathbf{I} - \mathbf{P}\mathbf{P}^T\right)\mathbf{x}\right] \quad (5)$$

$$Q_\alpha = \theta_1 \left[\frac{h_0 c_\alpha \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{(1/h_0)} \quad (6)$$

where, $\theta_i = \sum_{j=k+1}^{n} (\mu_j)^{2i}$, $h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$, c_α is the normal deviate corresponding to $(1-\alpha)$ percentile and μ_j is j^{th} singular value. When a fault occurs that results in change in covariance structure of the normal operating data, it gets reflected by a high Q value.

2.2. *Dynamic PCA*

Monitoring using PCA statistics implicitly assumes that the measurements at one time instant are statistically independent to the measurements at the past time instances. The assumption is generally not valid for most processes due to dynamics of the plant. The PCA method can be extended to take into account the serial correlations, by augmenting each observation vector with a few past observations and stacking the data in a bigger matrix.

$$\mathbf{X}_A = \left[\mathbf{X}(t)\mathbf{X}(t-1)\cdots\mathbf{X}(t-l)\right] \quad (7)$$

By performing PCA on the augmented data matrix ($\mathbf{X}_A$), a multivariate auto regressive (AR) model is extracted directly from the data (Ku *et al.*, 1995). This however, requires working with considerably larger data matrices than the conventional PCA. The T^2 and Q statistics and their control limits can be generalized directly to DPCA.

3. CORRESPONDENCE ANALYSIS

The aim of correspondence analysis is to develop simple indices to highlight associations between the rows and the columns. Unlike PCA, which canonically decomposes the total variance in the matrix $\mathbf{X}$, CA decomposes a measure of row-column association, typically formulated as the total χ^2 value, to capture the dependencies. CA can be presented in terms of weighted Euclidean space as follows. In general, through an optimization procedure, we seek a lower dimension (say k) approximation of the matrix $\mathbf{X}$ in an appropriate space $\boldsymbol{S}$. In terms of the row and column points, each row of $\mathbf{X}$ can be represented as a point $\mathbf{x}_i$ $(i = 1, 2..m)$ in an n-dimensional space. When one seeks to estimate the lower dimensional space (approximation) $\boldsymbol{S}$ that is closest to this cloud of row points, one could solve optimization problems that are formulated in several possible ways. One such optimization problem to determine the space $\boldsymbol{S}$ could then be minimize a weighted Euclidean distance defined as,

$$d^2 = (\mathbf{x} - \overline{\mathbf{x}})^T \mathbf{D} (\mathbf{x} - \overline{\mathbf{x}}) \quad (8)$$

It can be shown (Greenacre, 1984) that the solution to the problem of minimizing the weighted distances in Equation (8) can be given by decomposition of inertia of row (or column) cloud, i.e. generalized SVD of the matrix $\left[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T\right]$. The vectors $\mathbf{r}$ and $\mathbf{c}$ are the vectors of row sums and column sums of $\left[(1/g)\mathbf{X}\right]$, respectively (Equation (9) & (10)).

$$\mathbf{r} = \left[(1/g)\mathbf{X}\right]\mathbf{1} \quad (9)$$

$$\mathbf{c} = \left[(1/g)\mathbf{X}\right]^T \mathbf{1} \quad (10)$$

where, $\mathbf{1}$ is a vector of all 1's of appropriate dimension. The matrix $\mathbf{D}_r$ is then defined as $\mathbf{D}_r = diag(\mathbf{r})$ and similarly, $\mathbf{D}_c = diag(\mathbf{c})$.

The inertia of the row cloud and the column cloud can be shown to be the same (Greenacre, 1984) and is given by the χ^2 value divided by g. The weight matrix $\mathbf{D}$ is chosen as diagonal matrix of the row sums ($\mathbf{D}_r$) or the column sums ($\mathbf{D}_c$). The generalized SVD of this matrix is defined as

$$\left[(1/g)\mathbf{X} - \mathbf{r}\mathbf{c}^T\right] = \mathbf{A}\mathbf{D}_\mu\mathbf{B}^T \quad (11)$$

such that, $\mathbf{A}^T\mathbf{D}_r^{-1}\mathbf{A} = \mathbf{I}_{m\times m}$ and $\mathbf{B}^T\mathbf{D}_c^{-1}\mathbf{B} = \mathbf{I}_{n\times n}$.

The generalized SVD results of Equation (11) can also be realized via the SVD of an appropriately scaled matrix $\mathbf{X}$, as explained below. We define the matrix $\mathbf{P}$ as,

$$\mathbf{P} = \mathbf{D}_r^{-1/2}\left[\left(1/g\right)\mathbf{X} - \mathbf{rc}^T\right]\mathbf{D}_c^{-1/2} \tag{12}$$

Then, the regular SVD of the matrix $\mathbf{P}$ gives the required singular vectors. The problem of finding principal axis for the row cloud and the column cloud are dual to each other and $\mathbf{A}$ and $\mathbf{B}$ define the principal axes for the column cloud and the row cloud respectively. In general, major part of the χ^2 value can be explained by retaining only first k ($k << m,n$) principal axes corresponding to the largest singular values. The co-ordinates (scores) of the row profile points and column profile points for the new principal axis can be computed by projection on $\mathbf{A}$ and $\mathbf{B}$ (only first k columns are retained), respectively.

$$\mathbf{F} = \mathbf{D}_r^{-1}\mathbf{A}\mathbf{D}_\mu \tag{13}$$

$$\mathbf{G} = \mathbf{D}_c^{-1}\mathbf{B}\mathbf{D}_\mu \tag{14}$$

3.1. *Singular values and inertia*

The sum of the squared singular values gives the total inertia of the cloud. The inertia explained by each principal axis can then be computed by

$$IN\left(i^{th}\ axis\right) = \frac{\mu_i^2}{\sum_{j=1}^{n}\mu_j^2} \tag{15}$$

where, μ_i is i^{th} singular value.

Similarly, cumulative inertia explained up to the i^{th} principal axis is the sum of inertias explained up to that principal axis. This gives a measure of accuracy (or quality of representation) of the lower dimensional approximation. Although several mathematical criteria do exist for selecting the number of principal axis, there is no generally fixed criterion proposed to determine how many principal axes should be retained.

4. PROCESS MONITORING USING CA

Correspondence analysis has been used to build statistical models for ecological problems, study of vegetation habit of species, social networks, etc. Here we propose to build the statistical model for the plant data pertaining to normal operation using CA. As discussed earlier, CA takes joint row-column association into account while decomposing the χ^2 value. CA has also been shown to give better aggregation and clustering (Detroja *et al.*, 2005). CA also scores over PCA, which assumes statistical independence of samples (rows), as well as DPCA, which requires augmentation of the data matrix.

Once the statistical model is built from the normal operation data, the next task is to define control limits which can be used for the purpose of online statistical process monitoring of the plant. Motivated by Q and T^2 statistics used in PCA and DPCA, we defined here similar statistics for CA.

For online process monitoring, when a new measurement arrives, it is projected onto the PCs to obtain the new row scores (co-ordinates). The new measurement vector $\mathbf{x}$ is given by

$$\mathbf{x} = \left[x_1\ x_2\ \cdots\ x_m\right]^T \tag{16}$$

The row sum of this measurement vector, r is given by

$$r = \sum_{i=1}^{m} x_i \tag{17}$$

and the new row scores can be obtained as

$$\mathbf{f} = \left[\frac{1}{r}\mathbf{x}^T\mathbf{G}\mathbf{D}_\mu^{-1}\right]^T \tag{18}$$

4.1. *T^2 statistic for CA*

Hotelling's T^2 statistic effectively captures normal operating region for the multivariate data in PCA. For the statistical models that are built using CA, a similar statistic can be used to characterize the normal plant behavior. The T^2 value for CA model is defined as in Equation (19).

$$\mathrm{T}^2 = \mathbf{f}^T\mathbf{D}_\mu^{-2}\mathbf{f} \tag{19}$$

where, $\mathbf{D}_\mu$ contains first k-largest singular values, which were retained.

Control limit for the T^2 statistic based on CA, follows from the Equation (4).

4.2. *Q statistic for CA*

As explained earlier, monitoring the plant using only T^2 statistics is not adequate for fault detection, as it only monitors the variation along the principal axes which were retained in the statistical CA model. Any significant deviation in the direction of $n-k$ PCs (corresponding to smallest singular values), is also indicative of a fault.

The value of Q statistic for CA is defined as in Equation (20).

$$Q = \left[\mathbf{Bf} - \left(\frac{1}{r} \mathbf{x} - \mathbf{c} \right) \right]^T \left[\mathbf{Bf} - \left(\frac{1}{r} \mathbf{x} - \mathbf{c} \right) \right] \quad (20)$$

The control limit for the Q statistic is chosen as 95% confidence limit from the normal operating residual values.

Correspondence analysis, along with the statistics defined here, can be very useful in fault detection. In the next section, we demonstrate the usefulness of CA for fault detection and compare the performance of statistics based on CA, PCA and DPCA.

5. APPLICATION TO TENNESSEE EASTMAN CHALLENGE PROBLEM

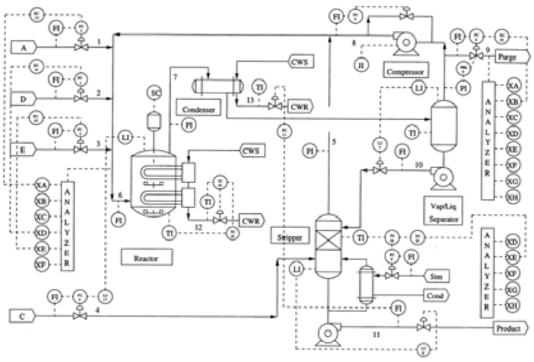

Figure 1: A diagram of the Tennessee Eastman process

The Tennessee Eastman process proposed by Downs and Vogel (1993) has been a benchmark problem for plant-wide control strategy and fault detection (Russell *et al.*, 2000). The test problem is based on an actual chemical process where only the components, kinetics and operating conditions were modified for proprietary reasons. Figure 1 shows a diagram of the process. The simulation code allows 21 pre-programmed major process disturbances as shown in Table 1. The plant-wide control structure recommended in Lyman and Georgakis (1995) and was used by Russell *et al.* (2000) for their study of fault detection using PCA and DPCA was used to generate the closed loop simulated process data for each fault.

The statistical models were built from the normal operation data consisting of 500 samples. All manipulated and measurement variables except the agitation speed of the reactor's stirrer for a total of 52 variables were used. The data was sampled every 3 minutes. Twenty-one testing sets were generated using the pre-programmed faults (IDV 1-21).

The normal operation data was used to build statistical model from PCA, DPCA and CA. Data compression is an important aspect of multivariate statistical tools. The number of PCs to be retained in PCA can be determined via several criteria such as cross validation or scree test. Earlier work (Russell et al., 2000) retained 11 PCs which explained approximately 55% of the variance in the data. When the analysis was done using CA, it was found that when 12 PCs were retained 96.65% of the total inertia was effectively captured. It should be noted here that these values of variance explained by PCA and inertia explained by CA can not be directly compared. Nevertheless, the representation given by CA would appear to be better as through modeling of the row-column associations it is better able to capture inter-relationships between variables and samples.

The objective of the fault detection technique is that it should be independent of the training set, sensitive to all the possible faults of the process, and prompt towards the detection of the fault. Since the fault alarms are inevitable, an out-of-control value of a statistic can be the result of a fault or of a false alarm. In order to decrease the rate of false alarms, a fault can be indicated only when several consecutive values of a statistic have exceeded the threshold. In this study, the fault is indicated only when six consecutive statistic values have exceeded the control limit, and the detection delay is recorded as the first time instance in which the threshold was exceeded. This was done exactly in accordance with what has been reported by Russell *et al.* (2000) so that results can be compared. The missed detection rates for faults 3, 9, and 15 were found to be fairly high, because no observable change in the mean or the variance could be detected by visually comparing the plots of each associated observation variable (Russell *et al.*, 2000). Therefore, these faults are not considered when comparing the methods.

Table 1: Process faults for the Tennessee Eastman process simulator

Fault	Description	Type
IDV(1)	A/C Feed ratio	Step
IDV(2)	B component	Step
IDV(3)	D feed temperature	Step
IDV(4)	Reactor cooling water (RCW) inlet temperature	Step
IDV(5)	Condenser cooling water (CCW) inlet temperature	Step
IDV(6)	A feed loss	Step
IDV(7)	C header pressure loss	Step
IDV(8)	A, B, C feed component	Random
IDV(9)	D feed temperature	Random
IDV(10)	C feed temperature	Random
IDV(11)	RCW inlet temperature	Random
IDV(12)	CCW inlet temperature	Random
IDV(13)	Reactor kinetics	Slow drift
IDV(14)	RCW valve	Sticking
IDV(15)	CCW valve	Sticking
IDV(16)	Unknown	
IDV(17)	Unknown	
IDV(18)	Unknown	
IDV(19)	Unknown	
IDV(20)	Unknown	
IDV(21)	The valve for Stream 4 was fixed (steady state position)	Constant position

The detection delays (in minutes) for all 18 faults (excluding fault 3, 9 and 15), are tabulated in Table 2. Statistic having minimum detection delay is shown bold faced. All faults could be detected by the statistics defined based on CA. It can also be seen that the Q and T^2 statistics based on CA performed better as compared to statistics based on PCA. In can also be seen that the CA statistics has also performed better than DPCA statistics, which is expected to perform (and have performed) better than PCA statistics. CA based Q statistic is relatively faster in detecting faults when compared with statistics generated via PCA and DPCA. An important observation needs to be made in relation to Fault 19. As seen in Table 2, this fault was not detected by any other statistic except the Q-DPCA and Q-CA. however, even here, CA is seen to detect the fault much more rapidly than the DPCA (30 v/s 246 minutes respectively). Detection delays are also seen to be reduced considerably for other fault cases as well. The false alarms were also fewer in CA when compared to PCA (results are not included due to brevity).

Table 2: Detection delays (in minutes)

Fault	PCA Q	PCA T^2	DPCA Q	DPCA T^2	CA Q	CA T^2
IDV(1)	9	21	15	18	**6**	21
IDV(2)	36	51	39	48	**24**	36
IDV(4)	9	--	**3**	453	**3**	--
IDV(5)	**3**	48	6	6	21	45
IDV(6)	**3**	30	**3**	33	**3**	**3**
IDV(7)	**3**	**3**	**3**	**3**	**3**	**3**
IDV(8)	60	69	63	69	**24**	63
IDV(10)	147	288	150	303	**75**	171
IDV(11)	33	912	21	585	**15**	567
IDV(12)	24	66	24	9	**6**	69
IDV(13)	111	147	120	135	**108**	135
IDV(14)	**3**	12	**3**	18	**3**	--
IDV(16)	591	936	588	597	**27**	84
IDV(17)	75	87	**72**	84	87	711
IDV(18)	**252**	279	**252**	279	261	303
IDV(19)	--	--	246	--	**30**	--
IDV(20)	261	261	252	267	**210**	252
IDV(21)	855	1689	858	1566	**717**	1548

6. CONCLUSION

A new approach to fault detection based on Correspondence Analysis was proposed in this paper. New statistics based on CA, which are similar to Q and T^2 statistics of PCA, were also defined. The Tennessee Eastman process simulation was used to compare the proposed approach to fault detection using CA against conventional PCA and Dynamic PCA.

The process model representation in CA is better as it takes joint row-column association into account without increasing the number of columns in the data. The simulation study also revealed that all the faults in Tennessee Eastman process could be detected. Detection delays for fault detection are significantly reduced for most of the faults when compared with PCA and DPCA statistics.

REFERENCES

Chiang L. H., Russell E. L. & Braatz R. D. (2000), Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis, *Chemometrics & Intelligent Laboratory Systems*, **50**, 243-252.

Detroja K. P., Gudi R. D. & Patwardhan S. C. (2005), Fault detection and isolation using correspondence analysis, *Proceedings of 16th IFAC World congress, Prague*, 2005.

Ding C., He X., Zha H. & Simon H. (2002), Unsupervised Learning: Self-aggregation in Scaled Principal Component Space, *6th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD 2002)*, **2431**, 112-124.

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

Greenacre M. J. (1984), *Theory and Application of Correspondence Analysis*, Academic Press Inc., London.

Greenacre M. J. (1993), *Correspondence Analysis in Practice*, Academic Press, London.

Hardle W. & Simar L. (2003), *Applied Multivariate Statistical Analysis*, Chapter 13, MD Tech, Berlin.

Jackson J. E. and Mudholkar G. S. (1979), Control procedures for residuals associated with Principal Component Analysis, *Technometrics*, **21**, 341--349.

Ku W., Storer R. H. & Georgakis C. (1995), Disturbance Detection & Isolation by dynamic principal components analysis, *Chemometrics & Intelligent Laboratory Systems*, **30**, 179-196.

Lyman P. R. and Georgakis C. (1995), Plant-wide control of the Tennessee Eastman problem, *Computers and Chemical Engineering*, **19**(3), 321-331.

Miller, P., Swanson, R. and Heckler, C. (1998), Contribution Plots: A Missing Link in Multivariate Quality Control, *Appl. Math. and Comp. Sci.*, **8**(4), 775–792.

Nomikos, P & MacGregor J.F. (1995), Multivariate SPC charts for monitoring batch processes, *Technometrics*, **37**(1), 44-59.

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