Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 14 (12): 1244-1251, 2014
ISSN 1812-5654 / DOI: 10.3923/jas.2014.1244.1251

Utilizing of Dissimilarity Scale-based PCA in Multivariate Statistical Process Monitoring Application

[1]Mohd. Yusri Mohd. Yunus and [2]Jie Zhang
[1]Process System Engineering Group, Faculty of Chemical and Natural Resources Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300, Gambang, Kuantan, Pahang, Malaysia
[2]Centre of Process Analitic and Chemometrics Technology, School of Chemical and Advanced Material, Newcastle University, Newcastle Upon Tyne, NE1 7RU, United Kingdom

Abstract: A new Multivariate Statistical Process Monitoring (MSPM) framework is proposed, in which the correlation among of the samples are determined by using dissimilarity scale structure. The typical MSPM system adopts linear-based Principal Component Analysis (cPCA) as the multivariate data compression method. Recently however, Classical Scaling-based (CMDS) technique has been proposed as an alternative for reducing the multivariate space, nonetheless, it demands new sets of monitoring schemes as well as statistics. This proposed approach still retains the conventional PCA as the main data compression technique as well as employs the original Hotelling's T^2 and SPE statistics for charting the monitoring status via Shewhart control chart. Therefore, the original conceptual applications of MSPM are greatly preserved to certain extent, without heavily focusing on new terminologies as can be experienced in the previous CMDS systems. There are twenty different cases of Tennessee Eastman Process (TEP) have been chosen for demonstration and the fault detection results of the proposed approach were comparatively analyzed to the outcomes of conventional MSPM based on two performance factors-total number of detected cases and also total number of fastest detection cases. The last two measures are determined through fault detection time. The overall outcomes show that the new technique produces almost comparable performances to the conventional MSPM based monitoring system in terms of number of cases detected, whereas, the City-block scale has been found the most efficient detection scheme among of all. More importantly, these effective monitoring outcomes can be performed based on lower number of PCS models.

Key words: Multivariate statistical process monitoring, process monitoring, fault detection, principal component analysis, dissimilarity scale

INTRODUCTION

Chemical-based industry is generally exposed with various instable conditions over the period of operation. Those instabilities, such as abnormal variation of raw material qualities, dynamical impact of disturbances, deviations of normal operating conditions as well as wear and tear of equipment or flawed readings are simply unavoidable and typically regarded as special event that significantly contributing to deterioration of product quality. Thus, it is always desirable to have a systematic mechanism which can routinely manage all of these abnormal situations automatically in a way that safe processing as well as normal operating variation is preserved that eventually meets the targeted productivity consistently. Such issues can be addressed quite effectively by the use of process monitoring system and Multivariate Statistical Process Monitoring (MSPM) can be regarded as the most practical method for handling complicated and large scale systems (Chiang *et al*., 2001). At the core of its fundamental application, there are two types of monitoring charts typically employed-Hotelling's T^2 and Squared Prediction Errors (SPE) (Raich and Cinar, 1996). The first represents conceptually the magnitude of deviation of the current sample from the centre, whereas, the second analyses the consistency of the current sample correlation according to the NOC model developed. The main task then, is to

Corresponding Author: Mohd. Yusri Mohd. Yunus, Process System Engineering Group, Faculty of Chemical and Natural Resources Engineering, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300, Gambang, Kuantan, Pahang, Malaysia
Tel: +6095492902 Fax: +6095492889

observe the progressions of both statistics on a control chart (usually Shewhart-type control chart) that constructed respectively.

MSPM is normally functioned to conduct fault detection, fault identification and fault diagnosis tasks (Qin, 2003). As the process is normally multivariate in nature, the system will typically develop a composite model that correlates all of the variables simultaneously by using a set of Normal Operating Condition (NOC) data obtained from the historical process archive. In this regard, the system normally utilizes conventional Principal Component Analysis (cPCA) as the main technique for multivariate data compression. However, cPCA is sometimes improperly used especially in modelling highly non-linear processes as a high number of Principal Components (PCS) is always involved (Dong and McAvoy, 1996). If large variables are involved, then the PCS may also be selected considerably. As a result, non-linear PCA (Zhang *et al*., 1997) which based on a combination between associative neural network and principal curve, is introduced but the computation is very demanding and it always requires a massive amount of data for creating the optimized NOC model. A number of other PCA-based extensions as well as multivariate techniques were also conducted in various studies, nevertheless, none of these approaches have been critically addressing the same issue suggested by the non-linear PCA.

Recently, Yunus (2012), have developed three main frameworks of MSPM by using the Classical Multidimensional Scaling (CMDS) approach. The motive of the works is almost similar to the non-linear PCA, by which, the primary goal is to summarize the multivariate data in such a way the number of compressed models is relatively smaller than the conventional approach. According to Borg and Groenen (1997), CMDS, in general, offers unconventional measure of variable correlation structure which specially applies dissimilarity scale (or inter-distance measure) instead of variance-covariance (or correlation matrix) association. Those CMDS studies have been perceived as an extended effort from the procedures developed by Cox (2001). Even though some improvements can be observed in terms of fault detection efficiency, those approaches employed different score projection (by way of variable scores) as opposed to the conventional PCA (by means of sample scores). This has created difficulties in understanding the real impact of different inter-distance scaling upon the fault detection performance between the two systems.

Alternatively, Cox and Cox (1994) have established the association between the eigenvectors between the sample and variable structure by means of major product moment and minor product moment expressions respectively. In other words, any measure of inter-distance scaling can be modified and converted into a set of correlation matrix-based eigenvectors that typically applied by PCA in projecting the scores. Therefore, this particular approach provides the bridge and it opens much wider comprehensive analyses between the two monitoring platforms. Thus, it is intended in this paper to (a) layout the procedures of obtaining the PCS loading factors from various types of inter-distance scaling, (b) propose a new MSPM framework that specifically employs the new scheme mentioned in (a) and lastly (c) perform the comparative analyses between the proposed against the conventional monitoring performances, particularly by means of fault detection time. In conducting task (c), there are three Performance Indicators (PI) have been Specified-false Alarm Rate (FAR), Number of Cases Detected (NCD) and also Number of Fastest Detection (NFD). The first the control limit robustness, meanwhile, the other two PIs represent the effectiveness and also the speed of detection performances, respectively that conducted through on-line operating environment.

MATERIALS AND METHODS

Establishing the PCS loading factors from the inter-distance scale structure: Cox and Cox (1994) started the procedures by defining the major product moment (MjPM) of X, that is $B = XX^T$ where, X is an'n' (samples) by 'm' (variables)and the data is assumed to be mean centered as well as scaled to unit variance. By applying the singular value decomposition on B, the following are obtained:

$$B_{ui} = \lambda_i u_i \quad (1)$$

$$XX^T u_i = \lambda_i u_i \quad (2)$$

where, λ_i and u_i are eigenvalues(scalar) and eigenvector vectors of B, respectively while i = 1,2,...m.

The eigenvectors of Minor Product Moment (MnPM) are then obtained through the following procedures:

$$X^T X X^T u_i = \lambda_i X^T u_i \quad (3)$$

$$Cq_i^* = \lambda_i q_i^* \quad (4)$$

where, $C = X_T X$ (MnPM) and $q_i^* = X^T u_i$ (the corresponding MnPM eigenvectors).

However, q_i^* should be normalized by $\lambda_i^{-0.5}$, as to get the true orthogonal loading factors as similar in PCA which are derived as follows:

Normalizing:

$$q_i^*, q_i = X^T u_i \lambda_i^{-0.5} \quad (5)$$

And therefore:

$$q_i^T q_i = \left[X^T u_i \lambda_i^{-0.5}\right]^T \left[X^T u_i \lambda_i^{-0.5}\right] = \lambda_i^{-0.5} \lambda_i^{-0.5} u_i^T X X^T u_i = \lambda_i^{-1} \lambda_i^{+1} = 1 \quad (6)$$

In this proposed method, the dissimilarity scales will be developed in the form of inter-distance measure. As a result, B must be able to be derived from any means of inter-distance measure that applied (all of the distances will usually takes a form of matrix, D, with size 'n' by 'n'). Borg and Groenen (1997) as well as Cox and Cox (1994) have established the relationships between scalar product (B) and Euclidean distance, (D) as shown in Eq. 7:

$$b_{ij} = -\frac{1}{2}\left(d_{ij}^2 - \frac{1}{n}\sum d_{ij}^2 - \frac{1}{n}\sum d_{ij}^2 + \frac{i}{n^2}\sum_i \sum d_{ij}^2\right) \quad (7)$$

where, b_{ij} = scalar product between variables 'i' and 'j', d_{ij} = Euclidean distance between variables 'i' and 'j'.

Hence, any type of inter-distance scales can be accordingly transformed into B via Eq. 7 which conceptually means that those scales are perceived as emulating the Euclidean distance. This is particularly necessary because the projection of multivariate scores always reconfigured in the form of euclidean space (Cartesian coordinates). Another modification is also compulsory, particularly in normalizing the q_i^*. For any scale other than Euclidean distance, the normalization will be performed according to its own vector length. Thus, normalization of q_i produces:

$$q_i = \frac{q_i^*}{|q_i^*|} \quad (8)$$

where, $|q_i^*|$ = vector length, so that, $a_i^T qi = 1$.

Simply applying the Eq. 5 will not lead to orthogonal transformation as those scales are not originally Euclidean-based.

The new MSPM framework based on dissimilarity-based PCA technique: The generic MSPM frameworks typically require two phases of development (Mason and Young, 2002) as shown in Fig. 1. The first involves with the formation of the NOC model which basically contains four main steps. The first step deals with collecting and standardizing the NOC data from the historical process archive. Then, in the second step, a set of linear-based model of NOC data is developed via cPCA approach. Next, the third step deals with calculating the monitoring statistics (T^2 and SPE) while lastly control limits will be set up defining the warning and also control limits, respectively (the fourth step).

In this study, the generic proposed framework follows the similar MSPM procedures, however some modifications are made particularly in step 2 and 4 (readers are advised to gain further information of step 1 and 3 from the following references: MacGregor and Kourti (1995) or Raich and Cinar (1996). In step 2, the main loading factors are obtained through the procedures discussed previously which connect the scalar product (MjPM) and the selected inter-distance measures. In particular, the main stages are modeling the multivariate

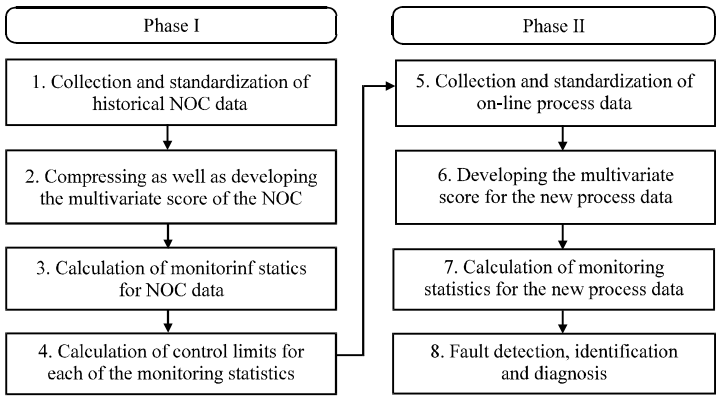

Fig. 1: Conventional MSPM framework

samples by way of inter-distance scales and producing matrix D, transforming D into B by using Eq. 7, finding the loading factors (q) as well as eigenvalues (squared root) which derived from the eigenvectors of B via Eq. 4 and finally, normalizing q through applying Eq. 8. In step 4, the monitoring limits are calculated based on chi squared distribution proposed by Nomikos and MacGregor (1995):

$$\lim_{\alpha} = (\nu/2\bar{m})\chi^2_{2\bar{m}^2/\nu,\alpha} \quad (9)$$

where, α equals to 0.05 and 0.01 for warning (95%) and control (99%) limits, respectively, $\bar{m}$ and $\acute{\imath}$ are, respectively representing the means and variances for each of the statistics.

The second phase of development which caters for on-line monitoring, is also involving four main steps. The first and the third basically follow the same procedures as in phase I. Once the loading factors have been obtained in step 2 of the phase I, the new scores can be easily projected by multiplying the new samples (standardized) with those loading factors specified the first phase. Unlike in phase I, the fourth step in phase II conducts fault detection operation. In particular, if a set of either statistics located outside the control boundary successively, the monitoring system will sound the alarm indicating that there is a special event happened in the process. Once the alarm triggered, fault identification will be performed in order to obtain a set of variables that may potentially contribute to the detected faulty condition. This will then followed by executing the fault diagnosis task particularly to determine clearly the true cause of the problem (but this normally demands higher complexity analyses and longer experiments).

RESULTS AND DISCUSSION

The Tennessee Eastman Process (TEP) has been chosen to demonstrate the monitoring capability of the proposed system. The comprehensive descriptions as well as the original simulation of TEP system were presented in Downs and Vogel (1993). This particular plant has been proposed in many works as a benchmark in evaluating the effectiveness of various schemes of controls, optimization techniques and also monitoring applications. In general, the system is consisted of five major unit operations including a reactor, a product condenser, a vapor-liquid separator, a recycle compressor and finally a product stripper. In particular, there are 21 manipulated and 41 measurement variables are monitored concurrently, as well as there are 20 types of abnormal operations considered in this study (as shown in Table 1).

All data including the NOC data (500 samples) as well as those 20 fault cases (each contains 960 samples) were obtained at http://brahms.scs.uiuc.edu, due to Chiang *et al.* (2001). In each of the fault batches, the fault was introduced in sample 160, whereby the fault will be detected, if and only if, 5 consecutive statistics located outside the 99% control limit. In particular, the new system has applied the City-block (City) and also Mahalanobis (Mahal) distances in scaling the original NOC data. All systems, including the conventional and also the proposed method, have been implemented based on 30 as well as 40 PCS respectively. Table 2 shows the proportion of transformed variation that modelled by each of the systems.

From Table 2, the results have found that cPCA has a slight advantage over the new systems as the numbers of transformed variations are comparatively higher. This can be understood by learning the fact that both of the new systems were using different scale from the Euclidean distance. Nonetheless, City-PCA has reached beyond 70% variation on both models and thus, the monitoring potential can be assumed to be equal to cPCA. Even though the percentage of Mahal-PCA was relatively lower compared to the other two systems particularly by using the 30 PCS, however, the 40 PCS has demonstrated higher

Table 1: List of abnormal operations of TEP system

Fault cases	Fault causes	Types
1-7	A/C feed ratio, B composition constant, B composition, A/C ratio constant, D feed temperature, reactor cooling water inlet temperature, condenser cooling water inlet temperature, A feed loss, C header pressure loss	Step
8-12	A, B, C feed composition, D feed temperature, C feed temperature, reactor cooling water inlet temperature, condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14-15	Reactor cooling water valve, condenser cooling water valve	Sticking
16-20	Unknown	-

Table 2: Proportion of transformed variation

	Transformed variation (%)		
No. of PCS	cPCA	City-PCA	Mahal-PCA
30	85	70	55
40	98	85	75

Table 3: False alarm rates for the conventional and new MSPM systems

		Percentages (%)					
		30 PCS			40 PCS		
Data	Monitoring statistics	cPCA	City-PCA	Mahal-PCA	cPCA	City-PCA	Mahal-PCA
NOC (original)	T^2	0.2	0.6	1.0	0.2	1.0	0.6
	SPE	0.4	0.8	1.2	1.6	1.6	0.8
NOC (testing)	T^2	1.6	22.0	2.2	2.8	5.8	2.2
	SPE	16.0	19.0	2.4	2.8	14.0	2.2

Table 4: The overall fault detection performances between the proposed and conventional MSPM system

	30 PCS			40 PCS		
Performance indicators	cPCA	City-PCA	Mahal-PCA	cPCA	City-PCA	Mahal-PCA
No. of cases detected	16	17	16	20	17	16
No. of fastest detection (relative)	4	12	7	6	15	2

percentage of transformation. All of these factors are crucial to be analysed thoroughly on the monitoring performances in order to obtain a correct understanding of the true implication made by the new approach. Meanwhile, the results of false alarm analysis based on the NOC data has been summarised as denoted in Table 3.

There are two forms of NOC data have been used in this study which are the original and testing sets. The first is normally applied for developing the multivariate model, whereas the second (960 samples) is typically utilised to assess the credibility of the developed control limits. The results in Table 3 indicate that the overall FAR for the original NOC is generally below 5 %,while the rates can reach as high as 22 % (City-PCA-T^2) for the testing set. The FAR values for the 40PCs are comparatively lower than the 30 PCS which conceptually prove that adopting higher number of PCS may help in obtaining excellent data projection as the transformed variations are also directly increased. Another important observation is that the FAR of City-PCA is generally higher than that of Mahal-PCA particularly for the 30 PCS model which basically promotes that the later scale is more robust compared to the former distance. Nonetheless, this study perceives the FAR performance of City-PCA as acceptable with moderate risk of false alarm (despite having large value of FAR).

The overall results of fault detection on those 20 cases are analysed and summarized in Table 4.

From Table 4, the generic performance of higher PCS has shown better outcomes compared to using smaller PCS based on both of the performance factors with regard to cPCA. This suggests that, adopting more PCS may enhance the performance capability of the conventional system, because more variations are transformed. Meanwhile, the generic performance of the proposed method (in both scales) has demonstrated almost equal performance between both PCS that selected, particularly in terms of number of faults detected. Nonetheless, the City-block scale has improved in terms of fault detection efficiency in contra to Mahalanobisscale, in which the results show that the variations transformed by the first scale are higher than the second scale by 10 % difference (the Mahalanobis scale has shown the lowest variation transformation than the other two methods).

In considering the overall performance, the City-block scale has been found the most efficient system than the other two in both of the PCS models that applied. This simply can be understood because the magnitude of FAR for City-PCA is comparatively larger than others and this has made the particular system to be highly sensitive. The conventional PCA can be considered the most effective detection system (40 PCS) while the 30 PCS outcomes has denoted the City-block scale as having the largest number of detections. However, it is also very important to report that, both of the new systems have also actually detected all the faults based on both of PCS settings, nonetheless, they were not counted due to the occurrence of minor false alarm events. Regarding cPCA, two of the cases were truly undetected, whereas the other two contain minor false alarms and all these have contributed to only 16 cases can be detected out of 20 based on 30 PCS.

Figure 2 and 3 denote the monitoring charts of T^2 and SPE, respectively of all systems for Fault 4 (F4). This particular fault occurs from a sudden temperature increase in the reactor as a result of inducing a step change reduction in the cooling water inlet of the reactor. The effect is then compensated immediately by a control mechanism and thus, bringing the process back to the normal condition. However, the root of the problem is still inherited in the system which eventually modifies the normal variation of the system dramatically.

In comparing both of the charts, SPE statistic has generally demonstrated consistent performance against T2 because most of the samples are located beyond the 99% limit starting from sample 161, particularly based on

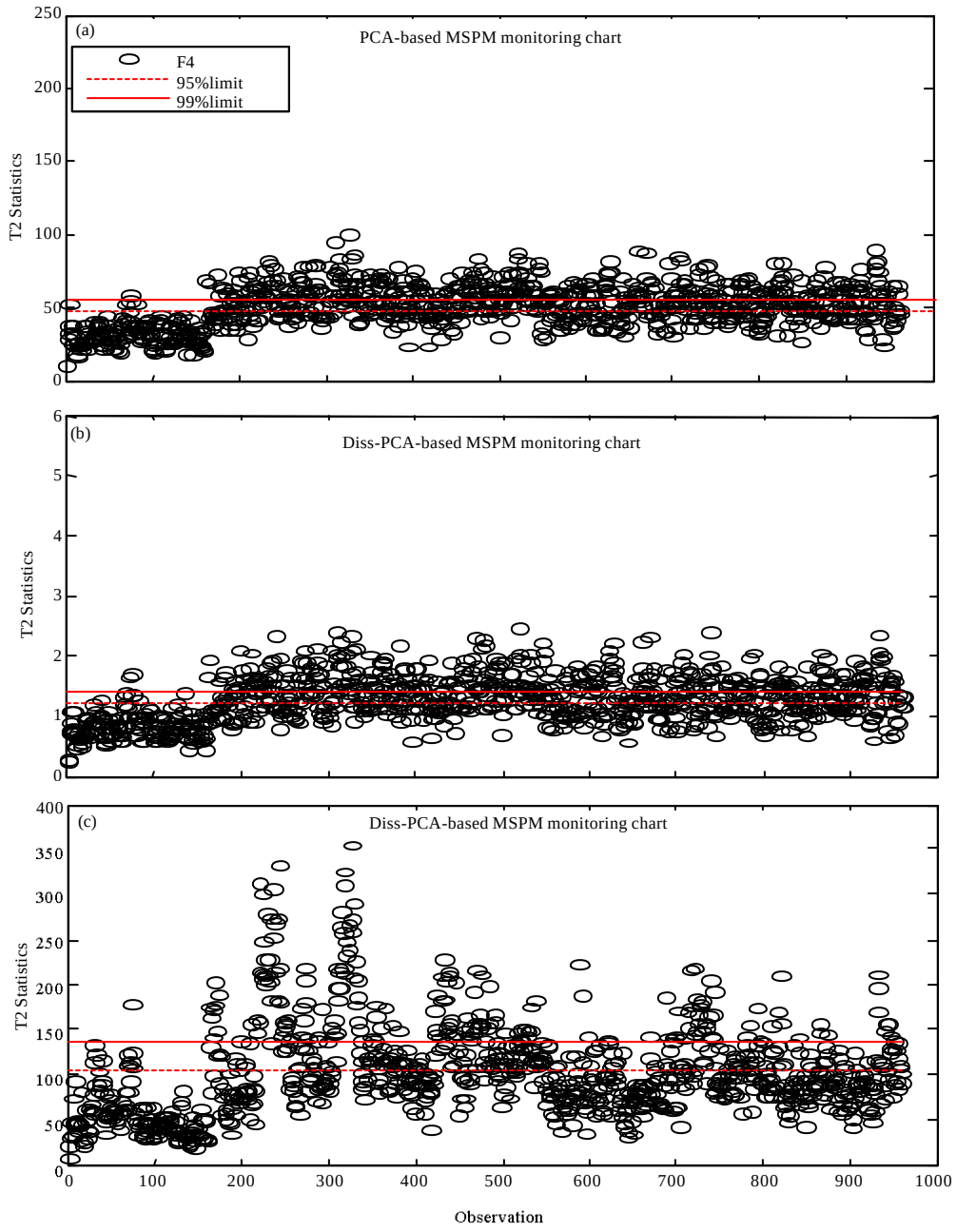

Fig. 2(a-c): T^2 charts for Fault 4 (F4) based on (a) 30 PCS of cPCA, (b) City-PCA and (c) Mahal-PCA

cPCA as well as City-PCA. Therefore, F4 has been detected efficiently as well as effectively by both cPCA and City-PCA systems which is 1 sampling time after the fault introduced into the process and this can be perceived as very excellent performance. More interestingly, this superb performance can be achieved by City-PCA by using only 30 PCS and that contributes around 75% variation. This suggests that the City-PCA system can potentially produce equal performance against cPCA, despite that the transformed variation is relatively smaller than the conventional system.

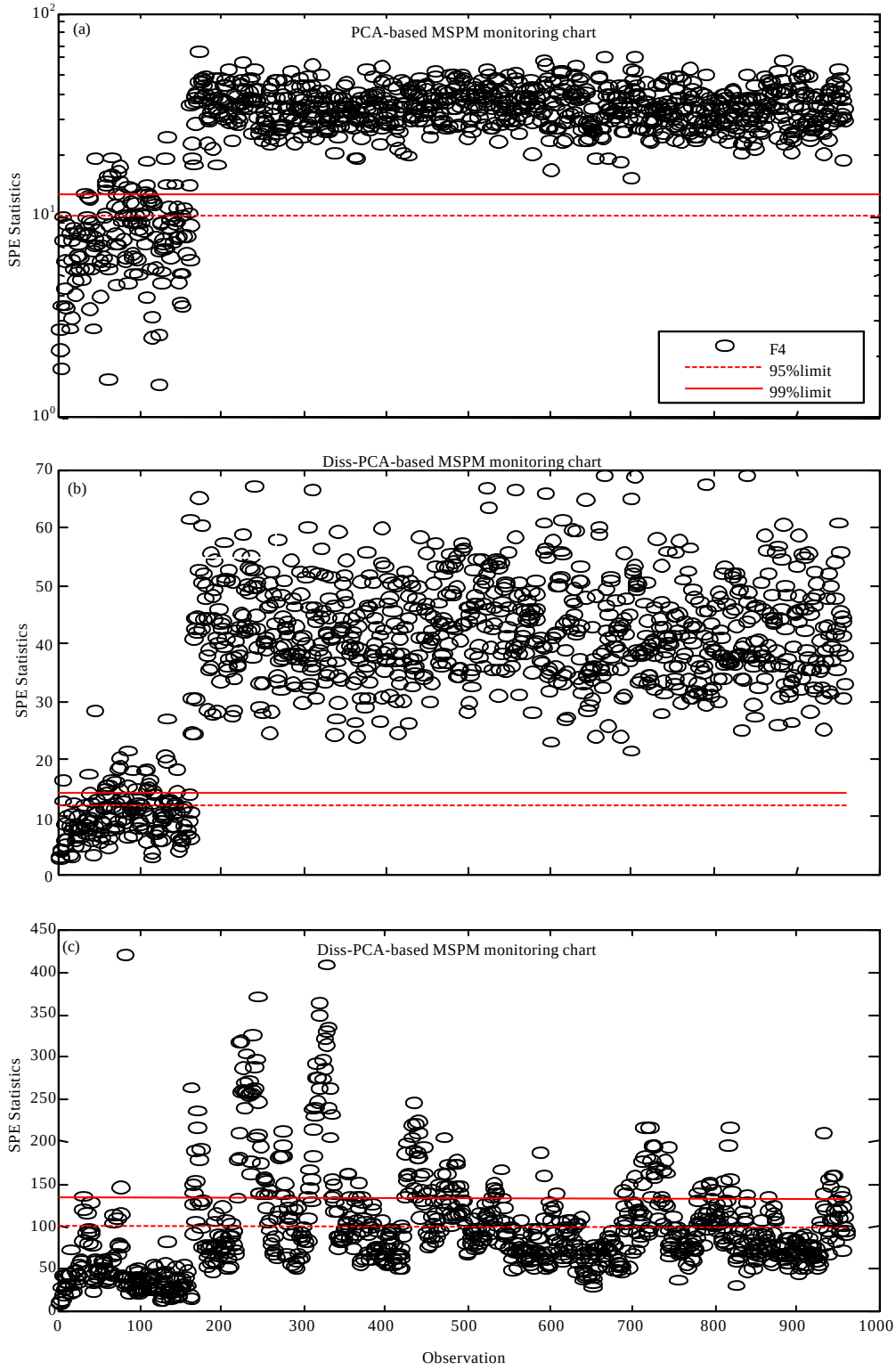

Fig. 3(a-c): SPE charts for Fault 4 (F4) based on (a) 30 PCs of cPCA (b) City-PCA and (c) Mahal-PCA

On the other hand, the general performance of Mahal-PCA has been found slower and also inconsistent compared to others. The first detection can be noticed at sample 217 (even though a number of samples have been found located outside the 99% limit much earlier but none of those samples denoting at least 5 samples in succession) and this is considerably delay in connection to both of previous performances. Besides, it can be also clearly seen that many of the samples accentuated below the control limits on both of T^2 and SPE charts which indicating problematic trending during the detection operation. Nonetheless, this particular method is capable of detecting the fault, although some weaknesses are observed.

CONCLUSION

In this study, a new MSPM framework is proposed by introducing the dissimilarity scale as means in producing the loading factors of PCA. The overall results on the TEP system have proved that the new system is reliable in detecting the specified faults by demonstrating high numbers of successful detections either by means of number of cases detected or number of fastest detection. Despite that few cases of minor false alarms have been observed during on-line monitoring, the rates are small and normally they are stabilized very quickly. All of these findings suggest that the new system has been sufficiently developed according to the concept of MSPM principles. The beauty of this newly method over the conventional approach is that it can deal both of the quantitative and also qualitative data simultaneously. This will then perhaps upgrade as well as expose the current practice of monitoring application with abundant amount of qualitative measurements and making the productivity becomes highly sensitive to any forms of malfunction operations.

ACKNOWLEDGMENT

Authors would like to dedicate a special gratitude to Universiti Malaysia Pahang for funding the study through Vot RDU 110371.

REFERENCES

Borg, I. and P. Groenen, 1997. Modern Multidimensional Scaling: Theory and Applications. Springler-Verlag, New York, USA.

Chiang, L.H., Russell and E.L. Braatz, 2001. Fault Detection and Diagnosis in Industrial Systems. Springer Verlag, Great Britain.

Cox, T.F., 2001. Multidimensional scaling used in multivariate statistical process control. J. Applied Statist., 28: 365-378.

Cox, T.F. and M.A.A. Cox, 1994. Multidimensional Scaling. Chapman and Hall, Great Britain, London.

Dong, D. and T.J. McAvoy, 1996. Nonlinear principal component analysis-based on principal curves and neural networks. Comput. Chem. Engin., 20: 65-78.

Downs, J.J. and E.F. Vogel, 1993. A plant-wide industrial process control problem. Comput. Chem. Engin., 17: 245-255.

MacGregor, J. F. and T. Kourti, 1995. Statistical process control of multivariate processes. Contr. Engin. Pract., 3: 403-414.

Mason, R.L. and J.C. Young, 2002. Multivariate Statistical Process Control with Industrial Applications. ASA-SIAM., USA.

Nomikos, P. and J.F. MacGregor, 1995. Multivariate SPC charts for monitoring batch processes. Technometrics, 37: 41-59.

Qin, S.J., 2003. Statistical process monitoring: Basics and beyond. J. Chemomet., 17: 480-502.

Raich, A. and A. Cinar, 1996. Statistical process monitoring and disturbance diagnosis in multivariable continuous processes. AIChE J., 42: 995-1009.

Yunus, M.Y.M., 2012. Multivariate statistical process monitoring using classical multidimensional scaling. Ph. D Thesis, Newcastle University, Newcastle Upon Tyne, UK.

Zhang, J., E.B. Martin and A.J. Morris, 1997. Process monitoring using non-linear statistical techniques. Chem. Engin. J., 67: 181-189.