High-Pressure Polyethylene Process Monitoring Using PCA Based Bayesian Classification

Jialin Liu

Department of Information Management, Fortune Institute of Technology,
125-8, Chyi-Wen Rd., Chyi-Shan 842 Kaohsiung Country, Taiwan, Republic of China
E-Mail: jialin@center.fjtc.edu.tw

Abstract: Using PCA based Bayesian Classification to monitor the real plant with different operated conditions is proposed. Since the process conditions are time-variant, as the PCA subspace cannot explain the data of new events, the PCA should be re-performed. In this work the method of updating Bayesian model is developed. Only the data of new events are trained in the newer subspace. The ability of PCA based Bayesian classification for monitoring different operating conditions is demonstrated using the real data from high-pressure polyethylene plant. *Copyright © 2004 IFAC*

Keywords: Process monitoring, principal component analysis, Bayesian classification

1. INTRODUCTION

A modern plant is well equipped with automation control system. Large amounts of operation data are generated and recorded. In plant history, the data of different operating conditions, normal and abnormal, are included. Data mining from historical data and finding out some of useful information provide to operators for decision-making, especially in abnormal situation that will effectively reduce plant accidents. Feature extraction from amounts of historical data is the first step of data mining. In complex chemical process, the measured variables are highly correlated with each other. Principal component analysis is one of popular approaches to compress highly dimensional variables to fewer principal components (Jackson, 1991). It is more efficient using extracted features for fault detection and isolation. Such as, Teppola et al. (1999) monitored the waste-water treatment plant in paper mill using PCA and fuzzy c-means clustering. Choi et al. (2003) predicted SOx and NOx from a power plant of steel mill through partial least square and credibilistic fuzzy c-means method.

Bayesian classification (Theodoridis and Koutroumbas, 1999) is a popular method of cluster analysis. Wang and McGreavy (1998) isolated the different operating conditions of a refinery fluid catalytic cracking process using automatic classification. Chen and Liu (1999) used mixture PCA to detect the abnormalities in different operating conditions. Generally, for clustering data, the expectation maximization (EM) algorithm is used for fitting the parameters of mixture model. However, EM algorithm suffers three significant difficulties. (1) Due to EM is a local maximum seeker; the different initial values will cause the different convergences. (2) The outliers easily affect the searching path that results in the lack of robustness of mixture model. (3) Predefined the number of clusters is required. Typically it needs multiple runs with various numbers of clusters for cluster validity. Ueda and Nakano (1998) proposed deterministic annealing EM (DAEM) algorithm in order to overcome the local maximum problem. They analogised the concept of maximum entropy of statistical mechanics, iteratively maximized entropy using a "temperature" parameter to achieve a stable solution. Medasani and Krishnapuram (2001) proposed robust mixture decomposition (RMD) algorithm retain "good" data for fitting parameters, reduced the effect of outliers. There are a lot of criteria for cluster validation, like Akaike's information criterion (AIC), Bayesian information criterion (BIC), normalized entropy criterion (NEC) and so no. McLachlan and Peel (2000) have compared the differences of these criteria. In this work, the BIC is used to determine the number of clusters.

This paper is organized as followed. The basic theory of PCA and Bayesian classification related algorithm is presented in section 2. Section 3 shows the method of PCA based Bayesian classification for process monitoring. Besides, the updating procedure of mixture model is proposed. Applied proposed process monitoring method to real plant data from high-pressure polyethylene plant is demonstrated in section 4. Finally the conclusions are given in section 5.

2. BASIC THEORY

2.1 Principal Component Analysis

The concept of PCA is linear combination of variables to perform a subspace. The dataset can be represented by less principal components or latent

variables. Therefore, numbers of variables are reduced, and meanwhile data are explained properly.

Considered the data matrix $\mathbf{Y} \in R^{m \times n}$ with m rows (observations) and n columns (variables), and each column are normalized to zero mean and unit variance.

$$\mathbf{X} = \left(\mathbf{Y} - \overline{\mathbf{Y}}\right)\mathbf{S}^{-1} \tag{1}$$

where $\overline{\mathbf{Y}}$ is a mean vector and $\mathbf{S}$ is a diagonal matrix of standard deviations. The eigenvectors of covariance matrix and the score vectors are performed as followed.

$$\begin{aligned} &\mathbf{S} = \frac{1}{m-1}\mathbf{X}^T\mathbf{X} \\ &\mathbf{S}\mathbf{p}_i = \lambda_i \mathbf{p}_i \,, i=1\ldots n \\ &\mathbf{t}_i = \mathbf{X}\mathbf{p}_i \,, i=1\ldots n \end{aligned} \tag{2}$$

where λ_i are the eigenvalues associated with eigenvectors $\mathbf{p}_i$ and arranged in descending order. The data matrix $\mathbf{X}$ can be decomposed as:

$$\begin{aligned} \mathbf{X} &= \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}} + \mathbf{t}_2\mathbf{p}_2^{\mathrm{T}} + \cdots + \mathbf{t}_K\mathbf{p}_K^{\mathrm{T}} + \mathbf{t}_{K+1}\mathbf{p}_{K+1}^{\mathrm{T}} + \cdots + \mathbf{t}_n\mathbf{p}_n^{\mathrm{T}} \\ &= \hat{\mathbf{X}} + \mathbf{E} \end{aligned} \tag{3}$$

The data matrix $\mathbf{X}$ is divided into two parts, systemic part $\hat{\mathbf{X}}$ and residual part $\mathbf{E}$. As the first K terms of components can explain data matrix properly, i.e. $\mathbf{X} \approx \hat{\mathbf{X}}$. The numbers of dimensions of data matrix are reduced from n to K.

There are amount of researches using PCA for fault detection (Russell *et al.*, 2001). The main concept is building PCA subspace via normal operating data from plant. In the on-line monitoring stage, the residuals of the new data $\mathbf{x}$ can be calculated from Eq. 3. The statistic Q is defined:

$$Q = \mathbf{e}\mathbf{e}^{\mathrm{T}} = \left(\mathbf{x} - \hat{\mathbf{x}}\right)\left(\mathbf{x} - \hat{\mathbf{x}}\right)^{\mathrm{T}} = \mathbf{x}\left(\mathbf{I} - \mathbf{P}_K\mathbf{P}_K^{\mathrm{T}}\right)\mathbf{x}^{\mathrm{T}} \tag{4}$$

The loading vectors $\mathbf{P}_K \in R^{m \times K}$ are the first K terms of eigenvectors of covariance matrix. Statistic Q is a measure of the approximated error. The confidence limit of Q is defined as followed:

$$\begin{aligned} &Q_\alpha = \Theta_1\left[\frac{c_\alpha\sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0\left(h_0 - 1\right)}{\Theta_1^2}\right]^{1/h_0} \\ &h_0 = 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2}, \Theta_i = \sum_{j=K+1}^{n}\lambda_j^i \qquad i = 1,2,3 \end{aligned} \tag{5}$$

The percentile α is the probability of Type I error in hypothesis testing and c_α is the normal deviate cutting off an area of the α upper tail of normal distribution. Another measuring difference between new data and the subspace of PCA is statistic T^2.

$$T^2 = \mathbf{x}\mathbf{P}_K\,?^{-1}\mathbf{P}_K^{\mathrm{T}}\mathbf{x}^{\mathrm{T}} \tag{6}$$

The diagonal matrix $?$ is K terms of eigenvalues. The confidence limit is defined:

$$T_\alpha^2 = \frac{K\left(m-1\right)}{m-K}F_{K,m-1,\alpha} \tag{7}$$

where $F_{K,m-1,\alpha}$ is the F distribution with degrees of freedom K and $m-1$. In this paper, statistic Q and T^2 are used to verify the score vector can represent the new data or not. If not, using score vector for classifying will possibly result in misclassification.

2.2 Bayesian Classification

Given a classification task of c classes, the priori probabilities are $P_j, j = 1\ldots c$. The joint probabilities density functions of $\mathbf{x}_i$ and class j are $p\left(\mathbf{x}_i, j\right)$. Then, the conditional probability density functions of $\mathbf{x}_i$ in class j are:

$$p\left(\mathbf{x}_i \middle| j;?\right) = p\left(\mathbf{x}_i, j\right)\Big/P_j \tag{8}$$

where $?$ are the parameter vectors of conditional probability density functions. According to total probability theorem, the probability density functions of $\mathbf{x}_i$ are $p\left(\mathbf{x}_i;?\right) = \sum_{k=1}^{c} p\left(\mathbf{x}_i \middle| k;?\right)P_k$. The posteriori probabilities can be gotten from Bayes rule.

$$P\left(j \middle| \mathbf{x}_i;?,\mathbf{P}\right) = \frac{p\left(\mathbf{x}_i \middle| j;?\right)P_j}{p\left(\mathbf{x}_i;?\right)} = \frac{p\left(\mathbf{x}_i \middle| j;?\right)P_j}{\sum_{k=1}^{c} p\left(\mathbf{x}_i \middle| k;?\right)P_k} \tag{9}$$

Here $\mathbf{P}$ are vector of priori probabilities, not the loading vector of PCA. If the parameters of conditional probability density functions and priori probabilities are known, the posteriori probabilities can be obtained from Eq. 9.

Expectation-Maximization Algorithm. The expectation maximization algorithm maximizes the expectation of loglikelihood function. The two steps of EM algorithm are:

E-Step:

$$\begin{aligned} Q\left(\mathbf{T};\mathbf{T}\left(t\right)\right) &\equiv E\left[\sum_{i=1}^{m} ln\left(p\left(\mathbf{x}_i \middle| j;?\right)P_j\right)\right] \\ &= \sum_{j=1}^{c}\sum_{i=1}^{m} P\left(j \middle| \mathbf{x}_i;\mathbf{T}\left(t\right)\right) ln\left(p\left(\mathbf{x}_i \middle| j;?\right)P_j\right) \end{aligned} \tag{10}$$

where $\mathbf{T}(t)$ are the parameters at the t^{th} step of iteration. The posteriori probabilities $P(j|\mathbf{x}_i;\mathbf{T}(t))$ can be obtained from Bayes rule.

M-Step:
Compute the next estimating parameters through maximizing the expectation of loglikelihood function.

$$\frac{\partial Q(\mathbf{T};\mathbf{T}(t))}{\partial \mathbf{T}} = 0 \tag{11}$$

Take a multivariate Gaussian probability density function as an illustrated example:

$$p(\mathbf{x}_i|j;?) = \frac{exp\left(-\frac{1}{2}(\mathbf{x}_i-\boldsymbol{\mu}_j)\mathbf{S}_j^{-1}(\mathbf{x}_i-\boldsymbol{\mu}_j)^T\right)}{(2\pi)^{n/2}|\mathbf{S}_j|^{1/2}} \tag{12}$$

where $\boldsymbol{\mu}_j$, $\mathbf{S}_j$ are the mean vector and covariance matrix of the j^{th} class. Using Lagrange multipliers the iterative steps can be derived:

$$P_j(t+1) = \frac{1}{m}\sum_{i=1}^{m} P(j|\mathbf{x}_i;\mathbf{T}(t)), j=1\ldots c \tag{13}$$

$$\boldsymbol{\mu}_j(t+1) = \frac{\sum_{k=1}^{m} P(j|\mathbf{x}_k;\mathbf{T}(t))\mathbf{x}_k}{\sum_{k=1}^{m} P(j|\mathbf{x}_k;\mathbf{T}(t))} \tag{14}$$

$$\mathbf{S}_j(t+1) = \frac{\sum_{k=1}^{m} P(j|\mathbf{x}_k;\mathbf{T}(t))(\mathbf{x}_k-\boldsymbol{\mu}_j(t))^T(\mathbf{x}_k-\boldsymbol{\mu}_j(t))}{\sum_{k=1}^{m} P(j|\mathbf{x}_k;\mathbf{T}(t))} \tag{15}$$

The fitting vector of parameters $\mathbf{T}$ can be obtained by repeating E and M steps until $\|\mathbf{T}(t+1)-\mathbf{T}(t)\| \le \varepsilon$, ε is the properly convergence criterion.

Deterministic Annealing EM Algorithm In general, the initial values of parameters $\mathbf{T}(0)$ are randomly selected, the posteriors are not reliable in early stage of the iteration. Relatively mislead the direction of seeking solution. However, the fitting parameters will be easily trapped into local maximum. Ueda and Nakano (1998) proposed deterministic annealing EM algorithm to resolve this problem. They introduced another posterior $F(j|\mathbf{x}_i;\mathbf{T}(t))$:

$$F(j|\mathbf{x}_i;\mathbf{T}(t)) = \frac{\left[p(\mathbf{x}_i|j;?(t))P_j(t)\right]^{\beta}}{\sum_{k=1}^{c}\left[p(\mathbf{x}_i|k;?(t))P_k(t)\right]^{\beta}},$$
$$0 \le \beta \le 1 \tag{16}$$

In a special case $\beta = 0$, the posterior is a uniform distribution unaffected by initial value of parameters. In another case $\beta = 1$, the posterior reduces to the original posterior. The β increases from 0 to 1, that means the posterior changes from a uniform distribution to original posterior. This procedure analogises to simulated annealing, $1/\beta$ is like the "temperature" of annealing. They add a β-loop on the EM-steps and replace posterior by Eq. 16.

The DAEM algorithm can be derived.

1. Setting initial value of β, $\beta \approx 0$.
2. The best solution of $\mathbf{T}_\beta$ can be obtained by iterating EM-steps until convergence.
3. If $\beta < 1$, increasing β and setting initial value of parameters through $\mathbf{T}_\beta$ of step 2. Repeat from step 2 until $\beta = 1$.

Robust Mixture Decomposition Algorithm Although the DAEM can overcome the local maximum problem, the outliers still affect the solution of DAEM. Unfortunately, outliers are unavoidable in real plant data. To resolve this difficulty, Medasani and Krishnapuram (2001) proposed the robust mixture decomposition (RMD) algorithm. They argued that the value of $p(\mathbf{x}_i|?)$ of the outlier $\mathbf{x}_i$ would have a low value when **?** is correctly estimated. Therefore, they trimmed a fraction of the data by sorting probability functions for each $\mathbf{x}_i$, i.e. $p(\mathbf{x}_1|?) \ge p(\mathbf{x}_2|?) \ge \cdots \ge p(\mathbf{x}_m|?)$ before each EM iteration, retain the first *r* terms of data for fitting parameters. That will effectively reduce the influence of outliers. The retention ratio (*r*/*m*) can be estimated from the data quality, or be included as one of the parameters to be optimised. The smaller retention ratio may underestimate the covariances, but can stabilize the solution for fitting parameters.

Bayesian Information Criterion Using the fitting parameters $\mathbf{T}$ approximate the expectation of loglikelihood. The BIC is defined as:

$$BIC_c = -2\times Q(\mathbf{T}) + \nu_c \ln(m) \tag{17}$$

where ν_c is number of fitting parameters. However, maximized the expectation of loglikelihood is to minimize BIC. Using multiple runs with various numbers of parameters can decide the number of cluster from the minimizing BIC.

3. PCA BASED BAYESIAN CLASSIFICATION

The PCA based Bayesian classification is to find out the clusters of different operating conditions using score vectors. Firstly, the subspace of PCA can be constructed from plant data. The highly dimensional

variables can be effectively reduced that will simplify the Bayesian model and depress the computing time of fitting parameters. In order to obtain the reasonable clusters, the score vectors of historical data are clustered through combination of DAEM and RMD algorithms in this work. However, the operating conditions of real plant are time-variant. As new data cannot be explained by existed subspace, the subspace should be reconstructed from all of data. It is more consistent using one PCA subspace describes the whole dataset than multiple PCA subspaces. Besides, since the new data don't belong to the existed subspace that must be independent with existed classes. Therefore, only the data of new events are clustered on the newer subspace, the trained dataset don't need to be performed clustering again. The method of updating Bayesian model is proposed; the Bayesian model of the previous subspace can be updated to the newer subspace.

3.1. Fault Isolation

The statistic Q and T^2 of on-line data can be conducted from Eq. 4 and Eq. 6. As $Q > Q_\alpha$ or $T^2 > T_\alpha^2$, indicate the new event has occurred, existed PCA cannot explain the new data properly. The subspace of PCA should be rebuilt. On the other hand, $Q < Q_\alpha$ and $T^2 < T_\alpha^2$, the score vectors of on-line data are used for calculating the posteriors $P(j|\mathbf{t}) = p(\mathbf{t}|j;\mathbf{?})\,P_j \Big/ \sum_{k=1}^{c} p(\mathbf{t}|k;\mathbf{?})P_k$. That can be used to identify the on-line data belonging to which operated condition has happened. If the posteriors are evaluated using score vectors without verify the on-line date can be explained by existed subspace, that could result in misclassifying the operating condition, since the new data might not belong to the existed subspace.

3.2. Scheme of Updating Bayesian Model

Adding the data of new events to the original data matrix, the total number of data is added from m to m^*. The data matrix $\mathbf{Y}^* \in R^{m^* \times n}$, $\mathbf{Y}^{*\mathrm{T}} = \left[\mathbf{Y}^{\mathrm{T}} \quad \mathbf{Y}_{\mathrm{new}}^{\mathrm{T}}\right]$, is normalized to $\mathbf{X}^* = \left(\mathbf{Y}^* - \overline{\mathbf{Y}}^*\right)\mathbf{S}^{*-1}$. The mean vector and diagonal matrix of standard deviation are $\overline{\mathbf{Y}}^*$ and $\mathbf{S}^*$. The first K^* ($K^* \geq K$, the original number of PCs) terms of loading vectors $\mathbf{P}^*$ and correspondingly score vectors $\mathbf{T}^*$ can be determined by $\mathbf{E}^* \approx 0$. Since the original dataset can be explained by both of subspaces.

$$\mathbf{Y} \approx \overline{\mathbf{Y}} + \mathbf{t}\mathbf{P}^{\mathrm{T}}\mathbf{S} \approx \overline{\mathbf{Y}}^* + \mathbf{t}^*\mathbf{P}^{*\mathrm{T}}\mathbf{S}^* \tag{18}$$

Besides, the values of conditional probability density functions should be equalized on both subspaces. For example, the conditional probability density functions of score vectors in the j^{th} class is:

$$p(\mathbf{t}|j) = exp\left(-\frac{1}{2}\left(\mathbf{t}-\boldsymbol{\mu}_j\right)\mathbf{S}_j^{-1}\left(\mathbf{t}-\boldsymbol{\mu}_j\right)^{\mathrm{T}}\right) \Big/ (2\pi)^{K/2}\left|\mathbf{S}_j\right|^{1/2}$$

On both of subspaces, the exponential parts have to be equalized.

$$\left(\mathbf{t}-\boldsymbol{\mu}_j\right)\mathbf{S}_j^{-1}\left(\mathbf{t}-\boldsymbol{\mu}_j\right)^{\mathrm{T}} = \left(\mathbf{t}^*-\boldsymbol{\mu}_j^*\right)\mathbf{S}_j^{*-1}\left(\mathbf{t}^*-\boldsymbol{\mu}_j^*\right)^{\mathrm{T}} \tag{19}$$

The new parameters of conditional probability density functions can be obtained.

$$\boldsymbol{\mu}_j^* = \left(\boldsymbol{\mu} - \Delta\overline{\mathbf{y}}\mathbf{S}^{-1}\mathbf{P}\right)\mathbf{A}^{+} \tag{20}$$

$$\mathbf{S}_j^* = \mathbf{A}^{+\mathrm{T}}\mathbf{S}_j\mathbf{A}^{+} \tag{21}$$

$$\mathbf{A} = \mathbf{P}^{*\mathrm{T}}\mathbf{S}^{*}\mathbf{S}^{-1}\mathbf{P} \tag{22}$$

where defines $\Delta\overline{\mathbf{y}} \equiv \overline{\mathbf{y}}^* - \mathbf{y}$, $\mathbf{A}^{+}$ is the pseudo inverse of $\mathbf{A}$, and updated priors are $P_j^* = m \times P_j \big/ m^*$. Through Eq. 20 to Eq. 22, the previous Bayesian model can be updated to the newer subspace.

4. EXAMPLE

This polyethylene plant is in Taiwan, there are five different operation lines; produce HDPE, LDPE, linear LDPE and so on. In this work the high-pressure process is studied, the operating conditions are more than 20 different grades in this process. The high purity ethylene (FI-001 in Figure 1) mixed with recycled gas before entering primary compressor, the amount of different modifiers (FIC-003 and FIC-004) will be added via the recipe. The pressure through secondary compressor (FI-002) is around 1500~2000 kg/cm^2-g. There are several reacted zones in reactor; the initiators maintain the temperature profile. The positions of temperature controllers (TIC-001, TIC-002 and TIC-003) are located closely on the top, middle and bottom of the reactor. The most important product quality is melt index (MI) that is significantly affected by temperature profile and pressure in reactor. The pressure is regulated by PIC-001 through the under flow of reactor. The mixture of polyethylene and unreacted gas will be flashed in two stages. Firstly, the higher-pressure gas is recycled to the entrance of secondary compressor from the top of high-pressure separator. Almost all of polymer and rest of unreacted gas will flow to extrusion hopper. PE will be pumped to extruder from bottom of extrusion hopper; the lower-pressure gas is recycled to the entrance of primary compressor from top of extrusion hopper. The reaction conversion is only around 10~15%, the most of unreacted gas continuously recycle in this process. The byproducts will be accumulated until the product quality is out of control limit and operator will be notified to purge recycled gas from FIC-006 by operator experience.

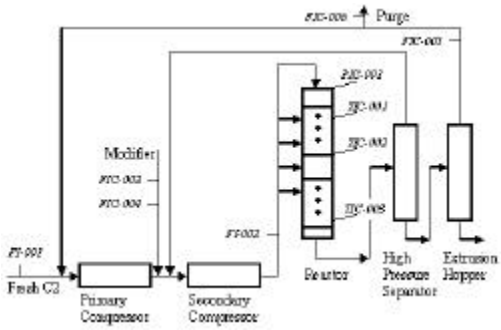

Fig. 1: Process flow diagram of high-pressure polyethylene plant

In this work, the reactor condition monitoring is focused. Not only the melt index is influenced by the reacted condition, but also this exothermic reaction is easily run away with slightly unstable temperature control. Besides, the pressure of reactor is higher than 1500 kg/cm^2-g, that is a most important issue to monitor the operating condition of reactor in this plant. The plant data have been collected every 5 minutes for a month. The data of shutdown has been trimmed, the training set contain 8928 data for fitting the parameters of Bayesian model. Thirteen process variables are collected, including three temperatures, respectively flow rates of initiators and the pressure in the reactor. Besides, the flow rates of two different modifiers, the top flow rate of extruder hopper and the other flow rates including purge, fresh ethylene, and secondary compressor. There exist seven different production grades in training dataset. The temperature profile and pressure for each grade are shown in Figure 2, the production grades are labelled from grade #1 to #7. For confidential concerned the process values are hidden, noteworthily the temperature difference can be greater than 70 degree C between different grades, such as grade #1 and #6. The variation of operating condition will be taken into account for production scheduling; in order to avoid the run away reaction will be caused as the transition of grades.

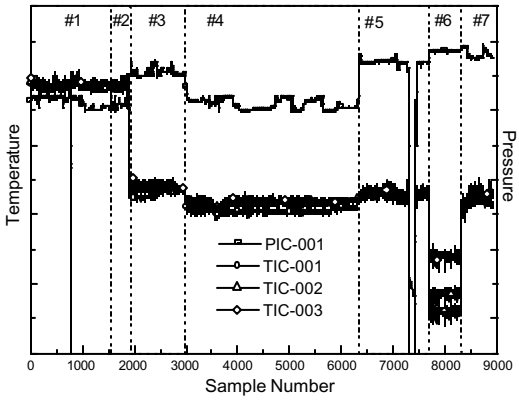

Fig. 2: Reactor temperatures and pressure for different grades.

Defined the subspace of PCA from plant data, the four principal components explain 90% of total variances. Due to the outliers of raw data, the retention ratio of RMD is set 0.8. In this example, the initial value of β in DAEM is set to 0.2 and the iterative procedure with $\beta_{new}=1.5\times\beta_{old}$ until $\beta=1$. The 9 clusters are determined with the minimizing BIC. The clustering result projected to the first two PCs is shown in Figure 3. Since the operating conditions of grade #1 and #2 are similar, those data share with the cluster C1&2. The data of grade #3 and #5 look like overlapped, but in the second and third PCs those data can be well separated. In the grade #4, there nearly exist two clusters C4_1 and C4_2. Consult with the expert of the plant; it is due to the different sources of ethylene feed. The purity of ethylene is slightly varied from different upstream vendors. That will cause the byproducts are easily accumulated as a little of impurity in feed ethylene. The operator will maintain the purity of recycled gas through purging the recycled gas. That is the reason of the different operated conditions in the same grade.

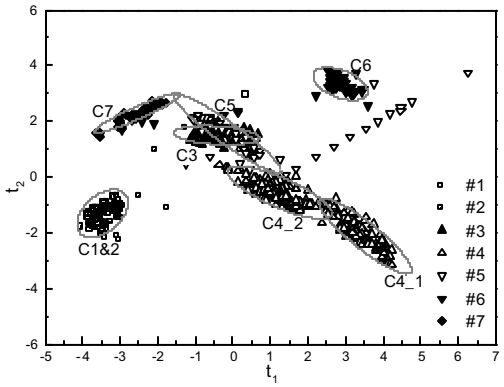

Fig. 3: First and second score vectors of trained data, and 9 clusters.

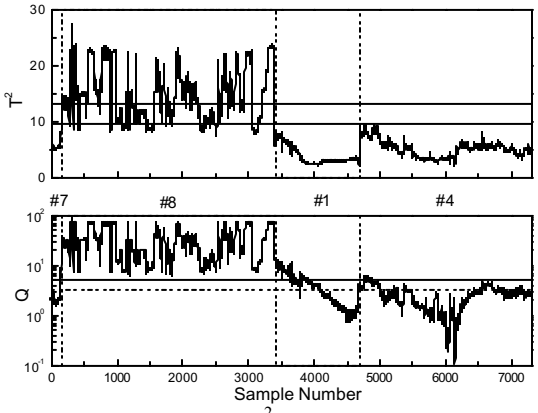

Fig. 4: The statistic Q and T^2 of the new data.

The new data are collected after training dataset the period is also one month. The dataset includes lasting data of grade #7, a new grade labelled #8, and grade #1, #4 have appeared in the training dataset. Before using the score vectors of PCA for classification, the statistic Q and T^2 of the new data have to be verified. In Figure 4, the statistic Q and T^2 of the new data for grade #7, #4 and part of #1 are within control limits. The data of grade #8 and leading part of grade #1 cannot be explained by existed subspace. Since grade #8 is a new event, it is

reasonable for PCA cannot explain these data. The melt index of grade #8 is much lower than grade #1; the ratio of two melt indexes is around 0.01. For the lower MI the impurity of recycled gas must be as less as possible. Therefore, the higher MI can suffer more impurity. Since that operator will not purge recycled gas in the transition from the lower MI to higher MI until the impurity reach the tolerable limit. It is the reason that the data of leading part of grade #1 cannot be explained by existed subspace.

Project the data can be explained by PCA to the subspace, the first two of score vectors are shown in Figure 5. That conform the previous result of the cluster. In order to include the new events, the training data set is added the data of new events. The subspace of PCA is reconstructed; the first 5 PCs explain 94% of total variances. Only the data of new events need to be clustered in the newer subspace. The clusters C8 and C1_2 can be gotten through DAEM and RMD algorithms. The previous clustering result can be updated to the new subspace using the proposed method described in section 3.2. In Figure 6, the clustering result in the new subspace is shown.

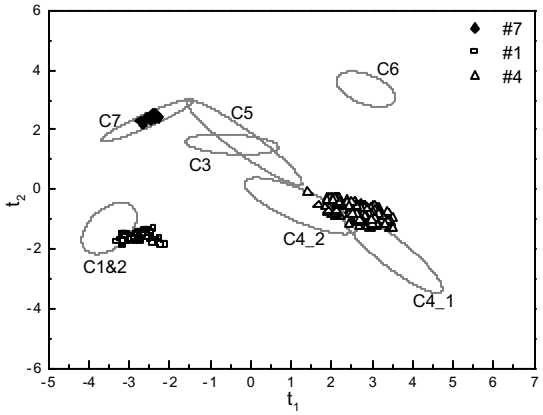

Fig. 5: On-line isolating the new data belonging to existed subspace.

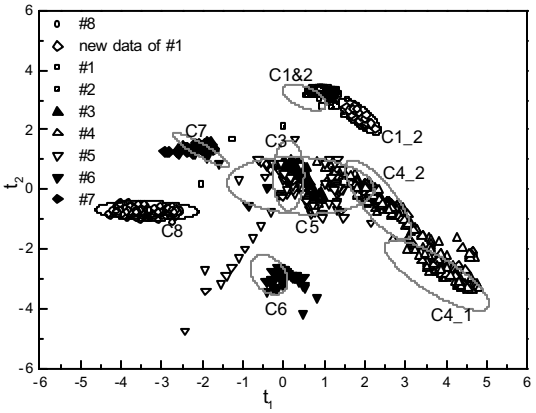

Fig. 6: Updating classified model and new clusters, C8 and C1_2.

5. CONCLUSIONS

The proposed method is successfully applied to the high-pressure polyethylene process monitoring. There exist several operating conditions for different grades in this process. Before using score vectors of PCA for on-line monitoring, the statistic Q and T^2 must be verified by hypothetically testing. As the data of new events cannot be explained by existed subspace, the newer subspace will be reconstructed by all of data for covering all of events. In the newer subspace, only the data of new events are clustered with DAEM and RMD algorithms. The number of clusters is determined through minimizing BIC. The previous clustering result can be directly updated to newer subspace with proposed method. Besides, the physical meaning of each cluster should be identified. Consult with the expert of plant or find out through operator log will be helpful to recognize the root causes.

ACKNOWLEDGMENT

This work is supported by the National Science Council, Republic of China, under Grant NSC-92-2213-E-268-001.

REFERENCES

Chen, J. and Liu, J. (1999) Mixture principal component analysis models for process monitoring. *Ind. Eng. Chem. Res*. **38**, 1478-1488.

Choi, S. W., Yoo, C. K. and Lee, I. (2003) Overall statistical monitoring of static and dynamic patterns. *Ind. Eng. Chem. Res*., **42**, 108-117.

Jackson, J. E. (1991). *A User's Guide to Principal Components*, Wiley, New York.

McLachlan, G. J. and Peel, D. (2000) *Finite mixture models*. Wiley, New York.

Medasani S. and Krishnapuram R. (2001) Categorization of image databases for efficient retrieval using robust mixture decomposition. *Computer Vision and Image Understanding.*, **83**, 216-235.

Nimmo, I. (1995). Adequately address abnormal situation operations. *Chem. Eng. Prog.* **91**, 36-45.

Russell, E. L., Chiang, L. H. and Braatz, R. D., (2001). *Fault Detection and Diagnosis in Industrial Systems.* Springer, London.

Teppola, P., Mujunen, S and Minkkinen, P. (1999) Adaptive Fuzzy C-Means clustering in process monitoring. *Chemometrics and Intelligent Laboratory Systems*, **45**, 23-38.

Theodoridis, S. and Koutroumbas, K. (1999). *Pattern Recognition*. Academic Press.

Ueda, N. and Nakano, R. (1998) Deterministic annealing EM algorithm. *Neural Networks*., **11**, 271-282.

Wang, X. Z. and McGreavy C. (1998) Automatic classification for mining process operational data. *Ind. Eng. Chem. Res* ., **37**, 2215-2222.