

# On-Line Soft Sensor for Polyethylene Process with Multiple Production Grades

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:** Since the online measurement of melt index (MI) of polyethylene is difficult, a virtual sensor model is desirable. However, a polyethylene process usually produces products with multiple grades. The relations between process and quality variables are highly nonlinear. In addition, a virtual sensor model in the real plant process with many inputs has to deal with the collinearity and the time-varying issues. A new recursive algorithm, which models the multivariable, time-varying and nonlinear system, is presented. Principal component analysis (PCA) is used to eliminate the collinearity. Fuzzy c-means (FCM) and fuzzy Takagi-Sugeno (FTS) modeling are used to decompose the nonlinear system into several linear subsystems. The effectiveness of the proposed method is demonstrated using the real plant data from a polyethylene process. *Copyright © 2005 IFAC*

**Keywords:** Principal component analysis, fuzzy c-means, fuzzy Takagi-Sugeno modeling

## 1. INTRODUCTION

Generally, the product quality variables of polyethylene (PE) are the density and the melt index (MI). Kiparissides (1993) stated the density of PE inversely proportions to the reaction conditions and the MI directly proportions to the exponents of the operating conditions. Therefore, the variation of MI is more significant than density. This is consistent with the real plant experience, which maintaining the MI specification is more difficult than density. Besides, while the measuring of MI is regulated, the density is on the product specification in the most of cases. Efficiently stabilizing the variation of MI has to rely on on-line analyzer. However, the hardware sensor is not so popular in plants. In the absence of on-line analyzer, the laboratory measurements of MI are performed every 2-4 hours; it makes the regulation of operating conditions is more difficult. Therefore, it is desirable to develop a virtual sensor model that uses the process variables to predict the MI for quality control. Such a virtual sensor model must be able to describe the nonlinear relations between process variable inputs and the MI, be applicable to the different grades of PE, and be easily updated to accommodate the time varying nature of the plant.

Qin (1998) proposed a block-wise recursive PLS algorithm to update the PLS model using the new data. Essentially PLS is a linear regression model.

To introduce nonlinearity, the inputs can be transformed using the nonlinear functions. However, there is no general guideline to determine the linearization functions.

Alternatively, nonlinear relations can be approximated by a number of local linear functions. Takagi and Sugeno (1985) suggested that fuzzy systems consisting of linguistic If-Then rules can be used to divide a nonlinear system into the several linear subsystems. The rules can be extracted either using the knowledge of experts or through the cluster analysis from process data, e.g.: the fuzzy c-means method.

In this paper, the fuzzy TS modeling and the historical data are used for building a piecewise linear virtual sensor model for inferring the MI using process input variables. Before building the inferential model, PCA is applied to remove the collinearity within inputs and alleviate the complexities of the fuzzy clustering. When the data for new events cannot be explained by the PCA subspace, the subspace has to be reconstructed to cover all of events. The fuzzy rules and the inferential models have to be updated to the new PCA subspace, and then, a recursive least square algorithm and the data for new events are used to correct the regression matrices.

This paper is divided into five sections. Section 2 presents the basic theories of PCA, fuzzy c-means and fuzzy TS modeling. A new recursive algorithm for updating the fuzzy models to the new PCA subspace is proposed in section 3. In section 4, the real plant data from polyethylene process are applied to demonstrate the effectiveness of the proposed method. The conclusions are given in the last section.

## 2. BASIC THEORY

### 2.1 Principal Component Analysis

Consider the data matrix  $\mathbf{W} \in R^{m \times n}$  with  $m$  rows of observations and  $n$  columns of variables. Each column is normalized to zero mean and unit variance:

$\mathbf{X} = (\mathbf{W} - \mathbf{1}\bar{\mathbf{W}}^T)\mathbf{S}^{-1}$  where  $\bar{\mathbf{W}}$  is a mean vector,  $\mathbf{1}$  is a column vector which elements are one, and  $\mathbf{S}$  is a diagonal matrix of standard deviation. The eigenvectors ( $\mathbf{P}$ ) of the covariance matrix can be obtained from the normalized dataset. The score vectors are the projection of the data matrix  $\mathbf{X}$  to each eigenvector.

$$\mathbf{t}_i = \mathbf{X}\mathbf{p}_i, \quad i=1 \dots n \quad (1)$$

The data matrix  $\mathbf{X}$  can be decomposed as:

$$\begin{aligned} \mathbf{X} &= \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \dots + \mathbf{t}_K\mathbf{p}_K^T + \mathbf{t}_{K+1}\mathbf{p}_{K+1}^T + \dots + \mathbf{t}_n\mathbf{p}_n^T \\ &= \hat{\mathbf{X}} + \mathbf{E} \end{aligned} \quad (2)$$

with  $\hat{\mathbf{X}}$  being the projection of the data matrix  $\mathbf{X}$  onto the subspace formed by the first  $K$  eigenvectors and  $\mathbf{E}$  being the remainder of  $\mathbf{X}$  that is orthogonal to the subspace.

The statistic  $Q$  is defined in order to examine the new data can be explained by the PCA subspace or not.

$$Q = \mathbf{e}\mathbf{e}^T = (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^T = \mathbf{x}(\mathbf{I} - \mathbf{P}_K\mathbf{P}_K^T)\mathbf{x}^T \quad (3)$$

The loading vectors  $\mathbf{P}_K \in R^{m \times K}$  are the first  $K$  terms of eigenvectors of the covariance matrix. The statistic  $Q$  is a measure of approximation error of the new data with the PCA subspace. The confidence limit of  $Q$  is defined as follows (Jackson, 1991):

$$\begin{aligned} Q_\alpha &= \Theta_1 \left[ \frac{c_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \quad (4) \\ h_0 &= 1 - \frac{2\Theta_1\Theta_3}{3\Theta_2^2}, \quad \Theta_i = \sum_{j=K+1}^n \lambda_j^i, \quad i=1,2,3 \end{aligned}$$

The percentile  $\alpha$  is the probability of a type I error in hypothesis testing and  $c_\alpha$  is the integral of the normal probability density function from  $\alpha$  to  $\infty$ . Another measure of the difference between new data and the PCA subspace is the statistic  $T^2$ .

$$T^2 = \mathbf{x}\mathbf{P}_K\mathbf{\Lambda}^{-1}\mathbf{P}_K^T\mathbf{x}^T \quad (5)$$

The diagonal matrix  $\mathbf{\Lambda}$  is the first  $K$  terms of eigenvalues,  $\mathbf{\Lambda} = \text{diag}[\lambda_1 \ \lambda_2 \ \dots \ \lambda_K]$ . The  $T^2$  - confidence limit is defined as:

$$T_\alpha^2 = \frac{K(m-1)}{m-K} F_{K,m-1,\alpha} \quad (6)$$

where  $F_{K,m-1,\alpha}$  is an  $F$  distribution with degrees of freedom  $K$  and  $m-1$ . The new data belong to the PCA subspace in the  $\alpha$  confidence limit only when  $Q < Q_\alpha$  and  $T^2 < T_\alpha^2$ .

### 2.2 Fuzzy C-Means Clustering

The objective of fuzzy clustering is to partition the dataset  $\mathbf{T}$  into  $c$  clusters with vague boundaries. The fuzzy clustering algorithm is based on minimization of the cost function  $J(\mathbf{T}; \mathbf{U}, \boldsymbol{\mu})$  with respect to the degrees of membership  $\mathbf{U}$  and the cluster centers  $\boldsymbol{\mu}$  (Bezdek, 1981). The object function is defined as:

$$\begin{aligned} \min_{\mathbf{U}, \boldsymbol{\mu}} J(\mathbf{T}; \mathbf{U}, \boldsymbol{\mu}) &= \sum_{i=1}^c \sum_{j=1}^m (u_{ij})^q D_{ij,\Sigma}^2 \quad (7) \\ D_{ij,\Sigma}^2 &= (\mathbf{t}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\mathbf{t}_j - \boldsymbol{\mu}_i) \end{aligned}$$

with subject to the constraints:

$$\begin{aligned} u_{ij} &\in [0,1], \quad 1 \leq i \leq c, \quad 1 \leq j \leq m \quad (8) \\ \sum_{i=1}^c u_{ij} &= 1, \quad 1 \leq j \leq m, \quad 0 < \sum_{j=1}^m u_{ij} < m, \quad 1 \leq i \leq c \end{aligned}$$

where  $u_{ij}$  is the membership value of the  $j^{\text{th}}$  data point sharing with the  $i^{\text{th}}$  cluster, the score vectors  $\mathbf{T}$  are used in fuzzy clustering instead of the original data. The fuzzifier  $q$  determines the fuzziness of the resulting clusters, when it is closer to 1 the boundaries of the clusters are crisper. The different types of distance ( $D_{ij,\Sigma}^2$ ) can be used to measure the distance between the  $j^{\text{th}}$  observation and the  $i^{\text{th}}$  cluster center. A common choice is  $\boldsymbol{\Sigma} = \mathbf{I}$ , which induces the  $D_{ij,1}^2$  is the Euclidean norm. The cluster shapes are hyperspherical, i.e., the clusters whose surfaces of constant membership are hyperspheres. If the covariance of the dataset is chosen to measure distances, i.e.  $\boldsymbol{\Sigma} = \frac{1}{m} \sum_{j=1}^m \mathbf{t}_j^T \mathbf{t}_j$ , the  $D_{ij,\Sigma}^2$  is the

Mahalanobis norm that generates hyperellipsoidal clusters whose orientations are all the same with the data spread. Gustafson and Kessel (1979) proposed each cluster has its own norm-inducing matrix in order to reflect the clusters of different geometrical

shapes in one dataset. The object function in Eq. 7 is modified as:

$$\min_{\mathbf{U}, \boldsymbol{\mu}, \boldsymbol{\Sigma}} J(\mathbf{T}; \mathbf{U}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^c \sum_{j=1}^m (u_{ij})^q D_{ij, \boldsymbol{\Sigma}_i}^2 \quad (9)$$

Using the Lagrange multiplier method, the Gustafson-Kessel algorithm can be derived:

1. Randomly initialize the degrees of membership following the constraints in Eq. 8.
2. Compute the cluster centers and covariances:

$$\boldsymbol{\mu}_i^{(k)} = \frac{\sum_{j=1}^m [u_{ij}^{(k-1)}]^q \mathbf{t}_j}{\sum_{j=1}^m [u_{ij}^{(k-1)}]^q}, \quad i = 1 \dots c \quad (10)$$

$$\boldsymbol{\Sigma}_i^{(k)} = \frac{\sum_{j=1}^m [u_{ij}^{(k-1)}]^q (\mathbf{t}_j - \boldsymbol{\mu}_i^{(k)})^T (\mathbf{t}_j - \boldsymbol{\mu}_i^{(k)})}{\sum_{j=1}^m [u_{ij}^{(k-1)}]^q} \quad (11)$$

3. Compute the distances and update the degrees of membership:

$$D_{ij, \boldsymbol{\Sigma}_i}^{2(k)} = \left\| \boldsymbol{\Sigma}_i^{(k)} \right\|^{1/n} (\mathbf{t}_j - \boldsymbol{\mu}_i^{(k)})^T \boldsymbol{\Sigma}_i^{-1(k)} (\mathbf{t}_j - \boldsymbol{\mu}_i^{(k)}) \quad (12)$$

$$u_{ij}^{(k)} = 1 / \sum_{i=1}^c (D_{ij, \boldsymbol{\Sigma}_i}^{2(k)} / D_{ij, \boldsymbol{\Sigma}_i}^{2(k-1)})^{2/(q-1)}, \quad i = 1 \dots c, \quad j = 1 \dots m \quad (13)$$

4. If the norm of the membership changes is larger than the predefined tolerance ( $\varepsilon$ ), i.e.,

$$\sum_{j=1}^m \sum_{i=1}^c \|u_{ij}^{(k)} - u_{ij}^{(k-1)}\| \geq \varepsilon, \quad k=k+1 \text{ back to step 2.}$$

### 2.3 Fuzzy TS Modeling

Takagi and Sugeno (1985) proposed a fuzzy rule-based model, which decompose a nonlinear system into  $c$  fuzzy rules:

$$R_i: \text{ If } \mathbf{x}_j \text{ is } A_i \text{ then } \mathbf{y}_j = f_i(\mathbf{x}_j), \quad i = 1 \dots c, \quad j = 1 \dots m \quad (14)$$

where  $\mathbf{x}_j$ ,  $\mathbf{y}_j$  are the  $j^{\text{th}}$  inputs and outputs,  $A_i$  and  $f_i$  respectively are the antecedent fuzzy set and the consequence inferential function of the  $i^{\text{th}}$  rule ( $R_i$ ).

In all the rules, the structures of the inferential functions are equal and the parameters are different to approximate a nonlinear system.

In this paper, fuzzy clusters extracted from data are considered as the rules of the fuzzy TS model. For each rule, the differences between score vectors and the cluster center are considered as the inputs of the inferential function. The outputs can be evaluated from the inputs and the regression matrix of the  $i^{\text{th}}$  rule,  $\mathbf{B}_i$ .

$$R_i: \text{ If } \mathbf{t}_j \text{ is } A_i \text{ then } \mathbf{y}_{ij} = (\mathbf{t}_j - \boldsymbol{\mu}_i) \mathbf{B}_i \quad (15)$$

where  $\mathbf{y}_{ij}$  is the inference of the  $j^{\text{th}}$  outputs using the  $i^{\text{th}}$  rule. The outputs can be obtained by defuzzification.

$$\mathbf{y}_j = \sum_{i=1}^c u_{ij} \mathbf{y}_{ij} / \sum_{i=1}^c u_{ij}, \quad j = 1 \dots m \quad (16)$$

If the dimensions of the PCA subspace are determined only by the inputs, the correlations between the inputs and the outputs are not taken into account. It may cause that the number of principal components (PCs) can properly explain the inputs, but some of information about the outputs are correlated with the “noises” beyond PCA subspace are neglected. Therefore, the number of PCs is determined by cross-validation in this paper. The data are split into the training and test sets. The relative root-mean-square errors (RRMSEs) of the test set are as a function of the number of PCs retained in the regression model formed with the training set.

$$\text{RRMSE} \equiv \left[ \frac{1}{m_{\text{test}}} \sum_{i=1}^{m_{\text{test}}} \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2 \right]^{0.5} \times 100\% \quad (17)$$

The optimum number of PCs is the number of PCs that produces the minimum RRMSEs of the test set. However, this method will determine more PCs than only using the data of inputs.

## 3. UPDATE INFERENCE MODEL

The PCA subspace has to be reconstructed in order to explain the data for new events. Liu (2004) proposed a method that updates the trained clusters to the new PCA subspace; eventually only the data for new events need to be clustered on the new subspace. Since the updated clusters may be compatible with the new clusters, the similarities between clusters have to be measured. If the similarity degree is high enough, the compatible clusters need to be merged. Besides, the fuzzy TS models are also adapted for a time-varying process. The regression matrices of all rules not only are updated, but also are corrected using the recursive least square algorithm with the new data.

### 3.1 Update Fuzzy Clusters to the New Subspace

Consider the addition of data for new events to the original data matrix. The number of observations is increasing from  $m$  to  $m^*$ . The data matrix  $\mathbf{W}^{*T} = [\mathbf{W}^T \quad \mathbf{W}_{\text{new}}^T]$  is normalized as  $\mathbf{X}^* = (\mathbf{W}^* - \overline{\mathbf{1W}}^{*T}) \mathbf{S}^{*-1}$ . The mean vector and the standard deviation matrix are  $\overline{\mathbf{W}}^*$  and  $\mathbf{S}^*$ . The

loading vectors  $\mathbf{P}^*$  and the correspondingly score vectors  $\mathbf{T}^*$  can be determined by assuming  $\mathbf{E}^* \approx 0$ . Since the original data can be explained by both of the subspaces.

$$\mathbf{W} \approx \mathbf{1}\overline{\mathbf{W}}^T + \mathbf{T}\mathbf{P}^T\mathbf{S} \approx \mathbf{1}\overline{\mathbf{W}}^{*T} + \mathbf{T}^*\mathbf{P}^{*T}\mathbf{S}^* \quad (18)$$

Hence the relation between the scores in the new and old subspaces is given by:

$$\begin{aligned} \mathbf{T}^* &\approx \mathbf{T}\mathbf{A} + \mathbf{1}\Delta\overline{\mathbf{W}}^T\mathbf{S}^{*-1}\mathbf{P}^* \\ \mathbf{A} &\equiv \mathbf{P}^T\mathbf{S}\mathbf{S}^{*-1}\mathbf{P}^*, \quad \Delta\overline{\mathbf{W}} \equiv \overline{\mathbf{W}} - \overline{\mathbf{W}}^* \end{aligned} \quad (19)$$

The above equations show that the scores for the old data can be transfer to the new subspace through a linear operation with the coordinate rotation ( $\mathbf{A}$ ) and shifting ( $\Delta\overline{\mathbf{W}}^T\mathbf{S}^{*-1}\mathbf{P}^*$ ). Therefore, the center of any class  $j$  in the original subspace ( $\boldsymbol{\mu}_j$ ) can be transferred to the newer subspace:

$$\boldsymbol{\mu}_j^* \approx \boldsymbol{\mu}_j\mathbf{A} + \Delta\overline{\mathbf{W}}^T\mathbf{S}^{*-1}\mathbf{P}^* \quad (20)$$

Assuming that there are  $m_j$  observations in class  $j$ , the covariance matrix ( $\boldsymbol{\Sigma}_j$ ) on the original subspace is:

$$\boldsymbol{\Sigma}_j = \frac{1}{(m_j - 1)} \sum_{i=1}^{m_j} (\mathbf{t}_i - \boldsymbol{\mu}_j)^T (\mathbf{t}_i - \boldsymbol{\mu}_j) \quad (21)$$

And the covariance on the newer subspace is:

$$\boldsymbol{\Sigma}_j^* = \frac{1}{(m_j - 1)} \sum_{i=1}^{m_j} (\mathbf{t}_i^* - \boldsymbol{\mu}_j^*)^T (\mathbf{t}_i^* - \boldsymbol{\mu}_j^*) \quad (22)$$

By substituting eqs 19-21 into the above equation, the covariance on the new subspace can be estimated from that of the original subspace as:

$$\boldsymbol{\Sigma}_j^* \approx \mathbf{A}^T \boldsymbol{\Sigma}_j \mathbf{A} \quad (23)$$

The centers and covariances of clusters of the original data can be updated to the newer subspace using Eq. 20 and 23. Only the data for new events need to be clustered on the new subspace. After that, the degrees of membership are recalculated from Eq. 12 and 13 using all of the clusters.

### 3.2 Merge Compatible Clusters

Frigui and Krishnapuram (1996) proposed a measure of cluster similarity it needs to be used in order to merge these clusters. The measure depends on the degree of sharing of each observation among all clusters. Let  $G_i = \{\mathbf{t}_k^* \in \mathbf{T}^* | u_{ik} > 0\}$  denote the set of all score vectors belonging to the  $i^{\text{th}}$  cluster, where

$\mathbf{T}^*$  are the score vectors of all observations on the new subspace. The similarity measure between cluster  $i$  and cluster  $j$  can be defined as

$$SIM_{ij} = 1 - \frac{\sum_{\mathbf{t}_k^* \in G_i \cup G_j} |u_{ik} - u_{jk}|}{\sum_{\mathbf{t}_k^* \in G_i} u_{ik} + \sum_{\mathbf{t}_k^* \in G_j} u_{jk}} \quad (24)$$

When  $SIM_{ij} = 1$ , cluster  $i$  and cluster  $j$  are identical, on the other hand  $SIM_{ij} = 0$ , cluster  $i$  and cluster  $j$  are disjointed, i.e.,  $G_i \cap G_j = \emptyset$ . A threshold value  $SIM_{thr}$  is used to determine whether two clusters should be merged. When  $SIM_{ij} \geq SIM_{thr}$ , cluster  $i$  and cluster  $j$  are merged to a new cluster with the center and covariance as follows:

$$\boldsymbol{\mu}_i^* = \frac{\sum_{\mathbf{t}_k^* \in G_i \cup G_j} u_{ij}^q \mathbf{t}_j^*}{\sum_{\mathbf{t}_k^* \in G_i \cup G_j} u_{ij}^q} \quad (25)$$

$$\boldsymbol{\Sigma}_i^* = \frac{\sum_{\mathbf{t}_k^* \in G_i \cup G_j} u_{ij}^q (\mathbf{t}_j^* - \boldsymbol{\mu}_i^*)^T (\mathbf{t}_j^* - \boldsymbol{\mu}_i^*)}{\sum_{\mathbf{t}_k^* \in G_i \cup G_j} u_{ij}^q} \quad (26)$$

### 3.3 Update and Correct Fuzzy TS Model

There are two steps for adapting the fuzzy TS model. Firstly, the regression matrix  $\mathbf{B}_i$  of the rule  $i$  is updated to the new subspace without recalculating the trained data. After that, using the recursive least square algorithm and the addition of data whose number of observation is  $(m^*-m)$  correct the regression matrices.

Classify the new quality data  $\mathbf{Z}_{\text{new}}$  into each class, denoted  $\mathbf{Z}_i^*$  is the quality data of the  $i^{\text{th}}$  class, i.e.,  $\mathbf{Z}_i^{*T} = [\mathbf{Z}_i^T \quad \mathbf{Z}_{i,\text{new}}^T]$ ,  $i = 1 \dots c$ . Recalculate the average of quality data for each class;  $\overline{\mathbf{Z}}_i^*$ , take the differences between the measures and average as outputs of fuzzy TS model,  $\mathbf{Y}_i = \mathbf{Z}_i^* - \mathbf{1}\overline{\mathbf{Z}}_i^{*T}$ . Assuming that there are  $m_i$  observations in the  $i^{\text{th}}$  class before adding the new data, the regression function is

$$\mathbf{Y}_{i,(m_i)} = (\mathbf{T}_{i,(m_i)}^* - \mathbf{1}\boldsymbol{\mu}_i^*) \mathbf{B}_{i,(m_i)}^* + \mathbf{E}_{i,(m_i)} \quad (27)$$

where  $\mathbf{E}_{i,(m_i)}$  is the regression error. From above equation, the regression matrix  $\mathbf{B}_{i,(m_i)}^*$  can be derived.

$$\mathbf{B}_{i,(m_i)}^* = (\mathbf{C}_{i,(m_i)}^{*T} \mathbf{C}_{i,(m_i)}^*)^{-1} \mathbf{C}_{i,(m_i)}^{*T} \mathbf{Y}_{i,(m_i)}, \quad \mathbf{C}_{i,(m_i)}^* \equiv \mathbf{T}_{i,(m_i)}^* - \mathbf{1}\boldsymbol{\mu}_i^* \quad (28)$$

The inverse matrix  $(\mathbf{C}_{i,(m_i)}^{*T} \mathbf{C}_{i,(m_i)}^*)^{-1}$  can be obtained from the model in the old subspace if the number of PCs has not increased:

$$\left(\mathbf{C}_{i,(m_i)}^{*T} \mathbf{C}_{i,(m_i)}^*\right)^{-1} = \mathbf{A}^{-1} \left[ \left(\mathbf{T}_i - \mathbf{1}\boldsymbol{\mu}_i\right)_{i,(m_i)}^T \left(\mathbf{T}_i - \mathbf{1}\boldsymbol{\mu}_i\right)_{i,(m_i)} \right]^{-1} \mathbf{A}^{-1} \quad (29)$$

After updating the fuzzy TS model for the old event data to the new subspace, the regression matrices need to be corrected using the addition data. Assuming that there are  $\Delta m_i$  observations are added into the class  $i$ ,  $\Delta m_i \equiv m_i^* - m_i$ . The matrix is corrected using the new data and the recursive least square algorithm (Ljung, 1999).

$$\mathbf{B}_{i,(m_i^*)}^* = \mathbf{B}_{i,(m_i)}^* + \mathbf{D} \left( \mathbf{Y}_{(\Delta m_i)} - \left( \mathbf{T}_{(\Delta m_i)}^* - \mathbf{1}\boldsymbol{\mu}_i^* \right) \mathbf{B}_{i,(m_i)}^* \right) \quad (30)$$

$$\mathbf{D} \equiv \left( \mathbf{C}_{i,(m_i)}^{*T} \mathbf{C}_{i,(m_i)}^* \right)^{-1} \mathbf{T}_{(\Delta m_i)}^* \left( \mathbf{I} + \mathbf{T}_{(\Delta m_i)}^* \left( \mathbf{C}_{i,(m_i)}^{*T} \mathbf{C}_{i,(m_i)}^* \right)^{-1} \mathbf{T}_{(\Delta m_i)}^{*T} \right)^{-1}$$

where  $\mathbf{B}_{i,(m_i^*)}^*$  and  $\mathbf{B}_{i,(m_i)}^*$  are the regression matrices respectively with  $m_i^*$  and  $m_i$  observations.

#### 4. ILLUSTRATING EXAMPLE

The polyethylene plant used for this example is located in Kaohsiung, Taiwan. In Fig. 1, the process flow diagram has been shown; the detailed description can be found in the previous work (Liu, 2004). In that paper, the reactor feeding temperature (TIC-001) was not taken into concern for process monitoring. However, the disturbance of feeding temperature is an important factor affecting the variation of MI through the change of reactor temperature profile. Hence it is included in this study as one of the input variables.

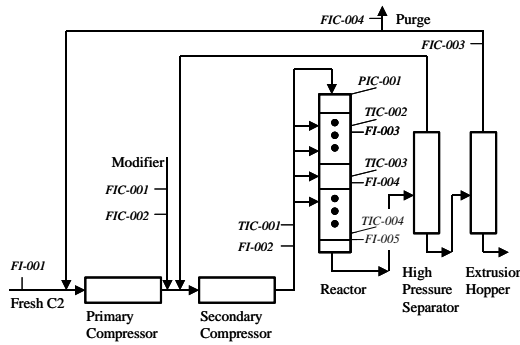


Fig. 1: Process flow diagram of high-pressure polyethylene plant.

The data of fourteen process variables in Fig. 1 and a quality variable (laboratory MI) are collected. Since the laboratory MI is measured every 4 hours, the data collection period are three months in order to collect enough data to reasonably represent the process behavior. There are 512 observations in the dataset for the first time modeling. Using cross-validation in section 2.3, the optimal number of PCs is found to be eight. There are 485 observations can be explained by the PCA subspace with 99% confidence limits. The clustering result of projection of first two score vectors is shown in Fig. 2. The data of inputs can be clustered into three groups; the solid lines represent

the limits of Mahalanobis distances equal to 1, in Eq. 12. In that period there are three different production grades labeled as #1, #2 and #3 in Fig. 2. The comparison of the regression results using the proposed method and the PLS model is shown in Fig. 3. For confidentiality, the values of logarithm of MI are presented in arbitrary units. The PLS results using a single linear model are also presented in Fig. 3. It is obvious that only one set of parameters of the linear model cannot properly explain outputs with three different grades of products.

Before on-line predicting the MI, it must be verified that process variables are in the control limits, i.e. the inputs are on the current PCA subspace. If either the statistic  $Q$  or  $T^2$  are out of the control limits, the operators will be informed to elucidate the operating condition (Liu, 2004). In practices, the inferring model is periodically updated and corrected by the proposed method in order to cover all of events on the PCA subspace. The additional 119 observations in one month following the period of the above training data were used for testing. The 100 observations of the dataset can be explained by the PCA subspace with 99% confidence limits. In Fig. 4, the model predictions of the proposed method and the PLS model *without coefficient updating* are shown. It is found that predicted MI using PLS, are higher than measured values for the products with the lowest MI, but lower for products with the other two MI specifications. For the product with the highest MI specifications, the PLS model errors become very large because the MI is proportional to the exponents of the operating conditions. Hence, the errors are exponentially amplified. In this test dataset, the RRMSEs of FTS and PLS are 12.58% and 38.38%.

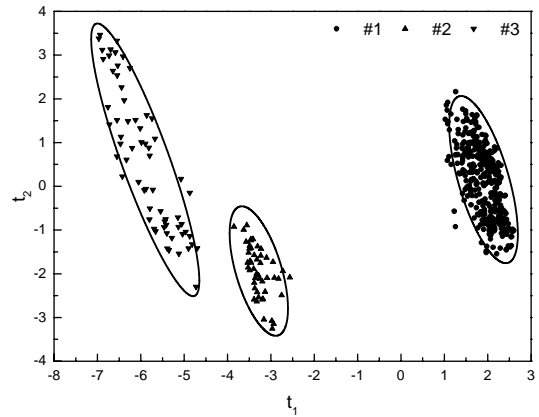


Fig. 2: Clustering results for grade #1, #2 and #3 in the PCA subspace.

A second test dataset is collected after first one; the period is also one month; there are 122 observations in the dataset, 110 observations among data can be explained by the PCA subspace, reconstructed with the data for new events, in 99% confidence limits. The FTS model is updated to the new PCA subspace and corrected using the previous test dataset. A new PLS model is also generated using the block-wise recursive PLS. The prediction results are shown in Fig. 5. The RRMSEs of the proposed method are

5.45% which is significantly better than that of the RPLS, 28.64%.

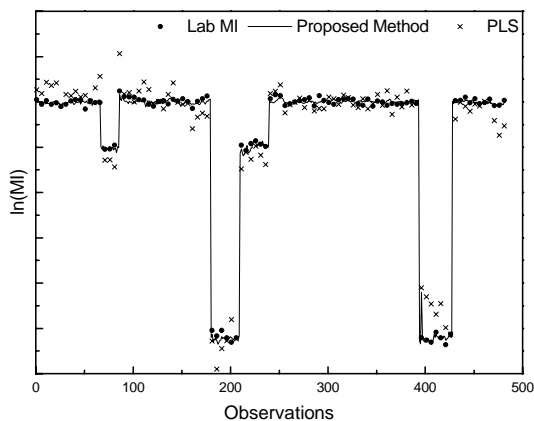


Fig. 3: Comparison of the regression results using proposed method and PLS.

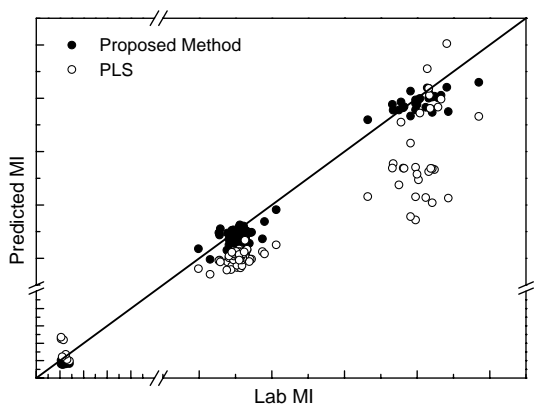


Fig. 4: Comparison of MI from laboratory with predictions using proposed method and PLS.

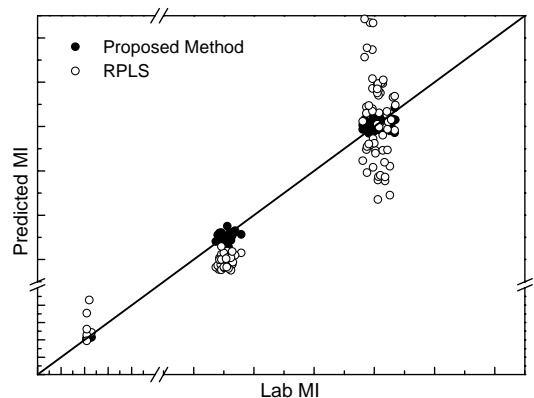


Fig. 5: Comparison of MI from laboratory with predictions using proposed method and RPLS.

## 5. CONCLUSIONS

A real plant process, the quality variables are affected by a number of process variables with collinearity. PCA can be used to eliminate the collinearity and reduce the system dimensions. Besides, the nonlinearity exists between inputs and outputs especially when there are multiple production grades. In this paper, FCM is used to decompose the operating space into several regions. Fuzzy TS model is used to build a local linear model for each region. In addition to nonlinearity, the

input/output relations may also be time-varying. When the data for new events cannot be explained by the current PCA subspace, the subspace has to be reconstructed. The classification and regression coefficients have to be updated as well. In this paper, a recursive algorithm for updating parameters of all the linear models is proposed. The proposed method is applied to the MI prediction of a polyethylene plant. Results show that nonlinearity and time varying characteristics of the polyethylene process plant can be dealt with effectively.

## ACKNOWLEDGMENT

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

## REFERENCES

- Kiparissides, C., Verros, G. and MacGregor, J. F. (1993), Mathematical modeling, optimization, and quality control of high-pressure polymerization reactors. *J.M.S-Rev. Macromol. Chem. Phys.*, **C33(4)**, 437-527.
- Qin, S. J. (1998), Recursive PLS algorithms for adaptive data modeling. *Comp. Chem. Engng.*, **22**, 503-514.
- Takagi, T. and Sugeno, M. (1985), Fuzzy identification of systems and its application to modeling and control. *IEEE Trans. Systems, Man and Cybernetics*, **15**, 116-132.
- Jackson, J. E. *A User's Guide to Principal Components*, Wiley, New York, 1991.
- Bezdek, J. *Pattern Recognition with Fuzzy Object Function*, Plenum Press, New York, 1981.
- Gustafson, D and Kessel, W. (1979), Fuzzy clustering with a fuzzy covariance matrix. *In Proc. IEEE CDC*, San Diego, CA, USA, 761-766.
- Liu, J. (2004), Process Monitoring Using Bayesian Classification on PCA Subspace, *Ind. Eng. Chem. Res.*, **43**, 7815-7825.
- Frigui, H. and Krishnapuram, R. (1996), A robust algorithm for automatic extraction of an unknown number of clusters from noisy data. *Pattern Recognition Letters*, **17**, 1223-1232.
- Ljung, L. *System Identification Theory For The User*, Prentice-Hall Inc., New Jersey, 1999.