MPCA FOR MONITORING EMULSION POLYMERIZATION PROCESS: ALTERNATIVE STRATEGIES FOR DECOMPOSING THREE-WAY DATA MATRICES

Carlos R. Alvarez, Adriana Brandolin and Mabel C. Sánchez

Planta Piloto de Ingeniería Química (CONICET – UNS
Camino La Carrindanga km 7, Bahía Blanca, 8000, Argentina
e-mail: {ralvarez-abrandolin-msanchez}@plapiqui.edu.ar

Introduction

Batch manufacturing processes are common in chemical, pharmaceutical, bio-technical and semiconductors industries. After charging the equipment with raw materials, the operation is initiated and, the observation of the first point is obtained. This corresponds to a vector of dimension *J*. The evolution of the batch is then registered measuring the same *J* variables at time intervals 2, 3,…, until *K*, when the operation is finished. Hence the information of *I* batch runs can be grouped in a three way data matrix $\underline{\mathbf{X}}$ (batch×variables×time).

Multivariate Statistical Process Control has been successfully applied for the monitoring and diagnostic of batch process during the last decade (Nomikos and Mc Gregor (1994, 1995), Wold *et al.* (1998)). These applications are based on Multiway Principal Component Analysis (MPCA) and Multiway Partial Least Square (MPLS) strategies proposed by Wold *et al.* (1987).

The unfolding method of the three-way data matrix $\underline{\mathbf{X}}$ (batch×variable×time) plays an important role in the required effort to develop the control charts, to process data on line during monitoring and to identify the source of faults. Generally $\underline{\mathbf{X}}$ is unfolded into a large two dimensional matrix $\mathbf{X}$, such that, each vertical time-slide of $\underline{\mathbf{X}}$ is put side by side to the right in $\mathbf{X}$, starting with the slide corresponding to the first time interval. Another arrangement has been also proposed by Wold *et al.* (1998) that consist in putting each vertical time-slide of $\underline{\mathbf{X}}$ under the previous one.

In the first unfolding strategy the whole batch is considered as one object. Thus each batch can be compared against a group of good batches to determine if it is a good batch or not. Since the mean trajectories of all process variables are removed, and consequently the main nonlinear and dynamic component of the data are not present any more, a PCA allows to study the systematic variation of variable trajectories about their mean trajectories.

In contrast, the approach developed by Wold *et al.* (1998) for the vertical unfolding only removes the grand mean of the variables for all batches and times, leaving the non-linear time-varying trajectories in the data. To avoid the capture of the deterministic behaviour of the process by the first principal components, Yoo *et al.* (2004) uses the vertical unfolding after centering and scaling the horizontal matrix $\mathbf{X}$. In this way the information regarding process variability for each time is maintained.

This work presents a comparative of analysis of performance between PCA techniques based on the horizontal unfolding and the vertical unfolding proposed by Yoo *et al*. (2004) for the modeling, on-line monitoring and fault identification stages. The study is carried out for a methyl-methacrylate emulsion polymerization reactor.

Multiway Principal Component Analysis: Horizontal Unfolding

Let us first consider the unfolding of the three-way data matrix $\underline{\mathbf{X}}$ ($I \times J \times K$) as it is shown in Fig. 1A. A matrix $\mathbf{X_1}$ of dimension (I x KJ) results that, prior to obtain an empirical model based on PCA, is mean centered and scaled forming matrix $\mathbf{Z_1}$. The subtraction of the mean trajectories of each variable removes the main nonlinear and dynamic components of the data, allowing to analyze the common cause variations in the time trajectories of all variables with respect to their mean trajectories. The scaling of variables to unit variance gives equal weight to all variables at each time interval.

Modeling

The matrix $\mathbf{Z_1}$ is decomposed into a summation of R products of score vectors $\mathbf{t_r}$ and loading vectors $\mathbf{p_r}$, plus a residual matrix $\mathbf{E_1}$ which is as small as possible in the least square sense

$$\mathbf{Z_1} = \sum_{r=1}^{R} \mathbf{t_r p_r^T} + \mathbf{E}_1 = \mathbf{T_1 P_1^T} + \mathbf{E}_1 \tag{1}$$

When the variables are highly correlated, a few principal components (P.C.) are used to express most of the variability of the data revealing similarities and differences among batches. Different criteria are established to determine the number of P.C. needed to represent the data in the latent variable space, for example, see the contributions by Jackson (1991), Nomikos and MacGregor (1995), etc.

Given a set of industrial data corresponding to I batches, first it is necessary to obtain a reference distribution against which future runs can be compared. This reference distribution should only include all batches that are subject to a common cause variation. If some batches from the original set reveal problematic operations or unacceptable products, that should be detected and alarm in the future, they are excluded from the original data set. To perform this selection the Hotelling D^2 and Q statistic are calculated for each batch. A run is taken out from the reference population if one or both statistic values are greater than the critical ones (Nomikos and MacGregor (1995)). The procedure is repeated until all runs described the normal batch operation. A population of only I' batches results and the PCA model is composed of the matrices $\mathbf{T_1'}$, $\mathbf{P_1'}$ and $\mathbf{E_1'}$ corresponding to the last post-analysis of batches. For the sake of simplicity, the reference population is considered constituted by I normal batches in the rest of this work.

The D^2 statistic measures the Mahalanobis distance of each batch with respect to the mean trajectory. Assuming the variables in matrix $\mathbf{Z_1}$ follow a multinormal distribution with zero mean vector and convariance

matrix $\mathbf{R_1}$, the $\mathbf{t}$ scores of all principal components also follow a multinormal distribution with zero mean vector and covariance matrix $\mathbf{S}$, which is diagonal due to the orthogonality of the scores. The D^2 statistic for each batch is calculated as follows

$$\boldsymbol{D_s} = \mathbf{t_R}\mathbf{S_R^{-1}}\mathbf{t_R^T}\left(\frac{\boldsymbol{I}}{(\boldsymbol{I}-1)^2}\right) \approx \boldsymbol{B}_{\boldsymbol{R}/2,(\boldsymbol{I}-\boldsymbol{R}-1)/2,a} \tag{2}$$

where $\mathbf{t_R}$ is the vector containing the coordinates of the batch in the reduce space formed by the R retained P.C. and, $\mathbf{S_R}$ is the corresponding (RxR) diagonal covariance matrix. The statistic follows a Beta distribution $B_{R/2,(I-R-1)/2,a}$ which can be approximated as

$$\boldsymbol{B}_{\boldsymbol{R}/2,(\boldsymbol{I}-\boldsymbol{R}-1)/2,a} \cong \frac{\dfrac{\boldsymbol{R}}{\boldsymbol{I}-\boldsymbol{R}-1}\boldsymbol{F}_{\boldsymbol{R},\boldsymbol{I}-\boldsymbol{R},a}}{1+\dfrac{\boldsymbol{R}}{\boldsymbol{I}-\boldsymbol{R}-1}\boldsymbol{F}_{\boldsymbol{R},\boldsymbol{I}-\boldsymbol{R},a}} \tag{3}$$

where $F_{R,(I-R),\,a}$ is the corresponding value of the F distribution for a certain level of significance α.

The Q statistic measures the representation error of the model defined in the subspace formed by the R latent variables and, it is calculated as the square prediction error (SPE)

$$\boldsymbol{Q} = \boldsymbol{e}^T\boldsymbol{e} \tag{4}$$

The critical value for Q is

$$\begin{aligned} \boldsymbol{Q}_a &= q_1\left[1-\frac{q_2\boldsymbol{h}_0(1-\boldsymbol{h}_0)}{q_1^{\,2}}+\frac{\boldsymbol{z}_a(2q_2\boldsymbol{h}_0^{\,2})^{1/2}}{q_1}\right]^{1/\boldsymbol{h}_0} \\ q_1 = \sum_{\boldsymbol{j}=\boldsymbol{R}+1}^{\boldsymbol{J}} l_j;\quad q_2 &= \sum_{\boldsymbol{j}=\boldsymbol{R}+1}^{\boldsymbol{J}} l_j^{\,2};\quad q_3 = \sum_{\boldsymbol{j}=\boldsymbol{R}+1}^{\boldsymbol{J}} l_j^{\,3};\quad \boldsymbol{h}_0 = 1-\frac{2q_1q_3}{3q_2^{\,2}} \end{aligned} \tag{5}$$

where z_a is the normal deviate cutting of an area of α under the upper tail of the distribution if h_0 is positive and, under the lower tail if h_0 is negative. Also λ_j stands for the j-th eigenvalue of matrix $\mathbf{Z_1}$.

Also the modeling stage comprises the procedure to define the control charts for the square prediction error statistic at time k (SPE_k), that is explained in the following section because it is related with the strategy selected for on-line monitoring.

On-Line Monitoring

The on-line monitoring of new batches is performed calculating the D^2 and SPE statistics for each k time. The first statistic is evaluated as follows

$$\boldsymbol{D}_k^2 = \left(\mathbf{t}_{\boldsymbol{Rk}} - \bar{\mathbf{t}}_{\mathbf{R}k}\right)\mathbf{S}_{\mathbf{R}k}^{-1}\left(\mathbf{t}_{\boldsymbol{Rk}} - \bar{\mathbf{t}}_{\mathbf{R}k}\right)' \approx \left(\frac{\boldsymbol{R}\left(\boldsymbol{I}^2 - 1\right)}{\boldsymbol{I}\left(\boldsymbol{I} - \boldsymbol{R}\right)}\right)\boldsymbol{F}_{\boldsymbol{R},\boldsymbol{I}-\boldsymbol{R},\boldsymbol{a}} \tag{6}$$

where $\mathbf{t}_{\boldsymbol{Rk}}$ is the score vector for the observations available up to the k-th time period of the batch under analysis in the R latent space of variables, $\bar{t}_{\mathbf{R}k}$ stands for the mean value of the same score vector in the reference population and, $\mathbf{S}_{\mathbf{R}k}$ represents the covariance matrix of these vectors

$$\mathbf{S}_{\boldsymbol{Rk}} = \frac{\mathbf{T}_1^{\mathrm{T}}\mathbf{T}_1}{(\boldsymbol{I}-1)} \tag{7}$$

The square quadratic error for the k-th observation is

$$\boldsymbol{SPE}_{\boldsymbol{k}} = \sum_{\boldsymbol{c}=(\boldsymbol{k}-1)\boldsymbol{J}+1}^{\boldsymbol{kJ}} \left(\boldsymbol{e}^{\boldsymbol{new}}(\boldsymbol{c})\right)^2 \approx (\boldsymbol{v}_{\boldsymbol{k}} / 2\boldsymbol{m}_{\boldsymbol{k}})\, c^{\,2}_{2\boldsymbol{m}_{\boldsymbol{k}}^2/\boldsymbol{v}_{\boldsymbol{k}},\boldsymbol{a}} \tag{8}$$

where $\mathbf{e}_{\boldsymbol{k}}^{\boldsymbol{new}}$ is the model representation error vector defined as follows

(9)

and

(10)

To calculate $\mathbf{t}_{Rk}^{new}$, a vector of complete standardized measurements $\mathbf{z}_{\mathbf{1}\boldsymbol{k}}^{\boldsymbol{new}}$ of dimension (1xKJ) should be available but, at time k only the first KJ columns are known, thus the remaining (K-k)J measurements are estimated using different techniques (Nomikos and MacGregor, (1995)).

The parameters m_k and v_k are the mean and variance of SPE_k, which are obtained during the model stage. Each normal run of the reference distribution is passed through a procedure to estimate future observations and, K groups of I batches containing JK variables are formed (Box *et al*. (1978)). These data are projected into the space of the latent variables and the parameters m_k and v_k are calculated as follows

$$\boldsymbol{m}_{\boldsymbol{k}} = \frac{\sum_{\boldsymbol{i}=1}^{\boldsymbol{I}} \sum_{\boldsymbol{c}=(\boldsymbol{k}-1)\boldsymbol{J}+1}^{\boldsymbol{kJ}} \boldsymbol{e}_{\boldsymbol{i},\boldsymbol{k}}(\boldsymbol{c})^2}{\boldsymbol{I}} \quad , \quad \boldsymbol{v}_{\boldsymbol{k}}{}^2 = \frac{\sum_{\boldsymbol{i}=1}^{\boldsymbol{I}} \left(\sum_{\boldsymbol{c}=(\boldsymbol{k}-1)\boldsymbol{J}+1}^{\boldsymbol{kJ}} \boldsymbol{e}_{\boldsymbol{i},\boldsymbol{k}}(\boldsymbol{c})^2 - \boldsymbol{m} \right)^2}{\boldsymbol{I}-1} \tag{11}$$

where $\mathbf{e}_{i,k}$ stands for the representation error of batch i for the time interval k.

Fault Identification

Once a special event has been detected, it is important to diagnose its cause. For example, the contribution of each measured variable to both statistics can be displayed on-line to diagnose the cause of an abnormal operation. These contributions to the D^2 and SPE statistics are calculated using Eq. (12) and Eq. (13) respectively (Westerhuis *et al.*, 2000)

(12)

(13)

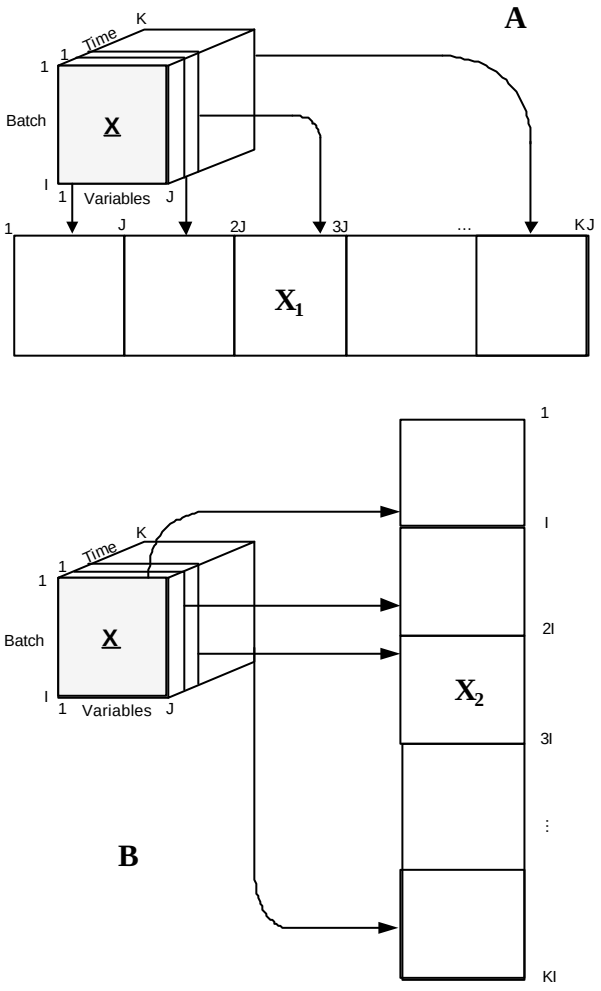


Fig. 1: Horizontal (A) and Vertical (B) Unfolding of the Three-Way Data Matrix $\underline{X}$

Multiway Principal Component Analysis: Vertical Unfolding

Let us consider the unfolding of matrix $\underline{\mathbf{X}}$ ($I \times J \times K$) as it is shown in Fig. 1B. A matrix $\mathbf{X_2}$ of dimension (I, KxJ) results. If $\mathbf{X_2}$ is centered (Wold *et al.* (1998)), only the grand mean of the variables for all batches and times are removed, leaving the non-linear time-varying trajectories in the data. In order to only maintain the information regarding process variability for each time, Yoo *et al.* (2004) used the vertical unfolding after centering and scaling the horizontal matrix $\mathbf{X_1}$. Thus a PCA is undertaken on matrix $\mathbf{Z_2}$(IKxJ). It should be noticed that matrix $\mathbf{X_1}$ is obtained after performing a post-analysis of the original run data to discard those that are not considered normal. This selection can only be done using the horizontal unfolding.

Modeling

The matrix $\mathbf{Z_2}$ is decomposed into a summation of R_2 products of score vectors $\mathbf{t_r}$ and loading vectors $\mathbf{p_r}$ plus a residual matrix $\mathbf{E_2}$

$$\boldsymbol{Z}_2 = \sum_{r=1}^{R_2} \boldsymbol{t}_r \boldsymbol{p}_r^T + \boldsymbol{E}_2 = \boldsymbol{T}_2 \boldsymbol{P}_2^T + \boldsymbol{E}_2 \tag{13}$$

These matrices have the following dimensions $\mathbf{T_2}$(IKxR_2), $\mathbf{P_2}$(Jx R_2) and $\mathbf{E_2}$(IKxJ).

The vertical unfolding allows a direct projection of the J observations for each interval k into the space of latent variables. Consequently no inference about future measurements is necessary to define the *SPE* control charts. The external reference distribution is based on the information contained in $\mathbf{E_2}$.The statistic SPE_k follows a $(\nu_k / 2m_k)c^2_{2m_k^2/\nu_k, a}$ distribution, where

$$m_k = \frac{\sum_{i=1}^{I}\sum_{c=1}^{J} \mathbf{e}_{i,k}(c)^2}{I} \quad , \quad \nu_k^2 = \frac{\sum_{i=1}^{I}\left(\sum_{c=1}^{J} \mathbf{e}_{i,k}(c)^2 - m\right)^2}{I-1} \tag{14}$$

$e_{i,k}(c)$ stands for the representation error of variable c for batch i at time k.

On-Line Monitoring

A vector of observations $\boldsymbol{z}_{2k}^{new}$ is straightforwardly projected into de latent variable space and, the statistics $\boldsymbol{D}_k^2$ and SPE_k are calculated. The vectors of scores and model prediction error for time k are

$$\mathbf{t}_k^{new} = \mathbf{z}_{2k}^{new} \mathbf{P}_2 \tag{15}$$

$$\mathbf{e}_k^{new} = \mathbf{z}_{2k}^{new} - \mathbf{t}_k^{new} \mathbf{P}_2^T \tag{16}$$

The SPE_k statistic for the new observation is

$$SPE_k = \sum_{c=1}^{J} e_k^{new}(c)^2 \tag{17}$$

Fault Identification

If the process is out of control, an identification stage of the fault source continues. For the vertical unfolding the contribution of each measurement to the statistics can also be calculated as in Westerhuis *et al.* (2000). Furthermore a graphic technique, the Biplots, can be applied because only *J* variables are taken into account and not (*K*x*J*) variables, as it is the case for the horizontal unfolding. The use of biplots enhances significantly the identification procedure. Although MPCA has been widely applied for monitoring batch processes, the use of accompanying biplots has not received similar attention.

Let us consider the following decomposition for matrix $\mathbf{Z_2}$

$$\boldsymbol{Z}_2 = \mathbf{GH'} = \begin{bmatrix} \mathbf{g}_1' \\ \mathbf{g}_2' \\ \vdots \\ \mathbf{g}_{IK}' \end{bmatrix} \begin{bmatrix} \mathbf{h}_1' & \mathbf{h}_2' & \cdots & \mathbf{h}_J' \end{bmatrix} \tag{18}$$

where $\mathbf{G}=\mathbf{U}\Gamma^{1/2}$ and $\mathbf{H}=\mathbf{V}\Gamma^{1/2}$. Matrices **U,** Γ and **V** come from the singular value decomposition of $\mathbf{Z_2}$ as $\mathbf{Z_2}=\mathbf{U}\ \Gamma\ \mathbf{V'}$. Furthermore the eigenvalue (Λ)and eigenvector ($\mathbf{P_2}$) matrices of $\mathbf{Z_2}$ are related with **U,** Γ and **V** as follows: $\Gamma=1/(IK-1)\Lambda^{1/2}$, $\mathbf{V}=\mathbf{P_2}$, $\mathbf{U}=1/(I-1)\mathbf{T}\Lambda^{-1/2}$? The projection of the i-th row of **G** on the j-th column of **H** represents z_{2ij}.

The biplot was introduced by Gabriel (1971) as a graphical display. Because the row and column vectors ($\mathbf{g_i}$, $\mathbf{h_j}$) are of dimension *J*, only 2 or 3 components are represented in the plane or space.

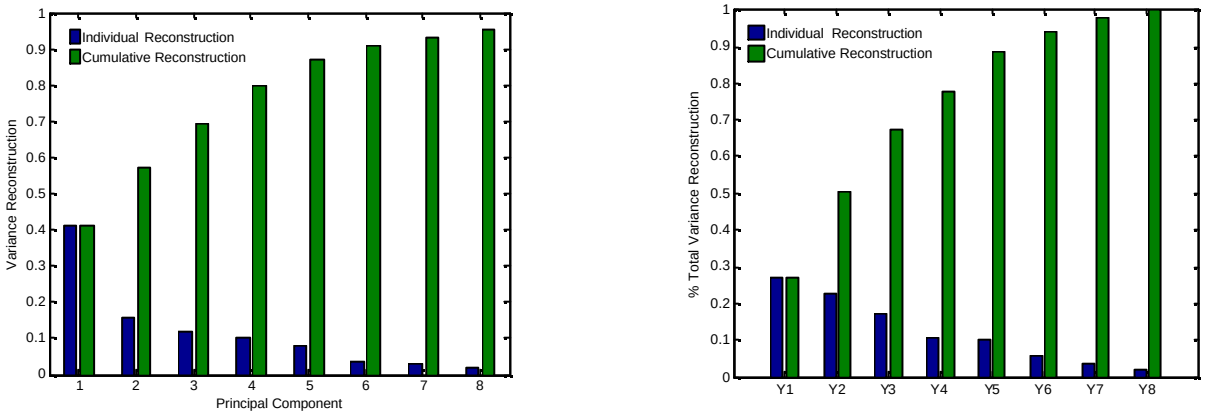
Results and Discussion

A non-isothermal semi-batch reactor model for methyl-methacrylate emulsion polymerization is developed (Alvarez *et al.*, 2006). The model includes the following equations: a) mass balances for initiator, surfactant, monomer, and radical and polymer molecules; b) population balances; c) energy balance; d) expressions to calculate: the average radical number per particle, radical entry into the particles, radical entry into micelles, radical desorption from particles, monomer conversion, monomer concentration in particles, monomer concentration in the aqueous phase, particle growth rate, total reactor volume, etc. The model also includes molecular weight calculations. gPROMS code environment (Process System Enterprise, Ltd.) is employed for modeling this process. Simulation results are validated using experimental data provided in the literature.

A reference normal data-base composed by 61 batches is obtained by simulation. It is assumed that eight measurements (molecular weight MW, monomer inlet flowrate Q0, surfactant concentration in aqueous phase SW, reactor temperature T, reactor-jacket temperature TREF, refrigerant inlent temperature TREF0, volume VR and conversion X) are sampled every 2 minutes for an 80 minutes run. These data were used to formulate the MPCA models that characterize the normal operation and to develop the control charts for each unfolding strategy.

Furthermore four operational faults are simulated: 1) an increment of monomer concentration after k=6 until the end of the batch (MB4), 2) an increment of the inlet flowrate of monomer, from k=8 until the end of the run (MB5), 3) a decrease of the monomer inlet flowrate between k=10 and k=21 (MB6), 4) a decrease of the refrigerant flowrate between k=4 and k=13 (MB7).

The percentage of total and individual variance reconstruction is shown in Fig. 2 for the horizontal and vertical unfoldings. Four P.C. are retained to get a 80% of total variance reconstruction in both cases.



Horizontal Unfolding Vertical Unfolding

Fig. 2: Percentage of variable reconstruction

Regarding the horizontal unfolding, in Fig. 3 the time evolution of the values for the statistics D^2 and *SPE* and their corresponding limits are shown for each batch. The SPE statistic detects all the faults with some delay for Batch MB4 and MB7, but the D^2 statistic only reveals the faults of Batch MB5 and MB6. For these two runs, the variable contribution plots to the D^2 statistic represented in Fig. 4 identify the right fault source.

Figure 5 shows the contributions of each measurement to the SPE statistic when the horizontal unfolding is considered. For Batch MB4, the fault occurs for an unmeasured variable, thus there is no a clear distinction of the fault source. The highest contribution corresponds to the aqueous concentration of surfactant that is in agreement with an increment of the inlet concentration of monomer. For Batches MB5 and MB6, the changes of the inlet flowrate of monomer are identified correctly. With respect to Batch MB7, the reduction of the unmeasured refrigerant flowrate is associated with the reactor temperature and conversion.

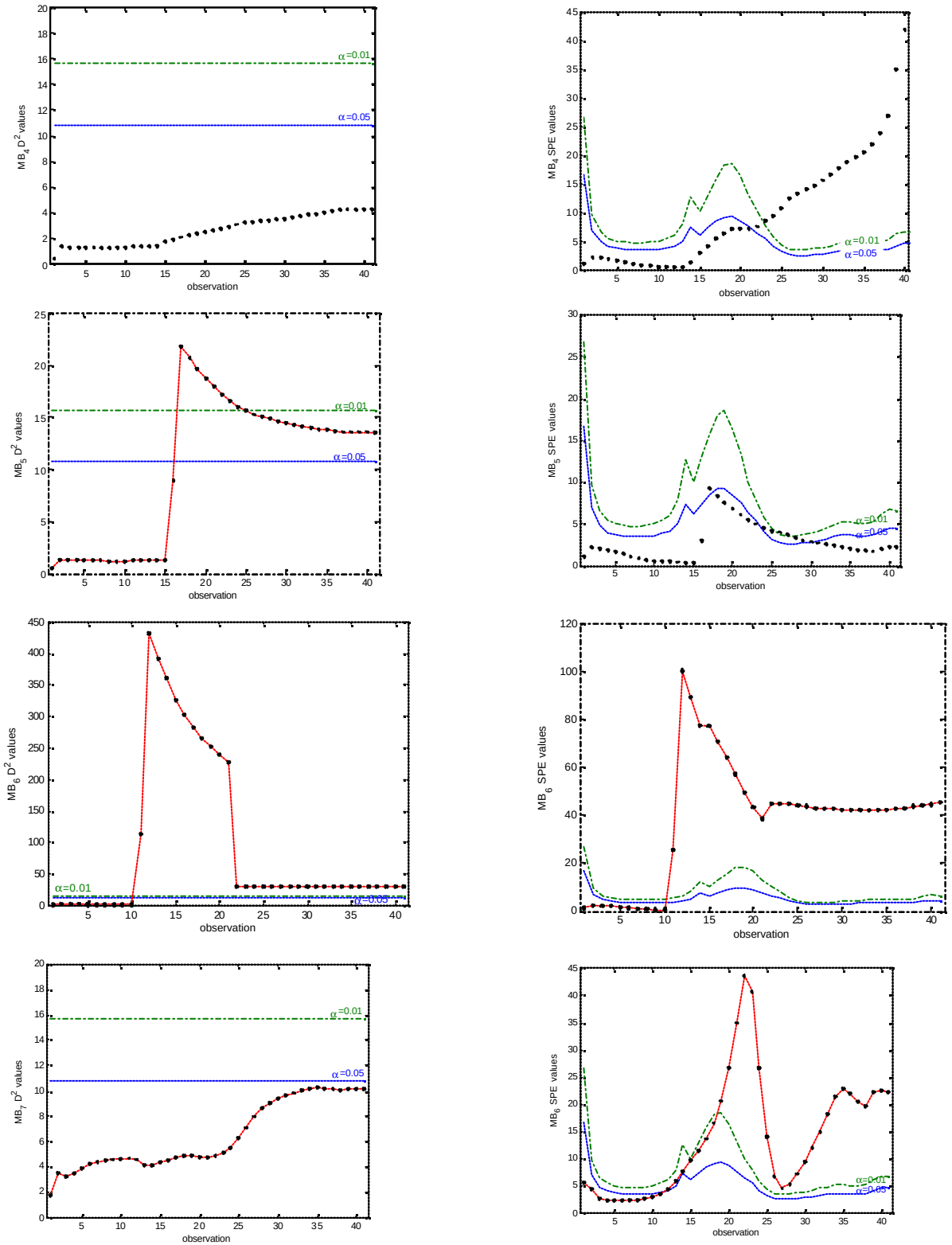


Fig 3: D^2 statistics for Batches MB 4:7 (left) - *SPE* statistics for Batches MB4:7 (right)

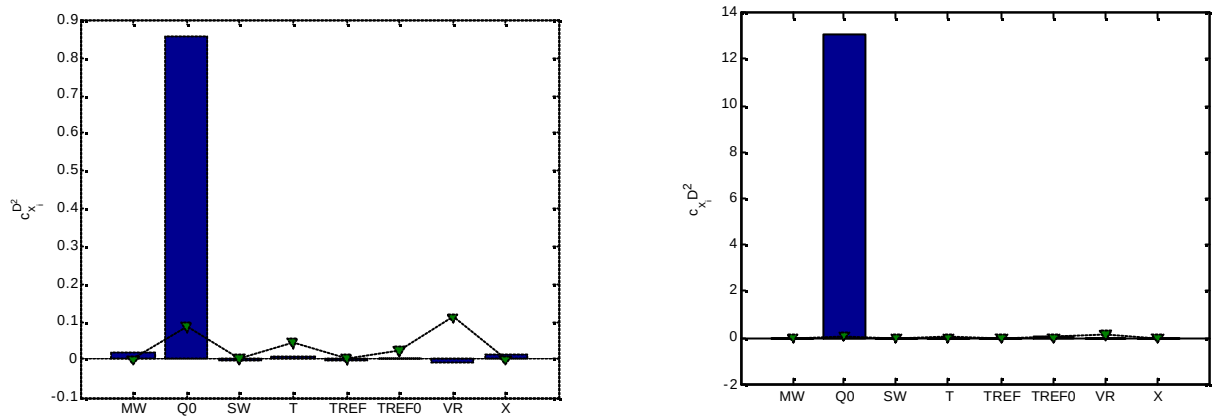


Fig 4: Contribution of each measurement to D^2 statistic (left: Batch MB5 k=18; right: Batch MB6 k=13)

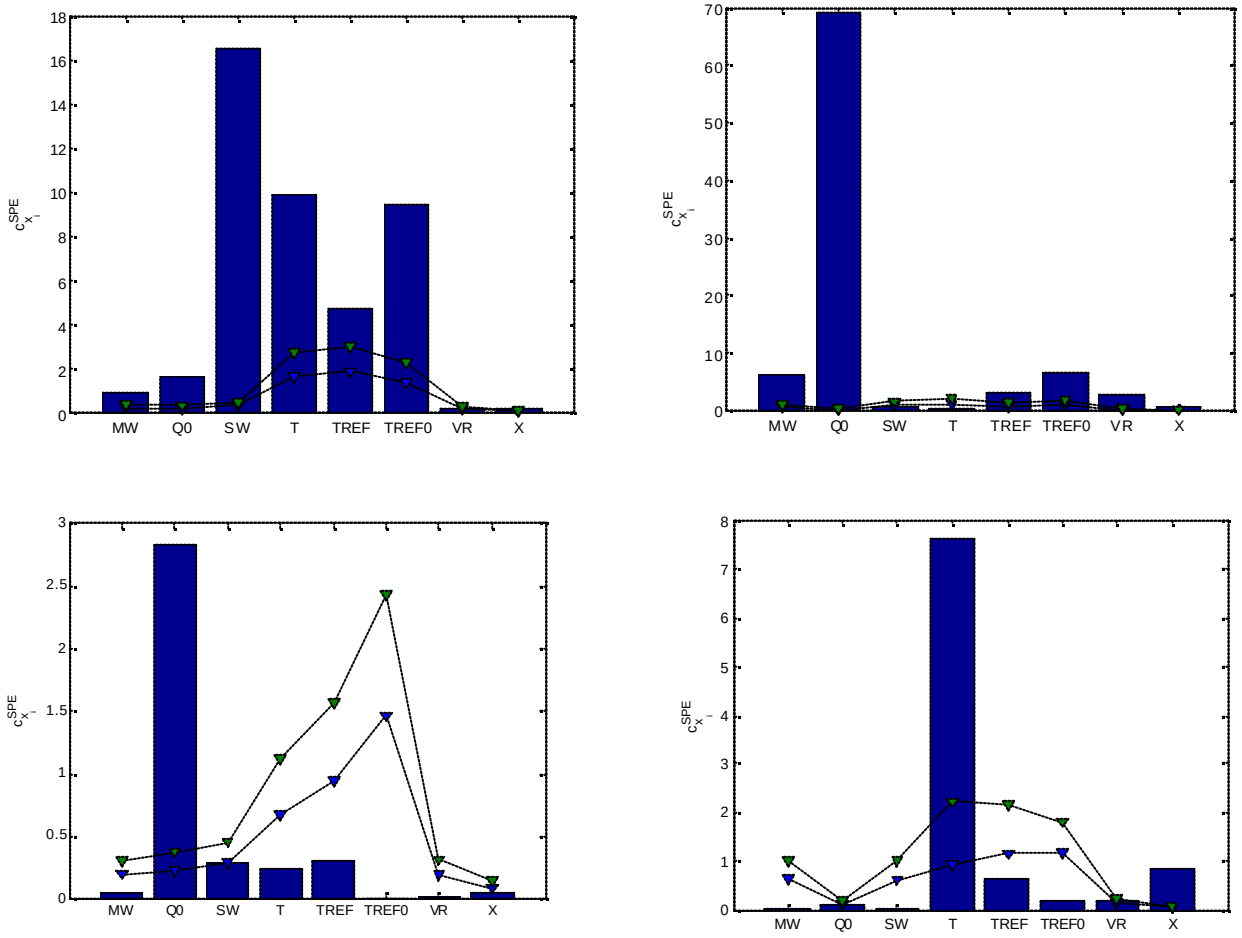


Fig 5: Contribution of each measurement to SPE statistic
(Left: Batch MB4 k= 24, Batch MB5 k=26; Right: Batch MB6 k=13, Batch MB7 k=15)

Regarding the vertical unfolding, Fig. 6 represents the time evolution of the values of statistics D^2 and *SPE* and their corresponding limits for each batch. For this unfolding, both statistics detect all faults. For Batch MB4, D^2 produces an alarm with a delay greater than *SPE* does, which detects the fault faster than the same statistic for horizontal unfolding. For Batch MB5 and MB6, the detection capabilities of both strategies are the same. For Batch MB7, modeling techniques have the same detection performance as the one shown for Batch MB4.

The contribution of each measurement to the statistic values is shown in Fig. 7 for the vertical unfolding technique. For Batch MB4, the fault occurs for an unmeasured variable. There is no a clear distinction of the fault source considering the contributions to D^2, but only the contributions of surfactant concentration and conversion are the highest for the *SPE* statistic, which is in agreement with an increment of the inlet concentration of monomer. For Batch MB5 and MB6, both types of unfolding provide the same identification using the contributions to D^2 statistic, but a clear distinction is not provide when the contributions to SPE are analyzed. For Batch MB7, the contributions to *SPE* statistic calculated with the vertical unfolding allow identifying that the fault is associated with the refrigerant, because the greatest contributions correspond to reactor temperature and reactor-jacket temperature.

Figure 8 represents the fraction of reconstruction of each measurement when they are represented in different score planes that come from the vertical unfolding model. A reconstruction fraction equal or grater than 0.5 is considered satisfactory. This information is applied to analyze the corresponding biplots, which are shown in Fig. 9. In these pictures the ellipses for α=0.01 are plotted. For two cases, the ellipse for α=0.05 is also included.

The information provided by the contribution plots (Fig. 7) regarding the fault source is enhanced with the analysis of biplots. They contain a point that represents the batch in the plane defined by two scores and one arrow for each variable. The heavy arrows correspond to variables with a high fraction of reconstruction in the plane. For Batch MB5, both the volume and inlet monomer flowrate may be considered as suspicious faults when the plane t_2-t_3 is analyzed, nevertheless the inspection of plane t_4-t_3 confirms that the fault is due to Q0. For Batch MB6, the same conclusion arises from the analysis of the batch projections on the arrows corresponding to Q0 and VR. For Batch MB7, the three biplots indicate the faults correspond to the reactor and reactor-jacket temperatures, that are associated with the refrigerant flowrate, which is the variable out of control.

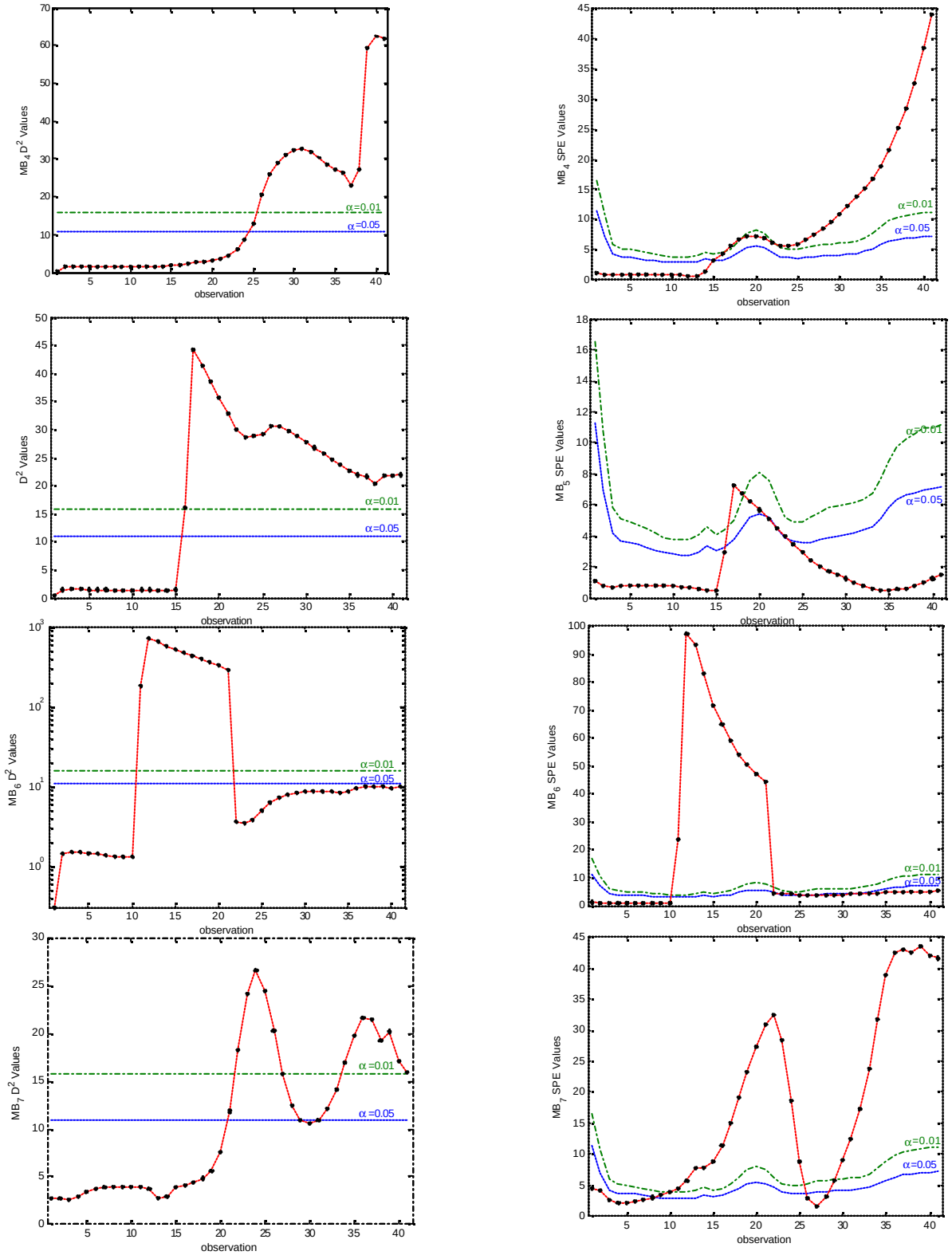


Fig 6: D^2 statistics for Batches MB4:7 (left) - *SPE* statistics for Batches MB4:7 (right)

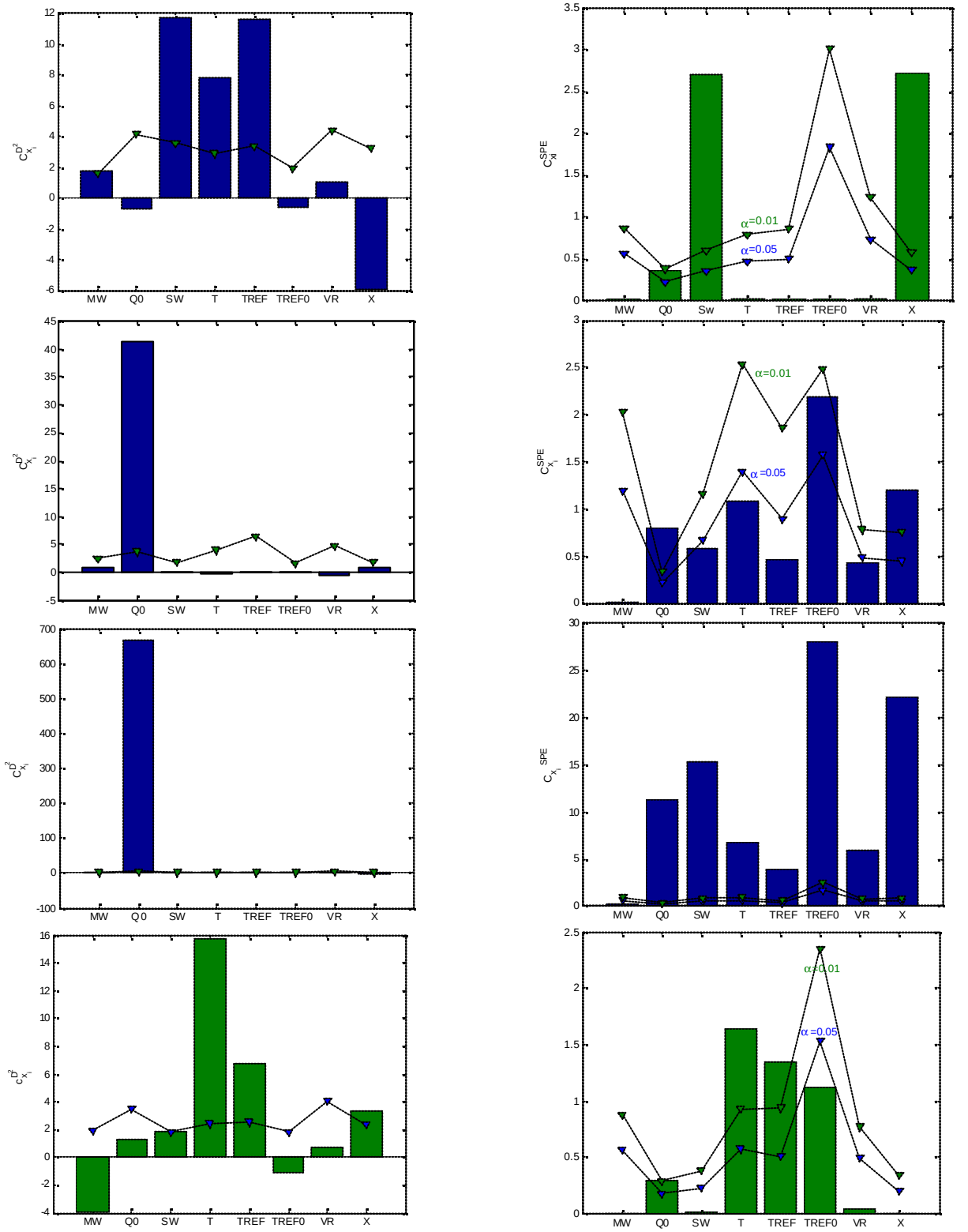


Fig 7: Contribution of each measurement to D^2 (left: Batches MB4:7) - *SPE* (right: Batches MB4:7)
Left: Batch MB4 k= 27, Batch MB5 k= 18, Batch MB6 k=13, Batch MB7 k=23
Right: Left: Batch MB4 k= 25, Batch MB5 k= 18, Batch MB6 k=13, Batch MB7 k=11

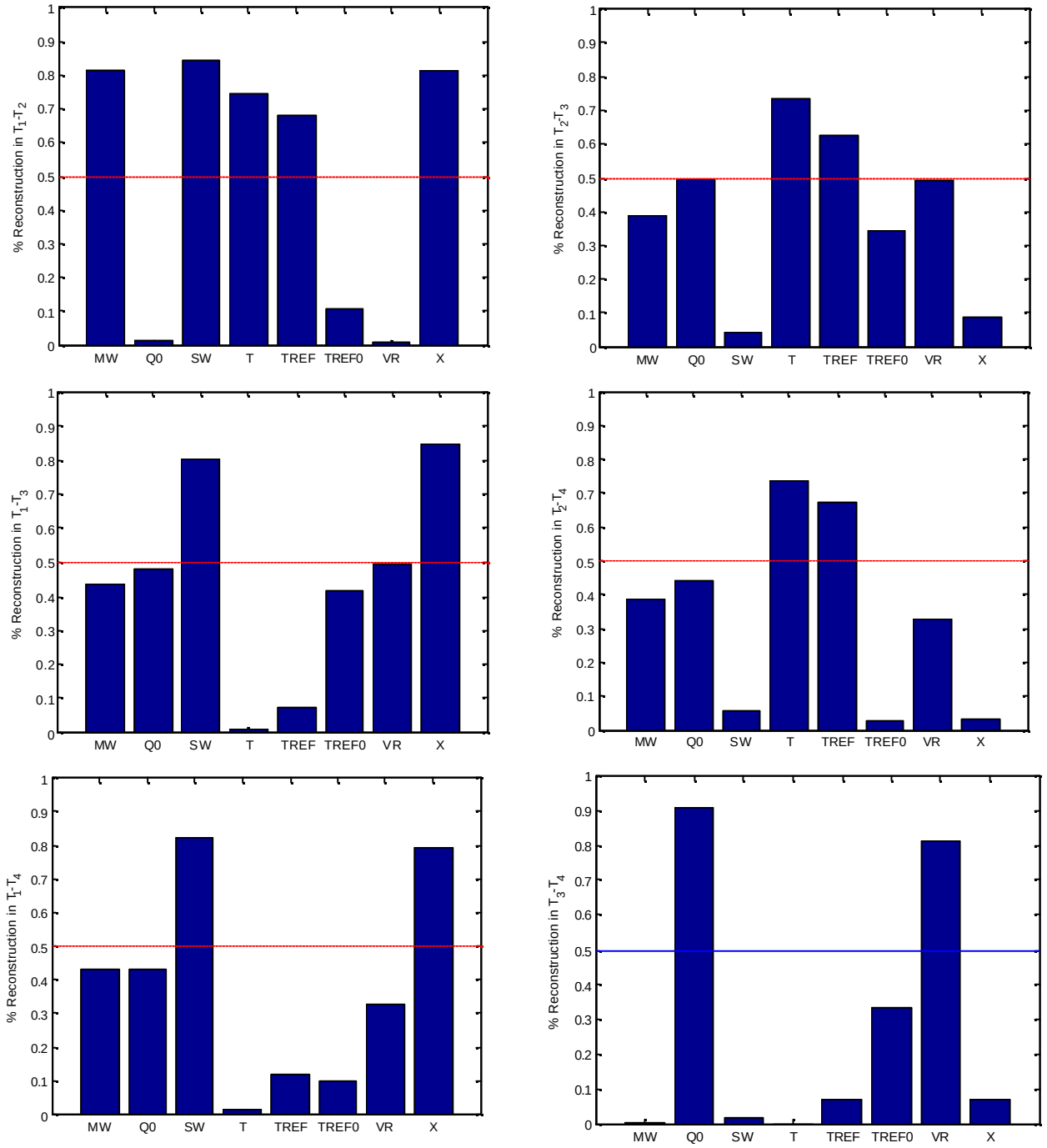


Fig 8: Variable reconstruction for different t planes

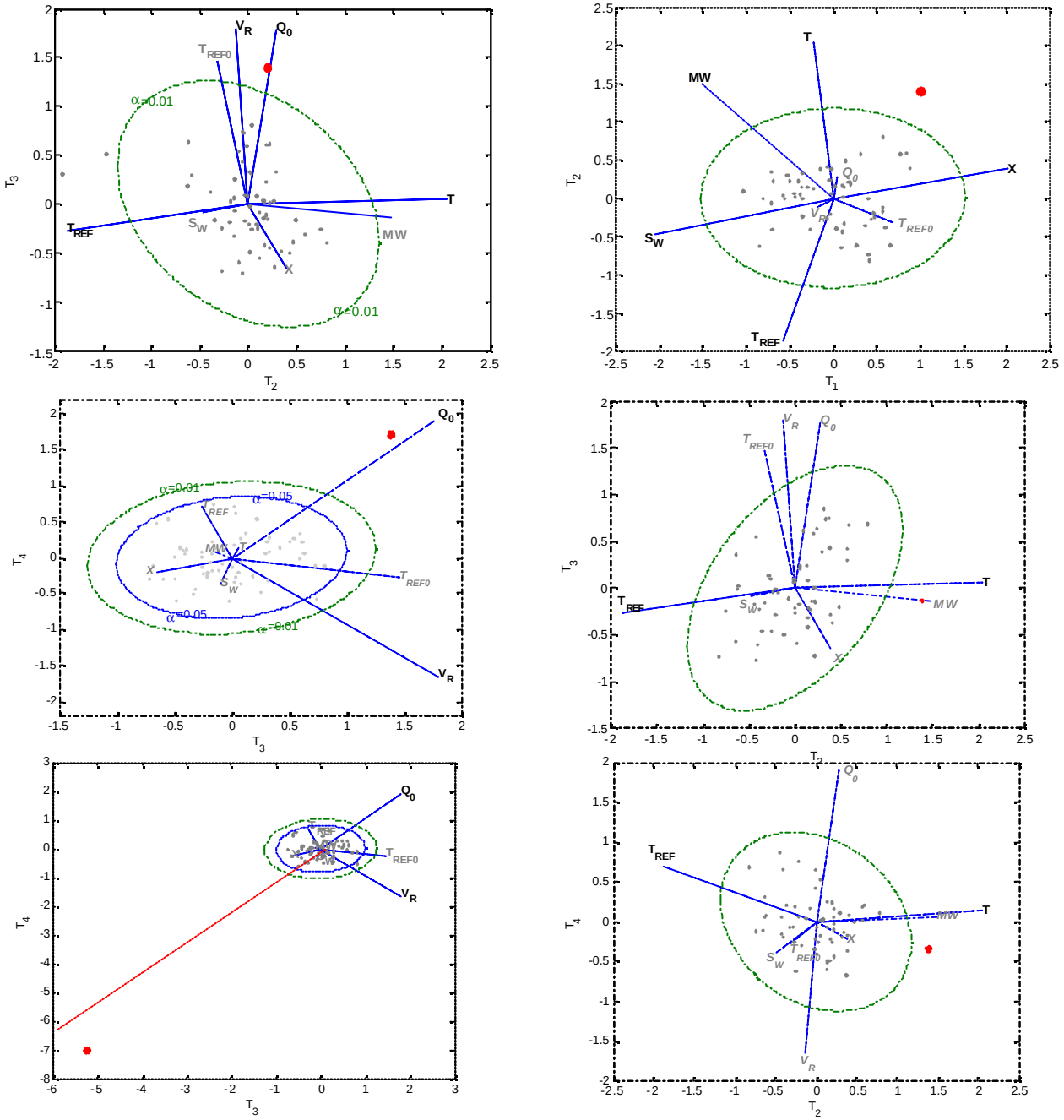


Fig 9: Biplots (Left: Batches MB5, MB5, MB6, Right: Batch MB7)

Conclusions

In this work a comparative study of the performance of the MPCA strategy for batch monitoring is performed considering two different techniques for the unfolding of the three-way array data matrix. The classic horizontal unfolding and the vertical unfolding of the mean centered and scaled horizontal matrix are studied, using simulated data of a methyl-methacrylate emulsion polymerization reactor. The analysis comprises the off-line modeling phase, and the on-line detection and identification phases.

The modeling phase involves the development of the empirical PCA model and the control charts. Given a set of industrial data, only the horizontal unfolding can be applied to determine a set of normal batches that behaves as a reference population. These data are then used to obtain different PCA empirical models depending on the selected unfolding strategy. The formulation of control charts for the *SPE* statistic requires a greater computation effort for the horizontal unfolding because estimates of future observations are necessary.

During the on-line detection phase, it is also needed to predict future observations for the horizontal unfolding. If these estimations are not sufficiently right, detection and identification faults may arise as in the provided examples. Furthermore, the vertical unfolding allows the use of biplots, a graphical tool that significantly enhances the identification of fault sources. Biplots can not be used with identification purposes for batch processes if the horizontal unfolding is applied.

References

1. Alvarez, C.R., Brandolin, A., Sánchez M. C., 2006. Production Monitoring in Polymerization Processes. MACRO 2006 World Polymer Congress, Río de Janeiro.
2. Box, G.E., Hunter, W.G., Hunter, J.S., 1978. *Statistics for Experiments*, New York, John Wiley.
3. Gabriel, K.R., 1971. The Biplot Graphical Display of Matrices with Applications to Principal Component Analysis. *Biometrika*, 58, 453.
4. Jackson, J.E., 1991. *A User's Guide to Principal Components*. John Wiley & Sons Inc., New York.
5. Nomikos, P.; MacGregor, J.F., 1994. Monitoring Batch Processes Using Multiway Principal Component Análisis, *AIChe Journal* **40**, 8, 1361-1375.
6. Nomikos, P.; MacGregor, J.F., 1995. Multiway Partial Least Squares in Monitoring Batch Processes, *Chemometrics & Intelligent Laboratory Systems* **30**, 97-108.
7. Westerhuis, J.A., Gurden, S.P., Smilde, A.K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems* **51**, 95.
8. Wold, S., 1987. Multi-Way principal components analysis, *Journal of Chemometrics* **1**, 41-56.
9. Wold, S.; Kettaneh, H.; Friden, H.; Holmberg, A., 1998. Modelling and Diagnosis of Batch Processes and Analogous Kinetic Experiments, *Chemometrics & Intelligent Laboratory Systems* **44**, 331-340.
10. Yoo, C.K., Lee, D. S., Vanrolleghem, A., 2004. Application of Multiway ICA for on-line process monitoring of a sequencing batch reactor. *Water Research* **38**, 1715.