# Monitoring a PVC Batch Process with Multivariate Statistical Process Control Charts

**Adriaan A. Tates, D. J. Louwerse, and Age K. Smilde***

*Department of Chemical Engineering, Process Analysis and Chemometrics, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

**Gerard L. M. Koot**

*Department of Engineering Support, Shell Research and Technology Centre Amsterdam, Badhuisweg 3, 1031 CM Amsterdam, The Netherlands*

**Harald Berndt**

*Shell Nederland Chemie B.V., PVC Manufacturing, Vondelingenweg 601, 3196 KK Rotterdam, The Netherlands*

Multivariate statistical process control charts (MSPC charts) are developed for the industrial batch production process of poly(vinyl chloride) (PVC). With these MSPC charts different types of abnormal batch behavior were detected on-line. With batch contribution plots, the probable causes of these abnormalities were located. Examples are given for two different types of abnormalities, i.e., a bias in the batch loading and a control valve problem during polymerization. These examples show that MSPC charts in combination with batch contribution plots are a simple and powerful method for fault detection and identification of their cause in the operation and control of batch processes.

## Introduction

Batch production processes play an important role in the chemical industry. Examples are the production of pharmaceuticals, biochemicals, and a large number of polymers. Detection of abnormalities in batch processes is very important, both for process safety and for product quality reasons.

Usually each batch run is different. This can be due to normal variation or special variation. Normal variation is caused by common variations in the operation (e.g., ambient temperature, feed quality); special variations are due to special causes (e.g., fouling, drift, incidental disturbances, or operational problems). If special causes are present and not detected and no countermeasures are taken, the operation may become nonoptimal and final product quality may deteriorate to off-spec levels. Detection of abnormalities directly after completion of a batch run may even prevent detrimental effects on the following batch run.

Typically, batch behavior is characterized by strong nonlinearities and non steady state operation, which limits the use of univariate statistical process control charts. Presently, consistent batch performance is obtained by monitoring a single variable or a few of the most relevant variables. Differences in batch performance, other than normal variation, and gradually changing process conditions are difficult to observe because historical process data and statistical control limits are not used.

There is a need for reliable and simple detection methods for abnormalities during a batch process and for means to assist in finding provable causes. In general, handling many strongly correlated variables can be done properly with statistical methods such as principal component analysis (PCA). Wise et al.[1] presented the theoretical basis for the use of principal component models for monitoring multivariate processes. Applications of PCA for monitoring batch processes, with process variables varying in time, were first presented by Nomikos and MacGregor.[2,3] Afterward Kosanovich et al.,[4,5] Kourti et al.,[6−9] and Martin et al.[10] have reported on their use of PCA to describe, analyze, and monitor batch processes. Neogi and Schlags[11] presented a very nice example of the use of multiway PCA (MPCA) for an emulsion batch process. They used reaction extent as a means to compare batches with varying time durations. They showed the power of using a multivariate statistical process control (MSPC) chart combined with a contribution plot to identify the process variables likely associated with the special variabilities in their examples. The use of contribution plots for continuous processes was first introduced by Miller,[12] who showed the power of contribution plots for two types of MSPC charts. In their paper, Neogi and Schlags used one type of MSPC chart while for batch processes two types of MSPC charts might give additional information.

This paper will show the advantages of monitoring batches with two types of MSPC charts, accompanied by their batch contribution plots, for operation and control of batch processes.

## Batch Production of PVC

PVC is produced by a suspension polymerization process. At the start, the raw materials are charged to the batch reactor. To start the polymerization reaction, initiator is added to the batch reactor. The propagation reaction takes place under moderate pressure and temperature. Because the reaction is highly exothermic, the excess of heat needs to be withdrawn from the

* To whom correspondence should be addressed. E-mail: asmilde@its.chem.uva.nl).

10.1021/ie9901067 CCC: $18.00 
Published on Web 11/12/1999

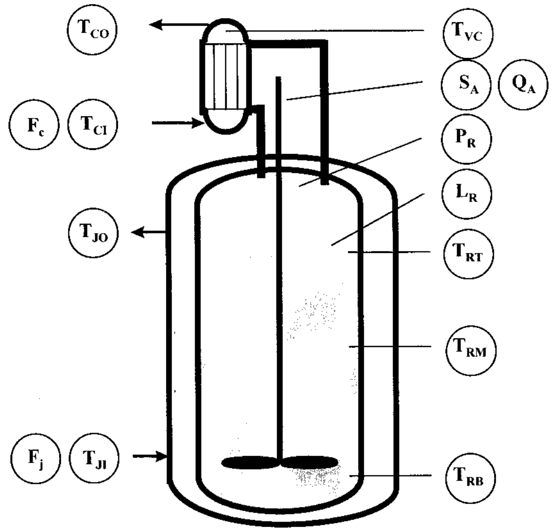


**Figure 1.** Schematic representation of a batch reactor and the relevant process variables.

**Table 1. Relevant Process Variables (*X*) for the PVC Reactor**

| number | process variables | description |
|---|---|---|
| 1 | $T_{CI}$ | inlet temperature cooling water of the condenser |
| 2 | $T_{CO}$ | outlet temperature cooling water of the condenser |
| 3 | $T_{VC}$ | condensation temperature at the top of the condenser |
| 4 | $Q_C$ | calculated duty of the condenser |
| 5 | $F_C$ | amount of cooling water through the condenser |
| 6 | $S_A$ | agitator speed |
| 7 | $Q_A$ | power supply to the agitator |
| 8 | $T_{RB}$ | temperature of the reactor at the bottom |
| 9 | $T_{RT}$ | temperature of the reactor at the top |
| 10 | $T_{RM}$ | temperature of the reactor in the middle |
| 11 | $T_{JO}$ | outlet temperature of the cooling water through the reactor jacket |
| 12 | $T_{JI}$ | inlet temperature of the cooling water through the reactor jacket |
| 13 | $L_R$ | liquid level in the batch-reactor |
| 14 | $F_J$ | amount of cooling water through the jacket |
| 15 | $P_R$ | pressure of the reactor |

process. Typical PVC reactors have a cooling jacket where cooling water flows through. Furthermore, a condenser on top of the reactor can also remove part of the heat of reaction by condensing vinyl chloride gas to liquid.

At conversions typically between 80% and 90%, the polymerization is finally stopped by adding a killing agent. A simplified process flow sheet of the batch reactor and the most relevant process variables which are measured during the polymerization are shown in Figure 1 and Table 1.

## Theory of PCA Modeling and On-line Monitoring

**Data Pretreatment.** The polymerization time varies from batch to batch. Because the batch process data are collected at fixed time intervals, this means that the data sets for different batches have different lengths. Multivariate analysis requires the data to be stacked in a matrix, so preprocessing of the data set is required. Linear interpolation techniques were used to transform the batch-time dimension into a reaction extent dimension. In this reaction extent dimension the batches are described with a fixed number of *K* points. Neogi and Schlags[11] used the same transformation, and this is believed to be an elegant and simple transformation of the batch data. More sophisticated transformations should only be used when this indisputably leads to major advantages.

**Building a PCA Model from Historical Batch Data.** A set of historical batch data can be pictured as a three-way array $\underline{\boldsymbol{X}}$ ($I \times J \times K$). For one batch, the data consists of $j = 1, 2, ..., J$ variables measured at $k = 1, 2, ..., K$ points throughout the batch. The same data are available for every batch $i = 1, 2, ..., I$ present in the data set. PCA is a two-way method, and this means that the three-way array of historical batch data must be rearranged into a two-way array, i.e., unfolding the three-way array. The left side of Figure 2 shows a schematic representation of the three-way array and of the unfolded two-way array (Wold et al.[13]).

The different batches are arranged along the vertical axis. The measured process variables as well as their time evolution are arranged along the horizontal axis. Each row in the unfolded matrix contains the total amount of data for one batch. Every column represents the values of one process variable at a certain time for all the batches.

The method of unfolding the three-way matrix in order to perform a PCA is a simple method for decomposing the data array into a lower dimensional space. Other methods for handling multiway data are for example the PARAFAC model and the Tucker model (Smilde[14]).

Before a PCA is performed on the batch data, the data are mean centered and scaled over the batches, thereby removing nonlinear time trajectories and assigning equal weights to all variables. After PCA, the batch process is described in a lower dimensional space, defined by principal components. This PCA decomposition can be described as follows:

$$\boldsymbol{X} = \sum_{r=1}^{R} \boldsymbol{t}_r \boldsymbol{p}_r^{\mathrm{T}} + \boldsymbol{E} = \boldsymbol{T}\cdot\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E} = \hat{\boldsymbol{X}} + \boldsymbol{E} \qquad (1)$$

The score vectors ($\boldsymbol{t}_r$) are orthogonal and the loading vectors ($\boldsymbol{p}_r$) are orthonormal. Usually, a small number of *R* principal components are sufficient to describe the data ($R \ll \min(I, J \times K)$). *R* is chosen such that most of the systematic variability of the process data is described by these principal components and that the residual matrix $\boldsymbol{E}$ contains no systematic structure anymore. Furthermore, the principal components are ordered such that the first principal component describes the largest amount of variation in the process data.

The loading vectors $\boldsymbol{p}_r$ define the reduced dimension space (*R*) and are the directions of maximum variability. Each loading vector summarizes the variation of the process variables about their average trajectories. The elements of the loading vector are the weights, of each process variable at a certain time, which give the *t*-scores for that batch.

The projection of a batch *i* on the reduced dimensional space, i.e., the point in the *R*-dimensional reduced space where batch *i* is located, corresponds with the elements of the *i*th row of the score matrix $\boldsymbol{T}$ ($\boldsymbol{T} = [\boldsymbol{t}_1...\boldsymbol{t}_R]$). Different batches can be compared with one another by

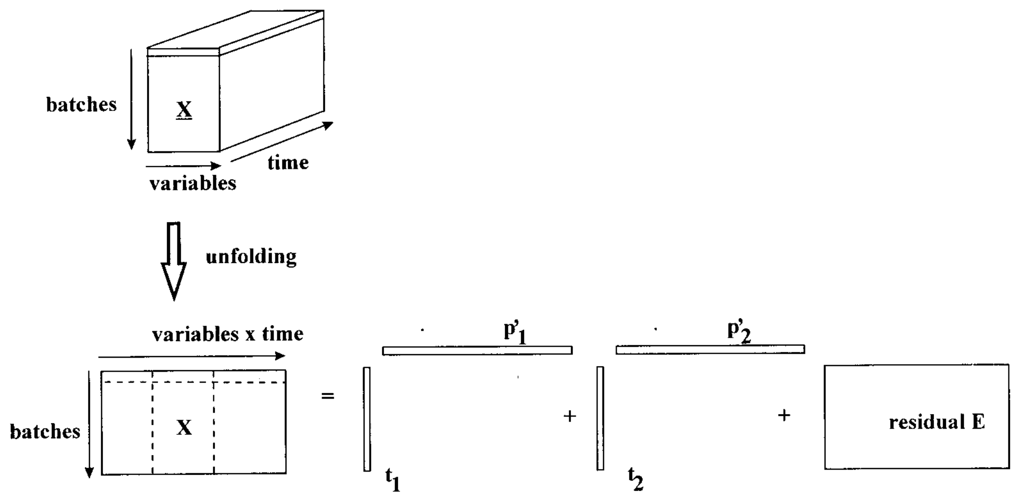


**Figure 2.** Three-way array $\underline{\boldsymbol{X}}$ filled with process data (top) and the unfolded two-way array $\boldsymbol{X}$ (left). A PCA is performed on the unfolded two-way array with $R = 2$ principal components. This results in two score vectors, two loading vectors, and a residual matrix $\boldsymbol{E}$.

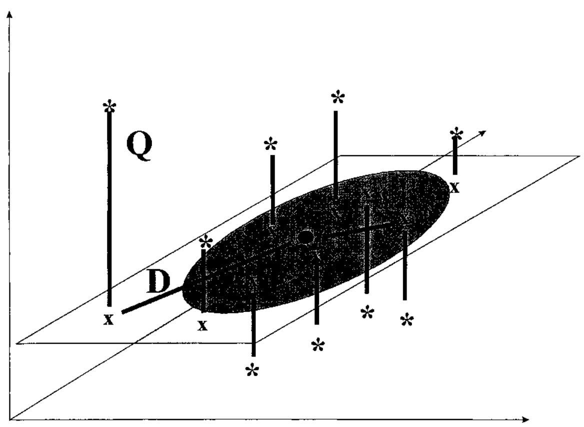


**Figure 3.** Crosses (x) in the reduced space, *t*-scores; distance between actual values (∗) and projected values (x), represented by the "Q". The distance between the projected values and the center of the reduced space (dot) is represented by the *D*-statistic.

these projections. If their projections are very close to each other, the resemblance between the batches is very large. On the other hand, if the distance between their projections is high, there are major differences between the batches.

The batches in the data matrix $\boldsymbol{X}$ can be compared with each other by comparing the score elements. This can be done in a scatter plot where the scores are plotted against each other, e.g., $\boldsymbol{t}_1$ plotted versus $\boldsymbol{t}_2$ (plane of Figure 3).

The batches can be quantified with Hotelling's $T^2$-statistic. Hotelling's $T^2$-statistic, named the *D*-statistic in this paper (following the notation of Nomikos and MacGregor[2,3]), is the multivariate analogue of the Student *t*-statistic. The Student *t*-statistic can be used to compare two univariate means. It is the difference between the means divided by the standard deviation. The *D*-statistic is used to compare two projections in the reduced space. It is the quadratic distance between the point in the principal component space where the batch is projected, i.e., the projection of that batch, and the center of the reduced space. This can be seen in Figure 3. The larger the *D*-statistic the further the batch is away from the center and the less "normal" that specific batch behaves. The *D*-statistic for a batch is calculated by

$$D_i = \boldsymbol{t}_i^{\mathrm{T}}(\boldsymbol{S})^{-1}\boldsymbol{t}_i \frac{I(I-R)}{R(I^2-1)} \quad (2)$$

where $\boldsymbol{S}$ represents the estimated covariance matrix of the *t*-scores of all batches. The distribution of the *D*-statistic for all batches can be approximated by a *F*-distribution, $F_{R,I-R}$, where $I$ is the total number of batches in the data set and $R$ is the number of principal components. With this *F*-distribution, the 95% and 99% confidence limits for the *D*-statistic are calculated. The name *D*-statistic was introduced to distinguish it from the $T^2$-statistic which works on the original variables (Nomikos and MacGregor[2,3]).

Besides the *D*-statistic, the batches can be quantified with the sum of squares of the residual matrix $\boldsymbol{E}$ from eq 1 of the model, called $Q$ (see Figure 3), and for batch number $i$ $Q$ is calculated as follows:

$$\boldsymbol{Q}_i = \sum_{j=1}^{J}\sum_{k=1}^{K}(\boldsymbol{x}_{ijk} - \hat{\boldsymbol{x}}_{ijk})^2 = \sum_{j=1}^{J}\sum_{k=1}^{K}(\boldsymbol{e}_{ijk})^2 \quad (3)$$

$x_{ijk}$, $\hat{x}_{ijk}$, and $e_{ijk}$ are the elements of $\underline{\boldsymbol{X}}$, $\underline{\hat{\boldsymbol{X}}}$, and $\underline{\boldsymbol{E}}$. $\underline{\hat{\boldsymbol{X}}}$ and $\underline{\boldsymbol{E}}$ are obtained by reshaping $\hat{\boldsymbol{X}}$ and $\boldsymbol{E}$ in the proper way. Likewise, $\boldsymbol{P}$ can be reshaped to obtain the three-way array of loadings $\underline{\boldsymbol{P}}$ $(J \times K \times R)$. $Q_i$ indicates the distance between the actual values of the batch and the projected values onto the reduced space. In other words, the residual accounts for any variability which is not described by the model. If $Q_i$ is very large, batch $i$ is poorly described by the PCA model and fits poorly into the reduced space defined by the principal components. $Q_i$ can be calculated for all $I$ batches in the data set. The distribution of the calculated $Q_i$ values can be approximated by a chi-squared distribution, $g\chi^2{}_h$, where $g$ is a constant and $h$ is the effective degrees of freedom of the chi-squared distribution (Box[15]). With this chi-squared distribution, the 95% and 99% confidence limits for the $Q$ value are calculated (Nomikos and MacGregor[2,3]).

**On-Line Monitoring of a Batch.** The historical data set of one reactor can be used to build an empirical model for normal operation of that reactor (yielding good product). To build such a model for normal operation, all abnormal batches must be removed from the set of the original batch data. If the set of data for normal operation is representative for all normal conditions, the upper control limit for the D-statistic is calculated (eq 2). For on-line monitoring, the squared prediction error ($\mathrm{SPE}_i$) for every batch in the data set is calculated for each time point $k$ individually, instead of $Q_i$ for the entire batch. For batch number $i$ and time point $k$, eq 3 then becomes

$$\mathrm{SPE}_{ik} = \sum_{j=1}^{J}(\boldsymbol{x}_{ijk} - \hat{\boldsymbol{x}}_{ijk})^2 = \sum_{j=1}^{J}(\boldsymbol{e}_{ijk})^2 \tag{4}$$

For every time point $k$ the distribution of the calculated $\mathrm{SPE}_{ik}$ values can also be approximated by a chi-squared distribution, and for every time point $k$ the 95% and 99% confidence limits are calculated with this distribution (Nomikos and MacGregor[2,3]).

In general, when a new batch is finished, it can be "projected" onto the reduced space defined by the principal components of the model. The loading matrix contains all the structural information about the model. With this loading matrix and with the new batch data, the score vector ($\boldsymbol{t}^{\mathrm{T}}$) for the new batch is calculated. Then, the $D$-statistic and the SPE value can be derived from this newly calculated score. If the new batch is similar to the previous batches, the difference between the model-predicted values and the actual batch data will be very small, and reversibly, if the batch behaves differently, the description by the model will be very poor and the differences will be large. This will show up as a large value for the $D$-statistic and/or the SPE value of that batch.

A problem with on-line monitoring is the fact that when a batch is in progress the measurements for the future time periods are unknown, and this means that the data set ($\boldsymbol{x}_{\mathrm{new}}$) is incomplete. Nomikos and MacGregor[2] presented three possible solutions for this problem. One of these solutions worked very well in practice and is the preferred choice of Nomikos and MacGregor.[3] In this solution, an assumption similar to the one made in model predictive control algorithms is used. This assumption is that the future deviations in $\boldsymbol{x}_{\mathrm{new}}$ from the mean trajectory will remain constant at their current values for the remainder of the batch. New methods of dealing with the "unknown future time periods" have been developed, based on consecutive processing of the different time points.[16,17] As said above, the method reported in Nomikos and MacGregor[3] works well in practice and is therefore used in this paper.

Based on the above assumption, at every time point $a$ ($a = 1, ..., K$), the future values of the process variables can be predicted. With the resulting matrix ($\boldsymbol{x}_{\mathrm{new},a}$), partly filled with known measurements and partly filled with predicted measurements, it is possible to project the current batch onto the model. With the on-line MSPC charts, i.e., the SPE plot and the $D$-statistic plot, it can be seen whether the new batch is in control or whether it deviates from the model. When the new batch deviates, the batch contribution plots can be used to determine the process variable which contributes most to this deviation. It should be noted that for the first few time points the results are not very accurate because there are too few real measurements. The stepwise procedure of the method for monitoring a new batch and calculating the contribution plots for every time point ($a$) is summarized in Figure 4.

1) determine $\boldsymbol{x}_{new,a}$

2) $t_{new,a} = \boldsymbol{x}_{new,a} . \boldsymbol{P}$

3) $\mathbf{D}_{new,a} = \mathbf{t}^{T}_{new,a} (\mathbf{S})^{-1} \mathbf{t}_{new,a} \cdot \dfrac{I(I-R)}{R(I^2-1)}$

4) $\hat{\mathbf{x}}_{new,a} = \mathbf{t}_{new,a}\mathbf{P}^{\mathbf{T}}$

5) $\mathbf{e}_{new,a} = \mathbf{x}_{new,a} - \hat{\mathbf{x}}_{new,a}$

6) $SPE_{new,a} = \sum_{j=1}^{J} (\mathbf{e}_{new,j,a})^2$

**Figure 4.** Stepwise method for calculating the $D$-statistic, the SPE value. These parameters are calculated for every time point $a = 1, ..., K$ of the new batch.

**Contribution Plots for On-Line Monitoring.** With the SPE and the $D$-statistic plot, it is possible to visualize if a batch is statistically out of control. If a batch is out of control, the values for the SPE and/or the $D$-statistic fall outside the 99% control limit (or 95% control limit).

If a batch is out of control in the $D$-statistic, the next logical step will be to examine the individual scores for that batch. For every batch $i$, the score $t_{ir}$ for dimension $r$ can be written as a linear combination of the elements of the $i$th row vector from $\boldsymbol{X}(=\boldsymbol{x}_i)$, multiplied by the loading vector $\boldsymbol{p}_r$ for that dimension.

$$t_{ir} = \boldsymbol{x}_i^{\mathrm{T}} \cdot \boldsymbol{p}_r = \sum_{j=1}^{J}\sum_{k=1}^{K} x_{ijk}p_{jkr} \tag{5}$$

where $x_{ijk}$ is an element of $\underline{\boldsymbol{X}}$ and $p_{jkr}$ is an element of the loading matrix $\underline{\boldsymbol{P}}$, described as the loading for process variable $j$, time point $k$, and dimension $r$. Thus, $t_{ir}$ can be decomposed into $J \times K$ terms $x_{ijk}p_{jkr}$. These $J \times K$ terms are the individual contributions to the score $t_{ir}$ (see Miller[11]). For batch $i$ and dimension $r$, each process variable $j$ has a contribution at every time point $k$, resulting in $K$ contributions for each process variable. The overall batch contribution of process variable $j$ in dimension $r$, is obtained by quadratic summing these $K$ individual contributions:

$$C_{D,ijr} = \sum_{k=1}^{K}(x_{ijk}p_{jkr})^2 \tag{6}$$

where $\boldsymbol{C}_{D,ijr}$ is defined as the batch contribution to the $D$-statistic, for batch $i$, process variable $j$, and dimension $r$. Similarly, when a batch is out of control in the SPE value, this SPE can be decomposed into the individual contributions. For a batch which is out of control at time point $k$, the $\mathrm{SPE}_k$ is given by eq 4. The individual contributions are given by the quadratic elements $e_j$ of the residual vector $\boldsymbol{e}$. The contribution of process variable $j$, to the SPE value at time point $k$, is

$$\boldsymbol{C}_{\mathrm{SPE},jk} = \boldsymbol{e}_{jk}^{\;2} \tag{7}$$

where $\boldsymbol{C}_{\mathrm{SPE},jk}$ is defined as the batch contribution to the SPE value for process variable $j$ at time point $k$. It must be emphasized that this is the contribution for process variable $j$ at time point $k$ in the batch and for every time point ($k = 1, ..., K$) a different contribution can be calculated.

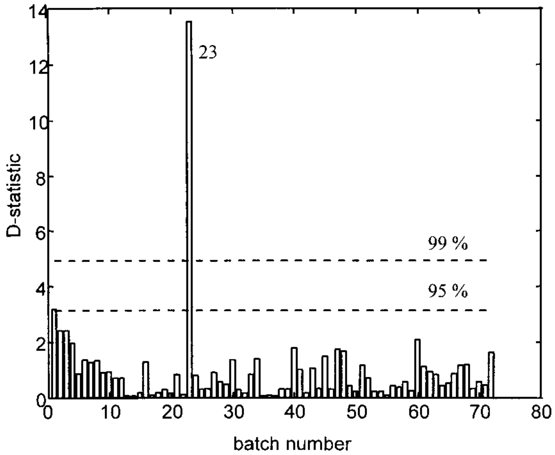


**Figure 5.** *D*-statistic plot for the data set of 72 batches. The PCA model is built with two principal components.

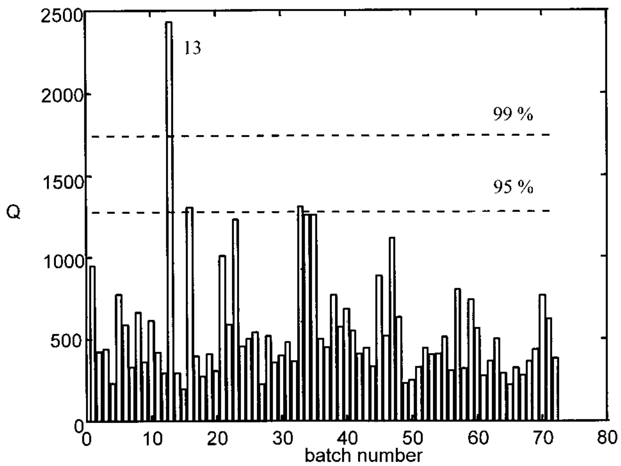


**Figure 6.** *Q*-plot for the data set of 72 batches. The PCA model is built with two principal components.

## Results and Discussion

**Multivariate Statistical Model for the PVC Batch Process.** First, the model was built from a historical data set of "normal" batches. A data set, containing 72 batches, was obtained from the PVC process. This data set was decomposed with PCA, and via cross-validation (Eastment and Krzanowski[15]) it was found that two principal components were needed to describe the data set. Figures 5 and 6 show the *D*-statistic and the SPE values of all the batches in the model. It can be seen that two batches, i.e., batch 13 and 23, were significantly different. These batches are left out, as well as one normal batch, i.e., batch 50, which is chosen arbitrarily. A new PCA model is built from the 69 remaining batches. The number of principal components is again determined with cross-validation. Two principal components were needed to build the model, and this model captures 38% of the variation in the process data set. This model contains only normal batches (no batch is above the 99% control limit), and all 69 batches resulted in good product quality. The amount of explained variation (38%) might seem low, but is often encountered in batch MSPC. This is due to the presence of a large amount of unsystematic variation in $\boldsymbol{X}$, due to, e.g., measurement error and sensor offsets.

With these 69 batches, the upper control limits for the *D*-statistic and the SPE are calculated for every time point (MSPC charts). The effectiveness of this multivariate statistical model is illustrated with the abnormal batches from the data set, i.e., batch numbers 13 and 23. These two outlying batches are first scaled with the mean and the standard deviation of the normal operating batches in the model, and then projected onto the model for every time point $k$ (according to the scheme in Figure 4). Furthermore, a third batch, batch number 50, is also projected for every time point to illustrate the projection of a "good" batch onto the model.

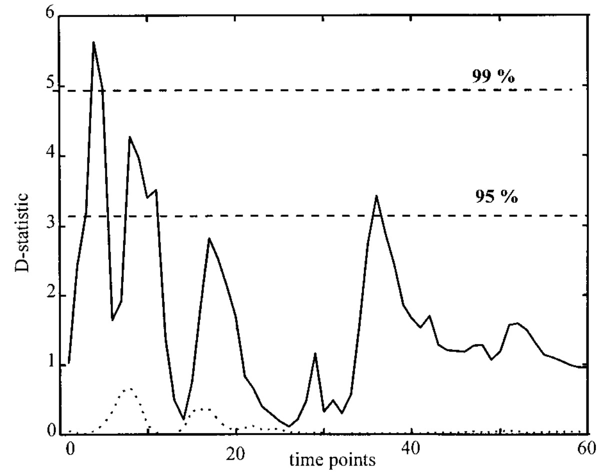


**Figure 7.** *D*-statistic plot and the upper control limits (- - -) for on-line monitoring abnormal batch 23 (—) and normal batch 50 (···).

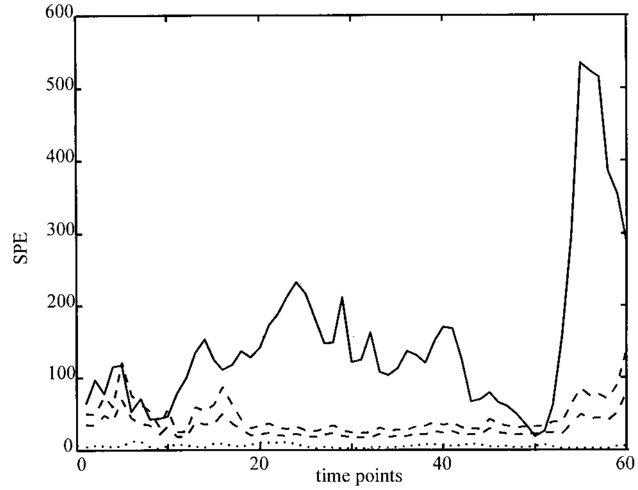


**Figure 8.** SPE plot and the upper control limits (- - -) for on-line monitoring abnormal batch 23 (—) and normal batch 50 (···).

**Monitoring Batches.** *Batch 50.* Batch 50 is monitored for every time point $k$ with the *D*-statistic plot (eq 2) and with the SPE plot (eq 7). The 95% and 99% control limits are calculated, and they define the normal operating region. The results from monitoring batch 50 are shown in Figures 7 and 8. It can be seen that batch 50 stays below the upper control limits, for both the *D*-statistic and the SPE. It follows that this batch has no abnormal behavior at any time during the polymerization. Therefore, this batch is described as being "in control" or "normal" and the polymerization is well described by the empirical model.

*Batch 23.* Batch 23 is also monitored for every time point $k$ with the *D*-statistic and the SPE plot. The results of monitoring batch 23 are also shown in Figures 7 and 8. From Figure 8 it can be seen that, in contradiction with monitoring batch 50, batch 23 immediately

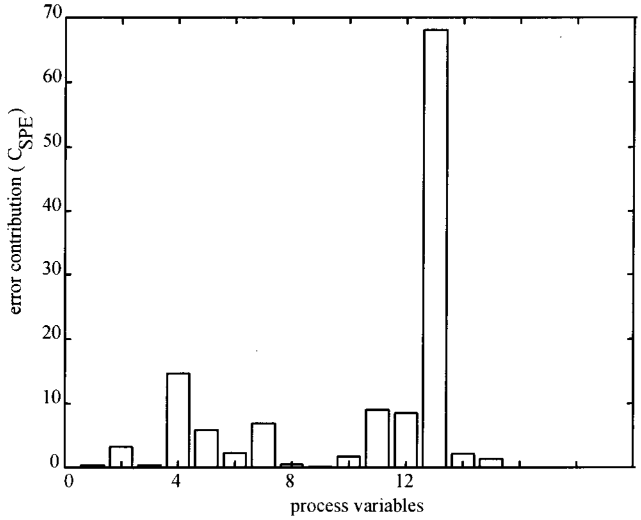


**Figure 9.** Contribution to the SPE value for all process variables at time point $k = 15$ for abnormal batch 23.

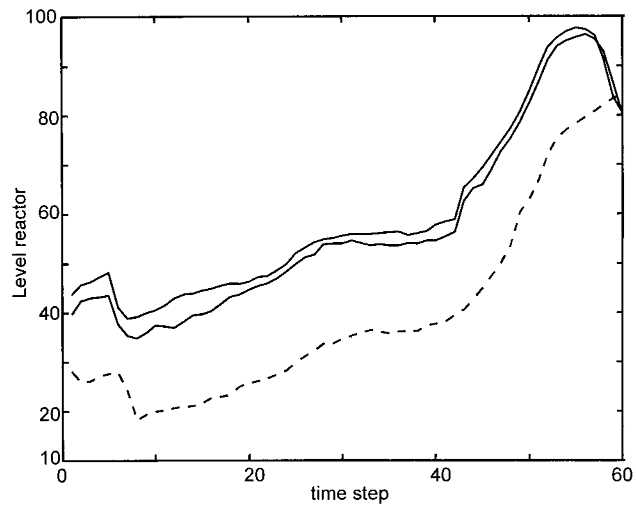


**Figure 10.** Univariate profile of the level in the reactor for abnormal batch number 23 (- - -) and for normal batches (—).

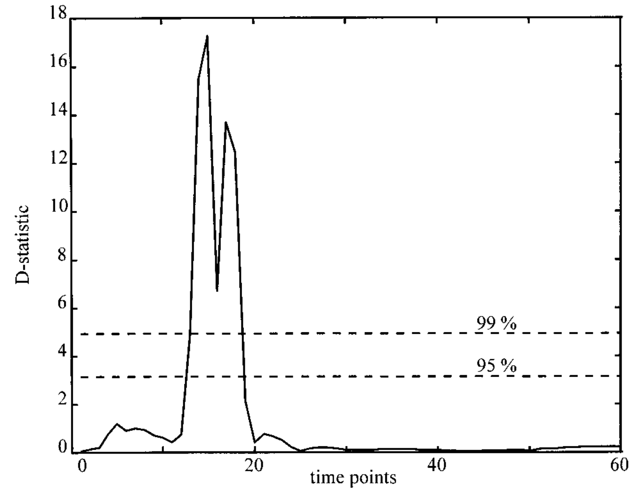


**Figure 11.** *D*-statistic plot and the upper control limits (- - -) for on-line monitoring abnormal batch 13 (—).

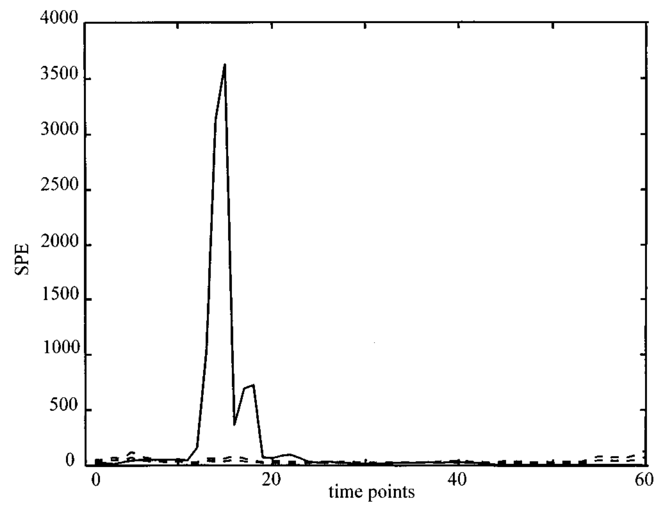


**Figure 12.** SPE plot and the upper control limits (- - -) for on-line monitoring abnormal batch 13 (—).

moves away from the normal operating regime; this indicates a special variation which is present from the start of the polymerization. This special variation is poorly described by the model, generating large SPE values, particularly at the end of the polymerization. However, in the model projection plane the batch does not exceed the control limits (e.g., the batch stays below the *D*-statistic control limits).

Immediately after the deviation is noticed the abnormal batch is studied in more detail. At time point number 15, the batch is out of control in the SPE value and not in the *D*-statistic. Figure 9 shows the calculated contributions to the SPE value at time point $k = 15$. It is obvious that the level in the reactor (see process variable number 13, Table 1) has the largest contribution to this SPE value. The contribution plot reveals that the level in the reactor causes this abnormal behavior, and this abnormal behavior is detected at an early point in the polymerization. Figure 10 shows the univariate plot of the level in the reactor for normal batches and for batch number 23. It can be seen that the level in batch number 23 is indeed structurally lower. This lower level was caused by a leaking valve in combination with vacuum being applied to another reactor: at the time that batch 23 was loaded, vinyl chloride monomer was sucked away into the other reactor. This operational error was not noticed during operation. The first reason is that the current univariate profiles of the process variables are not compared with those from previous "normal" batches. Therefore, the difference in the level was not noticed. A second reason is that not all process variables are monitored because this results in too many univariate charts. These problems are overcome by using MSPC. Only three MSPC charts (the *D*-statistic, the SPE chart, and the batch contribution chart) are necessary to detect the error at an early stage, and to assign the process variable that causes the error.

*Batch 13.* Another example of detecting an operational problem is shown by monitoring batch number 13. The results are shown in Figures 11 and 12. These MSPC charts clearly detect the out-of-control situation between the tenth and the twenty-fifth time points. During this period, the model description of the batch is very poor, resulting in a large SPE and *D*-statistic.

Again, the cause of these large *D*-statistic and SPE values is found by using the contribution plots. The flow of the cooling water through the condenser (process variable number 5) has the largest contribution, both in the *D*-statistics and in the SPE. This is obvious from Figures 13 and 14. Univariately looking at this process variable validates this finding (Figure 15). It shows that

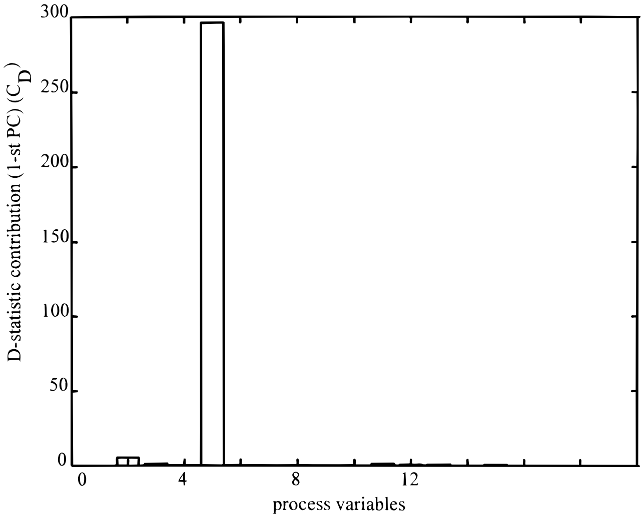


**Figure 13.** Weight contribution of the score ($C_D$) for time point $k$ = 15 for abnormal batch 13.

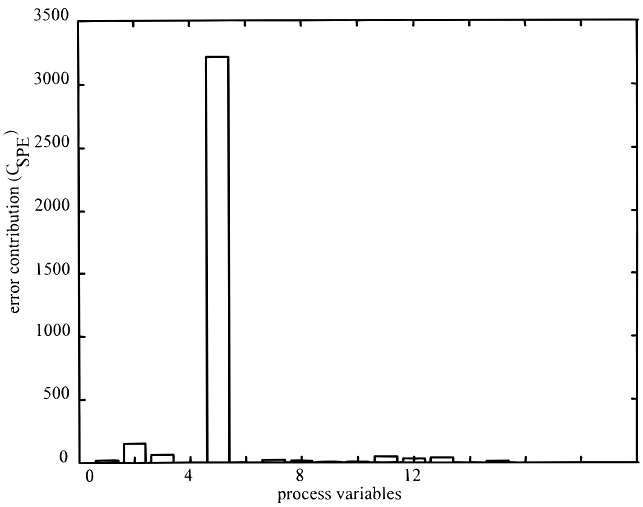


**Figure 14.** Error contribution ($C_{SPE}$) for all process variables at time point $k$ = 15 for abnormal batch 13.

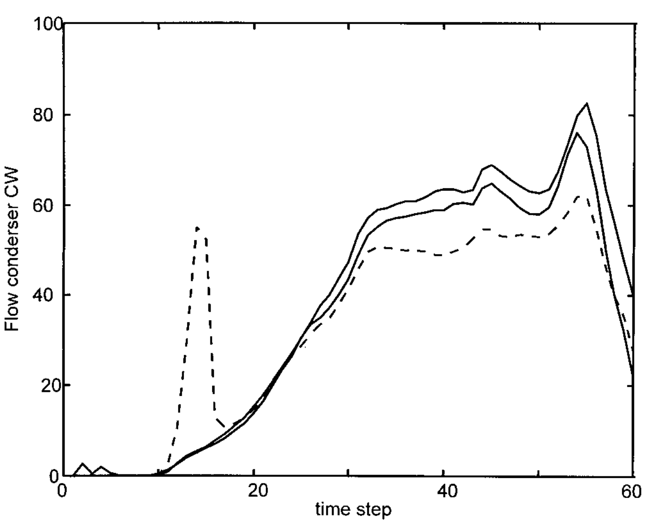


**Figure 15.** Univariate profile of the flow of the cooling water through the condenser for abnormal batch 13 (- - -) and for normal batches (—).

an abnormal peak in the flow of the cooling water is present between the tenth and the twenty-fifth time points. This abnormal peak was caused by the controller valve in the cooling water pipe. The valve was opened too fast, and this resulted in an overflow of coolant. Furthermore, it shows that this overshoot is corrected by the controller and the final product was still within specification. This operational disturbance was also not noticed by the operators. Although the disturbance was corrected by the controller, it is important to detect these disturbances. At the time of the malfunctioning valve, the heat removal via the condenser is not controlled and this is an undesired situation. When this occurs more often, corrective actions have to be taken (e.g., maintenance of the valve) to prevent the problem to occur in future batches.

## Conclusion

In this paper, a multivariate statistical model is built from historical data of a real PVC polymerization process. With this model, it is possible to compare future batches with previous “normal” batches. This is based on the idea that a future normal batch behaves similarly to the historical normal batches. It is shown that running batches can be monitored on-line with MSPC charts and abnormal behavior is immediately detected. Furthermore, the cause of this abnormal behavior can be found fairly easily with on-line batch contribution plots. Two examples of erroneous batches are discussed: a batch with an initial loading error and a batch with a malfunctioning control valve.

MSPC charts prove to be a powerful and simple method for detecting erroneous batches and for identifying the causes of these errors. The authors believe that such MSPC charts have the potential of becoming powerful aiding tools for process operation and process control of batch processes. Several problems of implementing these charts in particular, and of similar statistical techniques in general, will have to be dealt with: problems such as changing operation points, slow drift of process condition, e.g. due to fouling, and seasonal influences. These problems must be studied in more detail, and solutions should be found in order to further develop these techniques and for their use in practice.

## Acknowledgment

We thank Shell Nederland Chemie B.V. and Shell Research and Technology Centre Amsterdam for providing the batch data of the PVC polymerization process and for their support in this study.

## Notation

$a$ = index for time point with on-line monitoring
$\boldsymbol{e}$ = residual vector for a new batch
$\boldsymbol{E}$ = residual matrix for a data set
$F_{R,I-R}$ = $F$ distribution with $R$ and $I-R$ degrees of freedom
$g$ = constant associated with the distribution of the squared prediction error
$h$ = degrees of freedom for chi-squared distribution
$i$ = index for batches
$I$ = total number of batches
$j$ = index for process variables
$J$ = total number of process variables
$k$ = index for time points
$K$ = total number of time points
$\boldsymbol{p}$ = loading vector
$\boldsymbol{P}$ = loading matrix
$r$ = index for principal components

$R$ = total number of principal components
$\boldsymbol{S}$ = covariance matrix of $t$-scores
$t$ = score for a new batch
$\boldsymbol{t}$ = score vector
$\boldsymbol{X}_{\mathrm{new}}$ = new batch
$\underline{\boldsymbol{X}}$ = historical three-way database

*Greek Letter*

$\chi^2{}_h$ = chi-squared distribution with $h$ degrees of freedom

*Matrix*

$x$ = scalar
$\boldsymbol{x}$ = vector
$\boldsymbol{X}$ = matrix
$\underline{\boldsymbol{X}}$ = three-way/multiway array
$\boldsymbol{X}^{\mathrm{T}}$ = transposed matrix of $\boldsymbol{X}$

## Literature Cited


(1) Wise, B. M.; Ricker, N. L.; Veltkamp, D. F.; Kowalski, B. R. A theoretical basis for the use of principal components models for monitoring multivariate processes. *Process Control Quality* **1990**, *1*, 41−51.

(2) Nomikos, P.; MacGregor, J. F., Multivariate SPC charts for monitoring Batch processes. *Technometrics* **1995**, *31* (1), 41−59.

(3) Nomikos, P.; MacGregor, J. F. Monitoring of batch processes using multi-way Principal Component Analyses. *AIChE J.* **1994**, *40*, 1361−1375.

(4) Kosanovich, K. A.; Piovoso, M. J.; Dahl, K. S.; MacGregor, J. F.; Nomikos, P. Multi-Way PCA Applied to an Industrial Batch Process. *Proc. Am. Control Conf.* **1994**, *2*, 1294−1298.

(5) Kosanovich, K. A.; Dahl, K. S.; Piovoso, M. J. Improved Process Understanding Using Multiway Principal Component Analysis. *Ind. Eng. Chem. Res.* **1996**, *35*, 138−146.

(6) Kourti, T.; MacGregor, J. F. Process Analysis, monitoring and diagnosis, using multivariate projection methods. *Chemom. Intell. Lab. Syst.* **1995**, *28*, 3−21.

(7) Kourti, T.; MacGregor, J. F. Multivariate SPC Methods for Process and Product Monitoring. *J. Quality Technol.* **1996**, *28* (4), 409−428.

(8) Kourti, T.; Nomikos, P.; MacGregor, J. F. Analysis, Monitoring and Fault Diagnosis of Batch Processes using Multiblock and Multiway PLS. *J. Process Control* **1995**, *5*, 277−284.

(9) Kourti, T.; Lee, J.; MacGregor, J. F. Experiences with industrial applications of projection methods for multivariate statistical process control. *Comput. Chem. Eng.* **1996**, *20*, S745−S750.

(10) Martin, E. B.; Morris, A. J.; Papazoglou, M. C.; Kiparissides, C. Batch process monitoring for consistent production. *Comput. Chem. Eng.* **1996**, *20*, S599−S604.

(11) Neogi, D.; Schlags, C. E. Multivariate Statistical Analysis of an Emulsion Batch Process. *Ind. Eng. Chem. Res.* **1998**, *37*, 3971−3979.

(12) Miller, P.; Swanson, R. E.; Heckler, Ch. E. Contribution Plots: A Missing Link in Multivariate Quality Control. *Appl. Math. Comput. Sci.* **1998**, *8*, 775−792.

(13) Wold, S.; Geladi, P.; Esbensen, K.; Ohman, J. Multi-way Principal Components and PLS Analysis. *J. Chemom.* **1987**, *1*, 41−56.

(14) Smilde, A. K. Three-way analysis: Problems and Prospects. *Chemom. Intell. Lab. Syst.* **1992**, *15*, 143−157.

(15) Box, G. E. P. Some Theorems on Quadratic Forms Applied in the Study of Analysis of Variance Problems: Effect of Inequality of Variance in One-Way Classification. *Ann. Math. Stat.* **1954**, *25*, 290.

(16) Rännar, S.; MacGregor, J. F.; Wold, S. Adaptive batch monitoring using hierarchical PCA. *Chemom. Intell. Lab. Syst.* **1998**, *41*, 73−81.

(17) Louwerse, D. J.; Smilde, A. K. Multivariate Statistical Process Control of Batch Processes using Three-way Models. *Chem. Eng. Sci.* **1999**, in press.

(18) Eastment, H. T.; Krzanowski, W. J. Cross-Validatory Choice of the Number of Components from a Principal Component Analysis. *Technometrics* **1982**, *24*, 73−77.



*Received for review* February 12, 1999
*Revised manuscript received* August 12, 1999
*Accepted* August 16, 1999

IE9901067