Statistical batch process monitoring

Ramaker, H.J.; van Sprang, E.N.M.

Citation for published version (APA):
CHAPTER 2 ♦ I.T.A.-TRAJECTORY*
2.1 General introduction

The basic ideas about SPC were briefly explained in the previous chapter. Within quality management programs such as the 6σ program or TQM (Total Quality Management), SPC is used as an essential tool. A similar program concentrating on the Dutch market is ZVP (Zakelijk Verbeter Programma) by (R.J.M.M. Does et al. [47]). The ZVP is based upon a collection of various quantitative quality programs and consists of a scheme to come to a faultless process, service or product. These steps read as follows:

1) Measure
2) Analyse
3) Improve
4) Guarantee

The same philosophy as ZVP is employed for chemical processes, in particular batch processes. The first initiative for a batch improvement program (Batch Verbeter Programma) is to assign phases according to the concepts of multivariate statistical process control. W.H. Woodall [48] stated that it is very important to improve the communication between practitioners and researchers concerning statistical process control methods. One way to establish an improved communication is to assign phases to the concepts of multivariate statistical process control. Multivariate statistical process control is carried out in three phases: The Initial, Training and Application phase, or the I.T.A.-trajectory. Here, the training and application phase refer to respectively phase 1 and phase 2 from standard SPC terminology. An illustration of the I.T.A.- trajectory is given in the following figure:
Each of the phases from the I.T.A.-trajectory will be discussed in the following chapters.

2.2 I: Initial phase

In the initial phase historical process data is collected and analysed to determine if the data is suited for process monitoring. Because of control actions and safety, it is common that various process variables are measured on-line. This is usually done by means of DCS (distributed control systems) and SCADA (supervisory control and data acquisition system). These systems control the underlying PLC’s (programmable logic controller) and properly store the collected data in a database. Quality measurements of the end-product are usually stored in a LIMS (Laboratory Information Management System) system. Therefore, data are spread in various systems and are not always accessible in an easy manner. Furthermore, it is common that sensors are logged continuously without any markers. The analyst will have to use batch reports to search for the actual time where the batch started. It is not difficult to imagine that this phase is a very time consuming step and often forms a stumbling block in batch monitoring. Another problem often encountered in the initial phase is
missing data because of sensor failure or other malfunction. There are several suggestions in the literature to deal with this problem.

2.2.1 EXPLORATIVE ANALYSIS
Once the data is extracted from the database an explorative analysis is performed. In this analysis the relation and correlation between process variables and end-product quality is determined. Moreover, it is analysed if measured process variables consist of process relevant information. The first step is to look at the trajectories of the measured process variables over time. Often, abrupt faults are easily detected by examining the plots of the individual trajectories. This already tells a lot about the process. Furthermore, in this stage, it is determined if the process is statistically in-control e.g. does the process operates around a steady working point with a certain level of variation. This will be will be defined as normal operating conditions (NOC). In batch process monitoring it is important to specify the normal operating conditions. A first selection is based on the operator knowledge, batch reports and end-product quality results. In the training phase, also multivariate statistical techniques are used to determine the NOC batches in the obtained batch data. This is necessary to detect more incipient faults that are not easily detectable by examining individual trajectories of the process variables. Such faults are typical in industry.

2.3 T: Training phase
The goal of the training phase is to build a well functioning model based on a set of NOC batches. From this model, the \( \text{SPE} \) and \( D \) control chart are constructed that are used in the application phase for on-line monitoring of completely new batches. The training phase is divided in a few steps. These steps will be described step by step.

2.3.1 NOC SELECTION OR OUTLIER DETECTION
The batches that are collected in the initial phase preferably meet the following requirements. First, the number of batches should be sufficient. The statistical properties of the model are more precise when large numbers of batches are available. Secondly, the batches represent normal operating conditions (NOC). This is very important, since the process variation in the NOC batches serves as a reference distribution. The more this process variation represents NOC conditions, the better
future faulty batches are detected. If e.g. a faulty batch is included in the NOC data, the total amount of process variation increases. As a result, the reference distribution now consists of wrong batches. Therefore, the model is less capable of detecting differences in process variation between the NOC data and a new wrong batch.

2.3.1.1 Post-batch analysis
A single batch can be represented by a matrix \( X \) (\( K \times J \)), where \( K \) is the number of observations and \( J \) the number of process variables. \( J \) can be either engineering variables such as temperature and pressures, but also channels from spectroscopic measurements. By repeating batch runs, the batches can be stored in a three-way array \( X^{I \times J \times K} \), where \( I \) is the number of batches. This is illustrated in Figure 12:

Before the batch process monitoring model is built, a selection of NOC batches from the three-way array \( X \) is required. This selection is based on multivariate statistical tools and will be referred to as post-batch analysis. An example is given how to select batches from a dataset \( X^{53 \times 9 \times 200} \). The dataset concerns a polymerisation of styrene butadiene. From this process, 53 batches are obtained and 9 process variables are measured for a period of 200 time intervals for every batch. The question is: select NOC data by using post batch analysis. The trajectories of the process variables are plotted in the following figure:
It is obvious from Figure 13 that distinguishing between normal and faulty batches is laborious and difficult. The multivariate approach is performed as follows:

**Scaling**
Since PCA is a linear technique, pre-processing of \( \mathbf{X} \) is required because of the highly non-linear behaviour of batch processes. In this example, the process data \( \mathbf{X} \) is column-centred and tube-scaled (H.A.L. Kiers [49]), removing the average trajectories of all the process variables and scaling the process variables to unit variance. The pre-processed data in \( \mathbf{X} \) now represents deviations from the average trajectory and is therefore approximately linear (P. Nomikos & J.F. MacGregor [6], A.K. Smilde [50]).

**Unfolding the data**
The data is arranged in a three-way data array \( \mathbf{X} \). For post batch analysis, \( \mathbf{X} \) is matricized in order to perform a PCA analysis. The matricizing is depicted in Figure 14:

![Figure 13](image)
*Raw data of engineering variables from SBR production.*

![Figure 14](image)
*Matricizing three way data.*
It can be seen that matricizing results in the matrix \( X^{1\times JK} = X^{33\times 1800} \).

**PCA**

The next step is to build a PCA model of \( X \).

\[
X = TP' + E
\]  

Where \( T \) are the scores, \( P \) the loadings and \( E \) the residuals. In this example, three principal components (\( R \)), as determined with cross validation, are sufficient to describe ±30% of the variation in \( X \).

**T and Q plot**

The scores and residuals that result from the PCA analysis are used to construct two simple graphs, a scatter plot of \( T \) and a \( Q \)-plot. First, a scatter plot of \( T \) is simply constructed by plotting the scores for the first column of \( T \) against the scores of the second column of \( T \). The score plot for the SBR data is given in Figure 15.

![Figure 15](image1.png)

*Scatter plot of the scores.*

![Figure 16](image2.png)

*\( Q \)-statistic of the SBR process with 99% confidence limits.*

It can be seen from the score plot that batch 52 and 53 can be regarded as outliers. An individual inspection of these batches will learn what the cause is of this deviating behaviour. In this example the deviations for both batches are caused by impurities in
the feed (P. Nomikos & J.F. MacGregor [6]). From the residuals, the sum of squares of these residuals, denoted as $Q$, is calculated as follows:

$$Q_i = \sum_{k=1}^{K} \sum_{j=1}^{J} e_{ij,k}$$  \hspace{1cm} (15)

Note that the sum of squares of the residuals for post batch analysis is referred to as $Q$. The $Q$-statistic can be approximated by a weighted chi-square distribution $g \cdot \chi^2(b)$ where $g$ is the weight and $b$ the degrees of freedom (J.E. Jackson & G.S. Mudholkar [24]). The UCL for the $Q$-statistic can be found by using this approximation. The $Q$-statistic for all 53 batches are given in Figure 16. From this figure, there are no peculiarities observed in the $Q$-statistic. Based on the results of the $T$ and $Q$-plot, batch 52 and 53 are removed from the dataset. The data that remains is said to be statistically in-control. This can be verified by looking at the $T$-plot and $Q$-plot for the dataset where batch 52 and 53 are removed. Figure 17 and Figure 18 represent the scatter plot and the $Q$-statistic respectively. Note that now the scatter plot is a random shot implying that all batches have some variation around a mean trajectory. Also, the $Q$ statistics do not show any abnormalities.

2.3.1.2 Model building: evolving models
Now that a set of NOC batches is obtained, a model can be constructed for on-line monitoring. The model is used to construct the control charts ($SPE$ and $D$-chart) for monitoring completely new batches. Different models can be chosen as will be
described in Chapter 4.3. For illustration purposes, a time evolving model (Figure 19) is chosen (see Chapter 4.3).

This model is constructed as follows: at every time interval \( k \), a model is built. For the time evolving model, a matrix \( X_k \) \((I \times Jk)\) is constructed from the \( k \) frontal slabs of \( \overline{X} \). This is illustrated in Figure 19.

It can be seen that \( X_k \) is the matricized part of \( \overline{X} \) until time interval \( k \). Obviously, the size of \( X_k \) increases with time. The evolving model gives \( K \) different model loadings \( P_k \) \((Jk \times R)\) as can be seen from Figure 19. The computational power of modern computers is no impeding anymore to build these models. The reference distributions for the test statistics are formed by \( T_k \) \((I \times R)\) and \( E_k \) \((I \times Jk)\).

### 2.3.1.3 D and SPE-statistic

The \( D \)-statistic is based on the scores \( T_k \) that are calculated at every time interval. The \( D \)-statistic for batch \( i \) calculated at time interval \( k \) is given by:

\[
D_{i,k} = t_i^T S^{-1}_k t_i \tag{16}
\]

Where \( S_k \) is the variance-covariance matrix of the scores \( T_k \). The \( D \)-statistic represents the Mahalanobis distance from the center of the model plane towards the projected sample in this plane.

The calculation of the \( SPE \)-statistic is not much different from the \( Q \)-statistic. The term \( SPE \) stands for sum of squared prediction error and is also based on the residuals. The \( Q \)-statistic was based on the residuals calculated for a finished batch.
This $SPE$-statistic is based on the sum of squared residuals of the current part (cp) of the calculated residuals. Suppose a batch evolved until 10 time intervals and the residuals for this batch are calculated using the evolving model. The residual part that is concerned with the measurement at $k = 10$ is taken out. This part is referred to as the current part (cp). From the current part, the sum of squared residuals is calculated and forms the $SPE$-statistic. Then, at time point eleven, the residuals are calculated again and the current part is removed and the $SPE$ is calculated again. Thus, the $SPE$ for batch $I$ at time interval $k$ is calculated as follows:

$$SPE_{I,k} = \mathbf{e}_i^\mathbf{e}_i$$

(17)

The $SPE$-statistic represents a part of the Euclidian distance from the perpendicular projection of the sample to the model plane. The reason for this is that in this way instantaneous deviations in the residual space are detected. If the entire residual matrix would be used, deviations in the residuals might not be detected or much later because of the contributions from previous observations (averaging out effect).

Once the test statistics are calculated from the NOC data, the control charts can be constructed. These control charts are being used for monitoring completely new batches in the application phase. The control limits for the $D$-chart are found according to equation 9 with $J$ is replaced by $R$. Like the $Q$-statistic, the $SPE$ follows a weighted chi-square distribution $g \cdot \chi^2(h)$. The residuals are fitted to this chi-square distribution where the weight $g$ and degrees of freedom $h$ are optimised. From this distribution, the UCL can be obtained. The control limits of the $SPE$-chart always vary over time, since at every time interval the parameters of the chi-square distribution are estimated using different residuals.

### 2.4 A: Application phase

In the last phase of the ITA-trajectory, new batches are monitored denoted as $x_{nr}$. These batches are independent from the NOC data. Preferably, these new batches are operated under the same conditions as the NOC data. If that is the case, the variation in the new batches is comparable to the NOC data. However, if the variation suddenly changes because of a process fault, fast detection of this process upset is
desired. In the following, it is explained how new batches are monitored in the $D$ and $SPE$-chart that were constructed in the training phase.

### 2.4.1 On-line Monitoring

It is explained in the training phase how NOC batches can be described by latent variables (scores and residuals) in a dimension-reduced space. This dimension-reduced space is defined by the model loadings $P$. Since a time evolving model is used, this dimension-reduced space differs at every time interval ($P_k$). From the residuals and scores of the NOC batches, the $SPE$ and $D$-control charts are constructed. In order to compare a new batch $x_{ew}$ with the NOC batches in the dimension reduced space, the measurements up until time interval $k$ ($x_{k,ew}$) are projected onto this space. Once the residuals and scores for the new batch are calculated, the $SPE$ and $D$-statistic can be calculated. The calculation of the $SPE$ and $D$-statistic using a time evolving model reads as follows:

$$
\begin{align*}
    t_{k,ew} &= (P_k P_k^T)^{-1} P_k x_{k,ew} = P_k^T x_{k,ew} \\
    \rightarrow \quad D &= t_{ew}^T S_{ew}^{-1} t_{ew} \\
    (\text{because } P_k P_k = I \text{ by construction})
\end{align*}
$$

These statistics then are plotted in the corresponding control charts. If the test statistics do not exceed the control limit, the process is said to be in-control. If this is not the case, further action is needed to investigate what caused the test statistic to exceed the control limit.

#### 2.4.1.1 Fault detection and Fault Diagnosis

In process monitoring, an important question is: how fast is a process fault detected in the control charts? The speed of detection is dependent of the quality of the model and the historical data. If fault detection is the most important objective of batch monitoring, it does not really matter if the fault is detected in the $SPE$ or $D$ chart. The detection power of a model is often tested by projection of a faulty batch. For such a tracer batch, the time of the process fault is exactly known. Then, this batch is
monitored and the time it takes to detect the fault is examined. The Action Signal Time (AST) is used to quantify the detection power of the model to pinpoint fault batches. The AST is defined as the time between the introduction of an error and the out-of-control signal.

Next, if a fault is detected, the cause is searched for. This is referred to as fault diagnosis. In statistical batch process monitoring, fault detection plays an important role. The test statistic itself only provides information whether the new batch is statistically in or out of control. However, several techniques exist to investigate the physical nature of the fault that causes the control chart to signal. First, the SPE-chart signals faults of a different nature compared to the D-chart. Secondly, the contribution of the individual process variables to the D and SPE-statistic can be traced back. This will be discussed in the following.

2.4.1.1 Complementarity of SPE and D-chart
The result of the projection of a sample is expressed in terms of the SPE and D-statistic. The question therefore is: what kind of relation is there between a real process fault and the result of this fault in the SPE and D-statistic? To answer this question, the process measurements are categorised in two groups: i) faults that break the correlation structure or ii) faults that obey the correlation structure but have a more than normal variation. As an example, think of the following situation. The temperature ($T_{\text{pm}}$) and pressure ($P_{\text{pm}}$) are measured in the inner tube of a car tyre, as well as the ambient temperature ($T_{\text{a}}$). Now, as $T_{\text{a}}$ increases because of the sunny weather, so will $T_{\text{pm}}$ and $P_{\text{pm}}$. Thus, the process variables are correlated and behave according to simple physical laws. This is depicted in Figure 20.
The white bars represent the point in time where the process behaves under normal conditions. The black bars show the reaction of $T_{pr}$ and $P_{pr}$ when $T_r$ is changing. This is according to the correlation between the process variables. Besides, the variation of the increment of the process variables is considered as acceptable.

The first group of faults is the following. In extreme situations where the surrounding temperature may become very high (e.g. the car has been parked in the burning sun), the pressure and temperature in the inner tube might become dangerously high. Although the process variables behave according to the law of physics, the variation of the process variables is more intense than under normal circumstances. This is illustrated in Figure 20 as the patterned black bar. Such a fault where the variation of the process variables is abnormal high but the correlation between the process variables remains intact will be denoted as intensified correlation for the remaining of this thesis. The other group of faults is of the following. Suppose the temperature in the inner tube $T_{pr}$ increases as a result of $T_r$, but because the valve of the car tube is leaking, the pressure $P_{pr}$ remains constant. This is not according to the correlation between the process variables, and therefore the correlation is broken. This can be seen from Figure 21. Such disturbances will be denoted as breakage of the correlation.

Often heard remarks in the literature (R. Dunia et al. [51], S. Albert & R.D. Kinley [43], J.V. Kresta et al. [10], P. Nomikos & J.F. MacGregor [6]) are that abnormal variation that still obeys the correlation structure of the process variables is described by the scores ($D$-statistic) while new events not present in the NOC data will represent itself in the residuals ($SPE$-statistic).
Statistical batch process monitoring

It is believed that an intensified correlation (e.g. process shifts) is detected in the $D$-statistic while the breakage of the correlation (e.g. sensor failure) is detected in the $SPE$-statistic and that both charts are therefore complementary. However, this is a simplification of how a process fault manifests itself in the control charts as will be discussed later (Section 0).

2.4.1.1.2 Contribution plots

An alarm signal in the control charts will tell the user that the process is no longer operating under the specified operating conditions. However, it is not only important to detect that the process is deviating, it is also important to search for the process variables responsible for the alarm and to determine the cause. One of the tools to search for the responsible process variables are contribution plots (P. Miller et al. [16], J.A. Westerhuis et al. [17]). Contribution plots compute the contribution of a single process variable to the monitoring statistics. The process variables with the largest contribution are responsible for the out of control signal and should be analysed further to determine the cause.

The contribution is computed for both the $SPE$-chart and $D$-chart. The contribution of process variable $j$ to the $SPE$ is computed by

$$\epsilon_{j,k}^{SPE} = \epsilon_{j,k,raw}^2$$

which is simply the squared residual of process variable $j$ at observation $k$. The contribution $\epsilon$ of the process variable $j$ to the $D$-statistic is given by

$$\epsilon_{j,k}^{D} = \sum_{j=k}^{K} t^TS_k^{-1} [x_{j,k}^T p_{j,k} (P_{2,k} P_{2,k})^{-1}]^T$$

Figure 21

Breakage of the correlation.
where $p_{jk}$ is the $j,k$-th row of model vector $P_k$ and $x_{jk}$ is the $j$-th element of the observation vector $x_k$.

Once the contributions are computed, the responsible process variables are examined and the cause of disturbance is explored and if possible removed.

**2.4.1.1.3 Example: monitoring a faulty batch with impurity of the feed**

In this example, a new batch is monitored for the SBR process of which it is known that a feed impurity occurred halfway the batch run. The control charts for this batch are given in the following figure:

![SPE and D-chart of a faulty batch.](image)

The first deviations are observed in the $SPE$-chart around $k = 15$ and $k = 35$. There, a few observations cross the 95 % confidence limits. However, a clear out of control signal is given starting around time interval $k = 100$. The fault is well detected in the $SPE$-chart. The AST (95% confidence limit) for the $SPE$-chart is 6 observation units.
That is, after six observations after the introduction of the impurity an out of control signal is given.

A contribution plot with an approximate confidence limit for the SPE-chart at time interval \( k = 106 \) (marked with the circle) is given. The contributions reveal that process variable four, five and six and nine have a high contribution to the SPE-statistic (see Figure 23).

The process variables four, five and six represent the temperature measurement inside the reactor vessel and process variable nine the energy release. In fact, around time interval hundred an impurity of 50% above the normal level enters the reactor. As a result, unwanted side reactions add extra heat to the reaction. As a result, the instantaneous rate of energy release (process variable nine) is much higher as expected. This example shows how process upsets can be detected and how contribution plots can help to find the cause of the process fault.

2.4.1.2 Process improvement

Once a process upset is detected and diagnosed, the process operator can take further action. It was already explained that process control of batch processes is difficult. However, if it turned out that a sensor failure occurred somewhere along the batch, it is not necessary to adjust the batch since only the equipment is malfunctioning. Sometimes the feed for semi batch processes can be adjusted, if possible, to get the process within specifications again. When no control actions are possible, it can be decided to terminate the batch and start a new one.