Chapter 6

Statistics Decomposition and Monitoring in Original Variable Space

The traditional process monitoring method first projects the measured process data into the principle component subspace (PCS) and the residual subspace (RS), then calculates T^2 and SPE statistics to detect the abnormality. However, the abnormality by these two statistics are detected from the principle components of the process. Principle components actually have no specific physical meaning, and do not contribute directly to identify the fault variable and its root cause. Researchers have proposed many methods to identify the fault variable accurately based on the projection space. The most popular is contribution plot which measures the contribution of each process variable to the principal element (Wang et al. 2017; Luo et al. 2017; Liu and Chen 2014). Moreover, in order to determine the control limits of the two statistics, their probability distributions should be estimated or assumed as specific one. The fault identification by statistics is not intuitive enough to directly reflect the role and trend of each variable when the process changes.

In this chapter, direct monitoring in the original measurement space is investigated, in which the two statistics are decomposed as a unique sum of the variable contributions of the original process variables, respectively. The monitoring of the original process variables is direct and explicit in the physical meaning, but it is relatively complicated and time consuming due to the need to monitor each variable in both SPE and T^2 statistics. To address this issue, a new combined index is proposed and interpreted in geometric space, which is different from other combined indices (Qin 2003; Alcalá and Qin 2010). The proposed combined index is an intrinsic method. Compared with the traditional latent space methods, the combined index-based monitoring does not require the prior distribution assumption to calculate the control limits. Thus, the monitor complexity is reduced greatly.

6.1 Two Statistics Decomposition

According to the traditional PCA method, the process variables $\mathbf{x}$ could be divided into two parts: principal component $\hat{\mathbf{x}}$ and the residual $\mathbf{e}$:

$$\mathbf{x} = \mathbf{t}\mathbf{P}^T + \mathbf{e} = \hat{\mathbf{x}} + \mathbf{e}, \quad (6.1)$$

where $\mathbf{P}$ is the matrix associated with the loading vectors that define the latent variable space, $\mathbf{t}$ is the score matrix that contains the coordinates of $\mathbf{x}$ in that space, and $\mathbf{e}$ is the matrix of residuals. T^2 and SPE statistics are used to measure the distance from the new data to the model data. Generally, T^2 and SPE statistics should be analyzed simultaneously so that the cumulative effects of all variables can be utilized. However, most of the literatures have only considered the decomposition of T^2 . Therefore, this chapter considered the SPE statistical decomposition to obtain the original process variables monitored in T^2 and in the SPE statistical space.

6.1.1 T^2 Statistic Decomposition

The statistic can be reformulated as follows:

$$T^2 := \mathbf{D} = \mathbf{t}\mathbf{A}^{-1}\mathbf{t}^T = \mathbf{x}\mathbf{P}\mathbf{A}^{-1}\mathbf{P}^T\mathbf{x}^T = \mathbf{x}\mathbf{A}\mathbf{x}^T = \sum_{i=1}^J \sum_{j=1}^J a_{i,j} x_i x_j \geq 0, \quad (6.2)$$

where $\mathbf{A} = \mathbf{P}\mathbf{A}^{-1}\mathbf{P}^T \geq 0$, $\mathbf{A}^{-1}$ is the inverse of the covariance matrix estimated from a reference population, and $a_{i,j}$ is the element of matrix $\mathbf{A}$.

One of the T^2 statistic decompositions (Birol et al. 2002) is given as follows:

$$\begin{aligned} \mathbf{D} &= \sum_{k=1}^J \frac{a_{k,k}}{2} \left[(x_k - x_k^*)^2 + (x_k^2 - x_k^{*2}) \right] \\ &= \sum_{k=1}^J a_{k,k} [(x_k^2 - x_k^* x_k)] \\ &= \sum_{k=1}^J c_k^D. \end{aligned} \quad (6.3)$$

The x_k^* is given as follows:

$$x_k^* = - \frac{\sum_{\substack{j=1 \\ j \neq k}}^J a_{k,j} x_j}{a_{k,k}},$$

where the c_k^D is the decomposed T^2 statistic of each variable x_k . Next, the T^2 statistic of each variable x_k can be calculated as follows:

$$c_k^D = a_{k,k} [(x_k^2 - x_k^* x_k)]. \quad (6.4)$$

The detailed T^2 statistic decomposition process is not shown in here, details can be found in Alvarez et al. (2007, 2010).

6.1.2 SPE Statistic Decomposition

The SPE statistic, which reflects the change of the random quantity in the residual subspace, also has a quadratic form:

$$\begin{aligned} \text{SPE} &:= \mathbf{Q} = \mathbf{e} \mathbf{e}^T = \mathbf{x} (\mathbf{I} - \mathbf{P} \mathbf{P}^T) (\mathbf{I} - \mathbf{P} \mathbf{P}^T)^T \mathbf{x}^T \\ &= \mathbf{x} \mathbf{B} \mathbf{x}^T = \sum_{i=1}^J \sum_{j=1}^J b_{i,j} x_i x_j, \end{aligned} \quad (6.5)$$

where $\mathbf{B} = (\mathbf{I} - \mathbf{P} \mathbf{P}^T) (\mathbf{I} - \mathbf{P} \mathbf{P}^T)^T$, $b_{i,j}$ is the element of matrix $\mathbf{B}$, and $b_{i,j} = b_{j,i}$. Similar to the decomposition of T^2 statistic, SPE statistics can also be decomposed into a series of new statistic of each variable.

Firstly, the SPE statistic $\mathbf{Q}$ can be reformulated in terms of a single variable x_k :

$$\mathbf{Q} = Q_k = b_{k,k} x_k^2 + \left(2 \sum_{j=1, j \neq k}^J b_{k,j} x_j \right) x_k + \sum_{i=1, i \neq k}^J \sum_{j=1, j \neq k}^J b_{i,j} x_i x_j. \quad (6.6)$$

The minimum value of Q_k can be calculated as

$$\frac{\partial Q_k}{\partial x_k} = 2b_{k,k} x_k^* + 2 \sum_{j=1, j \neq k}^J b_{k,j} x_j = 0 \Rightarrow x_k^* = - \sum_{j=1, j \neq k}^J b_{k,j} x_j / b_{k,k} \quad (6.7)$$

$$Q_k^{\min} = -b_{k,k} x_k^{*2} + \sum_{i=1, i \neq k}^J \sum_{j=1, j \neq k}^J b_{i,j} x_i x_j. \quad (6.8)$$

The difference between the SPE statistic of x_k and $Q_k^{\min}$ is

$$\mathbf{Q} - Q_k^{\min} = b_{k,k} (x_k - x_k^*)^2. \quad (6.9)$$

The sum of the $Q_k^{\min}$ for $k = 1, 2, \dots, J$ is

$$\begin{aligned}
\sum_{k=1}^J Q_k^{\min} &= \sum_{k=1}^J \left(-b_{k,k} x_k^{*2} + \sum_{i=1, i \neq k}^J \sum_{j=1, j \neq k}^J b_{i,j} x_i x_j \right) \\
&= (J-2) Q + \sum_{k=1}^J b_{k,k} (x_k^2 - x_k^{*2}).
\end{aligned} \tag{6.10}$$

The SPE statistic obtained from (6.10) can be evaluated as the sum of the contributions of each variable x_k :

$$\begin{aligned}
Q &= \sum_{k=1}^J \frac{b_{k,k}}{2} \left[(x_k - x_k^*)^2 + (x_k^2 - x_k^{*2}) \right] \\
&= \sum_{k=1}^J b_{k,k} \left[(x_k^2 - x_k^* x_k) \right] \\
&= \sum_{k=1}^J q_k^{\text{SPE}}.
\end{aligned} \tag{6.11}$$

The original process variables of the SPE statistic are used to monitor the system status:

$$q_k^{\text{SPE}} = b_{k,k} \left[(x_k^2 - x_k^* x_k) \right]. \tag{6.12}$$

So the novel SPE statistic can be evaluated as a unique sum of the contributions of each variable q_k^{SPE} ($k = 1, 2, \dots, J$), which is used for original process variable monitoring.

6.1.3 Fault Diagnosis in Original Variable Space

Similar to other PCS monitoring strategies, the proposed original variable monitoring technique consists of two stages that are executed offline and online. Firstly, the control limits of the two statistics (T^2 and SPE) for each time interval are determined by reference population of normal batches in the offline stage. Next, two statistics are calculated at each sampling during the online stage. If one of statistics exceeds the established control limit, then a faulty mode is declared.

The historical data of the batch process are composed of a three-dimensional array $\mathbf{X}(I \times J \times K)$, where I , J , and K are the number of batches, process variables, and sampling times, respectively. The three-dimensional process data must be unfolded into two-dimensional forms $\mathbf{X}_k(I \times J)$, $k = 1, 2, \dots, K$ before performing the PCA operation. The unfolding matrix $\mathbf{X}_k$ is normalized to zero mean and unit variance in each variable. The main nonlinear and dynamic components of the variable are still left in the scaled data matrix $\mathbf{X}_k$.

The normalized data matrix $\mathbf{X}_k$ is projected into principal component subspace by loading matrix $\mathbf{P}_k$ to obtain the scores matrix $\mathbf{T}_k$:

$$\mathbf{X}_k = \mathbf{T}_k \mathbf{P}_k^T + \mathbf{E}_k,$$

where $\mathbf{E}_k$ is the residual matrix. The two statistics associated with the i th batch for the j th variable in k th time interval are defined as $c_{i,j,k}^D$ and $q_{i,j,k}^{\text{SPE}}$.

The control limit of a continuous process can be determined by using the kernel density estimation (KDE) method. Another method has been used for calculating the control limit for batch process, which is determined by the mean and variance of each statistic (Yoo et al. 2004; Alvarez et al. 2007). The mean and variance of $c_{i,j,k}^D$ are calculated as follows:

$$\begin{aligned} \bar{c}_{j,k}^D &= \sum_{i=1}^I c_{i,j,k}^D / I \\ \text{var}(c_{j,k}^D) &= \sum_{i=1}^I (c_{i,j,k}^D - \bar{c}_{j,k}^D)^2 / (I - 1). \end{aligned} \quad (6.13)$$

The control limit of statistic $c_{i,j,k}^D$ is estimated as

$$c_{j,k}^{\text{limit}} = \bar{c}_{j,k}^D + \lambda_1 (\text{var}(c_{j,k}^D))^{1/2}, \quad (6.14)$$

where λ_1 is a predefined parameter. Similarly, the control limit of statistic is

$$q_{j,k}^{\text{limit}} = \bar{q}_{j,k}^{\text{SPE}} + \lambda_2 (\text{var}(q_{j,k}^{\text{SPE}}))^{\frac{1}{2}}, \quad (6.15)$$

where λ_2 is a predefined parameter,

$$\begin{aligned} \bar{q}_{j,k}^{\text{SPE}} &= \sum_{i=1}^I q_{i,j,k}^{\text{SPE}} / I \\ \text{var}(q_{j,k}^{\text{SPE}}) &= \sum_{i=1}^I (q_{i,j,k}^{\text{SPE}} - \bar{q}_{j,k}^{\text{SPE}})^2 / (I - 1). \end{aligned} \quad (6.16)$$

As above, the control limit calculation is very simple. Although the calculation increases, the extra calculations can be performed offline, there is no restriction during the online monitoring stage. The proposed monitoring technique corresponding to the offline and online stages is summarized as follows:

A. Offline Stage

1. Obtain the normal process data of I batches $\mathbf{X}$, unfold them into two-dimensional time-slice matrix $\mathbf{X}_k$, and then normalize the data.

2. Perform the PCA procedure on the normalized matrix X_k of each time slice and obtain the loading matrices P_k .
3. Calculate the statistics $c_{i,j,k}^D$ and $q_{i,j,k}^{SPE}$ of each variable in all the interval times for all batches, then calculate the variable contributions at each time interval using (6.4) and (6.12).
4. The control limits of statistics $c_{i,j,k}^D$ and $q_{i,j,k}^{SPE}$ are estimated as (6.14) and (6.15).

B. Online Stage

1. Collect new sampling time-slice data x_{new} , and then normalize based on the mean and variance of prior normal I batches data (modeling data).
2. Use P_k to calculate the new statistics $c_{i,j,k}^D$ and $q_{i,j,k}^{SPE}$ of new sampling, and judge whether these statistics exceed the control limit. If one of them exceeds the control limit, then fault identification is performed to find the faulty variable that exceeds the control limit much greater than others; if none of them exceeds the control limit, then the current data are normal.

6.2 Combined Index-Based Fault Diagnosis

The monitoring method in the original process variables can avoid some of the disadvantages of traditional statistic approach in the latent variable space, such as indirectly monitoring (Yoo et al. 2004). However, the original variable monitoring method is relatively complicated due to the monitoring of each variable in both SPE and T^2 statistics. It means that each variable should be monitored twice, which increases the calculation. Thus, a new combined index, composed of the SPE and T^2 statistics, is proposed to decrease monitoring complexity.

6.2.1 Combined Index Design

In this section, we use symbol $X(I \times J)$ to substitute the unfolding process data matrix $X_k(I \times J)$ for general analysis. Similarly, P_k, T_k, E_k are substituted by P, T, E . The process data X could be decomposed into PCS and RS when performing PCA:

$$X = TP^T + E = \hat{X} + E, \quad (6.17)$$

where $\hat{X}$ is the PCS and E is the RS. If the principal number is m , then a PCS with m -dimension and a RS with $(J - m)$ -dimension can be obtained. When new data x are measured, they are projected into the principal subspace:

$$t = xP. \quad (6.18)$$

Fig. 6.1 Graphical representation of T^2 and SPE statistics

The principal component (PC) score vector $\mathbf{t} (1 \times m)$ is the projection of new data $\mathbf{x}$ in the PCS. Subsequently, the PC score vector is projected back into the original process variables to estimate the process data $\hat{\mathbf{x}} = \mathbf{t} \mathbf{P}^T$. The residual vector $\mathbf{e}$ is defined as

$$\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}} = \mathbf{x} (\mathbf{I} - \mathbf{P} \mathbf{P}^T). \quad (6.19)$$

Residual vector $\mathbf{e}$ reflects the difference between new data $\mathbf{x}$ and modeling data $\mathbf{X}$ in the RS. A graphical interpretation of T^2 and SPE statistics is shown in Fig. 6.1.

To describe the statistics clearly in the geometry, the principal component subspace is taken as a hyperplane. The SPE statistic checks the model validity by measuring the distance between the data in the original process variables and its projection onto the model plain. Generally, the T^2 statistic is described by the Mahalanobis distance of the project point $\mathbf{t}$ to the projection center of normal process data, which aims to check if the new observation is projected into the limits of normal operation. The residual space is perpendicular to the principal hyperplane. The SPE statistic shows the distance from the new data $\mathbf{x}$ to the principal component hyperplane.

A new distance index φ from the new data to the principal component projection center of the modeling data is given in the following. It can be used for monitoring instead of the SPE and T^2 indicators. Consider the singular value decomposition (SVD) of the covariance matrix $\mathbf{R}_x = \mathbb{E}(\mathbf{X}^T \mathbf{X})$ for given normal data $\mathbf{X}$,

$$\mathbf{R}_x = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T,$$

where $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_m, \mathbf{0}_{J-m}\}$ is the eigenvalue of $\mathbf{R}_x$. The original loading matrix $\mathbf{U}_{J \times J}$ is a unitary matrix and $\mathbf{U} \mathbf{U}^T = \mathbf{I}$. Each column of the unitary matrix is a set of standard orthogonal basis in its span space. The basis vectors of principal component space and residual space divided from matrix $\mathbf{U}$ are orthogonal to each other. Furthermore,

$$\mathbf{U} = [\mathbf{P}, \mathbf{P}_e], \quad (6.20)$$

where $\mathbf{P} \in R^{J \times m}$ is the loading matrix. $\mathbf{P}_e \in R^{J \times (J-m)}$ can be treated as the loading matrix of residual space. Thus, $\mathbf{P}$ and $\mathbf{P}_e$ are presented by $\mathbf{U}$ as follows:

$$\mathbf{P} = \mathbf{U} \mathbf{F}_1, \quad \mathbf{P}_e = \mathbf{U} \mathbf{F}_2, \quad (6.21)$$

where

$$\mathbf{F}_1 = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{J-m} \end{bmatrix}_{J \times m}, \quad \mathbf{F}_2 = \begin{bmatrix} \mathbf{0}_m \\ \mathbf{I}_{J-m} \end{bmatrix}_{J \times m}, \quad (6.22)$$

where $\mathbf{I}_m$ and $\mathbf{I}_{J-m}$ are the m and $J - m$ dimension unit matrices, respectively, and $\mathbf{0}_m$ and $\mathbf{0}_{J-m}$ are the m and $J - m$ dimension zero matrices, respectively. Furthermore, the SPE and T^2 statistics are denoted by $\mathbf{U}$:

$$\begin{aligned} \mathbf{e} &= \mathbf{x} (\mathbf{I} - \mathbf{P} \mathbf{P}^T) = \mathbf{x} (\mathbf{U} \mathbf{U}^T - \mathbf{U} \mathbf{F}_1 \mathbf{F}_1^T \mathbf{U}^T) \\ &= \mathbf{x} (\mathbf{U} \mathbf{U}^T - \mathbf{U} \mathbf{E}_1 \mathbf{U}^T) = \mathbf{x} \mathbf{U} (\mathbf{I} - \mathbf{E}_1) \mathbf{U}^T = \mathbf{x} \mathbf{U} \mathbf{E}_2 \mathbf{U}^T, \end{aligned} \quad (6.23)$$

where

$$\mathbf{E}_1 = \begin{bmatrix} \mathbf{I}_m & \mathbf{0}_{m, J-m} \\ \mathbf{0}_{J-m, m} & \mathbf{0}_{J-m} \end{bmatrix}, \quad \mathbf{E}_2 = \begin{bmatrix} \mathbf{0}_m & \mathbf{0}_{m, J-m} \\ \mathbf{0}_{J-m, m} & \mathbf{I}_{J-m} \end{bmatrix}. \quad (6.24)$$

Define $\mathbf{y} = \mathbf{x} \mathbf{U}$, then

$$\begin{aligned} \text{SPE} &:= \mathbf{Q} = \mathbf{e} \mathbf{e}^T = \mathbf{x} \mathbf{U} \mathbf{E}_2 \mathbf{U}^T \mathbf{U} \mathbf{E}_2 \mathbf{U}^T \mathbf{x}^T \\ &= \mathbf{x} \mathbf{U} \mathbf{E}_2 \mathbf{U}^T \mathbf{x}^T = \mathbf{y} \mathbf{E}_2 \mathbf{y}^T = \sum_{i=m+1}^J y_i^2. \end{aligned} \quad (6.25)$$

Similarly, we can describe the T^2 statistic as follows:

$$\begin{aligned} T^2 &:= \mathbf{D} = \mathbf{t} \mathbf{\Lambda}_m^{-1} \mathbf{t}^T = \mathbf{x} \mathbf{P} \mathbf{\Lambda}_m^{-1} \mathbf{P}^T \mathbf{x}^T \\ &= \mathbf{x} \mathbf{U} \mathbf{F}_1 \mathbf{\Lambda}_m^{-1} \mathbf{F}_1^T \mathbf{U}^T \mathbf{x}^T = \mathbf{x} \mathbf{U} \mathbf{\Lambda}^{-1} \mathbf{U}^T \mathbf{x}^T \\ &= \mathbf{y} \mathbf{\Lambda}^{-1} \mathbf{y}^T = \sum_{i=1}^m y_i^2 \sigma_i^2, \end{aligned} \quad (6.26)$$

where

$$\mathbf{\Lambda}_m^{-1} = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\}, \quad \mathbf{\Lambda}^{-1} = [\mathbf{\Lambda}_m^{-1}, \mathbf{0}_{(J-m) \times (J-m)}].$$

The new combined index could be obtained directly by composing the two statistics as

$$\varphi = \mathbf{D} + \mathbf{Q} = \sum_{i=1}^m y_i^2 \sigma_i^2 + \sum_{i=m+1}^J y_i^2. \quad (6.27)$$

It is proved via mathematical illustration that the two decomposed statistics can be geometrically added together directly. This result demonstrates that T^2 and SPE statistic can be combined primarily and that is an intrinsic property. Thus, the combined index is a more general and geometric representation compared with the other combined index. The monitoring strategy with the novel index is introduced in the next subsection.

6.2.2 Control Limit of Combined Index

In Sect. 6.1, the T^2 and SPE statistics are decomposed into two new statistics for each variable. To reduce the calculation of process monitoring, the two new statistics are combined into a new statistic φ to monitor the process.

$$\varphi_{i,j,k} = c_{i,j,k}^D + q_{i,j,k}^{\text{SPE}}, \quad (6.28)$$

where $\varphi_{i,j,k}$ is the combined statistic at sampling time k for the j th variable. The method mentioned in Sect. 6.1.3 can be used to calculate the control limit of the new statistic,

$$\varphi_{j,k}^{\text{limit}} = \bar{\varphi}_{j,k} + \kappa(\text{var}(\varphi_{j,k}))^{1/2}, \quad (6.29)$$

where κ is a predefined parameter, and

$$\begin{aligned} \bar{\varphi}_{j,k} &= \sum_{i=1}^I \varphi_{i,j,k} / I \\ \text{var}(\varphi_{j,k}) &= \sum_{i=1}^I (\varphi_{i,j,k} - \bar{\varphi}_{j,k})^2 / (I - 1). \end{aligned} \quad (6.30)$$

The online process monitoring can be performed according to comparing the new statistic and its control limit. There are several points to highlight for readers when the proposed control limit is used. Firstly, the mean and variance may be inaccurate for a small number of samples. As a result, a sufficient number of training samples should be collected during the offline stage. Secondly, the predefined parameter is important and it is designed by the engineers according to the actual process conditions. The tuning method regarding κ is similar to the Shewhart control chart. Equation (6.29) illustrates that the effect of variance depends on the predefined parameter κ and the fluctuation of control limits also relies on it on each sample. For example, the control limit is smooth when κ is selected to be a smaller value, and the control limit fluctuates when κ is selected to be a larger value.

If the combined statistic of the new sample has a significant difference from those of the reference data set, then a fault is detected. As a result, a fault isolation procedure is set up to find the fault roots. This fault response process is one of advantages in original process variable monitoring as each variable has a unique formulation and physical meaning. The proposed monitoring steps are similar as that in Sect. 6.1.2.

6.3 Case Study

A fed-batch penicillin fermentation process is considered in case study, and its detailed mathematical model is given in Birol et al. (2002). A detailed description of the fed-batch penicillin fermentation process is available in Chap. 4.

6.3.1 Variable Monitoring via Two Statistics Decomposition

Firstly, the original process variable monitoring algorithm mentioned in Sect. 6.1.2 is tested. The monitoring results of all variables would be interminable and tedious, so only several typical variables are shown here for demonstration or comparison. The monitoring result of variable 1 in a test normal batch is shown in Fig. 6.2. None of the two statistics ($c_{1,k}^D$ and $q_{1,k}^{SPE}$) exceeds its control limit, and the statistics ($c_{j,k}^D$ and $q_{j,k}^{SPE}$, $j = 2, \dots, 11$) of all the other variables do not exceed the control limits as well. The monitoring results of other variables are similar to that of variable 1, so we omitted them due to the restriction of the book length. These results show that proposed algorithm do not have a false alarm when it is used to monitor the normal batch.

Next, the fault batch data are used to test the proposed monitoring algorithm of the original process variables, and two types of faults are chosen here.

Fault 1: step type, e.g., a 20% step decrease is added in variable 3 at 200–250h.

Fig. 6.2 Original variables monitoring for normal batch (variable 1)

Fig. 6.3 Monitoring result for Fault 1 (variable 1)

The monitoring results are shown as follows. Figure 6.3 shows the monitoring result of variable 1 for fault 1, the statistics changes obviously during the fault occurrence. However, the statistics do not exceed the control limit, i.e., the process status exhibits changes, but variable 1 is not the fault source. The monitoring results of variables 2, 4, 8, 9, and 11 are almost the same as the result of variable 1, and these results are not presented here.

The monitoring results of variable 3 and variable 5 are shown in Figs. 6.4 and 6.5, respectively. Both of the variable statistics exceed the control limit at the sampling time 200h. Regarding the other variables of 6, 7, and 10, the statistics of these variables also exceed the control limit, and the simulation results of these variable are nearly the same as that of variable 5 (the results are not presented here).

The question is: which variable is the fault source, variable 3, 5, or others? From the amplitude of Figs. 6.4 and 6.5, it is easy to see that the two statistics for variable 3 exceed the control limits to a much greater extent than those for variable 5 and other variables. In particular, the Q statistic of variable 3 is 40 times greater than its control limit. From this perspective, variable 3 can be concluded to be the fault source, as it makes contribution to the statistics obviously. Note that there is no smearing effect in the proposed method. The smearing effect means that non-faulty variables exhibit larger contribution values, while the contribution of faulty variables is smaller. Because the statistics are decomposed into a unique sum of the variable contributions, each monitoring figure is plotted against the decomposed variable statistics. Furthermore, the proposed method may identify several faulty variables if they have larger contributions at close magnitudes.

Fig. 6.4 Monitoring result for Fault 1 (variable 3)

Fig. 6.5 Monitoring result for Fault 1 (variable 5)

Fig. 6.6 Relative contribution rate of R_c for Fault 1

To confirm the monitoring conclusion, the relative statistical contribution rate of the j th variable at time k is defined as

$$R_c^{j,k} = c_{j,k}^D / \sum_{j=1}^J c_{j,k}^D$$

$$R_q^{j,k} = q_{j,k}^{\text{SPE}} / \sum_{j=1}^J q_{j,k}^{\text{SPE}}.$$

The relative statistic contribution rates of 11 variables are shown in Figs. 6.6 and 6.7. It is clear that variable 3 is the source of Fault 1. It is found that variables 9, 10, and 11 still have the higher contribution when the fault is eliminated because the fault in variable 3 causes the change of the other process variables. The effects on whole process still continue, even if the fault is eliminated, and the fault variable evolves from the original variable 3 to other process variables.

Fault 2: ramp type, i.e., fault involving a ramp increasing with a slope of 0.3 in variable 3 at 20–80 h.

The two monitor statistics of variable 3 are shown in Figs. 6.8 and 6.9. It can be seen that both of the two statistics exceed the control limits at approximately 50 h. The alarming time lags relative to the fault occurrence time (approximately 20 h) are found because this fault variable changes gradually. When the fault is eliminated after 80 h, the relationship among the variables changes back to normal. The T^2 statistic obviously declines under the control limit, while the SPE statistic still exceeds the control limit because the error caused by Fault 2 still exists.

Fig. 6.7 Relative contribution rate of R_q for Fault 1

Fig. 6.8 Fault 2 monitoring by c statistic (variable 3)

Fig. 6.9 Fault 2 monitoring by q statistic (variable 3)

6.3.2 Combined Index-Based Monitoring

The same test data in Sect. 6.3.1 are used to test monitoring effectiveness of the new combined index. Considering a normal batch, the monitoring result of φ statistic is shown in Fig. 6.10. Variable 1 is still monitored in this section, as was the case in Sect. 6.3.1 for comparison. It is shown that the new index φ of variable 1 is far below its control limit, as is the case for the new index values of the other variables. This method shows some good performances, and the number of false alarms is zero in normal batch monitoring. The new index is more stable than the two statistics, and it is easy to observe for operators.

Fault 1: step type, e.g., a 20% step decrease is added in variable 3 at 200–250h.

The new statistic φ of variable 1 does not exceed the control limit in Fig. 6.11, although it changes from 200h to 250h during the fault. The values of new statistic φ of variables 2, 4, 8, 9, and 11 also do not exceed the control limit. The corresponding monitoring statistics are omitted here. Thus, these variables have no direct relationship with the fault variable, i.e., they are not the fault source.

Furthermore, the monitoring results of variables 3 and 5 are shown in Figs. 6.12 and 6.13, respectively. The value statistics of variables 3 and 5 exceed their control limits obviously, as well as those of variables 6, 7, and 10. As discussed in Sect. 6.3.1, one can see that the statistic φ of variable 3 changes to a greater extent than other variables, so variable 3 is the potential fault source. This result shows that the proposed approach is an efficient technique for fault detection.

Fig. 6.10 Original variables monitoring based on combined index for normal batch (variable 1)

Fig. 6.11 Fault 1 monitoring based on combined index (variable 1)

Fig. 6.12 Fault 1 monitoring based on combined index (variable 3)

Fig. 6.13 Fault 1 monitoring based on combined index (variable 5)

The relative contribution of the new statistic is used to confirm the fault source, which is defined as

$$R_{\varphi}^k = \varphi_{j,k} / \sum_{j=1}^J \varphi_{j,k}.$$

The relative contribution of variable 3 is nearly 100%, as shown in Fig. 6.14. So variable 3 is confirmed as the fault source. It is found that variables 9, 10, and 11 still have a higher contribution when the fault is eliminated because the fault in variable 3 causes the change of the other process variables and the effect on whole process still continues, even if the fault is eliminated.

Note that the relative contribution plot (RCP) is an auxiliary tool to locate the fault roots. It is only used for comparison with the proposed monitoring method to confirm diagnostic conclusions. Furthermore, the RCP is completely different from the traditional contribution diagram in this work. The RCP in this work is calculated using the original process variables, i.e., there is no smearing effect of the RCP. The contribution of each variable is independent of the other variables. Therefore, the proposed method is a novel and helpful approach in terms of original process variable monitoring. Furthermore, the color map of the fault contribution is intuitive. As a result, the map will promote the operator's initiative to find the fault source, and engineers can find some useful information to avoid more serious accidents.

Fault 2: ramp type, i.e., fault involving a ramp increasing with a slope of 0.3 in variable 3 at 20–80 h.

Fig. 6.14 Relative contribution rate of φ statistic for Fault 1

Fig. 6.15 Fault 2 monitoring of variable 3 by φ statistic

The monitoring result of variable 3 is shown in Fig. 6.15. It can be seen that the new statistic φ exceeds the control limit at approximately 50 h, and then it falls below the control limit after 80 h. The result shows that the combined index can detect different faults.

6.3.3 Comparative Analysis

The monitoring performances of different methods are compared. Several performance indices are given to evaluate the monitoring efficiency. False alarm (FA) is the number of false alarms during the operation life. Time detected (TD) is the time that the statistic exceeds the control limit under the fault operation, which can represent the sensitivity.

The monitoring results of the proposed method are compared with that of the traditional sub-PCA method (Lu et al. 2004) in latent space and the soft-transition sub-PCA (Wang et al. 2013) to illustrate the effectiveness. The FA and TD results for other 12 faults are presented in Tables 6.1 and 6.2, respectively. Fault variable numbers (1, 2, and 3) represent the aeration rate, agitator power, and substrate feed rate, as shown in Chap. 4. The fault type and occurring time for the variables are given in Table 6.1, and those input conditions are as same as those in Sects. 6.3.1 and 6.3.2.

Table 6.1 Monitoring results of FA for other faults

Fault ID	Var. No.	Fault type	M/S (%)	Fault time (h)	Original variables monitoring			Trad. sub-PCA (Lu et al. 2004)	Soft sub-PCA (Wang et al. 2013)
					c	q	φ		
1	2	Step	−15	20	0	0	0	9	0
2	2	Step	−15	100	0	173	0	1	0
3	3	Step	−10	190	0	95	0	11	0
4	3	Step	−10	30	0	57	0	5	0
5	1	Step	−10	20	0	0	0	1	0
6	1	Step	−10	150	16	0	0	2	0
7	1	Ramp	−5	20	2	1	0	1	0
8	2	Ramp	−20	20	4	0	0	6	0
9	1	Ramp	−10	20	2	0	1	10	0
10	3	Ramp	−0.2	170	1	0	0	3	0
11	2	Ramp	−20	170	4	0	0	1	0
12	1	Ramp	−10	180	2	0	0	2	0

It can be seen from Table 6.1 that there are multiple false alarms applying the traditional sub-PCA method to detect faults, while the original process variable monitoring method shows less false alarms based on the combined index φ in this chapter. Among the three indices of the original spatial monitoring, the c and q statistics may have a large number of false alarms for different reasons, but the new combined index φ is more accurate because it can balance the two indices.

Table 6.2 indicates that the original process variable monitoring has accurate and timely detection results comparing with the other two detection methods. The detection delay is more than 10h for Fault 4, 7, 8 and 11 in the traditional sub-PCA and the soft-transition sub-PCA. Such a delay is inconceivable in a complex industrial process. While the difference between the detected time and the real fault time for the proposed approach is less than 10h, except for fault 4. This result is helpful and meaningful in practice. As a result, the proposed approach could provide more suitable process information to operators. Thus, the proposed monitoring method based on a combined index shows advantages of rapid detection and fewer false alarms compared with the traditional or soft-transition sub-PCA approaches, whose monitoring operation is in the latent space but not the original measurement space.

Table 6.2 Comparing the time of fault detected

Fault ID	Fault time (h)	Original process variables monitoring			Trad. sub-PCA (Lu et al. 2004)		Soft-trans. sub-PCA (Wang et al. 2013)	
		c	q	φ	SPE	T^2	SPE	T^2
1	20	20	20	20	20	None	20	28
2	100	100	100	100	100	101	100	100
3	190	191	190	190	190	213	190	199
4	30	45	45	45	81	45	48	45
5	20	20	20	20	20	48	20	20
6	150	151	150	150	150	151	150	151
7	20	27	26	25	28	41	28	40
8	20	30	26	26	44	34	31	45
9	20	24	22	23	21	28	24	30
10	170	171	170	170	170	173	171	171
11	170	179	175	175	177	236	181	195
12	180	184	182	182	185	185	184	188

6.4 Conclusions

A new multivariate statistical method for the monitoring and diagnosis of batch processes, which operates on the original process variables, was presented in this chapter. The proposed monitoring method is based on the decomposition of the T^2 and SPE statistics as a unique sum of each variable contribution. However, problems may arise if the number of variables is large when the original process variables technique is applied. To reduce the workload of the monitoring calculation, a new combined index was proposed. A mathematical illustration was given to prove that the two decomposed statistics can be added together directly. Compared to the traditional PCA method in latent space, the proposed method is sufficiently direct, and only one statistical index is utilized, thereby decreases the calculation burden.

The new original variable space monitoring method can detect a fault with a clear result based on each variable. The fault source can be determined directly from the statistical index rather than using the traditional contribution plot. Furthermore, the control limit of the new combined statistics is very simple, and it does not need to assume that it follows some probability distribution. The simulation results show that the new combined statistics can detect the fault efficiently. As the new statistic index is the combination of two decomposed statistics, it can avoid many problems introduced by the use of a single statistic, such as false alarms or missing alarms.

References

- Alcala CF, Qin SJ (2010) Reconstruction-based contribution for process monitoring with kernel principal component analysis. *Ind Eng Chem Res*, pp 7849–7857
- Alvarez CR, Brandolin A, Sanchez MC (2007) On the variable contributions to the D-statistic. *Chemom Intell Lab Syst* 88:189–196
- Alvarez CR, Brandolin A, Sanchez MC (2010) Batch process monitoring in the original measurement's space. *J Process Control* 20:716–725
- Birol G, Undey C, Cinar A (2002) A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng*, pp 1553–1565
- Liu JL, Chen DS (2014) Fault isolation using modified contribution plots. *Comput Chem Eng*, pp 9–19
- Lu N, Gao F, Wang F (2004) Sub-PCA modeling and on-line monitoring strategy for batch processes. *AIChE J*, pp 255–259
- Luo LJ, Bao SY, Mao JF, Tang D (2017) Fault detection and diagnosis based on sparse PCA and two-level contribution plots. *Ind Eng Chem Res*, pp 225–240
- Qin SJ (2003) Statistical process monitoring: basics and beyond. *J Chemom*, pp 480–502
- Wang J, Ge WS, Zhou JL, Wu HY, Jin QB (2017) Fault isolation based on residual evaluation and contribution analysis and contribution analysis. *J Franklin Inst* 354:2591–2612
- Wang J, Wei HT, Cao LL, Jin QB (2013) Soft-transition sub-PCA fault monitoring of batch processes. *Ind Eng Chem Res*, pp 9879–9888
- Yoo CK, Lee JM, Vanrolleghem PA, Lee IB (2004) On-line monitoring of batch processes using multiway independent component analysis. *Chemom Intell Lab Syst*, pp 15163

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.