

Chapter 4

Simulation Platform for Fault Diagnosis



The previous chapters have described the mathematical principles and algorithms of multivariate statistical methods, as well as the monitoring processes when used for fault diagnosis. In order to validate the effectiveness of data-driven multivariate statistical analysis methods in the field of fault diagnosis, it is necessary to conduct the corresponding fault monitoring experiments. Therefore this chapter introduces two kinds of simulation platform, Tennessee Eastman (TE) process simulation system and fed-batch Penicillin Fermentation Process simulation system. They are widely used as test platforms for the process monitoring, fault classification, and identification of industrial process. The related experiments based on PCA, CCA, PLS, and FDA are completed on the TE simulation platforms.

4.1 Tennessee Eastman Process

The original TE industrial process control problem was developed by Downs and Vogel in 1993. It is used for the open and challenging control-related topics including multi-variable controller design, optimization, adaptive and predictive control, non-linear control, estimation and identification, process monitoring and diagnosis, and education. TE process model is established according to the actual chemical process. It has been widely used as a benchmark for control and monitoring research process. Figure 4.1 shows the flow diagram of TE process with five major units: reactor, condenser, compressor, vaporliquid separator, and stripper. Four kinds of gaseous material *A*, *C*, *D*, and *E* are input for reaction. In addition, a small amount of inert gas *B* is contained besides the above feeds. The final products are three liquid including *G*, *H*, and *F*, where *F* is the by-product.

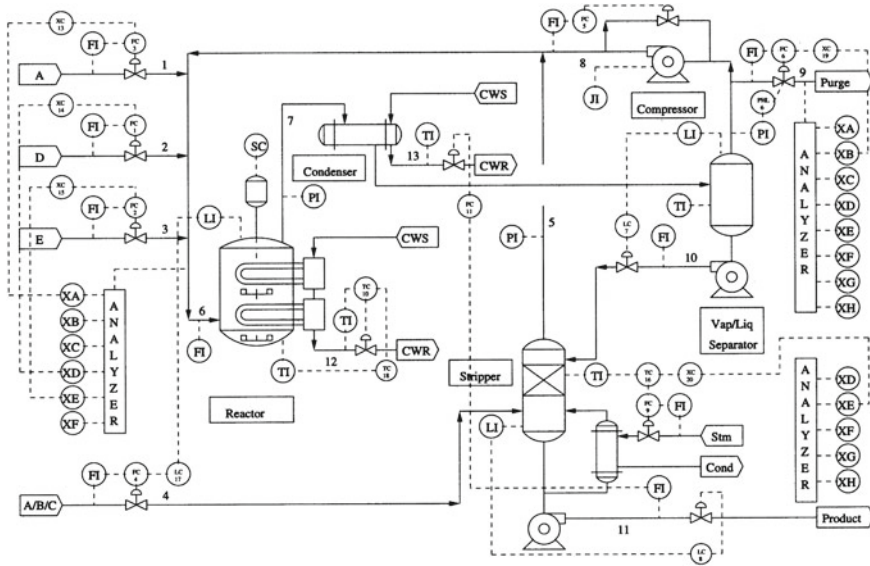
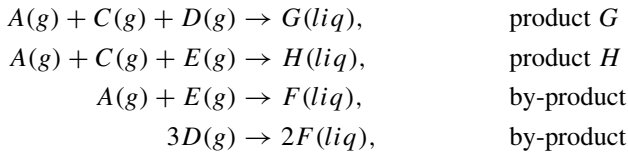


Fig. 4.1 Tennessee Eastman process



Briefly, TE process consists of two data modules: XMV module containing 12 manipulated variables (XMV(1)-XMV(12): $x_{23} - x_{34}$) and XMEAS module consisting of 22 process measured variables (XMEAS(1)-XMEAS(22): $x_1 - x_{22}$) and 19 component measured variables (XMEAS(23)-XMEAS(41): $x_{35} - x_{53}$), as listed in Tables 4.1 and 4.2.

In this book, the code provided is available on the website *online* at <http://depts.washington.edu/control/LARRY/TE/download.html>. Also, the code and data sets can be downloaded. The Simulink simulator allows an easy setting and generation of the operation modes, measurement noises, sampling time, and magnitudes of the faults. It is thus very helpful for the data-driven process monitoring study. 21 artificially disturbances (considered as faulty operations for fault diagnosis problem) in the TE process are shown in Table 4.3. In general, the entire TE data consists of training set and testing set, and each set includes 22 kinds of data under different simulation operations. Each kind of data has sampled measurements on 53 observed variables.

In the data set given in the web link above, d00.dat to d21.dat are training sets, and d00_te.dat to d21_te.dat are testing sets. d00.dat and d00_te.dat are samples under

Table 4.1 Monitoring variables in the TE process($x_1 - x_{34}$)

No.	Variable name	Units	No.	Variable name	Units
x_1	A feed (stream 1)	kscmh	x_{18}	Stripper temperature	°C
x_2	D feed (stream 2)	kg h^{-1}	x_{19}	Stripper steam flow	kg h^{-1}
x_3	E feed (stream 3)	kg h^{-1}	x_{20}	Compress work	KW
x_4	A and C feed (steam 4)	kscmh	x_{21}	Reactor cooling water outlet temperature	°C
x_5	Recycle flow (stream 8)	kscmh	x_{22}	Condenser cooling water outlet temperature	°C
x_6	Reactor feed rate (stream 6)	kscmh	x_{23}	D feed flow valve (stream 2)	%
x_7	Reactor pressure	kPa gauge	x_{24}	E feed flow valve (stream 3)	%
x_8	Reactor level	%	x_{25}	A feed flow valve (stream 1)	%
x_9	Reactor temperature	°C	x_{26}	A and C feed flow valve (stream 4)	%
x_{10}	Purge rate (stream 9)	kscmh	x_{27}	Compressor recycle valve	%
x_{11}	Product separator temperature	°C	x_{28}	Purge valve (stream 9)	%
x_{12}	Product separator level	%	x_{29}	Separator pot liquid flow valve (stream 10)	%
x_{13}	Product separator pressure	kPa gauge	x_{30}	Stripper liquid product flow valve (stream 11)	%
x_{14}	Product separator underflow (stream 10)	m^3h^{-1}	x_{31}	Stripper steam valve	%
x_{15}	Stripper level	%	x_{32}	Reactor cooling water flow valve	%
x_{16}	Stripper pressure	kPa gauge	x_{33}	Condenser cooling water flow valve	%
x_{17}	Stripper underflow (stream 11)	m^3h^{-1}	x_{34}	Agitator speed	

the normal operation conditions. The training samples of d00.dat are sampled under 25h running simulation. The total number of observations is 500. The d00_te.dat test samples are obtained under 48h running simulation, and the total number of observation data is 960. d01.dat–d21.dat (for training) and d01_te.dat–d21_te.dat (for testing) are sampled with different faults, in which the numerical label of the data set are corresponding to the fault type.

All the testing data set are obtained under 48h running simulation with the faults introduced at 8h. A total of 960 observations are collected, in which the first 160 observations are in the normal operation. It is worth to point out that the data sets

Table 4.2 Monitoring variables in the TE process($x_{35} - x_{53}$)

No.	Variable name	Stream	No.	Variable name	Stream
x_{35}	Composition <i>A</i>	6	x_{45}	Composition <i>E</i>	9
x_{36}	Composition <i>B</i>	6	x_{46}	Composition <i>F</i>	9
x_{37}	Composition <i>C</i>	6	x_{47}	Composition <i>G</i>	9
x_{38}	Composition <i>D</i>	6	x_{48}	Composition <i>H</i>	9
x_{39}	Composition <i>E</i>	6	x_{49}	Composition <i>D</i>	11
x_{40}	Composition <i>F</i>	6	x_{50}	Composition <i>E</i>	11
x_{41}	Composition <i>A</i>	9	x_{51}	Composition <i>F</i>	11
x_{42}	Composition <i>B</i>	9	x_{52}	Composition <i>G</i>	11
x_{43}	Composition <i>C</i>	9	x_{53}	Composition <i>H</i>	11
x_{44}	Composition <i>D</i>	9			

Table 4.3 Disturbances for the TE process

IDV	Process variable	Type
1	A/C feed ratio, B composition constant (stream 4)	Step
2	B composition, A/C feed ratio constant (stream 4)	Step
3	D feed temperature (stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss (stream 1)	Step
7	C header pressure loss—reduced availability (stream 4)	Step
8	A, B, C feed composition (stream 4)	Random
9	D feed temperature (stream 2)	Random
10	C feed temperature (stream 4)	Random
11	Reactor cooling water inlet temperature	Random
12	Condenser cooling water inlet temperature	Random
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	Valve position (stream 4)	Constant

once generated by Leoand et al. (2001) is widely accepted for process monitoring and fault diagnosis research. The data sets are smoothed, filtered, and normalized. The monitored variables are variables $x_1 - x_{53}$.

4.2 Fed-Batch Penicillin Fermentation Process

Fed-batch fermentation processes are widely used in the pharmaceutical industry. The yield maximization is usually considered as the main goal in the batch fermentation processes. The different characteristics of batch operation from the continuous operation include strong nonlinearity, non-stationary conditions, batch-to-batch variability, and strong time-varying conditions. These features result that the yield is difficult to predict. Therefore, the fault detection, classification, and identification of batch/fed-batch processes shows more difficulties compared with the continuous TE process.

The model of fed-batch penicillin fermentation process is described by Birol et al. (2002)

$$\begin{aligned} X &= f(X, S, C_L, H, T) \\ S &= f(X, S, C_L, H, T) \\ C_L &= f(X, S, C_L, H, T) \\ P &= f(X, S, C_L, H, T, P) \\ CO_2 &= f(X, H, T) \\ H &= f(X, H, T), \end{aligned}$$

where X , S , C_L , P , CO_2 , H and T are biomass concentration, substrate concentration, dissolved oxygen concentration, penicillin concentration, carbon dioxide concentration, hydrogen ion concentration for pH ($[H^+]$), and temperature, respectively. The corresponding detailed mathematical model is given in Birol et al. (2002).

The research group with the Illinois Institute of Technology has developed a dynamic simulation of penicillin production based on an unstructured model, PenSim V2.0. This model has been used as a benchmark for statistical process monitoring studies of batch/fed-batch reaction process. The flow chart of the fermentation process is depicted in Fig. 4.2. The fermentation unit consists of a fermentation reactor and a coil-based heat exchange unit. The pH and temperature are automatically controlled by two PID controllers by adjusting the flow rates of acid/base and cold/hot water. The glucose substrate is fed continuously into the fermentation reactor in open-loop operation in the fed-batch operation mode.

Fourteen variables are considered in PenSim V2.0 model, shown in Table 4.4: 5 input variables (1–4, 14) and 9 process variables (5–13). Since variables 11–13 are not measured online in industry, only 11 variables are monitored here.

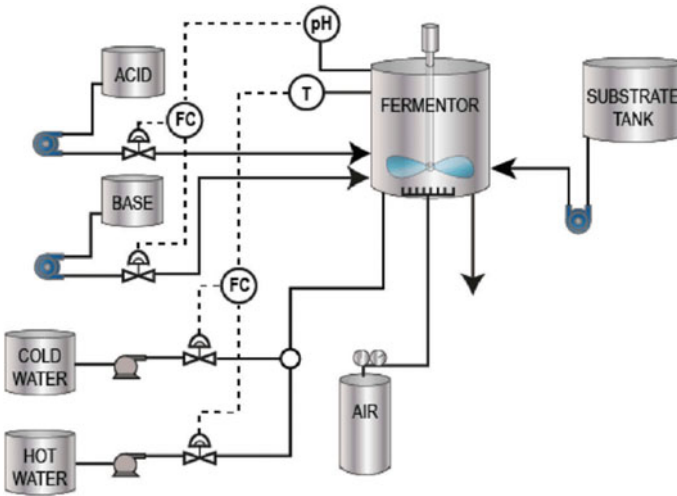


Fig. 4.2 Flow chart of the penicillin fermentation process

Table 4.4 Variables in penicillin fermentation process

No.	Variable
1	Aeration rate (L/h)
2	Agitator power input (W)
3	Substrate feed rate (L/h)
4	Substrate feed temperature (K)
5	Dissolved oxygen concentration (% saturation)
6	Culture volume (L)
7	Carbon dioxide concentration (mmol/L)
8	pH
9	Temperature in the bioreactor (K)
10	Generated heat (kcal/h)
11	Cooling water flow rate (L/h)
12	Penicillin concentration
13	Biomass concentration
14	Substrate concentration

4.3 Fault Detection Based on PCA, CCA, and PLS

This section tests the effectiveness of various multivariate statistical methods for the TE process. Faults in the standard TE data set are introduced at the 160 sampling. For comparison purposes, the normal operation data d00_te is chosen as to train the statistical model and faulty operation data d01_te-d21_te is used to test model and

detect fault. In the experiments for the PCA and PLS methods, the process variable matrix X consists of process variables (XMEAS (1–22)) and manipulated variables (XMV (1–11)). XMEAS (35) is used as the quality variable matrix Y for PLS. In the CCA experiment, the process variables (XMEAS (1–22)) are used as one data set, and the manipulated variables (XMV (1–11)) as another data set.

The fault detection rate (FDR) and false alarm rate (FAR) are defined as follows:

$$\begin{aligned} \text{FDR} &= \frac{\text{No.of samples}(J > J_{th}|f \neq 0)}{\text{total samples}(f \neq 0)} \times 100 \\ \text{FAR} &= \frac{\text{No.of samples}(J > J_{th}|f = 0)}{\text{total samples}(f = 0)} \times 100. \end{aligned} \quad (4.1)$$

Experiment and model parameters are determined as follows. The principal components of PCA are determined by the cumulative contribution of 90%. The number of principal components of PLS is selected as 6. T^2 and Q statistics are used to monitor process faults. It should be noted that in the monitoring of CCA, (3.18) and (3.19) are used as monitoring indices and the corresponding monitoring results are slightly different. For 21 fault types, the FDR for PCA, CCA, and PLS based on the control limit with 99% confidence level are shown in Table 4.5. It can be seen that the multivariate statistical methods listed in this section (including PCA, CCA, and PLS) can accurately detect the significant process faults.

Figures 4.3, 4.4, and 4.5 show the different monitoring results base on PCA, CCA, and PLS model for typical faults IDV(1), IDV(16), and IDV(20), respectively. Here, the black line is the statistic calculated from the real-time data and the red line is the normal statistic threshold from the offline model calculation.

It is easy to find that CCA has better detection for certain fault types from Table 4.5, such as faults IDV(10), IDV(16), IDV(19), and IDV(20). The monitoring results for faults IDV(16) and IDV(20) are shown in Figs. 4.4 and 4.5. Why does CCA show better detection capabilities than the other two methods in certain faults? Let's check the setting of process variable X for three methods. In contrast to PCA and PLS, CCA splits its X -space directly into two parts and extracts the latent variables by examining the correlation between these two parts, i.e., the latent variables extracted by CCA can better characterise the changes in the process.

4.4 Fault Classification Based on FDA

To further test the effectiveness of fault classification, samples from the 161th to the 700th of the 21 fault data sets and the normal data sets are used for training FDA model. The corresponding data from the 701th to the 960th samples are used to test FDA model and its classification ability. FDA in Sect. 2.2 is a classical method to validate the classification effect and identify the fault types. The following distance metric index is introduced to further quantify the difference between different faults:

Table 4.5 FDRs of PCA, CCA and PLS

IDV	PCA		CCA		PLS	
	T ²	SPE	T ₁ ²	T ₂ ²	T ²	SPE
1	99.13	99.88	99.38	99.63	99.75	99.38
2	98.38	95.13	95.63	96.13	98.63	97.75
3	1.00	3.00	0.25	0.50	3.75	1.88
4	50.88	99.88	100.00	97.38	40.63	96.88
5	23.75	23.88	100.00	100.00	25.50	25.88
6	99.00	100.00	100.00	100.00	99.25	100.00
7	100.00	100.00	100.00	83.00	99.13	100.00
8	97.00	86.25	87.00	92.25	96.88	96.75
9	1.50	2.00	0.13	0.13	2.13	2.25
10	27.88	36.13	78.75	79.38	57.00	31.25
11	52.50	61.63	77.00	56.88	41.88	65.75
12	98.38	90.25	97.00	99.00	99.00	96.75
13	93.75	95.13	94.38	94.25	95.50	94.25
14	99.88	98.88	100.00	99.88	99.88	100.00
15	1.25	2.00	0.63	0.75	4.50	1.13
16	12.13	36.25	85.00	86.63	29.75	19.25
17	79.50	95.88	91.38	95.25	80.13	89.75
18	89.13	90.50	89.50	89.50	89.50	89.50
19	11.63	16.50	84.38	84.25	1.63	13.38
20	31.13	52.75	70.38	75.50	41.75	45.38
21	41.25	48.75	26.63	36.88	56.38	43.00

$$D_2 = \left\| \text{FDA}_i - \text{FDA}_j \right\| ,$$

where FDA_i denotes the FDA feature vector of the i th fault.

The simulation results are shown in Fig. 4.6. The 22 kinds of data (including the normal operation and 21 faulty operation) can be roughly divided into two major categories: the first category is the faults that are significantly different from other faults, which contains faults IDV(2) (line with \diamond), IDV(6) (line with $*$), and IDV(18) (line with \circ); the other category is the set of faults whose characteristics are relatively close to each other.

The faults IDV(1), IDV(2), IDV(6), and IDV(20) are further analyzed. The FDA results for fault classification are shown in Fig. 4.7. The D_2 indices for these faults vary considerably, as the classification results clearly illustrated. Conversely, certain faults have very small differences in D_2 indices. For example, faults IDV(4), IDV(11), and IDV(14) have the similar FDA D_2 indices, shown in Fig. 4.8. These faults are difficult to classify accurately based on FDA model, as shown in Fig. 4.9.

Fig. 4.3 PCA, CCA, and PLS monitoring results for IDV(1)

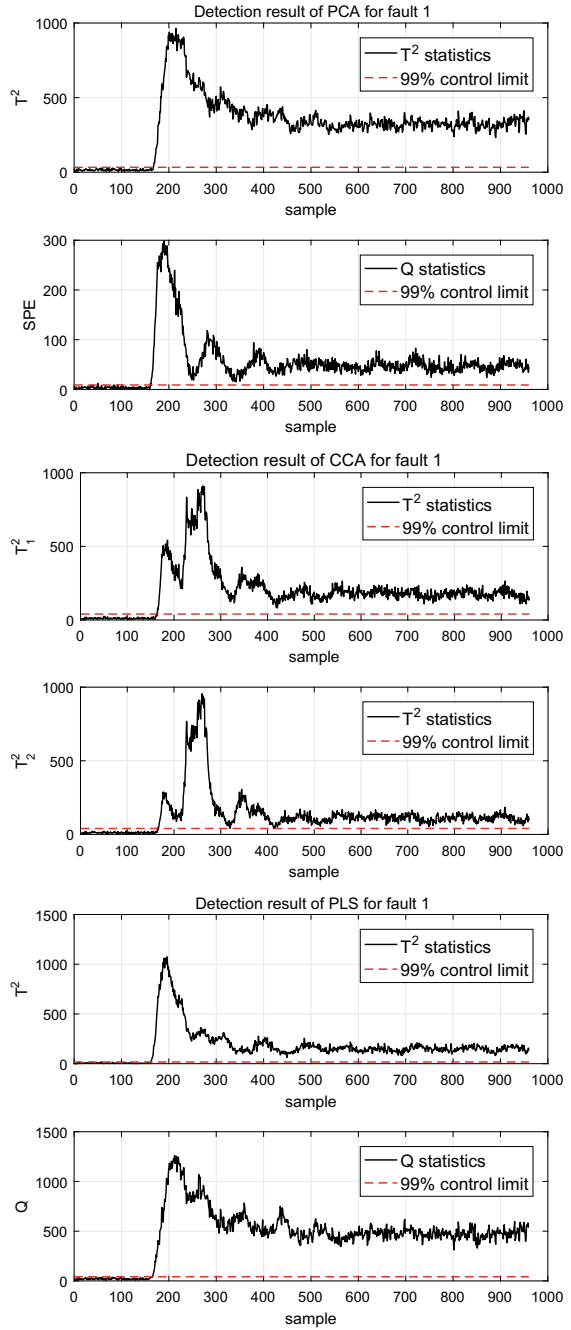


Fig. 4.4 PCA, CCA, and PLS monitoring results for IDV(16)

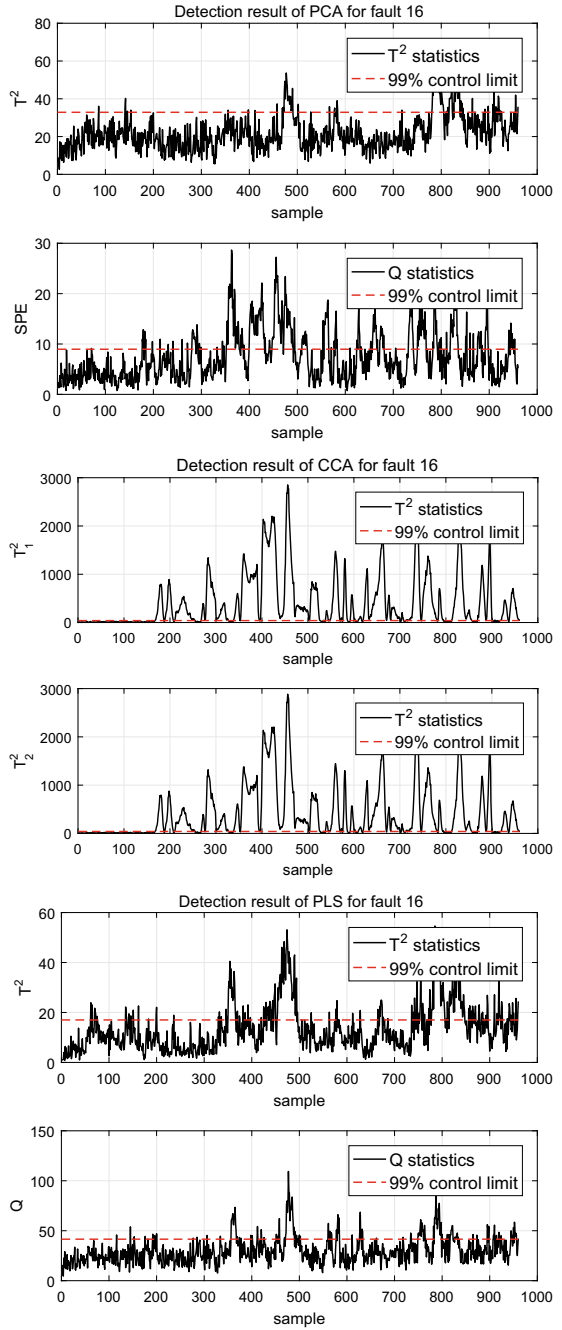
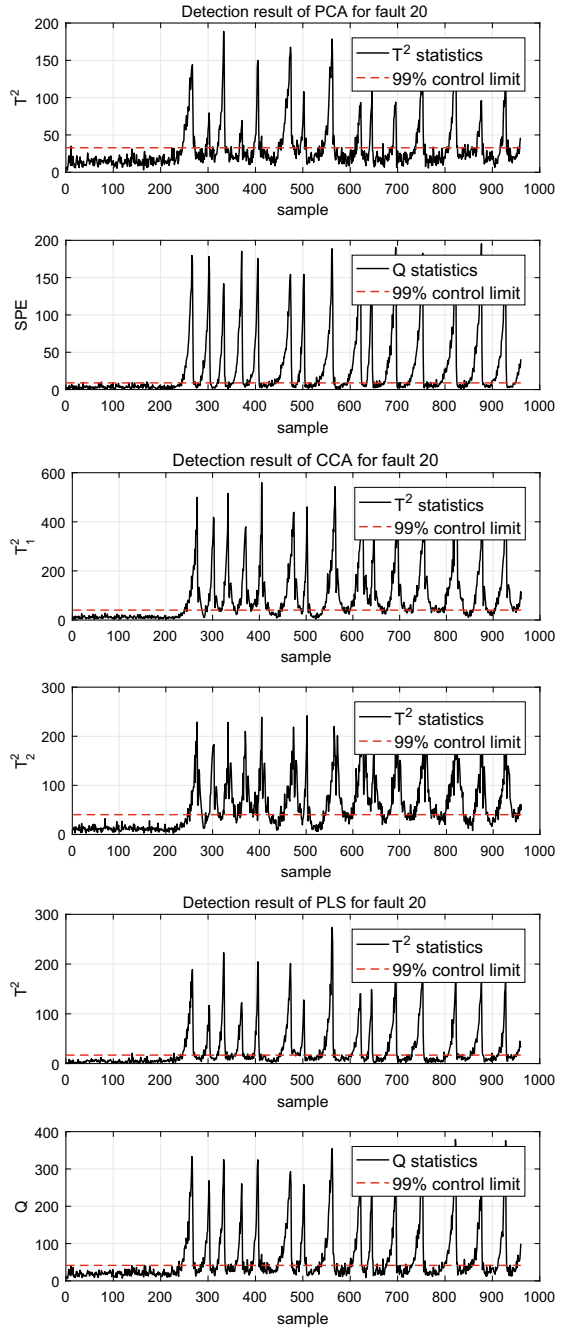


Fig. 4.5 PCA, CCA, and PLS monitoring results for IDV(20)



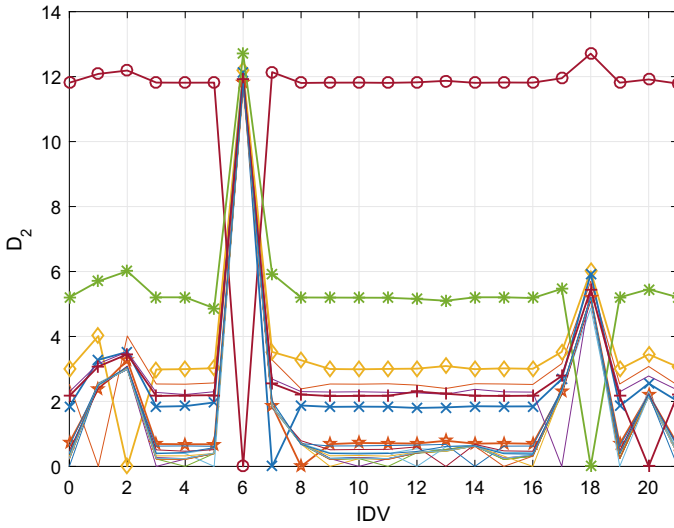


Fig. 4.6 D_2 index for different faults

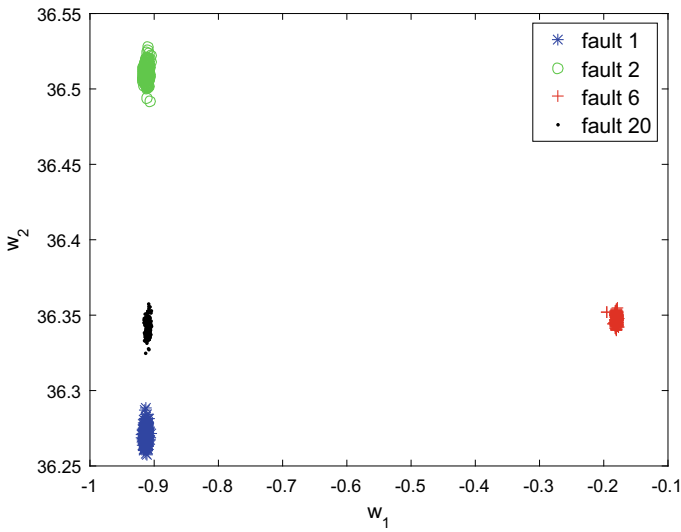


Fig. 4.7 FDA identification result for the fault 1, 2, 6, and 20

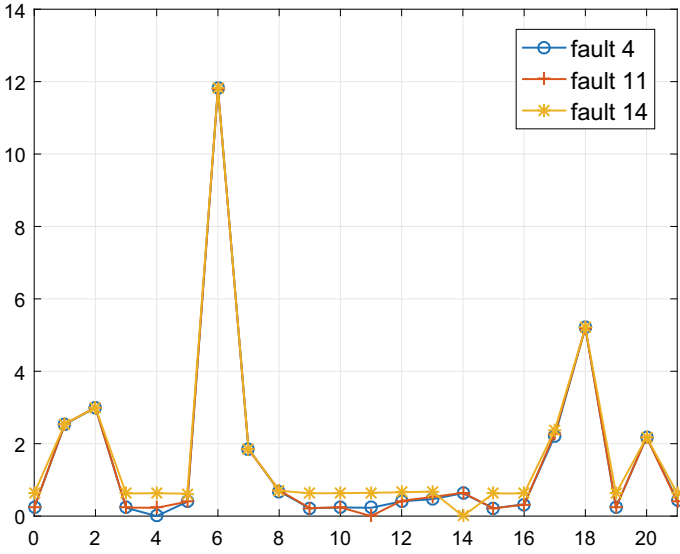


Fig. 4.8 D₂ indices for fault 4, 11, and 14

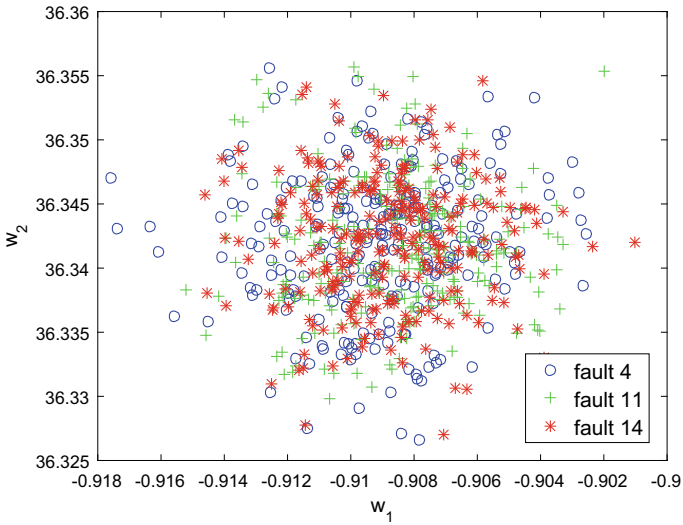


Fig. 4.9 FDA identification result for the fault 4, 11, and 14

4.5 Conclusions

Two kinds of simulation platforms are introduced for verifying the statistical monitoring methods and several experiments based on the traditional methods, PCA, PLS, CCA, and FDA, are finished. These basic experiments illustrate the characteristics of several methods and their fault detection effects. Actually, there are lots of improved methods to overcome the shortcomings and deficiencies of the original multivariate statistical analysis methods. Each method has its own conditions and scope of application. No one method completely outperforms the others in terms of performance. Furthermore, data-based fault detection methods need to be combined with the actual monitoring objects, and existing methods need to be improved according to its knowledge and characteristics. So this book focus on the fault detection (discrimination) strategies for batch processes and strong nonlinear systems.

References

- Birol G, Undey C, Cinar A (2002) A modular simulation package for fed-batch fermentation: penicillin production. *Comput Chem Eng*, pp 1553–1565
- Leoand M, Russell E, Braatz R (2001) *Tennessee eastman process*. Springer, London

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.

