Fault Detection Analysis Before and After Dynamic Model Reforming on the Benchmark Tennessee Eastman Process

Bo Chen, Zhu Wang, Xiong-Lin Luo*

Department of Automation, China University of Petroleum Beijing, 102249, China
luoxl@cup.edu.cn

To make the simulation object closer to the actual chemical process, the author improved the model of Tennessee Eastman (TE) process in their previous work. In this paper, based on above improved model, to further analyse its effect on the fault detection performance, a research on fault data of the improved model and original model is done, based on principle component analysis (PCA), the detection rate of Hotelling's T^2 statistic, Q statistic and support vector machine (SVM) integrated particle swarm optimization (PSO) approach are used to reflect their detection performance on the two models. From the detection rates, the detection performance gets worse when detecting the fault of the improved model. The analysis indicates that when the detection methods are used to detect the faults in actual chemical process, the detection performance will be influenced and may not be as effective as described in literature.

1. Introduction

With the rapid development of modern industrial technology, the structure of various large-scale automation systems is becoming more and more complex; fault detection of systems has always been the focus of academic attention (Amin et al., 2010). In fact, the fault detection methods or control algorithms are always tested on the simulation models. However, there are always a large number of assumptions which are taken as precondition when building the simulation models. Tennessee Eastman (TE) process is proposed according to an actual chemical process of Eastman chemical company in 1993, its simulation model is widely used as a test object, however, modelling the TE process presented a trade-off between rigor and model stiffness due to the vapour phase dynamics during pressure change (Downs and Vogel, 1993), in other words, many detection methods are questionable, because the TE model is also based on certain assumptions and simplifications. Thus, the author (Chen et al., 2018) improved the model to restore the vapour phase dynamics during pressure change, the aim was to make TE model sufficiently complex and realistic to improve the quality of control algorithms and fault detection and diagnosis methods. In fact, people always compare different detection methods on the same one model, there is hardly any literature discussing the performance of the same detection methods on model and actual process.

Thus, in this paper, based on the previous work, to analyse the specific differences of fault detection method on actual process and simplified model, in other words, the aim is to study whether the previous fault detection method has the same performance or not when the model is sufficiently complex and realistic, a research on fault data of the improved and original model is done. First, a control structure that can stable the system is used to produce the data, the data sets of normal condition are used as training sets, and the fault data are used as testing sets. Then, to avoid the dimension disaster and decrease calculation load, principal component analysis (PCA) is used to deal with the data sets, the Hotelling's T^2 statistic and the Q statistic are used to detect the faults, their correct classification rates (CCR) and false alarm rate (FAR) are used to reflect their detection performance on the two models. The detection results indicate that when the model is more complex and realistic, the fault detection methods get worse, to make above results more reliable, the SVM integrated PSO approach in recent literature is used too. Finally, the results show when the detection methods are used to detect the faults in actual chemical process, the detection performance may not be as effective as described in literature.

2. Preliminaries

2.1 TE process

TE process is a plant-wide process control problem, it is widely used as a test problem, and it contains eight components. The process includes five major units: a reactor, a condenser, a separator, a recycle compressor and a product stripper. The process has 53 monitoring variables including 41 measurements and 12 manipulated variables, once a fault is taken on, all variables will be influenced. More details can be found in (Downs and Vogel, 1993). The process flow diagram of TE process is shown in Figure 1. However, the original model simplifies the vapour phase dynamics during pressure change, there is only one temperature variable and the temperature within the whole device is represented by liquid temperature, the synchronization between vapour and liquid dynamic does not accord with practical process, thus, as mentioned in the introduction, the author (Chen et al., 2018) improved the TE model by taking into account the vapour phase energy balance, then the fast changes of pressure and vapour temperature are restored, in this way, the improved model can reflect the dynamic behaviour of vapour and liquid phase better and it also presents the differences in physical properties of vapour and liquid phase which is ignored in the original model, it indicates that the responses of some variables are also different from that of the original model and the improved model is closer to the actual process.
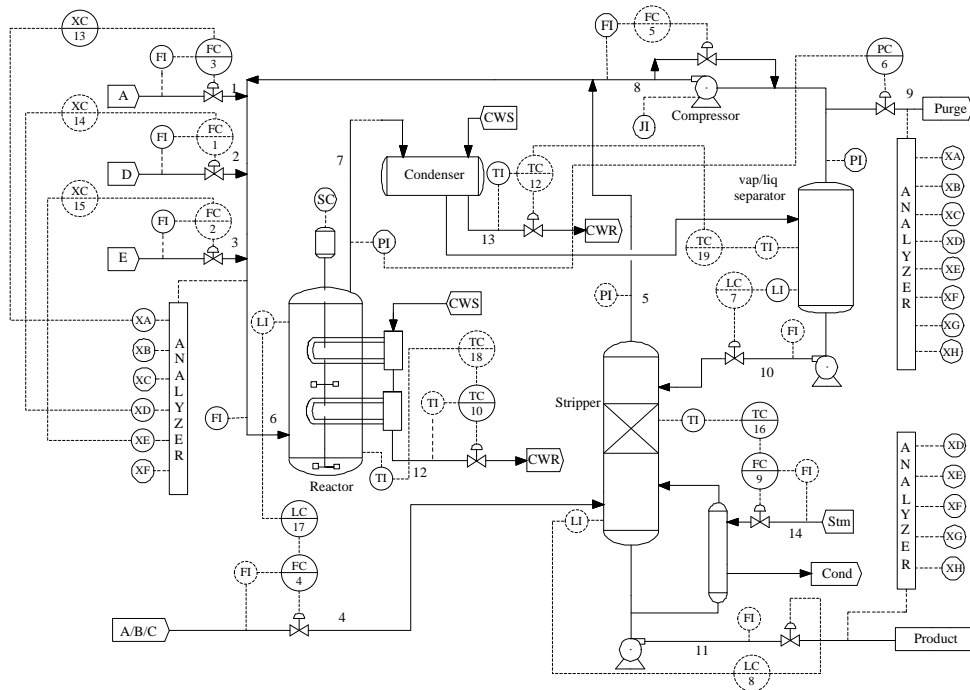

Figure 1: Process flow diagram of TE process including a control structure

2.2 PCA based Hotelling'sT^2 statistic and Q statistic

PCA is a dimension reduction technique, it projects data from a high dimensional data space to a lower dimensional subspace, and transforms a set of correlated variables into a set of uncorrelated variables, it is described briefly below. Set a data set X with n observations and m process variables denoted as a matrix $X^T=[x_1,...,x_n] \in R^{m \times n}$, $x \in R^m$, then the covariance matrix $S=X^TX/(n-1)$ is obtained. Then, singular value decomposition (SVD) is applied to S and the loading matrix $P=[P_{pc},P_{res}] \in R^{m \times m}$ is generated. Finally, the observations in X are projected to the lower-dimensional space $Z_{pc}^T=P_{pc}^TX^T$. The detailed can be found in literature (Wang and Yin, 2015). Hotelling's T^2 statistic and Q statistic are typical methods to monitor process; T^2 measures the magnitude of variations that are inside the PCs, Q statistic measures variability that breaks the normal process correlation, which often indicates an abnormal situation. Based on PCA, the T^2 statistic and Q statistic can be calculated, their confidence limits are calculated with a level of significance, α. More details can be found in (Yin et al., 2012).

2.3 PSO and SVM

SVM is widely used in classification task and a lot of scholars have studied it. Gaussian radial basis function (RBF) kernel SVM is adopted in this paper. The objective function is as following.

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{m}\xi_i \tag{1}$$

$$\text{s.t. } y_i(w^T x_i + b) \geq 1 - \xi_i \tag{2}$$

$$\xi_i \geq 0 \tag{3}$$

where (x_i, y_i) donates the training data, $i=1,...,m$, $x_i \in R^n$ and $y \in \{-1,1\}$, the parameter C is a penalty parameter of error term, it is represented by the parameter ξ_i.

The PSO algorithm was initially inspired by the regularity of bird cluster activities, and then a simplified model based on swarm intelligence was established. Set the $X_i=(x_{i1},x_{i2},...,x_{in})(i=1,..,m)$ donates the location of the ith particle, $P_i=(p_{i1},p_{i2},...,p_{in})$ donates the best position of the ith particle, $P_g=(p_{g1},p_{g2},...,p_{gn})$ donates the best position in all m particles, $V_i=(v_{i1},v_{i2},...,v_{in})$ donates the pace of the ith particle moving to another position. The particles are moved according to the following equations.

$$V_i(k+1)=wV_i(k)+c_1r_1(P_i-X_i(k))+c_2r_2(P_g-X_i(k)) \tag{4}$$

$$X_i(k+1)=X_i(k)+V_i(k+1) \tag{5}$$

where k is the number of iterations, w is inertia weight. If w is chosen appropriately, the number of iterations required can be small, c_1 and c_2 are acceleration constants, r_1 and r_2 are defined randomly between 0 and 1. The PSO is used to obtain the best parameters of SVM. Both Liu (2016) and Li et al. (2016) used the PSO to optimize parameters. Zhang and Guo (2016) used the PSO-SVM for diagnosis and the result indicated that it is an effective method.

3. Results and discussion

In this section, to study the detection performance of the above detection methods on two models, the training data sets and the testing data sets from the original model and the improved model of TE process are collected, respectively. Because the TE process is open-loop unstable, the system must be operated under closed loop, the control structure is shown in Figure 1. The training data sets and the testing data sets have 52 monitoring variables, including 41 measurements and 11 manipulated variables, in which the agitation speed is not included because it is not manipulated; Table 1 shows the final steady state values of 11 manipulated variables, it indicates that their final steady state values are almost exactly the same. The data set of normal condition is used as the training data set, it has 500 observation samples, and the testing data set has 960 samples, for the fault data, only the first 160 samples are in normal status. The FAR and CCR are used to evaluate the detection performance. When one of the test data exceeds the threshold, it is identified as a fault.

Table 1: The final steady state values of 11 manipulated variables in two models

MV	OM (%)	IM (%)	MV	OM (%)	IM (%)	MV	OM (%)	IM (%)	MV	OM (%)	IM (%)
1	63.1	63.1	4	61.2	61.2	7	38.1	38.1	10	41.1	41.1
2	54.0	54.0	5	22.2	22.2	8	46.5	46.5	11	18.1	18.1
3	24.6	24.6	6	40.1	40.1	9	47.4	47.5			

where MV is manipulated variables, OM is the original model, IM is the improved model.

The study includes 4 cases, the training sets and testing sets will come from two models respectively, as shown in Table 2, case 1: the training sets comes from the original model (Training sets 1), the testing sets comes from the original model (Testing sets 1); case 2: the training sets comes from the original model, the testing sets comes from the improved model (Testing sets 2); case 3: the training sets comes from the improved model (Training sets 2), the testing sets comes from the original model; case 4: the training sets comes from the improved model, the testing sets comes from the improved model. Case 1 is used to study the detection performance on the simplified model, case 4 is used to study the detection performance on the complex model,

the other two cases (case 2 and case 3) aim to show the difference of two models. Their FAR or CCR will be compared, they are calculated by Eq(6).

Table 2: Detail information of 4 cases

Case	Training sets 1	Testing sets 1	Training sets 2	Testing sets 2
Case 1	√	√		
Case 2	√			√
Case 3		√	√	
Case 4			√	√

$$FAR = \frac{\text{No. of normal samples identified as fault}}{\text{total No. of normal samples}} \times 100\% \quad CCR = \frac{\text{No. of correctly classified samples}}{\text{total samples}} \times 100\% \quad (6)$$

First, the normal condition is tested to reflect the FAR. As shown in Figure 2, Figure 2a and Figure 2d show that they can detect normally, only a few number of data points exceed the threshold, their FARs are close to 5%, which accords with the level of significance (α=0.05), as shown in Table 3, the FAR between case 1 and case 4 has little difference (the FAR of case 4 is slightly higher), because no faults occurs, all variables fluctuate near their steady-state values. Figure 2b and Figure 2c show that when the training set and the testing set come from two models respectively, the FAR is 100%. In fact, even though they have the same final steady state values, after adding the vapour phase energy balance equation and the exist of white noise, the correlation between variables has been changed and is reflected by the data, in other words, after trained by PCA, their covariance matrices are different too, their models are also different, thus all of the data points exceed the threshold.

Table 3: The confidence limits and false alarm rates of T^2

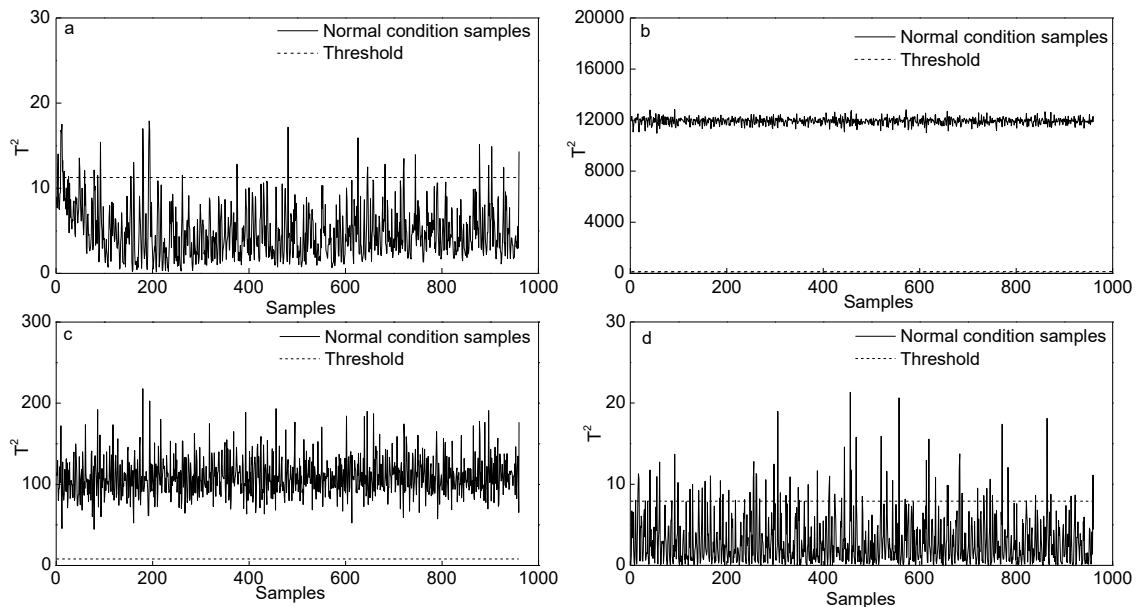| Case	Confidence limit of T^2	FAR of T^2 (%)	Confidence limit of Q	FAR of Q (%)
Case 1	11.3	3.65	7.96	3.65
Case 2	11.3	100	7.96	100
Case 3	7.92	100	13.1	100
Case 4	7.92	7.92	13.1	5.21

Figure 2: PCA based T^2 plots of normal condition in 4 cases, a: case 1, b: case 2, c: case 3, d: case 4

The analysis of FAR indicates that the PCA based T^2 and Q statistic can be used to detect the fault of two model, but case 2 and case 3 cannot be used to compare the detection performance, thus in the next section, only the case 1 and case 4 are used to reflect the CCR.

A fault occurs, such as fault 4 (reactor cooling water inlet temperature). The training data have 500 samples of normal condition; the testing data have 960 samples of fault 4. The T^2 and Q plot are illustrated in Figure 3. As shown in Figure 3, in case 4, T^2 has more data points that are not classified accurately, a lot of samples that should be identified fault samples are identified normal samples. Their CCR are shown in Table 4, their CCR of T^2 are 98.4 % and 60.8 % respectively, the CCR of Q in case 4 is also lower, it is obvious that the detection performance is worse when detecting the fault in case 4.

Table 4: The CCR of T^2 and Q statistic for fault 4

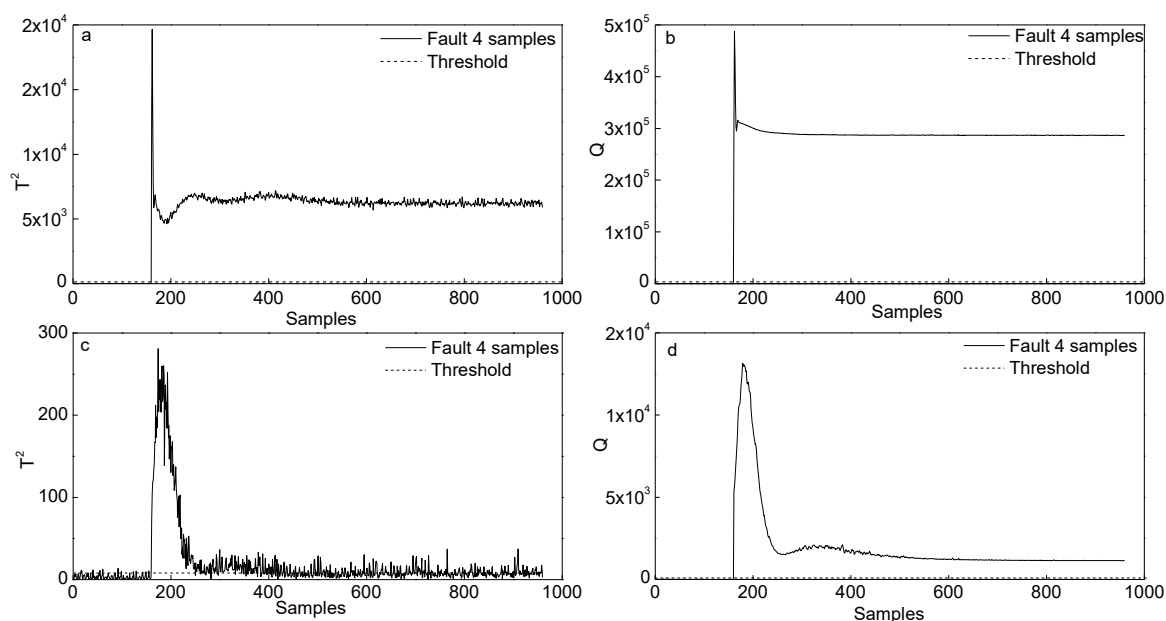| Case	CCR of T^2 (%)	CCR of Q (%)
Case 1	98.4	99.5
Case 4	60.8	98.6

Figure 3: PCA based T^2 and Q plots of fault 4 in case 1 and case 4, a and b: case 1, c and d: case 4

To avoid the chance of above detection method, the PSO-SVM is used to diagnose the fault 4 too. The training data set considered here has 1460 samples including 500 of normal condition, 480 of fault 4 and 480 of fault 5 (condenser cooling water inlet temperature), the fault 4 is the testing data set including 960 samples. Before detecting, the data are processed by PCA. The prediction results are shown in Figure 4. As shown in Figure 4a, the samples which are predicted to be from normal condition are labeled as '0', while the samples predicted from fault 4 are labeled as '1' and the samples predicted from fault 5 are labeled as '2'. When the condition is normal, both of them diagnose accurately and have the same number (160) of label '0', when fault 4 occurs, the case 1 has 742 labels of '1' and 58 labels of '2', but the case 4 has 690 labels of '1' and 110 labels of '2', which means that case 1 has more fault points that are detected accurately. As shown in Figure 4b, the CCR of case 1 and case 4 are 94% and 88.5% respectively, case 4 has lower CCR; the results also show that the performance of the fault detection method will be worse when applied to the improved model.

For some faults, both case 1 and case 4 have high CCR, because these faults are normal faults with significant symptoms and they are easy to be detected compared with incipient faults. The incipient faults are usually difficult to detect, because their magnitudes are extremely small and tend to be buried by either the process trend or the measurement noise (He et al., 2018). To reflect the detection performance of detection method on small faults for two models, the magnitudes of fault 1, 3, 5, 11 are reduced, the CCR of these faults are shown in Table 5.

As shown in Table 5, for case 1, the detection method still has higher CCR, but for case 4, fault 3, 5, 11 cannot be detected effectively when the magnitudes are reduced and nearly half of the data points have not been correctly detected. The results indicate that when the same incipient faults happen, getting good results in the simulation model does not mean that it will get the same results in a more complex actual chemical process.
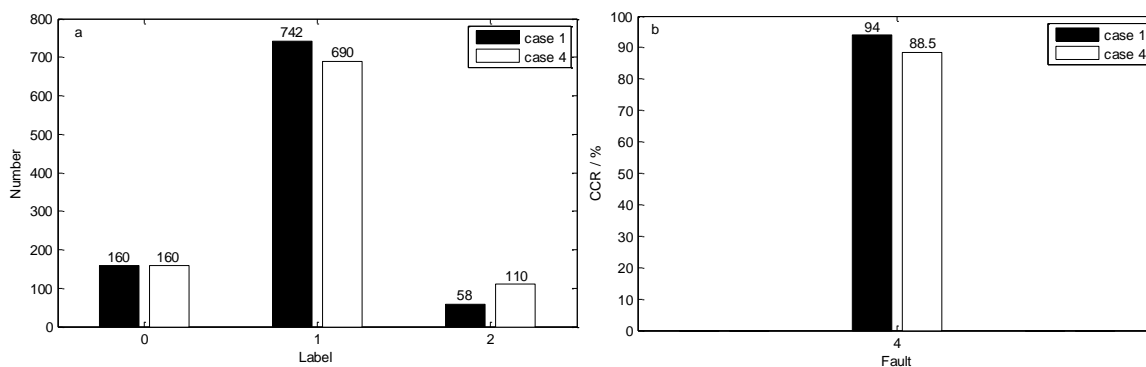
Figure 4: The prediction results of PSO-SVM for fault 4, a: the number of different predicted class labels for case 1 and case 4, b: the CCR of case 1 and case 4 for fault 4

Table 5: The CCR of other faults of case 1 and case 4

Fault Description	Fault type	Case 1 (%)	Case 4 (%)
1: A Feed Ratio, B Composition Constant in stream 4	Step	98.4	98.3
3: D Feed Temperature in stream 2	Step	98.3	57.1
5: Condenser Cooling Water Inlet Temperature	Step	80.0	49.7
11: Cooling Water Inlet Temperature	Random	96.8	46.7

4. Conclusions

In this paper, the aim is to compare the detection performance of the same fault detection method on simplified model and complex model. The performance is always good when detecting the simplified model but the research results indicate that when the model is complex and realistic, the performance of fault detection method gets worse. In other words, when the detection methods are used to detect the faults in the more complex actual chemical process, the detection performance will not be as effective as described in literature. In general, many methods cannot be applied in the actual process directly.

Acknowledgments

References

Amin M.T., Imtiaz S., Khan, F., 2018, Process System Fault Detection and Diagnosis Using a Hybrid Technique, Chemical Engineering Science, 192(2), 191–211.

Chen B., Lan F.W., Wang Z., Luo X. L., 2018, Dual-time scale based extending of the benchmark Tennessee Eastman process[M], Computer Aided Chemical Engineering, Elsevier, 44, 529–534.

Downs J.J., Vogel E.F., 1993, A plant-wide industrial process control problem, Computers and Chemical Engineering, 17(3), 245–255.

He, Z., Shardt, Y. A. W., Wang, D., Hou, B., Zhou, H., Wang, J., 2018, An incipient fault detection approach via detrending and denoising. Control Engineering Practice, 74, 1-12.

Li W., Wang X.C., Wang X.S., Wang H., 2016, Endpoint Prediction of BOF Steelmaking based on BP Neural Network Combined with Improved PSO, Chemical Engineering Transactions, 51, 475–480.

Liu Y.Y., 2016, The Design and Application of Quantum-Behaved Particle Swarm Optimization Based on Levy Flight, Chemical Engineering Transactions, 51, 499–504.

Wang G., Yin S., 2015, Quality-related fault detection approach based on orthogonal signal correction and modified PLS, IEEE Transactions on Industrial Informatics, 11(2), 398–405.

Yin S., Ding S.X., Haghani A., Hao H., Zhang P., 2012, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process, Journal of Process Control, 22(9), 1567–1581.

Zhang Z., Guo H., 2016, Research on Fault Diagnosis of Diesel Engine Based on PSO-SVM, Proceedings of the 6th International Asia Conference on Industrial Engineering and Management Innovation, Atlantis Press, DOI: 10.2991/978-94-6239-145-1_48.