2020 IEEE 9th Data Driven Control and Learning Systems Conference
November 20-22, 2020, Liuzhou, China

Application of ARIMA Model in Fault Diagnosis of TEP

Wei Mu[1], Aihua Zhang[1], Wenxiao Gao[1], Xing Huo[1]

1. College of Engineering, Bohai University, Jinzhou China, 121013
E-mail: muwei_8650@163.com

Abstract: In fault prediction, ARMA is a commonly used and important method to study time series, but it also has some problems. When time series is non-stationary, ARMA prediction effect is not accurate. Therefore, on this basis, the ARIMA model is used to transform the non-stationary time series into stationary time series by the difference method. In view of the three fault states in TEP quality, the ARIMA model is established by using Python, and a TEP quality fault diagnosis system is further established, and the quality variables are predicted. The results show that ARIMA model can predict the quality of TEP in a short time, which has the advantages of simple modeling and accurate prediction. The ARIMA model is reliable for the quality fault diagnosis. Numerical simulations show that the model can accurately describe the change of quality overtime when a fault occurred, and can make judgment and early warning of the fault in a short period of time.

Key Words: ARIMA model; Fault diagnosis; TEP; Time series

1 Introduction

As the reliability and security of the underlying system are increasingly high, it is very important to find the abnormal situation in the system as early as possible. With the maturity of model-based fault diagnosis technology, many industrial application methods have been well developed in the past few years [1-3]. TEP (Tennessee Eastman Process) is a simulation platform developed by Eastman's company, which simulates a real chemical process [4-6]. Downs first introduced the use of TEP as a benchmark for academic research in 1993[7]. Since then, TEP has been widely used in the simulation and verification of various control and process monitoring methods.

In recent years, fault detection methods are based on data-driven technology has become a hot topic [8-11]. But for the fault prediction, ARMA (Auto-Regressive and Moving Average) is a powerful method for capturing trends in time series and predicting their future values. It also allows us to naturally capture the empirical characteristics of time-dependent data. Some authors [12-14] use the ARMA method to predict failure rates. Other authors [15-17] use ARMA directly to predict device fault. If the time series is stationary, the ARMA model will be good for predicting faults. If the time series is non-stationary, there will be no point in using the ARMA model to predict faults. Therefore, on this basis, this paper further proposes the concept of difference, and establishes a new model called ARIMA (Auto-Regressive Integrated Moving Average) model. By using ARIMA model, non-stationary time series can be transformed into stationary time series for fault prediction.

This paper uses Python to establish a time series model-ARIMA model. Time series analysis is one of the most active branches in the field of data processing and analysis. This method is based on mathematical statistics and stochastic theoretical processes [18].The sample data of the time series are arranged in chronological order and the same time interval sequence $\{X_t\}$. Different from the general cross-section data, its biggest characteristic is the consistency and correlation of the data, so we can use a large number of historical data to predict the future sequence values. Through observe and study the quality change in a period of time, we obtain the law of its change and development, so as to improve the quality fit and improve the accuracy of its forecast. It can timely remind the relevant staffs to repair and avoid losses to the maximum extent possible. In summary, we can use historical data to predict future values, which is the advantage of the ARIMA model. Based on Auto-Regressive (AR) model, Moving Average (MA) model and Auto-Regressive Moving Average (ARMA) model, an Auto-Regressive Integrated Moving Average (ARIMA) model is established. The steps are smooth the data and establish the model after identifying the model type, testing the model and applying the model after the test to predict the quality value. Finally, the accuracy of ARIMA model is verified by numerical simulation and the application of TEP.

2 Establishment of ARIMA Prediction Model

The ARMA model was a famous time series prediction method proposed by Box and Jenkins at the beginning of 1970 [19]. It is a model that analyzes and scientifically predicts the sequence values in chronological order. On this basis, we obtain the ARIMA model through difference.

ARMA model can be divided into three forms: AR model, MA model and ARMA model.

A) AR model

The AR model expresses the relationship between a variable X_t in the time series $\{X_t\}$ and the previous p moments, which is called the AR model (p), where the general formula is as follow:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} \tag{1}$$

In the formula, X_t -quality at time t; t-time number; φ_i -Auto-Regressive coefficient; i-the natural number less than or equal to the Auto-Regressive coefficient p, such as, i

*This works is partly supported by the Natural Science Foundation of Liaoning, China under Grant 2019MS008, Education Committee Project of Liaoning, China under Grant LJ2019003.

=1,2,3..., p; p-Auto-Regressive order; X_{t-i} -delay variable; ε_t -white noise at time t in the noise sequence.

The sequence $\{\varepsilon_t\}$ will be a white noise sequence if the mean value and covariance of the independent random variables are 0 at different times. It is random error with a mean value of 0 and covariance of 0. In particular, the residual of the model is equal to the actual value minus the predicted value, and if the residual approximation has the property of white noise, the test will pass during the process of testing the model.

B) MA model

MA can be interpreted as the weighted average sum of multiple white noise sequences of time series, which is called MA model (q). The general formula is as follows:

$$X_t = \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{2}$$

In the formula, θ_j -Moving Average coefficient; j-the natural number less than or equal to the Moving Average coefficient q, such as, j=1,2,3...,q; q-Moving Average order; ε_{t-j} - white noise at time t –j in the noise sequence.

C) ARMA model

ARMA (p, q) is the expression of the combined AR model and the MA model. The general formula is as follows:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \theta_j X_{t-j} \tag{3}$$

It applies to a smooth time series ARMA model, however, in practical problems, the quality of time series is not smooth. Therefore, it cannot use ARMA model to describe directly. So you need use the ARMA model after d order difference treatment model, which is the ARIMA (p, d, q) model where d order number for the difference.

3 Specific Steps to Build the Model

3.1 Station Treatment

In practical engineering, time series are not stationary, so the time series should be stabilized by differential transformation. Generally speaking, the first-order difference can achieve a stable trend, which is defined as follows:

$$\nabla X_t = X_t - X_{t-1} \tag{4}$$

$$\nabla^d = (1-B)^d \tag{5}$$

In the formula, ∇ -difference operator; B-delay operator; d-the order of the difference.

Python is used to input sample data, and the analyzed data is time defined. Then the non-stationary time series is transformed into stationary time series by differential transformation. Through this stage, the value of parameter d can be determined: if a stationary sequence is obtained after the d-order difference processing of the time series, then ARIMA (p, d, q) model will be selected to make the order of the time series after stabilization as d in the model.

3.2 Build a Model

After the differential stabilization is completed, a stationary time series is input to obtain the graphs of the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF). First, identify the type of model:

A)If the autocorrelation coefficient is trailing and the partial autocorrelation coefficient is truncated, the sequence is suitable for the AR (p) model.

B)If the autocorrelation coefficient is truncated and the partial autocorrelation coefficient is trailing, the sequence is suitable for the MA (q) model.

C)If the autocorrelation coefficient and partial autocorrelation coefficient are trailing, the sequence is suitable for ARMA (p, q) model.

Here, the trailing property refers to the gradual decay of ACF and PACF to 0 with the increase of time delay, and the truncated property refers to the sudden decay to 0 after certain order. If the time series needs to be stabilized, the autocorrelation coefficient and partial autocorrelation coefficient after differential stabilization will are both trailing. ARIMA (p, d, q) model is established according to the ARMA (p, q) model selected above, and the difference order d has been determined through the steps in section 3.1.Then p and q in the model are determined by the delay order in the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF) graphs. The delay order refers to that after certain order; the sequence value no longer has any correlation, so this order is called the delay order. If the autocorrelation coefficient is delayed from order m, that is, the autocorrelation coefficient after order m is all within the confidence interval, the value of q is going to be between 1 and m. If the partial autocorrelation coefficient is within the confidence interval after order n, then the value of p is going to be between 1 and n. Since the method of judging p and q is subjective, multiple tests combined with the Bayes information criterion can yield results. The smaller BIC value is, the better fitting degree is, so we try to select the optimal combination of p and q.

3.3 Model Test

After the establishment of the corresponding model, it will be verified whether the residual sequence is white noise sequence. When the residual of the model is white noise sequence, it indicates that the information of the time series has been extracted and the model fits well. Otherwise, we select the appropriate p and q through the steps in section 3.2. Figuring out the autocorrelation coefficient and partial autocorrelation coefficient of the residual sequence and then observing the autocorrelation coefficient and partial autocorrelation coefficient of each delay order, if they are all within the confidence interval, it will means that the test has passed, and the residual sequence is a white noise sequence.

3.4 Predict Quality Value

According to the model after passing the test, the predicted graph of time series and the predicted value of quality are made. The relative error caused by using ARIMA model to predict quality is defined as the predicted value of quality minus the actual value and then divided by the actual value. The relative error is used to evaluate the predicted results. If the fitting value in the prediction result graph almost coincides with the observed value, it indicates that the model achieves good prediction effect and can predict the quality in the future period.

4 TEP Process and Simulation

TEP process variables can be divided into two parts, a part has 41 measurements, XMEAS module (1-41), which includes 19 sampling process measurements and 22 continuous process measured values. The other part includes 12 manipulated variables, XMV module (1-12).This experiment also is the same as some of the author [20], with XMEAS (35) as a variable quality.

In the test of fault detection, the regression model is first established by using 500 samples under normal working conditions, and then 500 samples are set up for detection. Fault is not set for 200 samples, while fault is set for the remaining 300 samples. In the simulation experiment, a total of 21 faults are designed. This time, quality-related faults IDV (10) and quality-unrelated faults IDV (14) are taken as examples. On the basis of PCR, SVD is further used for decomposition, and finally the fault is verified according to the fault logic judgment. The test results are shown in Figure 1 and Figure 2. On this basis, the ARIMA model is further used to verify the accuracy of the results. We can see relative error is within ±5% in Table 1 and Table 2.

The experimental results of these two experiments show that the relative error of the predicted results is less than 5% from Table 1 and Table 2, indicating that the predicted results of ARIMA model are relatively accurate.

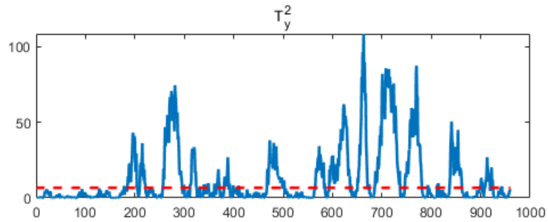

Figure 1.Testing report under IDV (10)

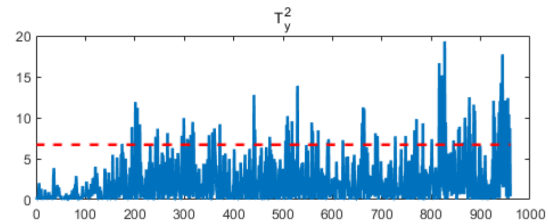

Figure 2.Testing report under IDV (14)

Table 1: Comparison between -actual value and predicted value

Time	Actual value	Predictive value	D-value	Relative error
0.00	4.8753	4.8131	-0.0619	-1.27
0.06	4.8753	4.8137	-0.0616	-1.26
0.12	4.7807	4.8089	0.0262	0.55
0.18	4.7807	4.8098	0.0291	0.61
0.24	4.7990	4.8065	0.0075	0.16
0.30	4.7990	4.8078	0.0088	0.18
0.36	4.8855	4.8057	-0.0798	1.63
0.42	4.8855	4.8072	-0.0783	1.60
0.48	4.9480	4.8060	-0.1420	-2.87
0.54	4.9480	4.8077	-0.1403	2.84

Table 2: Comparison between -actual value and predicted value

Time	Actual value	Predictive value	D-value	Relative error
0.00	4.7996	4.9031	0.1035	2.16
0.06	4.7996	4.9036	0.1040	2.17
0.12	4.9843	4.9037	-0.0806	-1.62
0.18	4.9843	4.9038	-0.0805	-1.62
0.24	4.9715	4.9029	-0.0686	-1.38
0.30	4.9715	4.9040	-0.0675	-1.36
0.36	4.9024	4.9040	0.0016	0.03
0.42	4.9024	4.9041	0.0017	0.03
0.48	5.0119	4.9025	-0.1094	-2.18
0.54	5.0119	4.9043	-0.1085	-2.16

5 TEP Engineering Applications

It can be seen from the above that ARIMA model has a good prediction effect, so we use ARIMA model to further predict TEP fault. There are three main types of TEP quality fault: step fault, random fault, and valve fault. The changes in the quality of the three faults are studied, and the data are recorded every 6 minutes for a total of 800 pieces.

The ARIMA model can not only identify fault types, but also predict quality in a short period of time based on historical data trends. This model is applied to the fault diagnosis system and an intelligent quality fault diagnosis system is established.

The process of quality fault diagnosis system is shown in Figure 3.

We have made three fault models：step fault, random fault, and valve fault. Their time series diagram are shown in Figure4, Figure5 and Figure6.Taking valve fault as an example, it can be seen from Figure 5 that the quality fluctuation is abnormal and has obvious tendency. Therefore, the stationary processing is the premise of modeling. Therefore, inputting the fault quality, solving the first-order difference according to equations (4) and (5), and getting the stationary time series, that is the differential sequence of the fault as shown in Figure 7.After differential treatment, the pressure oscillates up and down the 0 scale, and the time series is stable. Thus, the order of the difference is 1, that is d = 1 in ARIMA model.

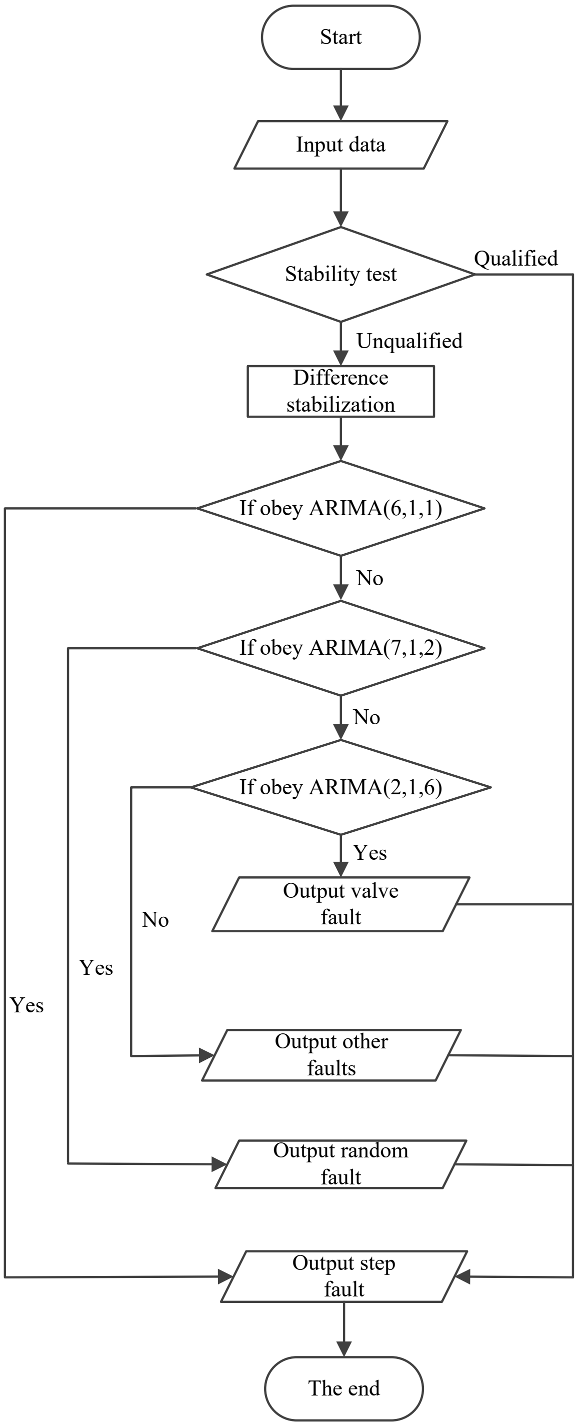

Figure 3. The process of quality fault diagnosis system

After the first difference, the stationary quality is input, and the quality autocorrelation coefficient graph and partial autocorrelation graph of the three faults are made. For an example of valve fault, seeing Figure 8 and Figure 9. Autocorrelation coefficient and partial autocorrelation coefficient figure are oscillation damping, showing the tail. And then, combining the Bayesian information criterion (BIC) to accurately choose ARIMA model of the optimal combination of the p and q. BIC value is smaller, that the higher the goodness of fit. It is available in the valve failure ARIMA model (2, 1, 6). Similarly, step fault of ARIMA model for (6, 1, 1), random fault of ARIMA model for (7, 1, 2).

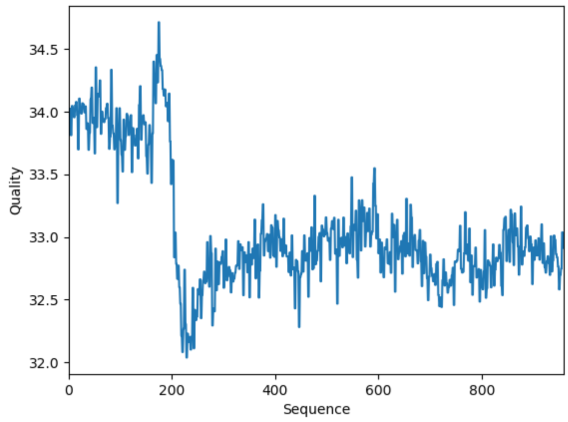

Figure 4. Step fault time series

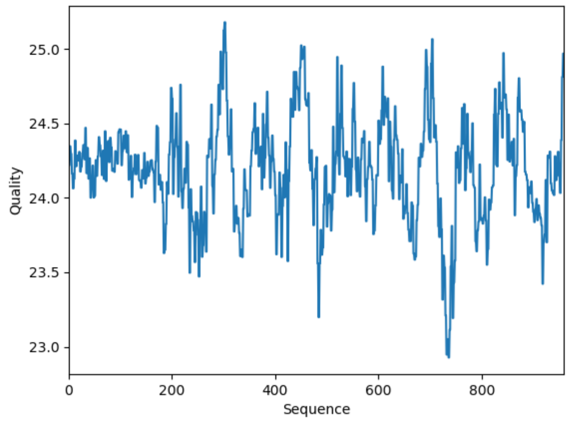

Figure 5. Random fault time series

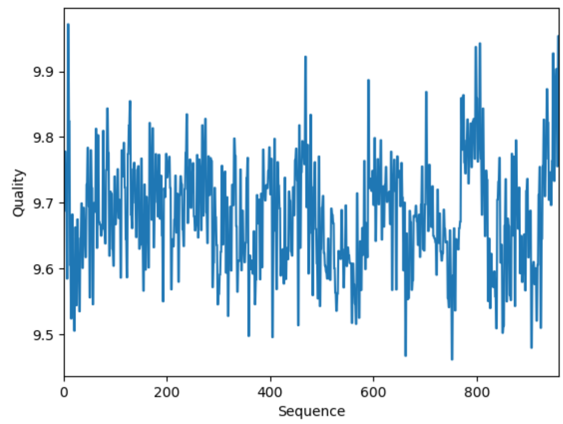

Figure 6. Valve fault time series

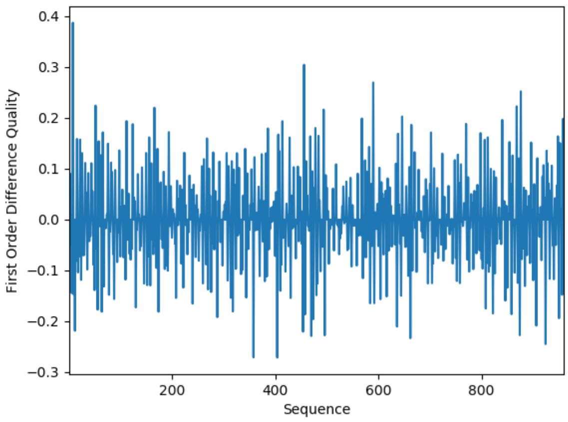

Figure 7. The differential sequence of the value fault

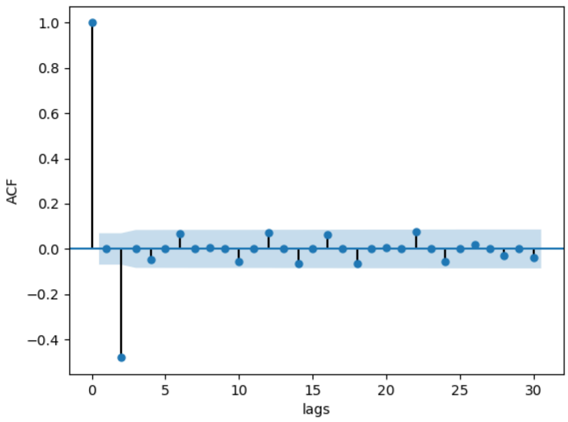

Figure 8. Autocorrelation coefficient

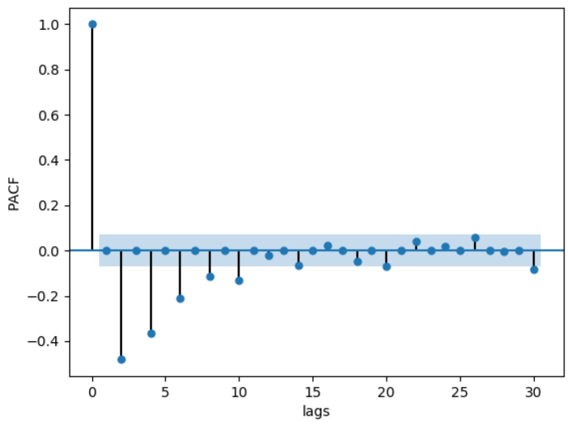

Figure 9. Partial autocorrelation coefficient

Checking the autocorrelation function graph and partial autocorrelation coefficient graph of the residual sequence, and seeing that the autocorrelation coefficient and partial autocorrelation coefficient of each delay order are within the confidence interval, indicating that relevant information has been extracted, the residual difference sequence has been listed as white noise sequence, and the model diagnosis has passed.

According to the tested models, ARIMA (6, 1, 1), (7, 1, 2) and (2, 1, 6) models are used to predict the time series of step fault, random fault and valve fault respectively. The actual pressure curve and the predicted pressure curve basically coincide (Figure10), indicating that the goodness of fit is good. Therefore, this model can approximately describe the trend of three typical faults over a period of time.

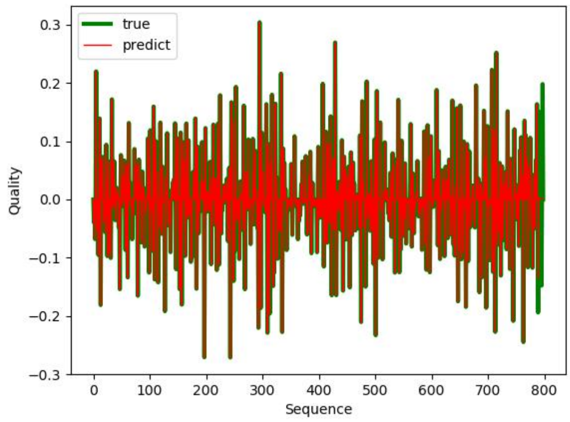

Figure10. Prediction of valve failure

Firstly, a series of values of the day's quality are obtained as the input quantity. If the quality fluctuation is basically stable, it is normal; then the sequence is non-stationary time series, and the first-order differential stabilization is carried out. Then, the model is tested to determine whether it conforms to the ARIMA (6, 1, 1) model, and to judge according to the principles mentioned above. If it conforms to the requirements, it will be a step fault. Otherwise, the next step will be verified whether it conforms to the ARIMA (7, 1, 2) model. If so, it will be a random fault. If not, then the ARIMA (2, 1, 6) model is determined. If so, it will be a valve fault. If not, other faults will be output to further determine the sources and problems of other faults. Other faults will not be discussed in this paper.

6 Conclusions

The ARIMA model is reliable for fault diagnosis of TEP quality. The model will accurately describe the change of quality value with time when quality fault occurs. ARIMA model can predict quality in a short period of time, which has the advantages of simple modeling and accurate prediction. An intelligent quality fault diagnosis system is based on ARIMA model which is established, which can make fault judgment and early warning in a short period of time.

References

[1] A. Ragab, M. El-Koujok, S. Yacout, et al. Fault Detection and Diagnosis in the Tennessee Eastman Process Using Interpretable Knowledge Discovery// *Reliability and Maintainability Symposium RAMS2017. IEEE, 2017.*

[2] Y. LI, Xiao-Chu T. Improved Performance of Fault Detection Based on, Selection of the Optimal Number of, Principal Components. *Journal of automation*, 2009(12):1550-1557.
[3] Mohd. Yunu M Y, J. Zhang. Utilizing of Dissimilarity Scale-based PCA in Multivariate Statistical Process Monitoring Application. *Journal of Applied Sciences*, 2014, 14(12):1244-1251.
[4] S. Yin, G. Wang, H. Gao. Data-Driven Process Monitoring Based on Modified Orthogonal Projections to Latent Structures. *IEEE Transactions on Control Systems Technology*, 2015:1-8.
[5] S. Yin, Steven X. Ding, Adel. Haghani, H. Hao, P. Zhang. A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process. *Journal of Process Control,* 2012, 22(9).
[6] X. Gao, J. Hou. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing*, 2016, 174.
[7] Downs JJ, Vogel EF. A plant-wide industrial process control problem. 1993, 17(3):245-255.
[8] Y. Wang. Data-driven Design of a Fault Detection and Isolation Scheme for Ground Vehicles with Dual-Clutch Transmission.*Process control committee of Chinese society of automation. Summary of the 30th China Process Control Conference (CPCC 2019).* 2019:205.
[9] W. Liu, P. Zhi, et al. A Review of Industrial Fault Diagnosis Based on Data-driven Methods// 2018.
[10] X. Wei, L. Jia, H. Liu. Data-driven fault detection of vertical rail vehicle suspension systems// *Control (CONTROL), 2012 UKACC International Conference on. IEEE*, 2012.
[11] Fravolini M L, Core G D, Papa U, et al. Data-Driven Schemes for Robust Fault Detection of Air Data System Sensors. *IEEE Transactions on Control Systems Technology*, 2017, PP (99):1-15.
[12] Singh N. Forecasting time-dependent failure rates of systems operating in series and in parallel. *Microelectronics Reliability,* 1994, 34(3):391-403.
[13] R Li, R. Kang. Research on failure rate forecasting method based on ARMA model. *Systems engineering and electronic technology*, 2008, 30(8):1588-1591.
[14] B. Li, J. Zhao, J. Guo. New indicators of equipment failure evaluation and prediction system based on ARMA. *Systems engineering and electronic technology*, 2011, 33(1):98-101.
[15] J. Zhao, L. Xu, L. Liu. Equipment Fault Forecasting Based on ARMA Model// *Mechatronics and Automation, 2007. ICMA 2007.International Conference on. IEEE*, 2007.
[16] H. Guo, H. Luo, B. Long, et al. Prediction Research About Small Sample Failure Data Based on ARMA Model// *IEEE Circuits & Systems International Conference on Testing & Diagnosis. IEEE Explore,* 2009.
[17] M. Baptista, S. Sankararaman, Medeiros IPD, et al. Forecastng Fault Events for Predictive Maintenance using Data-driven Techniques and ARMA Modeling. *Computers & Industrial Engineering*, 2017, 115:41-53.
[18] H. Li, H. Liu. The applications of mathematical statistics and stochastic process in scientific research. *Machine Design & Manufacturing Engineering,* 2017, 46(09):116-118.
[19] G. Janacek. Time series analysis forecasting and control. *Journal of Time*, 2010, 31(4):303-304.
[20] C. Sun, J. Hou. An Improved Principal Component Regression for Quality-Related Process Monitoring of Industrial Control Systems. *IEEE Access,* 2017, 5:21723-21730.