

## The Application of Data-driven Modelling and State Monitoring in Steam Pipe Networks in Iron& Steel Enterprises

Xianxi Luo<sup>a</sup>, Wei Wang<sup>b</sup>, Hongwei Deng<sup>c</sup>, Zehao Wu<sup>d</sup>

Jiangxi Province Engineering Research Center of New Energy Technology and Equipment (East China University of Technology), Nanchang, Jiangxi, China

<sup>a</sup>Xianxi\_luo@yahoo.com, <sup>b</sup>543881900@qq.com, <sup>c</sup>1650427313@qq.com, <sup>d</sup>654852846@qq.com

**Keywords:** Data Driven Modelling, Monitoring, PCA, Steam Pipe Networks.

**Abstract.** The steam pipe network is a complex object to be modelled and monitored with traditional mathematical model. In this paper, the mathematical model of single steam pipe is analyzed and an innovated data-drive modelling method with piecewise performed principal analysis (PCA) is proposed. In this method, the variable space is separated into smaller sections and perform PCA respectively to generate the corresponding approximate linear PCA model. By monitoring the statistical variables  $T^2$  and SPE, the abnormality and fault data can be easily discovered. The simulation result shows the effectiveness of the proposed method.

### Introduction

Steam networks in the iron& steel enterprises mainly have the function to recycle the waste heat energy and reduce the comprehensive energy consumption[1]. It's vital to overall monitor the performance and secure the demand and the production of steam balance all the time, or the main production line will be seriously influenced or even halted. The management and operation are usually based on the data come from the instruments in the field. However, the data maybe contaminated by the disturbances of the channel problems. Modelling the steam pipe networks is helpful to setup the relationship of the variables and improve the data quality[2]. Thus, the system can be effectively monitored.

However, it's difficult to setup an accurate model for the steam pipe networks. Even we have the mathematical model, the complex properties of the steam ambient disturbance induce the deviation from the models.

With the large amount of data collected by energy management system (EMS), the data-driven model of the steam can be feasible to setup. In recent years, the process of monitoring including fault detection based on data, or multivariate statistical process control has been rapidly made[3]. For the high-dimensional and redundant properties, Principal component analysis (PCA) is applied to deal with linear correlated, multi-dimensional and Gaussian process data by mapping the data set onto a lower dimensional subspace [4].

The PCA-based process monitoring scheme assumes that the process behaves linearly. For the cases of the nonlinear system as steam networks, the monitoring of a process with a linear PCA model could not perform normally [5]. To solve the problem, several nonlinear PCA methods have been proposed. Kernel principal component analysis (KPCA) [6] has been targeting the problems and the system dynamics have been discussed in [7]. Although the method can be effective, it cannot easily be applied in practice because of the complex procedure. We propose a piecewise performed PCA, an innovated method based upon PCA, to setup the model of steam pipe networks and monitor the process.

The followed sections are organized as follows. Firstly, the mathematical model and the characteristics are investigated. After that, the procedure of the proposed piecewise performed PCA and the fundamentals monitoring with PCA model are presented. The simulation is demonstrated to show the effectiveness through the case study for a single steam pipe. Finally, the conclusion is presented.

## Investigation of the steam pipe mathematical model

In this investigation, a single pipe carrying steam is studied (as shown in figure 1). According to the previous research, the physical model of the steam pipe consists of hydraulic equation and thermal equation. The models are coupled and should be combined to solve the variables and eliminate confliction[8].

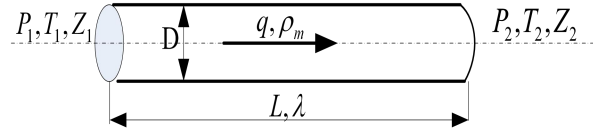


Fig. 1 The representation of single steam pipe

### Hydraulic model

The variables and parameters are shown in Fig. 1 and ignore the minor influencing factors, the hydraulic model can be simplified as [9]:

$$P_1^2 - P_2^2 = 1.25 \times 10^8 \frac{\lambda q^2 P_1 (1 + \eta) L}{D^5 \rho_m} \quad (1)$$

In the equation,  $P_1, P_2, T_1, T_2$  represent the inlet and outlet steam pressure and temperature,  $q$  is the mass flow rate,  $\lambda$  is the coefficient of frictional resistance,  $L$  and  $D$  are the length and the inner diameter of the pipe, and  $\rho_m$  is the weighted mean density of the carried steam. According to Альтшуль equation, the coefficient of frictional resistance  $\lambda$  is:

$$\lambda = 0.11 \times \left( \frac{\Delta}{D} + \frac{68}{\text{Re}} \right)^{0.25} \quad (2)$$

Here,  $\Delta$  is the equivalent absolute roughness;  $\text{Re}$  is the Reynolds number of the steam in the pipe. The Reynolds number can be calculated with the following equation:

$$\text{Re} = \frac{D u \rho_m}{\mu} = 354 \frac{q}{D \mu} \quad (3)$$

### Thermal model

By the law of energy conversation, the thermal model is written as equation (4).

$$T_1 - T_2 = \frac{(1 + \beta) q_l L}{1000 c_p q \times \frac{1000}{3600}} = \frac{(1 + \beta) q_l L}{278 c_p q} \quad (4)$$

The definitions and units of  $T_1, T_2, L, q$  are the same as those used in the hydraulic model.  $\beta$  is the appending heat loss coefficient, which is induced by the appendix of pipes, valves and supports. According to the ways of pipes being laid,  $\beta$  value can be selected as 0.15-0.25.  $c_p$  represents the heat capacity at a specific pressure. IF-97 [10] shows the way to calculate it.  $q_l$  represents the amount of heat loss per unit pipe length and can be calculated with

$$q_l = \frac{T_0 - T_a}{\frac{1}{2\pi\epsilon} \ln \frac{D_o}{D} + \frac{1}{\pi D_0 a_w}} \quad (5)$$

where  $\epsilon$  denotes the coefficient of thermal conductivity of the heat insulated material;  $D_0, D$  denote the outer and inner diameters of the steam pipe;  $T_0, T_a$  are the pipe's out surface temperature and the environment temperature;  $a_w$  is determined by the velocity of environmental air.

The characteristics of the model

To show the characteristics of the model more clearly, a study case was designed as follows.

A section of a steam pipe having 300mm inner diameter ( $D=300\text{mm}$ ) and 600m in length ( $L=600\text{m}$ ) with standard coating was selected to produce the training and testing data. The variation ranges of the variables were set by referring to an actual steam pipe line in an iron& steel plant in China. About 2% random noise was added to the flow rate magnitude at each point to simulate the actual condition.

Figure 2 shows the influence of differential pressure ( $\Delta P = P_1 - P_2$ ) upon the flow rate in case of  $P_1 = 0.85\text{Mpa}$ ,  $T_1 = 290^\circ\text{C}$ ,  $T_2 = 260^\circ\text{C}$ . As the calculated data are not in a line, the segmented line is for the approximation. Differential pressure significantly affects the flow rate of the steam.

Figure 3 shows the influence of the outlet temperature upon the flow rate when  $P_1 = 0.85\text{Mpa}$ ,  $T_1 = 290^\circ\text{C}$ ,  $P_2 = 0.7\text{Mpa}$ ,  $T_2 = 260\sim 280^\circ\text{C}$ . Compared to the influence of differential pressure, the temperature affects much less.

Figure 4 shows the influence of the input pressure upon the flow rate. As we investigated  $\Delta P = P_1 - P_2 = 0.15\sim 0.30\text{Mpa}$  When the condition of case (1)  $P_1 = 0.85\text{Mpa}$ ,  $T_1 = 290^\circ\text{C}$ ,  $T_2 = 260\sim 280^\circ\text{C}$ , and case (2)  $P_1 = 0.75\text{Mpa}$ ,  $T_1 = 283^\circ\text{C}$ ,  $T_2 = 260\sim 280^\circ\text{C}$ . These Two groups of scatters are shown in Fig.4, the input pressure and the differential pressure significantly affect the flow rate. Further study shows the approximation cannot be made with a single line.

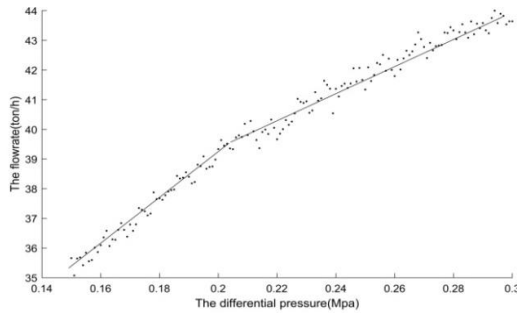


Figure 2. The influence of differential pressure

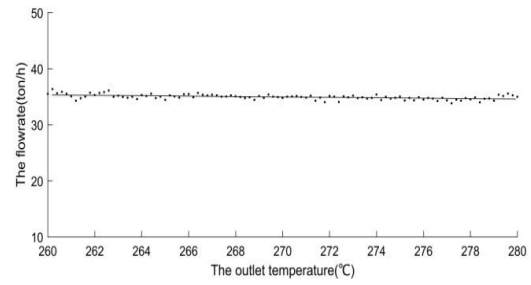


Fig 3. The influence of outlet temperature upon flow rate

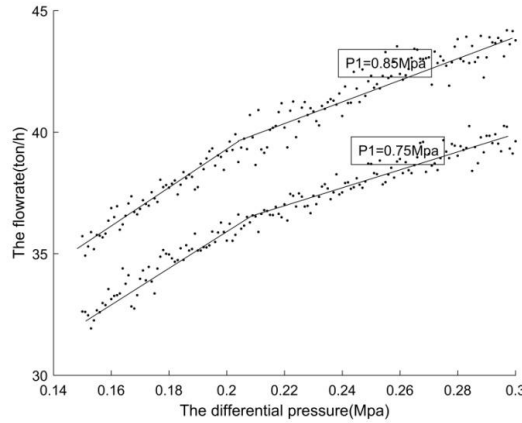


Fig. 4 The influence of inlet pressure upon flow rate

As the case study shown, the relationships between the variables of  $P_1, T_1, P_2, T_2, q$  are nonlinear. Besides the complexity and strong coupling properties of the model, several parameters are influenced by the ambient conditions and the current state of the steam.

In this paper, the model is used to generate the simulation data analogous to the state of actual state. After that, the data will be adopted to test the effectiveness of the proposed algorithm, data driven modelling and state monitoring.

## PCA Fundamentals for PCA Modelling and Process Monitoring

Principal Component Analysis, PCA is one of the technology of data-driven modelling and fault diagnosis. With PCA, the linear correlated multi-dimensional and Gaussian process data are mapped onto a lower dimensional space with little information loss. If given  $X = [x_1, x_2, \dots, x_m] \in R^{n \times m}$  to represent process data of  $n$  samples and  $m$  variables,  $X$  can be decomposed as [11],

$$X = \hat{T}\hat{P}_k^T + E \quad (6)$$

The pre-process is to standardize each row of  $X$  into the samples with zero mean and unit variance.  $\hat{P}_k \in R^{m \times k}$  represents the selected loading matrix, the columns of the matrix are eigenvectors

corresponding to the eigenvalues  $\lambda_i$  of the covariance matrix and mutually orthogonal.  $\hat{T} \in R^{n \times k}$  denotes the score matrix representing the projection of samples upon the column vector of loading matrix.  $E \in R^{n \times k}$  is the residual matrix denoting the influence of ambient noise or disturbance.  $k$  is the number of the remained principal components and is determined by the inequality of the cumulative percent variance (CPV):

$$\sum_{i=1}^k \lambda_i / \sum_{i=1}^m \lambda_i \geq 85\% \quad (7)$$

For one of the observation  $x$  (row vector of  $X$ ), the score vector  $t$  can be expressed as:

$$t = x^T \hat{P}_k \quad (8)$$

The estimation of the observation can be derived as  $\hat{x} = t \hat{P}_k^T$ . In order to monitor a multivariate process, two statistic indices are defined.

$$T^2 = t \Lambda_k^{-1} t^T \quad (9)$$

$$SPE = ee^T \quad (10)$$

in the equations,  $\Lambda_k = \text{delt}(\lambda_1, \lambda_2, \dots, \lambda_k)$ ,  $e = x - \hat{x}$ . The recommended control limits are given as,

$$T^2 \leq \frac{k(n-1)}{n-k} F_{k, n-k, \alpha} \quad (11)$$

$$SPE \leq \theta_1 \left( \frac{c_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0}} \quad (12)$$

In equations (11) and (12),  $F_{k, n-k, \alpha}$  represents an F-distribution with degree of freedom  $k$  and  $n-k$  under the confidence level  $\alpha$ .  $c_\alpha$  represents the threshold value of the standard normal distribution with

confidence level  $\alpha$ . In addition,  $\theta_i = \sum_{j=A+1}^m \lambda_j^i (i=1,2,3)$ ,  $h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2}$  are defined.

To monitor the performance status, the control charts of SPE (referred as  $Q$  in some literatures) and  $T^2$  are produced. When the value of SPE or  $T^2$  is beyond the control limit, some abnormality may have happened and an alarm signal should be shown up.

Assume that the system is linear, and the variables follow Gaussian distribution, classical PCA can be applied to setup the linear relationship of the variables. The approximate data-driven model to represent the relations of the discussed variables can be derived from equations (6) and (8),

$$E = X - \hat{T} \hat{P}_k^T = X(I - \hat{P}_k \hat{P}_k^T) \approx 0 \quad (13)$$

As the variables are standardized by centering the data to zero, theoretically, the model is linear for the variables in full ranges. This model can perform excellently if the studied system is linear and the variables are normally distributed. However, if the system is nonlinear, the farther the variables deviate from the original point, the larger the errors will be produced in the data-driven model.

## Piecewise Performed PCA

In the current application, the target is to monitor five variables,  $P_1, T_1, P_2, T_2, q$ . As the investigation of the mathematical model (mechanism model), the nonlinearity is obvious, and could not apply the traditional PCA to setup a data-driven linear model or monitor the state of the movement.

To overcome the shortage of classical PCA for a such an application, the authors propose the method of piecewise PCA, which means by separating the variable space into several smaller zones. If the small zones are properly divided, the characteristic of the system in the zones could be approximately linear. Different PCA linear models can be setup for these zones. As the new data come, the normality of the data could be determined by the corresponding PCA model and control limits respectively.

Theoretically, for a  $n$ -dimensional system, if the interval of each variable is divided into  $m$  sections in the variable range, there will be  $m^n$  zones in total. The problem will be got very complicated. In practice, we can choose critical variables and the number of sections to divide according to the actual requirements of the system.

Thus the procedure of the piecewise performed PCA can be described as the following steps[12],

1) Collect the data from a real system or simulation system as training data. Try to pick out the normal data.

- 2) Analyze the distribution, the changing range of the training data for each variable.
- 3) Figure out the highest concerned variable (HCV) for the process.
- 4) Determine the most critical influencing variables (CIV) which has the significant influence upon the highest concerned variable.
- 5) Preset the number of the zones to divide and the section boundaries of each section according to the statistical distribution of HCV and CIVs.
- 6) Group the collected data into corresponding zones. Standardize the grouped data into zero means and unit deviations. Save the actual means and deviations of each group.
- 7) Perform PCA analysis for each group data according to the aforementioned fundamentals. note down the important parameters of the model,  $\lambda_i \hat{P}_k$  and the control limits for  $T^2$  and SPE.
- 8) Test the model with the standardized training data by substituting the data into the models and monitor the SPE and  $T^2$  control chart. Normally, there should be no alarm. Otherwise, return to step 3).
- 9) Test the manually set data with the magnitude of certain percentage variation for one variable of one sample. Make sure whether the performance is acceptable. Otherwise, return to step 1) from the very beginning by adding the manually set data because of the low completeness of the training data.
- 10) Test the newly produced data.

## Case Study

To assess the effectiveness of the piecewise performed PCA method applied in the steam networks, the authors considered the problem of monitoring the flow rates in a single steam pipe. By monitoring the variables of this section of the steam network was to discover abnormality of the instruments or the steam leakage along the pipe line.

From the background of the application and the analysis result of the collected data, we chose  $q$  as the highest concerned variable, and  $P_1, \Delta P$  as the critical influencing variables. According to the samples generated by the mathematical model, the variation range of these critical influencing variables can be determined. The space was separated into four zones as shown in figure 5. The definition intervals were written as, Zone 1:  $P_1 \leq 0.79, \Delta P \leq 0.18$ ; Zone 2:  $P_1 \leq 0.79, \Delta P > 0.18$ ; Zone 3:  $P_1 > 0.79, \Delta P \leq 0.18$ ; Zone 4:  $P_1 > 0.79, \Delta P > 0.18$ .

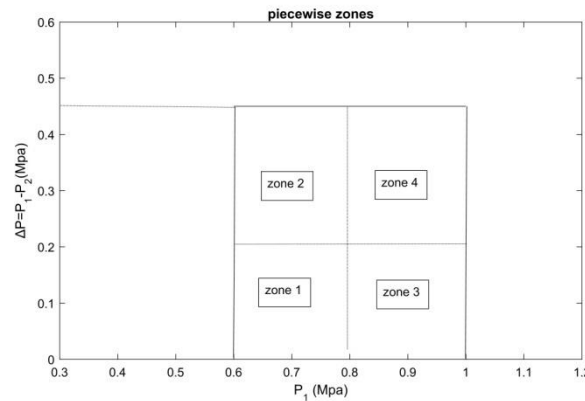


Fig 5. The piecewise divided zones for PCA

The variables were vectored as,

$$X = (x_1, x_2, x_3, x_4, x_5) = (P_1, T_1, \Delta P, T_2, q) \quad (14)$$

The samples were grouped according to the definition domains of the four zones, four groups of samples were derived. 100 samples were selected from each group to perform PCA and to draw SPE and  $T^2$  control charts as shown in figure 6-9. In these figures, the line above the curve in each chart is the control limit for the statistic variable investigated. Each group of samples were trained to obtain the parameters,  $\lambda_i \hat{P}_k$  and the control limits for  $T^2$  and SPE. The trained samples constrain  $T^2$  and SPE under their limits. Basically, for different zones, the limits were different. After that, 210 samples were chosen to test the effect of the method. The result is shown in figure 10.

The first 120 samples were the test with the previously deployed training data and the SPE and  $T^2$  are derived with PCA model. The following 80 samples were originally normal data. To test the

method and the PCA model, and make the figure have a nice discernment, the data were modified once every 4 samples by changing the amplitude of a single variable by 10% around its original value to represent the occurrences of different faults. The last 20 samples were from zone 1 and zone 2, respectively. The PCA models in zone 4 and zone 3 were applied to calculate SPE and  $T^2$ .

As the figure shown, when using the generated model to test the training samples, there's no alarm of exceeding the SPE and  $T^2$  limits. That testifies the effectiveness of the model.

If one of the variables is changed in a certain amplitude, the SPE or the  $T^2$  will exceed the limits to alarm. However, the two statistical variables of SPE and  $T^2$  alarms usually do not happen simultaneously. If the two alarms are both concerned, the faults should not be missed.

As the last 20 samples shown, SPEs are significantly beyond the SPE limit. It means If a PCA model isn't applied at the corresponding zone, the method cannot perform properly. That shows the necessity to divide the piecewise for PCA linearization respectively. By using the proposed method, the linearization error is constrained in specific range.

Although the method is just tested in a single steam pipe, surely it can be expand into the case of steam network which is comprised of several single pipes. But there will be more variables and the practical application will be more complicated.

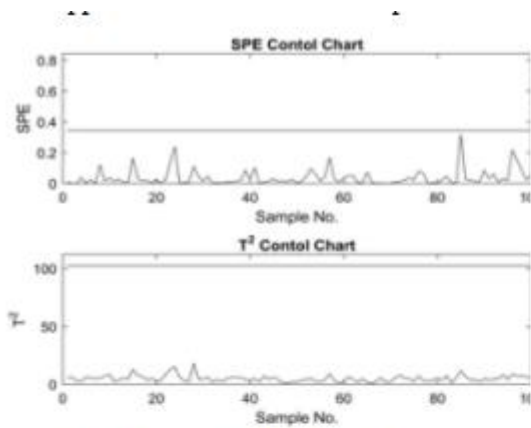


Fig. 6 Control charts for zone 1

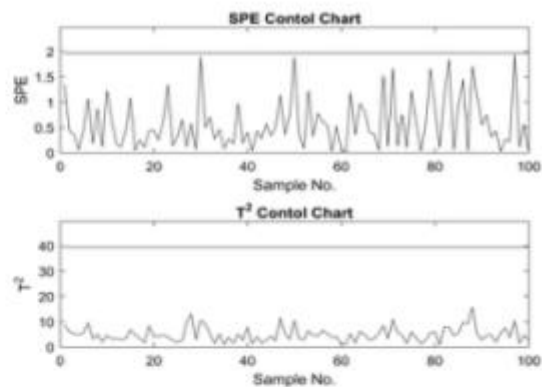


Fig. 7 Control charts for zone 2

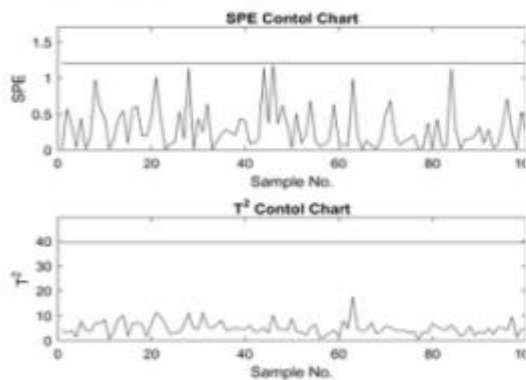


Fig. 8 Control charts for zone 3

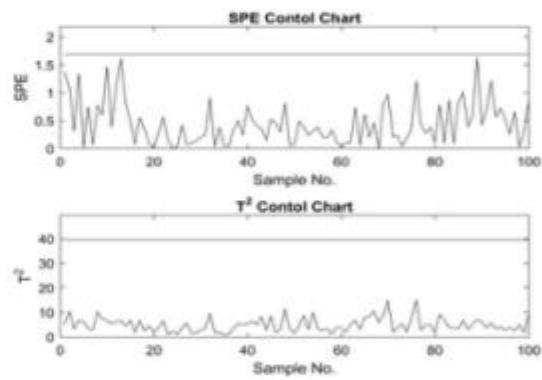


Fig.9 Control charts for zone 4

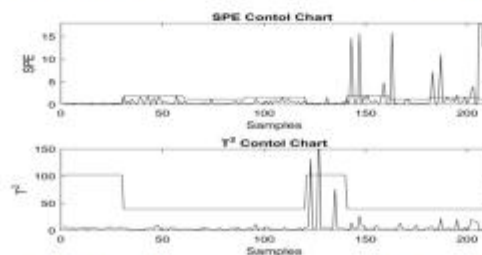


Fig.10 The effect of control chart of monitoring

## Conclusion

Steam networks in the iron& steel enterprises mainly have the important function to reduce the comprehensive energy consumption. It's essential to monitor the state to assure the normal performance. Setting up the models is helpful to partly overcome the problem of data quality.

The mechanism mathematical model of steam pipe is described, and the characteristics are investigated with a study case to show the complexity and nonlinearity of the object.

The paper proposed an innovated method to setup the data-driven model with piecewise performed PCA. The method separates the variable space into different zones and training the data to generate different model for each zone to solve the problem of linearization errors. When monitoring the movement of the system, the data were tested with the corresponding model in which zone the data locate.

For the case study, the data generated by the mathematical model are applied to the proposed method. The result shows the proposed method is effective and sensitive to process faults. The method could be expanded to the steam network with complex structure.

## Acknowledgement

This work is supported under the grant projects of Chinese National Sciences Foundation (No. 61463003 , 51409047), the Chinese Scholarship Council(No.201508360120) and project from Jiangxi Province (JXNE2017-01) .

## References

- [1] Chen, G., An overview of the production and operation of China's iron and steel industry in 2011. *Metallurgical Management*, 2012. (4) : p. 16—19.
- [2] Xianxi, L. and Y. Ming-zhe, Data Processing Approaches for the Measurements of Steam Pipe Networks in Iron& Steel Enterprises, in *Energy Efficiency*. 2012, INTECH: Rijeka,Croatia.
- [3] Shen Yin, S.X.D., Xiaochen Xie,, A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS*,, 2014. 60(11): p. 11.
- [4] S.J. Qin, R.D., Determining the number of principal components for best reconstruction. *Process Control* 2000. 10(2000): p. 6.
- [5] Alcalá, C.F. and S.J. Qin, Reconstruction-Based Contribution for Process Monitoring with Kernel Principal Component Analysis. *Industrial & Engineering Chemistry Research*, 2010. 49(17): p. 7849-7857.
- [6] Choi, S.W., et al., Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and Intelligent Laboratory Systems*, 2005. 75(1): p. 55-67.
- [7] Choi, S.W. and I.B. Lee, Nonlinear dynamic process monitoring based on dynamic kernel PCA. *Chemical Engineering Science*, 2004. 59(24): p. 5897-5908.
- [8] Xianxi;, L., Y. Mingzhe;, and W. Hong;, On Steam Pipe Network Modeling and Flow Rate Calculation. *Procedia Engineering*,, 2012. 2012( 29): p. 1897-1903.
- [9] Congli, M., S. Hongye, and C. Jian, An NT-MT Combined Method for Gross Error Detection and Data Reconciliation. *Chinese Journal of Chemical Engineering*, 2006. 14(5): p. 592-596.
- [10]Tjoa, I.B. and L.T. Biegler, Simultaneous strategies for data reconciliation and gross error detection of nonlinear systems. *Computers and Chemical Engineering*, 1991. 15(10): p. 679-690.
- [11]Huang, J. and X.F. Yan, Dynamic process fault detection and diagnosis based on dynamic principal component analysis, dynamic independent component analysis and Bayesian inference. *Chemometrics and Intelligent Laboratory Systems*, 2015. 148: p. 115-127.
- [12]Luo, X. and S. Li. Nonlinear system monitoring with piecewise performed principal component analysis. in *36th Chinese Control Conference*. 2017. Dalian, China.