Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(7):338-346

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Fault detection of batch process based on multi-way Kernel T-PLS

Zhao Xiaoqiang*, Xue Yongfei and Wang Tao

College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, China

---

ABSTRACT

Because the different batch of batch processes has different raw materials and changeable control conditions, measurement data may lead to drift and it's difficult to obtain complete sampling data at any time during the reaction process. So a multi-way kernel T-PLS (MKT-PLS) algorithm was proposed to improve the fault diagnosis accuracy of batch processes. This algorithm firstly unfolds three dimensional process data matrix by sampling time sequence, then fills process variable data according to certain rules to form complete sample for data missing problem, so obtained appropriate data matrix is used to fault detection by MKT-PLS algorithm. Simulation results of Pensim V 2.0 simulation platform show that the fault detection rate of the proposed algorithm is higher than the other algorithm for detecting faults affect the quality of final products. This algorithm is more suitable to monitor the real-time batch process.

Key words: fault detection; batch process; multi-way kernel T-PLS; Pensim V 2.0

INTRODUCTION

Modern chemical process especially batch process is very important for industrial production. Batch production can satisfy the diversified requirement of different customers and it has high flexibility and better economic effect. In order to maintain batch process in normal condition and manufacture qualified product, process monitoring is necessary and indispensable. Fault diagnosis technology has been studied by many scholars and experts. There are two main approaches of fault diagnosis methods, including first-principle models and process data analysis. The most popular method, multivariate statistical process monitoring (MSPM), can detect abnormal operating situations and diagnosis faults without the exact modeling of the process [1]. MSPM only needs measured data which can obtain easily by sensors in industrial field, so it's not difficult to meet the requirements in modern industrial process. Principal component analysis (PCA) [2] and partial least squares (PLS) [3] are basic projection methods of MSPM for continuous process. They use normal history process data to build predefined model, detect fault with the real-time measurement data. The major advantages of these methods are their abilities to handle large number of highly correlated variables, measurement errors, and missing data [1]. In batch process, measurement data is 3-dimensional form containing the following information: batch number, variable number and sample number. Many scholars studied batch process and proposed excellent strategies to solve the problems, such as multi-way PCA, multi-way PLS, multi-stage PCA, multi-stage PLS. Recently, Zhou [4-6] et al proposed an improved structure, namely total projection to latent structures (T-PLS), for fault diagnosis in continuous process. They make a further decomposition for the PLS model which has better effect than PCA when detecting quality-relevant fault. Zhao [7] et al give a suitable extension for T-PLS model, single space total projection to latent structures (SsT-PLS) and multi space total projection to latent structures (MsT-PLS) were analyzed and compared in their paper. They used T-PLS model to monitor continuous or batch process and simulation results showed that the new method's fault detecting rate is higher than the other's.

However, both Zhou and Zhao's T-PLS models are linear models. It's difficult to obtain accurate fault diagnosis result of actual industrial process when measurement data have strong nonlinear relationship. Kernel function is a kind of tools which have better effect when dealing with nonlinear data. So a multi-way kernel T-PLS algorithm was proposed in this paper. This algorithm use kernel function to solve the nonlinear problem of input data by projecting them onto feature space first, and then build linear T-PLS model with the kernel matrix in high-dimensional feature space. Hotelling's T^2 statistic (including T_y^2 , T_o^2 and T_r^2) and Q_r statistic are structured and calculated in four different subspace to detect different kind fault respectively. Because process data and quality data were preprocessed by kernel function, quality-relevant fault detection rate of multi-way kernel T-PLS algorithm is higher than that of multi-way T-PLS.

This paper is organized as follows. Multi-way kernel T-PLS model is proposed in section 2. Then, quality-relevant fault diagnosis methods based on Multi-way kernel T-PLS model is described in section 3. In section 4, a simulation study of penicillin fermentation process based on Pensim V 2.0 is illustrated and results of Multi-way kernel T-PLS model and Multi-way kernel PLS model are compared. Finally, conclusions are presented in the last section.

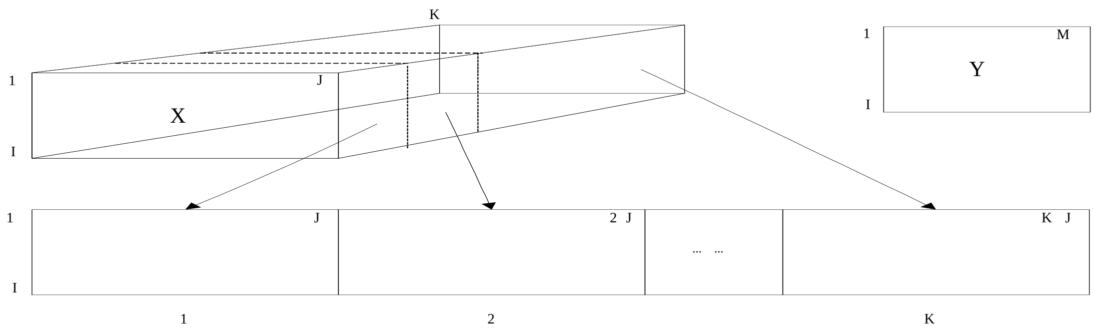
MULTI-WAY KERNEL T-PLS(MKT-PLS)

Similar to the multi-way principal component analysis and multi-way partial least squares method, multi-way kernel T-PLS modeling contains three steps. Firstly, three dimensional batch process data switch to two dimensional plane data according to the order of sampling time. Secondly, linearize the process data by projecting them onto high dimensional feature space using kernel function. Finally, build linear T-PLS model in high-dimensional feature space, calculate monitoring statistics and their corresponding control limits, compare them detect quality-relevant fault.

Batch Process Data Preprocessing

Batch process' measurement data is stored as stereoscopic data matrix, and it contains the following information of process: batch number, measured variables number and sampling number. Commonly, there are two methods to make the stereoscopic data matrix planarization. First, partition stereoscopic matrix is determined according to the batch number. Second, partition stereoscopic matrix is determined according to the sampling number. In this paper, we choose the second method to preprocess batch measured data. We cut the stereoscopic matrix X into slices in accordance with the sampling time sequence from the first sampling point to the K-th sampling point, and then tile them into a plane data matrix orderly. The schematic plan of batch process data preprocess showed as fig. 1.

Fig.1: Schematic plan of stereoscopic data matrix preprocessing



$I*KJ$ dimensional data matrix can obtain from the above complanation method, but it's very difficult to use them modeling directly due to its huge dimension. Since the k-th sample of the j-th variable has been measured I times totally and the I times measurements obey certain statistical characteristics, so it can use the average value of weighted moving window $\bar{x}_{\mathrm{jk}} = \frac{1}{I}\prod_{\mathrm{i}=1}^{I}(x_{ij\mathrm{k}}\lambda^{I-\mathrm{i}})$ instead. $I*KJ$ dimensional data matrix simplify to $1*KJ$ dimensional row vector. Partitioning and rearranging the row vector according the order $J, 2J, \cdots, KJ$, new process data matrix obtain as the following formula.

$$\bar{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_J \\ \bar{x}_{J+1} & \bar{x}_{J+2} & \cdots & \bar{x}_{2J} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{(K-1)J+1} & \bar{x}_{(K-1)J+2} & \cdots & \bar{x}_{KJ} \end{bmatrix} \tag{1}$$

As quality data can only be measured at the end of each batch off-line, so $I*M$ dimensional data matrix Y can be obtained after I batch production. Calculate the average value of final product quality measurements from first batch to I -th batch by weighted moving window. New quality data matrix $\overline{Y}$ can calculate as the following formula, where $\overline{\mathrm{y}}_{\mathrm{m}} = \frac{1}{I}\prod_{\mathrm{i}=1}^{I}(\mathrm{y}_{im}\lambda^{I-\mathrm{i}})$.

$$\overline{Y} = [\overline{\mathrm{y}}_1 \quad \overline{\mathrm{y}}_2 \quad \cdots \quad \overline{\mathrm{y}}_M] \tag{2}$$

It can detect fault for the batch process after structuring kernel T-PLS model using process data matrix $\overline{\mathrm{X}}$ and quality data matrix $\overline{Y}$. But it is worth noting that outliers exist at each batch process, in order to ensure the fault detection accuracy of batch process fundamentally, appropriate algorithms should be used to remove outlier value.

Multi-way Kernel T-PLS

Compared with continuous process, batch process' data is stored stereoscopically. So it's necessary to planarize the stereoscopic data before modeling. Using the above method obtain process data matrix $\overline{\mathrm{X}}$ and quality data matrix $\overline{Y}$, and then build MKT-PLS model with these matrix. Cover theorem states that "a set of training data that is not linearly separable, one can with high probability transform it into a training set that is linearly separable by projecting it into a higher dimensional space via some non-linear transformation"[8]. Kernel matrix $K = \Phi^T\Phi$ can obtain and be used to build linear T-PLS model for fault diagnosis after process data matrix $\overline{\mathrm{X}}$ mapping to a high dimensional feature space by radial basis function. Nonlinear model MKT-PLS between input-output space equal to linear model MT-PLS between feature-output space theoretically. Concrete steps to establish MKT-PLS model are as follows:

Step1: obtain $K*J$ dimensional process data matrix $\overline{\mathrm{X}}$ and $1*M$ dimensional quality data matrix $\overline{Y}$ using above method；

Step2: obtain process data matrix $\hat{\mathrm{X}}$ and quality data matrix $\hat{Y}$ by standardizing $\overline{\mathrm{X}}$ and $\overline{Y}$;

Step3: project matrix $\hat{\mathrm{X}}$ onto feature space F , $\Phi : x_i \in R^N \rightarrow \Phi(x_i) \in F$, structure kernel matrix $K = \Phi^T\Phi$;

Step4: $i=1$, $K_i = K$, $Y_i = \hat{Y}$, extract convergent u_{i} from Y_i ;

(1) u_{i} equal to a column of Y_i randomly;

(2) $t_i = K_i u_i$, $t_i \leftarrow t_i / \|t_i\|$;

(3) $q_i = Y_i^T t_i$;

(4) $u_i = Y_i q_i$, $u_i \leftarrow u_i / \|u_i\|$;

(5)verdict: if u_i convergence, go to Step5；otherwise, go to (1);

Step5: calculate the loading matrix of K_i : $p_i = K_i^T t_i$;

Step6: extract all principal component, calculate T, P, U and Q;

(1) $K_{i+1} = K_i - t_i p_i^T$, $Y_{i+1} = Y_i - u_i q_i^T$ ；

(2) $i = i+1$, repeat Step4 and Step5 until all principal component have been extracted, the number of principal component determined by cross validation;

(3) $T = [t_1, \cdots, t_A]$, $P = [p_1, \cdots, p_A]$, $U = [u_1, \cdots, u_A]$, $Q = [q_1, \cdots, q_A]$;

Step7: $K = TP^T + E$, $Y = UQ^T + F$;

Step8: run PCA algorithm on TP^T , $TP^T = t_y p_y{}^T + T_0 P_0{}^T$, the number of principal component is A-1;

Step9: run PCA algorithm on E, $E = T_{\mathrm{r}} P_r{}^T + E_r$, the number of principal component refer to [9].

Process data matrix $\hat{\mathrm{X}}$ and quality data matrix $\hat{Y}$ of batch process can be decomposed as follows after running MKT-PLS algorithm.

$$\hat{X} = t_y p_y^{\ T} + T_0 P_0^{\ T} + T_r P_r^{\ T} + E_r \qquad (3)$$

$$\hat{Y} = t_y + F \qquad (4)$$

$$E_r = E(I - P_r P_r^{\ T}) \qquad (5)$$

Where t_y , T_0 and T_r are score vectors of different subspaces, E_r is final residual that represents noise. The decomposition of MKT-PLS model is shown in fig.2.

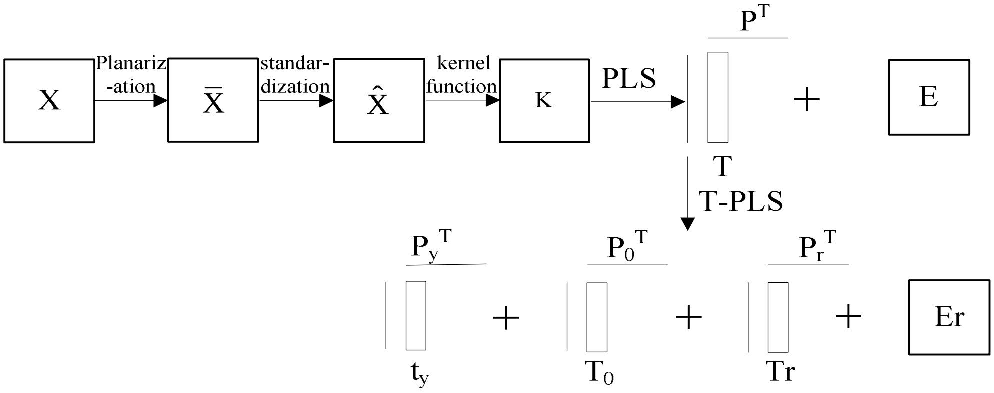


Fig.2: Decomposition structure of MKT-PLS model

FAULT DETECTION STEPS BASED ON MKT-PLS

Fault detection based MKT-PLS model includes two steps: first, establish MKT-PLS model using normal historical offline data, compute monitoring statistics and determine their corresponding control limits; Second, calculate D and Q statistic using online measured data, detecting quality-relevant fault in real time by comparing statistic and its control limit.

Data Supplement

Because batch process' measurement data is stored as whole, it's hard to get the whole sample until the batch production process finish. This is a disadvantage of real-time monitoring for batch process. However, only detecting fault in time and taking appropriate actions can avoid unnecessary losses in actual industrial production process. In order to solve this problem, Nomikos et al[10] proposed the following three filling algorithm to predict defaults from current time to the end of entire batch process. (1)0 value fill method, the data of future is considered not deviating from the average trajectory; (2)current value fill method, the data of future is considered have the same deviation from the average trajectory; (3) projection method, the data of future is determined by the current value projected onto a particular space. This paper chose the third method to fill the unsampled data * of the process variable data matrix $\bar{X} = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \cdots & \bar{x}_J \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_{kJ+1} & \bar{x}_{kJ+2} & \cdots & \bar{x}_{(k+1)J} \\ * & * & \cdots & * \end{bmatrix}$, so as to obtain a complete sample of batch process for fault detection.

Statistics And Control Limits

For batch process, there are many differences between each batch, such as different raw materials, changed production environment, operating point drift of equipments. All of these should be considered carefully before establishing MKT-PLS model. Weighted moving window was used to obtain the latest normal batch process' weighted average measurement data matrix $\bar{X}$, and then it was mapped to feature space non-linearly obtain kernel matrix K, score and residual vector of K are as the following formula.

$$
\begin{aligned}
\bar{t}_y &= qR^T K \in R^1 \\
\bar{t}_o &= P_o^T (P - p_y q) R^T K \in R^{A-1} \\
\bar{t}_r &= P_r^T (I - PR^T) K \in R^{A_r} \\
\bar{E}_r &= (I - P_r P_r^T)(I - PR^T) \in R^m
\end{aligned}
\qquad (6)
$$

MKT-PLS model's statistics $\bar{T}_y^2$, $\bar{T}_o^2$, $\bar{T}_r^2$ and $\bar{Q}_r$ are calculated using the weighted average value of I normal batch history data. If the measured variables conform to normal distribution, statistics and its corresponding control limits are shown in Table.1.

Table.1: Statistics and its corresponding control limits

Statistic	computational formula	control limits formula	Statistic	computational formula	control limits formula
$\bar{T}_y^2$	$\bar{t}_y^T \Lambda_y^{-1} \bar{t}_y$	$\frac{n+1}{n} F_{1,n-1,\alpha}$	$\bar{T}_r^2$	$\bar{t}_r^T \Lambda_r^{-1} \bar{t}_r$	$\frac{A_r(n^2-1)}{n(n-A_r)} F_{A_r,n-A_r,\alpha}$
$\bar{T}_o^2$	$\bar{t}_o^T \Lambda_o^{-1} \bar{t}_o$	$\frac{(A+1)(n^2-1)}{n(n-A+1)} F_{A-1,n-A+1,\alpha}$	$\bar{Q}_r$	$\lVert \bar{E}_r \rVert^2$	$g\chi^2_{h,\alpha}$

Off-line Modeling

MKT-PLS algorithm calculates four statistics and their corresponding control limits in different sub-space to monitor the batch process, detect fault by analyzing and comparing them in sub-space monitoring chart. Similar to T-PLS algorithm, the main variation of process uses Mahalanobis distance to measure while the residual part of process uses Euclidean distance. The detailed steps of modeling offline based on MKT-PLS algorithm are as following.

Step1: obtain the latest I batch process data matrix $\bar{\mathrm{X}}$ and quality data matrix $\bar{Y}$ using above method;

Step2: obtain process data matrix $\hat{\mathrm{X}}$ and quality data matrix $\hat{Y}$ by standardizing $\bar{\mathrm{X}}$ and $\bar{Y}$;

Step3: project matrix $\hat{\mathrm{X}}$ onto feature space F non-linearly, structure kernel matrix $K = \Phi^T \Phi$;

Step4: decompose kernel matrix K , calculate $t_y, p_y, T_o, P_o, T_r, P_r, E_r, u_y, q_y, F$;

Step5: structure statistics $\bar{T}_y^2$, $\bar{T}_o^2$, $\bar{T}_r^2$ and $\bar{Q}_r$, choose their corresponding control limits according table.1 with suitable degrees of freedom and confidence.

On-line Detection

Since the complete sampling of batch process can't obtain until the whole reaction finished, in order to realize real-time monitoring, unsampled values must be filled as the above method before detecting fault. MKT-PLS model uses the complete sample matrix which has been filled with appropriate data as input, calculates monitoring statistics in four specific sub-space, compares them with their corresponding control limits to detect fault. The concrete steps of on-line detection are as follows:

Step1: project the measured value $[X_1, X_2, \cdots, X_k]$ of sample time $1T, 2T, \cdots, (k-1)T, kT$ onto a special space, obtain the value $[X_{k+1}, \cdots, X_K]$ of sample time $(k+1)T, \cdots, KT$. The complete matrix X_{new} =$[X_1, X_2, \cdots, X_k, \ X_{k+1}, \cdots, X_K]$ has obtained, the concrete algorithm can refer to [65];

Step2: standardize process variable matrix X_{new} to $\hat{X}_{new}$;

Step3: use $\hat{X}_{new}$ as MKT-PLS model's input , calculate the following value respectively $t_{y,new}, p_{y,new}, T_{o,new}, P_{o,new}, T_{r,new}, P_{r,new}, E_{r,new}, u_{y,new}, q_{y,new}, F_{new}$;

Step4: calculate statistics $\bar{T}^2_{\mathrm{y,new}}$, $\bar{T}^2_{\mathrm{o,new}}$, $\bar{T}^2_{\mathrm{r,new}}$ and $\bar{Q}_{r,new}$ respectively;

Step5: detect quality-relevant fault: if statistics $\bar{T}^2_{\mathrm{y,new}}$ and $\bar{Q}_{r,new}$ beyond their corresponding control limits, the fault can affect final product's quality;

if statistics $\overline{T}^2_{\mathrm{o,new}}$ and $\overline{T}^2_{\mathrm{r,new}}$ beyond their corresponding control limits, the fault can't affect final product's quality;

SIMULATION RESULTS AND ANALYSIS

Penicillin fermentation process is a typical batch production process, which is widely used to evaluate the control strategies and algorithms for batch process. This batch process comprises nine variables (ventilation rate, stirring power, acceleration of the bottom stream, temperature of the fermentation vessel, pH, acceleration of the cooling water flow, medium volume, oxygen saturation and carbon dioxide concentration) and four quality variables (penicillin concentration, substrate concentration, biomass concentration and reaction heat). This paper uses Pensim V 2.0 simulation software for testing, which was developed by Professor Cinar and his research team of Illinois Institute of Technology from 1998 to 2002 [11], Pensim V 2.0 simulation software has high practical value. It can not only simulate penicillin fermentation process realistically, but also obtain a series of parameters of the fermentation process easily. Pensim V 2.0 has become an useful tools to validate scientific approach based on data-driven for diagnose fault and monitor batch process[12-14]. Initialization relevant parameters should be executed before simulating penicillin fermentation process using Pensim V 2.0 software. The default values and the range of relevant parameters are shown in Table.2 and Table.3.

Table.2: Default value and value range of Pensim V 2.0

Variable name	Unit	Default value	Value range
substrate concentration	G/L	15	14-18
oxygen concentration	G/L	1.16	1-1.2
biomass concentration	G/L	0.1	0
penicillin concentration	G/L	0	0
medium volume	L	100	100-104
carbon dioxide concentration	G/L	0.5	0.5-1
hydrogen ion concentration	MOL/L	$10^{-5.1}$	
pH		5	4.5-5.5
temperature of the fermentation vessel	K	298	295-301

Table.3: Control parameter of Pensim V 2.0

Variable name	Unit	Default value	Range
ventilation rate	L/H	8.6	8-9
stirring power	W	30	29-31
acceleration of the bottom stream	L/H	0.042	0.039-0.045
substrate heat	K	296	295-296
pH		5	4.95-5.05
reaction heat	K	298	297-298

Penicillin fermentation process simulation study base on Pensim V 2.0 contains those steps: set the initial values of variables and parameters, set the controller of temperature and pH, set reaction conditions (normal/fault), export real-time monitor plot of each process variable, export sample data matrix of each process variable. The detailed steps see Fig.3.

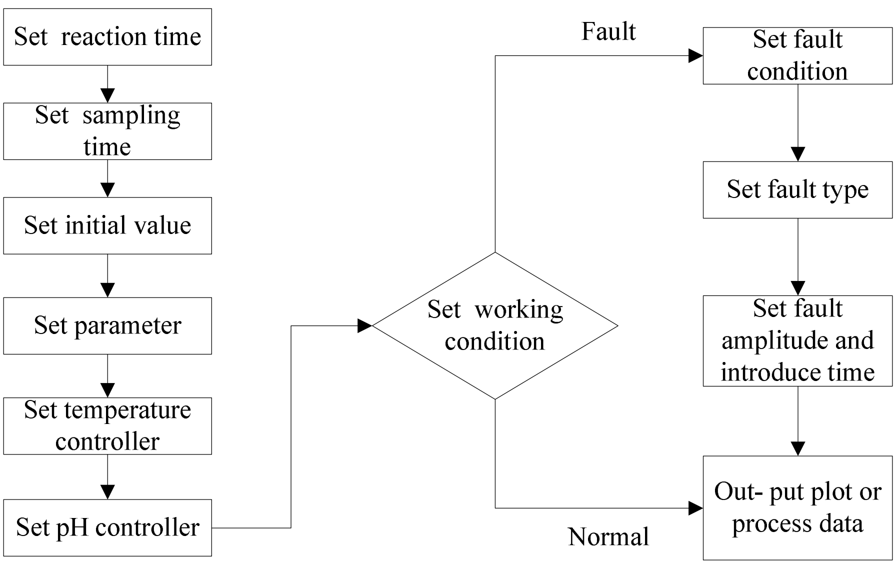


Fig.3: Flow chart of Penicillin fermentation process based on Pensim V 2.0

The actual penicillin fermentation process often occur valve leakage or pump failure, Pensim V 2.0 simulate three types of failures during the actual fermentation process. The fault's introduction time and end time can be set artificially, step and ramp signal can be set as the fault type, the default faults of Pensim V 2.0 are described in Table 4.

Table 4 The Default fault of Pensim V 2.0

Fault Number	Fault Description	Fault type
Fault 1	Aeration Rate Fault	Step / Ramp
Fault 2	Agitator Power Fault	Step / Ramp
Fault 3	Substrate Feed Rate Fault	Step / Ramp

In order to obtain different batch's process variables, 20 batches of normal penicillin fermentation process simulation have been done by adjusting the value of each parameter in Table.2 and Table.3. Fig.4 shows the trajectory of each process variable of penicillin fermentation process in normal condition based on Pensim V 2.0.

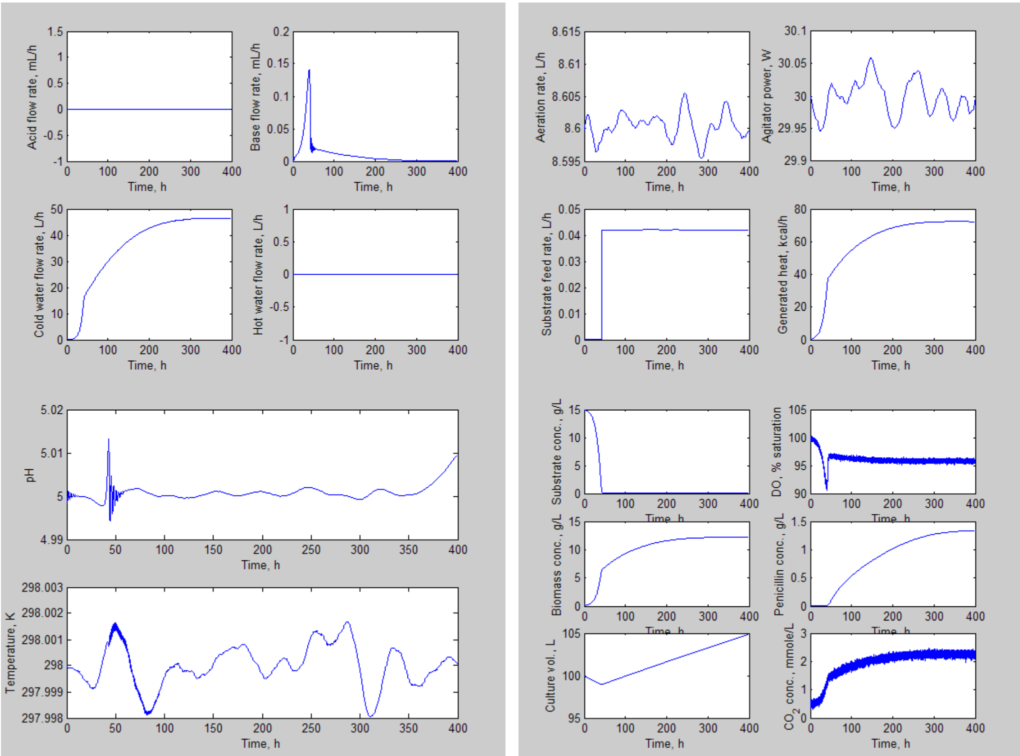


Fig.4 :Trajectory of each process variable of penicillin fermentation process (normal condition)

The 20 batches normal process data were used to build off-line modeling, and control limit of each monitoring statistic determined as section 3.3. In order to obtain fault samples, 'fault introduce time, fault end time, fault type, fault amplitude' should be set in type 9 of Pensim V 2.0 initialization. This paper set simulation time: 400h, sampling time: 0.2h, fault introduce time: 200h, fault end time: 400h, fault type: step, fault amplitude: -50. 2000*18 dimensional fault sample data matrix $X_{2000\times18}$ can obtain after the whole process finished. Trajectory of each process variable of penicillin fermentation process in fault 3 condition based on Pensim V 2.0 are showed in Fig.5.

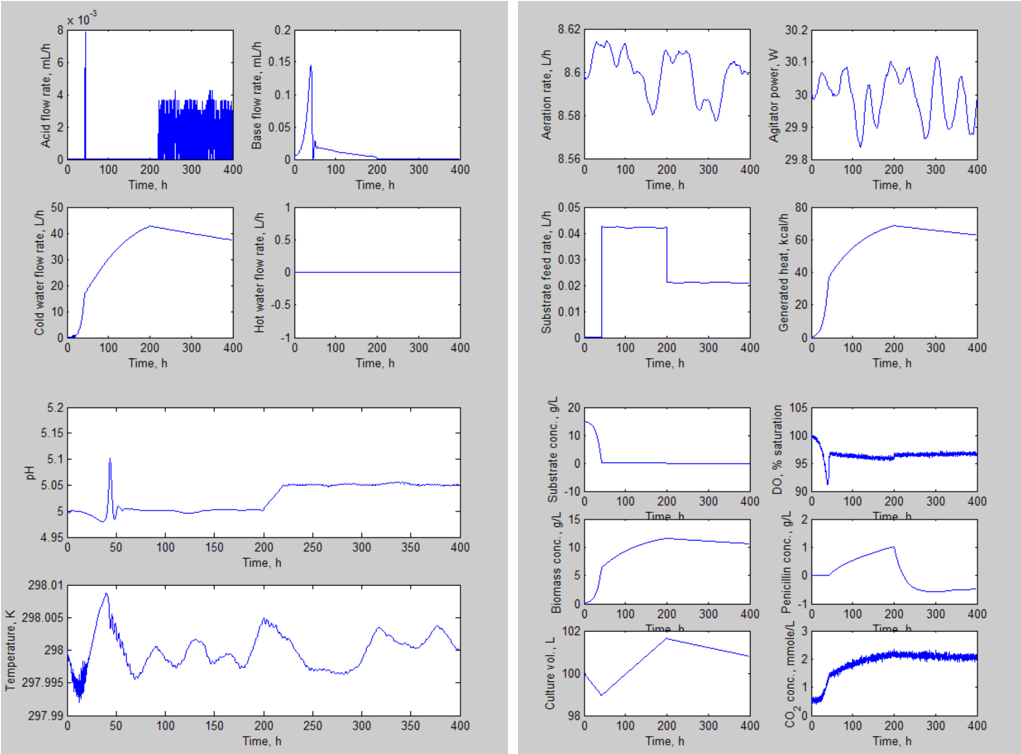
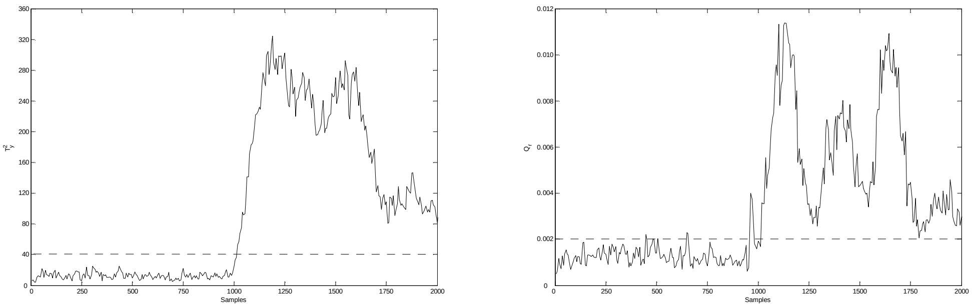


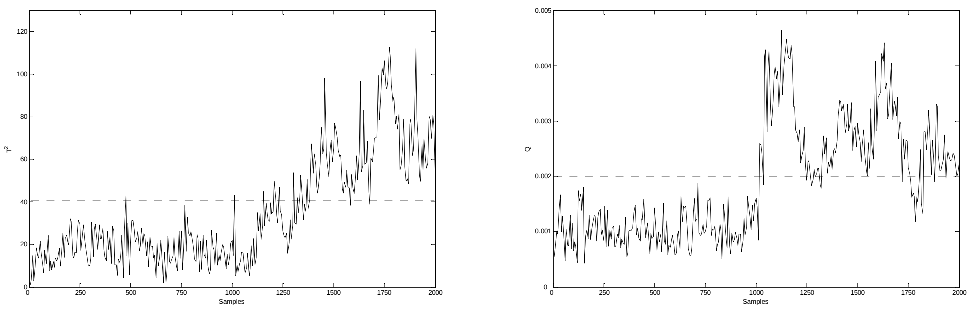
Fig.5: Trajectory of each process variable of penicillin fermentation process (fault 3 condition)

Use the above fault data matrix $X_{2000\times18}$ as MKT-PLS model's input, process monitor diagram based on MKT-PLS is showed in Fig.6. In order to highlight the superiority of proposed method, comparison algorithm was run with the same input. Fig.7 is process monitor diagram based on MKPLS.



(a) (b)

Fig.6: Monitor diagram based on MKT-PLS



(a) (b)

Fig.7: Monitor diagram based on MKPLS

Obviously, fault detection rate of MKT-PLS model is higher than that of MKPLS model, because MKT-PLS model decompose the orthogonal portion of variation from projection subspace and obtain real noise in residual subspace.

CONCLUSION

This paper proposed a suitable model, MKT-PLS which is an appropriate extended of KT-PLS algorithm, to detect fault for batch process. Fault diagnosis based on MKT-PLS must obtain appropriate data matrix first, so some data preprocess method was introduced and used. What's more, due to the shortcoming that complete sample can't obtain until whole process finish, supplementary data technique was used to fill the blank. Simulation results of penicillin fermentation process show that, MKT-PLS algorithm has higher fault detection rate, which means that it's more suitable to detect quality-relevant fault and monitor batch process real-time.

Acknowledgments
The authors wish to thank the National Natural Science Foundation of China for contract 51265032 and61263003, the university basic scientific research project of Gansu province for contract 1203ZTC061.

REFERENCES

[1] Donghua Zhou, Gang Li. *AIChE Journal*, **2010**, 56, 168-178.
[2] Ningyun Lu, Yuan Yao, Furong Gao, Fuli Wang. *AIChE Journal*, **2005**, 51, 3300-3304.
[3] Yingwei Zhang, Zhiyong Hu. *AIChE Journal*, **2011**, 89, 2667-2678.
[4] Gang Li, S. Joe Qin, Yindong Ji, Donghua Zhou. *Acta Automatica Sinica*, **2009**, 35, 759-765.
[5] Gang Li, Donghua Zhou, S. Joe Qin. Output-relevant fault reconstruction based on total PLS. In: Proceedings of *the 8th World congress on Intelligent Control and Automation (WCICA). Jinan, China*, **2010**, 1718-1722.
[6] Gang Li, S Joe Qin, Donghua Zhou. *Industrial and Engineering Chemistry Research*, **2010**, 49, 9175-9183.
[7] C Zhao, Y Sun. The multi-space generalization of total projection to latent structures (MsT-PLS) and its application to online process monitoring. 2013 10th *IEEE International Conference on Control and Automation (ICCA), Hangzhou, China*, **2013**, 1441-1446.
[8] T. M. Cover. *IEEE Transactions on Electronic Computers*, **1965**, 14(3), 326-334.
[9] S. Joe Qin, Dunia R. *Journal of Process Control*, **2000**, 10(2), 245-250.
[10] Nomikos Paul, Gregor John F. *Technometrics*, **1995**, 37(1), 41-59.
[11] Gulnur Birol, Cenk Undey, Ali Cinar. *Computers and Chemical Engineering*, **2002**, 26(11), 1553-1565.
[12] Yew Seng Ng, Rajagopalan Srinivasan. *Computers and Chemical Engineering*, **2009**, 33(4), 887-902.
[13] Xiang Gao, Wenfeng Yi, Yumei Sun. *Scientific Journal of Control Engineering*, **2013**, 3(6), 384-390.
[14] Shallon Stubbs, Jie Zhang, Julian Morris. I*nd. Eng. Chem. Res*., **2013**, 52(35), 12399-12407.