INVESTIGATION OF DYNAMIC MULTIVARIATE PROCESS MONITORING

Lei Xie, Shu-qing Wang, Jian-ming Zhang

National Key Laboratory of Industrial Control Technology,
Zhejiang University, Hangzhou 310027, People's Republic of China

Abstract: Chemical process variables are always driven by random noise and disturbances. The closed-loop control yields process measurements that are auto & cross correlated. The influence of auto & cross correlations on statistical process control (SPC) is investigated in detail. It is revealed both auto and cross correlations among the variables will cause unexpected false alarms. Dynamic PCA and ARMA-PCA are demonstrated to be inefficient to remove the influences of auto & cross correlations. Subspace identification based PCA (SI-PCA) is proposed to improve the monitoring of dynamic processes. Through state space modelling, SI-PCA can remove the auto & cross correlations efficiently and avoid unexpected false alarms. The application in Tennessee Eastman challenge process illustrates the advantages of the proposed approach. Copyright © 2005 IFAC

Keyword: Multivariate statistical processes control (MSPC); Subspace identification, False alarms rate (FAR), Dynamic processes

1. INTRODUCTION

With the advent of improved instrumentation and automation, chemical processes now produce large volumes of information which are highly correlated. Several multivariate statistical methods have been developed to identify the correlations between variables and create a reduced set of variables in the orthogonal axes that capture most of the variability in the collected information. One of most popular MSPC methods is principal component analysis (PCA), which also has been applied to chemical processes (Nomikos 1995, Wise 1996, Cinar 1999, Venkatasubramanian 2003).

However, PCA is based on the assumption that the process variables are independent and identically normally distributed (iid), i.e., stationary or uncorrelated in time (Jackson 1991, Ku 1995). In practice, this assumption is always violated, as chemical process variables are driven by random noise and disturbances. Due to the feedback control, the impact of disturbances propagates through to both the input and output variables. Thus the variables will move around the steady state and exhibit some degrees of auto-correlation and cross-correlation.

In order to monitor the process dynamics, PCA has been extended to include the time-series structures of variables (Ku 1995, Negiz 1997, Callao 2003, Simoglou 2002). Among these extensions, dynamic PCA (DPCA) by Ku et al (1995) is widely adopted and can be treated as a multivariate AR-like time series modelling approach. Although applications in Tennessee Eastman (Ku 1995) and some batch processes monitoring (Chen 2002) have demonstrated the efficiencies of dynamic PCA, it is proved recently that dynamic PCA can not eliminate the auto & cross correlations of variables (Kruger 2004). If a dynamic PCA is used, the score variables will be auto and cross-correlated even when the process variables are neither auto nor cross correlated. In other words, dynamic PCA will always induce the dynamics into the score variables. In order to overcome this question, Kruger et al involved ARMA filters to remove the auto correlations from the PCA scores. But unfortunately, the cross correlations still remain in the filtered score variables.

The contributions of this paper are as follows. First, the influences of auto & cross correlations of process variables are investigated through a numerical experiment. It is revealed that the presence of auto and cross correlations will both cause the false alarms Secondly, criterions to determine whether variables are auto or cross correlated are introduced. The ARMA filtering approach suggested by Kruger is inefficient to reduce the cross correlations of PCA scores. Thirdly, a subspace identification modelling approach combined with PCA is proposed to remove the entire dynamics from the score variables. A novel information based criteria is presented to determine the order of relative state-space model. Fourthly, the effectiveness of proposed approach is demonstrated using the Tennessee Eastman process. Finally, some remarks and conclusions are presented.

2. MONITORING CROSS CORRELATED PROCESS VARIABLES

2.1 PCA, Dynamic PCA and ARMA-PCA

Given the measurements, normal operating process data are collected and put in a two-dimensional data matrix $\boldsymbol{X} \in R^{N*n}$ with N samples and n variables. PCA decomposes the data matrix $\boldsymbol{X}$ in terms of r linear principal components with $r \le n$:

$$\begin{aligned}\boldsymbol{X} &= \boldsymbol{XPP}^T + \boldsymbol{X}\widetilde{\boldsymbol{P}}\widetilde{\boldsymbol{P}}^T \\ &= \boldsymbol{TP}^T + \boldsymbol{E}\end{aligned}, \qquad (1)$$

where $\boldsymbol{P} \in R^{n*r}$ and $\boldsymbol{T} = \boldsymbol{X} * \boldsymbol{P} \in R^{N*r}$ are defined as the principal component loadings and scores, respectively. $\widetilde{\boldsymbol{P}}$ contains the retained principal component directions. $\boldsymbol{E}$ is the residual matrix. When PCA is applied to monitor a process, T^2 and SPE statistics are commonly used. A more detailed analysis of PCA refers to Jackson (1991).

Dynamic PCA arranges the process variables to form an autoregressive (AR) structure:

$$\underline{\boldsymbol{X}} = [\boldsymbol{X}_0 \ \boldsymbol{X}_{-1} \cdots \boldsymbol{X}_{-d}] \in R^{(N-d)*(d+1)n}, \qquad (2)$$

where $\underline{\boldsymbol{X}}$ is an augmented set of variables, representing an AR model structure of order d and the subscripts 0, 1, d refer to the backshifts applied. PCA is applied to $\underline{\boldsymbol{X}}$ and the corresponding T^2 and SPE statistics can be obtained.

It is demonstrated by Kruger (2004) that the retained scores variables are auto-correlated irrespective of whether the process variables are auto-correlated or not. In order to overcome the deficiencies of dynamic PCA, Kruger et al (2004) incorporate ARMA filters in the PCA analysis (ARMA-PCA). The ARMA-PCA method applies traditional PCA method to original data matrix $\boldsymbol{X}$ first and r ARMA filters are identified to remove the auto correlations of each score variable. Although the auto correlations are efficiently eliminated, the cross correlations still exist and the filtered scores are not independent yet.

2.2 Performance investigation of ARMA-PCA

Type I error rate or false alarms rate refers to the percentage of statistics violating its confidence bound when monitoring normal operating process. For PCA, the ideal type I error rate for T^2 statistic is equal to the significant level α. In practice, false alarms rate is one of the most important parameters to determine. Too high false alarms rate will result in poor acceptance among shop-floor personnel while too low rate will make the monitoring system insensitive to potential process or sensor faults.

The influence of auto-correlation on process monitoring results has been well studied in Kruger (2004). In this section, it will be illustrated that not only auto- but also cross- correlation will result in higher false alarms rate in T^2 monitoring chart.

In order to study the cross correlation effect solely, it is assumed that PCA score variables has been filtered with the PCA-ARMA approach presented in Kruger (2004). And there are 3 filtered scores have the following description:

$$\begin{cases} t_i^1 = 3w_i \\ t_i^2 = 2w_{i-1} \\ t_i^3 = w_{i-2} \end{cases} \qquad (3)$$

, where $w_i \in N(0,1)$ is a sequence of normally distributed values with zero mean and unit variance. The filtered score variables are all shifted sequences of w_i. It is clear that the filtered scores are only cross correlated. In order to study the influence of cross correlation on the T^2 statistic, 2000 samples are generated with (3). The first 1000 samples are selected as reference data and the remaining 1000 samples served as testing data.

The T^2 monitoring chart is given in Fig.1 (solid line represents the 99% control limit) and shows that the number of violations exceeds 1% significantly around the 100th sample. Consecutive false alarms also appear around the 150th samples which will mislead the process engineers.

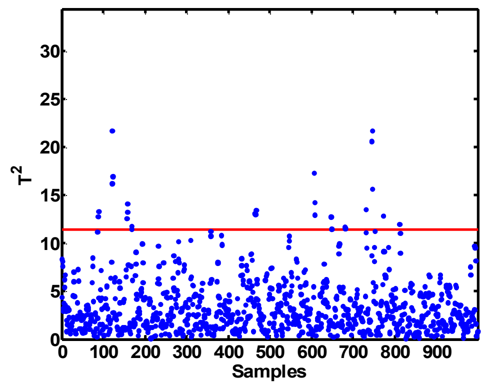

Fig.1 False alarms caused by cross correlations among PCA score variables

Further more, because ARMA-PCA approach ignores the cross correlations of score variables, this will lead to ARMA models with higher orders than necessary. For example, in the example 1 studied by Kruger (2004), the underlying model order is 4 while the AR orders selected are all larger than 4 except for the last score variable.

3. SUBSPACE IDENTIFICATION BASED PCA

3.1 Auto & cross correlation coefficients

Consider a set of random time series $\boldsymbol{u} = [u_1\ u_2 \cdots u_r]$, the auto & cross covariance coefficient is defined as:

$$\rho_{i,j}(\tau) = \frac{Cov(u_i(t), u_j(t-\tau))}{\sqrt{Cov(u_i(t), u_i(t))Cov(u_j(t), u_j(t))}},$$
$$1 \le i, j \le r$$
, (4)

$$Cov(u_i(t), u_j(t-\tau)) =$$
$$\begin{cases} \frac{1}{N-\tau}\sum_{t=\tau+1}^{N}(u_i(t) - \overline{u_i})(u_j(t-\tau) - \overline{u_j}), & \tau > 0 \\ Cov(u_j(t), u_i(t-(-\tau))), & \tau < 0 \end{cases},$$
$$\overline{u_i} = \frac{1}{N}\sum_{t=1}^{N} u_i(t),$$
(5)

where $Cov(u_i(t), u_j(t-\tau))$ is the cross covariance between u_i and u_j, N is the number of samples. For multi-normally distributed random variables (independent), $\rho_{i,j}(\tau) = 0$ when $\tau \ne 0$.

The auto & cross correlation coefficients can be defined in a similar manner. In the following sections, it is assumed process is stationary and variables are zero-centred, so the correlation coefficient equals the covariance coefficient and both of them are denoted as ACFs.

The following two theorems are introduced to determine whether variables are auto or cross-correlated:

Theorem 1: Suppose $u_i(t)$ is a Gauss white noise sequence (so independent) and $N >> M$. Then the auto correlation coefficient $\rho_{i,i}(\tau)$ is approximately normally distributed, i.e. $\sqrt{N}\rho_{i,i}(\tau) \sim N(0,1), 1 \le |\tau| \le M$.

Theorem 2: Suppose $u_i(t)$ is a Gauss white noise sequence and independent of $u_j(t)$ and $N >> M$. Then the cross correlation coefficient $\rho_{i,j}(\tau)$ is approximately normally distributed, i.e.

$$\sqrt{N}\rho_{i,j}(\tau) \sim \sqrt{\sum_{m=-M}^{M}\rho_{i,i}(m) \cdot \rho_{j,j}(m)} \cdot N(0,1),$$
$$1 \le |\tau| \le M$$
.

The detailed proof of theorem 1 and 2 refers to Ljung (1999, Section 16.6), the condition of theorem 2 can be relaxed as $u_i(t)$ being the linear combination of time shifted Gauss white noises.

3.2 Subspace identification based PCA (SI-PCA)

As mentioned above, Kruger's ARMA-PCA approach ignores the cross correlations of score variables. In order to describe the auto & cross correlations simultaneously, one should consider that the ARMA approach can be extended to multivariate case and a multivariate time series model should be established. However, the multivariate ARMA model is much more difficult to analyze because all the coefficients in univariate model will become matrices and the model complexity will grow rapidly as the model order increases (George 1994).

An alternative approach to analyze multivariate time series is state-space modelling. As argued by Ljung (2002) that "Generally speaking, it is preferable to work with state-space models in the multivariate case, since the model structure complexity is easier to deal with" and any linear model structure (ARX, ARMA etc.) can be represented by state-space model (Ljung 1999).

A state-space model for time series is given by:

$$\begin{aligned} \boldsymbol{z}_{k+1} &= \boldsymbol{A}\boldsymbol{z}_k + \boldsymbol{K}\boldsymbol{e}_k \\ \boldsymbol{t}_k &= \boldsymbol{C}\boldsymbol{z}_k + \boldsymbol{e}_k \end{aligned}, \qquad (6)$$

where $\boldsymbol{t}_k \in R^{r*1}$ is the corresponding time series, $\boldsymbol{z}_k \in R^{nz*1}$ is the vector of state variable, $\boldsymbol{A}$, $\boldsymbol{C}$ are the system matrices, $\boldsymbol{K}$ is the Kalman gain. Note that there are no inputs in this model. The residuals of state-space model $\boldsymbol{e}_k \in R^{r*1}$, which are assumed i.i.d., are employed to process monitoring instead of correlated scores.

Subspace identification (SI) algorithms have been widely adopted to identify the state-space model from input-output data because it does not need iterative, nonlinear optimization as maximum likelihood method and SI is very easy to implement (Overschee 1996).

For subspace algorithms, it is crucial to estimate the state $\boldsymbol{z}_k$ which is defined as a linear combination of past outputs,

$$\begin{aligned} \boldsymbol{p}_k &= [\boldsymbol{t}_{k-1}^T\ \boldsymbol{t}_{k-2}^T \ldots \boldsymbol{t}_{k-d}^T]^T \\ \boldsymbol{z}_k &= \boldsymbol{J}\boldsymbol{p}_k \end{aligned} \qquad (7),$$

where d is the number of lags as mentioned in dynamic PCA. Once $\boldsymbol{J}$ is determined, the $\boldsymbol{z}_k$ can be calculated by Eq. (6) and the state-space matrices can be estimated via linear squares regression. Different approach to calculate $\boldsymbol{J}$ distinguishes various derivations of subspace algorithms including CVA, N4SID and PLS etc. More details about subspace identification algorithms refer to Ljung (1999) and Overschee(1996).

To determine the order of system in Eq. (6), a number of approaches have been proposed. For instance, N4SID determines the system order by checking the singular values. Akaike information criterion (AIC) is also employed to determine the model order automatically. In this context, however, the purpose of modelling is to remove the auto & cross correlations from the score variables as much as possible, i.e., reduce the ACFs $\rho_{i,j}(\tau)$ of $\boldsymbol{e}_k$ as close to zero as possible when $\tau \neq 0$. To the end, the following Akaike like information criterion is suggested:

$$AICx = \log(V(nz)) + \frac{2q}{N} \tag{8}$$

$$V(nz) = \frac{1}{2Mr^2}\sum_{i=1}^{r}\sum_{j=1}^{r}\sum_{\substack{\tau=-M \\ \tau\neq 0}}^{M}\rho_{i,j}^2(\tau), \tag{9}$$

where q is the number of estimated parameters in Eq. (6), N is the length of modelling data and $M = 3 * nz$ is the maximum time lag. $\frac{2q}{N}$ is included to avoid over-fitting. The system order nz is then selected to minimize the $AICx$ objective function.

Once the state-space model is identified and the residual $\boldsymbol{e}_k$ is generated, we have the following T^2 statistic:

$$T_k^2 = \boldsymbol{e}_k^T \boldsymbol{S}_e^{-1} \boldsymbol{e}_k \sim \frac{r(N^2-1)}{N(N-r)} F_{\alpha,r,N-r}, \tag{10}$$

where $\boldsymbol{S}_e = \frac{1}{N-1}\sum_{i=1}^{N}\boldsymbol{e}_i\boldsymbol{e}_i^T$ is the correlation matrix of $\boldsymbol{e}_k$.

Note that not all scores are necessarily included in the state-model. If a score is independent on itself and other scores, it should be excluded to reduce the complexity of state-space model. In this situation, the T^2 statistic will become:

$$T_k^2 = \begin{bmatrix}\hat{\boldsymbol{e}}_k \\ \tilde{\boldsymbol{t}}_k\end{bmatrix}^T \boldsymbol{S}_{et}^{-1} \begin{bmatrix}\hat{\boldsymbol{e}}_k \\ \tilde{\boldsymbol{t}}_k\end{bmatrix} \sim \frac{r(N^2-1)}{N(N-r)} F_{\alpha,r,N-r}, \tag{11}$$

where $\boldsymbol{S}_{et} = \frac{1}{N-1}\sum_{i=1}^{N}\begin{bmatrix}\hat{\boldsymbol{e}}_i \\ \tilde{\boldsymbol{t}}_i\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{e}}_i \\ \tilde{\boldsymbol{t}}_i\end{bmatrix}^T$ is the correlation matrix of $\hat{\boldsymbol{e}}_k$ and $\tilde{\boldsymbol{t}}_k$. $\hat{\boldsymbol{e}}_k$ is the residual of auto & cross correlated scores $\hat{\boldsymbol{t}}_k$ and $\tilde{\boldsymbol{t}}_k$ is the independent part.

Remark: One could also apply the suggested state-space modelling approach to the original process variables first and then establish a PCA model. However, this might change the linear relationship among the variables and cause the variable reduction by PCA inefficient. Furthermore, if a set of process variables are driven by same dynamic processes, e.g., the reactor may have several temperature sensors on different locations, applying PCA first will result fewer dynamic scores. In section 4.2, by the application in TE process, it is revealed that PCA does not only concentrate the variation information but also the dynamic information in the scores and the computation cost to establish the state-space model is therefore less demanding.

3.3Dynamic process monitoring framework based on SI-PCA

The procedures of off-line and on-line monitoring using SI-PCA are as follows:

Off-line: develop the normal operating condition model (NOC)

1. Collect an operating data set during normal operation $\boldsymbol{X}$.

2. Apply PCA to $\boldsymbol{X}$ and obtain the score variables $\boldsymbol{t}_k = \begin{bmatrix}\hat{\boldsymbol{t}}_k \\ \tilde{\boldsymbol{t}}_k\end{bmatrix}$. The number of components can be determined by cross validation or other criteria. The independence of excluded principal components relative to SPE statistic should also be checked. If dynamics exist, state-space model can also be employed.

3. Subspace identification method in section 3.2 is employed to remove the dynamics of $\hat{\boldsymbol{t}}_k$. The confidence interval of T^2 statistic is determined based on the residual $\hat{\boldsymbol{e}}_k$, $\tilde{\boldsymbol{t}}_k$.

On-line monitoring:

1. For new observation, obtain the score values via $\boldsymbol{t}_k = \boldsymbol{P}^T \boldsymbol{x}_k$.

2. Apply the identified subspace model to calculate the residual $\hat{\boldsymbol{e}}_k$.

3. Determine the T^2 and SPE statistics and compare with the confidence intervals.

4. APPLICATION STUDIES

Since the introduction in 1993 by Downs and Vogel, Tennessee Eastman (TE) process has been widely studied in the literatures (Ku 1995, Ricker 1996, Chiang 2000, Kano 2000, Russel 2000). The TE model includes 5 process units: a reactor, a condenser, a vapour-liquid separator , a recycle compressor and a product stripper. There are 41 measurements and 12 manipulated variables. The process is open-loop unstable and requires regulatory controllers. TE process includes 20 programmed disturbances including composition step change in reactants, random react cooling water inlet temperature random disturbance and valve sticking etc. Detailed description about the operation of the TE process can be found in Downs (1993).

In this paper, Ricker's TE simulator based on Matlab® 6.5 was used to generate the data set. The closed-loop control strategy of Ricker (1996) was also adopted. The reference dataset to construct NOC model includes 2000 samples of 22 continuously measured variables which were recorded at 0.1 h interval.

It is somewhat cumbersome to plot the ACFs of all the 22 variables. In Fig.2, only the auto correlation coefficients of variables are illustrated. It can be seen that ten variables are strongly auto correlated including feed A (1), reactor pressure (7), purge rate (10), separator temperature (11), separator pressure (13), stripper pressure (16), stripper temperature (18), compressor power (20), reactor outlet coolant temperature (21), separator outlet coolant temperature (22).

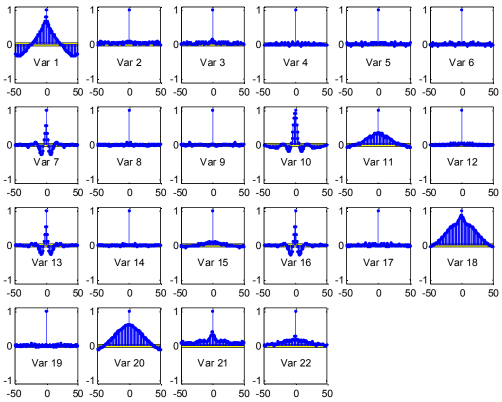

Fig.2 Auto correlation plots of TE process variables

PCA was first applied to the NOC data and 15 scores which explain 87.6% of total variation were retained. The auto correlation coefficients of scores are illustrated in Fig.3. As described in section 3.2, not only the variation information but also the dynamics information are concentrated in the first 5 scores.

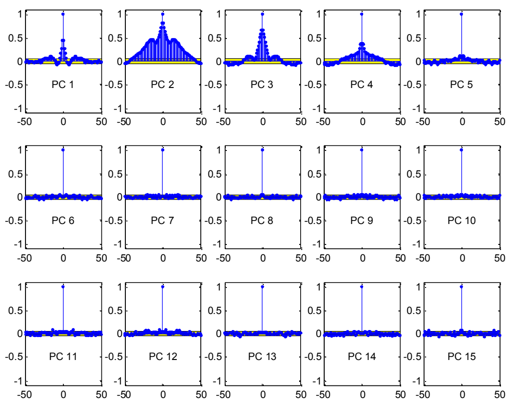

Fig.3 Auto correlation plots of PCA scores for TE process

Based on the criterion presented in section 3.2, a 5-order state-space model was identified to remove the dynamics efficiently. Fig.4 confirms that residuals are no longer auto or cross correlated. In

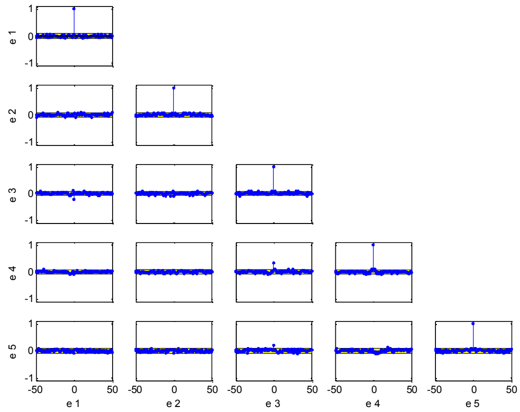

Fig.4 ACFs of residuals by SI-PCA (TE process)

contrast, the ARMA-PCA had to use as high as 20 order filters to remove the auto correlations of the first two scores.

To demonstrate the capability of SI-PCA to detect abnormal behaviour, excessive variation of the reactor cooling water temperature was simulated (disturbance type 11) to generate a fault dataset. The fault dataset also contains 2000 samples and the fault was injected after the 1000 sample.

The T^2 control chart is presented in Fig.5 and the data around the control limit are zoomed in Fig.6 for clear illustration. The last 1000 values of T^2 statistic produce an excessive number of violations. So SI-PCA detects the out-of-control situation correctly.

5. CONCLUSIONS

This paper studies the influences of process dynamics on false alarms rates of statistical process control. It is shown that the presence of cross correlation will also cause false alarms. Applications illustrate that two improved PCA methods, dynamic PCA and ARMA-PCA, do not remove the effects of time-series structures of variables.

Motivated by above facts, a subspace identification approach based PCA (SI-PCA) is proposed to remove the auto & cross correlations of score variables simultaneously. Akaike information like criterion is introduced to determine the state-space model.

Application in TE process fault detection reveals that the SI-PCA is efficient in removing the auto & cross correlation among variables and detecting the process abnormal behaviour.

This paper focus on the behaviour of T^2 statistic and the influence of auto & cross correlations on SPE statistic is also worth further study.

ACKNOWLEDGEMENTS

This project is supported by National High-Tech Program of China under grant No. 2001AA413110.

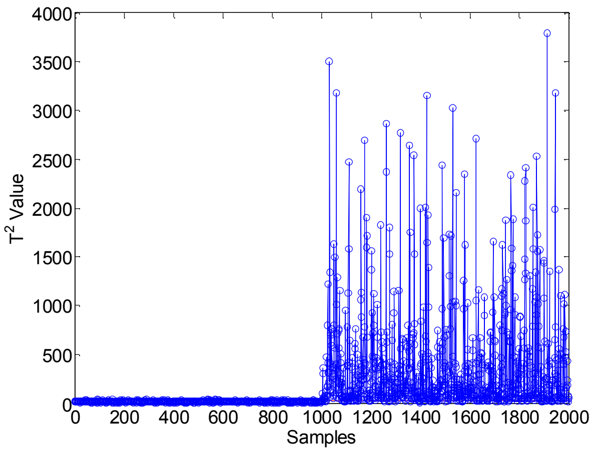

Fig.5 T^2 control chart for reactor faults by SI-PCA

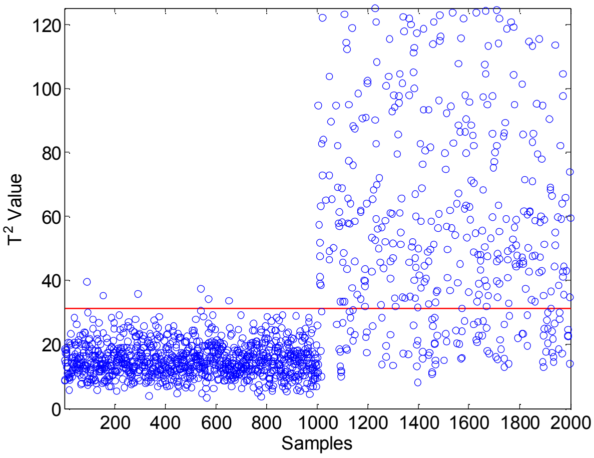

Fig.6 Zooming in the data of Fig.10 around the 99% control limit (Solid line)

REFERENCES

Callao M.P., A. Rius (2003), Time series: a complementary technique to control charts for monitoring analytical systems, *Chemometrics and Intelligent Laboratory Systems*, **Vol.66**, 79-87.

Chen J., K. Liu (2002), On-line batch process monitoring using dynamic PCA and dynamic PLS models, *Chemical Engineering Science*, **Vol.57**, 63-75.

Chiang L.H. (2000), *Fault detection and diagnosis in industrial systems*. Springer-Verlag, Heidelberg.

Cinar A., Cenk U. (1999). Statistical process and controller performance monitoring: A tutorial on current methods and future direction, *Proceedings of American Control Conference*, 2625-2639.

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

George B., M.J. Gwilym, R. Gregory (1994), *Time series analysis: forecasting & control (3rd edition)*, Prentice Hall, Englewood Clifs.

Jackson J.E (1991), *A user's guide to principal components*, Wiley Inter-science, New York.

Kano M., K. Nagao, S. Hasebe, I. Hashimoto, et al (2000), Comparison of statistical process monitoring method: application to the Eastman challenge problem, *Computers and Chemical Engineering*, **Vol.24**, 175-181.

Kruger U., Y. Zhou, W.I. George (2004), Improved principal component monitoring of large-scale processes, *Journal of Process Control*, **Vol.14**, 879-888.

Ku W., R.H. Storer, C. Georgakis (1995), Disturbance rejection and isolation by dynamic principal component analysis, *Chemometrics and Intelligent Laboratory Systems*, **Vol.30,** 179-196.

Larimore, W. E. (1983), System identification, reduced order filtering and modeling via canonical correlation analysis, *Proceedings of the American Control Conference*, 445-451.

Ljung L. (1999), *System identification – theory for the user*, Prentice Hall, Englewood Clifs.

Ljung L. (2002), *System identification toolbox – for use with MATLAB*, Mathwork, Natick.

Negiz A, A. Cinar (1997), Statistical monitoring of multivariate dynamic processes with sate-space models, *AIChE Journal*, **Vol.43**, 2002-2020.

Nomikos P., MacGregor, J. F. (1995), Multivariate SPC charts for monitoring batch processes. *Technometrics*, **Vol.37**, 41-59.

Overschee P.V. (1996), *Subspace identification for linear systems: Theory Implementation Applications*, Kluwer Academic Publishers, Boston/London/Dordrecht.

Ricker N. L.(1996), Decentralized control of the Tennessee Eastman challenge process, , *Journal of Process Control*, **Vol. 6**, 205-221.

Russell E.L., L.H. Chiang, R.D. Braatz (2000), Fault detection in industrial processes using canonical variate analysis and dynamic principal component analysis, *Chemometrics and intelligent laboratory systems*, **Vol.51**, 81-93.

Simoglou A., E.B. Martin, A.J. Morris (2002), Statistical performance monitoring of dynamic multivariate processes using state space modeling, *Computers and Chemical Engineering*, **Vol.26**, 909-920.

Venkatasubramanian V., R. Rengaswamy, S.N. Kavuri, K. Yin (2003), A review of process fault detection and diagnosis Part III: Process history based methods, *Computers and Chemical Engineering*, **Vol.27**, 327-346.

Wise B, N.B. Gallagher (1996), The process chemometrics approach to process monitoring and fault detection, *Journal of Process Control*, **Vol.6**, 329-348.