Two-Dimensional Dynamic PCA for Batch Process Monitoring

Ningyun Lu, Yuan Yao, and Furong Gao
Dept. of Chemical Engineering, Hong Kong University of Science and Technology, Clear water bay, Kowloon, Hong Kong SAR, P. R. China

Fuli Wang
Faculty of Information Science and Engineering, Northeastern University, Shenyang, P. R. China

DOI 10.1002/aic.10568
Published online September 23, 2005 in Wiley InterScience (www.interscience.wiley.com).

Keywords: dynamic principal component analysis, 2-D modeling, batch processes, process monitoring, statistical analysis

Introduction

Batch processes play an important role in many industries for flexible manufacturing of high value-added products. Online monitoring of batch processes is of critical importance in ensuring operation safety and quality consistency. Multivariate statistical methods for continuous processes, such as principal component analysis (PCA) and partial least square (PLS), have been extended to batch process monitoring with some successes, for example, multiway PCA/PLS,[1,2] hierarchical PCA,[3,4] and stage-based sub PCA modeling method.[5]

The above PCA- or PLS-based monitoring methods have been originally developed for concerning with static rather than dynamic relationships among process variables. These methods require implicitly a statistical assumption of batch independence, that is, the variation in the trajectories among the historical reference batches are purely common-caused variation.[1] Batch processes, however, are inherently dynamic processes. Dynamic behaviors may exist not only within a batch run, but also from batch to batch. For example, in injection molding, variations in materials and slowly-changing conditions such as mold temperature can change process characteristics, that is, process dynamics, in both time and batch-to-batch sense. Two batch dynamics, namely within-batch dynamics and batch-to-batch dynamics, are needed to describe such processes. In general, a batch process involves both within-batch and batch-to-batch dynamics at the same time. It is desirable to develop a batch process monitoring scheme that can capture both within-batch and batch-to-batch dynamics simultaneously.

We propose to do so by representing a batch process in a two-dimensional (2-D) space. Process data are double indexed by the sampling time within a batch, and the batch number of successive batch runs. 2-D dynamic modeling techniques that have been well developed for digital data filtering and image processing, can, therefore, be adopted for the modeling and monitoring of batch processes. This article demonstrates, for the first time, the feasibility of such a modeling. A novel batch monitoring scheme, a 2-D dynamic principal component analysis (2-D-DPCA), is proposed to model and monitor simultaneously the time- and batch-wise dynamics.

There exist several methods extending 1-D dynamic modeling methods for capturing batch dynamics: batch dynamic principal component analysis (BDPCA)[6] which concerns only with within-batch dynamics, the method of incorporating multiway PCA/PLS model with prior batch information for capturing batch-to-batch dynamics,[7] and subspace identification using a lifted state-space model.[8] Compared with these methods, the proposed 2-D-DPCA is a truly 2-D model; it can capture both within-batch and batch-to-batch dynamics at the same time using a parsimonious model structure; and online implementation of 2-D-DPCA is much easier. Simulations show that the proposed 2-D-DPCA based batch monitoring can be very effective for detecting even small changes in correlation or process drifts.

2-D dynamic PCA

2-D dynamic modeling has been originally developed for data filtering and image processing. 2-D dynamic model structures can be categorized into two groups: 2-D state-space models by Roesser[9] and Fornasini and Marchesini[10,11] and

Correspondence concerning this article should be addressed to F. Gao at kefgao@ust.hk.

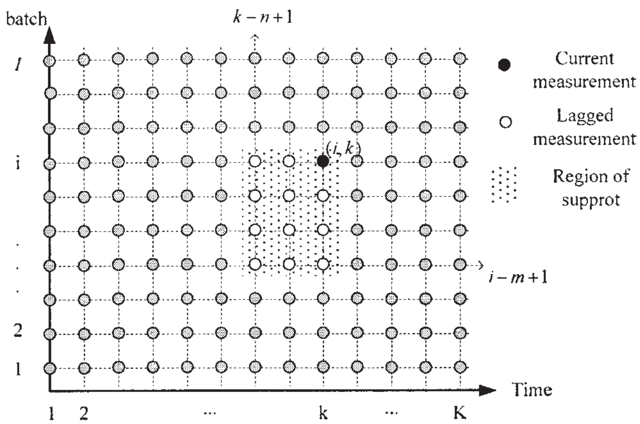

Figure 1. 2-D representation of batch process data.

lagged regression models such as 2-D auto-regressive (AR), moving average (MA), or autoregressive moving average (ARMA) models. 2-D AR model structure is selected in this article for augmenting process data to include lagged measurements in time and batch directions. PCA is then applied on the 2-D augmented data matrix for capturing auto-correlations in the two directions, and cross-correlations of process variables. We name this approach as 2-D dynamic principal component analysis (2-D-DPCA).

Consider process data generated by I number of successive batchs, J variables and K sampling intervals. Let $x_j(i, k)$ be process measurement of variable j at sampling interval k in batch run i ($i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, K$), which can be arranged in a 2-D field with two directions i and j standing for batch and time, respectively, as illustrated in Figure 1.

In the case of the batch process with 2-D dynamics, the current values of the variables $x_j(i, k)$ will depend on not only the past values in time direction, $x_j(i, k-1)$, $x_j(i, k-2) \ldots x_j(i, k-n)$; but the past values in batch direction, $x_j(i-1, k)$, $x_j(i-2, k) \ldots x_j(i-m, k)$, and even in the cross direction, $x_j(i-1, k-1) \ldots x_j(i-m, k-n)$, where m and n are the autoregressive orders in the two directions. We define the region covering the above lagged values as region of support (ROS), as shown in Figure 1. For statistical monitoring of the 2-D dynamic batch process, PCA can be applied to an augmented data matrix defined below to extract simultaneously 2-D auto-correlated and cross-correlated relationships

$$\mathbf{X} = \begin{bmatrix} X_{m,n} \\ \vdots \\ X_{i,k} \\ \vdots \\ X_{I,K} \end{bmatrix} \tag{1}$$

where

$$X_{i,k} = [\mathbf{x}(i, k), \mathbf{x}(i, k-1), \ldots, \mathbf{x}(i, k-n+1), \mathbf{x}(i-1, k), \mathbf{x}(i-1, k-1), \ldots, \mathbf{x}(i-1, k-n+1), \ldots, \mathbf{x}(i-m+1, k), \mathbf{x}(i-m+1, k-1), \ldots, \mathbf{x}(i-m+1, k-n+1)]\mathbf{x}(i, k) = [x_1(i, k), \ldots, x_j(i, k), \ldots, x_J(i, k)]$$

The dimension of the 2-D augmented matrix $\mathbf{X}$ is $(I-m+1)(K-n+1) \times mnJ$. Its columns are composed of the current measurements $\mathbf{x}(i, k)$, and the lagged process measurements in its ROS. PCA on X can be formulated by

$$\mathbf{t}(i, k) = X_{i,k}P \tag{2}$$

where P is the loading matrix with dimensions of $(mnJ \times A)$, A is the number of retained principal components, and $\mathbf{t}(i, k)$ is the vector of principal component scores at sampling interval k of bath run i. The reconstructed process data by PCA and the corresponding residuals are computed, respectively, by

$$\hat{\mathbf{x}}(i, k) = \mathbf{t}(i, k)P_{ik} \tag{3}$$

$$\mathbf{e}(i, k) = \mathbf{x}(i, k) - \hat{\mathbf{x}}(i, k) \tag{4}$$

where P_{ik} is the first J rows of P, that is, the loadings for $\mathbf{x}(i, k)$. The above modeling procedure, Eq. 2–4, is named 2-D-DPCA method, which can be directly used for the following online batch monitoring.

Two important parameters need to be determined in 2-D-DPCA: 2-D autoregressive orders (m, n), and the number of retained principal components A. Implicitly, the 2-D-DPCA may be considered as an equivalent to a multivariate 2-D-AR model; therefore, order estimation methods for 2-D-AR models can be borrowed for the determination of the 2-D autoregressive orders (m, n). Most well-known 1-D model order determination methods, such as Akaike information criterion (AIC), minimum descriptive length (MDL) and minimum eigenvalue (MEV), have been extended for the order estimation for 2-D AR models.[12] In this article, 2-D-AIC is selected, as it is simple and effective even for processes with significant cross-correlations. The order pair consisting of the maximum in the two directions is selected as the 2-D autoregressive orders of 2-D-DPCA. As for the number of principal components to be retained in 2-D-DPCA, cross-validation method[13] can be directly used.

2-D-DPCA based batch process monitoring

In the earlier 2-D-DPCA modeling, with a proper 2-D orders, and a proper number of principal components retained, the transformed principal component scores can extract all linear auto- and cross-correlated relationships in process data. The T^2 statistic, calculated from the principal component scores, are not independent in time- and batch-to-batch sense, which violates the statistical assumption for process monitoring. The residuals obtained by 2-D-DPCA modeling, however, are independent in time series and batch runs. This provides a basis for the statistical monitoring based on the squared prediction error (SPE). A monitoring scheme based on SPE in residual subspace is proposed for 2-D dynamic batch processes. SPE value at sampling interval k of batch run i is calculated by

$$SPE_{i,k} = \mathbf{e}(i, k) \cdot \mathbf{e}(i, k)^T, \quad (i = 1, \ldots, I; k = 1, \ldots, K) \tag{5}$$

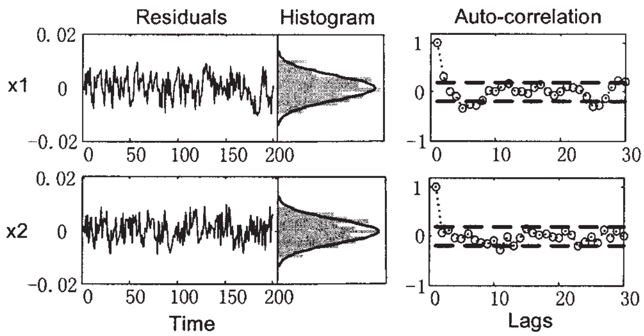

Figure 2. Residuals, histograms, and autocorrelation coefficients produced by the proposed 2-D-DPCA modeling for a normal batch.

Online monitoring based on 2-D-DPCA is conducted in a way similar to the traditional PCA-based methods. First, the control limits of SPE are estimated from the reference data. Adopting the works of Jackson and Mudholkar,[14] the distribution of SPE can be approximately described by a weighted Chi-squared distribution, $g\chi_h^2$, where the weight g, and the freedom degree h can be obtained following the approach of Nomikos and MacGregor.[1] That is, g and h at sampling interval k are estimated from $\mathbf{SPE_k} = \{SPE_{1,k}, SPE_{2,k}, \ldots, SPE_{I,k}\}$, $g_k = v_k/2m_k$ and $h_k = 2(m_k)^2/v_k$, where m_k is the average of $\mathbf{SPE_k}$ and v_k is the corresponding variance. The control limits of SPE with confidence level α at sampling time k can be approximated by

$$SPE_{k,\alpha} = (v_k/2m_k)\chi^2_{2(m_k)^2/v_k,\alpha} \tag{6}$$

Online monitoring can be conducted by judging whether or not the Q-statistic of the coming measurements in an evolving batch is below the control limits. A significant Q-statistic indicates confidently a process abnormality. The contribution plot,[15] a commonly-used and effective diagnosis tool in PCA-based methods, can be directly used to diagnose the abnormalities in the proposed 2-D-DPCA based method.

Illustration

2-D-DPCA modeling

The proposed 2-D-DPCA based process monitoring is tested with a batch process having the following 2-D dynamics

$$\begin{aligned}
x_1(i, k) &= 0.8*x_1(i-1, k) + 0.5*x_1(i, k-1) \\
&\quad - 0.33*x_1(i-1, k-1) + w_1 \\
x_2(i, k) &= 0.44*x_2(i-1, k) + 0.67*x_2(i, k-1) \\
&\quad - 0.11*x_2(i-1, k-1) + w_2 \\
x_3(i, k) &= 0.65*x_1(i, k) + 0.35*x_2(i, k) + w_3 \\
x_4(i, k) &= -1.26*x_1(i, k) + 0.33*x_2(i, k) + w_4
\end{aligned} \tag{7}$$

where x_1 and x_2 are two independent signals; x_3 and x_4 are linear combinations of x_1 and x_2; w_j ($j = 1, 2, 3, 4$) are Gaussian random variables with variance 0.01. For the first batch, the trajectories of x_1 and x_2 are initialized by two 1-D first-order dynamic signals. For the first sample of each batch, the values of x_1 and x_2 are set as $x_1(i, 1) = x_1(i-1, 1) + w_1$ and $x_2(i, 1) = x_2(i-1, 1) + w_2$. There are totally 100 batch runs with 200 samples in each batch, that is, $I = 100$, $J = 4$, and $K = 200$.

The 2-D autoregressive orders of the four variables are all determined as (1, 1) by 2-D AIC method, therefore, the 2-D augmented data vector for principal component analysis consists of $[\mathbf{x}(i, k), \mathbf{x}(i, k-1), \mathbf{x}(i-1, k), \mathbf{x}(i-1, k-1)]$, where $\mathbf{x}(i, k) = [x_1(i, k), x_2(i, k), x_3(i, k), x_4(i, k)]$. Cross-validation algorithm indicates that four of sixteen principal components should be retained in the 2-D-DPCA model, which can extract over 99 percent of variation in the reference data. Figure 2 shows the residuals of an arbitrarily selected batch run, the corresponding histograms with superimposed fitted normal density, and the autocorrelation coefficients of the residuals. It is clear that the residuals after a proper 2-D-DPCA modeling are nearly pure noise, indicating the effectiveness of the method in extracting the dynamic correlations. The statistical assumption of both time and batch independence are satisfied. The values of SPE for normal batch data are then confidently described by a weighted chi-square distribution, providing a reference distribution of normal batch operation against which future batches can be compared for process monitoring.

2-D-DPCA based batch monitoring

To test the effectiveness of the proposed method, two groups of faulty batches are generated to simulate a change in correlation structure and a small process drift, respectively.

The first fault is simulated by changing correlation structure. Variable x_2 is changed to the following from batch 61

$$\begin{aligned}
x_2(i, k) &= 0.67*x_2(i-1, k) + 0.8*x_2(i, k-1) \\
&\quad - 0.47*x_2(i-1, k-1) + w_2^*
\end{aligned} \tag{8}$$

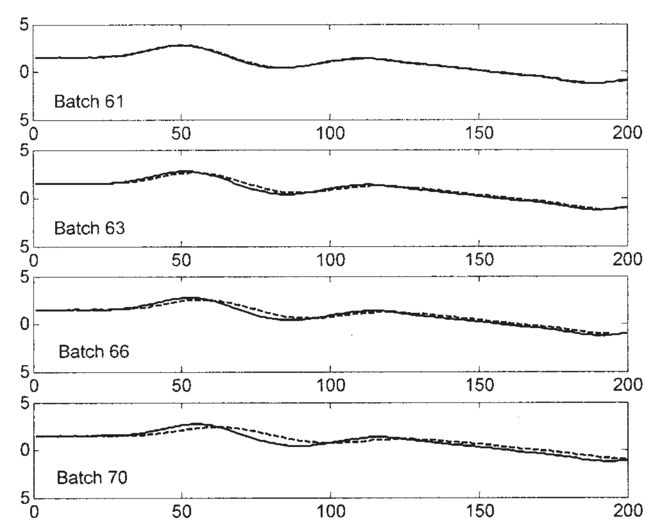

Figure 3. Faulty trajectories of variable x_2 with changed correlation structure.

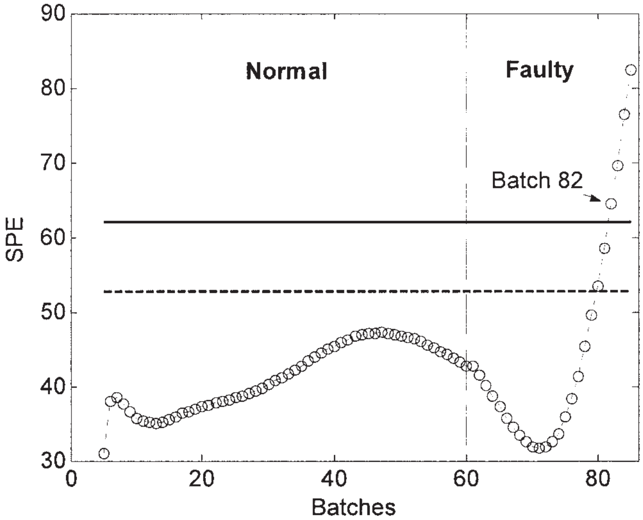

Figure 4. Offline batch-to-batch monitoring of the process for a changed correlation structure from batch 61 by MPCA.

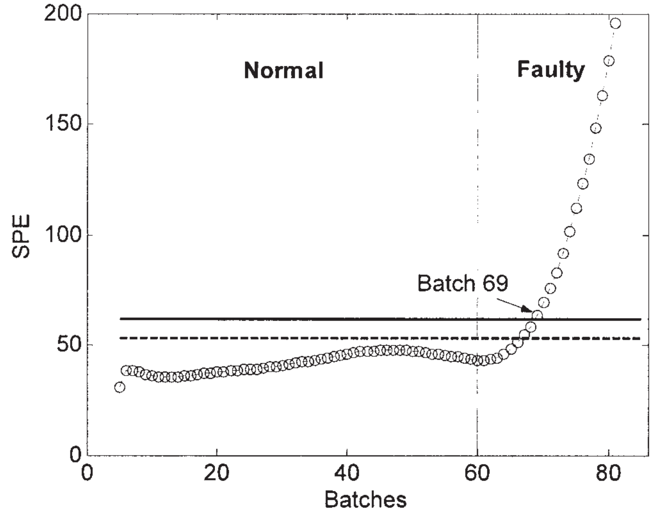

Figure 6. Offline batch-to-batch monitoring of the process for a small process drift from batch 61 by MPCA.

Despite the significant change in the 2-D autoregressive parameters for x_2, the numerical differences between the new and normal x_2, however, only become significant after 5–10 batch runs later, as shown in Figure 3. Prompt detection of this fault is a challenge for MPCA-based batch monitoring methods. Figure 4 shows the offline batch-to-batch monitoring result using the MPCA model incorporated with batch-to-batch information.[7] The fault can be detected only after batch 82 by SPE monitoring chart, 22 batches after the occurrence of the fault. The proposed 2-D-DPCA based method, however, can detect the fault without any delay, as shown in Figure 5. This prompt fault detection is attributed to the inherent advantages of the proposed 2-D-DPCA model, as it can model autocorrelation in time and batch at the same time, and the cross-correlation among process variables can be also effectively modeled. Any change of correlation (auto- or cross-) in process data will make the 2-D-DPCA model fail to explain the changed correlations, the unexplained information will then make Q-statistic significant and excess the control limits.

The second group of faulty batches are generated to simulate a small process drift on variable x_2 from batch 61 by adding a signal that increases slowly with time and batch. Figures 6 and 7 show the monitoring results by MPCA model incorporated with prior batch information and 2-D-DPCA methods, respectively. Similarly, MPCA-based method has a significant delay of 9 batches in fault detection; while the proposed 2-D-DPCA based batch monitoring scheme can detect it immediately in batch 61. From Figure 7, the SPE values are increasing slowly with time and batch, implying that the fault is a small process drift.

The above simulations show that the proposed 2-D-DPCA based batch monitoring is effective and sensitive for the detection of changes in correlation structures or small drift signals.

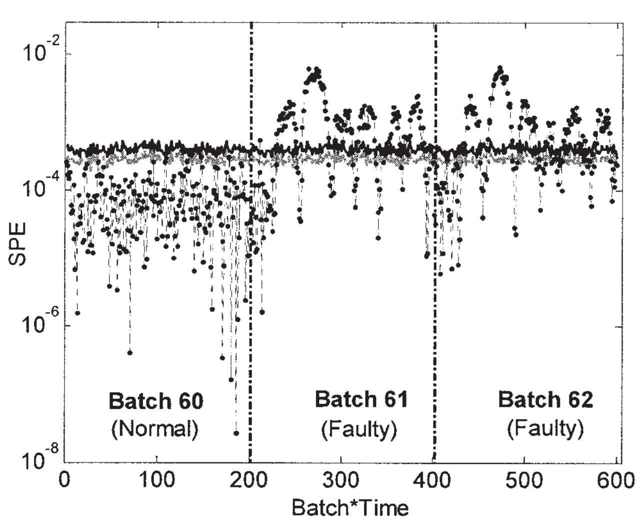

Figure 5. Online monitoring of the process for a changed correlation structure by 2-D-DPCA.

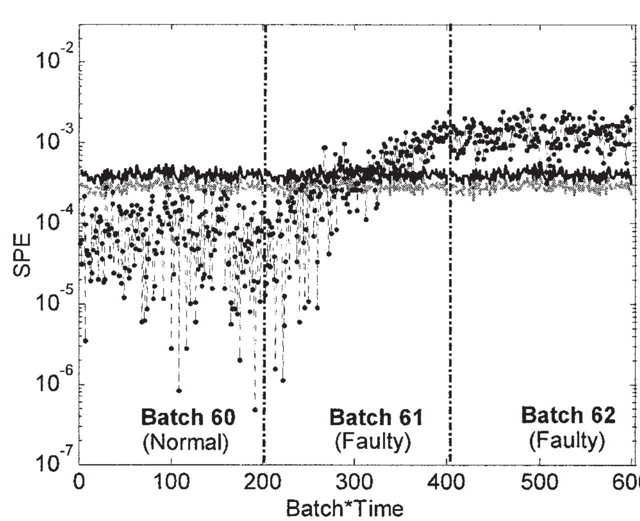

Figure 7. Online monitoring of the process for a small process drift from batch 61 by 2-D-DPCA.

Conclusions

Motivated by developing models to capture batch dynamics both in time and batch sense, a 2-D-DPCA modeling has been proposed for online batch process monitoring. Firstly, process data is augmented to include lagged measurements in both time and batch directions. PCA is then applied on the 2-D augmented data matrix to capture not only cross-correlations of process variables, but also autocorrelations in both time and batch directions. It is shown that the proposed 2-D-DPCA model can extract all correlated relationships in process data; while the residuals can provide an effective statistical basis for online monitoring of 2-D dynamic batch processes.

Acknowledgment

This work is supported in part by a competitive Earmarked Research Grant of Hong Kong Research Grant Council, under project number 601104.

Literature Cited

1. Nomikos P, MacGregor JF. Monitoring of batch processes using multiway principal component analysis. *AIChE J.* 1994;40:1361–1375.
2. Nomikos P, MacGregor JF. Multiway partial least squares in monitoring batch processes. *Chemometrics and Intelligent Laboratory Systems.* 1995;30:97–108.
3. Wold S, Kettaneh N, Tjessem K. Hierarchical multiblock PLS and PC models for easier model interpretation and as an alternative to variable selection. *J of Chemometrics.* 1996;10:463–482.
4. Rännar S, MacGregor JF, Wold S. Adaptive batch monitoring suing hierarchical PCA. *Chemometrics and Intelligent Laboratory Systems.* 1998;41:73–81.
5. Lu N, Gao F, Wang F. Sub PCA modeling and monitoring strategy for batch processes. *AIChE J.* 2004;50:255–259.
6. Chen J, Liu KC. On-line batch process monitoring using dynamic PCA and dynamic PLS models. *Chem Eng Sci* 2002;57:63–75.
7. Flores-Cerrillo J, MacGregor JF. Multivariate monitoring of batch processes using batch-to-batch information. *AIChE J.* 2004;50:1219–1228.
8. Lee, JH, Dorsey AW. Monitoring of batch processes through state-space models. *AIChE J.* 2004;50:1198–1210.
9. Roesser RP. A discrete state-space model for linear image processing. *IEEE Trans on Automatic Control.* 1975;AC-20:1–10.
10. Fornasini E, Marchesini G. State-space realization theory of two-dimensional filters. *IEEE Trans on Automatic Control.* 1976;AC-21: 484–492.
11. Fornasini E, Marchesini G. Doubly-indexed dynamic systems: state-space models and structural properties. *Math Syst Theory.* 1978;12: 59–72.
12. Aksasse B, Radouane L. Two-dimensional autoregressive (2D-AR) model order estimation. *IEEE Trans on Signal Processing.* 1999;47: 2072–2077.
13. Wold S. Cross-validation estimation of the number of components in factor and principal component models. *Technometrics.* 1978;20:397–405.
14. Jackson JE, Mudholkar GS. Control procedures fro residuals associated with principal component analysis. *Technometrics.* 1979;21:341–349.
15. Miller P, Swanson RE, Heckler CE. Contribution plots: a missing link in multivariate quality control. *Appl Math & Comp Sci.* 1998;8:775–792.

Manuscript received Feb. 22, 2005, and revision received Apr. 14, 2005.