Preprints of the 18th IFAC World Congress
Milano (Italy) August 28 - September 2, 2011

Dynamic latent variable modeling for statistical process monitoring

Gang Li [*], **Baosheng Liu** [**], **S. Joe Qin** [***], **Donghua Zhou** [*]

[*] *Department of Automation, Tsinghua University, Beijing 100084, P. R. China. Email: dlg00@mails.tsinghua.edu.cn., zdh@tsinghua.edu.cn*
[**] *Equipment Academy of Airforce, Beijing 100085, P. R. China. Email: liubaosheng1981@tsinghua.org.cn*
[***] *The Mork Family Department of Chemical Engineering and Materials Science, Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089, USA. Email: sqin@usc.edu*

Abstract: Dynamic principal component analysis (DPCA) has been widely used in the monitoring of dynamic multivariate processes. In traditional DPCA, the dynamic relationship between process variables are implicit and hard to interpret. To extract explicit latent factors that are dynamically correlated, a new dynamic latent variable model is proposed. The new structure can improve modeling of dynamic data and enhance the process monitoring performance. Fault detection indices are developed based on the proposed model. A case study is given to illustrate the effectiveness of the proposed new dynamic factor model.

Keywords: dynamic latent variable model, dynamic principal component analysis, subspace method, process monitoring

1. INTRODUCTION

Principal component analysis (PCA) and partial least squares (PLS) are the most important techniques in the area of statistical process monitoring (Kresta et al., 1991; Qin, 2003; Bersimis et al., 2007; Macgregor and Kourti, 1995). Static PCA models are suitable for discrete manufacturing processes where the measured variables are assumed independent and normally distributed. However, these assumptions do not hold any more for measurements form most continuous processes due to their dynamic and nonlinear nature. In these dynamic processes, variables are often driven by random uncontrollable disturbances. Therefore, they show the evident property of autocorrelation.

In order to deal with auto correlated measurements, time series models are used to make a prediction and the prediction errors are used for monitoring. However, this approach is often applied to the univariate process monitoring (Negiz et al., 1994). For multivariate time series with high cross correlated variables, dynamic factor models are preferred for process monitoring and prediction.

In the past twenty years, dynamic multivariate projections were developed for these cases. Ku et al. proposed a lagged versions of PCA to process multivariate variables with dynamic property (Ku et al., 1995), called dynamic PCA model. Dynamic PCA model is based on conducting singular value decomposition on an augmented data matrix, including the time lagged process variables. The use of time lagged variables is seen in many other models, such as dynamic nonlinear PLS (Qin and McAvoy, 1996) and two-dimensional dynamic PCA(Lu et al., 2005). However, there are limitations for this technique. One issue is that cross-correlation is not explicitly extracted, the other issue is that the number of dynamic factors which have nonzero singular values are not the minimum dimension of a linear dynamic system.

To solve these problems, some methods based on subspace modeling were proposed. Negiz and Cinar used a canonical variate (CV) state space model to describe dynamic processes, which is equivalent to a vector autoregressive moving-average time-series model (VARMA) (Negiz and Çinar, 1997, 1998). They used a stochastic realization based on canonical variate analysis to handle a large number of variables that are autocorrelated, cross-correlated and collinear, and constructed a T^2 statistic from the CV state variables for process monitoring. The CV state variables are linear combinations of the past measurements which can explain future variability the most. As PLS

[⋆] This work was supported by National 973 project under grants 2010CB731800 & 2009CB320602, the NSFC under grants 60721003, 60736026 & 60931160440, and the Texas-Wisconsin- California control consortium.

is able to model the relationship between two data sets, the subspace model identified by PLS algorithm can also be used for statistical performance monitoring (Simoglou et al., 2002). The comparison between CVA and PLS shown that CVA can provide more rapid detection for the faults. However, due to the sensitivity of small eigenvalues of the covariance matrix, PLS can obtain more stable results than the CVA method.

Subspace identification method (SIM) is a different system modeling scheme from PCA, which identifies a state space model for the process Qin (2006). PCA can be used to develop a consistent model estimates under the errors-in-variables situation (Wang and Qin, 2002). Li and Qin investigated the relationship between dynamic PCA and SIM under state space description of dynamic processes (Li and Qin, 2001). With the presence of process and measurement noise, they proposed a consistent dynamic PCA algorithm, namely indirect dynamic PCA (IDPCA) and established consistency conditions. Recently, Ding et al. combined SIM and model based fault detection technologies to propose another fault detection scheme (Ding et al., 2009). Using the linear dynamic state space description of processes, the residual generator of the parity space method after identifying a subspace model is equivalent to indirect dynamic PCA modeling for normal data.

However, when a state space model is taken for the monitoring, only dynamic relationships are focused for monitoring. In static PCA models, only static correlation between variables at zero lags are extracted. It is desirable to consider both dynamic auto-correlation and static cross-correlations for monitoring. In the area of time series prediction, dynamic factor analysis (DFA) has been proposed to restrict the dynamic variability in a reduced subspace. Motivated by DFA, a new dynamic statistical model is proposed in this paper, called dynamic latent variable (DLV) model.

The rest of the paper is organized as follows. In Section 2, existing dynamic PCA models and subspace based methods are reviewed briefly. Then, we propose a new dynamic latent variable model in Section 3. Following that, fault detection schemes are developed based on the proposed model in Section 4. In Section 5, we use a case study to illustrate the effectiveness of fault detection and compare it with two existing methods.

2. DYNAMIC PCA AND STATE SPACE MODELING

2.1 Direct modeling

Suppose a sample vector $\mathbf{x}_k \in \mathbb{R}^m$ consists of m sensor measurements at sampling time k. With the effect of dynamic processes and closed loop control, different samples $\mathbf{x}_k$ are not independent at different k, which indicates that they can be auto-correlated and cross-correlated. To capture the dynamic relations inside the variables, dynamic PCA performs PCA on the following augmented data matrix (Ku et al., 1995)

$$\bar{\mathbf{X}}_k = \begin{bmatrix} \mathbf{x}_{k+s} & \mathbf{x}_{k+s+1} & \cdots & \mathbf{x}_{k+n+s} \\ \mathbf{x}_{k+s-1} & \mathbf{x}_{k+s} & \cdots & \mathbf{x}_{k+n-1+s} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_k & \mathbf{x}_{k+1} & \cdots & \mathbf{x}_{k+n} \end{bmatrix} \tag{1}$$

where s is the lagged number of the data matrix, and $n + s + 1$ is the number of samples used in the modeling. If measurement noise terms have identical variance, then we can perform the following singular value decomposition, and select the largest A singular directions as the principal components.

$$\frac{1}{n}\bar{\mathbf{X}}_0\bar{\mathbf{X}}_0^T = \mathbf{UDV}^T \tag{2}$$

Ku et al. suggested a method to determine the order of dynamic process (Ku et al., 1995). Let $\mathbf{P}_D = \mathbf{U}(:,1 : A)$, $\tilde{\mathbf{P}}_D = \mathbf{U}(:, A + 1 : m(s + 1))$. Then, the principal components are

$$t_k = \mathbf{P}_D^T\bar{\mathbf{x}}_k \tag{3}$$

where $\bar{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \vdots \\ \mathbf{x}_{k-s} \end{bmatrix}$.

2.2 Indirect DPCA modeling

If the measurement noise does not have identical variance, Li and Qin suggested to use an indirect dynamic PCA algorithm based on SIM to eliminate the effect of noise (Li and Qin, 2001). The direct modeling of DPCA is referred as DDPCA in (Li and Qin, 2001). In their work, a SIM based indirect dynamic PCA is used for extracting the party vector $\tilde{\mathbf{P}}_I$. Their method performs SVD as follows:

$$\frac{1}{n}\bar{\mathbf{X}}_{s+1}\bar{\mathbf{X}}_0^T = \mathbf{U_2D_2V_2}^T \tag{4}$$

and $\mathbf{P}_I = \mathbf{U}_2(:,1 : A)$, $\tilde{\mathbf{P}}_I = \mathbf{U}_2(:, A + 1 : m(s + 1))$. The principal components can be represented similar to (3). Li and Qin used the Akaike information criterion for choosing the proper lagged number s (Li and Qin, 2001). The number of dynamic principal factors A is determined as the number of diagonal elements of $\mathbf{D}_2$ that are zero or nearest to zero.

2.3 State space modeling

Besides the above indirect dynamic PCA modeling, some other techniques are used to build a dynamic relationship between a future data set $\bar{\mathbf{X}}_{s+1}$ and a past data set $\bar{\mathbf{X}}_0$. Negiz and Çinar used the CVA technique to build a state space model for dynamic process monitoring (Negiz and Çinar, 1997, 1998). It provided the principal directions of variability of a linear dynamic system through the canonical variables, which are orthogonal vectors of the space of the past measurements that are highly correlated to the space of the future measurements. The objective of CVA is to search for correlation in the future and past data set, which has the smallest angle between them:

$$\begin{array}{ll}\max\limits_{\mathbf{w},\mathbf{c}} & \mathbf{c}^T\bar{\mathbf{X}}_{s+1}^T\bar{\mathbf{X}}_0\mathbf{w} \\ s.t. & \|\bar{\mathbf{X}}_{s+1}\mathbf{c}\| = \|\bar{\mathbf{X}}_0\mathbf{w}\| = 1\end{array} \tag{5}$$

The CVA is to perform the following SVD,

$$(\bar{\mathbf{X}}_{s+1}^T\bar{\mathbf{X}}_{s+1})^{-1/2}\bar{\mathbf{X}}_{s+1}^T\bar{\mathbf{X}}_0(\bar{\mathbf{X}}_0^T\bar{\mathbf{X}}_0)^{-1/2} = \mathbf{U}_3\mathbf{D}_3\mathbf{V}_3^T \tag{6}$$

and the CV state (or dynamic factors) is obtained by

$$\mathbf{t}_k = \mathbf{D}_3^{1/2}\mathbf{V}_3^T\mathbf{R}_0^{-1/2}\bar{\mathbf{x}}_k \tag{7}$$

where the SVD matrices only include singular values and vectors corresponding to A canonical variables retained in the model. $\mathbf{R}_0 = \frac{1}{n}\bar{\mathbf{X}}_0^T\bar{\mathbf{X}}_0$ means the covariance matrix of $\bar{\mathbf{x}}$ at zero lag.

Partial least squares (PLS) is conceptually similar to CVA in building a relationship between future and past data set (Simoglou et al., 2002). PLS algorithm searches the following objective:

$$\begin{array}{ll}\max\limits_{\mathbf{w},\mathbf{c}} & \mathbf{c}^T\bar{\mathbf{X}}_{s+1}^T\bar{\mathbf{X}}_0\mathbf{w} \\ s.t. & \|\mathbf{c}\| = \|\mathbf{w}\| = 1\end{array} \tag{8}$$

and the PLS score is obtained by

$$\mathbf{t}_k = (\mathbf{W}^T\mathbf{P})^{-1}\mathbf{W}^T\bar{\mathbf{x}}_k = \mathbf{R}^T\bar{\mathbf{x}}_k \tag{9}$$

where $\mathbf{W}, \mathbf{P}$ and $\mathbf{R}$ are the weighting and loading matrices from the PLS algorithm(Hóskuldsson, 1988).

3. DYNAMIC LATENT VARIABLE MODELING

The aforementioned dynamic latent variable models represent the dynamic linear relations among variables, and use them for residual generation. The idea guarantees the modeling residual is time-independent and suitable for statistical monitoring. However, there are still two problems. On one hand, there exist static and dynamic correlations, which should be monitored, respectively. These existing dynamic latent variable models do not distinguish dynamic relations and static relations from each other, which reduces the detection sensitivity to some tiny faults. On the other hand, they do not extract the dynamic factors according to their auto-covariance. In order to describe the dynamics as well as their static cross-correlations in process variables, an auto-regressive PCA algorithm is proposed to extract dynamic principal components from the original process data .

3.1 Auto-correlation PCA algorithm

We use an auto-regressive PCA algorithm to find the dynamic principal component that has most auto-covariance. Denote $\mathbf{X}_k = [\mathbf{x}_k, \mathbf{x}_{k+1}, ..., \mathbf{x}_{k+n}]^T$. The proposed algorithm maximizes the following objective:

$$\begin{array}{ll}\max\limits_{w} & \mathbf{w}^T\mathbf{X}_{s+1}^T\mathbf{X}_0\mathbf{w} \\ \mathbf{s.t.} & \mathbf{w}^T\mathbf{w} = 1\end{array} \tag{10}$$

Remark 1. This objective searches the direction of the largest auto-covariance with $s+1$ delay. If there is no prior knowledge, the largest auto-covariance usually reaches with one step, which indicates s=0. It is assumed that auto-correlation vanishes as delay time increases. Thus, the auto-covariance with one step delay is a convenient choice. Since $\mathbf{x}_k$ is a stationary time series from a normal process, this is a reasonable assumption.

Using the Lagrangian multiplier, we derive the following objective:

$$J = \mathbf{w}^T\mathbf{X}_{s+1}^T\mathbf{X}_0\mathbf{w} + \lambda(1 - \mathbf{w}^T\mathbf{w}) \tag{11}$$

Taking derivatives with respective to $\mathbf{w}$,

$$\frac{\partial J}{\partial \mathbf{w}} = (\mathbf{X}_{s+1}^T\mathbf{X}_0 + \mathbf{X}_0^T\mathbf{X}_{s+1})\mathbf{w} - 2\lambda\mathbf{w} = \mathbf{0} \tag{12}$$

As $\mathbf{w}^T\mathbf{w} = 1$, $\mathbf{w}$ is the solution of the following eigenvector problem:

$$\frac{1}{2}(\mathbf{X}_{s+1}^T\mathbf{X}_0 + \mathbf{X}_0^T\mathbf{X}_{s+1})\mathbf{w} = \lambda\mathbf{w} \tag{13}$$

Noting that $\mathbf{w}^T\mathbf{X}_{s+1}^T\mathbf{X}_0\mathbf{w} = \mathbf{w}^T\mathbf{X}_0^T\mathbf{X}_{s+1}\mathbf{w}$, the optimal objective can be obtained as

$$J = \frac{1}{2}\mathbf{w}^T(\mathbf{X}_{s+1}^T\mathbf{X}_0 + \mathbf{X}_0^T\mathbf{X}_{s+1})\mathbf{w} = \mathbf{w}^T\lambda\mathbf{w} = \lambda \tag{14}$$

The eigenvector in (13) with the largest eigenvalue λ_{max} is the optimal solution of objective (10).

Remark 2. Note that the covariance in (10) may be negative. If so, we should minimize $\mathbf{w}^T\mathbf{X}_{s+1}^T\mathbf{X}_0\mathbf{w}$ instead. The final result is to search the eigenvectors of (14) with respect to the largest absolute eigenvalues $|\lambda|_{max}$.

Then, the dynamic score vector is further calculated as

$$\mathbf{t} = \mathbf{X}_0\mathbf{w} \tag{15}$$

After that, the loading vector $\mathbf{p}$ can be obtained by

$$\mathbf{p} = \mathbf{X}_0^T\mathbf{t}/\mathbf{t}^T\mathbf{t} \tag{16}$$

which generates the residual uncorrelated to dynamic factors,

$$\mathbf{E} = \mathbf{X}_0 - \mathbf{t}\mathbf{p}^T \tag{17}$$

Different from static PCA, the dynamic principal components are extracted according to the degree of auto-covariance. The residuals should be left with very little auto-covariance, which corresponds to near zero λ. The whole procedure can be summarized as follows:

Algorithm 1 (Auto-correlation PCA)

(1) Let $i = 1$, $\mathbf{X}_0^{(i)} = \mathbf{X}_0, \mathbf{X}_{s+1}^{(i)} = \mathbf{X}_{s+1}$
(2) Solve the eigenvalue decomposition of $(\mathbf{X}_0^{(i)T}\mathbf{X}_{s+1}^{(i)} + \mathbf{X}_{s+1}^{(i)T}\mathbf{X}_0^{(i)})$, then obtain the eigenvector with the largest absolute eigenvalue, denote it as $\mathbf{w}_i$.
(3) $\mathbf{t}_i = \mathbf{X}_0^{(i)}\mathbf{w}_i$.
(4) $\mathbf{p}_i = \mathbf{X}_0^{(i)T}\mathbf{t}_i/\mathbf{t}_i^T\mathbf{t}_i$.
(5) Set $\mathbf{X}_0^{(i+1)} = \mathbf{X}_0^{(i)}(\mathbf{I} - \mathbf{w}_i\mathbf{p}_i^T)$, $\mathbf{X}_{s+1}^{(i+1)} = \mathbf{X}_{s+1}^{(i)}(\mathbf{I} - \mathbf{w}_i\mathbf{p}_i^T)$, $i = i + 1$. Return to step 2, until $i > A$.

The proposed auto-correlation PCA resembles the traditional partial least squares algorithm in the procedure. Thus, we can describe the structure on $\mathbf{X}$ space which is provided by the auto-correlation PCA algorithm similar to PLS. Denote $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_A], \mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_A], \mathbf{T} = [\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_A]$. $\mathbf{R} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$. We have

$$\mathbf{T} = \mathbf{X}_0\mathbf{R} \tag{18}$$

Given a sample of process vector $\mathbf{x}_k$, the score predictor and residual are calculated as

$$\begin{cases} \mathbf{t}_k = \mathbf{R}^T\mathbf{x}_k \\ \hat{\mathbf{x}}_k = \mathbf{P}\mathbf{t}_k = \mathbf{P}\mathbf{R}^T\mathbf{x}_k \\ \mathbf{e}_k = (\mathbf{I}-\mathbf{P}\mathbf{R}^T)\mathbf{x}_k \end{cases} \tag{19}$$

With Algorithm 1, it is guaranteed that $\mathbf{e}_k$ has minimum or no auto-covariance, which is desirable for use in process monitoring.

3.2 Dynamic modeling of the latent scores

As the latent variables contain most dynamics in the data, it is certainly auto-correlated. Therefore, it is necessary to build a dynamic model to reflect the auto-correlation inside the latent variables. As normal processes are often stationary which is operating around a set point, it is possible to describe $\mathbf{t}_k$ with a vector autoregressive (VAR) model as follows:

$$\mathbf{t}_k = \sum_{j=1}^{p} \boldsymbol{\alpha}_j \mathbf{t}_{k-j} + \mathbf{v}_k = \boldsymbol{\Theta}^T\boldsymbol{\varphi}_k + \mathbf{v}_k \tag{20}$$

where $\mathbf{v}_k$ is assumed to be an *i.i.d.* random process with zero mean and constant variance, representing the independent driving source of the normal variation, and α_i is the parameter matrix corresponding to score $\mathbf{t}_{k-i}$. Let $\boldsymbol{\Theta} = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_p]^T$, $\boldsymbol{\varphi}_k = [\mathbf{t}_{k-1}^T, \ldots, \mathbf{t}_{k-p}^T]^T$. p is the model order, which is determined by AIC (De Waele and Broersen, 2003).

The parameters of above model can be estimated by a multivariable least squares algorithm directly as follows (Ljung, 1999).

$$\hat{\Theta} = \left(\sum_{i=p+1}^{p+n} \boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^T\right)^{-1} \left(\sum_{i=p+1}^{p+n} \boldsymbol{\varphi}_i\mathbf{t}_i^T\right) \tag{21}$$

3.3 Dynamic latent variable modeling procedure

Denoting $\mathbf{E} = [\mathbf{e}_1, ..., \mathbf{e}_n]^T$, then $\mathbf{E}$ represents the static variation in the process data. As $\mathbf{E}$ could still contain large non-auto-correlated variability, it is necessary to decompose $\mathbf{E}$ further by static PCA. The whole procedure can be summarized as follows:

(1) Initialization. Center the variables to zero mean and scale them to unit variance.
(2) Use the auto-correlation PCA algorithm (**Algorithm 1**) to build an outer model of X and extract dynamic latent variables, represented by (19). Determine the numbers of components with a criterion.
(3) Use the multivariable LS algorithm to build the inner model of dynamic latent variables. Determine the order with AIC.
(4) Perform PCA on residual $\mathbf{E}$, $\mathbf{E} = \mathbf{T}_s\mathbf{P}_s + \mathbf{E}_r$, where $\mathbf{T}_s$ contain A_s components and A_s is determined by a PCA based method(Valle et al., 1999).

The ultimate space decomposition of $\mathbf{X}$ space can be represented as:

$$\mathbf{X}_0 = \mathbf{T}\mathbf{P}^T + \mathbf{T}_s\mathbf{P}_s + \mathbf{E}_r \tag{22}$$

And for a sample vector $\mathbf{x}_k$, the decomposition can be described as

$$\begin{cases} \mathbf{x}_k = \mathbf{P}\mathbf{t}_k + \mathbf{P}_s\mathbf{t}_{s,k} + \mathbf{e}_{r,k} \\ \mathbf{t}_k = \sum_{j=1}^{p} \boldsymbol{\alpha}_j \mathbf{t}_{k-j} + \mathbf{v}_k \end{cases} \tag{23}$$

We refer to (23) as the dynamic latent variable (DLV) model. DLV models focus dynamic variation of process in a low-dimensional latent space, which is beneficial for process monitoring. Although it omits some static variations, the DLV model captures the most of the dynamic variation first, then extracts large static variations. On the other hand, the DLV model represents dynamic and static relationships in the process variables explicitly, rather than implicitly in traditional DPCA.

4. FAULT DETECTION BASED ON DLV

Conventional PCA based fault detection uses squared prediction error (SPE) and the Hotelling T^2 control charts for monitoring processes. For the DLV model based process monitoring, scores and residual are calculated first as follows:

$$\begin{aligned} \mathbf{t}_k &= \mathbf{R}^T\mathbf{x}_k \\ \mathbf{t}_{s,k} &= \mathbf{P}_s^T(\mathbf{I}-\mathbf{P}\mathbf{R}^T)\mathbf{x}_k \end{aligned} \tag{24}$$

and

$$\begin{aligned} \mathbf{v}_k &= \mathbf{t}_k - \sum_{j=1}^{p} \boldsymbol{\alpha}_j \mathbf{t}_{k-j} \\ \mathbf{e}_{r,k} &= (\mathbf{I}-\mathbf{P}_s\mathbf{P}_s^T)(\mathbf{I}-\mathbf{P}\mathbf{R}^T)\mathbf{x}_k \end{aligned} \tag{25}$$

where $\mathbf{t}_k$ and $\mathbf{v}_k$ represent the score and residual in dynamic part of the process variation, and $\mathbf{t}_{s,k}$ and $\tilde{\mathbf{x}}_{r,k}$ represent the score and residual in static part of the process variation. However, $\mathbf{t}_k$ is auto regressive in the time domain which is not suitable for monitoring by a direct control limit. If one chooses to monitor $\mathbf{t}_k$ directly, it would leads to high missed alarm rates. Therefore, we choose to monitor $\mathbf{v}_k$ instead of $\mathbf{t}_k$ in the dynamic latent space. Therefore, we construct three indices to monitor the process based on the DLV model, which is listed in Table 1.

Table 1. Fault detection indices

Statistics	Calculation	Control limit
T_d^2	$\mathbf{v}^T\Lambda_d^{-1}\mathbf{v}$	$\frac{A(n^2-1)}{n(n-A)}F_{A,n-A,\alpha}$
T_s^2	$\mathbf{t}_s^T\Lambda_s^{-1}\mathbf{t}_s$	$\frac{A_s(n^2-1)}{n(n-A_s)}F_{A_s,n-A_s,\alpha}$
Q_r	$\|\mathbf{e}_r\|^2$	$g\chi^2_{h,\alpha}$

n: number of training samples, A: number of dynamic principal components; A_s: number of static principal components; $\Lambda_d = \frac{1}{n-1}\mathbf{V}^T\mathbf{V}$; $\Lambda_s = \frac{1}{n-1}\mathbf{T}_s^T\mathbf{T}_s$; $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_n]^T$; For Q_r, $g = S/2\mu$, $h = 2\mu^2/S$, μ is the sample mean of Q_r, and S is the sample variance of Q_r

There are explicit meanings of these detection indices for statistical process monitoring. T_d^2 represents the statistically independent information in dynamic principal latent variables, which contains the dynamic variation in the data, while T_s^2 reflects the static principal variation, which contains the static variation in the data. Subsequently, Q_r measures the variation of unmodelled subspace that is unexcited normally. In this framework, a DLV model can be seen as a natural extension of static the PCA model.

5. CASE STUDY ON CSTR

The continuous stirred tank reactor (CSTR) is a dynamic process, which is widely used to evaluate dynamic control and monitoring methods. The CSTR process can be described by the following group of differential equations (Yoon and MacGregor, 2001):

$$\begin{aligned} \frac{\mathrm{d}C_A}{\mathrm{d}t} &= \frac{q}{V}(C_{Af} - C_A) - k_0 \exp(-\frac{E}{RT})C_A + v_1 \\ \frac{\mathrm{d}T}{\mathrm{d}t} &= \frac{q}{V}(T_f - T) + \frac{-\Delta H}{\rho C_p} k_0 \exp(-\frac{E}{RT})C_A \\ &\quad + \frac{UA}{V\rho C_p}(T_c - T) + v_2 \end{aligned} \tag{26a}$$

where C_A is outlet concentration, T is reaction temperature, T_c is temperature of cooling water, q is input fluent velocity of reactant, C_{Af} is input reactant concentration, T_f is input reactant temperature, and v_1, v_2 are independent system noise process, where $v_i(k) \sim N(0, \sigma_{vi}^2)$. Others are constant parameters for the processes. In the simulation, C_A, T are the controlled output variables, and T_c, q, T_f are input variables. Normally distributed measurement noise is added to all these input and output variables. The negative feedback inputs are added to $[q, T_c]^T$ as in (Li et al., 2010).

After mean-centering all variables and scaling them into unit variance, we build a DLV model as well as DDPCA and IDPCA models with 1000 normal samples. Based on the auto-correlation test, two dynamic factors are kept for the DLV model, while 7 and 8 state variables are needed for DDPCA and IDPCA models, respectively. Further, the order of VAR model is determined as 2 according to AIC. Two kinds of faults are introduced to this simulation at 300 samples. Fault 1 is a decrease of two percents in input reactor concentration C_{Af}, which then affects all variables via closed loop control. Fault 2 considers a slow scaling process in the reactor container, which makes the volume of the reactor smaller from a given time.

For Fault 1, the detection results using different schemes are shown in Figs. 1-2. In this fault case, it is observed that DLV based statistics can detect faults in T_s^2 and Q_r, while only DDPCA based on Q manages to detect the fault. For Fault 2, the detection results are depicted in Figs. 3-4, which shows that the DLV based detection is quicker than DDPCA and IDPCA based detection. This is because DPCA models can not separate static and dynamic relations, which lowers the detection sensitivity. Further, the DLV model includes very few dynamic factors, which is suitable for the interpretation of the abnormal situation.

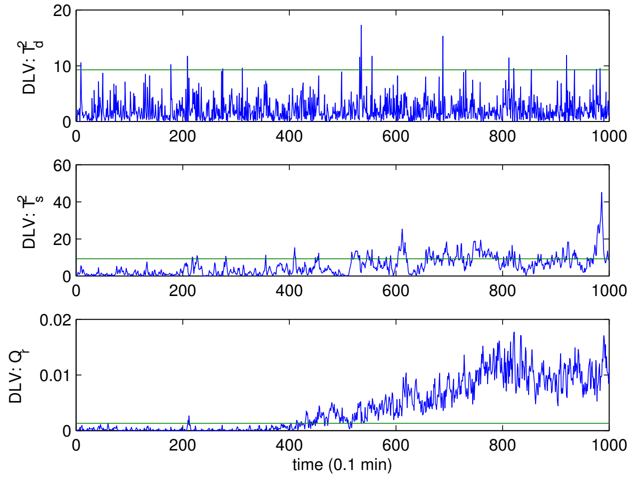

Fig. 1. Fault detection indices of the DLV model for Fault 1

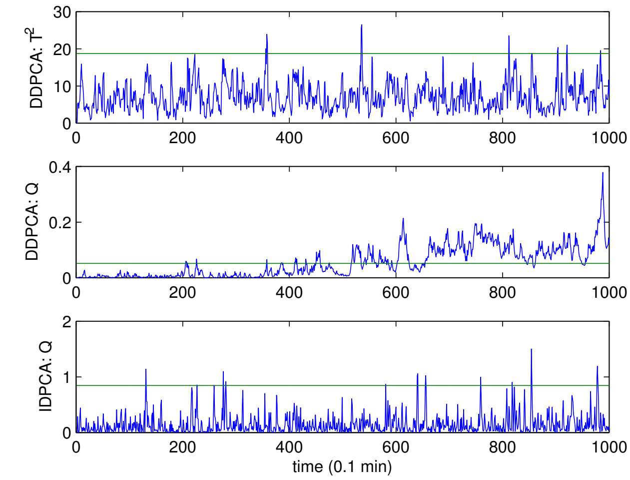

Fig. 2. Fault detection indices of DDPCA and IDPCA models for Fault 1

6. CONCLUSIONS

In this paper, a new dynamic latent variable model is proposed for the dynamic process monitoring. The DLV model extracts the dynamics in the variable space into auto-correlated latent factors. An auto-correlation PCA algorithm is proposed to extract dynamic principal factors according to their auto-covariance. This step separates the dynamic part and static part of the process variations. For the dynamic principal components, a vector AR model is adopted to extract an innovation-like residual for monitoring. In addition, PCA decomposition is applied to the remaining static variation. T^2 statistics are used to monitor the normally excited process variations, while Q_r is used for detecting when normal relations are broken. The DLV model has a clear interpretation in terms of subspace, which can be used for further diagnosis. A case study on CSTR shows the effectiveness of the proposed methods.

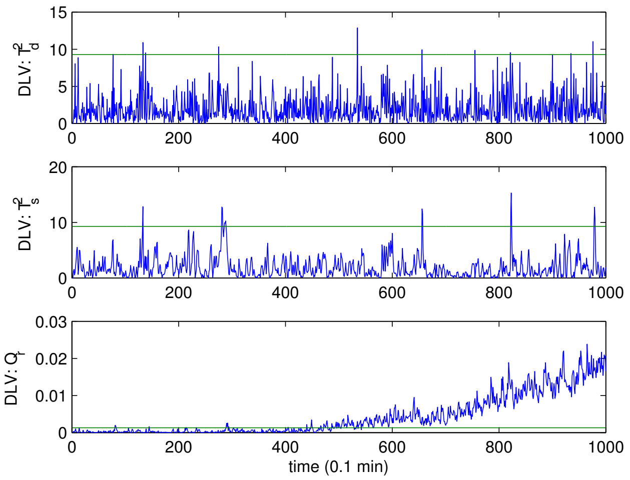

Fig. 3. Fault detection indices of the DLV model for Fault 2

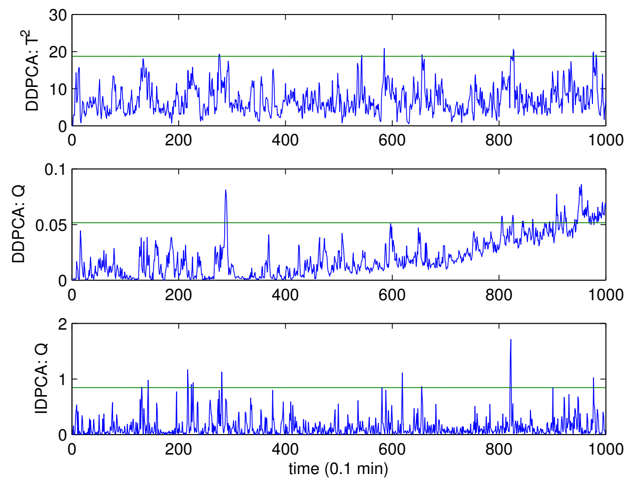

Fig. 4. Fault detection indices of DDPCA and IDPCA models for Fault 2

REFERENCES

Bersimis, S., Psarakis, S., and Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23(5), 517–543.

De Waele, S. and Broersen, P. (2003). Order selection for vector autoregressive models. *IEEE Transactions on Signal Processing*, 51(2), 427–433.

Ding, S., Zhang, P., Naik, A., Ding, E., and Huang, B. (2009). Subspace method aided data-driven design of fault detection and isolation systems. *Journal of Process Control*, 19(9), 1496–1510.

Hóskuldsson, A. (1988). PLS regression methods. *Journal of Chemometrics*, 2, 211–228.

Kresta, J.V., Macgregor, J.F., and Marlin, T.E. (1991). Multivariate statistical monitoring of process operating performance. *Canadian Journal of Chemical Engineering*, 69(1), 35–47.

Ku, W., Storer, R., and Georgakis, C. (1995). Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and intelligent laboratory systems*, 30(1), 179–196.

Li, G., Qin, S., Ji, Y., and Zhou, D. (2010). Reconstruction based fault prognosis for continuous processes. *Control Engineering Practice*, 18, 1211–1219.

Li, W. and Qin, S. (2001). Consistent dynamic PCA based on errors-in-variables subspace identification. *Journal of Process Control*, 11(6), 661–678.

Ljung, L. (1999). System identification: Theory for the user. *PTR Prentice Hall, Upper Saddle River, NJ.*

Lu, N., Yao, Y., Gao, F., and Wang, F. (2005). Two-dimensional dynamic PCA for batch process monitoring. *AIChE Journal*, 51(12), 3300–3304.

Macgregor, J.F. and Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403–414.

Negiz, A. and Çinar, A. (1997). Statistical monitoring of multivariable dynamic processes with state-space models. *AIChE Journal*, 43(8), 2002–2020.

Negiz, A. and Çinar, A. (1998). Monitoring of multivariable dynamic processes and sensor auditing. *Journal of Process Control*, 8(5-6), 375–380.

Negiz, A., Lagergren, E., and Çinar, A. (1994). Statistical quality control of multivariable continuous processes. In *American Control Conference,29 June-1 July 1994*, 2, 1289–1293.

Qin, S.J. (2003). Statistical process monitoring: basics and beyond. *Journal of Chemometrics*, 17(8-9), 480–502.

Qin, S. (2006). An overview of subspace identification. *Computers & Chemical Engineering*, 30(10-12), 1502–1513.

Qin, S. and McAvoy, T. (1996). Nonlinear FIR modeling via a neural net PLS approach. *Computers & Chemical Engineering*, 20(2), 147–159.

Simoglou, A., Martin, E., and Morris, A. (2002). Statistical performance monitoring of dynamic multivariate processes using state space modelling. *Computers & Chemical Engineering*, 26(6), 909–920.

Valle, S., Li, W., and Qin, S. (1999). Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods. *Ind. Eng. Chem. Res*, 38(11), 4389–4401.

Wang, J. and Qin, S. (2002). A new subspace identification approach based on principal component analysis. *Journal of Process Control*, 12(8), 841–855.

Yoon, S. and MacGregor, J.F. (2001). Fault diagnosis with multivariate statistical models Part I: using steady state fault signatures. *Journal of Process Control*, 11(4), 387–400.