A New Method of Dynamic Latent Variable Modeling for Process Monitoring
Gang Li, S. Joe Qin, Fellow, IEEE, and Donghua Zhou, Senior Member, IEEE

Abstract—Dynamic principal component analysis (DPCA) is widely used in the monitoring of dynamic multivariate processes. In traditional DPCA where a time window is used, the dynamic relations among process variables are implicit and difficult to interpret in terms of variables. To extract explicit latent variables that are dynamically correlated, a dynamic latent variable model is proposed in this paper. The new structure can improve the modeling and the interpretation of dynamic processes and enhance the performance of monitoring. Fault detection strategies are developed and contribution analysis is available for the proposed model. The case study on Tennessee Eastman process is used to illustrate the effectiveness of the proposed methods.

Index Terms—Contribution plots, dynamic latent variable model, dynamic principal component analysis, process monitoring and fault diagnosis, subspace identification method

I. INTRODUCTION

As the complexity of industrial processes grows quickly, process monitoring problem has become a popular topic in the area of process control [1], [2]. Among different process models, multivariate statistical process monitoring provides a data-driven framework for monitoring the industrial process without accurate physical models, which is convenient for implementation. Principal component analysis (PCA) and partial least squares (PLS) are two of the most widely used models in statistical process monitoring [3]–[6]. Static PCA models are suitable for manufacturing processes where the measured variables are fairly independent over time. However, applying static PCA to dynamically correlated data tends to have unnecessarily high missed detection rates. In order to deal with autocorrelated measurements, time series models can be used to generate the residual for monitoring. However, this approach is often applicable to the process monitoring where there is little cross-correlation among variables [7]. For multivariate time series with highly cross-correlated variables, dynamic factor models are preferred for appropriate process monitoring and diagnosis.

In the last two decades, dynamic multivariate projections were developed for this objective. Ku et al. proposed a lagged versions of PCA to process multivariate variables with dynamic property [8], called dynamic PCA model. Dynamic PCA model is to conduct a singular value decomposition on an augmented data matrix which includes lagged measurements within a time window. However, there are limitations for this technique, one of which is that auto-correlations are mixed with cross-correlations, making the number of model parameters overly large. The mixing also leads to implicit dynamic relations, which are difficult to interpret.

There are also some ideas based on subspace modeling. Negiz and Cinar used a canonical variate state space model to describe dynamic processes, which is equivalent to a vector autoregressive moving-average time-series model (VARMA) [9]. They used a stochastic realization based on canonical variate analysis (CVA) to handle a large number of variables that are autocorrelated, cross-correlated and collinear. The constructed state variables are linear combinations of the past measurements which can explain future variabilities in the data. As PLS 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 [10]. The comparison between CVA and PLS shows that CVA can provide more rapid detection for the faults. However, due to the sensitivity of small eigenvalues of the covariance matrix, PLS is often able to obtain more numerically stable results than the CVA method.

The subspace identification method (SIM) identifies a state space model for the process [11]. PCA can be used to develop a consistent state space model using a parity space under errors-in-variables situations [12]. Li and Qin investigated the relationship between dynamic PCA and SIM under state space descriptions of dynamic processes [13]. With the presence of process and measurement noise, they proposed a consistent dynamic PCA algorithm, namely indirect dynamic PCA (IDPCA) with consistency conditions. Recently, Ding et al. combined SIM and model based fault detection technologies to propose observer-based fault detection scheme [14]. 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. Another dynamic probabilistic model termed linear gaussian state-space model was proposed for dynamic process monitoring by Wen et al. [15], which uses the expectation maximum (EM) algorithm for the structure

0278-0046 (c) 2013 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
selection and the Kalman filter for the innovation estimation.

However, these subspace based methods all ignore cross-correlations among variables, which are useful for process monitoring. It is desirable to extract both dynamic auto-correlations and cross-correlations among variables to build a compact or parsimonious dynamic factor model for process monitoring. In this paper, a new dynamic latent variable (DLV) model is proposed to extract auto-correlations and cross-correlations explicitly, which gives the dynamic model clear insights.

The rest of the paper is organized as follows. In Section 2, existing dynamic PCA and subspace based methods are reviewed briefly. Then a new dynamic latent variable model is proposed in Section 3. Fault detection and diagnosis schemes are developed based on the proposed model in Section 4. In Section 5, a case study on the Tennessee Eastman Process (TEP) is used to illustrate the effectiveness of the proposed methods. Conclusions are given in the last section.

II. DYNAMIC PCA AND STATE SPACE MODELING

Suppose a sample vector \( x_k \in \mathbb{R}^m \) consists of \( m \) sensor measurements at sampling time \( k \). With the effect of dynamic processes and closed loop control, measurement samples at different \( k \) are not independent, which can be both auto-correlated and cross-correlated.

A. Direct DPCA modeling

To capture the dynamic relations inside the variables, dynamic PCA performs PCA procedure on the following augmented data matrix with \( k=0 \) [8]

\[
Z_k = \begin{bmatrix}
  x_{k+q} & x_{k+q+1} & \ldots & x_{k+n+q-1} \\
  x_{k+q-1} & x_{k+q} & \ldots & x_{k+n+q-2} \\
  \vdots & \vdots & \ddots & \vdots \\
  x_{k+1} & x_{k+2} & \ldots & x_{k+n}
\end{bmatrix}^T
\]

where \( q \) is the lagged number of the data matrix, \( n \) is the number of samples. The objective of DPCA is to find the solution of the following optimization:

\[
\max_{w_z} \quad w_z^T Z_q^T Z_0 w_z \\
\text{s.t.} \quad \| w_z \| = 1
\]

The following singular value decomposition can be performed and the largest \( A \) singular directions are selected as the dynamic principal components.

\[
\frac{1}{n} Z_q^T Z_0 = U_1 D_1 V_1^T
\]

Ku et al. suggested a method to determine the order of dynamic process [8]. If measurement noise terms have identical variances for all variables, the direct DPCA method is unbiased. Let \( D = U_1(:,1 : A) \), \( \hat{D} = U_1(:, A + 1 : mq) \). Denote \( z_k = [x_k^T, \ldots, x_{k-q+1}^T]^T \), then the dynamic principal components and the residual are as follows

\[
t_k = P_{D}^T z_k \\
r_k = P_{D}^T \hat{D}^T z_k
\]

B. Indirect DPCA modeling

If the measurement noise does not have identical variances, DPCA model is biased. Li and Qin suggested to use an indirect dynamic PCA algorithm based on instrument variables to eliminate the effect of noise [13]. In their work, a SIM based indirect dynamic PCA is used for extracting the party vector \( \mathbf{P}_I \). Their method performs SVD as follows:

\[
\frac{1}{n} Z_q^T Z_0 = U_2 D_2 V_2^T
\]

and \( \mathbf{P}_I = U_2(:,1 : A) \), \( \hat{P}_I = U_2(:, A + 1 : mq) \). Li and Qin used the Akaikie information criterion for choosing the proper lagged number \( q \) [13]. The number of dynamic principal factors \( A \) is determined as the number of diagonal elements of \( D_2 \) that are zero or nearest to zero.

C. 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 \( Z_t \) and a past data set \( Z_0 \). Negiz and Çınar used the CVA technique to build a state space model for dynamic process monitoring [9]. It provided the principal directions of a linear dynamic system through the canonical variables, which are combination of the past measurements that are highly correlated to the future measurements. The objective of CVA is to search for correlation in the future and past data set,

\[
\max_{w_z, c} \quad c_z^T Z_q^T Z_0 w_z \\
\text{s.t.} \quad \| Z_q c_z \| = \| Z_0 w_z \| = 1
\]

The procedure is to perform the following SVD,

\[
(Z_q^T Z_0)^{-1/2} Z_q Z_0 (Z_q^T Z_0)^{-1/2} = U_3 D_3 V_3^T
\]

and the CV states are obtained by

\[
t_k = P_c^T R_0^{-1/2} z_k
\]

where \( P_c = V_3(:,1 : A) \) includes singular vectors with respect to \( A \) largest singular value, and \( R_0 = \frac{1}{n} Z_0^T Z_0 \) means the covariance matrix of \( z_0 \). The order \( A \) can be determined by the AIC [16].

Partial least squares (PLS) is conceptually similar to CVA in building a relationship between future and past data set [10], which has the following objective,

\[
\max_{w_z, c} \quad c_z^T Z_q^T Z_0 w_z \\
\text{s.t.} \quad \| c_z \| = \| w_z \| = 1
\]

and the PLS scores are obtained by

\[
t_k = (W^T P)^{-1} W^T z_k
\]

where \( W, P \) are the weighting and loading matrices from the PLS algorithm [17].
III. DYNAMIC LATENT VARIABLE MODELING

The aforementioned dynamic process models extract only the dynamic relations among variables, and use them for residual generation. However, there are two problems. On one hand, there exist static and dynamic correlations, which should be both extracted. The existing dynamic modeling methods do not distinguish dynamic relations from static relations. On the other hand, the augmentation of many lagged variables in an extended vector does not retain the structure of the original variable space. In order to retain the variable space in a dynamic model, a structured dynamic PCA algorithm is proposed to extract dynamic latent variables from the original data space in this section. Furthermore, a dynamic latent variable model is used to characterize the dynamics as well as static cross-correlations in process variables.

A. Structured dynamic PCA algorithm

Let \( X_k = [x_{q-k}, x_{q-k+1}, \ldots, x_{q-k+n-1}]^T \) \((k = 0, \ldots, q - 1)\) represent the process data with \( k \) time lags, where \( q \) is the maximum of lagged number and \( n \) is the number of samples used for training the model. The proposed algorithm maximizes the following objective:

\[
\max_{\mathbf{w}, \beta} \mathbf{w}^T (\beta_0 \mathbf{X}_0^T + \cdots + \beta_{q-1} \mathbf{X}_{q-1}^T) (\beta_0 \mathbf{X}_0 + \cdots + \beta_{q-1} \mathbf{X}_{q-1}) \mathbf{w}
\]

\[
\text{s.t.} \\
\|\mathbf{w}\| = 1, \|\beta\| = 1
\]

(12)

where \( \beta = [\beta_0, \ldots, \beta_{q-1}]^T \). \( \beta_k \) is the weight coefficient for \( X_k \). This objective represents the chief dynamic variation from process variable space, by searching the weight vector \( \mathbf{w} \) and coefficient vector \( \beta \). This dynamic extension does not increase the dimension of \( \mathbf{w} \) and characterizes dynamic correlations in a compact latent structure. Notice that \( \mathbf{Z}_0 = [\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_{q-1}] \), the above objective can be rewritten as

\[
\max_{\mathbf{w}, \beta} \mathbf{w}^T (\mathbf{\beta} \otimes \mathbf{w})^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{\beta} \otimes \mathbf{w})
\]

\[
\text{s.t.} \\
\|\mathbf{w}\| = 1, \|\beta\| = 1
\]

(13)

where \( \mathbf{\beta} \otimes \mathbf{w} = [\beta_0 \mathbf{w}^T, \beta_1 \mathbf{w}^T, \ldots, \beta_{q-1} \mathbf{w}^T]^T \) is the Kronecker product.

Remark 1: The above objective is reduced to static PCA when \( q = 1 \), and the result is the same as the static PCA algorithm. If the number of variables \( m = 1 \), the objective is to find the maximum variance of the linear combination of lagged variables. Comparing optimization in Eq. (2) and Eq. (13), it can be found that the objects are the same except for the restrictions. The optimization (2) has no restriction for the form of \( \mathbf{w}_z \), while optimization (13) uses a predetermined structure of \( \mathbf{w}_z = \mathbf{\beta} \otimes \mathbf{w} \), which also obeys the norm restriction in (2).

To maximize the structured dynamic PCA objective, Lagrange multipliers are used

\[
J = (\mathbf{\beta} \otimes \mathbf{w})^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{\beta} \otimes \mathbf{w}) + \lambda_w (1 - \mathbf{w}^T \mathbf{w}) + \lambda_\beta (1 - \mathbf{\beta}^T \mathbf{\beta})
\]

(14)

From the properties of Kronecker products, we have the following equations:

\[
\mathbf{\beta} \otimes \mathbf{w} = (\mathbf{I}_q \otimes \mathbf{w}) \mathbf{\beta} = (\mathbf{\beta} \otimes \mathbf{I}_m) \mathbf{w}
\]

(15)

Taking derivatives with respective to \( \mathbf{w} \) and \( \mathbf{\beta} \), leads to

\[
\frac{\partial J}{\partial \mathbf{\beta}} = 2[(\mathbf{I}_q \otimes \mathbf{w})^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{I}_q \otimes \mathbf{w})] \mathbf{\beta} - \lambda_\beta \mathbf{\beta} = 0
\]

(16)

\[
\frac{\partial J}{\partial \mathbf{w}} = 2[(\mathbf{\beta} \otimes \mathbf{I}_m)^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{\beta} \otimes \mathbf{I}_m) \mathbf{w} - \lambda_w \mathbf{w}] = 0
\]

(17)

Define

\[
\mathbf{S}_w \equiv \mathbf{Z}_0^T \mathbf{Z}_0 \in \mathbb{R}^{mq \times mq}
\]

\[
\mathbf{S}_w \equiv (\mathbf{I}_q \otimes \mathbf{w})^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{I}_q \otimes \mathbf{w}) \in \mathbb{R}^{q \times q}
\]

\[
\mathbf{S}_\beta \equiv (\mathbf{\beta} \otimes \mathbf{I}_m)^T \mathbf{Z}_0^T \mathbf{Z}_0 (\mathbf{\beta} \otimes \mathbf{I}_m) \in \mathbb{R}^{m \times m}
\]

The above necessary conditions can be summarized as

\[
\mathbf{S}_w \mathbf{\beta} = \lambda_\beta \mathbf{\beta}
\]

\[
\mathbf{S}_\beta \mathbf{w} = \lambda_w \mathbf{w}
\]

(19)

According to Eq. (13), multiplying \( \mathbf{\beta}^T \) and \( \mathbf{w}^T \) to both sides in Eq. (19), the following result can be obtained when the objective reaches the maxima at optimal solution \( \mathbf{\beta}_{\text{opt}} \) and \( \mathbf{w}_{\text{opt}} \):

\[
J_{\text{max}} = \lambda_w = \lambda_\beta
\]

(20)

However, the optimal solution cannot be calculated directly based on Eq. (19), because two eigenvalue decompositions are coupled with each other. In order to determine the optimal solution, an iterative algorithm is proposed to search \( \mathbf{w} \) and \( \mathbf{\beta} \) by setting an initial value of \( \mathbf{w} \). Then the dynamic latent score vector is calculated as

\[
\mathbf{t} = \mathbf{X} \mathbf{w}
\]

(21)

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

\[
\mathbf{p} = \mathbf{X}^T \mathbf{t} / \mathbf{t}^T \mathbf{t}
\]

(22)

which generates the residual by deflating the data with the extracted dynamic latent scores,

\[
\mathbf{E} = \mathbf{X} - \mathbf{t} \mathbf{p}^T
\]

(23)

To sum up, the structured dynamic PCA procedure is listed as follows:

Algorithm 1 (Structured Dynamic PCA algorithm):

Center the raw data of \( \mathbf{X} \) to zero mean and scale them to unit variance, and set \( t = 1, \mathbf{X} = \mathbf{X} \).

1) Set \( \mathbf{w}_i = [1, 0, \ldots, 0]^T \), calculate \( \mathbf{S}_w \), and solve the following eigenvalue decomposition, and let \( \beta_i \) be the eigenvector corresponding to the largest eigenvalue.

\[
\mathbf{S}_w \beta_i = \lambda_i \beta_i
\]

(24)

2) Calculate \( \mathbf{S}_\beta \), and solve the following problem, and let \( \mathbf{w}_i \) be the eigenvector corresponding to the largest eigenvalue.

\[
\mathbf{S}_\beta \mathbf{w}_i = \lambda_w \mathbf{w}_i
\]

(25)

Return to Step 1 until \( \mathbf{w}_i \) converges.

3) \( \mathbf{t}_i = \mathbf{X}^i \mathbf{w}_i \)

4) \( \mathbf{p}_i = \mathbf{X}^i \mathbf{t}_i / \mathbf{t}_i^T \mathbf{t}_i \)

5) \( \mathbf{X}^{i+1} = \mathbf{X}^i - \mathbf{t}_i \mathbf{p}_i^T, \mathbf{Z}_0^{i+1} = [\mathbf{X}_0^{i+1}, \mathbf{X}_1^{i+1}, \ldots, \mathbf{X}_{q-1}^{i+1}] \),

\[
\mathbf{S}_z = \mathbf{Z}_0^{i+1} \mathbf{Z}_0^{i+1}
\]

Return to Step 1 until \( i > A \).
The convergence of the structured dynamic PCA algorithm is given in Appendix A. As the above algorithm resembles the
partial least squares algorithm in the procedure, the structure on X space provided by the structured dynamic PCA is
also similar to PLS. Denote \( W = [w_1, w_2, \ldots, w_m] \), \( P = [p_1, p_2, \ldots, p_A] \), \( T = [t_1, t_2, \ldots, t_A] \), \( R = W(P^TW)^{-1} \), then
\[
T = XR
\]
(26)

Given a sample of process vector \( x_k \), the score and residual are calculated as
\[
\begin{align*}
\mathbf{t}_k &= R^T x_k \\
\mathbf{e}_k &= (I - PR^T)x_k
\end{align*}
\]
(27)

B. Dynamic modeling of the latent scores

As the latent variables contain most dynamics in the data, it is necessary to build a dynamic model to characterize the auto-
correlation inside the latent variables. Assuming processes under the normal operation are stationary, it is reasonable to
describe \( t_k \) with a vector autoregressive model as follows:
\[
t_k = \sum_{j=1}^{p} \alpha_j t_{k-j} + \nu_k = \Theta^T \varphi_k + \nu_k
\]
(28)
where \( \nu_k \) 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 \( \Theta = [\alpha_1, \ldots, \alpha_p]^T \), \( \varphi_k = [t_{k-1}^T, \ldots, t_{k-p}^T]^T \), and \( p \) is the model order. The
parameters of above model can be estimated by a multivariable least squares algorithm directly as follows [18].
\[
\hat{\Theta} = \left( \sum_{i=p+1}^{p+n} \varphi_i \varphi_i^T \right)^{-1} \left( \sum_{i=p+1}^{p+n} \varphi_i t_i^T \right)
\]
(29)

C. Further modeling after extracting dynamics

After the extraction of dynamic factors, \( e_r \) has very little auto-covariance, which is desirable for further extraction of
static cross correlations. Denoting \( E = [e_1, \ldots, e_n]^T \), \( E \) represents the time-independent variations in the process data
after extracting the dynamic latent factors. This leaves \( E \) with mostly static cross correlations as the dynamic correlations
have been extracted. Thus it is necessary to decompose \( E \) further by the static PCA. The whole procedure of DLV
modeling can be summarized as follows.

1) Determine the lagged number \( q \) and the number of dynamic factors \( A \). Use the structured dynamic PCA algorithm to extract dynamic latent variables, represented by Eq. (27).

2) Determine the model order \( p \), use the multivariable LS algorithm to build the inner model of dynamic latent variables.

3) Determine the static principal components number \( A_s \). Perform PCA on residual \( E \) as \( E = T_s P_s + E_r \), where
\( T_s \) contains \( A_s \) components.

The ultimate space decomposition of \( X \) space can be represented as:
\[
X = TP^T + T_s P_s + E_r
\]
(30)

And for a sample vector \( x_k \), the decomposition can be described as
\[
\begin{align*}
x_k &= PT_k + P_t s_k + e_r \\
t_k &= \sum_{j=1}^{p} \alpha_j t_{k-j} + \nu_k
\end{align*}
\]
(31)

We refer to Eq. (31) 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. The DLV model captures the dynamic variation
first, then extracts dominant static variations. As a result, the DLV model characterizes dynamic and static relationships
in the process data explicitly, rather than implicitly as in traditional DPCA. Note that auto-correlations are first extracted before cross correlations are extracted. The geometric interpretation of DLV models is analyzed as follows.

**Lemma 1:** The DLV model induces an oblique decomposition of the \( X \)-space:
\[
\begin{align*}
X &= \tilde{X} + \tilde{X}_s + e_r \\
\tilde{X} &= PR^T X \in S_d \\
\tilde{X}_s &= P_s P_s^T (I - PR^T)X \in S_s \\
e_r &= (I - P_s P_s^T) (I - PR^T)X \in S_r
\end{align*}
\]
(32)

**Lemma 1** is easily proven and is omitted due to page limitation. A similar proof procedure can be found in [19]. The
meaning of different subspaces are summarized in Table I.

<table>
<thead>
<tr>
<th>subspace</th>
<th>dimension</th>
<th>description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_d )</td>
<td>( A )</td>
<td>subspace of ( X )-space that has dynamics, or strong auto-correlations.</td>
</tr>
<tr>
<td>( S_s )</td>
<td>( A_s )</td>
<td>subspace of ( X )-space that is static, or time independent.</td>
</tr>
<tr>
<td>( S_r )</td>
<td>( m - A - A_s )</td>
<td>subspace of ( X )-space that is not excited in the data or reflects the noise</td>
</tr>
</tbody>
</table>

D. Determination of model parameters

In the DLV model, there are some parameters to be decided. First of all, the lagged number \( q \) in Eq. (12) should be

determined before structured DPCA can be performed. The method used here was first suggested by Ku [8]. Considering
that there are \( r_0 \) static relations in \( m \) original variables \( X_0 \), there must be \( 2r_0 \) static relations and \( r_1 \) dynamic relations in \( 2m \) augmented data \( [X_0, X_1] \), the rest can be done in the same manner. If for \( (q + 1)m \) augmented data, there are not new relations any more, i.e. \( r_q = 0 \), then it can be seen that \( q \) is enough for covering all dynamic correlations. If there is no noise, both static and dynamic relations can be easily
determined as the right null space of \( Z_0 \). However, there is always noise in practice. Thus, the relations can be identified by the smallest singular values of \( Z_0 \), which are near zero.

Similarly, when a dynamic factor is extracted, the number of dynamic relations will be reduced. After \( A \) dynamic
factors are extracted, the dynamic relations should be removed totally, which reflects that there are only static relations in
\( Z_0^{A+1} \). When all dynamic factors are extracted, a vector autoregressive model is adopted to describe the dynamics. And
the model order $p$ is determined by the Akaike information criterion (AIC) [18].

Finally the static PCA model is built on the residual after the structured dynamic PCA. The number of components in the static PCA model is determined by cumulative percentage of variance (CPV) [20].

IV. FAULT DETECTION AND DIAGNOSIS BASED ON DLV

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

\[
\begin{align*}
    t_k &= R^T x_k \\
    v_k &= t_k - \sum_{j=1}^{p} \alpha_j t_{k-j} \\
    e_{r,k} &= (I - P_{s}P_{s}^T)(I - PR)^T x_k \\
    t_{s,k} &= P_{s}^T (I - PR)^T x_k
\end{align*}
\]

where $t_k$ and $v_k$ represent the score and residual in dynamic process variations, and $t_{s,k}$ and $e_{r,k}$ represent the score and residual in static process variations. Note that $t_k$ is not time-independent series, which is not proper for monitoring. Therefore, three indices are constructed to monitor the process with DLV model, which are listed in the Table II. There are explicit meanings of these statistics. $T^2_d$ monitors the innovation process of dynamic factors, which is statistically independent of historical dynamic variation. On the other hand, $T^2_s$ reflects the variation of static principal components, while $Q_r$ measures the noise level and modeling error, including unmodeled dynamics.

As $T^2_d$ and $Q_r$ have a quadratic form of $x_k$, which can lead to contribution analysis directly. Motivated by [21], reconstruction based contribution (RBC) for variable $j$ and sample $x_k$ is defined as

\[
    \text{RBC}_{j,k} = \frac{(\xi_j^T \Phi x_k)^2}{\xi_j^T \Phi \xi_j}
\]

where $\xi_j$ is the $j^{th}$ column of the identity matrix $I_m$ and $\Phi = (I - PR)^T(P_{s}A_{s}^{-1}P_{s}^T)(I - PR)^T$ for $T^2_d$, $\Phi = (I - PR)^T(I - P_{s}P_{s}^T)(I - PR)^T$ for $Q_r$.

V. CASE STUDY ON THE TE PROCESS

In this section the effectiveness of the proposed method for process monitoring and fault diagnosis is investigated by case study on the Tennessee Eastman (TE) process. TEP was created to provide a benchmark for evaluating process control and monitoring approaches, including PCA, PLS, and Fisher discriminant analysis (FDA) [22]. Yin et al. compared basic data-driven fault diagnosis techniques based on the benchmark of TE process, including PCA, DPCA, FDA, PLS, T-PLS, MPLS, ICA, and SAP (subspace aided approach) [23]. TEP contains 12 manipulated variables and 41 measured variables. In this study, 22 process measurements and 11 manipulated variables, i.e. XMEAS(1-22) and XMV(1-11), are chosen as $X$. Other measured variables are concentrations of different components, which are measured with a delay and can be seen as quality variables. There are 15 known faults in TEP, which includes seven step faults, five random faults and three sticking and slow change faults, as listed in the Table IV.

Firstly, 480 normal samples are centered to zero mean and scaled to unit variance. Then these normal samples are used to determine the parameters and build the data model. Parallel analysis is used to determine the model parameters $q$. When lagged number is 0, there are 20 relations which are all static. When the lagged number increases to 1, the number of new relations increase to 25, where 5 new dynamic relations are found in $X_g$ and other 20 relations are still static as the same as in the previous case. When the lagged number grows to 2, the new relations do not increase anymore and stabilize around 25, which means there are only 5 added dynamic relations and the lagged number of 1 is enough for describing the dynamic of TE process. Then, the structured dynamic PCA algorithm is built with $q = 2$. Then, parallel analysis is used again for choosing the proper $A$. After six dynamic factors are extracted from the original data, the relations do not increase for the augmented matrix. Thus, $A = 6$ is enough for describing the dynamic for TEP. Table III lists the iteration error and optimization result of $J_k$ for each step in iterations 1-6. From the table, it can be seen that the iteration process converges fast to the optimal value, and iteration steps are not more than 3 in this case. The optimal $J_{max}$ for each iteration is always below the upper bound $\lambda_{max}(S_z)$, which is consistent with the analysis in the Appendix. The results show that the computational cost of structured DPCA algorithm is considerably acceptable even for industrial implementation. Following that, the order of the vector auto regressive model is selected as $p = 4$ using AIC and $A_s = 16$ is decided for static PCA according to CPV.

After modeling, the correlation plots are plotted to check the availability of the proposed dynamic latent variable model. Figs. 1 - 3 depict the autocorrelations of the first six components of dynamic factors ($t$), innovations($v$), and static factors($t_s$), respectively. The results reflect that there are evident autocorrelation and cross correlation in the dynamic factors. And in the innovation part $v$, the autocorrelation and cross correlation are both removed. Meanwhile, there is no autocorrelation in static component $t_s$, while its cross correlation is extracted by a PCA model.

To monitor the process with DLV model, three statistics indices are developed as described in Section 4. Based on the test data of normal case (IDV 0) and faulty data with respect to known faults (IDV 1-15), the fault detection rates...
model, the number of principal components is selected as 19 according to co percent variance test including about 95% variation information. In the DPCA model, $q = 2$ and $A = 35$ are selected which contains 95% variation. In the SIM based method, $q = 2$ and $A = 16$ as suggested in [23]. In CVA method, $q = 2$ and $A = 29$ as suggested in [16]. For multiple indices in a method, if any statistic exceeds its control limit, the fault is detected. From the FDRs for all faults (except IDV 3,9,15), it can be seen that DLV based method can detect faults more stably than DPCA and outperforms PCA and SIM based methods in most cases. Although CVA based method has higher FDR, its FAR is also high, which is caused by the inversion of the small values of $R_0$ in Eq. (8).

However, DLV model is more convenient for the interpretation of the abnormal situations in the original space. Take IDV(10) as an example, Fig. 4 shows the detection result using DLV based indices. When this fault occurs, a random change is added in the temperature of C feed, which is compensated by feedback control. Fig. 5 plots the RBC$_j$ to $Q_r$ at the 400th sample, where the horizontal axis represents the variable index, and the vertical axis is the contribution. The variables corresponding to high contributions are significant candidates responsible for the fault. It is clear to find that variable 18 and 19 have the highest contributions, which are the measurements of the stripper temperature and the stripper steam flow, respectively. Further, it can be concluded that variable 18, 19 are most responsible for the fault. In order to confirm the fact, we plot the RBC for all samples after the fault is detected in Fig. 6, where horizon axis represents the sample index, and vertical axis is variable index. The RBC for each sample and each variable is drawn by a color tape, where the darker color represents the higher RBC. From the plot, it is observed that Variables 18 and 19 are indeed the most significant candidates for this fault, which is consistent with the conclusions in [15], [24]. However, traditional dynamic methods such as DPCA, SIM and CVA do not reveal a structure in the lagged variables, which makes the diagnosis analysis difficult to implement.

**Fig. 3.** Autocorrelation coefficients of static variables ($t_s$)

**Fig. 1.** Autocorrelation coefficients of dynamic latent variables ($\gamma$)

**Fig. 2.** Autocorrelation coefficients of innovation variables ($v$)

(FDR) and false alarm rates (FAR) with DLV model and four existing methods are compared in the Table IV. In the PCA

<table>
<thead>
<tr>
<th>dynamic factors No.</th>
<th>iteration steps</th>
<th>$J_1$ (error)</th>
<th>$J_2$ (error)</th>
<th>$J_3$ (error)</th>
<th>$\lambda_{max}(S_z)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1.3223</td>
<td>10.5580</td>
<td>10.5580</td>
<td>10.5637</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.3114</td>
<td>5.8093</td>
<td>(0.0000)</td>
<td>5.8203</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.2707</td>
<td>3.8436</td>
<td>3.8436</td>
<td>3.8823</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>1.2298</td>
<td>3.2170</td>
<td>3.2170</td>
<td>3.2540</td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>0.5014</td>
<td>2.3619</td>
<td>2.3619</td>
<td>2.5044</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>0.4957</td>
<td>2.3092</td>
<td>2.3092</td>
<td>2.4913</td>
</tr>
</tbody>
</table>

$J_1$ is the optimization object in (14) for r-th step, (error) means the normal error of $w$ between Eq. (24) and Eq. (25) for each step with the threshold $1e-5$, iteration steps is the total No. of iteration steps, $\lambda_{max}(S_z)$ is the largest eigenvalue of $S_z$ in (18).

**Fig. 4.** Detection result using DLV based indices.
Moreover, the decomposition structure can be used for further fault estimation and reconstruction.

Fig. 4. Fault detection result with $T_2^d$, $T_2^s$ and $Q_r$

VI. CONCLUSIONS

In this paper, a new dynamic latent variable model is proposed for data-driven modeling of dynamic processes. A structured dynamic PCA algorithm is proposed to extract dynamic latent variables, which cover the most of dynamic variations. This procedure separates dynamic autocorrelations from static cross correlations of process variables. For dynamic latent variables, a vector AR model is adopted to construct an innovation-like residual. In addition, a PCA decomposition is applied to extract the static variation from the remaining part.

Several statistics are developed based on the DLV model. $T_2^d$ and $T_2^s$ are used to monitor the dynamic and static variations, respectively, and $Q_r$ is used for detecting the whole residual of the process. The DLV model has a clear interpretation in variable space, which can be used for diagnosis with contribution analysis directly. The case study on TEP shows clear advantages of the proposed modeling and related methods.

Fig. 5. RBC to $Q_r$ at 400th sample

Fig. 6. RBC to $Q_r$ during the whole process. Variables with high contribution are dark in the color.

APPENDIX

A. Convergence of the structured DPCA algorithm
Denote $J_k = J(w^k, \beta^k)$ as the optimization objective during the $k$ iteration of Step 1 and 2 in structured dynamic PCA algorithm. First of all, observing the optimization objective in Eq. (2), it can be concluded that a solution of Eq. (13) is also a feasible solution of Eq. (2), as $\| (\beta \otimes w^m) \|^2_w = \sum_{j=0}^{q-1} \beta_j^2 (\sum_{i=1}^{m} w_{j i}^2)^2 = 1$. According to Eq. (2),

$$0 < J_k \leq \max_{w} \left( w^T S_z w \right) = \lambda_{\max}(S_z) \quad (k = 1, 2, \ldots)$$

which means $J_k$ is a monotone increasing sequence. Therefore, $J_k$ must be convergent and hence the algorithm.

REFERENCES


Gang Li received his bachelor degree in mechanics from Tsinghua University, China. He obtained his B.S. and M.S. degrees in Automatic Control from Tsinghua University in 1998 and 1990, respectively. He is currently a professor and the head of the Department of Electrical Engineering. His research interests include fault diagnosis and condition monitoring, model predictive control, model predictive control, predictive maintenance. He is a member of IFAC Technical Committee on Fault Diagnosis and Safety of Technical Processes, a senior member of IEEE, and an associate editor of Journal of Process Control, the associate Chairman of Chinese Association of Automation (CAA), and a council member of CAA.